REAL-TIME CONTROL STRATEGIES IN TRANSIT OPERATIONS: MODELS AND ANALYSIS

by

XU JUN EBERLEIN

B.Eng. Mech. Eng., Chongqing University, P.R. China, 1982
M.S. Systems Sciences, Chinese Academy of Sciences, P.R. China, 1988
M.S. Transportation, Massachusetts Institute of Technology, 1992

Submitted to the Department of Civil and Environmental Engineering
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June, 1995

© 1995 Massachusetts Institute of Technology
All rights reserved

Signature of Author

Certified by

Nigel H. M. Wilson
Professor of Civil and Environmental Engineering
Thesis Supervisor

Certified by

David H. Bernstein
Assistant Professor of Civil Engineering and Operations Research
Princeton University
Thesis Supervisor

Accepted by

Joseph M. Sussman
Chairman, Departmental Committee on Graduate Studies
REAL-TIME CONTROL STRATEGIES IN TRANSIT OPERATIONS: MODELS AND ANALYSIS

by

XU JUN EBERLEIN

Submitted to the Department of Civil and Environmental Engineering in April of 1995 in Partial Fulfillment of the Requirements for the Degree of Doctor of Philosophy in Transportation

ABSTRACT

Transit operations are subject to various influences and conditions that can result in variable headways, long waiting times and generally poor performance. To accommodate this transit agencies have used and continue to use a variety of real-time control strategies intended to improve performance. Practice shows that it is difficult to apply these strategies effectively based on personal judgment and poor real-time information. On the other hand, very few of these strategies have been properly studied in order to develop effective control policies or to understand their impacts on transit performance, especially with real-time information availability. This dissertation studies the strategies of deadheading, expressing and holding both singly and in combination. Models are formulated around two types of transit systems to study these strategies. The first type is a simplified abstract transit operation environment for which properties and solutions of the control problems can be studied analytically. The second type removes several simplifying assumptions and, though the control problems cannot be solved analytically, they are shown also to have important properties related to those of the simpler system. For expressing and deadheading the degree of correspondence between the two systems depends heavily on passenger demand patterns. For all strategies in both systems, the optimal control policies are strongly impacted by vehicle headway patterns at the initial station. The understandings gained are used to develop algorithms and formulate effective and efficient real time control policies. Using data from the MBTA Green Line these algorithms are implemented and studied. The advantages, disadvantages, and conditions for effectiveness of each strategy are investigated. The most effective and least disruptive control policies are found to result from the use of coordinated combined strategies. The control algorithms developed are also tested in situations where there are significant stochastic disturbances which are not captured in the models. Under these random conditions performance is found to decline only moderately. The results in this thesis provide important information regarding the design and implementation of real time control systems and help shed light on the potential improvements achievable with more advanced control systems.

Thesis Supervisors: Dr. Nigel H. M. Wilson
Title: Professor of Civil and Environmental Engineering, MIT
Dr. David Bernstein
Title: Assistant Professor of Civil Engineering and Operations Research, Princeton University
ACKNOWLEDGMENTS

I am lucky to have a supportive and encouraging thesis committee. My supervisor Professor Nigel Wilson is a dedicated transit expert and an honorable and devoted educator. His thoughtful comments and meticulous reading have greatly enhanced the quality of this work. Professor David Bernstein, my other supervisor, asked many insightful questions and provided great assistance in developing and refining the formulations that are fundamental to this work. Professor Cynthia Barnhart provided wonderful ideas at the initial stage of this research that helped me shape the overall approach.

Richard Macchi of the MBTA, who did pioneering work on expressing while at MIT, kindly provided operating information about the Green Line as well as all of the AVI data that are used in the computational testing of control strategies. Ventana Systems provided the Vensim software and adapted it to make vehicle trajectory display graphs possible.

I would like to thank the people who I have shared office space with in various stages of my study at MIT: Tilly Chang, Prodyut Dutt, Francisco Jauffred, Dale Lewis and Ibrahim El Sanhouri who provided warm encouragement. Special thanks to Adriana Bernardino, Denise Brehm, Hai-Ping Xu and Dinesh Gopinath who provided friendship, support and sympathetic ears as the work progressed.

Finally, I would like to thank my husband Bob and my daughter Sonya (now six) whose love, support and typographic assistance have sustained me throughout this long undertaking.
# TABLE OF CONTENTS

Abstract, 2  
Acknowledgments, 3  
Table of Contents, 4  
List of Tables, 7  
List of Figures, 9  

## Chapter 1 INTRODUCTION  
1.1. Motivation and Research Objectives, 11  
1.2. Overview of Prior Research on Control Strategies, 13  
  1.2.1. Holding, 14  
  1.2.2. Speed Control, 19  
  1.2.3. Station Skipping, 20  
1.3. Research Scope and Methodology, 23  
  1.3.1. Objective of Control, 23  
  1.3.2. Control Environment and Process, 24  
  1.3.3. Elements of Concern in Modeling, 27  
  1.3.4. Deterministic Models and Stochastic Tests, 36  
  1.3.5. Research Philosophy and System Types, 37  
1.4. Input Data Preparation, 38  
1.5. Thesis Organization, 40  

## Chapter 2 PROBLEM SETTING AND NOTATION  
2.1. The Transit Network, 41  
2.2. Demand Characteristics, 44  
2.3. Passenger/Vehicle Interaction, 46  
2.4. System Types, 46  
2.5. Objective Function, 47  

## Chapter 3 SYSTEM TYPES AND THEIR PROPERTIES  
3.1. System F, 49  
3.2. System G, 52  
  3.2.1. Dwell Time Effects, 54  
  3.2.2. Dispatching Headway Effects, 56  
  3.2.3. Trajectory Change Diminishing Effect, 60  
  3.2.4. Alternating Sign Effect, 62  

## Chapter 4 THE REAL-TIME DEADHEADING PROBLEM  
4.1. Introduction, 64  
4.2. The General RTDP Model, 65  
  4.2.1. Decision Variables and Objective Function, 65  
  4.2.2. The RTDP Formulation, 67
4.3. Optimal Deadheading in System F, 69
   4.3.1. The RTDPF Formulation, 70
   4.3.2. Properties of RTDPF, 71
   4.3.3. Subproblem 1-DH and Solution Properties, 76
   4.3.4. Exact Solution Algorithm for RTDPF, 84
   4.3.5. Numerical Examples, 88
   4.3.6. Effects of Adjacent Deadheading, 92
   4.3.7. Computational Tests with Green Line Data, 97

4.4. Deadheading in System G, 104
   4.4.1. The RTDGG Formulation, 105
   4.4.2. Effects of Impact Set Size, 109
   4.4.3. Comparison between Different Solutions Methods, 113
   4.4.4. Cost Characteristics, 115
   4.4.5. Effectiveness of Deadheading vs. Choice of m, 122
   4.4.6. Conditions for a Beneficial Deadheading Policy, 124
   4.4.7. Comparison between System G and System F Results 128
   4.4.8. An Efficient Algorithm for RTDGG, 132

Chapter 5 THE REAL-TIME EXPRESSING PROBLEM 135
   5.1. Introduction, 135
   5.2. The general Model of RTEP, 136
   5.3. Expressing in System F, 139
      5.3.1. The 1-EXP Model and Properties, 139
      5.3.2. Comparison with Deadheading, 143
      5.3.3. Computational Results with Green Line Data, 146
   5.4. Expressing in System G, 151
      5.4.1. Variable Definitions, 151
      5.4.2. Solution Structure of 1-EXP subproblem in RTEPG, 156
      5.4.3. Riding Time Savings, 169
      5.4.4. Impact of Capacity Constraint, 171
      5.4.5. Comparison between Expressing and Deadheading, 173
      5.4.6. Computer Algorithm and Empirical Guidelines for Expressing, 178

Chapter 6 THE HOLDING PROBLEM 181
   6.1. Introduction, 181
   6.2. Optimal Holding in System F, 183
      6.2.1. Model Formulation of HPF, 183
      6.2.2. Solution Algorithm for HPF, 185
      6.2.3. Numerical Examples, 190
      6.2.4. Computational Results, 195
6.3. Holding in System G, 200
   6.3.1. Model Formulation, 200
   6.3.2. Solution Algorithm, 201
   6.3.3. Computational Results for Holding at Station k0, 203
   6.3.4. Solution Property of HPG, 210
   6.3.5. Choice of Holding Station, 211
   6.3.6. Impact of Holding on Passenger Riding Times, 217

Chapter 7  THE COMBINED CONTROL PROBLEM  218

7.1. Introduction, 218
7.2. Combined Control in System F, 219
   7.2.1. Model Formulation and Properties, 219
   7.2.2. Algorithms for CCPF, 227
   7.2.3. Computational Results, 230
7.3. Combined Control in System G, 233
   7.3.1. Heuristic Algorithm for CCPG, 234
   7.3.2. Computational Results, 235
7.4. Effects of Randomness, 243
   7.4.1. Sensitivity Test Design, 244
   7.4.2. Results of Sensitivity Tests, 250

Chapter 8  CONCLUSIONS AND FUTURE RESEARCH  254

8.1. Conclusions, 254
   8.1.1. System Properties, 255
   8.1.2. Properties and Effectiveness of Control Strategies in System F, 256
   8.1.3. Properties and Effectiveness of Control Strategies in System G, 257
   8.1.4. Model Sensitivity to Randomness, 260
   8.1.5. Extensions of the Models, 261
8.2. Suggestions for Future Research, 262
   8.2.1. Extensions of Research, 262
   8.2.2. Related Areas, 263
8.3. Concluding Remarks, 263

Appendix A: Summary of Notation  265
Appendix B: Demand Profile  267
Appendix C: Dispatching Headway Profiles  268

References  272
## List of Tables

<table>
<thead>
<tr>
<th>Table</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1.1</td>
<td>Unit Composite Travel Times Computed from Green Line AVI Data</td>
<td>32</td>
</tr>
<tr>
<td>Table 3.1</td>
<td>No-Control Status: System F vs System G</td>
<td>56</td>
</tr>
<tr>
<td>Table 4.3.1</td>
<td>Optimal Solution to 1-DH subproblems in Example 4.3.2</td>
<td>90</td>
</tr>
<tr>
<td>Table 4.3.2</td>
<td>Results of Example 4.3.2 with Adjacent Deadheading</td>
<td>92</td>
</tr>
<tr>
<td>Table 4.3.3</td>
<td>Headway Pattern of Adjacent Deadheading Pairs in Example 4.3.2</td>
<td>93</td>
</tr>
<tr>
<td>Table 4.3.4</td>
<td>System F: Test 1 Results (without Adjacent Deadheading)</td>
<td>98</td>
</tr>
<tr>
<td>Table 4.3.5</td>
<td>System F: Test 2 Results (with Adjacent Deadheading)</td>
<td>99</td>
</tr>
<tr>
<td>Table 4.3.6</td>
<td>System F: Discarded Adjacent Deadheading in Test 1</td>
<td>100</td>
</tr>
<tr>
<td>Table 4.3.7a</td>
<td>System F: Adjacent Deadheading Pairs (m=3)</td>
<td>101</td>
</tr>
<tr>
<td>Table 4.3.7b</td>
<td>System F: Adjacent Deadheading Pairs (m=1)</td>
<td>101</td>
</tr>
<tr>
<td>Table 4.4.1</td>
<td>RTDPG: Total Cost Reduction by Rolling Horizon Size</td>
<td>112</td>
</tr>
<tr>
<td>Table 4.4.2</td>
<td>Comparison between Different Solution Methods for RTDPG</td>
<td>114</td>
</tr>
<tr>
<td>Table 4.4.3</td>
<td>System G: Adjacent Deadheading Cases in 3-DH Solutions</td>
<td>114</td>
</tr>
<tr>
<td>Table 4.4.4</td>
<td>System G: Adjacent Deadheading Cases in Control-the-First Solutions (m=3)</td>
<td>115</td>
</tr>
<tr>
<td>Table 4.4.5</td>
<td>Vehicle Sets with Non-quasiconvex Set Cost when m=3</td>
<td>119</td>
</tr>
<tr>
<td>Table 4.4.6</td>
<td>System G: Deadhead Vehicles with ( h_{i,k}/h_{i+1,k} &lt; 1 ) at ( k=k_0 ), m=3</td>
<td>125</td>
</tr>
<tr>
<td>Table 5.3.1</td>
<td>Example 5.3.1: Comp. between Deadheading and Expressing in System F</td>
<td>145</td>
</tr>
<tr>
<td>Table 5.3.2</td>
<td>System F: Expressing Results from Green Line Data</td>
<td>147</td>
</tr>
<tr>
<td>Table 5.3.3</td>
<td>System F: Comparison between Expressing and Deadheading (1)</td>
<td>148</td>
</tr>
<tr>
<td>Table 5.3.4</td>
<td>System F: Comparison between Expressing and Deadheading (2)</td>
<td>149</td>
</tr>
<tr>
<td>Table 5.3.5</td>
<td>Optimal Initiation Station in 1-EXP2 Subproblems</td>
<td>149</td>
</tr>
<tr>
<td>Table 5.4.1</td>
<td>System G: Optimal Express Segments, m=1</td>
<td>159</td>
</tr>
<tr>
<td>Table 5.4.2</td>
<td>Exceptional Express Segments in Dir. 2: Conditions before Expressing</td>
<td>162</td>
</tr>
<tr>
<td>Table 5.4.3</td>
<td>System G: Expressing Results for Different Sets of ks</td>
<td>164</td>
</tr>
<tr>
<td>Table 5.4.4</td>
<td>Expressing Results with Different Impact Set Sizes</td>
<td>168</td>
</tr>
<tr>
<td>Table 5.4.5</td>
<td>System G: Impacts of Threshold Benefit on Expressing Policies</td>
<td>169</td>
</tr>
<tr>
<td>Table 5.4.6</td>
<td>System G: Cost Structure with Different Objective Functions</td>
<td>171</td>
</tr>
<tr>
<td>Table 5.4.7</td>
<td>System G: Expressing under Capacity Constraint (m=3)</td>
<td>172</td>
</tr>
<tr>
<td>Table 5.4.8</td>
<td>System G: Comparison between Expressing and Deadheading (m=1)</td>
<td>174</td>
</tr>
<tr>
<td>Table 5.4.9</td>
<td>System G: Comparison between Expressing and Deadheading (m=3)</td>
<td>176</td>
</tr>
<tr>
<td>Table 5.4.10</td>
<td>System G: Waiting Time vs. Riding Time Savings</td>
<td>177</td>
</tr>
<tr>
<td>Table 5.4.11</td>
<td>System G: Waiting Time vs. Riding Time Savings under Cap. Constraint</td>
<td>178</td>
</tr>
<tr>
<td>Table 6.2.1</td>
<td>HPF Results with Terminal Schedule Constraint (m=3)</td>
<td>196</td>
</tr>
<tr>
<td>Table 6.2.2</td>
<td>HPF Results without Terminal Schedule Constraint (m=3)</td>
<td>196</td>
</tr>
<tr>
<td>Table 6.2.3</td>
<td>System F: Comparison between Rolling Sizes</td>
<td>198</td>
</tr>
<tr>
<td>Table 6.2.4</td>
<td>System F: Comparison between Holding and Deadheading</td>
<td>199</td>
</tr>
<tr>
<td>Table 6.3.1</td>
<td>HPG: Overall Results by Rolling Horizon Size</td>
<td>204</td>
</tr>
<tr>
<td>Table 6.3.2a</td>
<td>HPG Results with Terminal Schedule Constraint (m=3)</td>
<td>206</td>
</tr>
<tr>
<td>Table 6.3.2b</td>
<td>HPG Results without Terminal Schedule Constraint (m=3)</td>
<td>207</td>
</tr>
<tr>
<td>Table 6.3.3</td>
<td>System G: Comparison between Holding and Deadheading</td>
<td>207</td>
</tr>
<tr>
<td>Table 6.3.4</td>
<td>Cost Reduction Resulted from Two Algorithms (m=3)</td>
<td>209</td>
</tr>
<tr>
<td>Table 6.3.5</td>
<td>Results from Two Holding Stations</td>
<td>215</td>
</tr>
<tr>
<td>Table 6.3.6</td>
<td>System G: Riding Time Saving Comparison</td>
<td>217</td>
</tr>
<tr>
<td>Table 7.2.1</td>
<td>CCPF Results with Terminal Schedule Constraint (m=3)</td>
<td>231</td>
</tr>
<tr>
<td>Table</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-------</td>
<td>--------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>7.3.1</td>
<td>Execution Time of Algorithm CCPG</td>
<td>236</td>
</tr>
<tr>
<td>7.3.2</td>
<td>CCPG Results (m=3)</td>
<td>237</td>
</tr>
<tr>
<td>7.4.1</td>
<td>Sensitivity Test Scenarios</td>
<td>248</td>
</tr>
<tr>
<td>7.4.2</td>
<td>Results of Randomness Tests</td>
<td>251</td>
</tr>
</tbody>
</table>
List of Figures

