ABSTRACT

Title of dissertation: Co-design for Multi-subsystem and Vehicle Routing-and-Control Problems

Tianchen Liu
Doctor of Philosophy, 2020

Dissertation directed by: Shapour Azarm, Advisor
Nikhil Chopra, Co-advisor
Department of Mechanical Engineering

Co-design refers to the process of integrating optimization of the physical design with a controller for a system. The challenge in co-design is that the optimization is simultaneously applied to both static/time-invariant (e.g., physical plant design) variables and dynamic/time-variant (e.g., state and control) variables, which can be coupled with each other.

The objective of this dissertation is to explore new formulations and approaches in co-design for multi-subsystem and vehicle routing-and-control problems. Specifically, four research questions are considered and resolved. In Research Question 1 (RQ1), the critical issue is how to formulate a class of multi-subsystem co-design problems with convex physical design subproblems and linear quadratic regulator control subproblems, and construct a decentralized solution approach for such problems. In Research Question 2 (RQ2), solution methods for a broader class

of multi-subsystem co-design problems than those considered in RQ1 are investigated. In Research Question 3 (RQ3), the question is whether, in the context of co-design, the combined routing and control costs of a fleet of vehicles can be improved if optimal control is introduced into the routing. Finally, an extension of RQ3 is considered in Research Question 4 (RQ4), where the possibility of constructing an integrated vehicle routing-and-control problem with load-dependent dynamics is investigated.

Beyond the articles published by the author of this dissertation, the proposed research questions, models and methods presented have not been considered elsewhere in the literature.

Co-design for Multi-subsystem and Vehicle Routing-and-Control Problems

by

Tianchen Liu

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2020

Advisory Committee:
Professor Shapour Azarm, Chair/Advisor
Professor Nikhil Chopra, Co-advisor
Professor Balakumar Balachandran
Professor Jin-Oh Hahn
Professor Nuno Martins, Dean's Representative
Professor Miao Yu

Acknowledgments

I am deeply grateful to my advisors, Dr. Shapour Azarm and Dr. Nikhil Chopra, for their continuous support, patient guidance and helpful suggestions throughout the time of my dissertation research. Their concern about the intellectual development of their students always inspired me to improve myself.

During my study at the University of Maryland, I benefited a lot from the courses I took and all activities I participated in. I would like to thank all the professors who made this journey possible for me.

In addition, I would like to express my gratitude to all my peers (Prasad Chanekar, Eliot Rudnick-Cohen, Randall Kania, Abu Kebbie-Anthony, Xiangxue Zhao, Ruchir Patel, Yimeng Dong and Kevin Quigley), especially for many helpful discussions and resource sharing.

Lastly, I thank my family for their unconditional support and love.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

In this chapter, the motivation and objective for this dissertation are provided. An overview of the research questions is given, and related literature is reviewed.

1.1 Motivation and Objective

The integrated optimization of a physical system and its controller is referred to as 'co-design'. Co-design aims to incorporate optimal features of both the physical plant design domain and control domain while taking into account the constraints arising in the system. Also, a system may be composed of multiple subsystems, each of which may include a co-design problem. Co-design exists and can be considered in many real-world applications. For instance, physical design and control of an automobile system problem consists of many co-design subsystem subproblems, such as engine, chassis, gearbox, steering wheel, and many others.

In this dissertation, centralized and decentralized optimization approaches for co-design problems are considered. A centralized approach means that all subproblems or subsystems are solved simultaneously, while a decentralized approach refers to the process that subsystems or smaller subproblems are solved individually and then their solutions are coordinated in order to obtain an overall solution. In some

situations, like routing-and-control for a fleet of vehicles, co-design may be solved in a centralized way. In other situations, it might be necessary or preferred to solve the co-design problem in a decentralized manner.

1.1.1 Objective

The overall objective of this dissertation is to explore new formulations and approaches in co-design for multi-subsystem and vehicle routing-and-control problems.

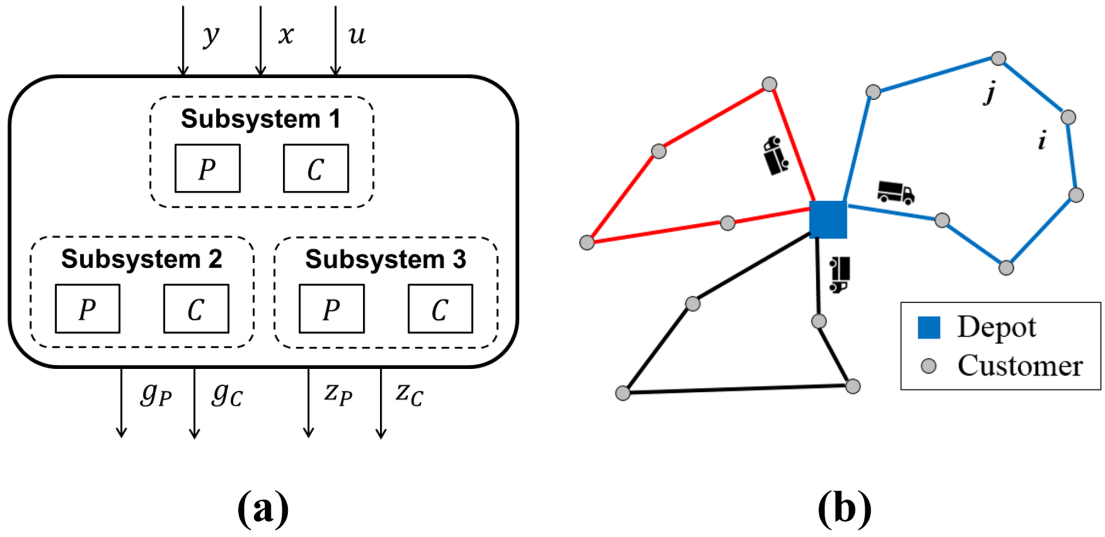


Figure 1.1: Co-design Problems: (a) Multi-subsystem, (b) Vehicle Routing-and-control

Fig. 1.1(a) shows a general multi-subsystem co-design problem in the optimization of both physical plant design (block 'P') and control (block 'C'). For such a problem, optimal solutions of input physical design and control variables for all subsystems are desired, while couplings can exist within and among different subsystems. As shown in Fig. 1.1(a), y, x, and u indicate plant design, state, and control

variables, respectively. For the plant design and control subproblems, the objective functions and the constraints are denoted as z_P and z_C, and g_P and g_C, respectively. This concept is investigated in Research Questions 1 and 2, and the outcomes include: 1) a multi-level decentralized solution approach using the indirect method, 2) two decentralized approaches using direct collocation and decomposition-based optimization methods.

Fig. 1.1(b) shows a proposed vehicle routing-and-control problem considered in the context of co-design. In a traditional vehicle routing problem, a sequence of locations to be visited by a fleet of vehicles is determined. Here, the control costs, such as the energy consumption costs of all vehicles, are also considered as discussed in Research Questions 3 and 4. The outcomes include: 1) a comparison of three different solution strategies for vehicle routing-and-control problems, and 2) a heuristic approach for solving routing-and-control problems with load-dependent vehicle dynamics.

1.2 Research Questions: Overview

In this section, the four research questions considered in this dissertation, and the adopted approaches and developed results are briefly reviewed.

Research Question 1 (RQ1): The critical question in RQ1, addressed in Chapter 2, is how to construct a decentralized approach for solving a class of multi-subsystem co-design optimization problems. The answer to RQ1 extends the prior co-design research from a single-system to a multi-subsystem problem. In each

subsystem of the proposed co-design problem, the physical design part has a convex objective function, with convex inequality and linear equality constraints. The control part has a finite time-horizon Linear Quadratic Regulator (LQR) feedback control. A multi-level decentralized approach is proposed for solving this class of problems that can obtain optimal or near-optimal solutions. The efficacy of the proposed formulation and approach is presented and demonstrated by a numerical and an engineering example. For both instances, the solutions obtained by the decentralized approach are compared against a centralized (all-at-once) approach. Also, a scalable version of the engineering example is formulated and solved. For the scalable example, it is shown that as the size of the problem increases, the computation time for the decentralized approach increases approximately linearly. In contrast that of the centralized one rises with a higher rate.

Research Question 2 (RQ2): The objective of RQ2 addressed in Chapter 3 is to determine whether it is possible to develop approaches that can obtain an optimal solution for more general multi-subsystem co-design problems than in RQ1, including those with nonlinear dynamic constraints. Following RQ2, two decentralized (multi-level and bi-level) approaches are formulated to solve multi-subsystem co-design problems. These are based on the direct collocation and decomposition-based optimization methods. In the multi-level approach, the problem is decomposed into two bi-level optimization problems, one for the physical plant and the other for the control part. In the bi-level approach, the problem is decomposed into subsystems, with each subsystem having the optimization model for physical plant and control parts together. In both approaches, the entire time horizon is discretized

to convert the continuous optimal control problem into a finite-dimensional nonlinear program. The optimality condition decomposition method is employed to solve the converted problem in a decentralized manner. The approaches are applied to similar examples as those considered in RQ1. A scalable case is also addressed to demonstrate that using a simulated parallelization with or without communication delays, the computational time of the decentralized approaches can outperform a centralized approach as the size of the problem increases.

Research Question 3 (RQ3): The objective of RQ3, addressed in Chapter 4, is to determine whether it is possible to optimize the combined routing and control costs for a fleet of vehicles from a co-design point of view. Accordingly, three solution strategies are considered: 1) the shortest route selection followed by a constant speed model, 2) the shortest route selection followed by optimal control, and 3) the optimal control-based route selection. The joint (or combined) costs, i.e., a combination of vehicles' travel time and control cost (energy consumption), are compared following these different strategies. Using data from a benchmark vehicle routing problem, it is shown that when compared with using a constant speed model, as commonly done in prior literature, the optimized cost can be improved by incorporating optimal control strategies. Also, the effect of wind on the vehicle dynamics is considered, and optimal results of the routing problem with different wind directions are compared.

Research Question 4 (RQ4): The objective of RQ4, addressed in Chapter 5, is to explore a more challenging vehicle routing-and-control problem than that considered in RQ3. Here it is assumed that the vehicle's dynamical model is depen-

dent on the loading condition, which is determined by the previous routing decision of the vehicle. In this problem of routing-and-control for a fleet of vehicles, the total cost is assumed to be a weighted sum of the traveling time and the energy consumption cost of the fleet. A sequential approach can be employed to solve this problem, in which the routing solution is obtained first, followed by finding optimal control costs of all vehicles. However, the sequential approach may not obtain an overall optimal solution since the routing decision affects the control cost and vice versa. As a result, a UCT (Upper Confidence bounds applied to Trees)-based approach has been implemented to obtain solutions of the routing-and-control problems. Two test examples are used to demonstrate the proposed approach: 1) a benchmark vehicle routing problem from the literature revised with an embedded optimal control problem, and 2) a notional integrated routing-and-control problem formulated for a small area from the New York City street map. For both examples, it is shown that the UCT-based approach obtains an improved solution when compared with the sequential approach.

1.3 Related Literature on Research Questions

1.3.1 Co-design: From Single-system to Multi-subsystem (RQ1 and RQ2)

Existing papers in co-design mainly focus on single-system problems, namely, physical and control design subproblems are considered as parts of a single system. For these problems, the traditional sequential method is to optimize the physical

system first followed by that for the control system. However, such a method does not fully handle the coupling between physical and control design variables in the co-design problems, so it can lead to sub-optimal solutions [1]. Co-design methods have been developed and applied to engineering examples in order to optimize all variables via a single-system co-design structure [2]. For example, Allison and Nazari [3] developed a decomposition-based approach to solve single-system co-design problems. Allison et al. studied co-design of the active suspension systems [4] and dynamic sustainable energy systems [5] based on a direct transcription method. Peters et al. [6, 7] derived control proxy functions for certain types of co-design problems, so that the optimal solution of the modified sequential approach matches with the simultaneous solution. Jiang et al. [8] proposed an iterative method to solve co-design problems with nonlinear control systems. Ravichandran et al. [9] adopted a metaheuristic evolutionary algorithm to obtain the optimal physical design values and set-point control of a two-link planar manipulator for carrying different payloads. Ricardez et al. [10] investigated optimization of the parameters and input of the dynamics of a chemical process with uncertainty. Optimal co-design of many other engineering applications are reported as well, including direct current motors [11, 12], wind turbines [13], four-bar mechanism [14, 15], vehicles [16, 17], and control system [18, 19] applications. The control and formation optimization of multiple unmanned aerial vehicles (UAVs) can also be regarded as a co-design problem [20]. *In many of the reported papers, the co-design problems considered do not contain a physical design objective function when evaluating the system's performance. Furthermore, no literature was found that specifically proposed and formulated optimal*

co-design problems with a multi-subsystem architecture in both design and control parts.

On the other hand, multi-subsystem (or decentralized) optimization models and methods have been developed and utilized to solve several (physical) engineering design problems. Examples of such methods include Benders' decomposition [21] and augmented Lagrangian decomposition [22, 23]. The basic idea behind these methods is to decompose the entire problem into multiple subproblems, optimize the subproblems and coordinate the solutions among the subproblems. Decentralized approaches have also been applied to control systems. For example, Geromel and Bernussou [24] proposed a feasible direction method using a matrix gradient to find the decentralized optimal controllers. Nedic et al. [25, 26] developed a decentralized subgradient method for multi-agent system optimization. Rantzer [27] verified that the global feedback control system can be solved with guaranteed optimality using the dynamic dual decomposition method. Maasoumy et al. [28] derived and applied a hierarchical control method to a multi-room HVAC system model. One approach that has been extended in RQ1 is the interactive prediction method [29, 30]. Cohen and Joalland [31] provided proofs of convergence of this method in the application of LQR feedback control. Smith and Sage [32] developed the interaction prediction method to solve nonlinear control problems. Applications of the interactive prediction method include optimal control of robot manipulators [33] and the water distribution network [34, 35]. *In short, the literature on decentralized optimization has focused on the decomposition-based optimization models and methods for either the physical or the control design problems but not both.*

With respect to RQ2, the direct methods have been used to numerically solve optimal control problems, e.g., [36, 37]. Some popular direct methods include direct shooting method [38, 39], pseudospectral optimal control [40, 41, 42], and direct collocation method [43]. In co-design, the direct transcription method has been applied to single-system problems [4, 44, 45, 46]. The direct collocation method has been employed in many optimal control applications. For example, Geiger et al. [47] investigated the application of the direct collocation method to optimal path planning of unmanned aerial vehicles. *However, no decomposition-based optimization technique in co-design has been employed using the direct method in a decentralized manner.*

1.3.2 Co-design: Vehicle Routing-and-control (RQ3 and RQ4)

Vehicle Routing Problems (VRPs) refer to a class of problems in which a fleet of vehicles visits a set of customer locations subject to certain constraints [48, 49]. The objective function in VRPs is a cost function that is minimized by determining a routing solution for all vehicles. The vehicles start and end at a single depot or multiple depots. In the traditional VRPs, the cost function usually minimizes the travel distance or time for all vehicles. Many variants of the VRPs have been reported in the literature. For example, the Capacitated VRP (CVRP) is considered when there are capacity limits (e.g., cargo weight or space limit) for the vehicles along each route, e.g., [50, 51, 52, 53]. VRPs with time windows have been investigated to find the best routes when time requirements must be satisfied [53, 54]. Other variations

include pickup and delivery [55], multiple depots with motion constraints [56, 57], and dynamic transport networks with distributed routing [58].

In the traditional VRPs, no vehicle dynamic system is considered. However, since the energy consumption and gas emissions of vehicle are increasingly important to consider recently, vehicle dynamical model has been introduced into routing decisions. Barth et al. [59, 60] describe routing models that consider CO_2 emissions. Bektaş and Laporte [61] propose a pollution-based routing problem with and without time window constraints. The effects of the speed and load (or mass) on energy consumption and greenhouse gas emissions are also studied [62]. Time-dependent VRPs are studied in [63, 64, 65, 66, 67], where the vehicle speed is used as the decision variable to optimize the energy cost. Bi-objective optimization problems concerning both travel time and gas emissions have been investigated [64, 68, 69]. Considering the limitation of the battery energy storage of the Electric Vehicles (EVs), the routing decision of EVs has also been studied. Baum et al. [70] present the optimization of speed and battery energy consumption in route planning algorithms. The recharging schedule of the EVs and locations of charging stations are incorporated into the routing problem [71, 72, 73, 74]. *In these previous papers, the vehicle speed is assumed to be constant, which can be sub-optimal if the control costs of the vehicles, e.g., energy consumption, are also considered in the entire problem.*

In RQ4, a load-dependent VRP is proposed. A similar problem is investigated in [75] but without the consideration of the vehicle dynamical model. The proposed problem is solved by a UCT (Upper Confidence bounds applied to Trees)-based approach, which can be regarded as an implementation of the Monte Carlo

Tree Search (MCTS) algorithms. In prior literature, the MCTS and the UCT approaches have been considered in the traditional VRPs [76, 77, 78]. *In these papers, similar problems and solution approaches as in RQ4 have been proposed. However, optimization of the control (energy consumption) costs is not investigated. Hence, the previous problems are less challenging due to no consideration of the coupling of the routing-and-control subproblems as in RQ4.*

1.4 Organization of the Dissertation

Fig. 1.2 shows the organization of this dissertation, and a suggested reading order of different chapters. The dissertation begins with this chapter, which includes an introduction to the research questions and a review of the related literature. Next, in Chapters 2 and 3, multi-subsystem co-design problems are considered. In Chapter 2, a class of multi-subsystem co-design problems is investigated and a multi-level solution approach is developed. In Chapter 3, certain assumptions of the problem formulations in Chapter 2 are relaxed, and two decentralized approaches are proposed to solve these more general multi-subsystem co-design problems.

Next, in Chapter 4 and 5, vehicle routing-and-control problems are studied. In Chapter 4, three solution strategies are considered and compared to show that the energy cost can be reduced by introducing vehicle optimal control into routing. In Chapter 5, a load-dependent vehicle routing-and-control problem is proposed. Finally, highlights of the results, contributions and possible future directions of the research are presented in Chapter 6.

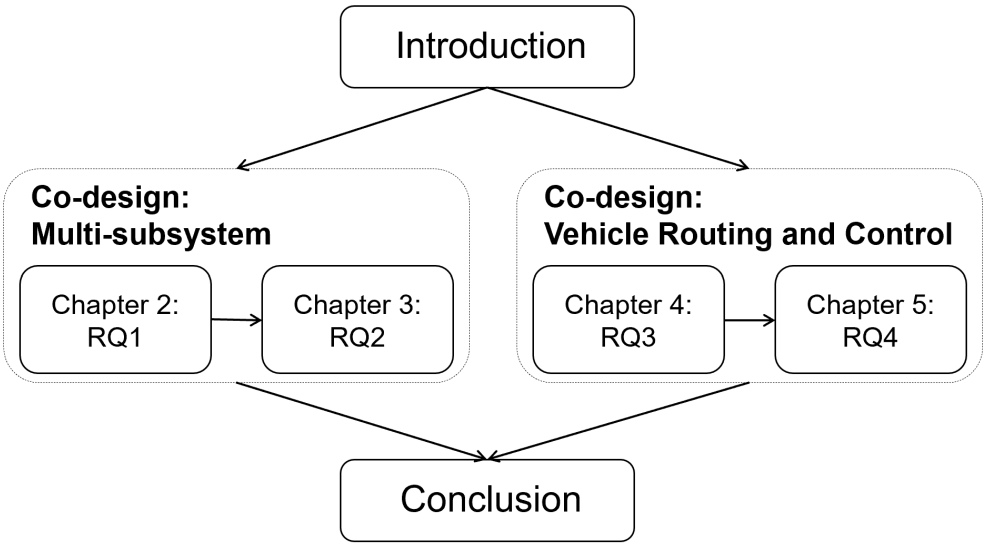


Figure 1.2: Organization of the Dissertation

The results from this dissertation research, as presented in chapters 2 to 5, are detailed in four articles. The author of this dissertation published these articles, as the lead co-author, with his co-advisors. These articles have been published or accepted in leading journals and conferences in the mechanical engineering field. These include: two full research papers in the ASME Journal of Mechanical Design (Chapters 2 and 3), one technical paper in the 2018 ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference (Chapter 4), and one full research paper in the ASME Journal of Dynamic Systems, Measurement, and Control (Chapter 5).

Chapter 2: Decentralized Optimization for a Class of Multi-subsystem Co-design Problems

In this chapter, the results from Research Question 1 are presented in detail[1].

The organization of this chapter is as follows. This chapter begins with a nomenclature section. Then, in Section 2.1, formulations and assumptions for the proposed class of multi-subsystem co-design problems are provided. A multi-level decentralized approach is proposed in Section 2.2, where the necessary optimality conditions, dual decomposition, and the solution steps are described. In Section 2.3, a numerical example and an engineering example is solved by the proposed decentralized approach, and results obtained from the decentralized approach is compared against a centralized approach. Next, a comparison of the computational times for a scalable engineering example using the decentralized and centralized approaches is presented. Finally, a summary of the chapter is given in Section 2.4.

[1]This chapter is based on the paper: Liu, T., Azarm, S., and Chopra, N., 2017, "On Decentralized Optimization for a Class of Multisubsystem Codesign Problems," ASME J Mech. Des., 139(12), p. 121404.

Nomenclature

A, B	Matrices in dynamic equations of control part
E	Consistency constraint matrix
e_C	Interaction error in the control's inner loop
e_R	Consistency constraint residual in the plant's inner loop
f	Physical inequality constraints
g	Physical equality constraints
H	Hamiltonian of the control's subproblems
h	Ancillary variables in the control's inner loop
L	Lagrangian of the physical design subproblems
N	Number of subsystems
n_c	Number of common shared physical variables
p	Co-state variables

Q, R, S	Weighting matrices in objective function in Linear Quadratic Regulator
t_f	Final time
u	Control variables
V	Set of pairs of connected subsystems
v	Coordinator variable in the control's inner loop
w_C	Weighting coefficient associated with the control objective function
w_P	Weighting coefficient associated with the physical design objective function
x	State variables
x_0	Initial conditions of state variables
y_i	Local physical design variables in i^{th} subsystem
$y_{si,j}$	Shared physical design variables in i^{th} and j^{th} subsystems

$\hat{y}_s$	Combined shared physical design variables
Z	System-level objective function
z_C	Control objective function
z_P	Design objective function
α	Coordinator variables in the control's inner loop
β	Step size in the plant's inner loop
γ, η	Lagrange multipliers of the physical design constraints
ϵ_C	Preset tolerance value of convergence for state variables
ϵ_P	Preset tolerance value of convergence for physical design variables
ϵ_R	Preset tolerance value of convergence for consistency constraint residual
λ	Ancillary co-state variables
μ	Dual variables for the physical design consistency constraints

σ Common value vector of shared physical variables

$(\cdot)_i$ Symbol with subscript i: variables/parameters in the i^{th} subsystem

2.1 Multi-subsystem Co-design Problem Formulation

Consider a multi-subsystem co-design problem with N coupled subsystems. As an example, Fig. 2.1 shows the structure of such a problem with three subsystems. The physical and control design variables are optimized in the problem. Two types of physical design variables are considered: local physical variables, y_i, which exist only in the i^{th} subsystem; and shared physical variables, $y_{si,j}$, which are shared between the i^{th} and j^{th} subsystems. The state and control variables of the i^{th} subsystem are represented by $x_i(t)$ and $u_i(t)$. When shared physical variables appear in more than two subsystems, for example, among i^{th}, j^{th} and k^{th} subsystems, consistency constraints $y_{si,j} = y_{sj,k}$ and $y_{sj,k} = y_{si,k}$ are introduced.

The interaction between the state variables of the i^{th} and j^{th} subsystems is determined by the coupling matrices A_{ij} in the dynamic equations. The set V is defined such that if A_{ij} is a non-zero matrix, then the integer pair $(i, j) \in V$; otherwise, if $A_{ij} = [0]$, then $(i, j) \notin V$.

The problem is formulated as a bi-objective optimization problem for each subsystem, which consists of one physical and one control design subproblem, and the objective function is the weighted sum of the two-part objective functions. The overall objective function of the problem is the summation of individual objective

functions of all N subsystems.

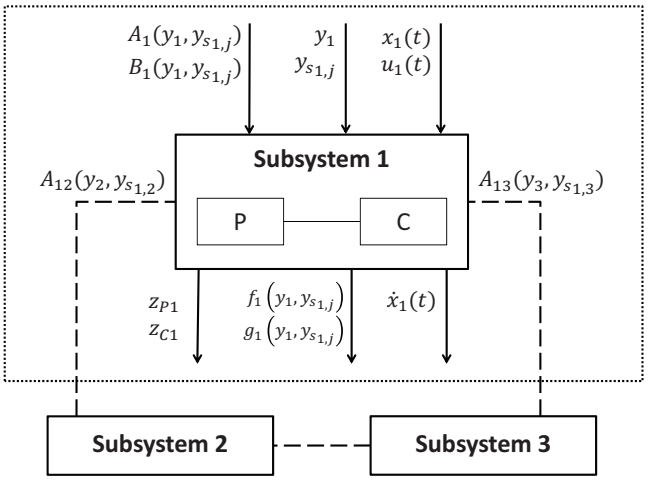


Figure 2.1: Problem Structure - Three Coupled Subsystems with Plant (P) and Control (C)

A solution to a class of co-design problems considered in RQ1 is developed under the following assumptions:

A1. The overall physical and control design objective functions are additively separable into the individual subsystem physical and control objective functions. This assumption simplifies the decomposition of the problem.

A2. The coupling is uni-directional. By uni-directional coupling it is meant that the physical design objective functions and constraints are dependent only on physical variables, and the control objective functions and constraints are dependent on physical and control variables [79].

A3. Physical design objective functions are convex, physical design inequality constraints are convex, and physical equality constraints are linear functions of

physical variables. In addition, active inequality and equality constraints are assumed to be linearly independent. This assumption is made to guarantee uniqueness of the solution in each step of the approach.

A4. The control part of each subsystem belongs to a class of continuous finite time-horizon LQR feedback control, where the time-horizon is within $[0,\ t_f]$. Cross-terms between the state and control variables of the same subsystem are considered in the control objective function, while the cross-terms between the state and control variables of different subsystems are not considered, for the sake of simplicity of the decomposition in the solution steps.

A5. The subsystems are controllable for all feasible values of the physical variables, which ensures the feasibility and existence of an optimal solution in the co-design problem.

For a system consisting of N subsystems, the co-design problem for the i^{th} subsystem is formulated as,

$$\min_{u_i(t),y_i,y_{s_{i,j}}} Z_i = w_{P_i} z_{P_i}(y_i, y_{si,j}) + \frac{1}{2} w_{C_i} \int_0^{t_f} \Big(x_i^T(t) Q_i x_i(t) + u_i^T(t) R_i u_i(t) + 2 x_i^T(t) S_i u_i(t) \Big) dt$$

subject to the constraints,

$$f_i(y_i, y_{si,j}) \leq 0 \quad \text{(Physical inequality constraints)} \tag{2.1}$$

$$g_i(y_i, y_{si,j}) = 0 \quad \text{(Physical equality constraints)}$$

$$\dot{x}_i(t) = A_i(y_i, y_{si,j}) x_i(t) + B_i(y_i, y_{si,j}) u_i(t) + \sum_{(i,j)\in V} A_{ij}(y_j, y_{si,j}) x_j(t)$$

(System dynamics)

$x_i(0) = x_{i0}$ (Initial conditions)

The weights associated with the physical and control objective functions of the i^{th} subsystem are w_{P_i} and w_{C_i}, where $w_{P_i} \in [0,1]$, $w_{C_i} \in [0,1]$ and $w_{P_i} + w_{C_i} = 1$ for all $i \in \{1, ..., N\}$. In the control objective function, Q_i and R_i are real symmetric positive definite weight matrices. The matrices A_i, B_i and A_{ij} representing the system dynamics can be linearly or nonlinearly parametrized in terms of the physical variables.

The overall co-design problem is formulated as,

$$\min_{u_i(t), y_i, y_{s_{i,j}}} Z = \sum_{i=1}^{N} w_{P_i} z_{P_i}(y_i, y_{s_{i,j}}) + \frac{1}{2} \sum_{i=1}^{N} w_{C_i} \int_0^{t_f} \Big(x_i^T(t) Q_i x_i(t) + u_i^T(t) R_i u_i(t) + 2 x_i^T(t) S_i u_i(t) \Big) dt$$

subject to the constraints,

$$f_i(y_i, y_{si,j}) \leq 0, \quad g_i(y_i, y_{si,j}) = 0, \quad y_{si,j} = y_{sj,i}, \quad \forall (i,j) \in V \tag{2.2}$$

$$\dot{x}_i(t) = A_i(y_i, y_{si,j}) x_i(t) + B_i(y_i, y_{si,j}) u_i(t) + \sum_{(i,j) \in V} A_{ij}(y_j, y_{si,j}) x_j(t)$$

$$x_i(0) = x_{i0}, \quad (i, j = 1, ..., N, j \neq i)$$

In the next section, a multi-level decentralized approach is developed to obtain optimal solution to the N-subsystem problem described by Eqn. (2.2).

2.2 Multi-level Decentralized Approach

As an example, Fig. 2.2 shows an overview of the proposed multi-level decentralized approach for solving a three-subsystem co-design problem. The approach consists the design of the physical and the control parts. Decomposition-based optimization techniques are applied in both parts.

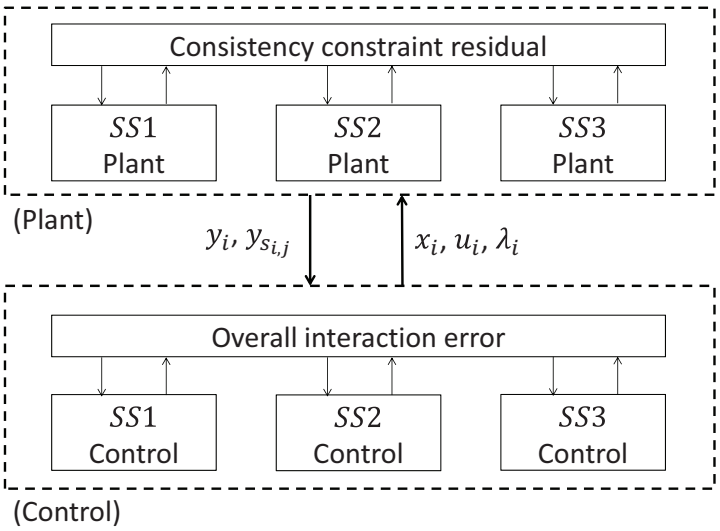


Figure 2.2: Scheme of Multi-level Decentralized Approach

First, the idea of interaction prediction method [29] is used in the control part (see Fig. 2.2, the 'Control' block), as described next. The bottom level of the control part solves for the individual optimal controller of each subsystem (SSi, $i = 1, 2, 3$). At the top level, the overall interaction error is calculated by the state and coordinator variables from interconnected subsystems. The coordinator variables are used to estimate the coupling between the subsystems. The coordinator variables are updated when the error is large, namely, the estimation of the coupling is not accurate

at the bottom level. When the error becomes sufficiently small, the decentralized solution of the control part is obtained.

For the physical part (see Fig. 2.2, the 'Plant' block), the physical variables are optimized using a dual decomposition scheme via a projected subgradient method [80]. At the bottom level, the partial Lagrangian is formulated with an initial dual variable vector. The local physical variables and copies of shared physical variables are optimized in each subsystem. At the top level, a consistency constraint residual is checked, and the dual variables are updated if the residual is not small enough.

In order to obtain the optimal or near-optimal solution of the entire co-design problem, the connection between the physical and the control parts needs to be established. The two parts are connected via the gradients of the Hamiltonian of the control part with respect to the physical variables. These values are computed from the necessary optimality conditions for the entire co-design problem. Then the physical and the control parts are solved iteratively to obtain a solution of the co-design problem.