<table>
<thead>
<tr>
<th>Figure</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Figure 1.1</td>
<td>AVI Records: Monday 10/16/93</td>
<td>30</td>
</tr>
<tr>
<td>Figure 1.2</td>
<td>Mean Unit Travel Time in 5 Weekday Morning Peaks</td>
<td>33</td>
</tr>
<tr>
<td>Figure 1.3</td>
<td>Comparison between Unit Travel Time and Demand</td>
<td>35</td>
</tr>
<tr>
<td>Figure 2.1</td>
<td>The Transit Network</td>
<td>42</td>
</tr>
<tr>
<td>Figure 3.1</td>
<td>System F: No-Control Trajectories. Wednesday, Dir 1</td>
<td>52</td>
</tr>
<tr>
<td>Figure 3.2</td>
<td>System G: No-Control Trajectories, Wednesday, Dir 1</td>
<td>53</td>
</tr>
<tr>
<td>Figure 3.3</td>
<td>System G: Headway Variation along the Route (with no Control)</td>
<td>55</td>
</tr>
<tr>
<td>Figure 3.4</td>
<td>Dispatching Headway Effects (i)</td>
<td>58</td>
</tr>
<tr>
<td>Figure 3.5a</td>
<td>System G: Headway Relations (Direction 1)</td>
<td>59</td>
</tr>
<tr>
<td>Figure 3.5b</td>
<td>System G: Headway Relations (Direction 2)</td>
<td>59</td>
</tr>
<tr>
<td>Figure 4.3.1</td>
<td>Trajectory of a Deadheaded Vehicle</td>
<td>75</td>
</tr>
<tr>
<td>Figure 4.3.2</td>
<td>Network Representing Problem of Finding Optimal Set of Deadhead Vehicles</td>
<td>86</td>
</tr>
<tr>
<td>Figure 4.3.3</td>
<td>Example 4.3.1: 1-DH</td>
<td>89</td>
</tr>
<tr>
<td>Figure 4.3.4</td>
<td>Example 4.3.2.: Vehicle Trajectories before Deadheading</td>
<td>91</td>
</tr>
<tr>
<td>Figure 4.3.5</td>
<td>Example 4.3.2.: Vehicle Trajectories after Deadheading</td>
<td>91</td>
</tr>
<tr>
<td>Figure 4.3.6</td>
<td>System F: Cost Change vs. Headway Ratio</td>
<td>103</td>
</tr>
<tr>
<td>Figure 4.3.7</td>
<td>System F: Optimal Solution Solution vs. Headway Pattern (Dir 1)</td>
<td>103</td>
</tr>
<tr>
<td>Figure 4.3.8</td>
<td>System F: Optimal Solution Solution vs. Headway Pattern (Dir 2)</td>
<td>104</td>
</tr>
<tr>
<td>Figure 4.4.1</td>
<td>RTDPG: Week Total Cost Reduction by Rolling Size</td>
<td>113</td>
</tr>
<tr>
<td>Figure 4.4.2</td>
<td>Quasiconvex Curves of 1-DHG Cost</td>
<td>118</td>
</tr>
<tr>
<td>Figure 4.4.3a</td>
<td>Cost Curve for Vehicle 20, f1 (m=3)</td>
<td>120</td>
</tr>
<tr>
<td>Figure 4.4.3b</td>
<td>Cost Curve for Vehicle 20, f1 (m=1)</td>
<td>120</td>
</tr>
<tr>
<td>Figure 4.4.4</td>
<td>Vehicle Trajectories under Different Control Policies</td>
<td>121</td>
</tr>
<tr>
<td>Figure 4.4.5</td>
<td>Example of Abnormal Headway Condition (m=3)</td>
<td>126</td>
</tr>
<tr>
<td>Figure 4.4.6</td>
<td>System G: Headheading Results and Condition (Dir 1)</td>
<td>127</td>
</tr>
<tr>
<td>Figure 4.4.7</td>
<td>System G: Headheading Results and Condition (Dir 2)</td>
<td>128</td>
</tr>
<tr>
<td>Figure 4.4.8a</td>
<td>Comparison between System F and System G (1)</td>
<td>129</td>
</tr>
<tr>
<td>Figure 4.4.8b</td>
<td>Comparison between System F and System G (2)</td>
<td>129</td>
</tr>
<tr>
<td>Figure 4.4.8c</td>
<td>Comparison between System F and System G (3)</td>
<td>130</td>
</tr>
<tr>
<td>Figure 5.1a</td>
<td>Cost Curve Family for Vehicle 11, m1</td>
<td>160</td>
</tr>
<tr>
<td>Figure 5.1b</td>
<td>Cost Curve Family for Vehicle 9, m1</td>
<td>161</td>
</tr>
<tr>
<td>Figure 5.2</td>
<td>Expressing Results and Condition (dir. 1)</td>
<td>165</td>
</tr>
<tr>
<td>Figure 5.3</td>
<td>Expressing Results and Condition (dir. 2)</td>
<td>165</td>
</tr>
<tr>
<td>Figure 6.1a</td>
<td>Comparison Between Rolling Sizes</td>
<td>198</td>
</tr>
<tr>
<td>Figure 6.1b</td>
<td>Comparison Between Rolling Sizes</td>
<td>198</td>
</tr>
<tr>
<td>Figure 6.2</td>
<td>Cost Reduction by Rolling Size</td>
<td>204</td>
</tr>
<tr>
<td>Figure 6.3a</td>
<td>Wednesday: Vehicle Trajectories after Holding with Terminal Constraint</td>
<td>208</td>
</tr>
<tr>
<td>Figure 6.3b</td>
<td>Wednesday: Vehicle Trajectories after Holding without Terminal Constraint</td>
<td>208</td>
</tr>
<tr>
<td>Figure 6.4a</td>
<td>Property of Holding Solutions, m1, Dir. 1, m=3</td>
<td>210</td>
</tr>
<tr>
<td>Figure 6.4b</td>
<td>Property of Holding Solutions, m1, Dir. 2, m=3</td>
<td>211</td>
</tr>
<tr>
<td>Figure 6.5a</td>
<td>Cost Reduction by Holding Station Location (dir 1)</td>
<td>212</td>
</tr>
<tr>
<td>Figure 6.5b</td>
<td>Cost Reduction by Holding Station Location (dir 2)</td>
<td>212</td>
</tr>
<tr>
<td>Figure 6.6a</td>
<td>Comparison of Headway Variation by Station (dir 1)</td>
<td>214</td>
</tr>
<tr>
<td>Figure 6.6b</td>
<td>Comparison of Headway Variation by Station (dir 2)</td>
<td>214</td>
</tr>
<tr>
<td>Figure 6.7</td>
<td>Choice of the Second Holding Station</td>
<td>216</td>
</tr>
<tr>
<td>Figure</td>
<td>Description</td>
<td>Page</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------------------------------------------------------------------------</td>
<td>------</td>
</tr>
<tr>
<td>7.1</td>
<td>Effectiveness of Control Strategies in System F</td>
<td>230</td>
</tr>
<tr>
<td>7.2</td>
<td>Comparison between Single and Combined Control in System F</td>
<td>232</td>
</tr>
<tr>
<td>7.3</td>
<td>Comparison between Single and Combined Control in System G</td>
<td>237</td>
</tr>
<tr>
<td>7.4</td>
<td>System G: Average Holding Time (m=3)</td>
<td>238</td>
</tr>
<tr>
<td>7.5</td>
<td>System G: Magnitude of Control (m=3)</td>
<td>239</td>
</tr>
<tr>
<td>7.6a</td>
<td>Vehicle Trajectories: Th1, System G, No Control</td>
<td>241</td>
</tr>
<tr>
<td>7.6b</td>
<td>Vehicle Trajectories: Th1, System G, After Deadheading</td>
<td>241</td>
</tr>
<tr>
<td>7.6c</td>
<td>Vehicle Trajectories: Th1, System G, After Expressing</td>
<td>242</td>
</tr>
<tr>
<td>7.6d</td>
<td>Vehicle Trajectories: Th1, System G, After Holding</td>
<td>242</td>
</tr>
<tr>
<td>7.6e</td>
<td>Vehicle Trajectories: Th1, System G, After Combined Control</td>
<td>243</td>
</tr>
<tr>
<td>7.7</td>
<td>Cost Reduction after Control under Randomness</td>
<td>241</td>
</tr>
</tbody>
</table>
CHAPTER 1

INTRODUCTION

1.1 Motivation and Research Objectives

It is well known that high frequency urban transit services are inherently irregular. Even if the scheduled headways are constant, the actual headways can be highly variable. Randomness in passenger demand, vehicle dwell times at stations, vehicle dispatching times and interstation running times, along with minor and major disruption from incidents and interference from other traffic, all contribute to service irregularity. Once a disturbance occurs, deviations from a given schedule are often amplified down the line. This can sharply increase passengers’ waiting time and frustration (see for example, Okrent (1974), Abkowitz et al (1978), and Engelstein (1983)).

Real-time operations control is intended to remedy such situations at the time they occur, and thereby reduce passengers cost and frustration. It is for this purpose that the U.S. Strategic Plan for Intelligent Vehicle/Highway Systems that was developed in 1992 identified real-time operations control as a high priority technology. In particular, it called for the development and deployment of this technology in the middle term (i.e., from 1992-2002).
More specifically, real-time operations control is intended to minimize passenger costs given a service schedule and fleet size\(^1\). In both bus and rail transit, the control strategies commonly used for this purpose include holding a vehicle at a station, deadheading or expressing a vehicle over a segment of the line, short-turning a vehicle, and adding vehicles held in reserve. In rail transit, splitting a train can also be used if each smaller train can be operated independently. While all these strategies can be quite effective, they can also negatively affect service quality if used inappropriately. For detailed descriptions of these real-time control strategies, the reader is referred to Wilson et al (1992) and Soeldner (1993).

In many U.S. transit agencies, control decisions are often made by field inspectors at various points along the routes, with only information of vehicles that have already passed at the point. In such a manual decision making process, since control personnel can only observe the local situation, an action intended to reduce current and local variation might result in higher variability in the future or elsewhere in the system. In addition, because the decisions are based on personal experience and judgment, performance can vary greatly across inspectors. Furthermore, in complicated situations it may be beyond human ability to make effective decisions. Finally, it is hard, if not impossible, to coordinate each inspector to make good system-wide decisions.

Over the past fifteen years automated systems, information systems, vehicle control systems, and related technologies have advanced greatly; the 90’s have become the “information era”. During this period, many public transportation agencies have begun employing these emerging technologies to improve their services. Increasingly, rail and bus systems are starting to make use of automatic vehicle location (AVL) systems, automatic vehicle identification (AVI) systems and automatic passenger counters (APC). These systems have the potential to provide accurate real-time information to support operations control. Lam (1994) provides a detailed discussion of applications of these technologies to public transportation.

---

\(^1\) The operating cost is a relatively unimportant issue in operations control mainly because of the fixed labor cost (Wilson et al, 1992).
One of the most important uses of the new technologies in public transportation is for *operations monitoring and control*. "Monitoring" refers to the real-time detection of service problems, and "control" refers to the remediation of these problems. However, while the availability of real-time information has increased, there has been little research on the theory and implementation of real-time control strategies. To date, most of the attention in Advanced Public Transportation Systems (APTS) program has been on the development of new information and communications technologies. In spite of their potential use in real-time control, existing AVL, AVI and APC systems are used more for monitoring than for control purposes. Currently, there are no fully automated real-time operations control systems in use in North America (Schweiger et al., 1994).

Unless research is conducted on the use of these technologies in real-time control, it is unlikely that the goal of the U.S. IVHS strategic plan will be realized. Motivated by this, this dissertation takes the first steps towards developing the necessary "soft" technologies for real-time transit operations control. The objective of this research is two fold: first, to develop models and algorithms for the commonly used control strategies deadheading, expressing, and holding (both singly and in combination), and to study the structure and properties of the optimal solutions. Second, to evaluate systematically these control strategies using realistic data. This evaluation will provide a better understanding of the nature and the effectiveness of the above strategies, and identify conditions under which each of the strategies should be applied. The hope is that with real-time information provided by AVI or AVL systems as input, the algorithms developed in this dissertation will generate optimal or near-optimal control policies efficiently in a computerized control system, and that the analysis of the strategies will also benefit non-computerized and computer aided control systems.

1.2 **Overview of Prior Research on Control Strategies**

In the most general sense, real-time control strategies in transit operations can be divided into three categories: station control, interstation control, and others. The first category
contains two main classes of strategies: holding and station skipping. The second includes speed control, traffic signal preemption, etc. The third includes strategies such as adding vehicles, splitting trains, etc. Strategies in the first category are by far the most often used and tend to be the most effective, and this dissertation focus solely on station control strategies. Within this category of control strategies we focus on holding and, among the station skipping strategies, deadheading and expressing. A common characteristics of the three strategies is that they do not necessarily involve changes in the order of vehicles, whereas another station skipping strategy, short-turning, often results in vehicle passing (Deckoff. 1990). Because of this, short-turning requires a separate form of analysis and is not addressed in this dissertation.

Most prior research on real-time control of bus transit focuses on holding alone, while research on rail transit is primarily on speed control. In fact, in the past 15 years, there was only one U.S. Ph.D. dissertation on transit operations control, Engelstein (1983), who also did not go beyond holding. Although the station skipping strategies such as deadheading, expressing and short turning are frequently used in practice, surprisingly few studies were found on them. Below we review the relevant literature.

1.2.1 Holding

Holding has attracted the most attention in the bus transit control literature, probably because it is the simplest and easiest to implement. It is also the first type of real-time control strategies to be studied analytically (it may even be the first type of real-time control strategies used in transit practice). Since the early 1970's, many papers have attempted to formulate mathematical or simulation models for finding optimal holding policies. Holding can be either schedule-based or headway-based. The former makes more sense for long-headway services, and the latter is appropriate for high frequency (headway < 10 minutes) services. Because most urban transit services are high frequency, we restrict our focus to this type of service in this dissertation. The following papers are notable in this regard: Osuna & Newell (1972), Barnett (1974, 1978), Newell (1974), Koffman et al (1978), Turnquist & Blume (1980), Engelstein (1983), Powell
(1985), and Abkowitz et al (1984, 1986). All of these papers formulate the holding problem to minimize a cost function containing either solely or primarily passenger waiting time, with the decision variables being the optimal threshold headway or holding time at one, or a number of, given stations.

Among the earliest research, Osuna & Newell (1972) used an infinite horizon dynamic programming formulation for an idealized single service point problem, where the objective is to minimize the total wait of all passengers over a long horizon, or the expected wait. Passenger arrivals and vehicle headways are assumed to be described by some continuous probability functions. The decision variable is the optimal threshold headway \( a \), such that if, at the control point, a vehicle's headway is smaller than \( a \), then it is held until \( a \); otherwise it is dispatched immediately. The optimal dispatching policy was found for the one-vehicle case, but they could not solve the problem for the case involving more than one vehicle.

Barnett (1974) considered a network with two terminals and one control stop between them. The main difference from Osuna & Newell (1972) is that, instead of using a general continuous probability distribution which presents serious difficulties, Barnett developed a much simpler two-point, discrete, approximate distribution of vehicle arrival delay. His objective function is to minimize a weighted sum of passengers waiting time and on-board passengers delay time at the control station, and the decision variable is a threshold holding time \( x \). He constructed an algorithm to obtain the optimal holding policy under the approximate distribution. Both Osuna and Newell (1972) and Barnett (1974) have been widely referenced by later studies. In later research, Barnett (1978) studied holding strategy for a one-vehicle shuttle service with a nonlinear cost function. He obtained optimal holding policies analytically in this simple case, but suspected that with such a nonlinear cost function strong analytic results will be rare for systems of several vehicles. Barnett and Kleitman (1978) considered a shuttle service with two terminals and one vehicle, and developed some weaker properties of the optimal holding policy, but concluded that "strong theorems about the form of the optimal policy are
unlikely to be forthcoming. The implications of this toward more detailed models of transportation systems are not altogether encouraging."

Commenting that Barnett (1974) "describes some control strategies designed to prevent pairing, but the magnitude of the pairing effects do not enter explicitly in his model". Newell (1974) for the first time included vehicle interstation run time and dwell time separately in his formulation. He considers a transit route with two vehicles and one control point, and assumes a constant passenger arrival rate during any vehicle headway. The decision variable and the objective function are the same as in Osuna & Newell (1972), but instead of using a continuous vehicle headway distribution, he used diffusion approximations of the first and second moments of vehicle interstation run time and dwell time. He derived an approximate solution of the problem, but noted that the theory described has very limited practical application and considerable work would be required to extend the theory to give recipes for control.