If the problem is to find an optimal co-design for only one system, the proposed approach becomes the same as the 'nested' strategy in [1]. Since a multi-subsystem co-design problem is considered, hierarchical optimization techniques in both physical and control parts are applied, and so the approach can be regarded as an extension of the nested strategy [1].

2.2.1 Necessary Optimality Conditions

In this section, the necessary optimality conditions for the proposed multi-subsystem co-design are derived using the Karush–Kuhn–Tucker (KKT) conditions and Pontryagin Maximum Principle (PMP). These conditions are used to solve for variable values later on in the solution steps of the proposed approach. The idea is motivated by the previous work from Fathy et al. [1], where necessary optimality conditions for single-system co-design problems were developed.

The Lagrangian and the Hamiltonian can be written as,

$$L = \sum_{i=1}^{N} \left(w_{P_i} z_{P_i}(y_i, y_{si,j}) + \eta_i^T f_i(y_i, y_{si,j}) + \gamma_i^T g_i(y_i, y_{si,j}) \right) + \mu^T (y_{si,j} - y_{sj,i}) \quad (2.3)$$

$$H = \sum_{i=1}^{N} \left[\frac{1}{2} \left(x_i^T(t) Q_i x_i(t) + u_i^T(t) R_i u_i(t) + 2 x_i^T(t) S_i u_i(t) \right) + p_i^T(t) \left(A_i(y) x_i(t) + B_i u_i(t) + \sum_{(i,j) \in V} A_{ij}(y) x_j(t) \right) \right] \quad (2.4)$$

The necessary optimality conditions for the overall co-design problem (Eqn. (2.2)) are,

$$f_i(y_i, y_{si,j}) \leq 0, \quad g_i(y_i, y_{si,j}) = 0 \quad (2.5)$$

$$\eta_i \geq 0, \quad \eta_i \circ f_i(y_i, y_{si,j}) = 0 \quad (2.6)$$

$$y_{si,j} = y_{sj,i}, \quad \forall (i,j) \in V \quad (2.7)$$

$$\frac{\partial H}{\partial u_i} = 0, \quad \frac{\partial H}{\partial x_i} = -\dot{p}_i(t), \quad \frac{\partial H}{\partial p_i} = -\dot{x}_i(t) \quad (2.8)$$

$$x_i(0) = x_{i0} \quad (2.9)$$

$$\frac{\partial L}{\partial y_i} + \frac{\partial H}{\partial y_i} = 0 \tag{2.10}$$

$$\frac{\partial L}{\partial y_{si,j}} + \frac{\partial H}{\partial y_{si,j}} = 0 \tag{2.11}$$

$$(i, j = 1, ..., N, j \neq i)$$

The notation '$\circ$' in Eqn. (2.6) refers to the Hadamard product, meaning an element-wise multiplication of two vectors or two matrices.

To devise an iterative approach between the physical and control parts, the necessary optimality conditions are considered with respect to the physical variables that involve the Hamiltonian from the control part, namely, Eqns. (2.10) and (2.11).

2.2.2 Dual Decomposition

A dual decomposition technique [80] is used to optimize the physical design variables of the plant (see Fig. 2.2). Since both the local and shared physical variables as well as the coupling constraints exist in the defined class of co-design problems considered here, the dual decomposition via projected subgradient method is used to solve the plant part in a decentralized way [80]. Each subsystem has the local physical variables, copies of the shared physical variables and the associated dual variables, which are locally optimized at the bottom level. At the top level, the dual variable vector is updated iteratively to ensure consistency of the copies of shared physical variables.

Suppose there are a total number of n_c shared physical variables. A vector $\sigma \in \mathbf{R}^{n_c}$ is introduced to represent the common values of the shared physical vari-

ables. Let $\hat{y_s}$ be the vector that includes all shared physical variables $y_{s_{i,j}}$ from all subsystems for all i, j. Then the consistency constraints can be written as $\hat{y_s} = E\sigma$, where E is a matrix representing the set of consistency constraints for shared physical variables, with the matrix entries $E_{ij} = 1$ when the physical variables are shared between the subsystems SSi and SSj, and $E_{ij} = 0$ otherwise.

The details of the solution steps for the proposed approach are given in the next section.

2.2.3 Approach: Solution Steps

The flowchart describing the approach is shown in Fig. 2.3. Step 1 (or **S1**) is to choose the initial values of the physical variables. The steps (**S2**)-(**S4**) are based on the interaction prediction method [29], which solves the control part in a decentralized way. The iterations to run the interaction prediction method are denoted as the 'Control Inner Loop', see Fig. 2.3. After the converged results are obtained in the control part, the gradients of the Hamiltonian with respect to the physical variables are computed, in the step (**S5**), which are the link between the plant and control parts. Steps (**S6**)-(**S7**) are denoted as the 'Plant Inner Loop', in which the plant part is optimized through the dual decomposition [80] via a projected subgradient method. The gradients of the Hamiltonian with respect to the physical variables are used at the bottom level to reveal the influence from the control part. At the top level, if the consistency constraint residual is not small enough, the dual variables are updated and the iteration is repeated from (**S6**).

Otherwise, the process goes to the step (**S8**), checking the difference between the converged physical design variable values obtained in the step (**S7**) and the values in the step (**S2**). If the difference is small enough, the approach ends and the converged results are obtained. The iterations in the steps (**S2**)-(**S8**) are denoted as the 'Outer Loop' of the proposed approach, see Fig. 2.3.

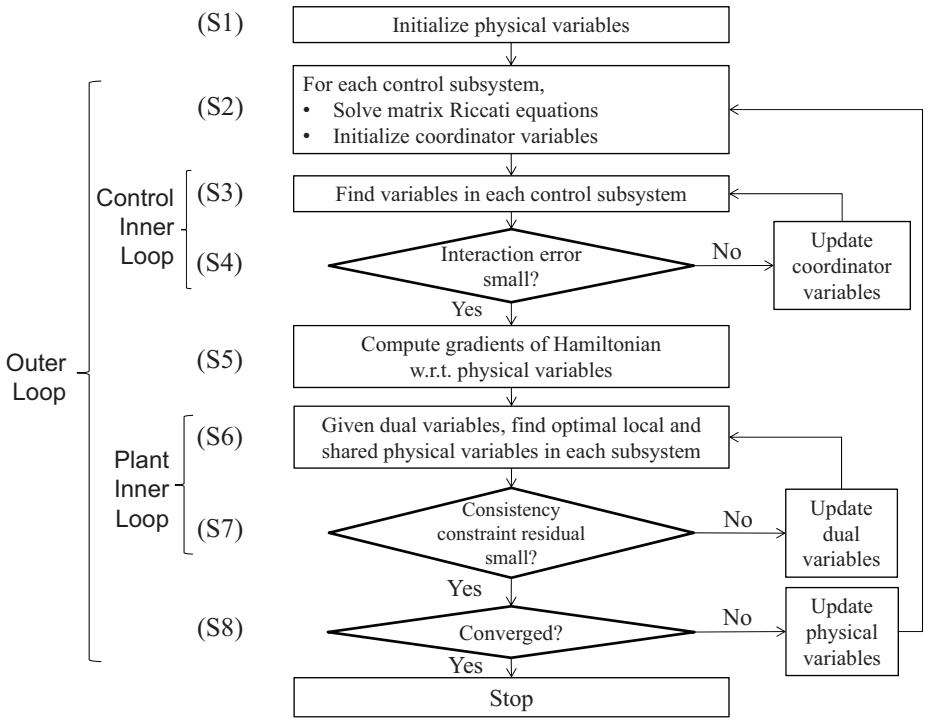


Figure 2.3: Flowchart of the Approach

In the multi-subsystem optimal co-design problem, given a set of fixed w_{P_i}'s and w_{C_i}'s, the parameter values in the matrices A_i, B_i, A_{ij}, Q_i, R_i and S_i, and initial conditions x_{i0} for all $i = 1, 2, ..., N$, the solution steps are detailed as in the following:

(**S1**) Select the initial values of y_{i0} and $y_{s_{i,j}0}$ for the physical variables of the i^{th} subsystem. Set the current physical variables' values, y_{i-c} and $y_{s_{i,j}-c}$, equal to y_{i0}

and $y_{si,j0}$ for all i.

(**S2**) Insert y_{i-c} and $y_{si,j-c}$ in the dynamic equations of the problem (Eqn. (2.2)) and find A_i, A_{ij} and B_i matrices. For the i^{th} subsystem ($i = 1, ..., N$), solve the independent matrix Riccati equation and obtain $P_i(t)$,

$$\begin{aligned}\dot{P}_i(t) = &-P_i(t)\big(A_i - B_i R_i^{-1} S_i^T\big) - \big(A_i^T - S_i R_i^{-1} B_i^T\big) P_i(t) \\ &+ P_i(t) B_i R_i^{-1} B_i^T P_i(t) - Q_i + S_i R_i^{-1} S_i^T \end{aligned} \tag{2.12}$$

with $P_i(t_f) = [0]$. Next, select the initial values of $\alpha_i^{(0)}(t)$, and compute $v_i^{(0)}(t)$ as,

$$v_i^{(0)}(t) \equiv \sum_{(i,j)\in V} A_{ij} x_{j0}, \quad \forall\, t \in [0, t_f] \tag{2.13}$$

where the superscript $(\cdot)^{(0)}$ in this step refers to the initial value of a variable, and in the following steps the superscript $(\cdot)^{(l)}$ refers to the value of a variable in the l^{th} inner loop.

(**S3**) For the l^{th} control inner loop, use $\alpha_i^{(l-1)}(t)$ and $v_i^{(l-1)}(t)$ to solve $h_i^{(l)}(t)$ and $x_i^{(l)}(t)$ with the boundary condition $h_i^{(l)}(t_f) = 0$ and the initial condition $x_i^{(l)}(0) = x_{i0}$ as,

$$\begin{aligned}\dot{h}_i^{(l)}(t) = &-\big(A_i^T - P_i(t) B_i R_i^{-1} B_i^T - S_i R_i^{-1} B_i^T\big) h_i^{(l)}(t) \\ &- P_i(t) v_i^{(l-1)}(t) + \sum_{(i,j)\in V} A_{ji} \alpha_j^{(l-1)}(t) \end{aligned} \tag{2.14}$$

$$\begin{aligned}\dot{x}_i^{(l)}(t) = &\big(A_i - B_i R_i^{-1} B_i^T P_i(t) - B_i R_i^{-1} S_i^T\big) x_i^{(l)}(t) \\ &- B_i R_i^{-1} B_i^T h_i^{(l)}(t) + v_i^{(l-1)}(t) \end{aligned} \tag{2.15}$$

Compute the co-state variables $\lambda_i^{(l)}(t)$,

$$\lambda_i^{(l)}(t) = P_i(t)x_i^{(l)}(t) + h_i^{(l)}(t) \tag{2.16}$$

Note that in (**S2**) and (**S3**), Eqn. (2.12) - (2.16) are calculated within each subsystem, so it is possible to compute these values in a parallel manner.

(**S4**) Compute the overall interaction error $e_C^{(l)}$,

$$e_C^{(l)} = \sum_{i=1}^{N} \int_0^{t_f} \left(\|v_i^{(l-1)}(t) - \sum_{(i,j)\in V} A_{ij}x_j^{(l)}(t)\|_2^2 \right) dt \tag{2.17}$$

If $e_C^{(l)} > \epsilon_C$, namely, the error of the estimation of the interaction between subsystems in the step (**S3**) is not sufficiently small, then update the coordinator variables to $\alpha_i^{(l)}(t)$ and $v_i^{(l)}(t)$,

$$\alpha_i^{(l)}(t) = -\lambda_i^{(l)}(t) \tag{2.18}$$

$$v_i^{(l)}(t) = \sum_{(i,j)\in V} A_{ij}x_j^{(l)}(t) \tag{2.19}$$

and repeat from the step (**S3**). Otherwise, continue to the step (**S5**). Let $x_i(t)$, $\lambda_i(t)$ and $u_i(t)$ denote the time-variant values of the state, co-state and control variables in the last inner loop for the i^{th} subsystem, where

$$u_i(t) = -R_i^{-1}\big(B_i^T P_i(t) + S_i^T\big)x_i(t) - R_i^{-1}B_i^T h_i(t). \tag{2.20}$$

(**S5**) Starting with this step, the approach moves to the physical part. Calculate the

gradient of the Hamiltonian of the control part with respect to the physical variables at y_{i-c} and $y_{si,j-c}$ values,

$$\left.\frac{\partial H}{\partial y_i}\right|_{y_{i-c},y_{s_{i,j-c}}} = \sum_{i=1}^{N}\left[w_{C_i}\int_0^{t_f}\lambda_i^T(t)\frac{\partial}{\partial y_i}\left(A_i(y_i,y_{si,j})x_i(t)+B_i(y_i,y_{si,j})u_i(t)\right.\right.$$
$$\left.\left.+\sum_{(i,j)\in V}A_{ij}(y_i,y_{si,j})x_j(t)\right)dt\right] \tag{2.21}$$

$$\left.\frac{\partial H}{\partial y_{si,j}}\right|_{y_{i-c},y_{s_{i,j-c}}} = \sum_{i=1}^{N}\left[w_{C_i}\int_0^{t_f}\lambda_i^T(t)\frac{\partial}{\partial y_{si,j}}\left(A_i(y_i,y_{si,j})x_i(t)+B_i(y_i,y_{si,j})u_i(t)\right.\right.$$
$$\left.\left.+\sum_{(i,j)\in V}A_{ij}(y_i,y_{si,j})x_j(t)\right)dt\right] \tag{2.22}$$

(**S6**) Set the initial dual variable vector $\mu^{(0)}=0$. For the i^{th} subsystem, the Lagrangian of the physical part is,

$$L_i = w_{P_i}z_{P_i}(y_i,y_{si,j})+\eta_i^T f_i(y_i,y_{si,j})+\gamma_i^T g_i(y_i,y_{si,j})+\mu^T y_{si,j} \tag{2.23}$$

Solve the physical variables $y_i, y_{si,j}$ by the necessary optimality conditions in the i^{th} subsystem,

$$\begin{aligned}
&\frac{\partial L_i}{\partial y_i}+\left.\frac{\partial H}{\partial y_i}\right|_{y_{i-c},y_{s_{i,j-c}}}=0\\
&\frac{\partial L_i}{\partial y_{si,j}}+\left.\frac{\partial H}{\partial y_{si,j}}\right|_{y_{i-c},y_{s_{i,j-c}}}=0\\
&\frac{\partial L_i}{\partial \gamma_i}=0\\
&f_i(y_i,y_{si,j})\le 0\\
&\eta_i\ge 0,\quad \eta_i\circ f_i(y_i,y_{si,j})=0
\end{aligned} \tag{2.24}$$

(**S7**) For the l^{th} plant inner loop, let $\hat{y}_s$ be the vector of all $y_{si,j}$ obtained in all

subsystems, compute the average of the shared physical variables by

$$\sigma = (E^T E)^{-1} E^T \hat{y}_s \tag{2.25}$$

and evaluate the consistency constraint's residual,

$$e_R^{(l)} = \|\hat{y}_s - E\sigma\|_2^2 \tag{2.26}$$

If $e_R^{(l)} > \epsilon_R$, then update the dual variable vector via a subgradient method with a fixed step size β,

$$\mu^{(l)} = \mu^{(l-1)} + \beta \hat{y}_s \tag{2.27}$$

and repeat from Eqn. (2.23). Otherwise, results in the plant's inner loop are obtained as y_{i-new} and $y_{si,j-new}$, and continue to step (**S8**).

(**S8**) If $\max\{\|y_{i-new} - y_{i-c}\|, \|y_{si,j-new} - y_{si,j-c}\|\} > \epsilon_P$ for $i = 1, ..., N$, update y_{i-c}, $y_{si,j-c}$ with the new physical design values y_{i-new}, $y_{si,j-new}$, and repeat from the step (**S2**). Otherwise, the approach has converged and the solution is obtained.

In the next section, two examples are solved using the proposed approach.

2.3 Examples

Two examples are given in this section. The first one is a numerical example, which is used to demonstrate the proposed approach of Section 2.2.3 step-by-step. The second example is an engineering model of a serial spring-mass-damper subsystems, in which each mass is separately controlled while the wire diameters

of the springs are the physical design variables. Finally, a scalable test problem is implemented by increasing the number of the spring-mass-damper subsystems.

2.3.1 Example 1: Numerical Example

The numerical example is modified from an example in Ref. [29], which is originally an optimal control problem. In order to make it as a co-design problem, five physical design variables (m_1 to m_5) are introduced in the example. The physical design objective functions and constraints are also added following the assumptions. The overall problem is formulated as,

$$\begin{aligned}\min_{m_1,\ldots,m_5,u_1(t),u_2(t)} Z = w_P\Big((m_1-1)^2+(m_2-2)^2+(m_3-1)^2+(m_4-2)^2+(m_5-3)^2\Big)\\ +\frac{1}{2}w_C\int_0^1\Big(x^T(t)Qx(t)+u^T(t)Ru(t)+2x^T(t)Su(t)\Big)dt\end{aligned}$$

subject to the constraints,

$$m_1^2+m_2^2+m_3^2+m_4^2-8\leq 0,\quad m_3+m_4+2m_5-8=0 \tag{2.28}$$

$$\dot{x}(t)=Ax(t)+Bu(t),\quad x_0=[-1,\,0.1,\,1,\,-0.5]^T$$

where

$$w_P=w_C=0.5$$

$$A = \begin{bmatrix} 2 & 0.1m_1^3 & 0.1m_4 & 0 \\ 0.2m_1 & -m_2 & 0.1 & -0.5 \\ 0.5 & 0.15 & m_5 & 0.5 \\ 0 & -0.2m_3 & -0.25 & -m_5^2 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0.1 & 0 \\ 0 & 0.5 \\ 0 & 0.25 \end{bmatrix}$$

$$Q = \text{diag}(2,\ 1,\ 1,\ 2), \quad R = \text{diag}(1, 2), \quad S = \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}^T$$

In order to solve the problem in a decentralized manner, the problem is decomposed into two subsystems ($SS1$ and $SS2$). The weights for the objective functions in each subsystem are $w_{P_1} = w_{P_2} = w_{C_1} = w_{C_2} = 0.5$. Local physical variables in the two subsystems are $y_1 = [m_1, m_2]^T$ and $y_2 = m_5$. The shared physical variables between the subsystems are $y_{s1,2} = y_{s2,1} = [m_3, m_4]^T$. The subsystems are then formulated as,

SS1:

$$\min_{y_1, y_{s1,2}, u_1(t)} Z_1 = w_{P_1}\left((m_1 - 1)^2 + (m_2 - 2)^2 + \frac{1}{2}(m_3 - 1)^2 + \frac{1}{2}(m_4 - 2)^2\right) + \frac{1}{2}w_{C_1}\int_0^1 \left(x_1^T(t)Q_1x_1(t) + R_1u_1^2(t) + 2x_1^T(t)S_1u_1(t)\right)dt$$

subject to the constraints,

$$f_1(y_1, y_{s1,2}) := m_1^2 + m_2^2 + m_3^2 + m_4^2 - 8 \leq 0 \tag{2.29}$$

$$\dot{x}_1(t) = A_1 x_1(t) + B_1 u_1(t) + A_{12} x_2(t), \quad x_{10} = [-1, 0.1]^T$$

$$A_1 = \begin{bmatrix} 2 & 0.1m_1^3 \\ 0.2m_1 & -m_2 \end{bmatrix}, \quad A_{12} = \begin{bmatrix} 0.1m_4 & 0 \\ 0.1 & -0.5 \end{bmatrix}$$

$$B_1 = [1,\, 0.1]^T, \quad Q_1 = \mathrm{diag}(2, 1), \quad R_1 = 1, \quad S_1 = [1,\, 1]^T$$

SS2:

$$\min_{y_2, y_{s_{2,1}}, u_2(t)} Z_2 = w_{P_2}\left(\frac{1}{2}(m_3 - 1)^2 + \frac{1}{2}(m_4 - 2)^2 + (m_5 - 3)^2\right) + \frac{1}{2} w_{C_2} \int_0^1 \left(x_2^T(t) Q_2 x_2(t) + R_2 u_2^2(t) + 2x_2^T(t) S_2 u_2(t)\right) dt$$

subject to the constraints,

$$g_2(y_2, y_{s_{2,1}}) := m_3 + m_4 + 2m_5 - 8 = 0 \tag{2.30}$$

$$\dot{x}_2(t) = A_2 x_2(t) + B_2 u_2(t) + A_{21} x_1(t), \quad x_{20} = [1, -0.5]^T$$

$$A_2 = \begin{bmatrix} m_5 & 0.5 \\ -0.25 & -m_5^2 \end{bmatrix}, \quad A_{21} = \begin{bmatrix} 0.5 & 0.15 \\ 0 & -0.2m_3 \end{bmatrix}$$

$$B_2 = [0.5,\, 0.25]^T, \quad Q_2 = \mathrm{diag}(1, 2), \quad R_2 = 2, \quad S_2 = [0,\, 1]^T$$

The tolerance values $\varepsilon_C = 0.01$, $\varepsilon_R = 0.01$, and $\varepsilon_P = 0.001$ are respectively used to check whether the control inner loop, plant inner loop and outer loop have converged. The problem is solved below following the steps of the proposed approach.

2.3.1.1 Approach: Solution Steps

(**S1**) Select the initial values of the physical variables $y_{10} = [1,1]^T$, $y_{20} = 1$, and $y_{s1,2\text{-}0} = [1,1]^T$ as the initial values for the physical variables. So $y_{1-c} = [1,1]^T$, $y_{2-c} = 1$, and $y_{s1,2-c} = [1,1]^T$.

(**S2**) Obtain A_1 and A_2 matrices using the initial values of physical design variables

$$A_1 = \begin{bmatrix} 2 & 0.1 \\ 0.2 & -1 \end{bmatrix}, \quad A_2 = \begin{bmatrix} 1 & 0.5 \\ -0.25 & -1 \end{bmatrix}$$
$$A_{12} = \begin{bmatrix} 0.1 & 0 \\ 0.1 & -0.5 \end{bmatrix}, \quad A_{21} = \begin{bmatrix} 0.5 & 0.15 \\ 0 & -0.2 \end{bmatrix} \tag{2.31}$$

Then solve Eqn. (2.12) for $i = 1, 2$ with boundary conditions $P_1(1) = P_2(1) = [0]$ to obtain $P_1(t)$ and $P_2(t)$. Next, set initial values of the coordinator variables of $\alpha^{(0)}(t)$ and $v^{(0)}(t)$ for $t \in [0,\ 1]$,

$$\begin{aligned} &\alpha_1^{(0)}(t) \equiv [0.5, 0.5]^T, \quad \alpha_2^{(0)}(t) \equiv [0.75, 0.75]^T \\ &v_1^{(0)}(t) \equiv A_{12}x_{20} \equiv [0.1, 0.35]^T, \quad v_2^{(0)}(t) \equiv A_{21}x_{10} \equiv [-0.45, -0.02]^T \end{aligned} \tag{2.32}$$

(**S3**) Compute $h_1^{(1)}(t)$ and $h_2^{(1)}(t)$ with boundary conditions $h_1^{(1)}(1) = h_2^{(1)}(1) = 0$ by Eqn. (2.14), and $x_1^{(1)}(t)$ and $x_2^{(1)}(t)$ with initial conditions $x_{10} = [-1, 0.1]^T$, $x_{20} = [1, -0.5]^T$ by Eqn. (2.15).

(**S4**) Check interaction error $e_C^{(1)}$ in Eqn. (2.17),

$$e_C^{(1)} = \int_0^1 \left(\|v_1^{(0)}(t) - A_{12}x_2^{(1)}(t)\|_2^2 + \|v_2^{(0)}(t) - A_{21}x_1^{(1)}(t)\|_2^2 \right) dt = 0.64 \tag{2.33}$$

Since $e_C^{(1)} > \epsilon_C$, update the coordinator variables to $\alpha_i^{(1)}(t)$ and $v_i^{(1)}(t)$ by Eqn. (2.18) and Eqn. (2.19),

$$\begin{aligned} \alpha_1^{(1)}(t) &= -\lambda_1^{(1)}(t) = -P_1(t)x_1^{(1)}(t) - h_1^{(1)}(t) \\ \alpha_2^{(1)}(t) &= -\lambda_2^{(1)}(t) = -P_2(t)x_2^{(1)}(t) - h_2^{(1)}(t) \\ v_1^{(1)}(t) &= A_{12}x_2^{(1)}(t) \\ v_2^{(1)}(t) &= A_{21}x_1^{(1)}(t) \end{aligned} \tag{2.34}$$

Repeat steps (**S3**) to (**S4**) until $e_C^{(l)} < \epsilon_C$ after the l^{th} iteration of the control's inner loop.

(**S5**) After $e_C^{(l)}$ becomes sufficiently small, using the values of $\lambda_i(t)$'s and $x_i(t)$'s in the last inner loop, compute the gradient of the Hamiltonian of the control part with respect to the physical variables,

$$\left.\frac{\partial H}{\partial y_1}\right|_{y_{1-c}, y_{s_{1,2-c}}} = 0.5 \int_0^1 \lambda_1^T(t) \frac{\partial}{\partial y_1} \Big(A_1 x_1(t) \Big) dt = \begin{pmatrix} -0.074 \\ \\ -0.048 \end{pmatrix} \tag{2.35}$$

$$\left.\frac{\partial H}{\partial y_2}\right|_{y_{2-c}, y_{s_{1,2-c}}} = 0.5 \int_0^1 \lambda_2^T(t) \frac{\partial}{\partial y_2} \Big(A_2 x_2(t) \Big) dt = 0.40 \tag{2.36}$$

$$\left.\frac{\partial H}{\partial y_{s_{1,2}}}\right|_{y_{1-c}, y_{s_{1,2-c}}} = \sum_{i=1}^{2} 0.5 \int_0^1 \lambda_i^T(t) \frac{\partial}{\partial y_{s_{1,2}}} \Big(A_i x_i(t) + \sum_{(i,j)\in V} A_{ij} x_j(t) \Big) dt$$

$$= \begin{pmatrix} 0.0035 \\ \\ -0.047 \end{pmatrix} \tag{2.37}$$

(**S6**) The Lagrangians of the physical parts in $SS1$ and $SS2$ are,

$$\begin{cases} L_1 = w_{P_1} z_{P_1}(y_1, y_{s1,2}) + \eta_1 f_1(y_1, y_{s1,2}) + \mu^T y_{s1,2} \\ \\ L_2 = w_{P_2} z_{P_2}(y_2, y_{s2,1}) + \gamma_2 g_2(y_2, y_{s2,1}) - \mu^T y_{s2,1} \end{cases} \tag{2.38}$$

Set the initial value of the dual variable vector $\mu = 0$. Solve for y_1 and $y_{s1,2}$, and y_2 and $y_{s2,1}$ separately in the two subsystems using the necessary optimality conditions, as in Eqn. (2.24),

SS1:

$$\begin{aligned} &\frac{\partial L_1}{\partial y_1} + \frac{\partial H}{\partial y_1}\bigg|_{y_{1-c}, y_{s1,2-c}} = 0 \\ &\frac{\partial L_1}{\partial y_{s1,2}} + \frac{\partial H}{\partial y_{s1,2}}\bigg|_{y_{1-c}, y_{s1,2-c}} = 0 \\ &f_1(y_1, y_{s1,2}) \leq 0, \quad \eta_1 \geq 0, \quad \eta_1 f_1(y_1, y_{s1,2}) = 0 \end{aligned} \tag{2.39}$$

SS2:

$$\begin{aligned} &\frac{\partial L_2}{\partial y_2} + \frac{\partial H}{\partial y_2}\bigg|_{y_{2-c}, y_{s1,2-c}} = 0, \quad \frac{\partial L_2}{\partial y_{s2,1}} + \frac{\partial H}{\partial y_{s1,2}} = 0 \\ &\frac{\partial L_2}{\partial \gamma_2} = 0, \quad g_2 = 0 \end{aligned} \tag{2.40}$$

which yield

$$y_1 = [0.97,\ 1.85]^T, \quad y_{s_{1,2}} = [0.82,\ 1.72]^T \tag{2.41}$$

$$y_2 = 2.53, \qquad\qquad y_{s_{2,1}} = [0.92,\ 2.02]^T \tag{2.42}$$

By the assumption A3, in this step, the group of equations (Eqn. (2.39) and (2.40)) have a unique solution. The combined shared physical variable vector is,

$$\hat{y_s} = [0.82,\ 1.72,\ 0.92,\ 2.02]^T \tag{2.43}$$

(S7) Since the consistency constraint is $y_{s_{1,2}} = y_{s_{2,1}}$, the matrix E is therefore

$$E = \begin{bmatrix} & & & \\ 1 & 0 & 1 & 0 \\ & & & \\ 0 & 1 & 0 & 1 \\ & & & \end{bmatrix}^T$$

Now, calculate σ by Eqn. (2.25) and evaluate the consistency constraint residual by Eqn. (2.26),

$$\sigma = (E^T E)^{-1} E^T \hat{y_s} = [0.87,\ 1.87]^T \tag{2.44}$$

$$e_R^{(1)} = 0.22 \tag{2.45}$$

Since $e_R^{(1)} > \epsilon_R$, μ is updated by Eqn. (2.27) with $\beta = 0.05$,

$$\mu = [-0.0026,\ -0.0075,\ 0.0026,\ 0.0075]^T \tag{2.46}$$

and repeat from step (**S6**) until the residual is small enough. The new values for the physical variables are

$$\begin{aligned} y_{1-new} &= [0.92,\ 1.76]^T \\ y_{2-new} &= 2.65 \\ y_{s1,2-new} &= [0.87,\ 1.82]^T \end{aligned} \tag{2.47}$$

(**S8**) Since $\max\{\|y_{1-new} - y_{1-c}\|,\ \|y_{2-new} - y_{2-c}\|, \|y_{s1,2-new} - y_{s1,2-c}\|\} > \epsilon_D$, update y_{1-c}, y_{2-c}, $y_{s1,2-c}$ with the values of y_{1-new}, y_{2-new}, $y_{s1,2-new}$, and repeat previous steps from step (**S2**). Stop when the maximal value of all the norms are less than ϵ_D.

2.3.1.2 Optimization Results and Discussion

The final optimal solutions obtained by different approaches are shown in Table 2.1. The centralized solution is obtained by TOMLAB™ [81], which solves for all variables simultaneously. The results show that the optimal solution obtained by the decentralized method matches well with the centralized solution, when using the same initial conditions. For this example, the decentralized approach takes less than or equal to 4 iterations for all control inner loops, around 15 iterations

for all plant inner loops and a total of 9 iterations for the outer loop to converge. As a comparison and as shown in Table 2.1, if the co-design problem is solved in a sequential manner, it obtains a solution which is worse than the result for the centralized and decentralized approaches.

	y_1^*	y_2^*	$y_{s1,2}^*$	Z^*
Decentralized	$[0.68, 1.29]^T$	2.31	$[1.33, 2.03]^T$	3.38
Centralized	$[0.7064, 1.3217]^T$	2.3375	$[1.3263, 1.9987]^T$	3.3353
Sequential	$[0.9313, 1.8626]^T$	2.7292	$[0.8052, 1.7365]^T$	4.1479

Table 2.1: Example 1: Optimal Solutions

Fig. 2.4 shows the trajectories of the optimal controller $u_2(t)$ of SS2. As shown, $u_2(t)$ obtained by the centralized and decentralized methods are nearly the same, while the magnitudes of both are smaller than that obtained from the sequential approach, leading to a better optimal value of objective function.

Note that different values of initial points for physical design variables are tested in this example, including both feasible and infeasible initial points. The test results indicate that the converged solutions for the physical variables are nearly identical, which means that the proposed approach is robust to an initial point.

2.3.2 Example 2: Spring-mass-damper System Model

This example is composed of three spring-mass-damper subsystems, which are connected to each other in series, as shown in Fig. 2.5.

In the control part, for the i^{th} subsystem, the state variable $x_i(t) = [x_{i1}(t), x_{i2}(t)]^T$ denotes position and velocity of the i^{th} mass at time t, and the control variable $u_i(t)$

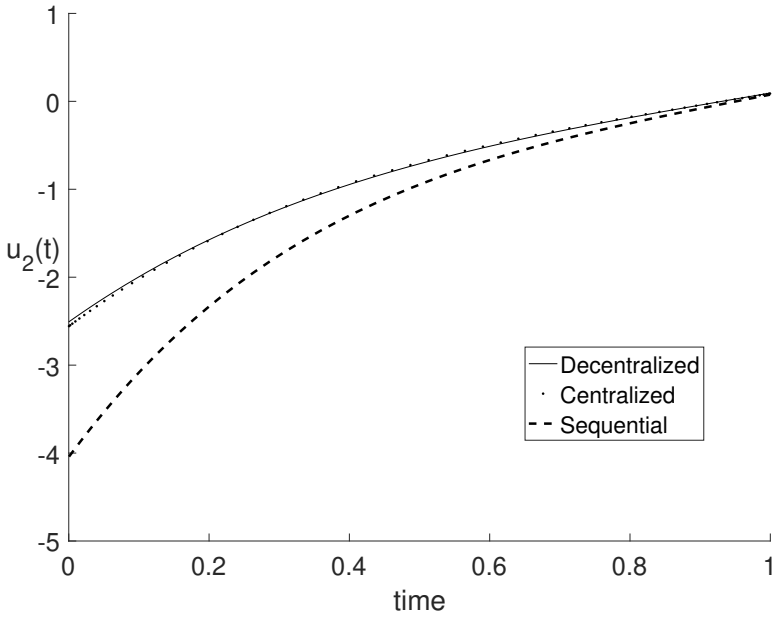


Figure 2.4: Example 1: $SS2$ Optimal Controller

is the input applied to the mass. The control objective function of each subsystem is a quadratic cost function, with the weighting matrices $Q_i = \text{diag}(1, 1)$ and $R_i = 1$ for $i = 1, 2, 3$. The system-level objective function is a weighted sum of all subsystems' physical and control design objective functions, with the associated weights chosen as $w_{P_i} = w_{C_i} = 0.5$ for all i. The final time is: $t_f = 5$ seconds. The tolerance values are set as $\varepsilon_C = 0.1$, $\varepsilon_R = 0.01$, and $\varepsilon_P = 0.001$.

In the physical design part, it is assumed that the damping coefficients and the masses are known as $c_1 = 10\,\text{kg} \cdot \text{s}^{-1}$, $c_2 = 15\,\text{kg} \cdot \text{s}^{-1}$, $c_3 = 20\,\text{kg} \cdot \text{s}^{-1}$, and $m_1 = m_2 = m_3 = 5\,\text{kg}$. The physical design part of the model considered in this example is based on a spring design optimization formulation in [82]. The wire diameters of springs are selected as the physical variables, i.e., $y_i = d_i\,(i = 1, 2, 3)$. The physical objective function in the i^{th} subsystem, z_{D_i}, is to maximize energy storage capacity. It is shown in [82] that if the spring index C is fixed at the

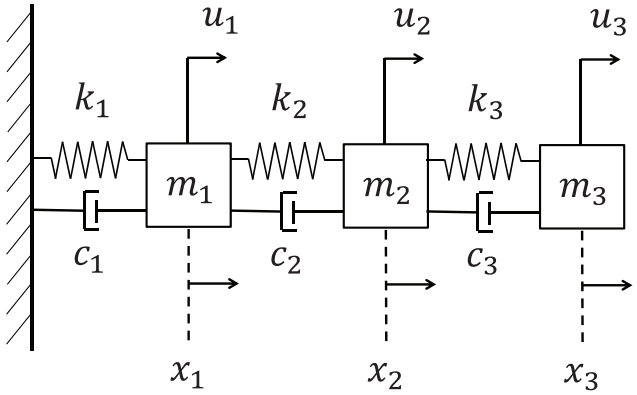


Figure 2.5: Example 2: Series of Three Spring-mass-damper Subsystems

maximum value, then the optimal spring design can be found by solving the following optimization problem,

$$\begin{aligned}
&\min \quad z_{D_i} = (10.24G)\left(\frac{\pi}{12.8F_u}\right)^2 C^{-2}y_i^4 \\
&\text{subject to the constraints,} \\
&C^{-1} - \kappa_1 C^{-1} y_i^{-1} \leq 1 \qquad \text{(Inside diameter)} \\
&\kappa_2 y_i^{-1} \leq 1 \qquad \text{(Lower bound on wire diameter)} \\
&\frac{0.8F_u}{D_s G} C^3 y_i^{-2} \leq 1 \qquad \text{(Clash allowance)}
\end{aligned} \tag{2.48}$$

where G is shear modulus ($G = 30$ MPa, for chrome silicon), F_u is maximum allowable force (N), C is spring index (dimensionless), κ_1 is the minimum allowable inside diameter (m), κ_2 is the lower bound on physical variables (m), and D_s is a

clearance constant (dimensionless). The spring constants can be calculated as [4],

$$k_i = \frac{d_i^4 G}{8 D_i^3 N_a \left(1 + \frac{1}{2C^2}\right)} \tag{2.49}$$

where $D_i = C d_i$ is coil diameter, and N_a is the number of active coils of the spring.

In this example, the parameters are set as $C = 8$, $F_u = 6$, $\kappa_1 = 0.05$, $\kappa_2 = 0.1$, $D_s = 0.25$, and $N_a = 100$. The problem is formulated as the following,

$$\begin{aligned}
\min_{y_i, u_i(t)} \quad & Z = \sum_{i=1}^{3} w_{P_i} (10.24G) \left(\frac{\pi}{12.8 F_u}\right)^2 C^{-2} y_i^4 \\
& + \frac{1}{2} \sum_{i=1}^{3} w_{C_i} \int_0^5 \left(x_i^T(t) Q_i x_i(t) + R_i u_i^2(t) \right) dt
\end{aligned}$$

subject to the constraints,

$$(C^{-1} - 1) y_i - \kappa_1 C^{-1} \leq 0, \qquad i = 1,\, 2,\, 3$$

$$\kappa_2 - y_i \leq 0, \qquad i = 1,\, 2,\, 3$$

$$\frac{0.8 F_u}{D_s G} C^3 y_i^{-2} \leq 1, \qquad i = 1,\, 2,\, 3 \tag{2.50}$$

$$\dot{x}_1(t) = \begin{bmatrix} 0 & 1 \\ -\frac{k_1+k_2}{m_1} & -\frac{c_1+c_2}{m_1} \end{bmatrix} x_1(t) + \begin{bmatrix} 0 & 0 \\ \frac{k_2}{m_1} & \frac{c_2}{m_1} \end{bmatrix} x_2(t) + \begin{bmatrix} 0 \\ \frac{1}{m_1} \end{bmatrix} u_1(t)$$

$$\dot{x}_2(t) = \begin{bmatrix} 0 & 1 \\ -\frac{k_2+k_3}{m_2} & -\frac{c_2+c_3}{m_2} \end{bmatrix} x_2(t) + \begin{bmatrix} 0 & 0 \\ \frac{k_2}{m_2} & \frac{c_2}{m_2} \end{bmatrix} x_1(t)$$

$$+\begin{bmatrix} 0 & 0 \\ \frac{k_3}{m_2} & \frac{c_3}{m_2} \end{bmatrix} x_3(t) + \begin{bmatrix} 0 \\ \frac{1}{m_2} \end{bmatrix} u_2(t)$$

$$\dot{x}_3 = \begin{bmatrix} 0 & 1 \\ -\frac{k_3}{m_3} & -\frac{c_3}{m_3} \end{bmatrix} x_3(t) + \begin{bmatrix} 0 & 0 \\ \frac{k_3}{m_3} & \frac{c_3}{m_3} \end{bmatrix} x_2(t) + \begin{bmatrix} 0 \\ \frac{1}{m_3} \end{bmatrix} u_3(t)$$

$$x_{i0} = [1, 1]^T, \qquad i = 1,\ 2,\ 3$$

Finally, it is assumed that a unique equilibrium exists for the entire multi-subsystem co-design problem. The proposed decentralized approach is applied to solve this problem. Note that all physical variables are shared variables since they exist in more than one subsystems. Thus, the copies of the shared physical variables are optimized in each subsystem.

2.3.2.1 Optimization Results and Discussion

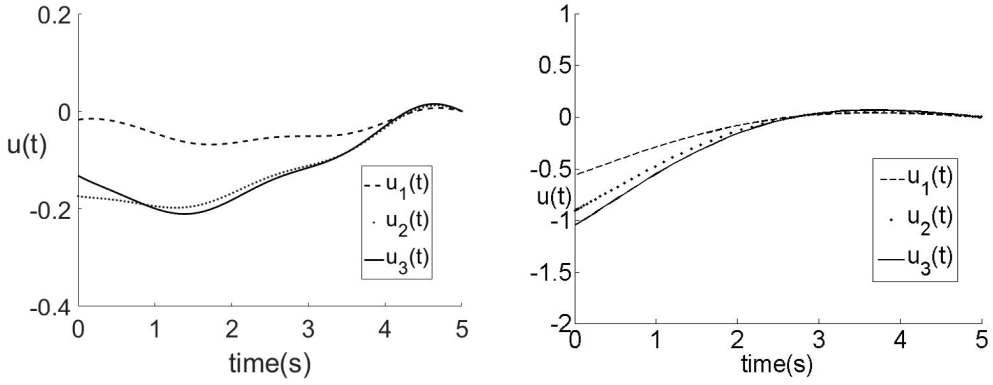
The problem is solved by the proposed decentralized approach and by the centralized approach using TOMLAB™ [81]. The solutions for the optimal damping coefficients are shown in Table 2.2. The profiles of optimal controllers $u_i(t)$ and velocities of all masses $x_{i2}(t)$ obtained by both centralized and decentralized methods are shown in Fig. 2.6 and Fig. 2.7. As is observed, the state variable profiles of the decentralized approach are very close to those of the centralized approach.

From Table 2.2, it can be observed that the optimal value of the entire objec-

	y_1^*	y_2^*	y_3^*	Z^*
Decentralized	0.10	0.10	0.10	7.34
Centralized	0.1166	0.1000	0.1000	6.8558
Sequential	0.1000	0.1000	0.1000	6.9860

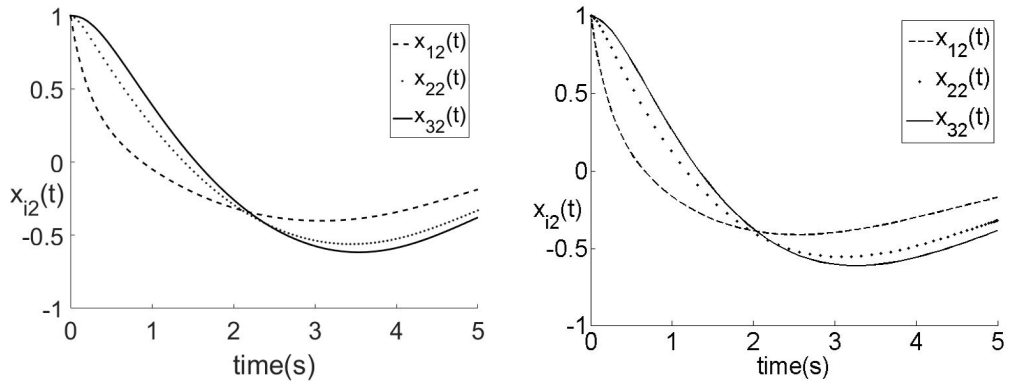
Table 2.2: Example 2: Optimal Solutions

tive function of the decentralized approach is slightly higher that of the centralized and the sequential methods. However, the limitation of a centralized or a sequential approach is that the optimization requires full information from all subsystems to solve all variables simultaneously. While in a decentralized manner, the full information is not needed, and smaller subproblems can be solved in the process. Take $SS1$ as an example, in the decentralized approach, $SS1$ can be optimized with only considering its coupling with $SS2$, and this subproblem has a smaller size than the overall problem.



(a) Decentralized (b) Centralized

Figure 2.6: Example 2: Optimal Control Variables $(u_i(t), i = 1, 2, 3)$



(a) Decentralized (b) Centralized

Figure 2.7: Example 2: Optimal State Variables - Velocity ($x_{i2}(t), i = 1, 2, 3$)

2.3.2.2 Computational Cost

In this section it is shown that, by way of a scalable test problem, as the size of the co-design problem (e.g., number of subsystems) increases, the proposed decentralized approach performs better in terms of computational cost than the centralized approach. The test problem is a spring-mass-damper with N subsystems (with each subsystem composed of a simple spring-mass-damper subproblem), as shown in Fig. 2.8. With the same objective functions and constraints as in the engineering example, in each subproblem, there exists two state variables (displacement and velocity of the mass), one control variable (force on the mass), and one physical plant variable (spring's wire diameter). The number of subsystems considered for each test problem is: 5, 10, 20, 30, all the way to 100. Each of these 11 test problems is solved by the centralized approach as well as the proposed decentralized approach using the same tolerance value for convergence. As an example, when the number of subsystems is equal to 100, there are 200 state, 100 control, and 100 physical plant

variables.

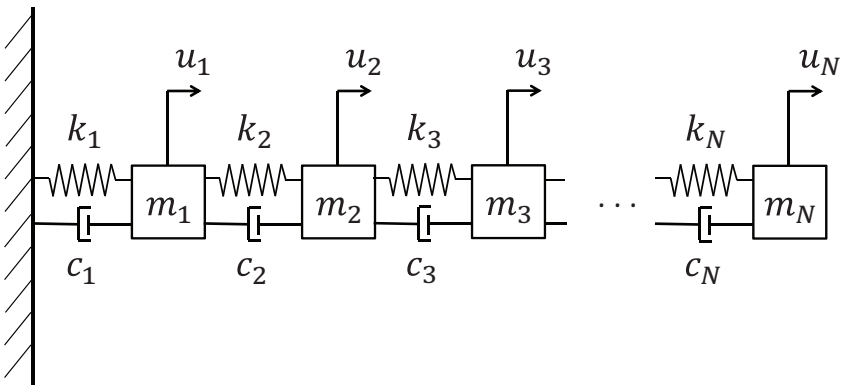


Figure 2.8: Series of Multiple Spring-mass-damper Subsystems

Fig. 2.9 shows a comparison of the computational (CPU) time between the centralized and decentralized approach for the solution of this test problem as the number of subsystems is increased. The computational time of the centralized approach is based on the CPU time of the solutions obtained from TOMLAB [81], which employs an efficient approach for solving co-design problems by an all-in-one approach. Fig. 2.9 clearly shows that as the number of variables increases, the computational time of the decentralized approach increases about linearly with respect to the number of subsystems. On the other hand, the computational cost of the centralized approach grows approximately nonlinearly.

2.4 Summary

In this chapter, the formulation for a class of multi-subsystem co-design problems is presented. In this class of problems, the physical design part in each subsystem has a convex objective function with convex inequality and linear equality con-

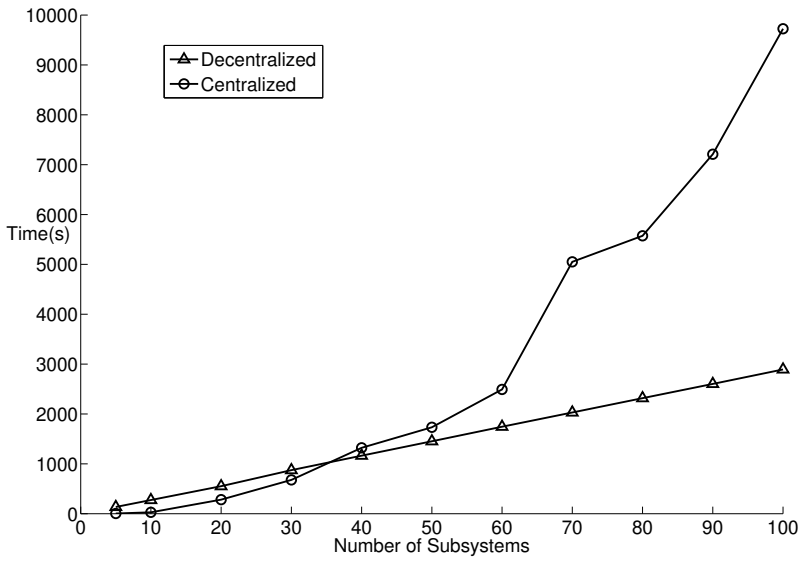


Figure 2.9: Comparison of Computational Time between Centralized and Proposed Decentralized Approaches

straints. The control part of each subsystem belongs to a class of finite time-horizon LQR feedback control. While many papers on co-design optimization problems have been published in recent years, co-design problems with a multi-subsystem structure that has both physical design and control domains in the same subsystem have not been specifically considered.

The proposed decentralized approach is composed of two hierarchical decomposition techniques, one for the physical part and the other for the control part of the co-design subsystems. The co-design problems are solved in a decentralized manner via a hierarchical structure inside each part. The necessary optimality conditions are derived for the overall problem, as well as for the subsystem subproblems. Derived from the necessary optimality conditions, the gradients of the Hamiltonian of the control part with respect to the physical variables bridge the physical and control parts, providing the quantitative information of the coupling between the

two parts.

The decentralized approach is applied to several examples. In the first numerical example, the approach is illustrated in a detailed step-by-step manner. In the second example, an engineering co-design model of a three spring-mass-damper system is solved. Finally, a scalable test problem is considered. For this problem, the test results show that the computational time of the proposed decentralized approach increases approximately linearly with respect to an increase in the number of subsystems while the computational cost of the centralized approach grows at a higher rate.

In the next chapter, the results from Research Question 2 (RQ2) are presented. In RQ2, more general co-design problems than those in this chapter are considered, and two decentralized solution approaches are proposed.

Chapter 3: Decentralized Multi-subsystem Co-design Optimization Using Direct Collocation and Decomposition-based Methods

In this chapter, the results from Research Question 2 are presented in detail[1].

The organization of this chapter is as follows. This chapter begins with a nomenclature section. Then in Section 3.1, formulations for the multi-subsystem co-design problem with relevant assumptions are detailed. In Section 3.2, the direct collocation method and the optimality condition decomposition are briefly introduced, and a decentralized implementation of the direct collocation method is described. Next, the multi-level and bi-level decentralized approaches are proposed in Section 3.3. The approaches are applied to a number of numerical and engineering examples, and the results are shown in Section 3.4. Finally, the summary of this chapter is presented in Section 3.5.

[1]This chapter is based on the paper: Liu, T., Azarm, S., and Chopra, N., 2020, "Decentralized Multi-subsystem Co-design Optimization Using Direct Collocation and Decomposition-based Methods." ASME. J. Mech. Des. doi: https://doi.org/10.1115/1.4046438

Nomenclature

E	Consistency constraint matrix
e	Norm of difference of two vectors
f	Dynamic equations of the system
g	Inequality constraints for physical plant design variables
H	Hamiltonian of control subproblems
h	Equality constraints for physical plant design variables
L	Lagrangian of physical plant design optimization subproblems
M	Number of subintervals in discretization
N	Number of subsystems
p	Vector of control/path constraints
q	Vector of initial or final conditions of state variables
s	Time difference between two grid points

t_f	Final time
u	Vector of control variables
V	Set of pairs of connected subsystems
w_C	Weighting coefficient associated with the control objective function
w_P	Weighting coefficient associated with the physical design objective
X	Combined optimizing variable vector
x	State variables
y_i	Local physical plant design variable vector in i^{th} subsystem
$y_{si,j}$	Shared physical plant design variable vector in i^{th} and j^{th} subsystems
$\bar{y}_s$	Vector of combined shared physical plant design variables
Z	System-level objective function
z_C	Control objective function

z_P Physical plant design objective function

β Step size constants

γ, η Lagrange multipliers of physical plant design constraints

ϵ_D Preset tolerance value of convergence for optimizing variables

ϵ_Z Preset tolerance value of convergence for objective function

ϵ_R Preset tolerance value of convergence for consistency constraint residual

ϵ_{y_s} Preset tolerance value of convergence for shared plant design variables

λ Co-state variables

μ Dual variable vectors associated with physical plant design consistency constraints

ν Dual variable vectors in bi-level approach

ρ	Lagrange multipliers associated with converted complicating constraints
σ	Common value vector of shared physical plant design variables
ϕ	Converted complicating constraints at collocation points
ψ	Scaling factors
$(\cdot)_i$	Symbol $(\cdot)$ with subscript i: variables/parameters in i^{th} subsystem
$(\cdot)_k$	Symbol $(\cdot)$ with subscript k: variables/parameters at grid point k

3.1 Multi-subsystem Co-design Problem Formulation

The multi-subsystem co-design optimization problem considered in RQ2 is similar to that in RQ1. As an example, Fig. 3.1, which is slightly different from Fig. 2.1, shows the structure of such a problem with three subsystems. Each subsystem has a plant design (block 'P') and a control (block 'C') part. The notations of the variables remain the same as in RQ1.

In the multi-subsystem co-design problems considered in RQ2, the control subproblems are relaxed. In this problem, the control objective functions, dynamics and state constraints can be nonlinear and non-convex. Hence, a decentralized for-

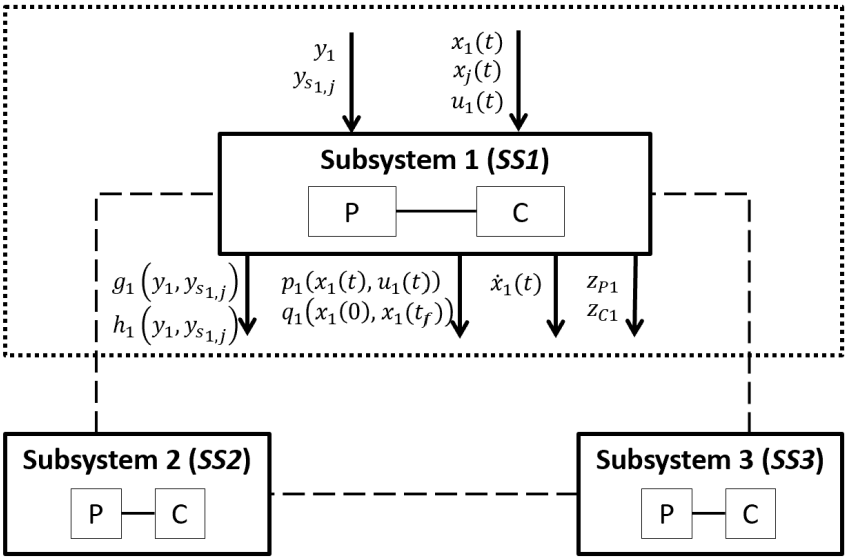


Figure 3.1: Problem Structure - Three Coupled Subsystems with Plant (P) and Control (C)

mulation of the i^{th} subsystem for a multi-subsystem co-design problem is as follows,

$$\min_{u_i, y_i, y_{s_{i,j}}} \quad Z_i = w_{P_i} z_{P_i}(y_i, y_{s_{i,j}}) + w_{C_i} \int_0^{t_f} z_{C_i}(x_i(t), u_i(t))dt$$

subject to the constraints,

$$g_i\left(y_i, y_{s_{i,j}}\right) \leq 0, \quad h_i\left(y_i, y_{s_{i,j}}\right) = 0$$

$$\begin{aligned} \dot{x}_i(t) &= f_i(x_i, u_i, x_j, y_i, y_{s_{i,j}}) \\ &= f_{ii}(x_i, u_i, y_i, y_{s_{i,j}}) + \sum f_{ij}(x_j, y_i, y_{s_{i,j}}) \end{aligned} \tag{3.1}$$

$$p_i\left(x_i(t), u_i(t)\right) \leq 0, \quad q_i\left(x_i(0), x_i(t_f)\right) = 0$$

$$(i = 1, ..., N)$$

$\dot{x}_i(t) = f_i$ represents the dynamic constraints, and p_i and q_i are path and control constraints.

3.2 Decentralized Implementation of Direct Collocation Method

In this section, the key formulations of a decentralized implementation of the direct collocation method [43] using the optimality condition decomposition [83] is described. The detailed derivation can be found in the Appendix A. The implementation is then utilized in both the multi-level and the bi-level approaches proposed in Section 3.3.

Consider the control part of the co-design problem (Eqn. (3.1)) without the plant design variables:

$$\min_{u_i} \quad \sum_{i=1}^{N} w_{C_i} \int_0^{t_f} z_{C_i}(x_i(t), u_i(t))dt$$

subject to the constraints,

$$\dot{x}_i(t) = f_i(x_i, u_i, x_j) \tag{3.2}$$

$$p_i\left(x_i(t), u_i(t)\right) \leq 0, \quad q_i\left(x_i(0), x_i(t_f)\right) = 0$$

$$(i = 1, ..., N)$$

With the direct method, the entire time-horizon $[0, t_f]$ is discretized into M subintervals by grid points, $0 = t_0 < t_1 < ... < t_M = t_f$, for all subsystems. The time difference s between two subsequent grid points is considered to be constant, i.e., a uniform grid with $s = t_k - t_{k-1} = t_f/M$, $(k = 1, ..., M)$. The control and state variables at all grid points are the optimizing variables in the converted problem. For convenience, state and control variables of SSi at t_k are denoted as $x_{i,k}$ and $u_{i,k}$.

Following the direct collocation method, the control variables between two grid points are approximated by linear interpolation, and the state variables are approximated by a cubic polynomial function. It can be shown [43] that the largest difference between the derivative of the approximation and that of the actual curve occurs at the midpoint of each interval, which is called a collocation point. Hence, a defect function $\phi_{i,k}$ for interval k in SSi is defined,

$$\phi_{i,k} = \dot{x}_{i,k+1/2} - f_i(x_{i,k+1/2}, u_{i,k+1/2}, x_{j,k+1/2}) \tag{3.3}$$

and is constrained as $\phi_{i,k} = 0$ in the converted problem in order to enforce that the approximating curve of each state variable matches with the actual one. Furthermore, following [43], $\phi_{i,k}$ can be expressed as,

$$\begin{aligned}\phi_{i,k} =& \frac{3(x_{i,k+1} - x_{i,k})}{2s} - \frac{1}{4}\left(f_{i,k} + f_{i,k+1}\right) \\ & - f_i(x_{i,k+1/2}, u_{i,k+1/2}, x_{j,k+1/2}) = 0\end{aligned} \tag{3.4}$$

The variables $u_{i,k+1/2}$ and $x_{i,k+1/2}$ of SSi at $t_{k+1/2}$ can be evaluated from $x_{i,k}$ and $u_{i,k}$. The evaluation of values of $x_{j,k+1/2}$ starts from the equations of variables at the collocation points,

$$x_{j,k+1/2} = \frac{1}{2}\left(x_{j,k} + x_{j,k+1}\right) + \frac{s}{8}\left(f_{j,k} - f_{j,k+1}\right) \tag{3.5}$$

Since $\dot{x}_j = f_j$, and with a two-point approximation of this differential equation [84],

$$f_{j,k} - f_{j,k+1} = \begin{cases} \frac{2x_{j,k+1} - x_{j,k} - x_{j,k+2}}{2s} & k = 1 \\ \frac{x_{j,k+1} + x_{j,k} - x_{j,k+2} - x_j}{2s} & 1 < k < M-1 \\ \frac{2x_{j,k} - x_{j,k+1} - x_{j,k-1}}{2s} & k = M \end{cases} \tag{3.6}$$

Hence, $x_{j,k+1/2}$ can be evaluated by plugging Eqn. (3.6) into Eqn. (3.5). If x_i is defined as the vector for the state and control variables at all grid points for SSi:

$$x_i = \left(u_{i,0}^T, \ ..., u_{i,M}^T, \ x_{i,0}^T, \ , ..., x_{i,M}^T \right)^T \tag{3.7}$$

then $\phi_{i,k}$ can be represented as a function of x_i and x_j, i.e., $\phi_{i,k}(x_i, x_j) = 0$. Hence, the control problem (Eqn. (3.2)) can be converted as a nonlinear program,

$$\begin{aligned} &\min_{x_i} \quad \sum_{i=1}^{N} w_{C_i} \bar{z}_{C_i}(x_i) \\ &\text{subject to the constraints,} \\ &\phi_{i,k}(x_i, x_j) = 0 \\ &p_i(x_i) \leq 0, \quad q_i(x_i) = 0 \quad (i = 1, \ ..., \ N) \end{aligned} \tag{3.8}$$

where $\bar{z}_{C_i}$ is the converted control objective function, in which the trapezoidal rule is used to approximate the integrals in the original objective function.

The converted problem (Eqn. (3.8)) can be solved by the optimality condition decomposition (OCD) method [83], which is an iterative bi-level approach. The

decomposed subproblem $SSi\,(i = 1,\ ...,\ N)$ is formulated as,

$$\begin{aligned} \min_{x_i} \quad & w_{C_i}\bar{z}_{C_i}(x_i) + \sum_{j=1, j\neq i}^{N} \hat{\varrho}_j^T \phi_j(x_i^\dagger) \\ & \text{subject to the constraints,} \\ & \phi_i(x_i^\dagger) = 0, \quad p_i(x_i) \leq 0, \quad q_i(x_i) = 0 \end{aligned} \tag{3.9}$$

where $x_i^\dagger = (\hat{x}_1, ..., \hat{x}_{i-1}, x_i, \hat{x}_{i+1}, ..., \hat{x}_N)$, ϱ_i are the Lagrange multipliers associated with the complicating (or coupling) constraints ϕ_i, and the ' ˆ ' (hat) notation stands for some fixed value of the variables.

In OCD, first the values $\hat{x}_i$ and $\hat{\varrho}_i$ are initialized. At the bottom level, x_i and ϱ_i can be optimized independently in SSi for $i = 1, ..., N$, as shown in Eqn. (3.9). At the top level, the difference of the optimized values x_i and the previous values $\hat{x}_i$ is calculated. If the difference is sufficiently small, the iteration stops. Otherwise, $\hat{x}_i$ and $\hat{\varrho}_i$ are updated with x_i and ϱ_i and the procedure continues to the next iteration.

3.3 Decentralized Approaches for Multi-subsystem Co-design Problems

In this section, a multi-level and a bi-level decentralized framework to solve multi-subsystem co-design problems are described. The direct collocation method [43] and the optimality condition decomposition method [83] are used in both frameworks.

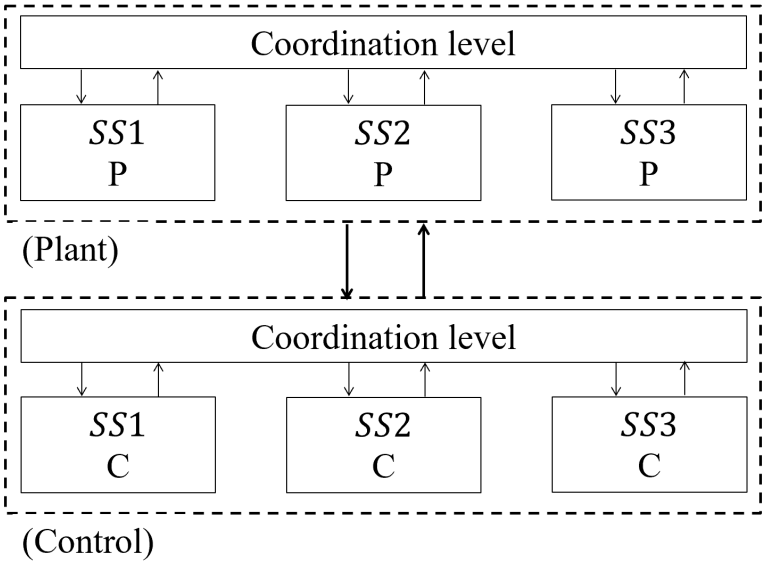


Figure 3.2: Scheme of Multi-level Decentralized Approach (A1)

3.3.1 Approach (A1): Multi-level Approach

To solve the multi-subsystem co-design problem (Eqn. (3.1)) in a decentralized manner, both the plant and control parts are decomposed and solved. A similar multi-level framework as in Section 2.2 in RQ1 is adopted, as shown in Fig. 3.2. In the plant design part, a dual decomposition approach [80] is used to solve for the plant design variables. In the control part, a decentralized implementation of the direct collocation method presented in Section 3.2 is employed.