Seemingly discouraged by difficulties with analytical methods, the research community in this area was quiet in the mid 70's. At this time, a number of automatic vehicle monitoring (AVM) systems had been tried in bus systems in Europe and the U.S (Koffman, 1978). Although these AVM systems were not very effective as a means of control, they opened the door to more types of control strategies, including station skipping and signal preemption. In the late 70's and early 80's, more articles on the bus holding problem appeared, and the simulation approach started to dominate the research. The transit systems studied in the research extended from one or two stops to an entire route. Mentioning that in an early simulation study of a London subway route Levene et al (1970) showed that holding alone did not significantly improve the overall level of service, Koffman (1978) developed a simulation model of a single direction bus route with traffic signals, and tested several strategies for real-time control of bus headways. Those strategies were holding, skipping loading (but not unloading), and signal preemption. At any stop, if the headway of a bus is less than a desired headway, it is held until that headway is achieved. If a bus load exceeds a desired level, it skips loading passengers although passengers can always alight. Bus dwell time was explicitly
computed as a linear function of passenger boardings and alightings. An empirical bus travel time distribution was assumed. Passengers were assumed to arrive at a stop at a constant rate. All stops, with the exception of the first, the last, and the middle one, were assumed equally likely to be origins and destinations. A disturbance of \( \pm 0.5 \) (scheduled headway) was added to the dispatching headways of buses. The simulation results showed that holding produces very minor or no improvements in wait times at the expense of longer travel times. The skip loading strategy worsens wait times by more than it improves travel times. Only signal preemption results in benefits to both waiting and travel times.

Turnquist & Blume (1980) adopted the idea of a two point discrete headway distribution from Barnett (1974), but used a more general probability model in their analysis. By analyzing two extreme cases, they provided upper and lower bounds on the effectiveness of holding. These bounds were expressed as a relationship between the headway coefficient of variation and the proportion of passengers delayed. Based on these results, they suggested a simple screening method for choosing the most effective control stop along the route.

In his Ph.D. dissertation, Engelstein (1983) attempted to use empirical and simulated data to establish factors that affect bus headway and running time variations, and evaluated holding strategies using this data. This work was latter published in the papers of Abkowitz & Engelstein (1984) and Abkowitz, Eiger & Engelstein (1986). They developed an empirical headway variation function based on Monte Carlo simulation, and used it to estimate passengers waiting time. The holding problem was formulated as total wait (along the route) minimization, and the decision variables were the optimal control point and threshold value for holding. The results show that (headway-based) holding is fairly effective: with a reduction in total passenger wait between 5% and 15%. The authors also suggested that the optimal holding point should be just before high demand stations.

To summarize, there are several common features of the above research:
(a) The holding policy is of a "threshold" type. That is, the optimization result specifies a fixed threshold headway. Regardless of the difference in no-control headways between vehicles, all vehicles will be held until such a fixed headway if the no-control headway is shorter, or not held at all otherwise. This approach is used probably because estimated headway probability functions are used rather than real-time headway information. With real-time information available, however, the holding time need not be specified as a single optimal value and can vary according to the actual headway of each vehicle.

(b) The authors all emphasize the minimal information requirements of their models. This is understandable because almost all this research was done in the 1970's and early 1980's, long before the "information era" of the 90's. Real-time information in bus systems was either not available or not reliable then. The situation now is quite different: we want to make use of the available real-time information as fully as possible.

(c) To make up for the lack of real-time information, all the authors used some form of probability density functions on vehicle headways and/or running times. These probability functions greatly complicate the holding problem. Even for the simplest type of strategies and idealized transit systems, researchers have shown the holding problem was difficult to analyze, and suggested that analytic models may not be very helpful. However, since today we can replace such probability functions by real-time information, the holding problem may become easier to analyze and solve - perhaps even analytically.

(d) Probably due to the lack of data, few papers explicitly evaluated the dwell time effect on headway variation, or considered it at all. Instead, composite vehicle running time, where the dwell time and interstation travel time were not separated, were used. Koffman (1978) was probably the only one who separated dwell time and interstation travel time in his simulation model. Although Engelstein (1983) concluded in his regression analysis that dwell time makes a significant contribution to bus running time and its variation, he used only the composite running time in his holding models. The missing dwell time effect in holding models is in sharp contrast to the well known fact that variable vehicle
dwell times, due to variability of passenger demand across stations, are a major cause of headway variation both along a route and in the time dimension. Thus such dwell time effects warrant closer examination.

We also want to point out that, among the researchers, opinions were quite varied and even contradictory on the effectiveness of holding and the significance of the dwell time effect on headway variation. In this dissertation we will draw our own conclusions on these issues.

1.2.2 Speed Control

Speed control is a dominant control strategy studied in the rail transit literature. In the 80's, many papers addressing rail transit operations control appeared in the literature under the keywords "urban railways" and "traffic control". This research usually attempted to find control policies influencing interstation running time of vehicles. While this dissertation focuses on station control strategies only, we are interested in the differences and similarities between speed control and holding (dwell time control), and also in the research methodology used.

Among these papers, three are most relevant to this research. To regulate vehicle headways on an urban railway, Sasama & Ohkawa (1983) developed a state feedback model, in which the state variables are the deviations of actual departure times from the no control-no disturbance departure times of vehicles at all stations, and the control variables are (+) changes of vehicle interstation travel times. The minimization criteria is a quadratic function of the state variables, control variables, and vehicle headways. They applied "linear regulator theory" to derive the optimal control policy, which is a linear function of the state variables. Campion et al (1985) and Van Breusegem & Campion (1989) proposed a different formulation for the same problem, which was claimed to have advantages over that of Sasama & Ohkawa (1983) in terms of real-time efficiency and information usage. The difference was in the formulation of the state vectors.
Unlike the research on holding in bus systems, the above research (implicitly) assumed computerized systems for real-time operations control, and the availability of vehicle position information. They also explicitly formulated vehicle movement equations in terms of both dwell time and interstation travel time. In these regards, their research is relevant to this effort. However, in the application of their models, the effectiveness of speed control is limited and cannot be used to replace station control strategies. First, speed change is restrained by many factors in high frequency urban rail transit systems, such as drivers' characteristics, vehicle types and ages, station spacing, etc. Trains may be able to slow down more easily than to speed up. Second, although slowing down or even interstation stopping can be used like holding to remedy headway irregularity, it is less desirable than holding. For one thing, holding at a station allows passenger boardings and alightings. For another, interstation stopping increases passenger anxiety and travel time, while holding at a station would not affect those passengers who alight at that station, and this stopping appears more natural. It is also important to note that their solutions do not apply to the holding problem. First the control variable in their formulation can take either negative or positive values, while dwell time change must be non-negative. Second, dwell time changes are allowed only at a few given control points.

In general, the linear feedback control approach involves computation of a "gain matrix", which is computationally burdensome in real time. To overcome this problem, these authors have either approximate control policies or subproblem optimization. That is, when considering the control of a vehicle at a station, only the state variables of one or two of its neighboring vehicles at neighboring stations are taken into consideration. In this case, the problem has been reduced to a simple small deterministic optimization problem.

1.2.3 Station Skipping

Strategies such as deadheading, expressing, and short-turning all involve deliberate station skipping in a transit route, and they are frequently used in transit systems in the U.S (and elsewhere), maybe more so than holding. A rough estimate on the MBTA
Green Line, Boston, showed that, one of these strategies were used on average every four or five minutes. While these strategies can be quite effective, due to the more dramatic change they can produce to vehicle headways than holding at a single station, they can also have larger negative effects when used inappropriately. Studies by Macchi (1990) on expressing and Deckoff (1990) on short-turning in the MBTA Green Line showed that approximately 1/4-1/3 of these (manual) decisions in the sample data set analyzed were "bad", or non-beneficial. Because all decisions on the Green Line were made by field inspectors on the basis of their experience and personal judgment, the bad decisions were partly due to lack of real-time information, and partly due to limits in human judgment in complex situations. This clearly shows the importance of research in this area for better decision making. However, such research is surprisingly scarce, which is a sharp example of the imbalance between theory and practice.

Before reviewing the few studies in this area, we need to distinguish real-time station skipping strategies from their pre-planned counterparts. There are a few papers on the pre-planned deadheading problem (Furth, 1985) and expressing problem (Araya et al., 1983). In these studies the deadheading and expressing vehicle trips and the skipped stations are planned in advance and are of a recurrent nature; they are known as "operations planning" strategies rather than real-time "operations control" strategies. These problems differ in their nature, objectives, decision variables, and input parameters. For example, in the study on deadheading alternate buses on a route by Furth (1985), the objective is to minimize the fleet size to meet a regular alternating deadheading schedule. The research by Araya et al (1983) was on the problem of order changes between local and express vehicles, where the "express vehicles" have fixed routes and schedules. In comparison, the real-time station skipping problems are to decide which vehicle should be controlled and which stations should be skipped depending on real-time conditions, in a way so as to minimize cost to passengers. Hence in this dissertation we do not consider the literature on pre-planned station skipping strategies.
Of the few studies on real-time station skipping strategies, a number of MIT students have studied the MBTA Green Line case. Macchi (1990) studied expressing; Deckoff (1990) studied short-turning; Fellows (1990) examined the uses of AVI information for headway control including both station skipping and holding; and Soeldner (1992) compared expressing and short-turning. All of these studies are empirical and use AVI data. Although their spreadsheet models did not consider dwell time effects (which is important when station spacing is dense and passenger demand varies greatly across stations), and were case specific, nonetheless their efforts provide a good starting point for better understanding these strategies. With the exception of Koffman (1978), who simulated a strategy that skips loading (but not unloading), Macchi (1990) and Soeldner (1992) are the only two prior studies we are aware of in the U.S on the real-time expressing problem. Also, they have provided good empirical data such as Green Line O-D matrix, which will be used for computational tests in this dissertation.

At the same time that this research was been conducted, a related effort was underway by Li (draft dated June 1994) in Canada. Li's problem setting was a highly undercapacity bus route (in Shanghai, China) with extremely short scheduled headways (1-2 minutes). Given this setting Li minimizes the number of passengers left behind by buses as a linear approximation to the nonlinear waiting time objective function. Li deals with the control decision at a single dispatching terminal and the solution relies on a given (fixed and small) set of possible control actions. Li developed two models: one assumes average demand at stations and average interstation running time for vehicles, while the other assumes constant dwell time and fixed headway at each station. These are both significant simplifications of reality and together with other assumptions, such as the requirement that no passengers arriving during dwell time can board, seem overly restrictive. Li uses tabu search to find feasible solutions for both models, which are not necessarily optimal.

While related, the results of Li's research have little applicability to the work presented here. Substituting the number of passengers left behind for waiting time is not appropriate for the research we are conducting. We are not modeling strategies that
depend on a small set of predetermined control stations or segments, and our interest is in more than working heuristics for a specific case. This research tries to provide, by both analytical and empirical means, a deeper understanding of the nature and properties of the transit systems and the control strategies, including why, how, and under what conditions which strategy is most beneficial. Solution algorithms are then developed based on the analysis.

1.3 Research Scope and Methodology

In this research we model, analyze, and evaluate the real-time control strategies deadheading, expressng and holding, both singly and in combination. Since these problems are complicated and little prior research exists on the station skipping strategies, this dissertation is one of the first attempts to model them, and we will focus on the most important aspects. In this section we define the bounds of this research.

1.3.1 Objective of Control

In this dissertation we consider the objective of control to be the minimization of total passenger waiting times in a high frequency urban transit service (scheduled headway < 10 minutes), where passengers arrive randomly at stations and the scheduled vehicle headway is constant during a period (e.g., rush hours). This objective function is most commonly used in transit research for evaluating transit service quality and reliability. It can be justified by two arguments. First, research on passenger demand has shown that transit users are more sensitive to waiting time than to line-haul or in-vehicle travel times, which suggests larger weights on waiting time than riding time in their utility functions (see evidence in, for example, Kemp, 1973; Ben-Akiva and Lerman, 1985). Second, although composite passenger travel time (waiting time plus riding time) can also be used, the riding time component of such an objective function is likely to have a much smaller contribution.
There may be a concern that, because expressing can reduce some passengers riding time by saving dwell times, and holding can increase passengers riding time by increasing dwell times, to exclude riding time in the objective function may underestimate expressing benefits and overestimate holding benefits. This will not be a big problem when the number of passengers on board when the vehicle is expressed or held is a small portion of all passengers in the route. In addition, dwell time change at a control station is likely to be much smaller than headway changes along the entire route beyond the control point. Such a difference is even sharper when the number of control stations is small. However, there may be situations where the riding time saving is significant. This issue can be investigated through computational testing.

For these reasons, we do not feel it is necessary to increase the modeling complexity to include passenger riding times in the objective function. However, we will examine the potential impact of including riding time in the objective function through empirical analysis.

Because passenger waiting times are closely related to vehicle headways, to minimize the sum of squared headway variation is also sometimes used as an objective function, primarily in rail transit research. However, the sum of squared headways is not the same as total passenger waiting time when passenger arrival rates vary across stations. Furthermore, when station skipping strategies are involved, it is not clear that the most even headways will result in the least passenger waiting times. Thus in this dissertation we do not use headway variation minimization to replace waiting time minimization.

The formulation of the objective function that minimizes total passenger waiting time will be given in Chapter 2.

1.3.2 Control Environment and Process

In U.S. transit systems, each work day each vehicle operator is assigned a vehicle and schedule of departure times at one or more terminals during his work period, called "a piece of work". The operator typically does not know (and may not care about) the
position of other vehicles, but tries to follow his own schedule. Hence operators generally do not make control decision by themselves except holding to meet the schedule at time points. A transit agency typically has some form of control system in place for real-time operations, either centralized, decentralized, or partially centralized, which can be either computerized, manual, or a mixture. Control decisions are either made at stations by field inspectors, or made by personnel in the control center, and then the instructions are relayed to the vehicle operators. Because the control strategies considered in this dissertation all need to be implemented at a station, a control decision on a vehicle should be made about the time the vehicle enters the control station. A computerized control system should be able to generate such decisions quickly. In the case of a manual control system, the field inspector makes a decision based on available vehicle and passenger information. In this dissertation we develop both algorithms for computerized control systems and guidelines for manual control systems, regardless of whether they are centralized or decentralized.

**Simultaneous vs. Sequential Control Process**

When a centralized and/or computerized control system is concerned, however, one issue arises about the control process: should the control process deal with all vehicles in the system simultaneously, or in a particular sequence? To be more specific, consider $M$ vehicles operating in the transit network at time $T$. Among them $m$ are to arrive at various stations in the next $t$ seconds. Suppose we want to give each operator of the $m$ vehicles control instruction in a time range of $0-\tau$ seconds, before each of them enters its next station. The computer procedure needs at most $\tau'$ seconds to run, hence we may start to run the computer procedure $\tau+\tau'$ seconds before any of the $m$ vehicles enters a station. In a simultaneous control process, all $M$ vehicles are considered in the computer procedure at the same time, so that the interaction between control of the $m$ vehicles are taken into consideration in order to generate the optimal control policy for all of them. In a sequential control process, however, only one vehicle is considered at a time in order of arrival time at a station, and, in the circumstance of two or more vehicles entering
different stations at the same time, in a FIFO order of their positions in the route. The impact of the control decision is then evaluated based on the current status of other vehicles.

There is no question that the complexity of the simultaneous control process is much higher and therefore is much more costly in terms of modeling and algorithm development effort and computation time. The question is, how much difference would it make in terms of system performance?

Theoretically, any vehicle at any station in the transit system may be controlled. While deadheading can start from two different terminals in a loop network, expressing may start from any station, and there may be more than one express segment for a vehicle in an optimal control sequence. Also, any vehicle may be held at any station. In practice, however, the number of control stations (segments) is limited by many factors, but principally because it is expensive and unnecessary to make every station in the network a control station. Previous research on holding has shown that there exist a few effective points for control in a network (see for example Abkowitz et. al., 1986). Station skipping strategies, on the other hand, are unlikely to be used more than once per vehicle trip. It is also unlikely to be desirable to skip the same station by more than one consecutive vehicle, for waiting passengers hate to see vehicles passing by without stopping. In addition, in a decentralized control system, a control action can be taken only where a field inspector is stationed.

When the number of control stations (segments) is small, there is a very small probability that two or more vehicles will arrival at different control stations at approximately the same time. When such circumstances do occur, if the control stations are far apart from each other, the impact of one control action on the other is likely to be small. Furthermore, due to the dynamics of the system, simultaneous decisions based on computed interactions between vehicles that are far apart may be a risk, or not necessarily better than sequential decisions.
For these reasons, this dissertation adopts a sequential decision-making approach. More specifically, we consider a rolling horizon scheme when modeling the control strategies. That is, we control one vehicle trip at a time, and consider the impact of controlling that vehicle on a small set of following vehicle trips. The size of such an "impact set", also called rolling horizon size, will be discussed in detail later. The resulting optimal control policy will then be applied to the vehicle considered. The process is rolled forward and the optimization is repeated for each vehicle trip in the system.

Data Requirements

We assume the monitoring and data processing part of the control system will provide the following information as input to control decisions:

- Departure times of the preceding vehicle at all stations
- The most recent departure times of vehicles under consideration at the dispatching terminal (or the most recently passed station)
- Mean vehicle interstation travel times between consecutive stations, and mean time for deceleration and acceleration when approaching and departing a station
- Expected passenger arrival rate and passenger alighting ratio at every station
- Parameters of the dwell time function
- Scheduled headway, minimal safe headway, upper and lower bounds of layover time at terminal(s), scheduled next departure time for each vehicle at terminal(s)

The first two types of data can be obtained from any AVI/AVL systems directly. The next three types of data can be estimated from historical data and surveys, the accuracy of which can be greatly enhanced by real-time APC systems should they become available. The last set of data is from the operating plan. The specific meaning and the uses of the above data will be elaborated upon in the following chapters.