In the multi-level framework, the plant and control parts are integrated by gradients of the Hamiltonian of the control part with respect to plant design variables at $\hat{y}_i$ and $\hat{y}_{s_{i,j}}$. Using a direct method, the state and control variables are discretized, hence the gradients are calculated by summation functions of variable values at the grid points (see below, step (S3)).

3.3.1.1 Approach Steps

The detailed steps, **(S1)** to **(S6)**, are as follows:

(S1) Initialize the plant design variable values y_{i0} and $y_{si,j0}$ of the i^{th} subsystem. Set the current variable values, $\hat{y}_i$ and $\hat{y}_{s_{i,j}}$, equal to y_{i0} and $y_{si,j0}$ for all i.

(S2) Apply the decentralized implementation of the direct collocation method in Section 3.2. Define x_i for the control and state variables of SSi at all grid points:

$$x_i = \left(u_{i,0}^T, ..., u_{i,M}^T, x_{i,0}^T, , ..., x_{i,M}^T \right)^T \tag{3.10}$$

The variable and multiplier values $\hat{x}_i$ and $\hat{\varrho}_i$, respectively, are initialized. At the bottom level of the control part, $\hat{y}_i$ and $\hat{y}_{s_{i,j}}$ are the current (fixed) values of the plant design variables. The SSi is then solved,

$$\min_{x_i} \quad w_{C_i} \bar{z}_{C_i}(x_i) + \sum_{(j,i)\in V} \sum_{k=0}^{M} \hat{\varrho}_{j,k}^T \phi_{j,k}(x_i, \hat{x}_j, \hat{y}_j, \hat{y}_{s_{j,i}})$$

subject to the constraints,

$$\begin{aligned} &\phi_{i,k}(x_i, \hat{x}_j, \hat{y}_i, \hat{y}_{s_{i,j}}) = 0 \\ &p_i(x_i) \leq 0, \quad q_i(x_i) = 0 \end{aligned} \tag{3.11}$$

to obtain $x_i^\ddagger$ and $\varrho_i^\ddagger$ for all SSi, where $x_i^\ddagger$ and $\varrho_i^\ddagger$ denote the optimal results of x_i and ρ_i in this step. At the top level, the difference between $x_i^\ddagger$ versus $\hat{x}_i$ is evaluated,

$$e = \sum_{i=1}^{N} \left(\| x_i^\ddagger - \hat{x}_i \|_2 \right) \tag{3.12}$$

If e is greater than some tolerance value ϵ_D, then $\hat{x}_i$ and $\hat{\varrho}_i$ are updated with $x_i^\ddagger$ and $\varrho_i^\ddagger$, and the iteration continues. Otherwise, the approach is considered to be converged for the control block. The approach continues to step (S3). The optimal variables and multipliers of the control part in the i^{th} subsystem are denoted as $x_i^\ddagger$ and $\varrho_i^\ddagger$.

(S3) The gradient of the Hamiltonian of the control part with respect to plant design variables at $\hat{y}_i$ and $\hat{y}_{s_{i,j}}$ can be obtained as follows,

$$\left.\frac{\partial H}{\partial y_i}\right|_{\hat{y}_i,\hat{y}_{s_{i,j}}} = \sum_{i=1}^{N} w_{C_i}\left[\sum_{k=0}^{M-1}\lambda_{i,k+1/2}\frac{\partial f_{i,k+1/2}(x_i^\ddagger)}{\partial y_i}\right] \tag{3.13}$$

$$\left.\frac{\partial H}{\partial y_{si,j}}\right|_{\hat{y}_i,\hat{y}_{s_{i,j}}} = \sum_{i=1}^{N} w_{C_i}\left[\sum_{k=0}^{M-1}\lambda_{i,k+1/2}\frac{\partial f_{i,k+1/2}(x_i^\ddagger)}{\partial y_{si,j}}\right] \tag{3.14}$$

$\lambda_{i,k+1/2}$ represents the approximating value of co-state variable at collocation points $t_{k+1/2}$,

$$\lambda_{i,k+1/2} = -\frac{3}{2}\psi_{i,k}\rho_{i,k}^\ddagger \quad (k = 0, 1, ..., M-1) \tag{3.15}$$

where $\psi_{i,k} = \frac{z_{C_{i,k+1/2}}}{2s}$ represents the scaling factor which depends on the discretization [43].

(**S4**) The dual variable vector is initialized: $\mu = 0$. For the i^{th} subsystem, the partial Lagrangian of the plant part is,

$$L_i = w_{P_i} z_{P_i}(y_i, y_{si,j}) + \eta_i^T g_i(y_i, y_{si,j}) + \gamma_i^T h_i(y_i, y_{si,j}) + \mu^T y_{si,j} \tag{3.16}$$

The plant design variables $y_i, y_{si,j}$ are then solved by the necessary optimality con-

ditions in the i^{th} subsystem,

$$\begin{aligned}
&\frac{\partial L_i}{\partial y_i} + \frac{\partial H}{\partial y_i}\bigg|_{\hat{y}_i, \hat{y}_{s_{i,j}}} = 0, \quad \frac{\partial L_i}{\partial y_{si,j}} + \frac{\partial H}{\partial y_{si,j}}\bigg|_{\hat{y}_i, \hat{y}_{s_{i,j}}} = 0 \\
&\frac{\partial L_i}{\partial \gamma_i} = 0 \\
&g_i(y_i, y_{s_{i,j}}) \leq 0, \quad \eta_i \geq 0, \quad \eta_i \circ g_i(y_i, y_{s_{i,j}}) = 0
\end{aligned} \tag{3.17}$$

where '$\circ$' denotes the Hadamard product.

(**S5**) Let $\bar{y}_s$ be the vector of all $y_{si,j}$ obtained in all subsystems. The consistency constraints can be represented as $\bar{y}_s = E\sigma$ [80], where E is a matrix with binary entries. $E_{ij} = 1$ when a plant variable is shared between SSi and SSj, and 0 otherwise. For the l^{th} iteration in the plant part, the approach computes the common values of shared plant design variables and evaluates the consistency constraint residual by

$$\sigma = (E^T E)^{-1} E^T \bar{y}_s \tag{3.18}$$

$$e = \|\bar{y}_s - E\sigma\|_2 \tag{3.19}$$

If $e > \epsilon_R$, it updates the dual variable vector via a subgradient method with step size β,

$$\mu = \mu + \beta \bar{y}_s \tag{3.20}$$

and repeat from Eqn. (3.16). Otherwise, the results in the plant part are obtained as y_{i-new} and $y_{si,j-new}$. Calculate the overall objective function value Z_{new} accordingly,

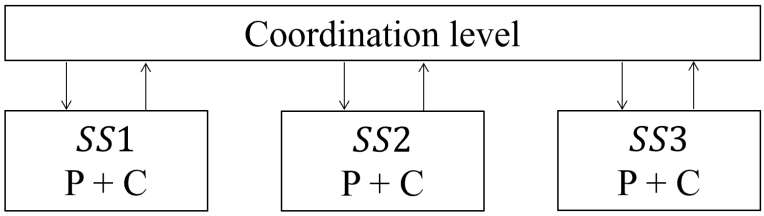


Figure 3.3: Scheme of Bi-level Decentralized Approach (A2)

and continue to the next step.

(**S6**) If $\|Z_{new} - Z_{current}\|_2 > \epsilon_Z$, update $\hat{y}_i$, $\hat{y}_{s_{i,j}}$ with the new plant design values y_{i-new}, $y_{si,j-new}$, and repeat from step (S2). Otherwise, the approach has converged and the solution is obtained.

3.3.2 Approach (A2): Bi-level Approach

In this framework, as shown in Fig. 3.3, the entire problem is decomposed in terms of subsystems and co-design of each subsystem. The solution steps are described on both cases when shared plant variables $y_{s_{i,j}}$ exist or do not exist.

3.3.2.1 Approach Steps (No Shared Plant Variables)

First, if there is no shared plant design variable, the decentralized implementation in Section 3.2 can be directly used. The bi-level framework is adopted from the optimality condition decomposition. At the bottom level, the co-design problem of each subsystem is solved by employing the direct collocation method. The optimizing variables are coordinated at the top level. The iterations stop when all variables converge.

The detailed steps **(S1)** to **(S3)** are described as in the following:

(S1) For each subsystem at the bottom level, define vector x_i for the optimizing variables in SSi as:

$$x_i = \left(u_{i,0}^T, ..., u_{i,M}^T, x_{i,0}^T, , ..., x_{i,M}^T, y_i^T \right)^T \tag{3.21}$$

which is the combining vector of the local control and state variables at all grid points, as well as the local plant design variables. Initialize the vectors of optimizing variables $\hat{x}_i$ and the vectors of Lagrange multipliers $\hat{\varrho}_{i,k}$ for all SSi.

(S2) By the decentralized implementation in Section 3.2, the formulation of SSi can be expressed as:

$$\min_{x_i} \quad w_{P_i} z_{P_i}(x_i) + w_{C_i} \bar{z}_{C_i}(x_i) + \sum_{(j,i)\in V} \sum_{k=0}^{M} \hat{\varrho}_{j,k}^T \phi_{j,k}(x_i, \hat{x}_j)$$

subject to the constraints,

$$g_i(x_i) \leq 0, \quad h_i(x_i) = 0 \tag{3.22}$$

$$\phi_{i,k}(x_i, \hat{x}_j) = 0$$

$$p_i(x_i) \leq 0, \quad q_i(x_i) = 0$$

where $\bar{z}_{C_i}$ is the converted control objective function. $x_i^\ddagger$ and $\varrho_{i,k}^\ddagger$ are then solved, denoting the optimal optimizing variables and the Lagrange multipliers associated with the complicating constraints $\phi_{i,k}$ for all SSi.

(S3) At the top level, check the difference of the previous optimizing variable values

$\hat{x}_i$ and the current optimized values $x_i^{\ddagger}$. If

$$\sum_{i=1}^{N} \| \left(\hat{x}_i - x_i^{\ddagger} \right) \|_2 < \epsilon_D \tag{3.23}$$

the converged co-design results are obtained and the iteration stops. Otherwise, update $\hat{x}_i$ and $\hat{\varrho}_{i,k}$ with $x_i^{\ddagger}$ and $\varrho_{i,k}^{\ddagger}$, and repeat from step (S2).

3.3.2.2 Approach Steps (with Shared Plant Variables)

When there are shared plant variables $y_{s_{i,j}(i)}$ between subsystems, both complicating variables and complicating constraints exist in the converted optimization problem by the direct method. Hence, a bi-level framework incorporating the optimality condition decomposition and the dual decomposition is devised. At the bottom level, the local state, control and plant design variables as well as a local copy of the shared plant variables are optimized in each subsystem. At the top level, the dual variables associated with the shared plant variables are updated. The iterations continue until the shared plant variable values converge.

The detailed steps **(S1)** to **(S3)** are described as in the following:

(S1) Similarly as in Section 3.3.1, in each subsystem at the bottom level, define vector x_i for the variables in the i^{th} subsystem as:

$$x_i = \left(u_{i,0}^T, ..., u_{i,M}^T, x_{i,0}^T, , ..., x_{i,M}^T, y_i^T \right)^T \tag{3.24}$$

In addition, the copy of the shared plant design variables is denoted as $y_{s_{i,j}(i)}$ in

the SSi. Initialize the vectors of optimizing variables $\hat{x}_i$, the vectors of Lagrange multipliers $\hat{\varrho}_{i,k}$, and the dual variable vectors $\nu_{i,j}$ for all subsystems.

(S2) By the decentralized implementation in Section 3.2, the formulation of SSi (i = 1, ..., N) can be expressed as:

$$\min_{x_i, y_{s_{i,j}(i)}} \quad w_{P_i} z_{P_i}(x_i, y_{s_{i,j}(i)}) + w_{C_i} \bar{z}_{C_i}(x_i) + \sum_{(i,j)\in V} \bar{\nu}_{i,j}^T y_{s_{i,j}(i)} + \sum_{(j,i)\in V} \sum_{k=0}^{M} \hat{\varrho}_{j,k}^T \phi_{j,k}(x_i, \hat{x}_j, \hat{y}_{s_{i,j}(j)})$$

subject to the constraints,

$$g_i(x_i, y_{s_{i,j}(i)}) \leq 0, \quad h_i(x_i, y_{s_{i,j}(i)}) = 0 \tag{3.25}$$

$$\phi_{i,k}(x_i, \hat{X}_j, y_{s_{i,j}(i)}) = 0$$

$$p_i(x_i) \leq 0, \quad q_i(x_i) = 0$$

where $\tilde{\nu}_{i,j} = \nu_{i,j}$ if $i < j$, and $\tilde{\nu}_{i,j} = -\nu_{i,j}$ if $i > j$. Hence, $x_i^\ddagger$ and $y_{s_{i,j}(i)}^\ddagger$ for the i^{th} subsystem can be solved.

(S3) At the top level, the consistency constraints $y_{s_{i,j}(i)}^\ddagger = y_{s_{i,j}(j)}^\ddagger$ are checked for the shared plant design variables. If

$$\sum_{(i,j)\in V} \| \left(y_{s_{i,j}(i)}^\ddagger - y_{s_{i,j}(j)}^\ddagger \right) \|_2 < \epsilon_{y_s} \tag{3.26}$$

the shared plant variable values have converged, and the co-design results are ob-

tained. Otherwise, update the dual variable vectors $\nu_{i,j}$ by

$$\nu_{i,j} = \nu_{i,j} - \beta(y^{\ddagger}_{s_{i,j}(i)} - y^{\ddagger}_{s_{i,j}(j)}) \tag{3.27}$$

where β is selected as a fixed step-size value. Also, $\hat{x}_i$ and $\hat{\varrho}_{i,k}$ are updated with $x^{\ddagger}_i$ and $\varrho^{\ddagger}_{i,k}$, and repeat from step (S2).

3.3.3 Notes on the Approaches

For the proposed multi-level approach (A1), since the co-design problem is decomposed into both plant and control parts, it is assumed that there exists an objective function for both parts for each subsystem to ensure all the solution steps are valid. If the decomposition techniques are not applied in both parts, the approach has the same structure as the 'nested' co-design strategy [1]. Hence, the proposed approach can be regarded as an extension of the nested strategy.

The bi-level approach (A2) aims to improve the performance of the multi-level approach. If no shared plant design variable exists, the convergence of the approach can be obtained by the convergence of the optimality condition decomposition method. In the presence of shared physical design variables, a bi-level approach is proposed to solve problems with both complicating variables and complicating constraints in a decomposed manner.

For both approaches, there is currently no convergence proof. The solution obtained can be a local optimum due to the non-convexity property of the considered problem. On the implementation side, the proposed approaches can be implemented

as black-box functions. While a guideline for selecting the tolerance values is not included, in general, if the tolerance values are made smaller, better optimized solutions are obtained, but then a longer computational time is required.

3.4 Examples

The decentralized approaches proposed in Section 3.3 are applied to a numerical and a spring-mass-damper system example in this section. In the numerical example, the dynamic system is linear and the objective function is quadratic, hence the decentralized approach in Chapter 2 is applicable. In the engineering example, a three-subsystem co-design problem with nonlinear dynamics is considered, which the approach in Section 2.2 is not applied. Also, a scalable version of the engineering example is studied. Results by the proposed decentralized approaches are compared against the centralized solution.

3.4.1 Numerical Example

The numerical example is slightly modified from the numerical example in Section 2.3. The problem has two subsystems, and m_1 to m_5 are the five physical plant design variables. The subsystems $SS1$ and $SS2$ are formulated as,

$$(SS1:)$$

$$\min_{m_1,m_2,m_3,m_4,u_1(t)} Z_1 = w_{P_1}\left((m_1-1)^2+(m_2-2)^2+\frac{1}{2}(m_3-1)^2+\frac{1}{2}(m_4-2)^2\right) + \frac{1}{2}w_{C_1}\int_0^1 \left(x_1^T(t)Q_1x_1(t)+R_1u_1^2(t)\right)dt$$

subject to the constraints,

$$g_1(m_1, m_2, m_3, m_4) := m_1^2 + m_2^2 + m_3^2 + m_4^2 - 8 \leq 0$$

$$\dot{x}_1(t) = \begin{bmatrix} 2 & m_1 \\ m_1 & -2m_2 \end{bmatrix} x_1(t) + \begin{bmatrix} 1 \\ 2 \end{bmatrix} u_1(t) + \begin{bmatrix} -0.01m_4 & 0 \\ 0.01 & -0.02 \end{bmatrix} x_2(t) \tag{3.28}$$

$$x_{10} = [-1, 0.1]^T, \quad Q_1 = \text{diag}(2, 1), \quad R_1 = 1$$

$(SS2:)$

$$\min_{m_3, m_4, m_5, u_2(t)} Z_2 = w_{P_2}\left(\frac{1}{2}(m_3 - 1)^2 + \frac{1}{2}(m_4 - 2)^2 + (m_5 - 3)^2\right) + \frac{1}{2} w_{C_2} \int_0^1 \left(x_2^T(t) Q_2 x_2(t) + R_2 u_2^2(t)\right) dt$$

subject to the constraints,

$$h_2(m_3, m_4, m_5) := m_3 + m_4 + 2m_5 - 8 = 0$$

$$\dot{x}_2(t) = \begin{bmatrix} -m_5 & 0.5 \\ 2 & -m_5 \end{bmatrix} x_2(t) + \begin{bmatrix} 2 \\ 5 \end{bmatrix} u_2(t) + \begin{bmatrix} 0.02 & 0.01 \\ 0 & -0.01m_3 \end{bmatrix} x_1(t) \tag{3.29}$$

$$x_{20} = [1, -0.5]^T, \quad Q_2 = \text{diag}(1, 2), \quad R_2 = 2$$

The weights in the objective function are set as $w_{P_1} = w_{P_2} = w_{C_1} = w_{C_2} = 0.5$. Local design variables in the two subsystems are defined as $y_1 = [m_1, m_2]^T$ and $y_2 = m_5$, respectively. The shared design variable between the subsystems $y_{s1,2} = y_{s2,1} = [m_3, m_4]^T$. It can be verified that for all values of m_1 to m_5, the two subsystems are controllable.

The problem is then solved by the proposed decentralized approaches, as well

	y_1^*	y_2^*	$y_{s1,2}^*$	Z^*
Centralized	$[1.11, 1.80]^T$	2.75	$[0.79, 1.70]^T$	0.91
Decentralized (RQ1)	$[1.13, 1.78]^T$	2.75	$[0.79, 1.71]^T$	0.93
Decentralized (A1)	$[1.14, 1.78]^T$	2.75	$[0.79, 1.71]^T$	0.93
Decentralized (A2)	$[1.11, 1.80]^T$	2.75	$[0.79, 1.70]^T$	0.91

Table 3.1: Example 1: Optimal Solutions

as the centralized (or all-at-once) method by TOMLAB [81], which solves for the variables from all subsystems simultaneously. Since the dynamic equations is linear in the example, it can also be solved with the decentralized approach proposed in Chapter 2. All the tolerance values are selected as 0.01.

The optimal solutions by the centralized and decentralized approaches are shown in Table 3.1. It is shown that the optimal solution obtained by each of the decentralized approach is close to the centralized solution. In particular, the solution by the bi-level approach is exactly identical to the centralized one.

In Fig. 3.4, the optimal control variables of $SS1$ and $SS2$ are plotted. As is observed, the decentralized optimal control solutions are again close to the centralized one, which demonstrates the validity of the proposed approaches. Finally, various feasible and infeasible initial points of the physical variables are used in the simulation, and the approximately identical converged results are obtained for all tests.

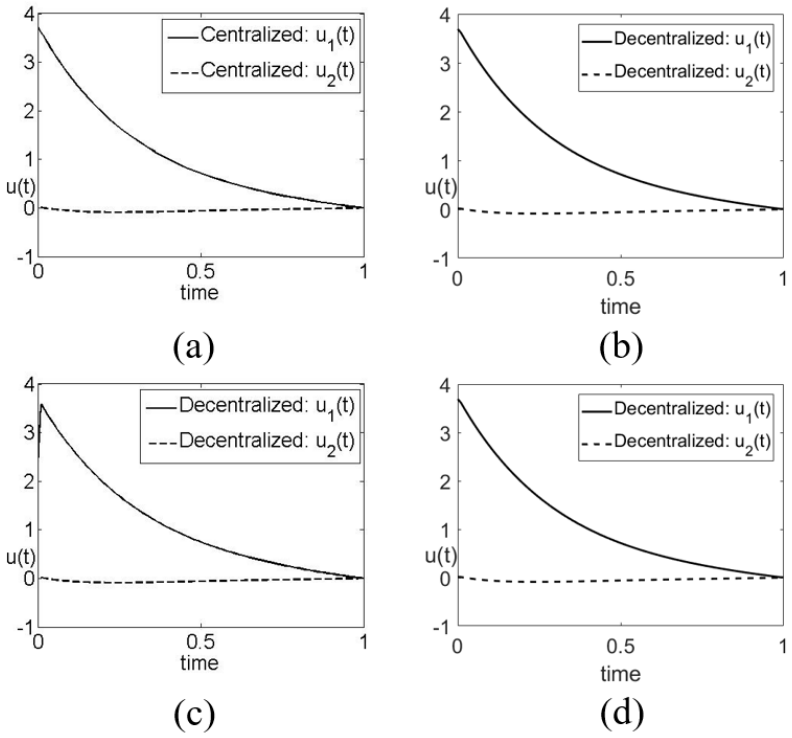


Figure 3.4: Example 1: Optimal Controllers, (a) Centralized, (b) Decentralized: Multi-level (RQ1), (c) Decentralized: Multi-level, (d) Decentralized: Bi-level

3.4.2 Engineering Example: Spring-mass-damper System Model

A similar engineering example in Chapter 2 is considered, as shown in Fig. 3.5, however, here nonlinearity is introduced into the dynamic equations. The plant design part of each subsystem consists of one mass (m), one spring (k) and one damper (c). The control part has a quadratic objective function. The state variable $x_i(t) = [x_{i1}(t), x_{i2}(t)]^T$ denotes the displacement and velocity of the mass m_i, and the control variable $u_i(t)$ is the force applied to the mass from 0 to 5 seconds.

The formulation of the plant design part has been adopted from the spring design optimization formulation [82]. The wire diameters of springs are selected as the plant design variables, i.e., $y_i = d_i$ for all i. The objective function in the i^{th}

subsystem, z_{P_i}, is to maximize energy storage capacity. It is shown in [82] that if the spring index C is fixed at the maximum value, then the optimal spring design can be found by solving the following optimization problem,

$$\begin{aligned}
&\min_{y_i} \quad z_{P_i} = (10.24G)\left(\frac{\pi}{12.8F_u}\right)^2 C^{-2} y_i^4 \\
&\text{subject to the constraints,} \\
&C^{-1} - \kappa_1 C^{-1} y_i^{-1} \leq 1 \qquad \text{(Inside diameter)} \\
&\kappa_2 y_i^{-1} \leq 1 \qquad \text{(Lower bound on wire diameter)} \\
&\frac{0.8F_u}{D_s G} C^3 y_i^{-2} \leq 1 \qquad \text{(Clash allowance)}
\end{aligned} \tag{3.30}$$

where G is shear modulus (MPa), F_u is maximum allowable force (N), C is spring index (dimensionless), κ_1 is minimum allowable inside diameter (m), κ_2 is the lower bound on the plant design variables (m), and D_s is a clearance constant (dimensionless). The spring constants can be calculated as [4],

$$k_i = \frac{d_i^4 G}{8 D_i^3 N_a \left(1 + \frac{1}{2C^2}\right)} \tag{3.31}$$

where $D_i = C d_i$ is coil diameter, and N_a is the number of active coils of the spring. With the proposed decentralized approaches using the direct collocation method, nonlinear dynamic equations can be handled in the co-design problems. In this example, a nonlinear force versus deflection for the springs is considered. In other words, it is assumed that each spring is increasingly less resilient as being elongated

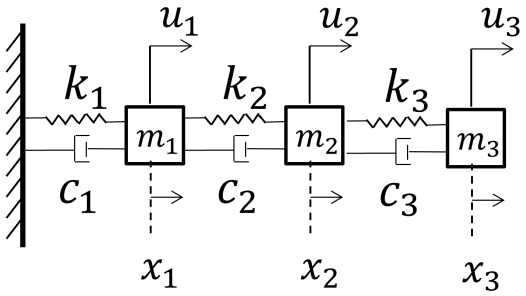


Figure 3.5: Example 2: Spring-mass-damper System Model

or depressed, hence the response force of the i^{th} spring can be expressed as

$$F_{k_i} = -k_i x_{i1} + \theta x_{i1}^3$$

where $\theta = 0.5$ is selected in this example. Then the co-design problem is formulated as the following,

$$\min_{y_i, u_i(t)} Z = \sum_{i=1}^{3} w_{P_i}(10.24G)\left(\frac{\pi}{12.8F_u}\right)^2 C^{-2} y_i^4 + \frac{1}{2}\sum_{i=1}^{3} w_{C_i}\int_0^5 \left(x_i^T(t)Q_i x_i(t) + R_i u_i^2(t)\right)dt$$

subject to the constraints,

$$(C^{-1}-1)y_i - \kappa_1 C^{-1} \le 0, \quad \kappa_2 - y_i \le 0$$

$$\frac{0.8F_u}{D_s G} C^3 y_i^{-2} \le 1, \quad x_{i0} = [1,1]^T$$

$$\dot{x}_{11} = x_{12}, \quad \dot{x}_{21} = x_{22}, \quad \dot{x}_{31} = x_{32}$$

$$u_1 = m_1\dot{x}_{12} + c_1 x_{12} + c_2(x_{12} - x_{22}) + k_1 x_{11} + k_2(x_{11} - x_{21})$$
$$- \theta x_{11}^3 - \theta(x_{11} - x_{21})^3 \tag{3.32}$$

$$u_2 = m_2\dot{x}_{22} + c_2(x_{22} - x_{12}) + c_3(x_{22} - x_{32}) + k_2(x_{21} - x_{11}) + k_3(x_{21} - x_{31})$$
$$- \theta(x_{21} - x_{11})^3 - \theta(x_{21} - x_{31})^3$$

$$u_3 = m_3\dot{x}_{32} + c_3(x_{32} - x_{22}) + k_3(x_{31} - x_{21}) - \theta(x_{31} - x_{21})^3$$

$$Q_i = \text{diag}(1, 1), \quad R_i = 1 \quad (i = 1, 2, 3)$$

The parameters are selected as $c_1 = 10$, $c_2 = 15$, $c_3 = 20$, and $m_1 = m_2 = m_3 = 5$, $G = 30$ (for chrome silicon), $C = 8$, $F_u = 6$, $\kappa_1 = 0.05$, $\kappa_2 = 0.1$, $D_s = 0.5$, and $N_a = 200$. All the tolerance values are selected as 0.01.

The problem is solved by the decentralized approaches and by the centralized approach using TOMLAB [81]. The solutions for the optimal wire diameters are shown in Table 3.2. It is shown that the optimal solution of the bi-level approach is identical to that of the centralized method. The optimal solution of the multi-level approach is close but slightly worse than the other two. The profiles of optimal controllers $u_i(t)$ obtained by all approaches are shown in Fig. 3.6. As is observed, the optimal control variables of the decentralized approaches are comparable with the ones of the centralized approach.

	y_1^*	y_2^*	y_3^*	Z^*
Centralized	0.13	0.10	0.10	10.14
Decentralized (A1)	0.10	0.10	0.10	10.58
Decentralized (A2)	0.13	0.10	0.10	10.14

Table 3.2: Example 2: Optimal Solutions

3.4.2.1 Computational Cost for Increasing Number of Subsystems

In this section, the previous spring-mass-damper model is reformulated as a scalable problem, as shown in Fig. 3.7. For the simplicity of the implementation,

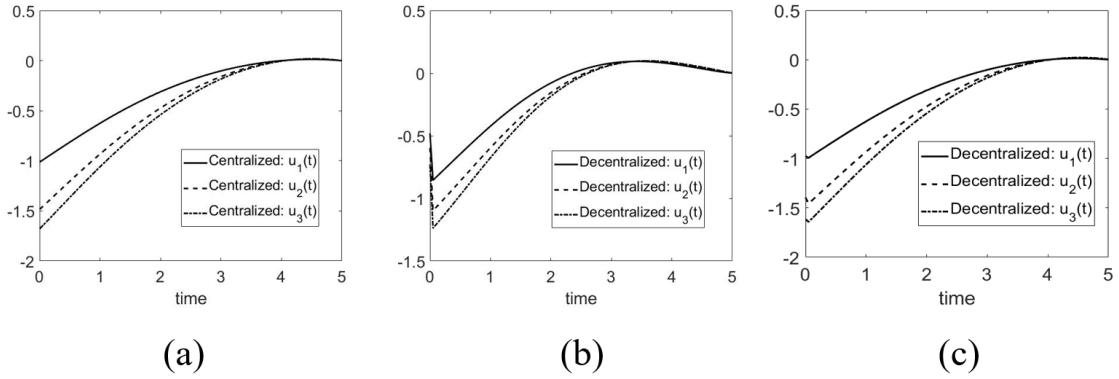


Figure 3.6: Example 2: Optimal Controllers, (a) Centralized, (b) Decentralized: Multi-level, (c) Decentralized: Bi-level

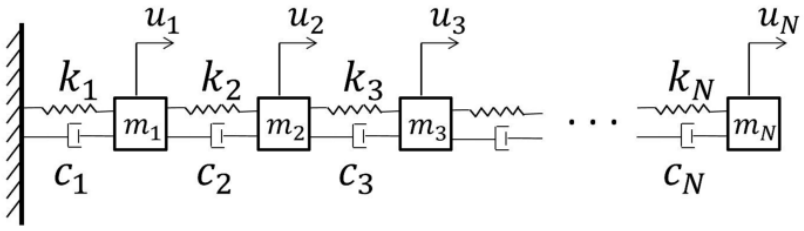


Figure 3.7: Series of Spring-mass-damper Subsystems

linear dynamic equations are used in this case , with all masses $m_i = 5$ and damping coefficients $c_i = 10$. The objective function of the plant design part of each subsystem is assumed to be $(y_i - 0.1)^2$. There exists a control input for each mass. The number of subsystems considered for each test problem is 5, 10, 20 to 80. Each test problem is solved in both the centralized (by TOMLAB [81]) and the proposed decentralized approaches.

Fig. 3.8 shows a comparison of the computational (CPU) time for the test problems when no communication costs and delays are considered. The centralized results are given by the CPU time obtained from TOMLAB [81], which, as expected, grow nonlinearly as the number of variables is increased.