1.3.3 Elements of Concern in Modeling

Passenger waiting times depend on vehicle headways at stations. At any station along the transit route, a vehicle's headway\(^2\) is determined by its dispatching headway and the trip times of its preceding vehicle and itself up to that station. Vehicle trip times are in turn

---

\(^2\) In this dissertation, whenever there is no specification, "vehicle headway" means "preceding headway of the vehicle".
composed of dwell times at stations and interstation running times. In general, vehicle
dwell times depend on passenger demand (boardings and alightings), while interstation
runtimes do not. This property has a large impact on control strategies, especially station
skipping strategies as we will see more clearly as we proceed.

In reality, vehicle dispatching headways, dwell times, and interstation running times are
all random variables, and this randomness contributes to the irregular headways. While
dispatching, running and dwell times interact with each other, there are important
independent causes for randomness in each variable. Dispatcher behavior (e.g., the way
dispatching decisions are made) and vehicle operator behavior (such as attendance,
schedule adherence performance, etc), as well as labor and operating rules (e.g., layover
times, break times, work pieces, operator relief time and location) are the primary
independent causes for dispatching headway irregularity. Variable passenger demand is
the primary independent cause for dwell time irregularity. Finally, vehicle characteristics,
route conditions, and operator driving habits may be the primary independent causes for
interstation travel time irregularity.

From the view point of modeling real-time control strategies, the randomness in vehicle
dispatching headways (in the absence of control) is of the least concern. While the
dispatching headways will be important input to the control models, they can be treated
as deterministic. This is because when considering a control at the dispatching terminal,
only the availability, not the actual departure times, of a small set of vehicles needs to be
known or forecast, and this information is the easiest to obtain. Furthermore, since only
the first vehicle in the set is to be dispatched, only the information for that vehicle needs
to be accurate. This is not a problem at all because the control decision can be made
when the vehicle is known to be available. On the other hand, when considering controls
at other stations, the relevant dispatching headways are likely already realized and
therefore effectively deterministic.

This leaves the randomness in vehicle dwell times and interstation running times as
elements of primary concern. These two elements need further discussion, because the
accuracy of the prediction of vehicle travel times beyond the control point directly impacts the control. While we do not have direct information on the magnitude of the variations of dwell time and interstation running times separately, we present some relevant observations from historical AVI data.

**The Data**

We obtained one-week of vehicle location data by time of day from the MBTA Green Line AVI system. The MBTA Green Line is a light rail transit system, which has network and operational characteristics similar to the type of system of interest here (see Chapter 2), except that it has four branches instead of a single loop. Only the data from its B line will be used, as we are studying a loop network and this branch forms a complete loop network itself. There are a total of 52 stations in the two directions of the B line, with 13 AVI detectors located along the route. The inbound (IB) direction goes from station 1 to station 26, and the outbound (OB) direction is from station 27 to 52. Each AVI detector is typically located close to a station, some near the entrances and some near the exits. Appendix B shows the correspondence between the AVI detectors and the stations. For notational convenience, hereafter we will use the station index to indicate the AVI detector locations. It can be seen from Appendix B that except for the two detectors placed at each end of the route (stations 1 and 52), all other detectors are concentrated in the middle of the route between stations 19 and 35. The AVI data for morning peak hours (about 6:30am-10:30am) from Monday to Friday in the week of October 16-20, 1993, are used in this dissertation. The scheduled headway during this period was 5 minutes. Fig. 1.1 depicts a subset of this AVI data, with linear interpolation between data points for each train.

It should be noted that the AVI data are records of realized vehicle movements, some of which may result from actual control actions such as station skipping and holding. Most of the control actions are difficult to detect from the AVI data, except short-turning. The point at which a vehicle can typically be short-turned is from station 24 (in Direction 1 or IB) to station 29 (in Direction 2 or OB). Since a short-turned vehicle will not pass
stations 25-28, it can be easily distinguished from other vehicles in the AVI records. In Fig. 1.1, there are 10 short-turned trains among the 36 shown. Although the missing data points in short-turned vehicle trips are not obvious in Fig. 1.1 because of the linear interpolation between station 25 and 29, most of the short-turned vehicles can still be observed because they passed one or more vehicles between stations 24 and 30.

![AVI Detector location](image)

**Fig. 1.1 AVI Records: Monday 10/16/93**

Green Line demand data is also available, as compiled by Macchi (1990) from a 1985 survey\(^3\). The compiled demand data contains both passenger arrival rates and passenger alighting ratios (as a proportion of vehicle arrival load) at each Green Line station by time period of the day, although only the morning peak demand data of the B line is used in this research. In Appendix B we illustrate the demand profile by plotting the mean passenger arrivals and alightings per minute. Again, data points are interpolated.

---

\(^3\) The data collection was performed by the Central Transportation Planning Staff (CTPS) under contract with MBTA.
Appendix B clearly shows that in the morning peak, passenger activity is highest at stations 25, 26, and 28, namely Park Street IB, Government Center IB, and Park Street OB, all of which are transfer stations (to the Red Line and the Blue Line). Station 26 is also the terminal in the inbound direction, where the number of alightings is the highest. Both passenger boardings and alightings at stations 1 to 18 and 34 to 52 are quite low. In fact, the average passenger arrival rate at each of those stations is less than 2 passengers per minute.

Although this survey data is somewhat dated, which means the expected value of passenger arrivals and alightings at each station may have changed (which we do not have any information about), the demand profile given in Appendix B is still largely true based on recent observations of the Green Line, in the sense that the shape of the demand curve (including both passenger boardings and alightings) along the route is high in the middle and low at both ends. More specifically, Park Street in both directions and Government Center IB (stations 25, 26, and 28) always have much higher passenger activity than other stations in the weekday morning peak period, because they are transfer stations with the Red Line and the Blue Line.

**Preliminary Data Analysis**

The AVI data contains only what we call "composite travel time" of vehicles in that the dwell time and interstation travel time are not reported separately. It is, however, possible to recover useful information about the contributions of vehicle dwell times and interstation running times from the AVI data.

Let us call any segment between two adjacent AVI detectors an "AVI segment". We first compute the "composite travel time" of each AVI segment. Because distances vary across AVI segments, to make the travel time compatible between segments we normalize the composite travel times in each AVI segment by dividing them by the corresponding segment length (distance between the two AVI detectors). In this way we obtain the "unit composite travel time" (or simply "unit travel time") for each segment.
Table 1.1 lists the mean and standard deviation (across vehicle trips) of the unit travel times by weekday for the AVI data.

A few observations on the unit travel times can be drawn from Table 1.1:

1. In each AVI segment, the standard deviation across vehicles (shown in the column labeled StDev) is much smaller than the mean. In most segments the coefficient of variation is around 0.1-0.2.

2. The same pattern repeats almost exactly each day of the week. Fig. 1.2 plots, at the ending point of each AVI segment, all the mean unit travel times (the data points are connected for easier interpretation of the pattern).

<table>
<thead>
<tr>
<th>Week Day</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Week Total</th>
</tr>
</thead>
<tbody>
<tr>
<td># Vehicle Trips</td>
<td>36</td>
<td>35</td>
<td>38</td>
<td>38</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>AVI segment</td>
<td>mean</td>
<td>StDev</td>
<td>mean</td>
<td>StDev</td>
<td>mean</td>
<td>StDev</td>
</tr>
<tr>
<td>Stations 1-19</td>
<td>0.11</td>
<td>0.01</td>
<td>0.12</td>
<td>0.01</td>
<td>0.12</td>
<td>0.01</td>
</tr>
<tr>
<td>Stations 19-20</td>
<td>0.08</td>
<td>0.01</td>
<td>0.09</td>
<td>0.03</td>
<td>0.08</td>
<td>0.01</td>
</tr>
<tr>
<td>Stations 20-22</td>
<td>0.09</td>
<td>0.02</td>
<td>0.09</td>
<td>0.02</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>Stations 22-24</td>
<td>0.10</td>
<td>0.02</td>
<td>0.10</td>
<td>0.02</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>Stations 24-26</td>
<td>0.12</td>
<td>0.03</td>
<td>0.13</td>
<td>0.02</td>
<td>0.13</td>
<td>0.02</td>
</tr>
<tr>
<td>Stations 26-27</td>
<td>0.18</td>
<td>0.08</td>
<td>0.18</td>
<td>0.07</td>
<td>0.19</td>
<td>0.07</td>
</tr>
<tr>
<td>Stations 27-28</td>
<td>0.13</td>
<td>0.03</td>
<td>0.14</td>
<td>0.03</td>
<td>0.14</td>
<td>0.04</td>
</tr>
<tr>
<td>Stations 28-29</td>
<td>0.18</td>
<td>0.05</td>
<td>0.20</td>
<td>0.11</td>
<td>0.20</td>
<td>0.06</td>
</tr>
<tr>
<td>Stations 29-31</td>
<td>0.10</td>
<td>0.01</td>
<td>0.12</td>
<td>0.05</td>
<td>0.10</td>
<td>0.01</td>
</tr>
<tr>
<td>Stations 31-33</td>
<td>0.06</td>
<td>0.01</td>
<td>0.07</td>
<td>0.01</td>
<td>0.06</td>
<td>0.01</td>
</tr>
<tr>
<td>Stations 33-34</td>
<td>0.11</td>
<td>0.02</td>
<td>0.11</td>
<td>0.02</td>
<td>0.11</td>
<td>0.02</td>
</tr>
<tr>
<td>Stations 34-52</td>
<td>0.11</td>
<td>0.01</td>
<td>0.11</td>
<td>0.01</td>
<td>0.11</td>
<td>0.01</td>
</tr>
<tr>
<td>Round Trip</td>
<td>0.11</td>
<td>0.01</td>
<td>0.11</td>
<td>0.01</td>
<td>0.11</td>
<td>0.01</td>
</tr>
</tbody>
</table>

Table 1.1 Unit Composite Travel Times Computed from Green Line AVI Data

3. The differences in unit travel times between AVI segments are large, as is clearly shown in Fig 1.2.
Fig. 1.2. Mean Unit Travel Time in the 5 Weekday Morning Peaks

Obviously, the AVI segments from station 26 to 27 and from station 28 to 29 are the two peaks in unit travel time, and this pattern is repeated every weekday. Since there is an AVI detector at every station from stations 26 to 29, and the distance between the detectors is small (0.12-0.24 miles), the composite travel time largely reflects dwell time at these stations as well as delay caused by interlockings (switches). Thus the peak unit travel times for AVI segments 26-27 and 28-29 reflect long dwell times at stations 26 and 28 respectively. The long dwell time at station 28 (Park Street OB) is obviously due to its high passenger boardings, while the long travel time in segment 26-27 is due to high alightings at station 26 (Government Center IB), and possibly also delay at this turnaround point caused by interlockings. The unit travel time in AVI segment 24-26 does not seem to fully reflect the long dwell time at station 25 (Park Street IB), because this segment spans the distance between three stations, and the long dwell time impact at station 25 has been averaged over this larger distance (0.57 miles) in the unit travel time. This is a deficiency of this simple approach. Even so, the unit travel time in this segment is still higher than in most AVI segments. The correspondence between demand volume and dwell time length is clear.
The above shows unit trip time variation across AVI segments. To see the travel time variation across vehicles, in Fig. 1.3 we plot the unit travel times of all vehicle trips during the Monday period in the same way as in Fig. 1.2 (data for the other four days look similar). For direct comparison, we also included the passenger demand profile in the figure. The similarity between the demand pattern and the unit travel time pattern is striking!

From Fig. 1.3, we can see that the variation of unit travel time across vehicles in the same segment is generally lower than the variation across segments. Furthermore, the variation across vehicles is high in high demand segments and low in low demand segments. This again indicates that the dwell time (and possibly also interstation stopping time caused by interlocking delays) has a stronger impact on both mean and variation of the composite travel time than the impact of interstation running times.

The strong impact of vehicle dwell time on trip time variation (hence headway variation) is thus empirically evident in the Green Line data. The implication is that, failure to consider dwell time explicitly in control models is unlikely to result in realistic and useful control policies. Hence dwell time will be a primary element of concern in our analysis.

On the other hand, since interstation running times are likely to have small variance across vehicles around the mean, it seems reasonable to assume a single free interstation running time for all vehicle trips for each station pair. In fact, the variance of interstation travel time across vehicles may be more controllable in the future, through various means such as vehicle and operation quality control, and technological innovation such as automatic speed control systems (especially for rail transit). However, variances caused by behavioral factors, for example dwell time variation across stations resulting from variable passenger demand, are less likely to be controllable. This again indicates the importance of dwell time in the control models.
Fig. 1.3 Comparison between Unit Travel Time and Passenger Demand
1.3.4 Deterministic Models and Stochastic Tests

While it is clear that real-time control problems are stochastic in nature, it does not follow that a stochastic model will provide the best results. In general, a stochastic model is much more costly in terms of modeling and algorithm development efforts; less efficient, especially with many types of randomness involved; and less tractable. For these reasons it may be possible to develop deterministic models which capture most of the important elements of the problem and require fewer unrealistic simplifying assumptions. One result of all of this is that the performance of a stochastic model may not be significantly better than the performance of a deterministic model.

The models in this dissertation will be developed using a deterministic framework. One can think of a deterministic framework as the limiting case of a stochastic framework in which the variance of the stochastic elements becomes very small. Given the real time nature of available information considered in this study many of the most important stochastic elements are both realized and measured. Thus, it is not unreasonable to expect the remaining variances to be small and to treat the problem as being deterministic. Nonetheless, it is important to understand what the implication of these deterministic assumptions are in a realistic stochastic environment. In this dissertation we analyze the impact of stochastic factors via sensitivity tests based on the following considerations.

First, as discussed in the previous subsection, the two most important stochastic factors are vehicle dwell time and interstation running time. In the deterministic models, we will use expected passenger arrival rates which are constant over time but vary across stations. Also, we will assume a single (free) interstation travel time for all vehicles in the same segment (the "actual" interstation travel time of a vehicle depends on whether it is blocked by its preceding vehicle). These expected values are often either readily available or can be easily estimated, and such a model should be sufficient to capture the most important impacts from dwell time variation across stations. What is not reflected in such a model is that passenger arrival rates at each station may change randomly with
time, and that free interstation travel time may vary across vehicles. The models
developed can, however, be tested by adding randomness to passenger arrival rates and
interstation travel times and seeing the impacts on performance.

Second, the dynamic modeling approach adopts a rolling horizon scheme, which places
more weight on the vehicle currently considered for control than its following vehicles.
Consequently, the dwell time variation along the route for the current vehicle is more
important than the dwell time variation across vehicles at the same station. Furthermore,
since dwell time variation along the route depends on the passenger demand profile,
updates of passenger demand information may reduce errors in dwell time prediction.
This is where modern information technologies such as APC systems can certainly play a
role.

Finally, we want to emphasize the importance of algorithm efficiency for the real-time
control models. The total computation time should be on the order of seconds. If a
stochastic model can provide somewhat better solution with much longer time than a
deterministic model, it may not be preferable.

Based on these considerations, in this dissertation we take a deterministic modeling
approach, while testing the impact of randomness via sensitivity analysis. A detailed
sensitivity test plan will be given in a later chapter.

1.3.5 Research Philosophy and System Types

Real-time control problems are very complicated in general, involving many interacting
factors. These are theoretical problems which challenge OR techniques, and they are
practical problems in which reality should not always be sacrificed for the sake of tractability. This makes the modeling and solution process interesting but difficult. On
the one hand, as for most practical problems, if the models are general and realistic, they
are not well enough behaved for elegant analytical results. On the other hand, while
simplification may increase tractability, it may also decrease the usefulness of the models
in addressing the problems they are designed for. While we will certainly not be content
with an unrealistically simplified although well behaved model, in order to solve realistic problems it is useful to look at simpler problems first to gain insights that help to capture the essence of the more general problems.

Following this philosophy, we take a general-simplified-general approach. In particular, we study each control strategy in two different types of transit systems: the one of ultimate interest is called System G, which stands for a "general" type of transit system in the U.S.; the other is called System F, which is a special case of System G, where F stands for "fixed" or constant parameters (see section 2.7). We first develop mathematical program formulations for each strategy in the general system G, and then simplify them based on the characteristics of System F so that they can be solved exactly. Analytical results and insights gained from the System F models are then utilized to help develop solution algorithms for the models in System G where exact solution may not be possible.

An important feature of the research methodology in this dissertation is, where analytical methods work, we will not stop just at algorithm development, but try to obtain a deeper understanding based on the nature of the systems and the problems. On the other hand, where analytical methods can not give a complete answer, we will base the algorithm development also on understanding the nature of the systems and problems. Hence empirical data analysis and computational testing are important parts of this dissertation.