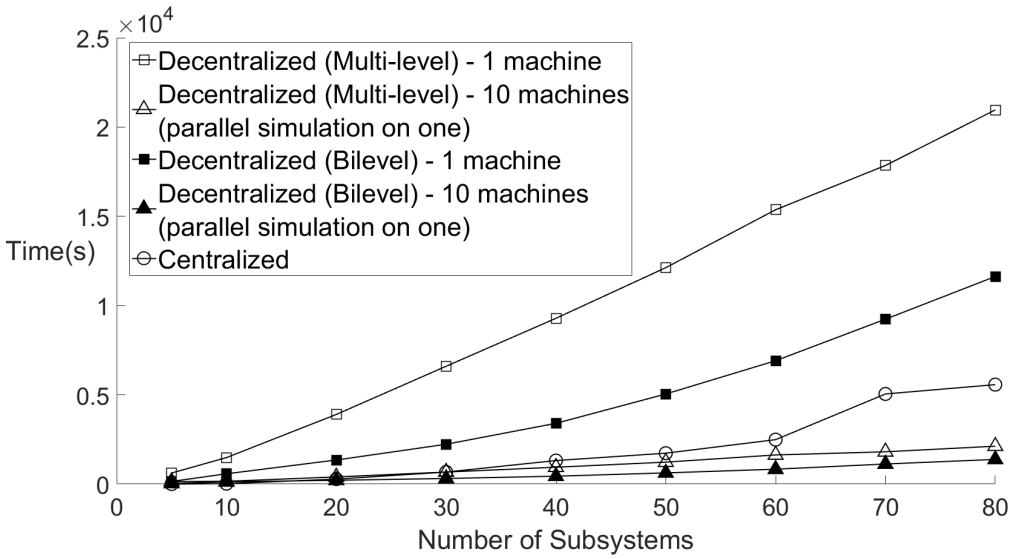


Figure 3.8: Comparison of Computational Time between Centralized and Decentralized Approaches

Using the decentralized approach, the results show that the computational time increases about linearly as the number of subsystems is increased. However, if all steps in this method are simulated on a single machine, i.e., subsystems are optimized sequentially one after another, the computational time values are greater than the centralized ones, as shown in Fig. 3.8. Two reasons can cause this phenomenon. First, due to the decomposition nature of the proposed approaches, more computation steps are performed. A longer computational time is expected if all steps are operated on a single machine. Second, the Matlab program written by the author is not as efficient as TOMLAB.

Hence, a 'parallelization' procedure is simulated, which assumes that some number of 'machines' can be used simultaneously. The machines are used to solve the decomposed subproblems in a parallel manner. As an example, this simulation can be utilized when multiple cores of a CPU are used for calculation, or when the

entire system is structurally impossible to be solved simultaneously. In this example, all the decomposed subproblems are divided into groups of 10 machines, so each group can be solved 'in parallel'. The largest value of the running time among the 10 machines is measured. By this time, all other 9 machines in the same group are assumed to have finished the tasks of optimizing the corresponding subproblems. Following this procedure, a second computational time plot is presented for this test problem, which shows that both the multi-level and the bi-level decentralized approaches are able to perform better when the number of subsystems increases. In addition, it is shown that the computational time of the bi-level approach is better than the multi-level one.

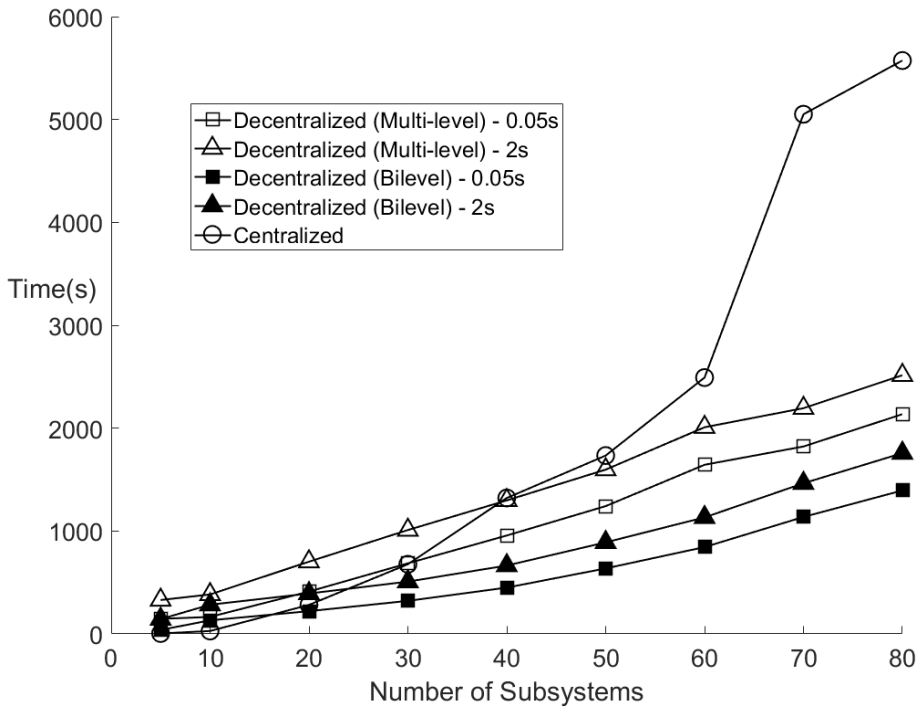


Figure 3.9: Comparison of Computational Time between Centralized and Decentralized Approaches with Various Communication Costs

In a parallel computing environment, communication costs and delays may be

important factors to consider, therefore a simulation is implemented to include communication costs among 10 machines in this example. A constant value, simulating the time to transfer information together with delays, is added in each iteration for the decentralized approaches. In Fig. 3.9, the constant value is selected as 0.05s and 2s, respectively. As is observed, when the number of subsystems is large, the decentralized approaches can outperform the centralized one.

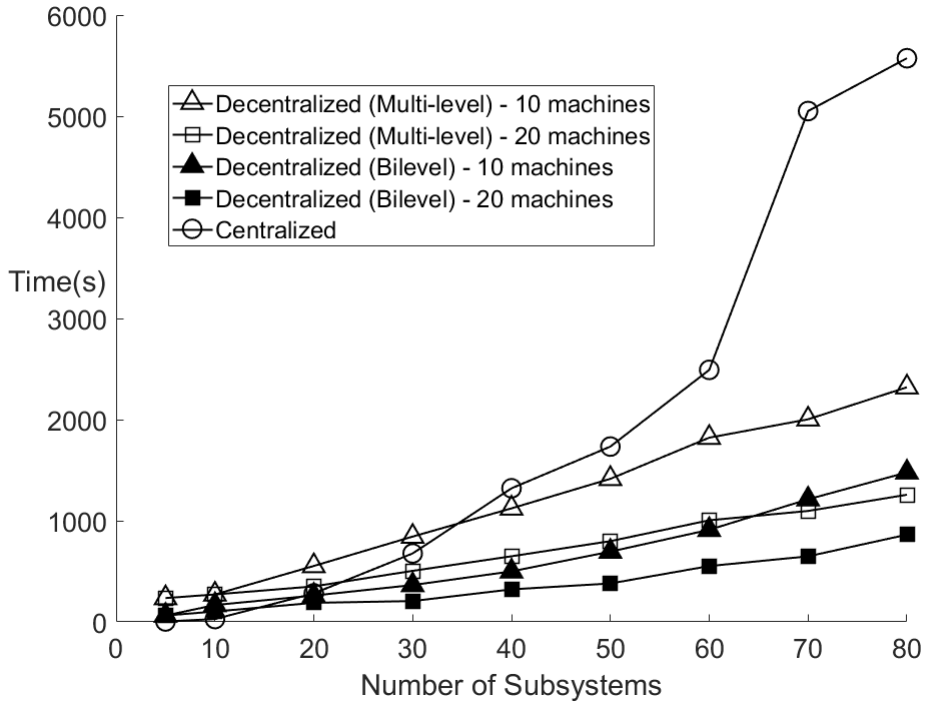


Figure 3.10: Comparison of Computational Time between Centralized and Decentralized Approaches with Different Computational Power and Constant Communication Costs

Finally, the number of machines can also influence the computational time of the decentralized approaches. If more subsystems can be optimized in parallel simultaneously, the total computational time to obtain optimal solution can be reduced. In the last simulation, the number of 'machines' is set as 10 and 20, and the time to transfer information with delays in each iteration is assumed to be 1s.

It is shown in Fig. 3.10 that as more machines are used at the subsystem level, the total computational time can be further reduced. On the contrast, despite that there is no communication cost during the optimization process, a centralized approach lacks the capability of utilizing more computational power when the problem has large number of variables. In all the tests, the crossing points of the computational time curves can provide useful information for making decisions.

3.5 Summary

In this chapter, two new decomposition-based numerical approaches are developed to solve more general multi-subsystem co-design problems. The direct collocation method is used to discretize the continuous-time dynamic constraints, and is incorporated with the optimality condition decomposition method to solve the problems in a decentralized manner. The proposed approaches are applied to a numerical and an engineering example, and the solutions are compared with the centralized one.

Finally, a scalable test problem is considered to show a potential advantage of the proposed decentralized approaches. It is shown that by simulating multiple parallel machines to solve the decomposed subproblems, the decentralized approaches take a shorter computational time than the centralized approach as the size of the problem increases. Besides, it can be observed that the bi-level approach outperforms the multi-level one in terms of both solution quality and the computational time.

In the next chapter, the results from Research Question 3 (RQ3) are presented. In RQ3, an idea about optimizing the combined routing and control costs of a fleet of vehicles is explored from a co-design perspective.

Chapter 4: Integrating Vehicle Routing and Control: A Comparison of Three Solution Strategies

In this chapter, the results from Research Question 3 are presented in detail.[1]

The organization of this chapter is as follows. This chapter begins with a nomenclature section. Then in Section 4.1, background knowledge of the Vehicle Routing Problems (VRPs), vehicle control model, a formulation of optimal control problem for a vehicle to travel along an edge and the multiple shooting method is presented. Three strategies for solving the proposed problem and a comparison of the results from different strategies are presented in Section 4.2. In Section 4.3, results are presented for a test example based on a benchmark problem dataset [85]. In Section 4.4, the optimal results of VRPs with different wind directions are discussed. Finally, a summary is given in Section 4.5.

[1]This chapter is based on the paper: Liu, T., Azarm, S., and Chopra, N., 2018, "Integrating Multi-Vehicle Routing and Control: A Comparison of Three Solution Strategies." Proceedings of the ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference. Quebec City, Quebec, Canada. August 26–29, 2018.

Nomenclature

A	Vehicle's wind resisting cross-sectional area (m^2)
C_d	Air drag resistance coefficient
C_r	Rolling resistance coefficient
c	Edge cost
D	Edge length (m)
g	Gravitational constant (m/s^2)
J	Joint cost
l	Vehicle's load (kg)
m	Vehicle's mass (kg)
N_C	Number of customers
N_V	Number of vehicles
Q	Vehicle capacity

r	Demand at customer's location
t_f	Final time to reach destination (s)
u	Vehicle control variable: forward force (N)
v	Vehicle speed (m/s)
$\dot{v}$	Vehicle acceleration (m/s^2)
v_s	Speed limit (m/s)
v_w	Wind speed (m/s)
v_{wp}	Projected wind speed along vehicle's velocity direction (m/s)
w_1,w_2	Weighting coefficients associated to time, control objectives, respectively
x	Vehicle displacement (m)
y	Binary variable for route selection
θ	Road angle (rad)

η Vehicle efficiency

ρ Air density (kg/m^3)

τ Normalized time horizon

φ Angle between wind and vehicle heading directions (rad)

ω Weighting parameter

4.1 Background

4.1.1 Vehicle Routing Problems (VRPs)

VRPs can generally be represented by a directed graph of nodes and edges, as shown in Fig. 4.1. The nodes in the graph represent customer locations that the vehicles must visit and serve, and the edges represent routes that the vehicles travel between the nodes. Single depot is assumed in this thesis. Each customer has a fixed demand r_i (e.g., units of products to be collected) and must be visited only once by a single vehicle k from a fleet of vehicles. M and N are the sets of vehicles and customer locations, respectively. For an edge that connects node i and j, the edge cost is denoted as c_{ij}. By minimizing the sum of costs for all edges and for all vehicles, each vehicle of the fleet starts from the depot, visits a subset of customers, and finally returns to the depot. VRPs are usually formulated and solved as a mixed-integer programming problem [48],

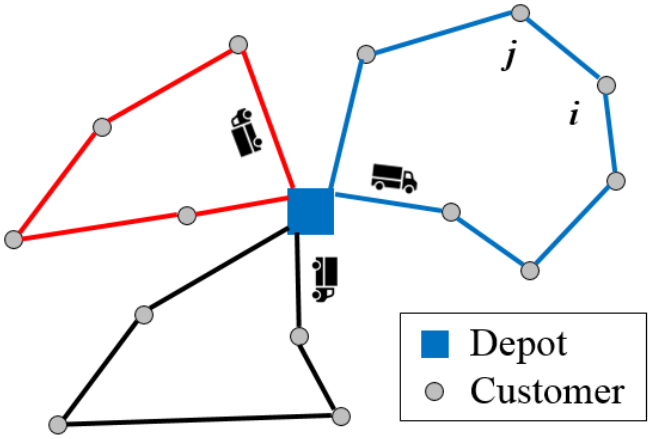


Figure 4.1: Vehicle Routing Problems (VRPs)

$$\min \sum_{k\in M}\sum_{i\in N}\sum_{j\in N} c_{ij}y_{ijk} \tag{4.1}$$

subject to,

$$\sum_{k\in M}\sum_{j\in N} y_{ijk} = 1, \quad \forall i \in N\backslash\{0\} \tag{4.2}$$

$$\sum_{j\in N} y_{0jk} = 1, \quad \forall k \in M \tag{4.3}$$

$$\sum_{i\in N} y_{ihk} - \sum_{j\in N} y_{hjk} = 0, \quad \forall k \in M, \forall h \in N\backslash\{0\} \tag{4.4}$$

$$\sum_{i\in N} y_{i(n+1)k} = 1, \quad \forall k \in M \tag{4.5}$$

$$\sum_{i,j\in S} y_{ijk} \leq |S| - 1, \quad \forall S \subseteq N, \forall k \in M \tag{4.6}$$

$$\sum_{i\in N} r_i \sum_{j\in N} y_{ijk} \leq Q_k, \quad \forall k \in M \tag{4.7}$$

Eqn. (4.1) is the objective function that minimizes the total cost of all vehicles. y_{ijk} is a binary variable, which equals 1 if node j is visited directly after node i by

vehicle k, and 0 otherwise. Eqn. (4.2) indicates that every node is only visited once. Eqns. (4.3) - (4.5) constrain each vehicle to start from the depot, arrive and depart from an assigned set of nodes and finally return to the depot. Note that the subscripts 0 and $n+1$ refer to the depot. Finally, Eqn. (4.6) is for the subtour elimination constraints, where S is a subset of nodes. In the capacitated vehicle routing problems (CVRPs), Eqn. (4.7) is added as the capacity constraints of the vehicles, where Q_k represents the capacity limit.

4.1.2 Vehicle Control Model

In this subsection, the vehicle control model used in RQ3 and RQ4 is described. For simplicity, all vehicles considered in the research are assumed to be homogeneous. For each vehicle, $v(t)$ is the vehicle's speed, and $u(t)$ is the control variable for the forward force exerted on the vehicle. $u(t) < 0$ means a backward force is exerted. The vehicle dynamics is shown as follows,

$$\dot{v}(t) = \frac{1}{m+l}\left(-(m+l)g\sin\theta - (m+l)gC_r\cos\theta - \frac{1}{2}\rho A C_d v^2(t) + u(t)\right) \tag{4.8}$$

$$= \frac{1}{m+l}(-\gamma - \beta v^2(t) + u(t)) \tag{4.9}$$

where $\gamma = (m+l)g\sin\theta + (m+l)gC_r\cos\theta$, representing the force to overcome rolling friction and gravity forces, and $\beta v^2(t) = \frac{1}{2}\rho A C_d v^2(t)$, representing the force to overcome air drag resistance.

It is assumed that the vehicle's deceleration is caused by a negative force

(i.e., $u(t) < 0$). The energy used for this deceleration also accounts for the energy consumption. This assumption is made so that the optimal control cost can be calculated without considering other factors like braking effects. Hence, the energy consumption W_C is calculated as,

$$W_C = \frac{1}{\eta} \int_0^{t_f} (u^2(t)v^2(t))^{\frac{1}{2}} dt \tag{4.10}$$

where η is a constant efficiency representing the ratio of the output work to move the vehicle forward over the input work or energy consumed.

Note that this vehicle dynamic model can also be derived from the vehicle power model as in [59, 60, 61], where it is assumed that the speed of the vehicle is constant, and the acceleration is zero. Since the speed and acceleration are assumed to be time-varying in RQ3 and RQ4, the output work equals the integration of the power over the entire time horizon, which provides the exact same dynamic model as in Eqns. (4.8) and (4.9).

It should be noted that this vehicle dynamic model is extremely simplified and ignores many parts, including the dynamics of powertrain, tires, suspension systems, etc.

4.1.3 Optimal Control Problem for a Vehicle Traveling along an Edge

In RQ3 and RQ4, a continuous-time optimal control problem for a vehicle to travel between two customer locations is introduced.

Two objective functions are considered for the proposed optimal control problem: the traveling time and the energy consumption. A weighted sum technique is used to compose the overall objective function, which is referred to as the "joint cost" in RQ3 and RQ4

$$J = w_1 \times \int_0^{t_f} 1\, dt + w_2 \times W_C = \int_0^{t_f} \left(w_1 \times 1 + w_2 \times \eta (u^2(t) v^2(t))^{\frac{1}{2}} \right) dt \quad (4.11)$$

where w_1 and w_2 are the weights associated with the two objectives. The optimizing variables are the optimal controller $u(t)$ and the final time t_f.

In terms of the path constraints, the distance between two nodes is equal to D, hence the displacement of the vehicle at the final time is D. The initial and final speed are both assumed to be 0. In the route, the speed is always positive and less than a pre-specified speed limit v_{ub}. The control variable $u(t)$ is assumed to be bounded within the range $[u_{lb}, u_{ub}]$.

In RQ3, a weighting parameter is introduced as $\omega = \frac{w_2}{w_1}\eta$, then the objective function is written as,

$$J = \int_0^{t_f} w_1 \left(1 + \omega (u^2(t) v^2(t))^{\frac{1}{2}} \right) dt \quad (4.12)$$

If η (or vehicle efficiency) is fixed, different choices in the value of ω provide different weightings for the two objectives. In this formulation, the magnitude of the energy consumption portion of the objective function is much larger than that of the time factor, hence a small value of ω is usually selected, which will be explored further

in the next sections. Hence, the proposed continuous-time optimal control problem is formulated as follows,

$$\min_{u(t),t_f} \int_0^{t_f} \left(1 + \omega(u^2(t)v^2(t))^{\frac{1}{2}}\right) dt \tag{4.13}$$

subject to,

$$\dot{x}(t) = v(t) \tag{4.14}$$

$$\dot{v}(t) = \frac{1}{m+l}(-\gamma - \beta v^2(t) + u(t)) \tag{4.15}$$

$$x(0) = 0, \quad x(t_f) = D, \quad v(0) = 0, \quad v(t_f) = 0 \tag{4.16}$$

$$0 \leq v(t) \leq v_{ub}, \quad u_{lb} \leq u(t) \leq u_{ub} \tag{4.17}$$

Eqn. (4.13) is the weighted-sum cost of travel time and energy consumption. Also, since w_1 is a constant, it is neglected in the above optimization problem. Eqns. (4.14) and (4.15) are the vehicle dynamics, Eqn. (4.16) provides the initial and final values of the displacement and speed. Eqn. (4.17) dictates the path constraints on the state and control variables. The parameters for the vehicle's dynamic model are fixed as follows: $g = 9.81 m/s^2$, $\theta = 0$ for all roads, $A = 5m^2$, $\rho = 1.2041 kg/m^3$, $C_r = 0.01$, $C_d = 0.7$, $u_{lb} = -6000N$, and $u_{ub} = 6000N$. Each vehicle has a fixed static load during the entire trip, $m + l = 3000kg$.

In RQ4, W_C in Eqn. (4.10) is replaced by $\int_0^{t_f} u^2(t)v^2(t)dt$, and the two objectives are properly scaled. Hence, the proposed problem is formulated as follows,

$$\min_{u(t),t_f} \int_0^{t_f} \left(w_1 \times \frac{1}{10^3} + w_2 \times \frac{u^2(t)v^2(t)}{10^{11}}\right) dt \tag{4.18}$$

subject to,

$$\dot{x}(t) = v(t) \tag{4.19}$$

$$\dot{v}(t) = \frac{1}{m+l}(-\gamma - \beta v^2(t) + u(t)) \tag{4.20}$$

$$x(0) = 0, \quad x(t_f) = D, \quad v(0) = 0, \quad v(t_f) = 0 \tag{4.21}$$

$$0 \leq v(t) \leq v_{ub}, \quad u_{lb} \leq u(t) \leq u_{ub} \tag{4.22}$$

where w_1 and w_2 are the weights associated with the two objectives, where $w_1 \geq 0$, $w_2 \geq 0$, and $w_1 + w_2 = 1$. Note that the two objectives are normalized by dividing the approximate magnitude of each objective.

4.1.4 Multiple Shooting Method

The multiple shooting method, e.g., [37, 84, 86], is implemented to solve the proposed problems in RQ3 and RQ4 in Section 4.1.3. Using this method, the time horizon $[0, t_f]$ is discretized into subintervals by grid points as $0 = t_0 < t_1 < t_2 < \cdots < t_N = t_f$. Since the final time t_f is unknown in this problem, the time horizon is scaled by dividing by t_f, hence the scaled final time is 1, i.e., $0 = \tau_0 < \tau_1 < \tau_2 < \cdots < \tau_N = 1$. The state and control variables are then approximated by discrete values at grid points and solved by a nonlinear optimization method. Since the multiple shooting method is a numerical approach to solve the optimal control problem, there is no guarantee of global optimality of the solution. For the proposed problems in Section 4.1.3, it is tested that the overshooting method provides the exact same solution as the direct collocation method, another numerical method

which is often used in an optimal control problem.

4.2 Solution Strategies

In the previous papers (e.g., [60, 61]), the speed of vehicles is usually assumed to be constant. While this assumption makes the problem easier to solve, traveling at a constant speed may not be the best option for saving energy or minimizing the joint cost. Since the speed of the vehicle can be changed with time, it is of interest to determine whether the use of an optimal controller can reduce the joint cost.

Three solution strategies for evaluating the joint cost of the entire routing-and-control problem are presented in this section, as shown in Fig. 4.2.

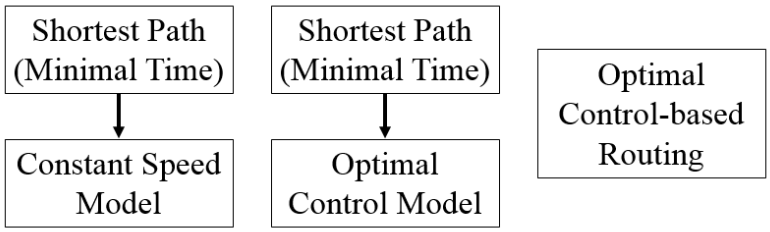


Figure 4.2: Schemes of Solution Strategies

4.2.1 Solution Strategy 1: Shortest Route Followed by Constant Speed

In the first strategy (Fig. 4.2, left), the routes are determined by the shortest traveling distance for all vehicles, which is obtained by solving the VRP (Eqns. (4.1) – (4.7)). Once the routes are determined, vehicles are operated at a constant speed to run from one node to another. In prior literature, the acceleration and deceleration of

vehicles have not been considered in the control cost function evaluation. However, since the vehicle speed is 0 at the starting and final positions, for the sake of making a fair comparison of the energy consumption with an optimal control model (see subsequent subsections), the constant speed model is slightly revised to consist of the following three stages, (S1) to (S3), as the vehicle travels along an edge: (S1) The vehicle accelerates from 0 to v_{ub} with a maximum input u_{ub}. (S2) The vehicle maintains the speed at v_{ub}. (S3) The vehicle decelerates from v_{ub} to 0 with a negative maximum input u_{lb}. Note that in stages (S1) and (S3) the maximum magnitude of the input is utilized. This strategy guarantees that the vehicle always reaches the destination in the shortest possible time. To evaluate the energy consumption, the time and traveled distance of all stages should be calculated. Given the vehicle dynamic model,

$$\frac{dv}{dt} = \frac{1}{m+l}(-\gamma - \beta v^2 + u) \tag{4.23}$$

The time and distance in stage (S1) are respectively denoted as t_{S1} and x_{S1}, and can be found by

$$t_{S1} = \int_0^{v_{ub}} \frac{1}{\frac{1}{m+l}(-\gamma - \beta v^2 + u_{ub})} dv \tag{4.24}$$

$$x_{S1} = \int_0^{v_{ub}} \frac{v}{\frac{1}{m+l}(-\gamma - \beta v^2 + u_{ub})} dv \tag{4.25}$$

Similarly, for stage (S3), t_{S3} and x_{S3} can be found by

$$t_{S3} = \int_{v_{ub}}^{0} \frac{1}{\frac{1}{m+l}(-\gamma - \beta v^2 + u_{lb})} dv \tag{4.26}$$

$$x_{S3} = \int_{v_{ub}}^{0} \frac{v}{\frac{1}{m+l}(-\gamma - \beta v^2 + u_{lb})} dv \tag{4.27}$$

The distance for stage (S2) is the difference of the total distance and the ones in (S1) and (S3),

$$x_{S2} = D - x_{S1} - x_{S3} \tag{4.28}$$

and the vehicle travels at the constant speed limit, hence the time for (S2) is

$$t_{S2} = \frac{x_{S2}}{v_{ub}} \tag{4.29}$$

with the acceleration in stage (S2) being 0,

$$\frac{dv}{dt} = \frac{1}{m+l}(-\gamma - \beta v_{ub}^2 + u_{S2}) = 0 \tag{4.30}$$

$$u_{S2} = \gamma + \beta v_{ub}^2 \tag{4.31}$$

Hence, the joint cost for the vehicle to travel along one edge is calculated as

$$\begin{aligned} J &= \int_0^{t_f} \left(1 + \omega (u^2(t)v^2(t))^{\frac{1}{2}}\right) dt \\ &= \int_0^{t_f} 1\, dt + \omega \int_0^{t_f} \left((u^2(t)v^2(t))^{\frac{1}{2}}\right) dt \\ &= t_{S1} + t_{S2} + t_{S3} + \omega(|u_{ub}|x_{S1} + |u_{S2}|x_{S2} + |u_{lb}|x_{S3}) \end{aligned} \tag{4.32}$$

4.2.2 Solution Strategy 2: Shortest Route Followed by Optimal Control

The first step of this strategy (Fig. 4.2, middle) is identical to that of Solution Strategy 1, i.e., finding the routes for vehicles to achieve the shortest total traveling distance. Following this first step, the optimal control problem (Eqns. (4.13) – (4.17)) is solved to obtain the total joint cost.

4.2.3 Solution Strategy 3: Optimal Control-based Route

In the third strategy (Fig. 4.2, right), a continuous-time optimal control problem of vehicles is considered. Here, the edge cost c_{ij} in the VRP is assumed to be the joint cost obtained by employing the optimal control rule for one vehicle traveling from node i to node j, which is the weighted sum of the minimum time and energy consumption objectives. Hence, the formulation for the combined problem is as follows,

$$\min_{y_{ijk},u_{ijk}(t),t_{f(ijk)}} \sum_{k\in M}\sum_{i\in N}\sum_{j\in N} c_{ijk}(u_{ijk}(t),t_{f(ijk)})y_{ijk} \tag{4.33}$$

subject to,

$$\sum_{k\in M}\sum_{j\in N} y_{ijk} = 1, \quad \forall i \in N\backslash\{0\} \tag{4.34}$$

$$\sum_{j\in N} y_{0jk} = 1, \quad \forall k \in M \tag{4.35}$$

$$\sum_{i\in N} y_{ihk} - \sum_{j\in N} y_{hjk} = 0, \quad \forall k \in M, \forall h \in N\backslash\{0\} \tag{4.36}$$

$$\sum_{i \in N} y_{i(n+1)k} = 1, \quad \forall k \in M \tag{4.37}$$

$$\sum_{i,j \in S} y_{ijk} \leq |S| - 1, \quad \forall S \subseteq N, \forall k \in M \tag{4.38}$$

$$\sum_{i \in N} r_i \sum_{j \in N} y_{ijk} \leq Q_k, \quad \forall k \in M \tag{4.39}$$

$$c_{ijk} = \min \int_0^{t_{f(ijk)}} \left(1 + \omega (u_{ijk}^2(t) v_{ijk}^2(t))^{\frac{1}{2}}\right) dt \tag{4.40}$$

subject to,

$$\dot{x}_{ijk}(t) = v_{ijk}(t) \tag{4.41}$$

$$\dot{v}_{ijk}(t) = \frac{1}{m_k + l_k}(-\gamma - \beta v_{ijk}^2(t) + u_{ijk}(t)) \tag{4.42}$$

$$x_{ijk}(0) = 0, \quad x_{ijk}(t_{f(ijk)}) = D_{ij}, \quad v_{ijk}(0) = 0, \quad v_{ijk}(t_{f(ijk)}) = 0 \tag{4.43}$$

$$0 \leq v_{ijk}(t) \leq v_{ub}, \quad u_{lb} \leq u_{ijk}(t) \leq u_{ub} \tag{4.44}$$

This problem combines the vehicle routing problem and the optimal control problem along each edge. Eqns. (4.40) – (4.44) represent the optimal control problem of the vehicle k along the edge i and j, which is inside the routing optimization problem. With the assumptions that vehicles are homogeneous and each vehicle has a fixed static load during the entire trip, the results of the optimal control problem are independent of the selections of routes, i.e., in this formulation, the joint costs of all edges $c_{ijk}(u_{ijk}(t), t_{(f(ijk))})$ can be solved first by the multiple shooting method, and the binary variables y_{ijk} of the routing problem are then determined. On the contrary, if the previous assumptions are relaxed, then the optimal values of the variables $u_{ijk}(t)$, $t_{(f(ijk))}$ and y_{ijk} are dependent on each other. This combined

problem cannot be solved sequentially, leading to a challenging optimization problem when the number of nodes is large.

4.2.4 Optimal Control Problem for a Vehicle Traveling along an Edge

In this subsection, the constant speed and optimal control models are employed to calculate the joint cost for a single edge.

4.2.4.1 Profile Patterns of Speed and Control Variables

Suppose the distance of an edge between two nodes is $D = 1800m$, and the speed limit is $v_{ub} = 20m/s$. Examples of the profiles for $v(t)$ and $u(t)$ using a constant speed model and an optimal control model are shown in Fig. 4.3 and Fig. 4.4.

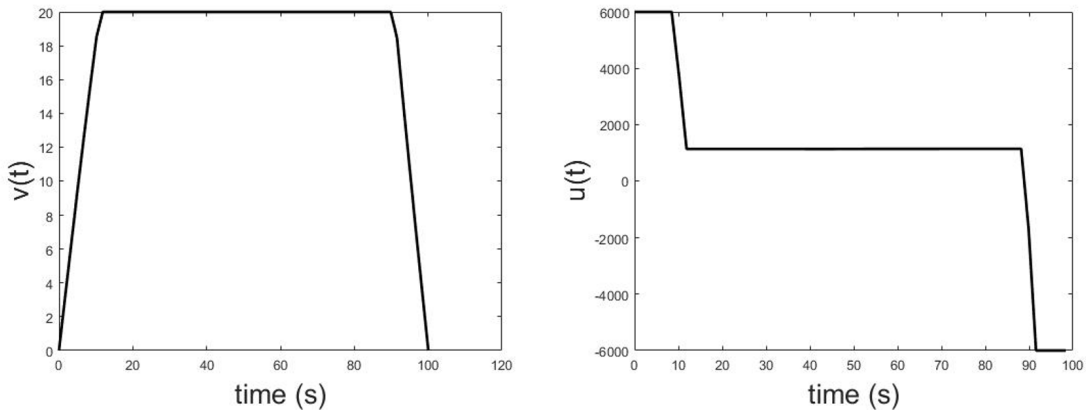


Figure 4.3: Typical Profiles of Speed (left) and Control (right) Variables in Constant Speed Model

As shown in Fig. 4.3 and 4.4, for the constant speed model, the vehicle accelerates and decelerates at the maximum rate at the start and end node, respectively, and maintains a constant speed limit in the middle of the trip. On the other hand,

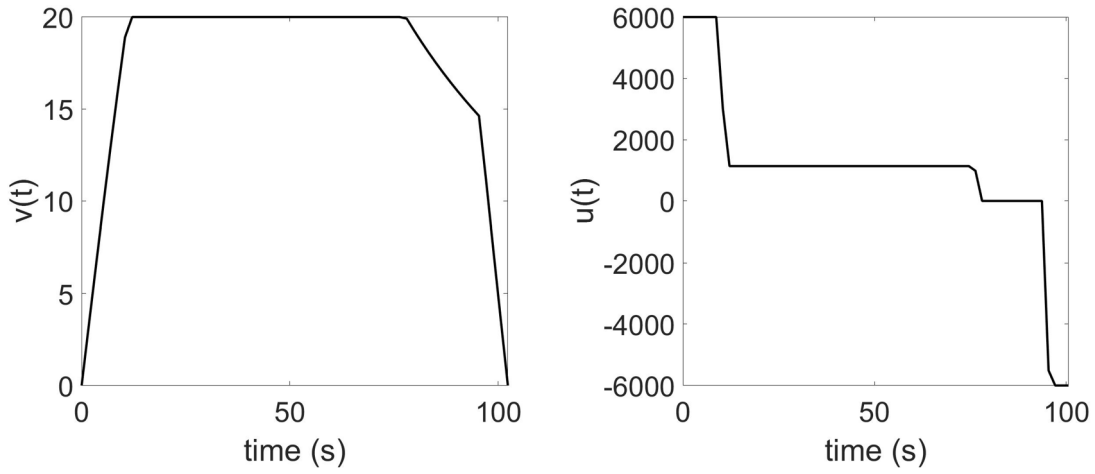


Figure 4.4: Typical Profiles of Speed (left) and Control (right) Variables in Optimal Control Model

for the optimal control model, the vehicle may not be operated at the maximum input at the acceleration stage, and the controller can stop the input for some time before reaching the destination during the decelerating stage. In this way, although the traveling time is increased, the joint cost can be potentially reduced due to the saving in energy consumption using the controller.

4.2.4.2 Effect of Distance on Energy Consumption Percentage of All Stages

By observing Fig. 4.4, it is expected that the stage (S2) accounts for most of the energy consumed if the distance becomes very large. Therefore, Fig. 4.5 is plotted to display the trends of the energy consumption percentages of all stages when the distance increases.

As shown in Fig. 4.5, the total energy consumption is mainly dependent on the constant speed stage (S2) for a long-distance situation. Therefore, it can be justi-

fied that for long traveling distances, the energy consumption can be approximately estimated by assuming constant speed throughout the trip and neglecting the acceleration and deceleration stages, as seen in the extant VRP literature (e.g., [60, 61]). However, if the traveling distance is relatively short, it is reasonable to reduce the energy consumed during the acceleration and deceleration periods.

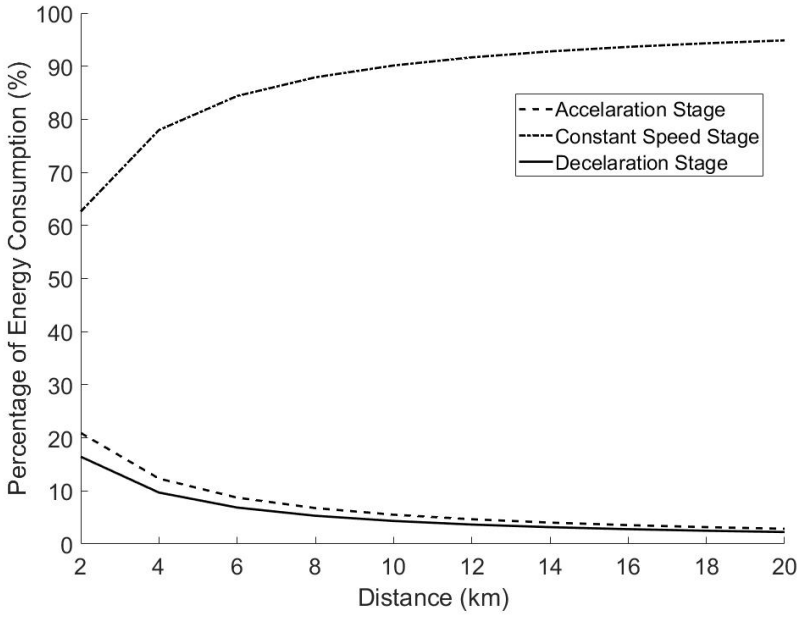


Figure 4.5: Percentage of Energy Consumption of All Stages Versus Distance

4.2.4.3 Effect of Speed Limit on Joint Cost

In the constant speed model with a fixed edge distance, the speed limit can affect the joint cost in two ways. First, the speed can influence the time needed for all stages. For example, a higher value of speed limit prolongs stages (S1) and (S3) but shortens the stage (S2). Second, the speed is related to the air drag force, which causes energy consumption to increase during the stage (S2). Therefore, different speed limit values are tested to find the joint cost. The distance is $D = 1800m$, and

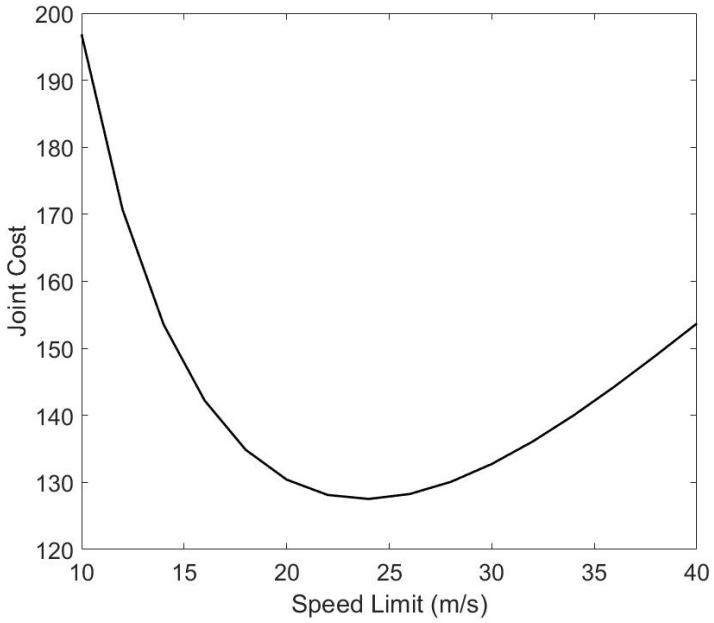


Figure 4.6: Joint Cost Versus Speed Limit

the weighting parameter is $\omega = 10^{-5}$. As shown in Fig. 4.6, it can be concluded that for a certain weighting parameter, as the speed increases, the joint cost of the constant speed model is reduced first and then rises.

4.2.4.4 Effect of Weighting Parameter on Joint Cost

The weighting parameter $\omega = \frac{w_2}{w_1}\eta$ defined in Section 4.1.3 contributes to balancing the optimization objectives of the traveling time and energy consumption. If ω is relatively large, the energy consumption is the dominant contributing factor in the joint cost, otherwise, the travel time contributes more significantly to the joint cost. With $D = 1800m$ and $v_s = 20m/s$ used in this test and varying ω in the range from 10^{-9} to 10^{-3}, the ratio of the joint cost of the optimal control model over that of the constant speed model is evaluated and plotted in Fig. 4.7.

When the weighting parameter ω is very small, the joint cost with the optimal

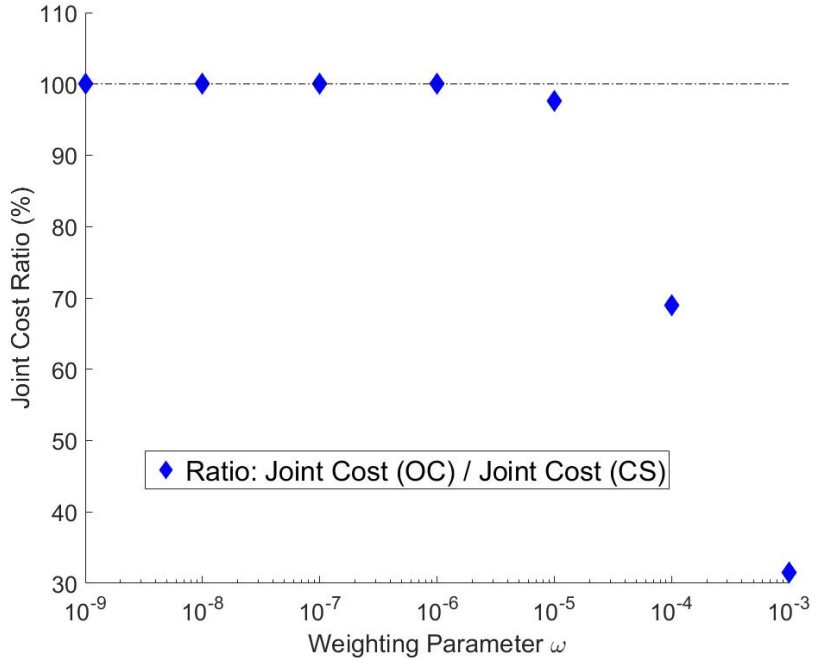


Figure 4.7: Joint Cost Ratio Versus Weighting Parameters (OC: Optimal Control Model; CS: Constant Speed Model)

control model is approximately equal to that with the constant speed model, meaning that the ratio is about 100%. As the weighting parameter increases, the joint cost with the optimal control model is better than that with the constant speed model. As is observed in Fig. 4.7, when $\omega = 10^{-3}$, the joint cost with the optimal control model becomes only around 30% of that with the constant speed model for this test problem.

4.3 Example

In this section, the three solution strategies from Section 4.2 are employed in a benchmark VRP test example (Solomon problem R1-type R104 [85]). For the given problem dataset, the first four columns: 1) customer number, 2-3) X and Y

coordinates and 4) demand are used. By the X and Y coordinates, the distance between every pair of nodes can be found, which is then used to find the joint cost of each edge and that of the entire problem. The real distance, measured in meter, is assumed to be 100 times the distance given for the X-Y coordinate values in the dataset. Customer number 1 refers to the depot, where the vehicles start from and finally return to. Each vehicle starts from the depot without any load. Fig. 4.8 shows a visualized map of the positions for the depot and the first 20 customer locations ($N_C = 20$) to be visited of the test example. For each node, the customer number is shown inside the parentheses, and followed by the number of cargo units at this location.

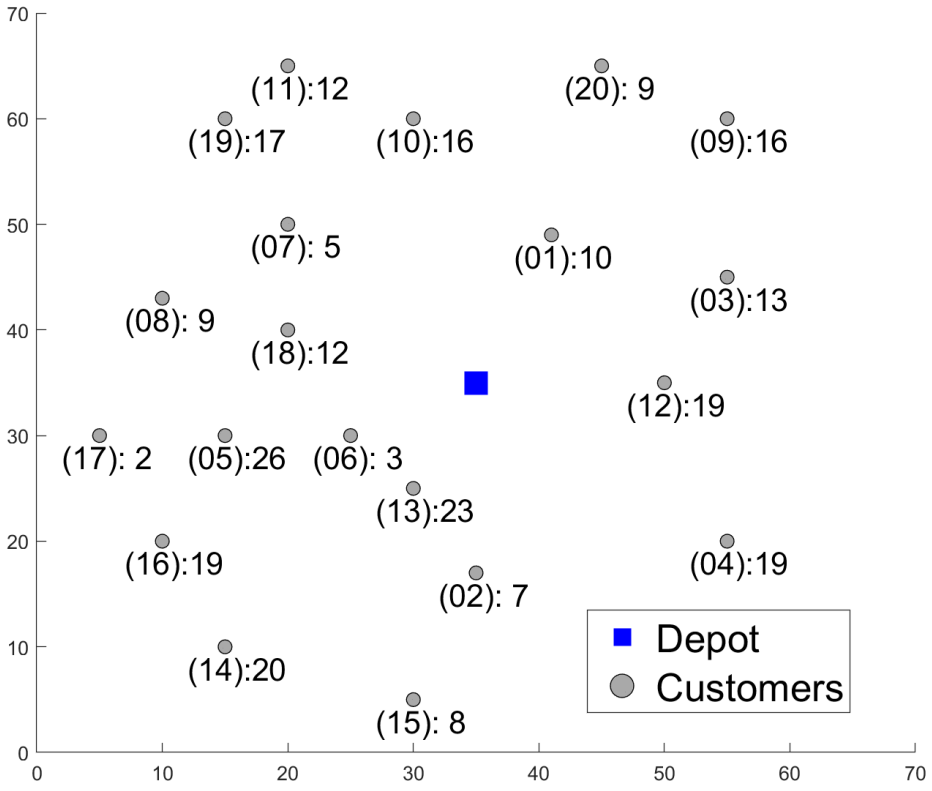


Figure 4.8: Benchmark VRP Example with 20 Customers

4.3.1 Results and Discussions

The problem is solved in Matlab and the Gurobi optimization solver [87] with the python interface. In this section, some results and discussions are presented with different parameter settings.

In the first set of test cases, the number of vehicles is considered to be $N_V = 4$. For each vehicle, the capacity is set to be 1.5 times the total demand divided by the number of vehicles. The weighting parameter is $\omega = 10^{-5}$. The speed limit is $v_s = 20m/s$. The number of customers N_C is 20, 60 and 100. Table 4.1 shows the computed joint cost from the three solution strategies (STR-1 stands for Strategy 1, STR-2 for Strategy 2 and STR-3 for Strategy 3).

Case	N_C	STR-1	STR-2	STR-3
1	20	2479.35	2404.72	2404.72
2	60	4574.19	4374.98	4374.98
3	100	6222.53	5881.73	5881.73

Table 4.1: Joint Cost Results: Various Number of Customers

From Table 4.1, it can be observed that the joint cost of the strategies 2 (STR-2) and 3 (STR-3), which employ a continuous-time optimal controller, is less than that of the strategy 1. Especially, if a higher value of the weighting parameter is selected, the joint cost can be significantly further reduced. The results for the strategies 2 and 3 are also found to be equal in all cases. The reason is that for the selected parameters and capacity constraint, the routing results of all strategies are identical. As an example, Fig. 4.9 shows the final routes of the vehicles when the

number of customers is 60.

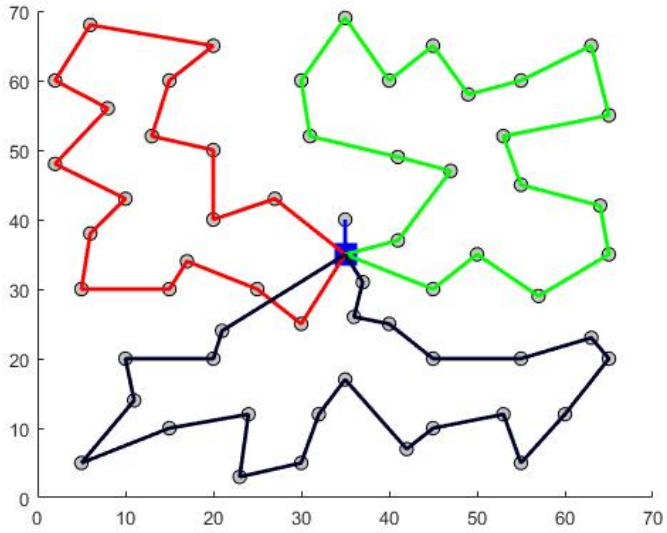


Figure 4.9: Routing Solution in Case 2 from All Strategies

In the second set, the number of customers is fixed as $N_C = 20$, and the number of vehicles N_V is 3, 4 and 5, for the case 4, 5 and 6, respectively. Other configurations are remained the same. Table 4.2 shows the joint cost in these cases.

Case	N_V	STR-1	STR-2	STR-3
4	3	2271.77	2200.27	2200.27
5	4	2479.35	2404.72	2404.72
6	5	2723.46	2645.85	2645.85

Table 4.2: Joint Cost Results: Various Number of Vehicles

As shown in Table 4.2, although the optimal routes of all strategies are still identical, the results of all cases again demonstrate that considering optimal control of the vehicle has the potential to reduce the joint cost for the entire problem. As an example, Fig. 4.10 shows the routing results in Case 6, which optimizes the joint cost in the absence of wind.

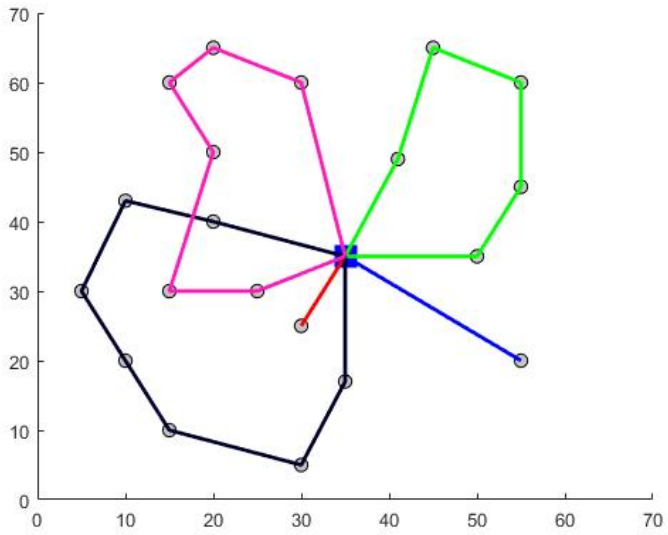


Figure 4.10: Routing Solution in Case 6 from All Strategies

4.4 Effects of Wind on Optimal Control-based Routing

4.4.1 Problem Formulation

In this section, wind is considered in the proposed vehicle control problem. It is to be noted that different wind directions can lead to heterogeneous air drag forces in the vehicular dynamic model, then the problem is formulated as follows,

$$\min_{u(t),t_f} \int_0^{t_f} \left(1 + \omega(u^2(t)v^2(t))^{\frac{1}{2}}\right) dt \tag{4.45}$$

subject to,

$$\dot{x}(t) = v(t) \tag{4.46}$$

$$\dot{v}(t) = \frac{1}{m+l}(-\gamma - \beta(v - v_{wp})^2(t) + u(t)) \tag{4.47}$$

$$x(0) = 0, \quad x(t_f) = D, \quad v(0) = 0, \quad v(t_f) = 0 \tag{4.48}$$

$$0 \leq v(t) \leq v_{ub}, \quad u_{lb} \leq u(t) \leq u_{ub} \tag{4.49}$$

where v_{wp} is the wind speed projected along a vehicle's velocity direction. As shown in Fig. 4.11, if a uniform constant west-wind field is assumed, i.e., the wind originates from the west and blows to east, then v_{wp} for the vehicle to travel from node A to node B can be calculated by

$$v_{wp} = v_w \cos\varphi \tag{4.50}$$

while v_{wp} for the vehicle to travel from B to A is

$$v_{wp} = -v_w \cos\varphi \tag{4.51}$$

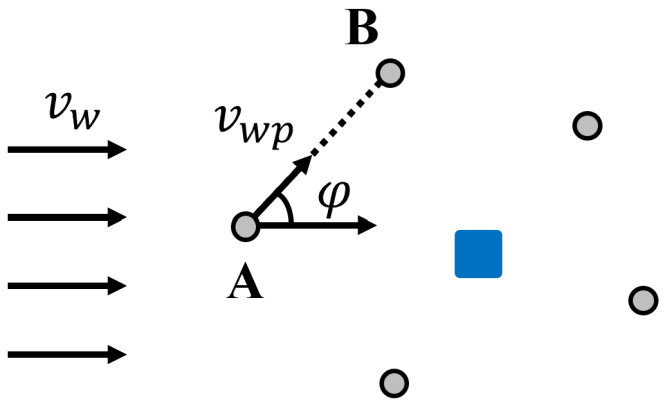


Figure 4.11: VRP with a Uniform Constant West-wind Field

Therefore, the edge costs are different for the vehicle to travel from A to B and from B to A. It is worth noting that the wind effect appearing in the vehicle dynamics is reasonably simplified. v_{wp} in Eqn. (4.47) only accounts for the portion

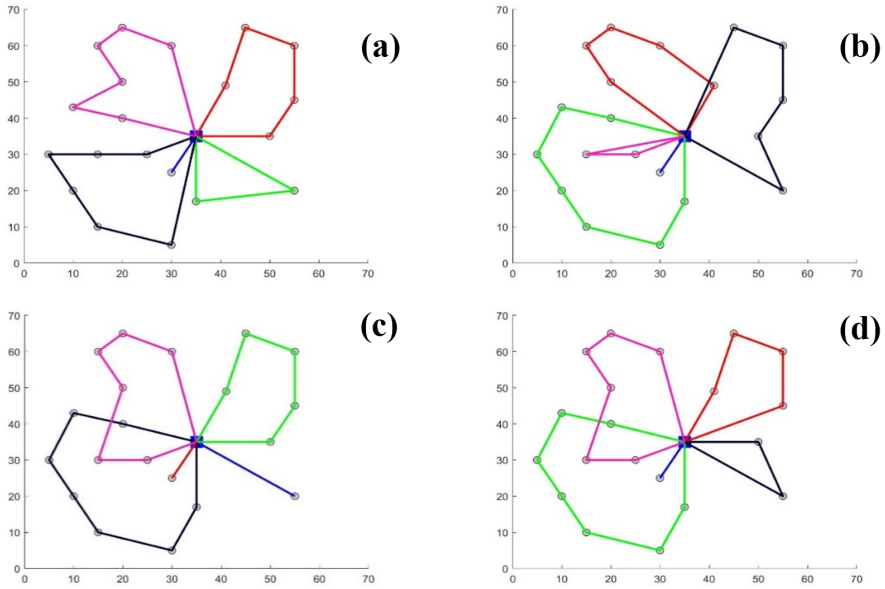


Figure 4.12: Routing Solution with Various Wind Directions from Strategy (STR-3): (a) East, (b) West, (c) North, (d) South

of the wind which hits the frontal area of the vehicle. More accurate models of the relationship between the aerodynamic drag and the vehicle fuel efficiency can be found in [88] and elsewhere.

4.4.2 Results and Discussions

In this section, different wind directions are considered in the optimal control-based routing problem. The wind speed v_w is assumed to be $5m/s$, and the directions are from east, west, north and south. The total number of customers is $N_C = 20$. The number of vehicles is considered to be $N_V = 5$. The weighting parameter is $\omega = 10^{-5}$ and the speed limit is $v_s = 20m/s$. The capacity is set to be 1.5 times the total demand divided by the number of vehicles.

The optimal joint costs are listed in Table 4.3. Despite the fact that the

Case	Wind Direction	STR-1	STR-2	STR-3
6	No Wind	2723.46	2645.85	2645.85
7	East	2724.14	2653.10	2646.76
8	West	2724.14	2653.10	2647.19
9	North	2724.18	2654.78	2653.02
10	South	2724.18	2654.78	2646.65

Table 4.3: Joint Cost Results: Various Wind Directions

vehicle dynamic model is simplified, it can be observed in Fig. 4.12 that routing results determined by the optimal-control based strategy are different for each wind direction, leading to a better joint cost compared with (STR-1) and (STR-2).

4.5 Summary

To the best of our knowledge, this research is the first to consider strategies for joint optimization of a continuous-time optimal control problem with a multi-vehicle routing problem. Results from a comparison study of the considered strategies are presented. These strategies are for 1) optimization of time-based routes followed by a constant speed control model for vehicles, 2) optimization of time-based routes followed by optimal control of vehicles, and 3) optimal control-based multi-vehicle routing. The objective of the optimization is to minimize a joint cost, which is a weighted sum of the travel time and energy consumed for all vehicles and for all routes.

The test results indicate that the optimal control-based routing strategy can potentially lead to a better value of a joint cost compared with the strategies of

finding the overall cost value when vehicles traveling at constant speed or following optimal control along the fastest routes, especially when wind is taken into the consideration.

In the next chapter, the results from Research Question 4 (RQ4) are presented. In RQ4, a more challenging vehicle routing-and-control problem than that considered in this chapter is investigated, and a heuristic solution approach is implemented.

Chapter 5: Integrating Optimal Vehicle Routing and Control with Load-dependent Vehicle Dynamics Using a UCT-based Approach

In this chapter, the results from Research Question 4 are presented in detail[1].

The organization of this chapter is as follows. This chapter begins with a nomenclature section. Then in Section 5.1, the integrated optimal vehicle routing-and-control problem is formulated. A new implementation of the UCT-based approach is described in Section 5.2. To demonstrate the proposed approach, in Section 5.3, results are presented and discussed for a benchmark VRP example, which is revised with the addition on an optimal control problem, as well as an integrated routing-and-control example for a small area in New York City. Finally, a summary of the chapter is given in Section 5.4.

[1]This chapter is based on the paper: Liu, T., Azarm, S., and Chopra, N., 2020, "Integrating Optimal Vehicle Routing and Control With Load-Dependent Vehicle Dynamics Using a Confidence Bounds for Trees-Based Approach." ASME. J. Dyn. Sys., Meas., Control, 142(4): 041006.

Nomenclature

A	Vehicle's wind resisting cross-sectional area (m^2)
C	Exploration coefficient in backpropagation step
C_d	Air drag resistance coefficient
C_r	Rolling resistance coefficient
c	Edge cost
D	Edge length (m)
g	Gravitational constant (m/s^2)
J	Joint cost
l	Vehicle's load (kg)
l_0	Mass of one unit of cargo (kg)
m	Vehicle's mass (kg)
N_C	Number of customers

N_n	Maximum number of tree nodes to stop expanding the tree
N_V	Number of vehicles
n	Counter of simulations in backpropagation step
p	Normalized probability in selection step
$\tilde{p}$	Pseudo-probability in selection step
q	Upper Confidence Bounds (UCB) value
$\hat{q}$	Minimal UCB value in selection step
r	Demand at customer's location
t_f	Final time to reach destination (s)
u	Vehicle control variable: forward force (N)
v	Vehicle speed (m/s)
$\dot{v}$	Vehicle acceleration (m/s^2)
w_1,w_2	Weighting coefficients associated to time, control objectives,

respectively

x Vehicle displacement (m)

y Binary variable for route selection

θ Road angle (rad)

ρ Air density (kg/m^3)

τ Normalized time horizon

λ Parameter value for exponential distribution in selection step

η Vehicle efficiency

5.1 Integrated Optimal VRP and Control Problem

A more challenging vehicle routing-and-control problem is proposed in RQ4, compared with that in RQ3. The problem integrates the VRP and control formulations described in Section 4.1, but takes the mass of loading into consideration of the dynamic model. Hence, the optimization of routing and control subproblems are strongly coupled in this problem.

The objective function of the entire problem is the sum of all edge costs, subject to the routing constraints. Each edge cost c_{ijk} is the joint cost obtained by evaluating the optimal control problem for the vehicle k traveling from node i

to node j, subject to the dynamics and path constraints. Hence, the integrated problem is formulated as follows,

$$\min_{y_{ijk},u_{ijk}(t),t_{f(ijk)}} J = \sum_{k\in M}\sum_{i\in N}\sum_{j\in N} c_{ijk}(u_{ijk}(t), t_{f(ijk)})y_{ijk} \tag{5.1}$$

subject to,

$$\sum_{k\in M}\sum_{j\in N} y_{ijk} = 1, \forall i \in N\backslash\{0\} \tag{5.2}$$

$$\sum_{j\in N} y_{0jk} = 1, \forall k \in M \tag{5.3}$$

$$\sum_{i\in N} y_{ihk} - \sum_{j\in N} y_{hjk} = 0, \forall k \in M, \forall h \in N\backslash\{0\} \tag{5.4}$$

$$\sum_{i\in N} y_{i(n+1)k} = 1, \forall k \in M \tag{5.5}$$

$$\sum_{i,j\in S} y_{ijk} \leq |S| - 1, \forall S \subseteq N, \forall k \in M \tag{5.6}$$

$$c_{ijk} = \min \int_0^{t_{f(ijk)}} \left(w_1 \times \frac{1}{10^3} + w_2 \times \frac{u_{ijk}^2(t)v_{ijk}^2(t)}{10^{11}} \right) dt \tag{5.7}$$

subject to,

$$\dot{x}_{ijk}(t) = v_{ijk}(t) \tag{5.8}$$

$$\dot{v}_{ijk}(t) = \frac{1}{m_k + \hat{l}_{ijk}}(-\gamma - \beta v_{ijk}^2(t) + u_{ijk}(t)) \tag{5.9}$$

$$x_{ijk}(0) = 0, x_{ijk}(t_{f(ijk)}) = D, v_{ijk}(0) = 0, v_{ijk}(t_{f(ijk)}) = 0 \tag{5.10}$$

$$0 \leq v_{ijk}(t) \leq v_{ub}, u_{lb} \leq u_{ijk}(t) \leq u_{ub} \tag{5.11}$$

It is assumed that some units of cargos r_i is to be collected from customer i, and the mass of each cargo is l_0. The edge cost c_{ijk} is influenced by $\hat{l}_{ijk}$ in Eqn. (5.9),

which represents the cumulative load mass on vehicle k as it is about to travel from customer i to customer j, which, at this instant, has collected the cargos from all previously visited customers. Hence, $\hat{l}_{ijk}$ acts as the coupling term of the routing-and-control parts of the model, and is expressed as

$$\hat{l}_{ijk} = \sum_{h} r_h l_0 \tag{5.12}$$

where subscript h is the index of customers that vehicle k has visited prior to traveling along the edge (i, j).

For a traditional VRP formulation with the objective function being the shortest total distance traveled by all vehicles, one might be able to identify a "good" solution by intuition for simple cases [48]. However, in the case when optimal control is taken into consideration, the edge cost c_{ijk} can be significantly different when vehicle k travels along altered routes prior to customer i. This coupling characteristic leads to a more complicated problem of optimizing the total routing-and-control cost. Meanwhile, the number of feasible solutions will also grow exponentially with the number of customers. Hence, to optimize the proposed problem and obtain a good solution with a reasonable computational cost, a new implementation of the UCT approach is considered for the integrated problem formulated in this section.

5.2 UCT-based Approach

A UCT-based approach, which is heuristic in nature, is presented in this section. It is worth noting that although the approach is employed specifically for the proposed integrated vehicle routing-and-control problem in this chapter, the approach can be modified and used to solve other (similar) integrated mixed integer programming and optimal control problems.

5.2.1 Tree Structure

A key issue in applying the UCT-based approach is to map the VRP into a tree structure which represents the routing solutions. In this way, each tree node represents a feasible instance of locations of all vehicles, referred to as a state. The tree starts by expanding from the root node, which denotes the state that all vehicles are to be dispatched from the depot. A branch is added to expand the tree from one state (parent node) to the other (child node), which denotes the action that one vehicle moves from one location to another, while other vehicles stop at their current locations. A leaf node refers to a routing solution that all customers have been visited and all vehicles have returned to the depot.

Fig. 5.1 (left) shows a possible tree structure for a small-scale example with the number of customers and vehicles as $N_C = 5$ and $N_V = 2$. Vehicle 1 and 2 are marked with the blue and red colors, respectively. The root node represents the state $[0, 0]$, meaning that the two vehicles are at the depot. For the first move, vehicle 1 may travel to customer 1, or 2 or 3, while vehicle 2 stays at the depot.

If the tree expands from the state $[0, 0, 1]$, as an example, then vehicle 2 travels to an unvisited customer (i.e., customer 4) from the depot, while vehicle 1 stays at customer 1. In this way, the tree can be expanded until all leaf nodes are found or the number of tree nodes exceeds a pre-set value N_n. Fig. 5.1 (right) shows a corresponding routing map for the specific state in the red rectangle on Fig. 5.1 (left). At this state, vehicle 1 travels from the depot to customer 1 and 2, and vehicle 2 travels from the depot to customer 4 and 3.