1.4 Input Data Preparation

In this dissertation we intend to use realistic data for all computational experiments and testing. For this purpose we collected data from the MBTA Green Line B line. Along with the 5 weekdays of AVI data and demand data mentioned earlier, we also obtained data for service schedules and route configuration (number of stations and distance between stations) of the B line.
As mentioned earlier, the B line contains 52 stations in two directions, and 13 AVI detectors along the route. For notational convenience, hereafter we will call the IB direction "Direction 1", and the OB direction "Direction 2". There are layover times at station 1, but not at station 27. The AVI data for each of the 5 morning peak periods contains 31-36 round-trips, with an average of 2 round-trips for each vehicle. Because one main objective in this research is to test control policies under as realistic as possible headway conditions, and as we will see later, the no-control dispatching headway pattern has great impact on control policies, what we want from the AVI data are the realistic no-control dispatching headways in each direction (i.e., at station 1 and station 27). It is particularly important to separate the directions in order to evaluate impacts from different demand patterns. Thus the plan was to extract from the AVI records the dispatching times of all vehicle trips in each direction, and form two input data sets for each morning (i.e., one for each direction per morning).

Since there are AVI detectors at both stations 1 and 27, at first this looked like an easy enough task. However, it turns out that the original data have much greater complexity than we thought. A serious problem was that for each morning there are quite a number of vehicle trips short-turned, thus station 27 was skipped by the vehicles. For example, Fig. 1.1 showed that 10 vehicle trips (out of 36) were short-turned in Monday morning. This makes the dispatching headways in the OB direction inappropriate as realistic no-control headways. Thus, we only obtained 5 input data sets (for direction 1) directly form the AVI data, and we need to "recreate" the no-control headways for direction 2 at station 27, so as to continue the trips from direction 1.

To do this, we used the no-control vehicle movement equations (2.1) to (2.6) defined in Chapter 2, together with the realistic vehicle departure time data at station 1, the demand data and the route configuration data, to compute the hypothetical trajectory of each vehicle trip. For this simulation, the dwell time function (see equation (2.5)) parameters were taken from Lin and Wilson (1992), which was empirically estimated for the Green Line.
This simulation results in the other 5 input data sets (for direction 2), one for each weekday morning. Thus we obtained 10 input data sets in total, each containing departure times at either station 1 or 27. The two input data sets for the same morning contain exactly the same number of vehicle trips. We index each input data set by weekday and direction. For example, input data set "m1" is for all vehicle trips in Direction 1 on Monday morning, "m2" is for all vehicle trips in Direction 2 on Monday morning, and so on. Later we will also refer to these input data sets as "the Green Line data sets". Appendix C shows dispatching headway patterns from these data sets.

1.5 Thesis Organization

This dissertation is organized as follows. In Chapter 2, we formally define conditions under which the control strategies are studied, and introduce basic notation and assumptions. In Chapter 3, we study free system status without control by system type. We prove and demonstrate several properties of the systems which are useful for both better understanding problem causality and algorithm development in later Chapters. Chapters 4 to 6 model, analyze, and compare three individual control strategies - deadheading, expressing and holding by system type. Chapter 7 develops models and algorithms for combined control strategies, and evaluate their effectiveness. Finally, Chapter 8 concludes with a summary of the results and suggestions for future research.
CHAPTER 2

PROBLEM SETTING AND NOTATION

In this chapter we formally define general conditions under which the control strategies are studied, and introduce basic notation and assumptions. More specific assumptions and notation regarding individual strategies will be given later in corresponding chapters. The reader can also find a complete summary of notation in Appendix 1.

2.1 The Transit Network

We consider a transit system that consists of a one-way loop network, as shown in Fig. 2.1. Such a network is representative of, for example, a single rail transit line, or a simple bus or trolley bus line. Letting $R_*$ and $R_{++}$ ($Z_+$ and $Z_{++}$) denote the sets of non-negative and positive reals (integers) respectively, $N \in Z_{++}$ denotes the number of stations in total. All vehicles start their first trip each day from a dispatching terminal indexed as station 1. Starting from station 1, the downstream stations are indexed as $1, 2, \ldots, N/2$, with the ending terminal in this direction being $N/2 \in Z_{++}$. In the reverse direction, we have stations $N/2+1, N/2+2, \ldots, N$. Since stations in both directions have a one-to-one correspondence, $N$ is always even and $N/2$ is always an integer. Except for terminal $N$, a station can accommodate only one vehicle at a time and no passing can occur at any point (other than $N$) in the network. Individual stations will be denoted using index $k$. For convenience, we sometimes use index $k>N$ and $k=0$. In such situations we adopt the
convention that if \( k > N \), then \( k \) represents station \( k \mod N \). Also, if \( k = 0 \) then \( k \) represents station \( N \).

Conceptually, the network in Fig. 2.1 can be seen as having two directions. We denote the set of stations from 1 to \( N/2 \) as direction 1, and the set of stations from \( N/2 + 1 \) to \( N \) as direction 2. When controlling a vehicle we consider one direction at a time, and such a direction is called the control direction of that vehicle. Given these two directions, we need to distinguish between different sets of stations. First, \( K \) denotes the ordered set of all stations in the network. That is, \( K = \langle 1, 2, \ldots, N/2, \ldots, N \rangle \). Obviously \( |K| = N \). Similarly, \( K_c \) denotes the ordered set of all stations in a control direction; that is, \( K_c = \langle k_0, k_1, \ldots, k_t \rangle \subseteq K \), where \( k_0 \) denotes the first station in the control direction and can only be either station 1 or \( N/2 + 1 \); \( k_t = k_0 + N/2 - 1 \), which is the ending terminal in the control direction. We also denote the direction opposite to the control direction as \( \bar{K}_c \). Obviously \( |K_c| = |\bar{K}_c| = N/2 \). In particular, the index \( c \) takes one of the two values: 1 and 2, which correspond to the two directions respectively. Thus, \( K_1 = \langle 1, 2, \ldots, N/2 \rangle \), and \( K_2 = \langle N/2 + 1, \ldots, N \rangle \).

![Fig. 2.1 The Transit Network](image)

**Vehicles and Operating Plans**

We consider a high frequency transit service with scheduled headways of less than ten minutes. There are a total of \( M \in \mathbb{Z}_{++} \) vehicles in operation in the network, each of which will be denoted by index \( i \) or \( j \). A control problem considers a group of \( m \) (\( 1 \leq m \leq M \)) vehicle trips denoted as \( I_m = \langle i, i+1, \ldots, i+m \rangle \), where vehicle \( i \) is to be controlled.
immediately applying the solution, and vehicle trip \( i+m \), as well as \( i-1 \), serve as boundaries of the control problem. A real-time control problem is *dynamic* in nature. At the time to make control decisions for the group \( I_m \), all the vehicles \( i \in I_m \) should either have arrived at or will arrive within a short time interval \( t \) at station \( k_{o-1} \) (i.e., station \( N \) or \( N/2 \)), so that their departure times at station \( k_{o-1} \) or arrival times at \( k_o \) are either known or can be fairly accurately determined before the control decision is made.

There may be a layover time for each vehicle/crew at terminal \( N \) and/or \( N/2 \), and such a layover time is of variable length depending on work rules, local practice and actual vehicle arrival times. The next layover terminal of the control direction \( c \) is denoted as \( k_c \).

If both terminals \( N \) and \( N/2 \) require layover times, then \( k_1 = N/2 \) and \( k_2 = N \). Otherwise, if only station \( N \) requires layover time, then \( k_c = N \). At the dispatching terminal next to \( k_c \), a scheduled departure time of each vehicle trip is assigned to the vehicle operator based on the scheduled headway, trip time and layover time. This scheduled departure time is denoted as \( t_{i,c} \) for each vehicle trip \( i \). The subscript \( c \) is direction index defined above, which indicates the next layover terminal for the control direction. That is, if both terminals in the network require layover times, then \( t_{i,1} \) is the scheduled departure time at \( N/2+1 \) and \( t_{i,2} \) is the scheduled departure time at 1 for vehicle trip \( i \). If there is only one layover terminal in the network, the subscript \( c \) is omitted. Each vehicle, controlled or not, must stop at both stations \( N/2 \) and \( N \).

Throughout this dissertation, \( h_c \in R_{++} \) denotes the *scheduled headway*, and \( h_{ik} \in R_{++} \) denotes the actual *departure headway* between vehicles \( i \) and \( i-1 \) at station \( k \). Note that \( h_{ik} \) is always positive because of the vehicle indexing conventions used. We also let \( d_{ik} \) denote the *departure time* of vehicle \( i \) from station \( k \), \( R_{ik} \) the *free interstation travel time* of vehicle \( i \) from station \( k-1 \) to station \( k \), and \( a_{ik} \) the *arrival time* for vehicle \( i \) at station \( k \). The *free trip time* of vehicle \( i \) between its departures from any two stations \( k \) and \( k' > k \) is then denoted as \( T_{kk'} \).

Throughout this dissertation, we assume the following if not otherwise noted:
**A1.** Vehicles can not pass each other during a round trip from station 1 to N. Therefore, vehicles are always operated in FIFO order after departing station N. A minimal safe headway, $h_0 > 0$, must be observed at all times such that we always have $h_{i,k} \geq h_0$.

**A2.** Each vehicle runs at a fixed speed, except when approaching (departing) a station, where extra time is required for acceleration (deceleration). If a vehicle $i$ skips any station, it saves a constant time $\delta_i$ in acceleration (deceleration) when approaching (departing) that station.

**A3.** Vehicle capacity constraint is not binding.

Assumption A1 generally holds for most urban transit systems, particularly rail and trolley bus systems. A2 is a reasonable approximation to reality and is introduced mainly for the sake of simplicity. While A3 is often true in the US transit systems and is also introduced for simplicity, consequences from violation of this condition will be addressed later.

It follows from assumptions A1 and A2, a vehicle can not enter a station until $h_0$ after the preceding vehicle’s departure. Furthermore, in the no-control case, the free running time of vehicle $i$ between stations $k-1$ and $k$ is the minimal running time plus the extra time in acceleration and deceleration, that is, $R_{i,k} + 2\delta_i$. Hence,

\[ a_{i,k} = \max(d_{i-1,k} + h_0, d_{i,k-1} + R_{i,k} + 2\delta_i) \]

This says that the arrival time of a vehicle at a station can depend on the preceding vehicle’s movement as well as its own unconstrained movement.

### 2.2 Demand Characteristics

The *passenger arrival rate* at station $k$ is denoted by $r_k$. We assume:

**A4.** Passengers arrive randomly at a constant rate $r_k$ at station $k$. 

44
Thus, the expected number of passengers arriving at station $k$ during a period of length $h_{i,k}$ is given by $r_k h_{i,k}$. This assumption is consistent with observed random passenger arrivals for short headway (< 10 minutes) services. See, for example, the empirical studies by Okrent (1974) and Engelstein (1983).

We let $B_{i,k}$ denote the number of passengers boarding vehicle $i$ at station $k$, and $A_{i,k}$ denote the number of passengers alighting from vehicle $i$ at station $k$. When there is no control, the number of passengers boarding vehicle $i$ at station $k$ is the number of passengers randomly arriving at $k$ during $h_{i,k}$. That is, the expected number of boarding passengers is:

\[(2.2) \quad B_{i,k} = r_k h_{i,k}\]

Further, we assume:

**A5.** When there is no control action taken on vehicle $i$ at station $k$, the number of passengers alighting at station $k$, $A_{i,k}$, is a fixed proportion $q_k$ of its arrival load at $k$, denoted $L_{i,k-1}$.

Hence, following Tsygalnitsky (1977) and Simon and Furth (1985):

\[(2.3) \quad A_{i,k} = q_k L_{i,k-1}\]

The *departure load* of vehicle $i$ from station $k$, $L_{i,k}$, is thus given by:

\[(2.4) \quad L_{i,k} = L_{i,k-1} + B_{i,k} - A_{i,k} = L_{i,k-1} (1-q_k) + B_{i,k}\]

This says that the departure load of a vehicle at a station is its arrival load plus passenger boardings less passenger alightings.
2.3 Passenger/Vehicle Interaction

Obviously, passengers and vehicles interact in a variety of ways to influence the performance of the system. Letting $s_{i,k}$ denote the dwell time of vehicle $i$ at station $k$, we assume:

A6. The general no-control dwell time function is of the form

$$(2.5) \quad s_{i,k} = c_0 + c_1 b_{i,k} + c_2 a_{i,k} \quad c_0 > 0, \quad 0 \leq c_i r_k < 1$$

where $c_0$, $c_1$, and $c_2$ are constant parameters typically empirically derived. For empirical evidence supporting this dwell function, the reader is referred to Lin and Wilson (1992). Given this, it follows that

$$(2.6) \quad d_{i,k} = a_{i,k} + s_{i,k}$$

In the event a control action is taken, we assume:

A7. Passengers do not change their destinations because of the control.

As a result of control actions on vehicle $i$ or because of capacity constraints, there may be leftover passengers, $P_{i,k}$, at a station $k$. Leftover passengers are those who waited at station $k$ but could not get on vehicle $i$. Of course, when there is no control, $P_{i,k}=0$. Particular definitions of $P_{i,k}$ depend on individual strategies and will be given later in the corresponding chapters.

2.4 System Types

In this dissertation we consider two different typical systems, characterized by different vehicle dwell time functions and passenger demand patterns. These systems are briefly described below.
**System F**: No-control dwell times of vehicles at any station are a constant independent of passenger activity. That is, \( s_{i,k} = c_0 \) for all \( i \) and \( k \). Free interstation running times are a constant across vehicles. That is, \( R_{i,k} = R_k \) for all \( i \).

**System G**: The dwell time function is of the form \( s_{i,k} = c_0 + c_1 B_{i,k} + c_2 A_{i,k} \). That is, vehicle dwell times depend on both passenger boardings and alightings. Also, free interstation running times are constant across vehicles. That is, \( R_{i,k} = R_k \) for all \( i \).

More about these two types of systems can be found in Chapter 3. Obviously, the complexity in modeling the strategies increases from System F to System G. On the other hand, both of them can be representative of real systems. By studying the systems in the above order, this dissertation proceeds from low to high level of complexity, and develops algorithms from exact to inexact ones. In this way, the simpler system provides useful insights to the more complex one.

### 2.5 Objective Function

Let \( w_{i,k} \) denote passenger waiting time at station \( k \) for vehicle \( i \), including both randomly arrival passengers and passengers leftover by previous vehicles, the objective function for minimizing total passenger waiting time for vehicle set \( I_m \) in the control direction is written in a general form as follows:

\[
(2.7) \quad f(h) = \sum_{i \in I_m} \left[ \sum_{k \in K_c} \left( r_k h_{i,k}^2 / 2 + P_{i-1,k} h_{i,k} \right) + u_c \sum_{i \in K_c} r_k h_{i,k}^2 / 2 \right]
\]

where \( 0 \leq u_c \leq 1 \).

The first term in the objective function represents waiting time of randomly arriving passengers during headway \( h_{i,k} \), and the second term represents extra waiting time of passengers left by the previous vehicle \( i-1 \). We define \( P_{0,k} = 0 \), where \( i=0 \) refers to the dummy vehicle preceding the first vehicle of the day. The third term represents the terminal condition at \( k_f \), where the actual passenger waiting time is zero because there are no boardings. The motivation of using a weight \( u_c \sum_{i \in K_c} r_k \) for vehicle headway at this
station is to take into consideration future cost in the opposite direction. If there is no layover at station $k$, any control action taken in the control direction will directly impact the other direction. In this case, we can set the parameter $u_c = 1$, and this term becomes the cumulative passenger arrival rates in the other direction, i.e., $\sum_{k \in K_c} r_k$. On the other hand, if there is a layover at $k$, we need a smaller weight because the impact of control on the other direction will be largely, or even completely, canceled by the layover. Such a cancellation effect depends on both the scheduled layover time and local practice, and the value of $u_c$ can be set between 0 and 1 to reflect such effects.
CHAPTER 3

SYSTEM TYPES AND THEIR PROPERTIES

In this chapter we describe the two types of transit systems and study their properties in the absence of control. We start with two definitions.

**Definition 3.1**

A trajectory of a vehicle $i$ is an ordered set of arrival and departure times at all stations $k \in K$, namely $\langle a_{i,k_0}, d_{i,k_0}, a_{i,k_0+1}, d_{i,k_0+1}, \ldots a_{i,k_r}, d_{i,k_r} \rangle$.

**Definition 3.2**

A segment of a transit network, denoted by $[k_1,k_2] \subseteq K$, is an ordered set of stations $<k_1, k_1+1, \ldots, k_2-1, k_2>$. 

3.1 System $F$

System $F$ is a transit system in which the following restrictions apply in addition to restrictions A1-A7:

A3.1 Dwell times at non-control stations are constant for all vehicles, i.e., $c_1 = c_2 = 0$, and hence $s_{ik} = c_0 > 0$ in (2.5).

A3.2 The passenger arrival rate is constant at every station other than $N$ and $N/2$, i.e., $r_k = r > 0$ for $k \neq N, N/2$. 

49
The passenger alighting ratio is constant at every station other than \( k_0 \) and \( k_r \), i.e.,

\[ q_k = q > 0 \text{ for } k \neq k_0, k_r. \]

Even though these assumptions make for the simplest type of transit system, they are not entirely unrealistic. In particular, assumptions A3.1 to A3.2 may apply to a rapid transit service with the following characteristics:

(i) The variation between passenger arrival rates at different stations is small.
(ii) Dwell times are relatively insensitive to the number of passenger boardings and alightings. This may be an acceptable assumption for some rapid transit systems with multiple doors per railcar, and the car floor level with the station platform. In this case the constant term in the dwell time function (2.5) may dominate and the general dwell time function may be reduced to \( c_0. \)

**Lemma 3.1 (Separability of System F)**

Given the definitions of vehicle arrival time and departure time in equations (2.1) and (2.6) respectively, and a fixed feasible arrival time of each vehicle \( i \) at station \( k_0 \) (i.e., \( a_{i,k_0} \)), an uncontrolled vehicle's trajectory is unaffected by trajectory changes of any other vehicles.

**Proof** A feasible arrival time of vehicle \( i \) at station \( k_0 \) satisfies

\[
(3.1) \quad a_{i,k} \geq d_{i,1,k} + h_0, \quad k=k_0.
\]

Also, the first vehicle in the system is not blocked by any other vehicle and satisfies

\[
(3.2) \quad a_{i,k} = d_{i,k-1} + R_k + 2\delta, \quad i=1, k > k_0
\]

Now assuming relation (3.1) holds at a station \( k-1 \geq k_0 \) for any vehicle \( i \), and relation (3.2) holds for a vehicle \( i-1 \geq 1 \) at any station \( k \), then

\[
a_{i,k+1} + c_0 + R_k + 2\delta \geq d_{i,1,k+1} + h_0 + R_k + 2\delta + c_0,
\]

which implies \( d_{i,k+1} + R_k + 2\delta \geq d_{i,1,k} + h_0 \) for vehicle \( i \geq 1 \) at station \( k > k_0 \). Thus we have

\[
A \text{ fixed dwell time is actually used as a headway control strategy in very high frequency transit lines (e.g., <3 min.) in Japan.}
\]

50
proved by recursion that \( d_{i,k-1} + R_t + 2\delta \geq d_{i,1,k} + h_o \) holds for any vehicle at any station in System F. And hence for System F., arrival time equation (2.1) is reduced to

\[(3.3) \quad a_{i,k} = d_{i,k-1} + R_t + 2\delta, \quad \forall (i,k).\]

Substituting (3.3) into \( d_{i,k} = a_{i,k} + c_o \) yields

\[(3.4) \quad d_{i,k} = d_{i,k-1} + R_{i,k} + 2\delta + c_o, \quad \forall (i,k)\]

Solving this recursive equation given \( a_{i,k_0} \), we have

\[(3.5) \quad d_{i,k} = a_{i,k_0} + \sum_{j=k_0+1}^{k} (R_j + 2\delta) + \sum_{j=k_0}^{k} c_0 = a_{i,k_0} + \sum_{j=k_0+1}^{k} R_j + (k - k_0)(2\delta + c_0) + c_0\]

Clearly all terms on the right-hand side of (3.5) are fixed values associated with vehicle \( i \) and station \( k \). Thus each \( d_{i,k} \) is uniquely defined by \( i \) and \( k \) and independent of any other vehicle’s movement. Also, it follows from (3.3) that \( a_{i,k} \) for \( k > k_0 \) depends solely on \( d_{i,k-1} \). Hence once \( d_{i,k} \) values are determined, each \( a_{i,k} \) is also determined. Thus the result follows. \( Q.E.D. \)

Lemma 3.1 is used in Chapter 4 for developing an exact solution to the real-time deadheading problem in System F.

Lemma 3.1 also shows that, once departed, a vehicle’s headway will stay constant along the route when there is no intervention on it or its preceding vehicle. Fig. 3.1 illustrates trajectories of 13 vehicle trips computed using equation (3.3) and (3.4), where the initial arrival times are taken from Green Line data set "w1". The constant dwell time used here is \( c_0 = 0.37 \) (minutes), and \( \delta = 0.15 \) (minutes). The constant headway pattern for each vehicle can be clearly seen from Fig. 3.1.
3.2 System G

System G is a general transit system under assumptions A1-A7. Here the dwell function takes the general form of equation (2.5). That is,

\[ s_{i,k} = c_0 + c_1 r_{i,k} h_{i,k} + c_2 q_i L_{i,k-1} \]

which says the dwell times depend on both passenger boardings and alightings. Because passenger alightings depend on vehicle arrival load, the dwell time function also requires a vehicle load equation:

\[ L_{i,k} = L_{i,k-1} + (h_{i,k} r_k - L_{i,k-1} q_i) \]

The main differences between System G and F are:

(i) Vehicle dwell time depends on passenger demand.

(ii) Passenger demand varies by station.
(iii) A vehicle may be blocked by its preceding vehicle.

As a consequence of these differences, a vehicle's headway is no longer constant along the route, which makes System G much more complicated than System F. Fig. 3.2 depicts no-control vehicle trajectories in System G, computed from the Wednesday morning Green Line data, where all the initial conditions are the same as in Fig. 3.1, except the dwell function parameters $c_0 = 0.2$ (minutes), $c_1 = 0.007$ (minutes/passenger) and $c_2 = 0.008$ (minutes/passenger). For source of the empirical values of Green Line dwell function parameters, see Lin and Wilson (1992). The trajectories are computed using equations (2.1), (2.6), (3.6) and (3.7). The demand pattern is given in Appendix B.

One can observe vehicle headway changes along the route from Fig. 3.2 (for example, the 4th vehicle). Also obvious is that at some stations a vehicle is blocked by its preceding vehicle (e.g., at station 25). While they do not appear in Fig. 3.1 for System F, these headway irregularities in System G come from the variable dwell times accentuating the
dispatching headway irregularities. In the following subsection we analyze both variable dwell time effects and dispatching headway effects in System G.

3.2.1 Dwell Time Effects

Given dispatching headways with non-zero standard deviation, headway variation at other stations along the route will increase due to variable dwell times. To show this, we computed headway variation at each station along the route of Green Line, after computing the vehicle trajectories (in System G) using equations (2.1), (2.6), (3.6) and (3.7). The vehicle departure times at station 1 (hence the dispatching headways) are actual data from the Green Line during the morning peak from Monday to Friday in the week of October 16-20, 1993. Fig. 3.3 plots the standard deviation of the 5 weekday morning peak headways along the route.

From Fig. 3.3 we can see that, without exception, headway variance increases along the route in every data set. The steepest increase occurs in the middle of the route, from station 25 (Park Street IB) to 31 (Copley OB). These stations are at the Government Center end of the B line, which form the peak demand segment (see Appendix B). Before station 25 and after station 31, the headway variance curves are quite flat, with only slight increases at each station. This is because demand at those stations is low, and hence dwell time differences between vehicles is small at each of those stations. At the end of the route, variance increases may also be limited due to the limits on headway changes. Fig. 3.3 clearly shows that the impact of dwell times (due to uneven demand) on headway variation is very significant - the standard deviation of headways starts under 2 minutes at dispatching, but end up between 5 and 7 minutes at the terminal station 52!
As a result of the headway variation due to dwell time effects, passenger waiting times increase significantly. Table 3.1 compares (in an approximate way) the difference between System F and System G, where the dispatching headways, the average dwell times (about 0.37 minutes), the average passenger arrival rates (about 2.4 psg./min.) between the two systems are the same. The resulting total passenger cost over the five morning peaks in System G is about 21% higher than in System F, and such a significant cost increase is solely due to the headway variance caused by variable dwell times.

This result also shows that we could reduce passenger waiting times simply by imposing constant dwell time at every station for all vehicles (provided such a dwell time is large enough for all passengers to board or alight). Of course, such a solution is far from optimal because it unnecessarily increases dwell times (and hence total trip times) at low demand stations. Nonetheless it shows the important role that variable dwell time plays in passenger waiting times.
<table>
<thead>
<tr>
<th>Data Set</th>
<th>M</th>
<th>Cost (psg-min)</th>
<th>Mean $h$ (min)</th>
<th>StDev of $h$</th>
<th>Cost (psg-min)</th>
<th>Mean $h$ (min)</th>
<th>StDev of $h$</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>36</td>
<td>49,114.70</td>
<td>4.66</td>
<td>1.32</td>
<td>62,466.84</td>
<td>4.61</td>
<td>1.84</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>50,409.10</td>
<td>4.85</td>
<td>1.13</td>
<td>62,700.51</td>
<td>4.99</td>
<td>1.73</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>55,021.20</td>
<td>5.02</td>
<td>1.66</td>
<td>67,756.57</td>
<td>4.97</td>
<td>2.17</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>55,002.80</td>
<td>4.87</td>
<td>1.39</td>
<td>66,082.73</td>
<td>4.86</td>
<td>1.92</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>43,244.70</td>
<td>4.81</td>
<td>0.99</td>
<td>52,427.08</td>
<td>4.91</td>
<td>1.52</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>38,361.50</td>
<td>4.60</td>
<td>3.94</td>
<td>38,820.45</td>
<td>4.57</td>
<td>5.74</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>37,232.50</td>
<td>4.92</td>
<td>3.53</td>
<td>37,792.15</td>
<td>5.12</td>
<td>5.30</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>39,846.70</td>
<td>4.95</td>
<td>4.00</td>
<td>38,118.50</td>
<td>4.94</td>
<td>5.55</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>38,018.00</td>
<td>4.82</td>
<td>3.49</td>
<td>38,640.17</td>
<td>4.86</td>
<td>5.40</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>30,820.70</td>
<td>4.87</td>
<td>3.27</td>
<td>29,938.91</td>
<td>5.00</td>
<td>4.79</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>437,071.90</td>
<td></td>
<td></td>
<td>494,743.91</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg.</td>
<td></td>
<td>87,414.38</td>
<td>4.84</td>
<td>2.47</td>
<td>98,948.78</td>
<td>4.87</td>
<td>3.60</td>
</tr>
</tbody>
</table>

Note: $H=5$ min. $c_i = 0.37$ min. in System F. Avg. cost is daily average.

Table 3.1 No-Control Status: System F vs System G

3.2.2 Dispatching Headway Effects

In System G, the overall headway variation along the route also depends on the magnitude and pattern of initial headways. Here we present some results regarding the effects of initial headways.

**Proposition 3.1 (Dispatching Headway Effects)**

(i) In System G, $h_{i,k}$ are monotonically increasing along the route if $h_{i,k_0} \geq \max_k (h_{i-1,k})$, and are monotonically decreasing along the route if $h_{i,k_0} \leq \min_k (h_{i-1,k})$, $k_0 \leq k \leq k_f$ until either vehicle $i$ or $i-1$ is blocked by its previous vehicle, or $i$ reaches the terminal of the route.

(ii) Given a vehicle set $I_m = (i, i+1, ... i+m)$, if all vehicles $j \in I_m$ have dispatching headways equal to the dispatching headway of $i-1$, that is, $h_{j,k_0} = h_{i-1,k_0}, \forall j \in I_m$, and vehicle $i-1$ also has equal headways at all stations along the route, that is, $h_{i-1,k} = h_{i-1,k_0}, \forall k \in K$, then all vehicles $j \in I_m$ have equal headways at all $k \in K$ regardless of demand pattern.
Proof: We first prove (i) for the "dispatching headway effects" when: \( h_{i,k_0} \geq \max(h_{i-1,k}) \). The proof for the case where \( h_{i,k_0} \leq \min(h_{i-1,k}) \) is identical except that "\( \geq \)" is replaced by "\( \leq \)" and "max" is replaced by "\( \min \)", and hence is omitted.

With constant interstation running times, when vehicle \( i \) is not blocked by \( i-1 \), the headway change of \( i \) between two consecutive stations \( k-1 \) and \( k \) is due solely to the dwell time difference

\[
(3.7) \quad h_{i,k} - h_{i,k-1} = s_{i,k} - s_{i-1,k} = c_1 r_k (h_{i,k} - h_{i,k-1}) + c_2 q_k (L_{i,k} - L_{i-1,k-1})
\]

which implies \( h_{i,k} = h_{i,k-1} - c_1 r_k h_{i,k-1} + c_2 q_k (L_{i,k-1} - L_{i-1,k-1}) / (1-c_1 r_k) \), and hence,

\[
(3.8) \quad h_{i,k} \geq h_{i,k-1} \iff h_{i,k-1} - c_1 r_k h_{i,k-1} + c_2 q_k (L_{i,k-1} - L_{i-1,k-1}) / (1-c_1 r_k) \geq h_{i,k-1} - 1 - c_1 r_k h_{i,k-1} + c_2 q_k (L_{i,k-1} - L_{i-1,k-1}) \geq 0
\]

Now if \( h_{i,k_0} \geq \max_k(h_{i-1,k}) \), we have \( h_{i,k_0} - h_{i-1,k_{0+1}} \geq 0 \), because \( h_{i-1,k_{0+1}} \leq \max_k(h_{i-1,k}) \). Also, it follows from (3.7) that \( L_{i,k_0} \geq L_{i-1,k_0} \) since \( h_{i,k_0} \geq h_{i-1,k_0} \). Thus we have \( c_1 r_k (h_{i,k_0} - h_{i-1,k_{0+1}}) + c_2 q_k (L_{i,k_0} - L_{i-1,k_0}) \geq 0 \) because \( c_1 r_k \geq 0 \) and \( c_2 q_k \geq 0 \). Hence, it follows from (3.8) that \( h_{i,k_0+1} \geq h_{i,k_0} \). Now suppose \( h_{i,k} \geq h_{i,k-1} \geq \ldots \geq h_{i,k_0} \) holds, because \( h_{i,k_0} \geq \max(h_{i-1,k}) \), we have \( h_{i,j} \geq h_{i,j} \) at all \( j \leq k \), and hence \( L_{i,j} \geq L_{i-1,j} \) at all \( j \leq k \). Because of the same reason we also have \( h_{i,k} \geq h_{i,k+1} \), thus \( c_1 r_k (h_{i,k} - h_{i,k+1}) + c_2 q_k (L_{i,k} - L_{i-1,k}) \geq 0 \) and it follows from (3.8) that \( h_{i,k+1} \geq h_{i,k} \). Thus we have proved (i) by induction.

The proof for (ii) is straightforward in light of (i). Given that \( h_{j,k_0} = h_{j-1,k_0}, \forall j \in I_m \), and \( h_{i-1,k_0} = \min_k(h_{i-1,k}) = \max_k(h_{i-1,k}), \forall k \), both \( h_{i,k_0} \leq \min_k(h_{i-1,k}) \) and \( h_{i,k_0} \geq \max_k(h_{i-1,k}) \) are true. Hence by (i) we have both \( h_{i,k} \geq h_{i,k-1} \) and \( h_{i,k} \leq h_{i,k-1} \iff h_{i,k} = h_{i,k-1} \) for all \( k \). The same can be shown for \( j > i, j \in I_m \). Q.E.D.

This proposition is very useful because we can use it to predict at station \( k_0 \) the headway pattern of any vehicle \( i \) down the line, with the readily available information of \( h_{i,k_0} \) and \( h_{i-1,k} \). This is very helpful when making control decisions at dispatching. It will be used in following chapters for analyses and developing algorithms for System G.

Fig. 3.4 illustrates Proposition 1 by plotting headways of four consecutive vehicles along the route, computed from their trajectories in System G using the Green Line data. It can
be clearly seen that, since vehicle 6's dispatching headway (at station 1) is larger than the maximal headway of vehicle 5, vehicle 6 headway increases all the way along the route. Vehicle 7's dispatching headway is smaller than the minimal headway of vehicle 6, so its headway decreases along the route, until it is blocked by vehicle 6 at station 26. (Note that once a vehicle is blocked at some point, its headway at later stations may no longer be monotone.) One can also observe that, if a vehicle's dispatching headway is between maximal and minimal headways of its preceding vehicle, its headway can change non-monotonically along the route. This is the case for vehicle: its headway first decreases, then starts to increase at station 47.

![Headway (min.)](image)

<table>
<thead>
<tr>
<th>i</th>
<th>h</th>
<th>hdw ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>5.29</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>4.61</td>
<td>0.87</td>
</tr>
<tr>
<td>6</td>
<td>5.64</td>
<td>1.22</td>
</tr>
<tr>
<td>7</td>
<td>3.98</td>
<td>0.7</td>
</tr>
</tbody>
</table>

Fig 3.4 Dispatching Headway Effects (i)

On the other hand, Proposition 3.1 and Fig. 3.4 also indicate that, if the dispatching headway of vehicle $i$ is between $\min_k(h_{i-1,k})$ and $\max_k(h_{i-1,k})$, its headway variation across stations will be smaller than otherwise. This is why the variance of vehicle 5 headway along the route is apparently smaller than other vehicles. Headway data computed from
no-control trajectories based on the Green Line data sets indeed verify this. Figs. 3.5a and 3.5b plot, for each direction, three headway related measures for vehicle trips in which interstation stopping occurs neither to them nor their preceding vehicle trips. The three measures are: standard deviation of headways across stations of vehicle trip \( i \); the difference between each vehicle's dispatching headway and the mid range of its preceding vehicle's headway, that is, \( h_{i,k} = \frac{1}{2} \{ \min_k(h_{i-1,k}) + \max_k(h_{i-1,k}) \} \), referred as the "relative dispatching headway" of \( i \); and the headway difference between stations \( k_0 \) and \( k_r \), called "terminal headway difference" of vehicle trip \( i \).