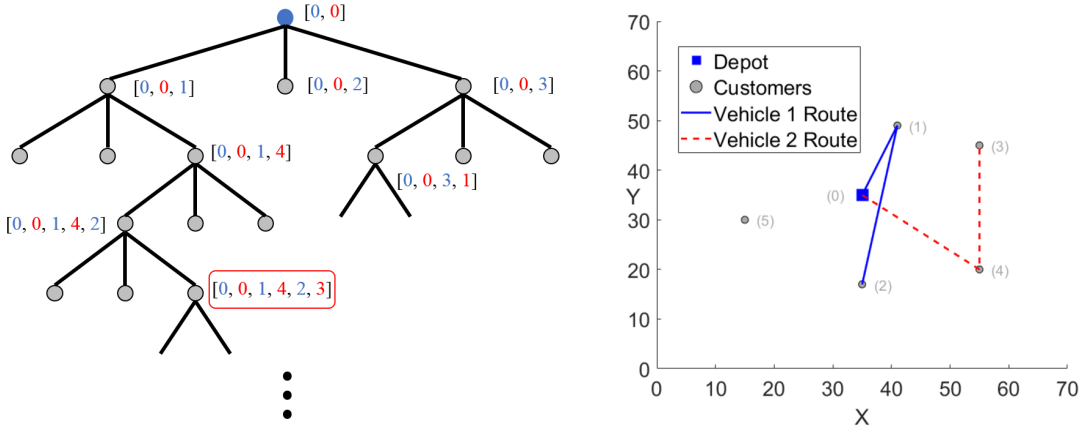


Figure 5.1: Demonstrative Example: (left) Tree Structure, (right) Routing Map

A feasible routing solution must follow the properties of a VRP: (1) Each customer is only visited once, hence the child nodes must be the customers which have not been visited at the current state. (2) For the first move of each vehicle, the depot, which has a customer node index 0, cannot be a candidate node. Starting from the second move of each vehicle, it can choose to visit another customer or return to the depot. (3) When a vehicle has returned to the depot, it is not dispatched again. In other words, the only candidate node for the next move of a returned vehicle must be 0. (4) If only one vehicle is on a route and all other vehicles have returned to the

depot at a current state, the vehicle cannot return before all remaining customers are visited. (5) If all customers have been visited, all vehicles must return to the depot.

For a small problem, it is possible to solve the problem by finding the objective function values of all leaf nodes. However, it is impractical to search exhaustively for the optimal solution of a relatively large problem. Therefore, a constant value N_n is set to terminate the approach when the number of the tree nodes reaches N_n, and the best routing solution at this instant is regarded as a "good" solution to the problem.

Following this tree structure, the UCT-based approach is implemented to find a solution to the integrated routing-and-control problem.

5.2.2 Approach Steps

In this section, the UCT-based approach is described. The approach iteratively expands the tree and updates an estimate to the objective function value. The approach consists of four steps: 1) selection, 2) expansion, 3) simulation and 4) backpropagation, which are presented in detail in the following subsections.

5.2.2.1 Step 1: Selection

The first step is to randomly select one node to expand the tree among all the tree nodes. In the first iteration, the only possible selection is the root node, denoting the state that all vehicles are at the depot. In the later iterations, a node

is randomly selected to fulfill the expansion step.

As an example, Fig. 5.2 shows a tree structure when a selection step is to be performed. First, the candidate nodes (Fig. 5.2, solid red circles in the tree) are identified, which have no child node at the instance. The selection step needs the information about the Upper Confidence Bound (UCB) values of the candidate nodes. In the proposed approach, the UCB value is an estimation of the upper bound of the total joint cost (Eqn. (5.1)).

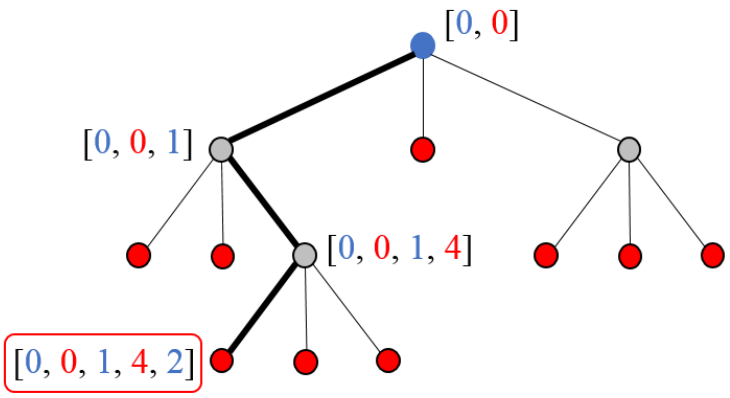


Figure 5.2: Step 1: Selection

Next, a probability mass function of the candidate nodes is generated for randomly selecting the node to expand the tree structure. Suppose the UCB values of the candidate nodes are $q_1, q_2, \ldots, q_k$, and the minimal value is denoted as $\hat{q}$ then an exponential distribution can be employed to generate the pseudo-probabilities $\tilde{p}_i$ of the candidate nodes,

$$\tilde{p}_i = \lambda e^{-\lambda(\frac{q_i}{\hat{q}}-1)} \tag{5.13}$$

By employing the exponential distribution, the probability of selecting one node is negatively correlated to its UCB values. Hence, the nodes with low UCB values are preferred. In other words, the algorithm tends to select the node with a lower estimated total joint cost, without any information on the true optimal value. In this way, this step differs from the previous literature [76] where a true optimal value is assumed to be known. In addition, choosing the parameter value λ can help to determine the exploration and exploitation of the tree. The pseudo-probabilities are then normalized to ensure that their sum equals to 1,

$$p_i = \frac{\tilde{p}_i}{\sum_{i=1}^{k} \tilde{p}_i} \tag{5.14}$$

Following the probability mass function (Eqn. (5.14)), a child node is then randomly selected. As an example, the state with a red rectangle (Fig. 5.2, left side) is chosen among the nine candidate nodes.

5.2.2.2 Step 2: Expansion

The next step is to expand the tree from the selected node, as shown in Fig. 5.3.

At the state selected in Step 1, vehicle 1 stays at customer 2 and vehicle 2 is ready to move. As shown in Fig. 5.3 (left), the three potential customers for vehicle 2 to move are 0 (depot), 3 and 5, then the tree is expanded to have three child nodes from the selected node. The corresponding routing map of the child nodes is shown Fig. 5.3 (right).

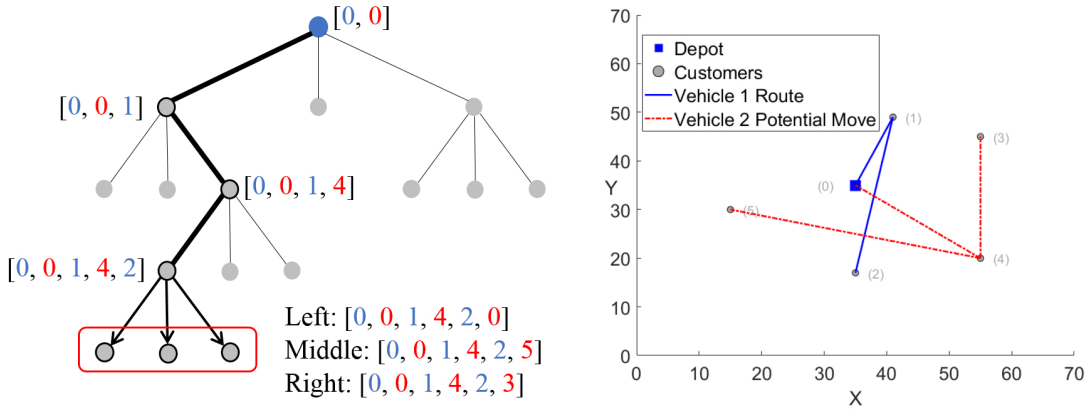


Figure 5.3: Step 2: Expansion, (left) Tree Presentation, (right) Routing Map

Note that all the unvisited customers can be the child nodes of the current state. However, when the number of customers N_C is large, the breadth of the tree will greatly increase, leading to a decrease of the depth when the largest number of tree nodes N_n is fixed, which can lead to a relatively bad solution. Hence, a fraction of unvisited "neighbors", according to the Euclidean distances of the locations, are added as the child nodes. Despite that it may filter out the global optimal solution, this approach can be efficient in obtaining a good solution for a large problem.

5.2.2.3 Step 3: Simulation

For each child node expanded in Step 2, a random path is generated from this node following the properties described in Section 5.2.1 to complete a feasible routing solution. The subsequent moves of the vehicles are randomly generated by employing a uniform distribution probability mass function at each step.

Fig. 5.4 shows one possible feasible routing solution (leaf node) and its corresponding routing map generated from the last child node in Step 2. This child node

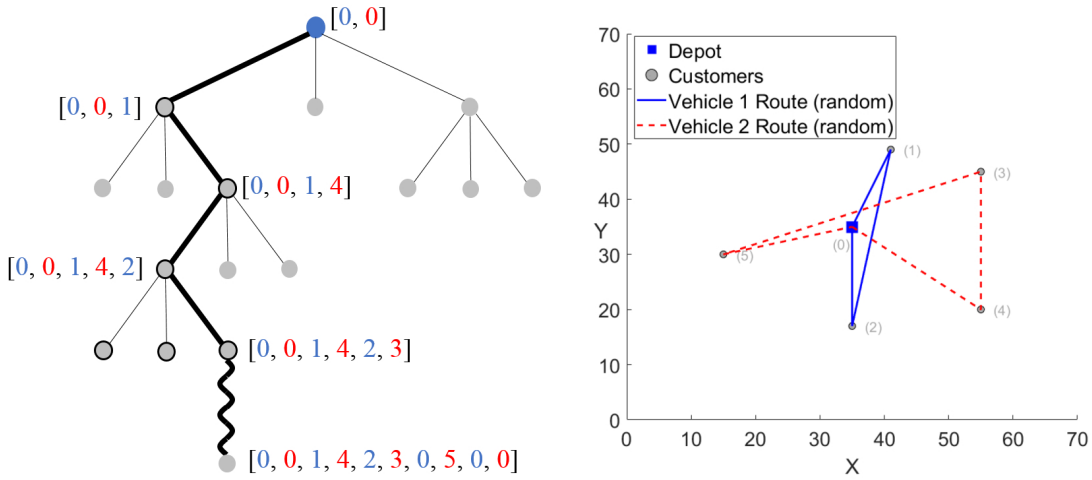


Figure 5.4: Step 3: Simulation, (left) Tree Presentation, (right) Routing Map

denotes the state that vehicle 1 and 2 are at customer 2 and 3, respectively. The vehicles randomly move to the other unvisited customers or return to the depot, till a feasible routing solution is obtained. As an example, the leaf node in Fig. 5.4 (left) represents the routing solution that vehicle 1 returns to the depot directly after visiting customer 2. Hence, vehicle 2 must visit both customers 3 and 5 before returning.

5.2.2.4 Step 4: Backpropagation

Given the simulated routing solution determined in Step 3 for each child node, the amount of cargos of each vehicle traveling between any two locations can be calculated. Hence, the total joint cost value (i.e., objective function) can be obtained. For each child node, the cost is added to the cumulative simulation cost of itself and those of its parent nodes, and the counters of simulations for itself and all its parent nodes are added by 1. With these updated values, the UCB value q_i for node i is evaluated by

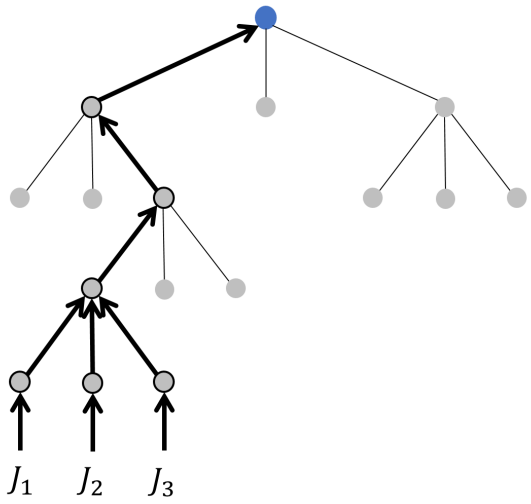


Figure 5.5: Step 4: Backpropagation

$$q_i = \bar{J}_i + C\sqrt{\frac{\log(n_{p_i})}{n_i}} \tag{5.15}$$

where $\bar{J}_i = \frac{\sum_{j=1}^{n_i} J_j}{n_i}$ is the average simulation cost value of the i-th node ($\sum_{j=1}^{n_i} J_j$ is the cumulative simulation cost of the i-th node so far), n_i and n_{p_i} are the counters of simulations that have tested the i-th node and its parent node, respectively. $C = 2$ is the exploration coefficient for the estimation.

As shown in Fig. 5.5, for the demonstrating example, J_1 to J_3 represent the total joint costs of the three child nodes after Step 3. Hence, for each child node, the average simulation cost is its simulation cost, and the counter of simulations is 1. For each of all the parent nodes (the direct parent node all the way up to the root node), the cumulative simulation cost is increased by $J_1 + J_2 + J_3$, and the counters of simulations are increased by 3. Finally, the process goes back to Step 1 for the next iteration, or stops when the number of tree nodes exceeds N_n.

5.3 Examples

5.3.1 Example 1: A Benchmark VRP with Optimal Control

The same benchmark VRP problem [85] as in Section 4.3 is selected and then revised by adding to it the optimal control model described in RQ4. Fig. 5.6 is the map for the depot and 20 customer locations ($N_C = 20$), marked with customer number and numbers of cargo units. The mass of a cargo unit is $l_0 = 100kg$. As an example, for customer 4, the demand amount of 19 is for a cargo mass of 1,900 kg. It is assumed that each vehicle starts from the depot without any load.

In the simulations, the following values are used for the model parameters: $g = 9.81m/s^2$, $m = 11,000kg$, $A = 10m^2$, $\rho = 1.2041kg/m^3$, $v_{ub} = 31.29m/s\,(70\,miles/hour)$, $C_r = 0.01$, $C_d = 0.7$, $u_{lb} = -10,000N$, and $u_{ub} = 10,000N$. It is also assumed that the roads are flat, so $\theta = 0$ for all roads.

In a sequential approach, first the shortest total travel distance is obtained by solving the mixed integer programming problem in Gurobi [87]. Hence, the routes for all vehicles are determined. The vehicles then start to visit the customers according to the assignments to meet the demands. When calculating the joint cost of a certain edge, the cumulative number of cargos is available for each vehicle. Hence, the total joint cost for the proposed problem can be obtained.

Next, the proposed UCT-based approach is employed. The goal is to find a "good" feasible routing solution which reduces the total joint cost of the entire problem. To avoid solving the optimal control problem every time in the simulation

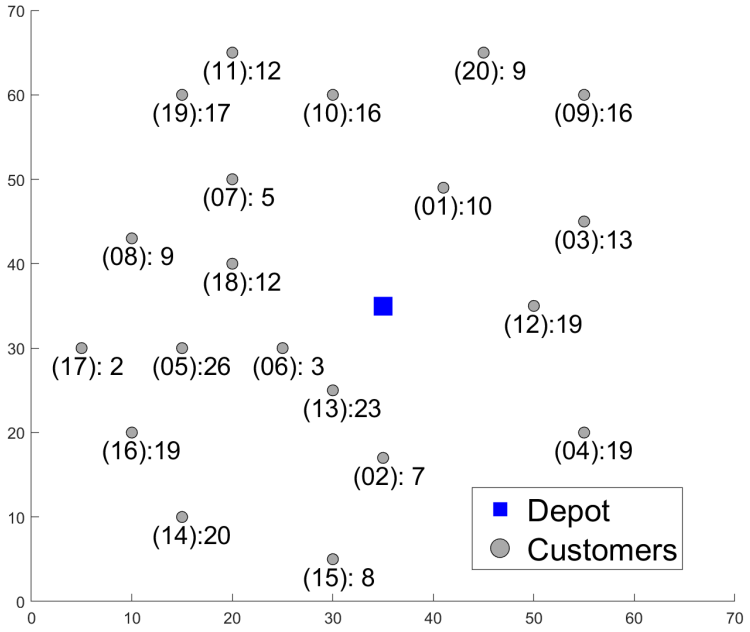


Figure 5.6: Benchmark VRP Example with 20 Customers

step, the optimal control cost is approximated by a function of the total mass of the vehicle (vehicle and cargos) and the distance of an edge $J_c(m_{total}, D)$. The optimal joint costs with the combinations of 10 different distances and 16 different total mass values are solved by the multiple shooting method. Then a multivariate linear regression technique is applied to obtain $J_c(m_{total}, D)$ for each combination of w_1 and w_2.

In the test cases, the number of vehicles is $N_V = 3$. In the selection step, λ is selected to be 5. Using the sequential approach, the routes are determined by the shortest total travel distance, and the total joint cost is calculated at the second step. For the UCT-based approach, the best estimating total joint cost is regarded as a good optimal result after the number of tree nodes exceeds $N_n = 15,000$ for each run. A total of 30 runs are conducted. The best and worst routing solutions

are selected, which leads to the minimum and maximum estimating objective values of the 30 runs, respectively. Then, the actual joint costs are re-evaluated on the best and worst routing solutions, which are listed in Table 5.1 along with the sequential solution.

w_1		0.1	0.3	0.5	0.7	0.9
Objective Value (Joint Cost)	**Sequential**	6.73	6.85	7.24	6.85	5.72
	UCT-based (best)	5.29	5.64	6.27	6.28	5.38
	UCT-based (worst)	6.37	6.51	7.07	6.45	6.02
Total Travel Time (10^3 seconds)	**Sequential**	7.40	7.29	6.42	5.65	4.86
	UCT-based (best)	9.14	8.53	7.10	5.64	4.58
	UCT-based (worst)	9.73	9.54	7.94	5.83	4.95
Total Energy Consumption (kJ)	**Sequential**	815.94	816.13	897.70	981.59	1161.49
	UCT-based (best)	697.40	663.68	737.32	881.40	1120.22
	UCT-based (worst)	774.38	722.23	786.57	888.58	1250.93

Table 5.1: Example 1: Optimization Results of Sequential and UCT-based Approaches

From Table 5.1, it can be observed that the best routing solution by the proposed UCT-based approach provides a better joint cost value, compared to the sequential approach. When w_1 is small, the total travel time is increased in order to reduce energy consumption and lower the weighted joint cost. When w_1 is large, the energy consumption is increased in order to shorten the total travel time for all the vehicles. In Table 5.2, the statistics of the objective function values of the proposed heuristic approach are listed. It can be observed that the sequential solutions are worse than the solutions obtained by the UCT-based approach.

In Fig. 5.7, the solutions obtained from both approaches are plotted. The x-axis is for the objective function for total travel time of all vehicles, and the y-axis

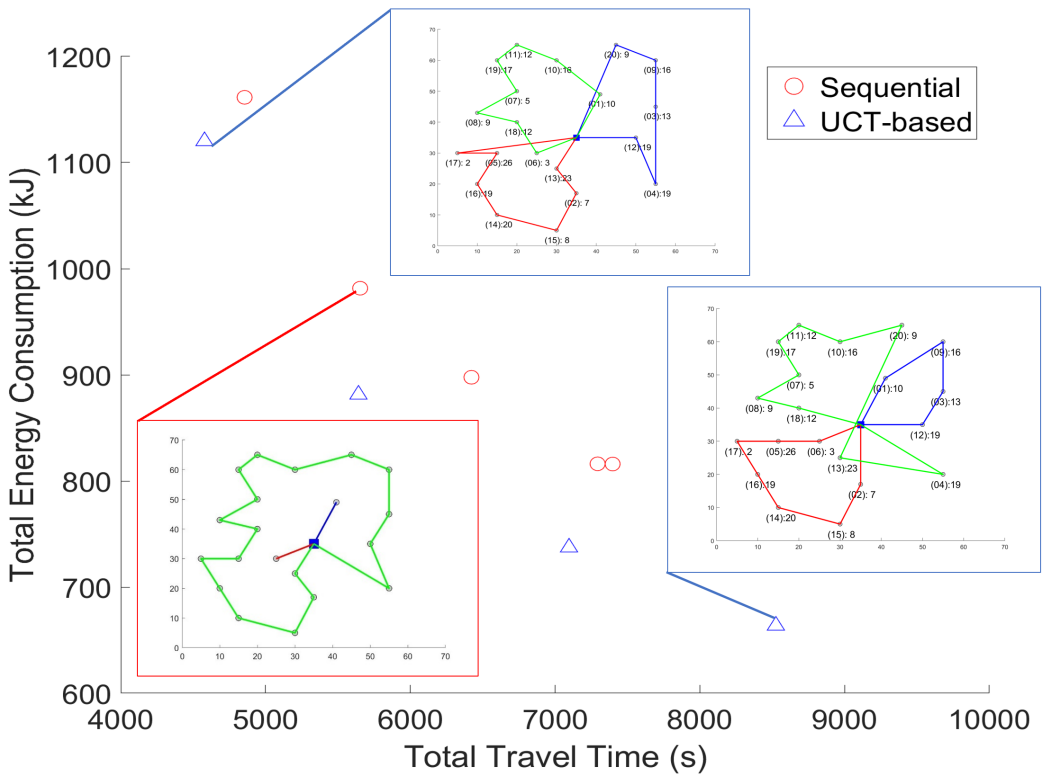


Figure 5.7: Solutions from Sequential and UCT-based Approaches, with Routing Solutions for Different Weighing Parameter w_1

$\mathbf{w_1}$	Mean	Standard Deviation	95% Confidence Interval	Objective (Sequential)
0.1	5.80	0.26	(5.70, 5.89)	6.73
0.3	6.08	0.22	(6.00, 6.16)	6.85
0.5	6.76	0.23	(6.68, 6.84)	7.24
0.7	6.38	0.05	(6.36, 6.40)	6.85
0.9	5.82	0.14	(5.77, 5.87)	5.72

Table 5.2: Example 1: Objective Function Value Statistics of UCT-based and Sequential Approaches

is for the objective of total energy consumption. Though the solution obtained by the UCT-based approach is generally not the global optimum, it can be observed that the joint cost is improved from the sequential approach with respect to both objectives.

In addition, the routing results for the approaches are shown in Fig. 5.7. For the sequential approach, since the routing results are obtained by the shortest total travel distance of all vehicles, the routes are identical regardless of the weighting parameter w_1. In contrast, for the UCT-based approach, the routing solutions are not only different from the sequential one, but also different depending on the choice of w_1. Referring to Fig. 5.7, the routing solutions are plotted when $w_1 = 0.1$ and $w_1 = 0.9$. It can be observed that when $w_1 = 0.1$, the routing solution contributes to a smaller energy consumption cost. Also, one vehicle (shown with green) travels a relatively longer distance for the sake of carrying fewer cargos to save energy cost. On the other hand, when $w_1 = 0.9$, the total travel time is the more important objective to optimize. Hence all vehicles tend to have shorter trips, leading to a better total travel time.

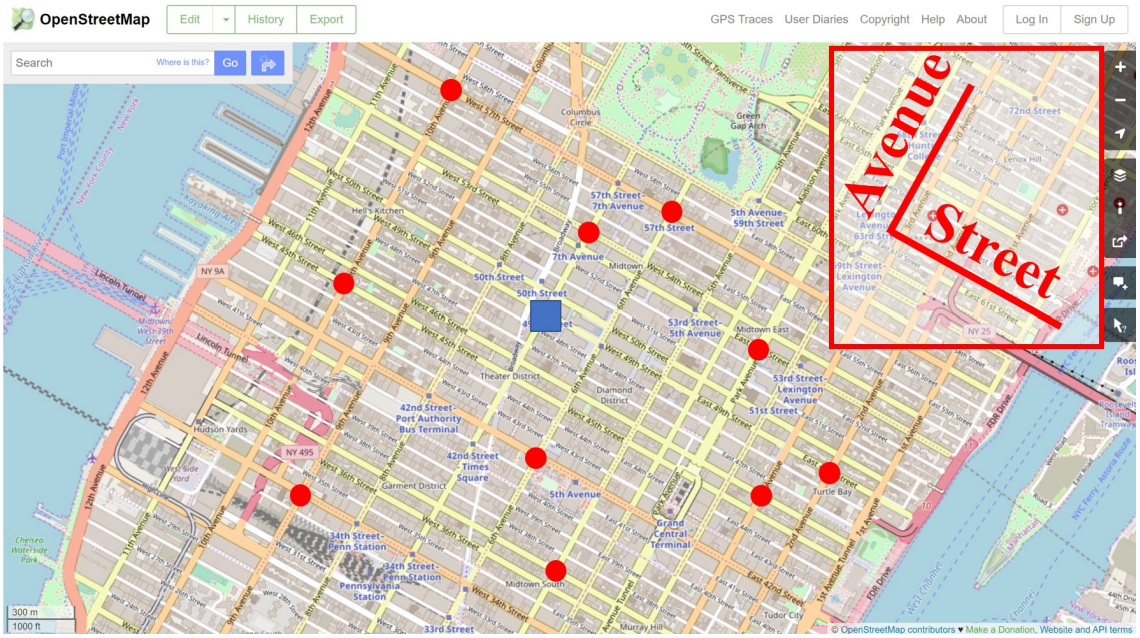


Figure 5.8: New York City VRP Example with 10 Pickup Locations

5.3.2 Example 2: A Vehicle Routing-and-Control Problem in NYC

For this example, a notional integrated routing-and-control problem is formulated and solved for a small area from the New York City (NYC) street map (captured from OpenStreetMap under Open Database License). For this problem, it is assumed that there are three shuttles which are to be dispatched from the depot (Fig. 5.8, blue square) to pick up customers and their packages (demands) at ten locations (Fig. 5.8, red circles), and finally return to the depot.

The information about customer location and demand is provided in Table 5.3. In the map shown on Fig. 5.8, a road is called a "Street" if it is running east-west, and an "Avenue" if running north-south.

Furthermore, the following assumptions are made to simplify the problem:

1. Each vehicle has a net mass of $5,000kg$. The upper and lower bounds on the

Customer	Location	Demand (kg)
0 (Depot)	49th St, 7th Ave	
1	57th St, 10th Ave	700
2	57th St, 6th Ave	700
3	54th St, 7th Ave	140
4	45th St, 10th Ave	1,050
5	53rd St, 4th Ave	700
6	46th St, 3rd Ave	140
7	49th St, 2nd Ave	700
8	42nd St, 6th Ave	140
9	37th St, 5th Ave	140
10	34th St, 9th Ave	700

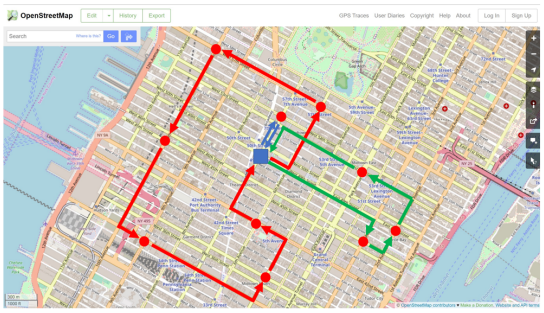
Table 5.3: Example 2: Pickup Location and Demand Information

control variable are $u_{lb} = -15,000N$, and $u_{ub} = 15,000N$.

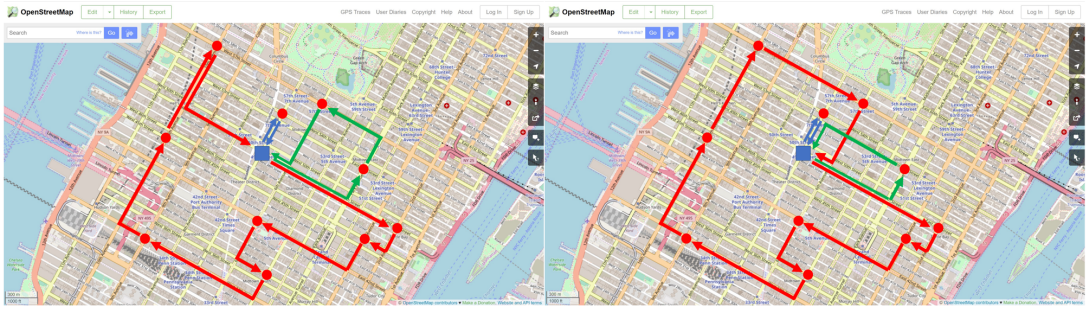
2. The road length is 80 meters between each two consecutive streets and 270 meters between two consecutive avenues. The speed limit is 40mph (17.88m/s). It is assumed that two-way traffic is allowed on each road.

3. Each vehicle makes at most one turn when traveling between two locations. If the vehicle has to make a turn, it is assumed to stop at the intersection. The vehicle will not stop at other intersections or be affected by other traffic.

The final routing solutions and the optimization results are shown in Fig. 5.9 and Table 5.4, respectively. It can be observed that the total travel distance of the solution by the UCT-based approach is increased, which affects the total travel time of all the vehicles. However, the total energy consumption of all vehicles is decreased



(a)



(b) (c)

Figure 5.9: New York City VRP Example Solutions: (a) Sequential, (b) UCT-based, $w_1 = 0.1$ and $w_1 = 0.5$, (c) UCT-based, $w_1 = 0.9$

by the routing solution by the UCT-based approach.

$\mathbf{w_1}$		0.1	0.5	0.9
Total Travel Distance (m)	**Sequential**	11,540	11,540	11,540
	UCT-based	12,240	12,240	12,140
Total Travel Time (10^2 seconds)	**Sequential**	13.31	9.62	8.33
	UCT-based	13.68	9.88	8.50
Total Energy Consumption (kJ)	**Sequential**	937.4	2086.2	3673.6
	UCT-based	935.1	2083.8	3453.1

Table 5.4: Example 2: Optimization Results of Sequential and UCT-based Approaches

5.4 Summary

To the best of our knowledge, the chapter presents the first attempt for formulating and solving an integrated routing-and-control problem with load-dependent dynamics. In the proposed problem, a continuous-time vehicle control model is embedded within the vehicle's routing model. As shown in the integrated model, the dynamical equations are dependent on each vehicle's load, which is determined by the demand at each customer's location that each vehicle visits. The problem is challenging to solve as the routing and control parts of the model are coupled with each other. A new implementation of a UCT-based approach is developed and used to solve the proposed model heuristically.

Two test examples are used to demonstrate the solutions obtained from the proposed integrated model and the UCT-based approach. For both examples, the integrated solutions are compared against those obtained by a sequential approach, where the VRP is solved first followed by the optimal control problem. In the first example, a benchmark VRP test problem from the literature is selected and revised by embedding in it an optimal control problem. For this integrated problem, formulated as a bi-objective optimization problem, it is shown that the joint cost obtained from the integrated approach is better than that obtained from the sequential approach. For the second example, a notional integrated routing-and-control problem for a small area from NYC street map is formulated and solved. For this example, based on the solutions, it is shown that the total travel distance of all the vehicles by the UCT-based approach is increased, which prolongs the total travel time, but

the total energy consumption can be reduced by such route decisions.

In the next chapter, the conclusions from this dissertation are provided. These include a summary of the work, highlight of the results, contributions made and possible future directions.

Chapter 6: Conclusion

This chapter presents the concluding remarks of the dissertation. In Section 6.1, a summary of the models and methods developed in this dissertation is provided, and results are highlighted. In Section 6.2, the main contributions of this dissertation are provided. Finally, some limitations of the current work and suggestions for extending the research in this dissertation are presented in Section 6.3.

6.1 Summary

In this section, the summary and highlights of the results of each research question are presented.

As a result of RQ1, a class of multi-subsystem co-design problems is formulated, where the physical design subproblems are convex, and the control subproblems are linear quadratic regulators. Notably, a multi-level decentralized approach is developed to solve the proposed class of the co-design problems. The approach is composed of two hierarchical decomposition techniques, one for the physical part and the other for the control part of the co-design subsystems. In particular, the indirect method is employed to solve the control part. Derived from the necessary optimality conditions, the gradients of the Hamiltonian of the control part with

respect to the physical variables bridge the physical and control parts.

When implemented on a numerical and engineering example, results from the proposed decentralized approach are comparable with those from a centralized (all-at-once) approach. Finally, a scalable test problem of the serial spring-mass-damper system is implemented. As the number of subsystems increases, the decentralized approach can outperform the centralized approach on the computational time.