![Fig 3.5a System G: Headway Relations (Direction 1)](image)

![Fig 3.5b System G: Headway Relations (Direction 2)](image)
There are a number of important observations from Fig. 3.5. First, we can see that, regardless of demand pattern differences, headway variation across stations of a vehicle trip \( i \) tends to increase when its dispatching headway is further away from the mid range of \( i-1 \)'s headways, \([\min_t(h_{i-1,k}) + \max_t(h_{i-1,k})]/2\). Second, terminal headway difference of vehicle trips overlap with their relative dispatching headway. Third, in each figure above, the three curves intersect at value "0". The implication of these observations is that, to minimize headway variation across stations for vehicle trip \( i \), its dispatching headway should be around the mid range of \( i-1 \)'s headway. Also, terminal headway difference can be used to replace standard deviation as a measure of headway change over the route, since it is much easier to calculate.

The above results on dispatching headway effects are significant when considering real-time control decisions. First, they allow us to foresee a vehicle's headway change tendency along the route with only past and current information - headways of the preceding vehicle trip and the initial headway of the present vehicle. Second, they imply that control at the dispatching station is very important and effective. Furthermore, the above results are not demand dependent, which means they are generally applicable in System G type transit systems, regardless of the particular demand pattern.

We should note that although in the above proof the initial station is given as the dispatching station, the proposition can be easily extended to any station, provided that the vehicle load difference at the initial station is taken into account. In later chapters, the properties presented above will be used in developing control algorithms for System G.

3.2.3 Trajectory Change Diminishing Effect

In a System G type transit services with \( c_i r_i < 0.5 \), the impact of a change in trajectory of vehicle \( i \) on its following vehicles diminishes quickly. We prove this informally below.

Let \( i \) and \( i+1 \) be two uncontrolled adjacent vehicles. The relation between the two vehicles departure times at station \( k \), when they are not blocking each other, is:
\[ d_{i+1,k} = a_{i+1,k} + c_0 + c_i r_k (d_{i+1,k} - d_{i,k}) \]
\[ \Rightarrow (1 - c_i r_k) d_{i+1,k} = a_{i+1,k} + c_0 - c_i r_k d_{i,k} \]
\[ \Rightarrow d_{i+1,k} = \frac{(a_{i+1,k} + c_0 - c_i r_k d_{i,k})}{(1 - c_i r_k)}, \quad k=k_0. \]

Now denoting the original departure times by \( d_{i,k}^0 \) and \( d_{i+1,k}^0 \) respectively, and letting \( d_{i,k} = d_{i,k}^0 + \Delta d_{i,k} \), then we have at \( k=k_0 \):

\[ d_{i+1,k} = a_{i+1,k} + c_0 + c_i r_k (d_{i,k}^0 + \Delta d_{i,k}) \]
\[ \Rightarrow (1 - c_i r_k) d_{i+1,k} = a_{i+1,k} + c_0 - c_i r_k (d_{i,k}^0 + \Delta d_{i,k}) \]
\[ \Rightarrow d_{i+1,k} = \frac{(a_{i+1,k} + c_0 - c_i r_k d_{i,k}^0)}{(1 - c_i r_k)} - \frac{[c_i r_k/(1 - c_i r_k)] \Delta d_{i,k}}{1 - c_i r_k}, \quad k=k_0 \]

That is,

\[ d_{i+1,k} = d_{i,k}^0 - \frac{[c_i r_k/(1 - c_i r_k)] \Delta d_{i,k}}{1 - c_i r_k}, \quad k=k_0 \]

This says that at station \( k_0 \), because the arrival time is fixed, the departure time change is due to the dwell time change alone, which is \([c_i r_k/(1 - c_i r_k)] \Delta d_{i,k}\). At a station \( k > k_0 \), the departure time change is due to both arrival time and dwell time change, and we have

\[ d_{i+1,k} = d_{i,k}^0 + \frac{(\Delta a_{i+1,k} - c_i r_k \Delta d_{i,k})}{(1 - c_i r_k)}, \quad k > k_0 \]

Then for \( i + 2 \) we have

\[ d_{i+2,k} = d_{i+2,k}^0 - c_i r_k \frac{([- c_i r_k \Delta d_{i,k}]/(1 - c_i r_k)]}{(1 - c_i r_k)}, \quad k = k_0. \]
\[ d_{i+2,k} = d_{i+2,k}^0 + [\Delta a_{i+2,k} - c_i r_k (\Delta a_{i+1,k} - c_i r_k \Delta d_{i,k})/(1 - c_i r_k)]/(1 - c_i r_k), \quad k > k_0 \]

In general, we have

\[ (3.9) \quad \Delta d_{i,m,k} = d_{i,m,k}^0 - d_{i+1,k}^0 = [- c_i r_k/(1 - c_i r_k)]^m \Delta d_{i,k}, \quad k = k_0 \]
\[ (3.10) \quad \Delta d_{i,m,k} = d_{i,m,k}^0 - d_{i+1,k}^0 = \sum_{j=1}^{m} \frac{(c_i r_k)_{m-j}}{(1 - c_i r_k)^{m-j}} \Delta a_{i+1,k}^m - [- c_i r_k/(1 - c_i r_k)]^m \Delta d_{i,k}, \quad k > k_0 \]

Because \( c_i r_k < 0.5 \), \( c_i r_k/(1 - c_i r_k) < 1 \) and \( \lim(c_i r_k/(1 - c_i r_k))^m = 0 \) as \( m \to \infty \). When \( m \geq 2 \), \([c_i r_k/(1 - c_i r_k)]^m \) becomes very small. In the Green Line case, the average \( c_i r_k = 0.02 \), and the largest \( c_i r_k = 0.27 \), so \([c_i r_k/(1 - c_i r_k)]^3 = 0.000008 \) in average, and the largest \([c_i r_k/(1 - c_i r_k)]^3 = 0.02 \). Considering the fact that the number of stations which have large arrival rates (such as transfer stations) is usually small in a transit network, the average
value is more indicative overall. Such a small value of \( \left( c_i r_k / (1 - c_i r_k) \right)^3 \) indicates that the trajectory change of vehicle \( i \) has an insignificant impact on trajectory of vehicle \( i+3 \). At station \( k_0 \) the impact of \( \Delta d_{i,k} \) on \( i+3 \) is insignificant as shown by (3.9). At a station \( k > k_0 \), the arrival time change of a vehicle \( j \), \( \Delta a_{i,j,k} \), is equal to its departure time change at \( k-1 \), thus it is actually the sum of \( \Delta d_{i,k} \) from station \( k_0 \) to \( k \). That is, \( \Delta a_{i,j,k} = \sum_{k_0}^k \Delta d_{j,k'} \), \( k_0 < k' < k \). As \( \Delta d_{i+m,k} \) at each station \( k \) gets smaller as \( m \) increases, so does \( \Delta a_{i+m,k} \). Also, in the term
\[
\sum_{j=1}^{m} \frac{(c_i r_k)^{m-j}}{(1 - c_i r_k)^{m-j+1}} \Delta a_{i+j,k} \text{ the weight of } \Delta a_{i+1} \text{ gets smaller as } m \text{ increases. This is because }
\]
\[
(c_i r_k)^{m-j} / (1 - c_i r_k)^{m-j+1} < (c_i r_k)^{m-2} / (1 - c_i r_k)^{m-1} \text{ when } c_i r_k < 0.5, \text{ so we have } 1/ (1 - c_i r_k) > (c_i r_k)
\]
\[
/ (1 - c_i r_k)^2 \geq 1/ \Delta a_{i+m,k} (c_i r_k)^{m-1} / (1 - c_i r_k)^m. \text{ So as the impact of } \Delta d_{i,k} \text{ diminishes, this term also diminishes.}
\]

The significance of the existence of such a small \( m \) in transit operations should not be underestimated. It implies that the future system performance (for vehicle \( i+3 \) and beyond) is largely independent of current control decisions on vehicle \( i \). This suggests that a rolling horizon scheme of dynamic control is reasonable and effective.

3.2.4 **Alternating Sign Effect**

It should be specially noted that (3.9) and (3.10) also show sign difference of headway changes along the vehicle dimension. That is, the sign of \( \Delta d_{i+m,k} \) caused by \( \Delta d_{i,k} \) changes with \( m \) being odd or even, provided that there is no interstation stopping of the vehicles. At \( k = k_0 \), \( \Delta d_{i+m,k} (m=0,1,2,...) \) will have alternating signs starting with the sign of \( \Delta d_{i,k} \). At \( k > k_0 \), the term \( \sum_{j=1}^{m} \frac{(c_i r_k)^{m-j}}{(1 - c_i r_k)^{m-j+1}} \Delta a_{i+j,k} \) for \( m > 1 \) is dominated by \( \Delta a_{i+m,k} / (1 - c_i r_k) \), which has the same sign as \( \Delta d_{i+m,k} \). Thus the vehicle trajectory changes also have alternating signs at \( k > k_0 \).

Such alternating signs have different effects on station skipping and holding strategies. To see this, let us consider a set of \( m+1 \) vehicles where \( i \) is the first vehicle in the set and is to be controlled, and the cost of control is computed over all vehicles in the set. Let \( \Delta w_m(i) \) denote the set cost change after control, and \( \Delta f_j, j=i,...,i+m \) denote the cost
change associated with vehicle $j$'s preceding headways in the direction. With station skipping strategies, $\Delta f_i$ has a negative value because $i$'s preceding headways are reduced, and $\Delta f_{i+1}$ has positive sign due to the extra dwell time for leftover passengers boarding $i+1$ at skipped stations. When $m=1$, $\Delta w_m(i) = \Delta f_i + \Delta f_{i+1} = -|\Delta f_i| + |\Delta f_{i+1}|$. A local minimum occurs at the number of skipped stations $n^*$ if $|\Delta f_{i+1}| < |\Delta f_i|$ at $n^*-1$ and $|\Delta f_{i+1}| > |\Delta f_i|$ at $n^*+1$. Since larger $n$ leads to larger $\Delta f_{i+1}$, the use of $m=1$ in cost evaluation leads to a more conservative control policy than when $m>1$ is used. When $m=2$, $\Delta w_m(i) = -|\Delta f_i| + |\Delta f_{i+1}| - |\Delta f_{i+2}|$. This time when $f_{i+1}$ is increasing $f_{i+2}$ will actually decrease, and if the headway of $i+2$ is large, $\Delta f_{i+2}$ may also make a large contribution to set cost reduction. In this case the use of $m=2$ may result in a more severe control action than otherwise. When $m=3$, $\Delta w_m(i) = -|\Delta f_i| + |\Delta f_{i+1}| - |\Delta f_{i+2}| + |\Delta f_{i+3}|$. $\Delta f_{i+3}$ would be quite small according to the "trajectory change diminishing effect", but its positive sign can cancel some of $\Delta f_{i+2}$ and help to balance the control action. When $m>3$, $\Delta f_{i+m}$ will be too small to have significant impact on $\Delta w_m(i)$.

With a holding strategy alone, $\Delta f_i$ has a positive sign since the headway of $i$ increases. In this case if $m=1$ is chosen the benefit of holding may be overestimated, while if $m=2$ is chosen the benefit may be underestimated. Again $m=3$ may be a balanced choice.

The alternating sign effects are subject to change when interstation stopping of vehicles is present. Such a change depends on the type of control strategy and will be explored in the following chapters.
CHAPTER 4

THE REAL-TIME DEADHEADING PROBLEM

4.1 Introduction

This chapter studies the real-time deadheading problem (RTDP). When a vehicle is deadheaded, it runs empty through a number of stations in order to save time and thus reduce headways at later stations. Deadheading starts at a terminal, after all passengers have already alighted (by choice) and the control decision is made before a vehicle leaves station $N$ ($N/2$). When a vehicle is deadheaded, station $1$ ($N/2+1$) is always the first station to be skipped. In comparison, another similar control strategy, expressing (see Chapter 5), in which a vehicle also skips a number of stations, can start at either a terminal or an intermediate station. The decision to express a vehicle is announced to passengers at the starting station, and passengers traveling to stations to be skipped must alight (or not board) before the vehicle starts to operate express. One advantage of deadheading is that it saves more time at the starting station, because there are no passenger alightings and boardings at all, and no explanations are required. In addition, it may be less frustrating since no passengers are forced to alight. The benefit of deadheading is that the waiting time of passengers at stations beyond the skipped stations can be reduced. The price of deadheading is attributed to the passengers who are skipped by the deadheaded vehicle. Another potential price of deadheading is the loss of capacity
over the deadhead segment. This can be a problem in peak hours when demand may approach vehicle capacity. The Real-time deadheading problem (RTDP) is to decide, at any given time, which vehicles should be deadheaded and how many stations should be skipped by each deadhead vehicle in such a way as to minimize the total passenger cost.

This chapter is organized as follows. In section 4.2, we develop a general model for the RTDP. In section 4.3, we study a special case of RTDP for System $F$, which can be solved optimally. In section 4.4, we study RTDP for System $G$. Numerical examples are provided in each section.

### 4.2 The General RTDP Model

#### 4.2.1 Decision Variables and Objective Function

**Decision Variables**

The decision variables in RTDP are binary, denoted as $y_{i,k}$, defined as follows:

\[
y_{i,k} = \begin{cases} 
1 & \text{if vehicle } i \text{ stops at station } k \\
0 & \text{otherwise}
\end{cases}
\]

Let $\mathbf{y} = (y_{i,k} : i \in I_m, k \in K_c) \in \{0, 1\}^{m \times k}$ denote the complete set of decision variables. Similarly, $\mathbf{y}_i = (y_{i,k} : k \in K_c)$ denotes the decision variables associated with vehicle $i$.

**Objective Function**

As discussed in both Chapters 1 and 2, our objective is the minimization of total passenger waiting time among a set of vehicle trips. For a vehicle set $I_m$, the objective function is written as

\[
f(h(\mathbf{y})) = \sum_{i \in I_m} \left\{ \sum_{k \in K_c} \left[ r_k (d_{i,d_{i-1,k}})^2 / 2 + P_{i-1,k}(d_{i,k} - d_{i-1,k}) \right] + \right. \\
\left. (u_c \sum_{k \in K_c} r_k (d_{i,k,i} - d_{i-1,k,i})^2 / 2) \right\}, \quad 0 \leq u_c \leq 1
\]

65
This objective function is the same as (2.7), except that here we replace $h_{i,k}$ by its definition, $d_{i,k} \cdot d_{i-1,k}$. The objective function captures the impacts of deadheading on two groups of passengers: the passengers at stations beyond the skipped stations; and the passengers at the skipped stations. The first term captures the waiting time of randomly arriving passengers. Under the assumption that passengers arrive at a constant rate (see A6), the expected total waiting time of passengers who arrive at station $k$ between two adjacent vehicle departures $d_{i-1,k}$ and $d_{i,k}$ is $w_{i,k} = r_k(d_{i,k} - d_{i-1,k})^2/2$. This waiting time is non-zero wherever $r_k$ is non-zero, since $d_{i,k} - d_{i-1,k}$ is always positive due to the minimal safe headway requirement.

The second term captures the waiting time of those passengers who were waiting for vehicle $i$ at station $k$ but could not board because it skipped $k$, and had to wait for the next vehicle $i+1$. This term is additive between vehicles. If there are no "leftover" passengers, this term is zero. Note that since the boundary vehicle $i+m$ is not controlled, $P_{i+m,k} = 0$ at every station. In general, $P_{i,k}$ is a function of the decision variable $y_{i,k}$, and such a function may vary by system type as well as problem type. For RTDP, the definition of $P_{i,k}$ as a function of $y_{i,k}$ is given in section 4.3.1 for System F, and in section 4.4.1 for System G.

The third term incorporates the terminal condition at station $k$, by setting a weight $u_c \sum_{k \in K_c} r_k$, which can be interpreted as an artificial passenger arrival rate at $k$ (where the actual arrival rate is always 0). The value of $u_c$ is an engineering choice in the range of 0 to 1. See Chapter 2, section 2.5 for a more detailed discussion.

There is one important consequence of objective function (4.2.2) that deserves some discussion. Specifically, stations with different passenger arrival rates may be treated differently. That is, stations with smaller arrival rates are more likely to be skipped, since this will result in less extra waiting time in the deadhead segment. Though this is an efficient way of using vehicles from the operator’s view point, it will not be desirable for passengers waiting at such stations. Since the scheduled headway is the same at every station, passengers are likely to be upset if the actual headways at such stations are
consistently higher. One way to avoid such a conflict is to redesign the operating plan at
the route definition and vehicle scheduling levels (off-line). For example, to schedule
alternating deadheading and local vehicles at some stations and publish the schedule in
advance.

4.2.2 The RTDP Formulation

Before developing the model in detail, we first introduce several concepts.

Definition 4.1

A deadhead segment \([k_0, k_e] \subseteq K_e \subseteq K\) for vehicle \(i\) is an ordered set of stations
\(<k_0, k_0+1, \ldots, k_e>\), in which \(i\) skips \(n\) stations starting at \(k_0\), and then stops at \(k_e = k_0+n\). The
stations \(\{k: k_0 \leq k \leq k_e-1\}\) are called control stations, and \(k_0\) can only be either station 1 or
station \(N/2+1\). The stations \(\{k: k \in K, k_e \leq k \leq k_f\}\) are called benefit stations and the
segment \([k_e, k_f]\) is called the benefit segment associated with the deadhead segment \([k_0, k_e]\).