In RQ2, certain assumptions of the control parts are relaxed when compared with the multi-subsystem co-design problems considered in RQ1. The direct collocation method is employed to solve the control part numerically. To solve the co-design problem, two decentralized schemes are proposed. The multi-level approach follows the procedure in RQ1. The bi-level approach simplifies the procedure by decomposing the entire system into multi-subsystem co-design subproblems.

For the similar examples as in RQ1, the results from the decentralized and centralized approaches are comparable. However, it can also be observed that the bi-level approach obtains better solutions than those from the multi-level one. In the scalable spring-mass-damper problem, multiple computerized 'machines' are simulated to solve the decomposed subproblems in parallel. In this way, the decentralized approaches are shown to outperform the centralized approach as the number of subsystems increases. The results also show that computational power can be utilized more effectively using the decentralized approaches.

In RQ3, a new integrated optimization of the routing and control for a fleet of vehicles is considered. In the literature for routing problems, the vehicle speed is assumed to be constant. To demonstrate that the consideration of vehicle speed

can be beneficial when introducing optimal control into the routing problem, three solution strategies are proposed and compared. These strategies are: 1) optimization of time-based routes followed by a constant speed control model for vehicles, 2) optimization of time-based routes followed by optimal control of vehicles, and 3) optimal control-based multi-vehicle routing. The objective of the optimization model for all three strategies is to minimize a weighted sum of the travel time and energy consumption costs. The results indicate that the optimal control-based routing strategy can potentially lead to a better joint cost compared with the other strategies, especially in the scenarios when wind is considered.

RQ4 extends the routing-and-control problem in RQ3. In the more challenging integrated problem, the dynamical equations are dependent on each vehicle's load, which is determined by the demand at each customer's location that each vehicle previously visited. Since the routing and control subproblems are coupled, an obvious approach is a sequential one, where the routing problem is solved first, followed by the optimal control solution of vehicles along the routes.

By implementing a UCT-based approach, it is possible to simultaneously optimize the routing decisions and control of all vehicles. Two examples are solved by both the sequential and the UCT-based approach. In the first benchmark VRP example, the results show that the UCT-based approach can obtain better combined cost than the sequential approach. In the second example of a simulating routing-and-control problem for a small area of NYC, the results show that the total travel distance by the UCT-based approach is longer compared to the sequential approach, but the total energy consumption costs can be reduced by such route decisions.

6.2 Contributions

As a result of RQ1, the following contributions are made:

- A new class of multi-subsystem co-design problems is studied for the first time. In the prior literature, the physical plant design and control are considered as part of a single system structure. In the proposed multi-subsystem structure, coupling of the physical design and control subproblems is studied within each subsystem, as well as across different subsystems.

- A novel multi-level decentralized approach is developed to solve the proposed multi-subsystem co-design problem, which can be regarded as an extension of the approaches for solving single-system co-design problems.

As a result of RQ2, the following contributions are made:

- More general multi-subsystem co-design problems are formulated as a result of RQ2, which relaxes certain assumptions considered on the control subproblems in RQ1. In RQ2, nonlinear dynamics and non-convex constraints on state and control variables can be considered.

- To the best of our knowledge, the direct collocation method has not been implemented in a decentralized manner. As a result of RQ2, the direct collocation method is combined with the optimality condition decomposition method. Accordingly, a multi-level and a bi-level approach is developed to solve the proposed new class of multi-subsystem problems.

As a result of RQ3, the following contribution is made:

- To the best of our knowledge, the outcome from RQ3 is the first study in the literature that considers the continuous-time optimal control of vehicles combined with routing problems.

As a result of RQ4, the following contributions are made:

- The outcome of RQ4 is a more challenging vehicle routing-and-control problem than that considered in RQ3. Here, a continuous-time vehicle control model is embedded within the vehicle's routing model, while the vehicle dynamics are dependent on the loading conditions during the operation. To the best of our knowledge, no such problem considering both routing and control costs has been formulated in the literature.

- A UCT-based approach is adapted to simultaneously solve the routing and control solution of the proposed problem. It is believed that with proper modifications, the UCT-based approach can also be applied to solve other co-design problems with discrete design variables.

6.3 Current Limitations and Possible Future Directions

6.3.1 Optimality Proof of Multi-subsystem Co-design Problems

In this dissertation, the approaches presented in Chapters 2 and 3 can be applied to solve multi-subsystem co-design problems in a decentralized manner. However, there are two limitations. First, the implementations of the approaches

are not optimized for efficiency. Second, there is no theoretical optimality proof of the obtained solution. To address this shortcoming, future study can be considered along two directions:

- To develop decentralized approaches for solving co-design problems, which can obtain a solution that is provably close to a global or local optimum under certain assumptions.

- To identify some classes of multi-subsystem co-design problems with special structures, so that the convexity of the entire problem can be proved globally or in a local neighborhood. This characteristic can be used as the sufficient condition to show that the obtained co-design solution from either centralized or decentralized approaches is globally or locally optimal.

6.3.2 Multi-subsystem Co-design with Model Predictive Control (MPC)

MPC is a numerical approach to obtain control solutions for a system using the system's exact or approximate dynamical model. MPC solves an online finite time-horizon optimal control problem subject to the constraints on the state and control variables, and then the system only executes control for the current timeslot. This procedure repeats until the task is completed. Many problems are yet to be investigated on co-design of engineering systems with MPC. In addition, when multiple subsystems are considered, a decentralized MPC implementation may be preferred under some conditions, e.g., when communications among subsystems are limited. The formulation, solution approaches and potential applications of co-

design with decentralized MPC can be further studied.

6.3.3 Vehicle Routing-and-Control Problems

In Chapters 4 and 5, vehicle routing-and-control problems are formulated and solved. One limitation is that the current UCT-based implementation is not well polished. The runtime of the approach is long if the size of the problem increases. Additionally, the vehicle control model and the routing problem are both assumed to be simple. A number of configurations can be added to make these problems more realistic, e.g.,

- A more realistic model can be considered, for example, comprising engine, transmission, drive shafts, suspension and wheels can be used to obtain energy consumption costs.
- Non-homogeneous vehicles can be considered, including different dynamic models for different vehicles, capacity constraints, restrictions on the cargo types, etc.
- Time issues can be included in the problem, e.g., waiting and loading time delays, and time window constraints for specific customers.
- Real-time traffic information can be included to assist in making routing-and-control decisions.

Finally, the robustness of vehicle planning during operation can be considered. As an example, the problem of optimal re-routing and control of vehicles can be

explored. This problem can become important when one or more vehicles fail to complete the routing assignment according to an initial plan. It is also important to find the optimal or a relatively good solution when considering energy consumption costs. Hence, an efficient algorithm for optimization of the combined routing and control costs for re-planning can be considered.

Appendix A: Derivation of the Decentralized Direct Collocation Method in Section 3.2

In Research Question 2, the discretization steps described in [43] are modified to fit a decentralized formulation.

In the direct collocation method, as shown in Fig. A.1, the value of control variables between every two grid points are approximated as a linear interpolating function, and the values of state variables are approximated by a cubic polynomial function, known as cubic collocation at Lobatto points [43]. Hence, for $t_k \leq t \leq t_{k+1}$,

$$u_i(t) = u_{i,k} + \frac{t - t_k}{s}\left(u_{i,k+1} - u_{i,k}\right) \tag{A1}$$

$$x_i(t) = \sum_{l=0}^{3} c_{i,k}^{(l)} \left(\frac{t - t_k}{s}\right)^l \tag{A2}$$

where $c_{i,k}^{(l)}$ $(l = 0, 1, 2, 3)$ are coefficients of the cubic polynomial,

$$\begin{aligned} c_{i,k}^{(0)} &= x_{i,k} \\ c_{i,k}^{(1)} &= s f_{i,k} \\ c_{i,k}^{(2)} &= -3x_{i,k} - 2s f_{i,k} + 3x_{i,k+1} - s f_{i,k+1} \\ c_{i,k}^{(3)} &= 2x_{i,k} + s f_{i,k} - 2x_{i,k+1} + s f_{i,k+1} \end{aligned} \tag{A3}$$

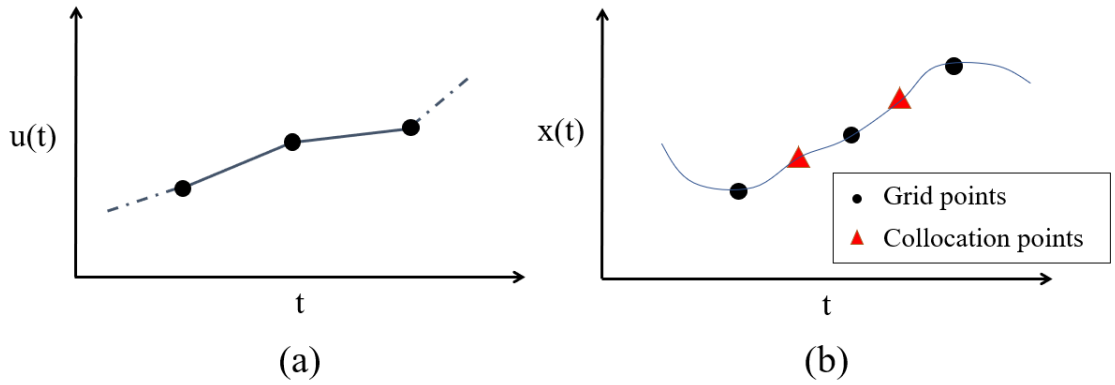


Figure A.1: Direct Collocation Method Discretization: (a) Control - Piecewise Linear (b) State - Piecewise Cubic Polynomial

$$f_{i,k} := f_i(x_{i,k}, u_{i,k}, x_{j,k}, y_i, y_{s_{i,j}})$$

It is then proved [43] that by finding the coefficients $c_{i,k}^{(l)}$, the dynamic equations are satisfied at all grid points. By matching the slopes of the cubic polynomials at the midpoints (or collocation points) at all discretized intervals, the approximating functions are regarded as the solution to the optimal control problem. In the subinterval $t_k \leq t \leq t_{k+1}$, the variables at the collocation point are denoted with the subscript $(k+1/2)$, and are evaluated as [43] $(k = 0, ..., M-1)$,

$$\begin{aligned} u_{i,k+1/2} &= \frac{1}{2}(u_{i,k} + u_{i,k+1}) \\ x_{i,k+1/2} &= \frac{1}{2}(x_{i,k} + x_{i,k+1}) + \frac{s}{8}(f_{i,k} - f_{i,k+1}) \\ \dot{x}_{i,k+1/2} &= \frac{3(x_{i,k+1} - x_{i,k})}{2s} - \frac{1}{4}(f_{i,k} + f_{i,k+1}) \end{aligned} \tag{A4}$$

Hence, the continuous-time problem is transformed to a nonlinear optimization problem, whose constraints include 1) the slopes of approximating polynomials matching

with the value of dynamic equations at the midpoint of each discretized interval, 2) the control constraints for variables at all grid points, and 3) the boundary conditions at $t_0 = 0$ and $t_M = t_f$. The constraints are formulated as,

$$\dot{x}_{i,k+1/2} = f_i(x_{i,k+1/2}, u_{i,k+1/2}, x_{j,k+1/2}) \tag{A5}$$

$$p_i(x_{i,k}, u_{i,k}) \leq 0, \quad q_i(x_{i,1}, x_{i,M}, t_M) = 0 \tag{A6}$$

A defect function $\phi_{i,k}$ for interval k in subsystem i is defined as following,

$$\phi_{i,k} = \dot{x}_{i,k+1/2} - f_i(x_{i,k+1/2}, u_{i,k+1/2}, x_{j,k+1/2}) \tag{A7}$$

Therefore, the constraints of Eq. (A5) can be rewritten as $\phi_{i,k} = 0$. Furthermore, by plugging in the variables in Eq. (A4), $\phi_{i,k}$ can be expressed as,

$$\begin{aligned}\phi_{i,k} =& \frac{3(x_{i,k+1} - x_{i,k})}{2s} - \frac{1}{4}\left(f_{i,k} + f_{i,k+1}\right) \\ & - f_i(x_{i,k+1/2}, u_{i,k+1/2}, x_{j,k+1/2}) = 0\end{aligned} \tag{A8}$$

Note that $\phi_{i,k}$ is related to the variable values of SSj at the grid and the collocation points, i.e., $x_{j,k}$, $x_{j,k+1/2}$ and $x_{j,k+1}$.

Bibliography

[1] Fathy, H. K., Reyer, J. A., Papalambros, P. Y., and Ulsoy, A. G., 2001. On the Coupling between the Plant and Controller Optimization Problems. In the Proceedings of the 2001 American Control Conference, IEEE, Jun 25–27, Arlington, VA, Vol. 3, pp. 1864–1869.

[2] Allison, J. T., and Herber, D. R., 2014. Multidisciplinary Design Optimization of Dynamic Engineering Systems. AIAA Journal, 52(4), pp. 691–710.

[3] Allison, J. T., and Nazari, S., 2010. Combined Plant and Controller Design using Decomposition-based Design Optimization and the Minimum Principle. In the Proceedings of the 2010 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, ASME, Aug. 15–18, Montreal, Quebec, Canada, Vol. 1, pp. 765–774.

[4] Allison, J. T., Guo, T., and Han, Z., 2014. Co-design of an Active Suspension using Simultaneous Dynamic Optimization. Journal of Mechanical Design, 136(8), p. 081003.

[5] Allison, J. T., Herber, D. R., and Deshmukh, A. P., 2014. Integrated Design of Dynamic Sustainable Energy Systems. In the Proceedings of the 20th International Conference on Engineering Design (ICED 15), Vol 1: Design for Life, Jul 27–30, Milan, Italy.

[6] Peters, D. L., Papalambros, P., and Ulsoy, A., 2011. Control Proxy Functions for Sequential Design and Control Optimization. Journal of Mechanical Design, 133(9), p. 091007.

[7] Peters, D. L., Papalambros, P. Y., and Ulsoy, A. G., 2013. Sequential Co-design of an Artifact and its Controller via Control Proxy Functions. Mechatronics, 23(4), pp. 409–418.

[8] Jiang, Y., Wang, Y., Bortoff, S. A., and Jiang, Z.-P., 2015, Optimal Codesign of Nonlinear Control Systems Based on a Modified Policy Iteration Method. IEEE Transaction on Neural Networks Learning Systems, 26(2), pp. 409–414.

[9] Ravichandran, T., Wang, D., and Heppler, G., 2006. Simultaneous Plant-controller Design Optimization of a Two-link Planar Manipulator. Mechatronics, 16(3), pp. 233–242.

[10] Sandoval, L.R., Budman, H.M. and Douglas, P.L., 2008. Simultaneous Design and Control of Processes under Uncertainty: A Robust Modelling Approach. Journal of Process Control, 18(7), pp. 735–752.

[11] Alyaqout, S. F., Papalambros, P. Y., and Ulsoy, A. G., 2011. Combined Robust Design and Robust Control of an Electric DC Motor. IEEE/ASME Transactions on Mechatronics, 16(3), pp. 574–582.

[12] Reyer, J. A., and Papalambros, P. Y., 1999. Optimal Design and Control of an Electric DC Motor. In the Proceedings of the 1999 Design Automation Conference, ASME, Sept. 12–15, Las Vegas, NV, USA.

[13] Deshmukh, A. P., and Allison, J. T., 2013. Design of Nonlinear Dynamic Systems using Surrogate Models of Derivative Functions. In the Proceedings of the 2013 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, ASME, Aug. 4–7, Portland, OR, USA.

[14] Zhang, W., Li, Q., and Guo, L., 1999. Integrated Design of Mechanical Structure and Control Algorithm for a Programmable Four-bar Linkage. IEEE/ASME Transactions on Mechatronics, 4(4), pp. 354–362.

[15] Yan, H.-S., and Yan, G.-J., 2009. Integrated Control and Mechanism Design for the Variable Input-speed Servo Four-bar Linkages. Mechatronics, 19(2), pp. 274–285.

[16] Sridharan, S., Echols, J., Rodriguez, A., Mondal, K., et al., 2014. Integrated Design and Control of Hypersonic Vehicles. In the Proceedings of the 2014 American Control Conference, IEEE, Jun 4–6, Portland, OR, Vol. 3, pp. 1371–1376.

[17] Kim, M.-J., and Peng, H., 2006. Combined Control/plant Optimization of Fuel Cell Hybrid Vehicles. In the Proceedings of the 2006 American Control Conference, IEEE, Jun 14–16, Minneapolis, MN, USA.

[18] Li, H., Yan, W., and Shi, Y., 2018, Triggering and Control Codesign in Self-Triggered Model Predictive Control of Constrained Systems: With Guaranteed Performance. IEEE Transaction on Automatic Control, 63(11), pp. 4008–4015.

[19] Xu, Y., Cervin, A., and Arzen, K.E., 2018. Jitter-Robust LQG Control and Real-Time Scheduling Co-Design. In the Proceedings of the 2018 American Control Conference, IEEE, Jun 27–29, Milwaukee, WI, pp. 3189–3196.

[20] Stipanovic, D.M., Inalhan, G., Teo, R. and Tomlin, C.J., 2004. Decentralized Overlapping Control of a Formation of Unmanned Aerial Vehicles. Automatica, 40(8), pp. 1285–1296.

[21] Geoffrion, A.M., 1972. Generalized Bender's Decomposition. Journal of Optimization Theory and Applications, 10(4), pp. 237–260.

[22] Tosserams, S., Etman, L.F.P. and Rooda, J.E., 2007. An Augmented Lagrangian Decomposition Method for Quasi-separable Problems in MDO. Structural and Multidisciplinary Optimization, 34(3), pp. 211–227.

[23] Tosserams, S., Etman, L.F.P. and Rooda, J.E., 2008. Augmented Lagrangian Coordination for Distributed Optimal Design in MDO. International Journal for Numerical Methods in Engineering, 73(13), pp. 1885–1910.

[24] Geromel, J.C. and Bernussou, J., 1982. Optimal Decentralized Control of Dynamic Systems. Automatica, 18(5), pp. 545–557.

[25] Nedic, A., and Ozdaglar, A., 2009. Distributed Subgradient Methods for Multi-agent Optimization. IEEE Transactions on Automatic Control, 54(1), pp. 48–61.

[26] Nedic, A., Ozdaglar, A., and Parrilo, P. A., 2010. Constrained Consensus and Optimization in Multi-agent Networks. IEEE Transactions on Automatic Control, 55(4), pp. 922–938.

[27] Rantzer, A., 2009. Dynamic Dual Decomposition for Distributed Control. In the Proceedings of the 2009 American Control Conference, IEEE, Jun 10–12, St. Louis, MO, pp. 884-888.

[28] Maasoumy, M., Pinto, A. and Sangiovanni-Vincentelli, A., 2011. Model-based Hierarchical Optimal Control Design for HVAC Systems. In the Proceedings of the 2011 Dynamic System Control Conference and Bath/ASME Symposium on Fluid Power and Motion Control, ASME, Oct 31–Nov 2, Arlington, VA.

[29] Jamshidi, M., 1996. Large-scale Systems: Modeling, Control, and Fuzzy Logic. Prentice-Hall, Inc.

[30] Takahara, Y., 1965. A Multi-level Structure for a Class of Dynamical Optimization Problems. Master's dissertation, Case Western Reserve University.

[31] Cohen, G. and Joalland, G., 1976. Coordination Methods by the Prediction Principle in Large Dynamic Constrained Optimization Problems. IFAC Proceedings Volumes, 9(3), pp. 539–547.

[32] Smith, N.J. and Sage, A.P., 1973. An Introduction to Hierarchical Systems Theory. Computers & Electrical Engineering, 1(1), pp. 55–71.

[33] Sadati, N. and Babazadeh, A., 2006. Optimal Control of Robot Manipulators with a New Two-level Gradient-based Approach. Electrical Engineering, 88(5), pp. 383–393.

[34] Osiadacz, A.J. and Bell, D.J., 1986. A Simplified Algorithm for Optimization of Large-scale Gas Networks. Optimal Control Applications and Methods, 7(1), pp. 95–104.

[35] Joalland, G. and Cohen, G., 1980. Optimal Control of a Water Distribution Network by Two Multilevel Methods. Automatica, 16(1), pp. 83–88.

[36] Betts, J. T., 1998. Survey of Numerical Methods for Trajectory Optimization. Journal of Guidance, Control, and Dynamics, 21(2), pp. 193–207.

[37] Rao, A. V., 2009. A Survey of Numerical Methods for Optimal Control. Advances in the Astronautical Sciences, 135(1), pp. 497–528.

[38] Gerdts, M., 2003. Direct Shooting Method for the Numerical Solution of Higher-Index DAE Optimal Control Problems. Journal of Optimization Theory and Applications, 117(2), pp. 267–294.

[39] Diehl, M., Bock, H. G., Diedam, H., and Wieber, P.-B., 2006. Fast Direct Multiple Shooting Algorithms for Optimal Robot Control. In Fast Motions in Biomechanics and Robotics, Springer, pp. 65–93.

[40] Fahroo, F., and Ross, I. M., 2002. Direct Trajectory Optimization by a Chebyshev Pseudospectral Method. Journal of Guidance, Control, and Dynamics, 25(1), pp. 160–166.

[41] Benson, D., 2005. A Gauss Pseudospectral Transcription for Optimal Control. Ph.D. dissertation, Massachusetts Institute of Technology, Cambridge, MA.

[42] Francolin, C., and Rao, A, 2011. Direct Trajectory Optimization and Costate Estimation of State Inequality Path-constrained Optimal Control Problems Using a Radau Pseudospectral Method. In the Proceedings of the 2012 Guidance, Navigation, and Control Conference, AIAA, Aug. 13–16, Minneapolis, MN, pp. 4528.

[43] Von Stryk, D. M. O., 1993. Numerical Solution of Optimal Control Problems by Direct Collocation. In Optimal Control, Springer, pp. 129–143.

[44] Herber, D.R., 2017. Advances in Combined Architecture, Plant, and Control Design. Ph.D. dissertation, University of Illinois at Urbana-Champaign, Champaign, IL.

[45] Chilan, C. M., Herber, D. R., Nakka, Y. K., Chung, S. J., Allison, J. T., Aldrich, J. B., and Alvarez-Salazar, O. S., 2016. Co-Design of Strain Actuated Solar Arrays for Precision Pointing and Jitter Reduction. AIAA Journal, 55(9), pp. 3180–3195.

[46] Spielberg, A., Araki, B., Sung, C., Tedrake, R., and Rus, D., 2017. Functional Co-optimization of Articulated Robots. In the Proceedings of the 2017 International Conference on Robotics and Automation, IEEE, Singapore, pp. 5035–5042.

[47] Geiger, B., Horn, J., DeLullo, A., Niessner, A., and Long, L., 2006. Optimal Path Planning of UAVs Using Direct Collocation with Nonlinear Programming. In the Proceedings of the 2006 Guidance, Navigation, and Control Conference, AIAA, Aug. 21–24, Keystone, CO, pp. 6199.

[48] Toth, P. and Vigo, D. Eds., 2014. Vehicle Routing: Problems, Methods, and Applications (Vol. 18). SIAM.

[49] Golden, B.L., Raghavan, S. and Wasil, E.A. eds., 2008. The Vehicle Routing Problem: Latest Advances and New Challenges (Vol. 43). Springer Science & Business Media.

[50] Laporte, G., 1992. The Vehicle Routing Problem: An Overview of Exact and Approximate Algorithms. European Journal of Operational Research, 59(3), pp. 345–358.

[51] Toth, P. and Vigo, D., 2002. Models, Relaxations and Exact Approaches for the Capacitated Vehicle Routing Problem. Discrete Applied Mathematics, 123(1-3), pp. 487–512.

[52] Ralphs, T.K., Kopman, L., Pulleyblank, W.R. and Trotter, L.E., 2003. On the Capacitated Vehicle Routing Problem. Mathematical Programming, 94(2-3), pp. 343–359.

[53] Baldacci, R., Mingozzi, A. and Roberti, R., 2012. Recent Exact Algorithms for Solving the Vehicle Routing Problem under Capacity and Time Window Constraints. European Journal of Operational Research, 218(1), pp. 1–6.

[54] Kallehauge, B., Larsen, J., Madsen, O.B. and Solomon, M.M., 2005. Vehicle Routing Problem with Time Windows. Column Generation (67-98). Springer, Boston, MA.

[55] Scanlon, R., Wang, Q. and Wang, J., 2016. Ant Colony Optimization Model for Vehicle Routing Problem with Simultaneous Pickup and Delivery. In the Proceedings of the 2016 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, ASME, Aug. 21–24, Charlotte, NC, USA.

[56] Manyam, S.G., Rathinam, S., Darbha, S. and Obermeyer, K.J., 2012. Computation of a Lower Bound for a Vehicle Routing Problem with Motion Constraints. In the Proceedings of the 2012 Dynamic Systems and Control Conference, ASME, Oct. 17–19, Fort Lauderdale, FL, Vol. 45301, pp. 695–701.

[57] Manyam, S.G., Rathinam, S. and Darbha, S., 2015. Computation of Lower Bounds for a Multiple Depot, Multiple Vehicle Routing Problem with Motion Constraints. Journal of Dynamic Systems, Measurement, and Control, 137(9), p.094501.

[58] Scholz-Reiter, B., Rekersbrink, H., Wenning, B.L. and Makuschewitz, T., 2008. A Survey of Autonomous Control Algorithms by Means of Adapted Vehicle Routing Problems. In the Proceedings of the 2008 Biennial Conference on Engineering Systems Design and Analysis, ASME, Jul. 7–9, Haifa, Israel, pp. 411–416.

[59] Barth, M., Younglove, T., Scora, G., 2005. Development of a Heavy-duty Diesel Modal Emissions and Fuel Consumption Model. Technical Report UCB-ITS-PRR-2005-1, California PATH Program, Institute of Transportation Studies, University of California at Berkeley.

[60] Barth, M., Boriboonsomsin, K., 2009. Energy and Emissions Impacts of a Freeway-based Dynamic Eco-driving System. Transportation Research Part D: Transport and Environment, 14 (6), pp. 400–410.

[61] Bektaş, T. and Laporte, G., 2011. The Pollution-Routing Problem. Transportation Research Part B: Methodological, 45(8), pp. 1232–1250.

[62] Xiao, Y., Zhao, Q., Kaku, I. and Xu, Y., 2012. Development of a Fuel Consumption Optimization Model for the Capacitated Vehicle Routing Problem. Computers & Operations Research, 39(7), pp. 1419–1431.

[63] Franceschetti, A., Honhon, D., Van Woensel, T., Bektaş, T. and Laporte, G., 2013. The Time-dependent Pollution-routing Problem. Transportation Research Part B: Methodological, 56, pp. 265–293.

[64] Jabali, O., Woensel, T. and De Kok, A.G., 2012. Analysis of Travel Times and CO_2 Emissions in Time-dependent Vehicle Routing. Production and Operations Management, 21(6), pp. 1060–1074.

[65] Qian, J. and Eglese, R., 2014. Finding Least Fuel Emission Paths in a Network with Time-varying Speeds. Networks, 63(1), pp. 96–106.

[66] Qian, J. and Eglese, R., 2016. Fuel Emissions Optimization in Vehicle Routing Problems with Time-varying Speeds. European Journal of Operational Research, 248(3), pp. 840–848.

[67] Fukasawa, R., He, Q., Santos, F. and Song, Y., 2016. A Joint Vehicle Routing and Speed Optimization Problem. INFORMS Journal on Computing, 30(4), pp. 694–709.

[68] Demir, E., Bektaş, T. and Laporte, G., 2014. The Bi-objective Pollution-routing Problem. European Journal of Operational Research, 232(3), pp. 464–478.

[69] Oberscheider, M., Zazgornik, J., Henriksen, C.B., Gronalt, M. and Hirsch, P., 2013. Minimizing Driving Times and Greenhouse Gas Emissions in Timber Transport with a Near-exact Solution Approach. Scandinavian Journal of Forest Research, 28(5), pp. 493–506.

[70] Baum, M., Dibbelt, J., Hübschle-Schneider, L., Pajor, T. and Wagner, D., 2014. Speed-consumption Tradeoff for Electric Vehicle Route Planning. In OASIcs-OpenAccess Series in Informatics (Vol. 42). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.

[71] Schneider, M., Stenger, A. and Goeke, D., 2014. The Electric Vehicle-routing Problem with Time Windows and Recharging Stations. Transportation Science, 48(4), pp. 500–520.

[72] Arslan, O., Yıldız, B. and Karaşan, O.E., 2015. Minimum Cost Path Problem for Plug-in Hybrid Electric Vehicles. Transportation Research Part E: Logistics and Transportation Review, 80, pp. 123–141.

[73] Lin, J., Zhou, W. and Wolfson, O., 2016. Electric Vehicle Routing Problem. Transportation Research Procedia, 12, pp. 508–521.

[74] Worley, O., Klabjan, D. and Sweda, T.M., 2012. Simultaneous Vehicle Routing and Charging Station Siting for Commercial Electric Vehicles. In Electric Vehicle Conference (IEVC), 2012 IEEE International (1-3). IEEE.

[75] Zachariadis, E.E., Tarantilis, C.D. and Kiranoudis, C.T., 2015. The Load-dependent Vehicle Routing Problem and its Pick-up and Delivery Extension. Transportation Research Part B: Methodological, 71, pp. 158–181.

[76] Mańdziuk, J. and Nejman, C., 2015. UCT-based Approach to Capacitated Vehicle Routing Problem. In International Conference on Artificial Intelligence and Soft Computing, Springer, Cham, pp. 679–690.

[77] Edelkamp, S., Gath, M., Greulich, C., Humann, M., Herzog, O. and Lawo, M., 2016. Monte-Carlo Tree Search for Logistics. In Commercial Transport. Springer, Cham, pp. 427–440.

[78] Mańdziuk, J. and Świechowski, M., 2017. UCT in Capacitated Vehicle Routing Problem with Traffic Jams. Information Sciences, 406, pp. 42–56.

[79] Peters, D. L., Papalambros, P. Y., and Ulsoy, A. G., 2009. On Measures of Coupling between the Artifact and Controller Optimal Design Problems. In the Proceedings of the 2009 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference, ASME, Aug. 30–Sept. 2, San Diego, CA, USA, pp. 1363–1372.

[80] Boyd, S., Xiao, L., Mutapcic, A., and Mattingley, J., 2007. Notes on Decomposition Methods. Stanford University, Stanford, CA, accessed Dec. 25, 2019, https://web.stanford.edu/class/ee364b/lectures/decomposition_notes.pdf.

[81] Rutquist, P. E., and Edvall, M. M., 2010. Propt-MATLAB Optimal Control Software. TOMLAB Optimization, Inc, Pullman, Washington.

[82] Azarm, S., and Papalambros, P., 1982. An Interactive Design Procedure for Optimization of Helical Compression Springs. Technical Report No. UMMEAM-82-7, University of Michigan, Ann Arbor.

[83] Conejo, A. J., Nogales, F. J., and Prieto, F. J., 2002, A Decomposition Procedure Based on Approximate Newton Directions. Mathematical Programming, 93(3), pp. 495–515.

[84] Moin, P., 2010, Fundamentals of Engineering Numerical Analysis, Cambridge University Press.

[85] Solomon, M. M., 1987. Algorithms for the Vehicle Routing and Scheduling Problems with Time Window Constraints. Operations Research, 35(2), pp. 254–265.

[86] Bock, H.G. and Plitt, K.J., 1984. A Multiple Shooting Algorithm for Direct Solution of Optimal Control Problems. IFAC Proceedings Volumes, 17(2), pp. 1603–1608.

[87] Gurobi Optimization, 2019. Gurobi Optimizer Reference Manual, Ver. 9, Gurobi Optimization, LLC.

[88] Hucho, W.H. and Sovran, G., 1993. Aerodynamics of Road Vehicles. Annual Review of Fluid Mechanics, 25(1), pp. 485–537.