As an example, Fig. 4.3.1 (in section 4.3.2) depicts trajectories for two vehicles, including
trajectories of vehicle \(i\) both with and without deadheading. The deadhead segment for
vehicle \(i\) is \([1, 8]\). That is, after the deadheaded vehicle \(i\) departs from station 0 (i.e.,
station \(N\)), it skips seven stations 1 to 7, and then stops at station 8. The vehicle dwell
time shown in Fig. 4.3.1 is a constant, \(c_0\), at non-control stations, and is zero at skipped
stations. The benefit segment is \([8, 22]\), which joins the deadhead segment at station 8.

Definition 4.2

A vehicle \(i\) is deadheadable if and only if there exists a segment \([k_0, k_e] \subseteq K_e\) with \(k_0
< k_e < k_0+N/2\), such that

\[
(4.2.3) \quad d_{i,k-1} + R_{i,k} + (y_{i,k} + y_{i,k-1}) \delta_i \geq d_{i,k} + h_0, \quad \forall k: \, k_0 < k < k_e
\]

where \(y_{i,k} = 1\), and \(y_{i,k} = 0\) for \(k: \, k_0 < k < k_e\). Such a segment is then called a feasible
deadhead segment for vehicle \(i\). A no-control policy or a policy that deadheads a vehicle
on a feasible deadhead segment is a feasible policy.
In other words, Definition 4.2 says that a deadheadable vehicle $i$ must not be blocked by its preceding vehicle anywhere during the course of its deadheading.

Now, incorporating assumptions A1 to A8, we present a general formulation of the real-time deadheading problem (RTDP) at a given dispatching terminal $k_0$, where either $k_0 = 1$ or $k_0 = N/2+1$.

[RTDP] Minimize (4.2.2)

Subject to:

(4.2.4) $a_{i,k}(y) - d_{i-1,k}(y) \geq h_0 \quad \forall (i,k) \in I_m \times K_c$

(4.2.5) $y_{i,k} - y_{i,k+1} \leq 0 \quad \forall (i,k) \in I_m \times K_c$

(4.2.6) $\sum_{k \in K_c} y_{i,k} < N/2 \quad \forall i \in I_m$

(4.2.7) $y_{j+m,k} = 1 \quad \forall k \in K_c$;

(4.2.8) $y_{i,k} \in \{0, 1\} \quad \forall (i,k) \in I_m \times K_c$

where $j$ is the first vehicle in $I_m$; $P_{j-1,k}$, $d_{j+1,k} \forall k \in K_c$, $d_{i,k_0-1} \forall i \in I_m$ and $k_0$ are given

Inequality (4.2.4) is the deadheadability constraint (see Definition 4.2 above). (4.2.8) is the binary integer constraint on the decision variables. (4.2.5) ensures that the skipped stations are consecutive. Since a deadheaded vehicle always starts to skip stations from $k_0$, the skipped stations always have smaller index values than the other stations. Also, since the control variable has a smaller value (i.e., $y_{i,k} = 0$) at a skipped station than at a non-control station (i.e., $y_{i,k} = 1$), if a non-control station $k$ is followed by a skipped station $k+1$, we would have $y_{i,k} = 1 > y_{i,k+1} = 0$. Thus (4.2.5) holds only if any skipped stations are consecutive. (4.2.6) says that the number of skipped stations can be at most $N/2-1$, which implies that a deadhead segment cannot change directions. (4.2.7) imposes the boundary condition that vehicle $j+m$ is not deadheaded. Note that by our station index convention $k_0-1$ is station $N$ when $k_0 = 1$.

Unfortunately, [RTDP] is, for all intents and purposes, mathematically intractable. Aside from the usual difficulties of a nonlinear integer program, one main difficulty is due to the recursive nature of the departure times $d_{i,k}$ and the tie between $d_{i,k}$ and $y_{i,k}$. This is to say that, each vehicle's $d_{i,k}$ and $y_{i,k}$ are interdependent, and they are also dependent on the
$d_{tk}$ and $y_{tk}$ values associated with other vehicles. On the other hand, the variables $d_{tk}$ are not independent variables, in the sense that once all $y_{tk}$ values are determined, each $d_{tk}$ is also determined. When the general dwell time function (2.5) has all non-zero parameters, we cannot even write $d_{tk}$ as an explicit function of all $y_{tk}$. Another difficulty is due to the cost characteristics of the deadheading problem. In both objective functions, the costs involve vehicle headways, so that the cost associated with each vehicle always involves its adjacent vehicles. This means that the RTDP is a non-separable problem in general, even though the objective functions are written in summation form.

Although the general RTDP is intractable, it provides important insights into a special case of the RTDP that is tractable and can be solved to optimality. In the following section, we derive from RTDP such a model for System $F$.

### 4.3 Optimal Deadheading in System $F$

System $F$ is a simplest type of transit system (see Section 3.1 for a description). In this section we show that a special case of RTDP in System $F$, called RTDP$^\gamma$, is tractable and can be solved to optimality. This special case requires one more restriction in addition to A1-A8, A3.1, and A3.2:

**A4.1.** No adjacent vehicles will be deadheaded in the same direction.

The justification for assumption A4.1 is that there are, in fact, likely to be few normal operating situations in which two adjacent vehicles should be deadheaded in the same direction. This is because the purpose of deadheading is usually to shorten a large headway, and deadhead segments of different vehicles in the same direction will be partially or fully overlapping. Therefore, if two adjacent vehicles with large headways are both deadheaded, passengers at the skipped stations will have to wait a long time, and this is further exacerbated by their increasing frustration when watching more than one vehicle pass by without stopping. Assumption A4.1 is likely to be violated when a vehicle
with a large preceding headway is tailed by two or more "bunched" vehicles. The effects of such violations will be investigated later in this section.

In the following we develop various formulations for RTDPF and its subproblems, study their properties, present solution algorithms, and evaluate the effectiveness of deadheading in System $F$ using the Green Line data sets.

### 4.3.1 The RTDPF Formulation

We first express variables $a_{t,k}$, $b_{t,k}$, and $P_{t,k}$ in System $F$ in terms of $y_{t,k}$:

(4.3.1) $a_{t,k} = d_{t,k-1} + R_k + (y_{t,k} + y_{t,k-1})\delta, \forall i,k$

(4.3.2) $d_{t,k} = a_{t,k} + y_{t,k}c_0, \forall i,k$

(4.3.3) $P_{t,k} = r(d_{t,k} - d_{t+1,k})(1-y_{t,k}), \forall i,k$

The arrival and departure time definitions (4.3.1) and (4.3.2) follow from equations (3.3) and (2.6) respectively. (4.3.1) says that, if a vehicle is deadheaded and only one of the two stations $k$ or $k-1$ is skipped, the running time of vehicle $i$ between them is $R_k + \delta$. On the other hand, if both stations are skipped, the running time will be $R_k$. Clearly the arrival times defined by (4.3.1) are nonlinear. We note that without assumptions A3.1 and A3.2, interstation stopping behavior of blocked vehicles will need to be considered in the formulation, and $a_{t,k}$ cannot be expressed in such a simple form. A3.1 and A3.2 result in the preceding headway of a vehicle being invariant at all stations when neither the vehicle nor its predecessor is controlled, that is, $h_{i,k}^0 = h_i$. Thus once a minimal safe headway is established at dispatching, it will be maintained throughout the network. This means no interstation stopping would occur in either the no-control or deadheading case, since by definition 4.2 feasible deadheading will not result in interstation stopping. (4.3.2) says that the actual dwell time at a skipped station will be zero. (4.3.3) defines leftover passengers at a station. When a station $k$ is skipped by vehicle $i$, it is not skipped by $i-1$ (assumption A4.1), and therefore there are no leftover passengers from $i-1$ at $k$. Thus at
any control station \( k \), \( y_{i,k} = 0 \) and we have \( P_{i,k} = B_{i,k} = r(d_{i,k} - d_{i-1,k}) \), while at any non-control station, \( y_{i,k} = 1 \) and \( P_{i,k} = 0 \).

With assumption A3.2 and equation (4.3.3) defining \( P_{i,k} \), we can rewrite objective function (4.2.2) as follows:

\[
 f(h(y)) = \sum_{k \in K_c^i} \sum_{i \in I_m} [r(d_{i,k} - d_{i-1,k})^2/2 + r(d_{i,k} - d_{i-1,k})(d_{i-1,k} - d_{i-2,k})(1 - y_{i-1,k})] \\
 + u_c |K_c^i| r(d_{i,k} - d_{i-1,k})^2/2 
\]

where \( K_c^i = K_c - \{k_i \} \). Dividing \( f(h(y)) \) by \( r(>0) \) yields

(4.3.4) \( f(h(y)) = \sum_{k \in K_c^i} \sum_{i \in I_m} [(d_{i,k} - d_{i-1,k})^2/2 + (d_{i,k} - d_{i-1,k})(d_{i-1,k} - d_{i-2,k})(1 - y_{i-1,k})] \\
 + u_c |K_c^i| (d_{i,k} - d_{i-1,k})^2/2 
\)

We now formulate the complete RTDPF:

[RTDPF]

Minimize (4.3.4)

Subject to (4.2.4) - (4.2.8) and

(4.3.5) \( y_{i,k} + y_{i-1,k} \geq 1, \ \forall (i,k) \in I_m \times K_c \)

Compared to [RTDP], [RTDPF] has one additional constraint, (4.3.5), which incorporates assumption A4.1. Since no adjacent vehicles are deadheaded, at least one of the two variables \( y_{i,k} \) and \( y_{i-1,k} \) must be 1. Assumptions A3.1 and A3.2 have been incorporated into the variable definitions (4.3.1) and (4.3.2), and assumption A3.2 into the objective function (4.3.4), as discussed above.

### 4.3.2 Properties of RTDPF

Interestingly, although the program [RTDPF] still looks fairly complicated, it is entirely separable due to the separability property of System F. The following results follow closely from Lemma 3.1 proved in Chapter 3.
Lemma 4.1

Given the definitions of vehicle arrival and departure times in (4.3.1) and (4.3.2) respectively, if the departure time of each vehicle \( i \) at station \( k_0 - 1 \) (i.e., \( d_{i,k_0-1} \)) is fixed, an uncontrolled vehicle’s trajectory is unaffected by controls on any other vehicles.

**Proof** When vehicle \( i \) is uncontrolled, \( y_{ik} = y_{ik-1} = 1 \) for all \( k \in K \). Substituting this into (4.3.1) yield

\[
(4.3.6) \quad a_{ik} = d_{ik-1} + R_k + 2\delta, \quad \forall k.
\]

Now substituting \( y_{ik} = 1 \) and (4.3.6) into (4.3.2) yields \( d_{ik} = a_{ik} + s_{ik} = d_{ik-1} + R_{ik} + 2\delta + c_0 \). Solving this recursive equation given \( d_{i,k_0-1} \), we have

\[
(4.3.7) \quad d_{i,k} = d_{i,k_0-1} + \sum_{j=k_0}^{k} (R_j + 2\delta + c_0) = d_{i,k_0-1} + \sum_{j=k_0}^{k} R_j + (k - k_0 + 1)(2\delta + c_0)
\]

Clearly all three terms are fixed values associated with vehicle \( i \) and station \( k \). Thus each \( d_{i,k} \) is uniquely defined by \( i \) and \( k \) and independent of any other vehicle’s movement. Also, it follows from (4.3.6) that \( a_{ik} \) depends solely on \( d_{i,k-1} \). Hence once \( d_{i,k} \) values are determined each \( a_{ik} \) is also determined. Thus the result follows. **Q.E.D.**

From Lemma 4.1, it is obvious that the optimal solution to [RTDPF] can be obtained by solving all subproblems in which at most one vehicle is deadheaded. We define the subproblem below.

**Definition 4.3**

Given a vehicle \( i \), a **1-DH** problem is to find the optimal deadhead segment for \( i \), such that the total cost across vehicle trips \( i \) and \( i+1 \) are minimized.

The optimal solution to [RTDPF] is completely determined by the set of optima to the 1-DH subproblems because the optimal solution to a 1-DH problem with respect to vehicle \( i \) depends solely on \( i \)'s own trajectory and the trajectories of its two adjacent vehicles \( i-1 \) and \( i+1 \). Once the optimal control policy and optimal trajectory for vehicle \( i \)
are determined, only trajectory changes of the two adjacent vehicles would affect optimality. But because \(i-1\) and \(i+1\) are uncontrolled and are not blocked by their preceding vehicles, it follows from Lemma 4.1 that their trajectories do not change under any deadheading policy on \(i\). Thus each 1-DH problem is independent of all other 1-DH problems, and the optimal 1-DH costs are additive. Therefore, we can examine one vehicle at a time, and the problem of solving [RTDPF] reduces to solving \(m-1\) 1-DH problems.

We now give a result describing the feasible region of a 1-DH problem. Hereafter we will denote the no-control values of variables with the superscript \(0\)’ where a distinction is needed. For example, \(d_{i,k}^0\) will be used to denote the value of \(d_{i,k}\) when \(i\) is not controlled.

**Lemma 4.2**

A segment \([k_0, k]\) is a feasible deadhead segment for vehicle \(i\) in [RTDPF] if and only if the following inequalities hold:

\[
\begin{align*}
(4.3.8) \quad & a_{i,k}^0 - (k_e - k_0)\Delta \geq d_{i-1,k}^0 + h_0 \\
(4.3.9) \quad & k_0 < k_e < k_0 + N/2
\end{align*}
\]

where \(\Delta = c_0 + 2\delta\), \(k_0 = 1\) or \(N/2+1\).

**Proof:** Since (4.3.9) is straightforward, we prove the Lemma by showing that, given variable definitions (4.3.1) and (4.3.2) in [RTDPF], inequality (4.3.8) is equivalent to inequality (4.2.3) in Definition 4.2. First observe that for each skipped station \(k\) except \(k_0\), \(y_{ik} = y_{ik+1} = 0\). At \(k = k_0\), we have \(y_{ik} = 0\) and \(y_{ik+1} = 1\) since \(k_0 - 1\) is not skipped. At station \(k = k_e\), \(y_{ik} = 1\) and \(y_{ik+1} = 0\) by Definition 4.1.

Combining (4.3.1) and (4.3.2) yields

\[
(4.3.10) \quad d_{i,k} = d_{i,k+1} + R_{i,k} + (y_{i,k} + y_{i,k+1})\delta + c_0 y_{i,k}
\]

Solving this recursive equation with given \(d_{i,k_0-1}\) we have

\[
(4.3.10) \quad d_{i,k} = d_{i,k_0-1} + \sum_{j=k_0}^{k} [R_j + (y_{j,k} + y_{j,k-1})\delta + c_0 y_{j,k}]
\]
Now observe that when \( i \) is deadheaded on segment \([k_0, k_e]\), at \( k = k_0 \), \( y_{i,k_0} = 0 \) and \( y_{i,k_1} = 1 \) since \( k_{r-1} \) is not skipped, that at each skipped station \( k: k_0 < k < k_e \), \( y_{i,k} = y_{i,k_1} = 0 \), and that at station \( k = k_e \), \( y_{i,k_e} = 1 \) and \( y_{i,k_1} = 0 \). Substituting all these into (4.3.10) yields

\[
(4.3.10') \quad d_{i,k_r} = d_{i,k_0-1} + (y_{i,k_0-1} + y_{i,k_0})\delta + c_0 y_{i,k_0} + (y_{i,k_r-1} + y_{i,k_r})\delta + c_0 y_{i,k_r} + \sum_{j=k_0}^{k_r} R_j
\]

\[
+ \sum_{j=k_0}^{k_{r-1}} [(y_{j,k} + y_{j,k-1})\delta + c_0 y_{j,k}] = d_{i,k_0-1} + \sum_{j=k_0}^{k_r} R_j + 2\delta + c_0
\]

Compared to \( d_{i,k_r}^0 \) in (4.3.7) when \( i \) is not controlled, a total amount of time \((k_e - k_0)(2\delta + c_0)\) is saved over \([k_0, k_e]\). That is, \( d_{i,k_r} - d_{i,k_r}^0 = (k_e - k_0)(2\delta + c_0) \). Since \( a_{i,k_r} = d_{i,k_r} - c_0 \), \( a_{i,k_r}^0 = a_{i,k_r} - c_0 a_{i,k_r}^0 = a_{i,k_r} = d_{i,k_r}^0 - d_{i,k_r} = (k_e - k_0)(2\delta + c_0) \). Thus, inequality (4.2.3) implies that for segment \([k_0, k_e]\) to be a feasible deadhead segment, we must have

\[
(4.3.11) \quad a_{i,k_r} = a_{i,k_r}^0 - (k_e - k_0)(c_0 + 2\delta) \geq d_{i-1,k_r}^0 + h_0
\]

which is (4.3.8). On the other hand, when \( i \) is deadheaded, at each \( k: k_0 \leq k < k_e \), (4.3.10) yields

\[
(4.3.12) \quad d_{i,k} = a_{i,k} = d_{i,k_0-1} + \delta + \sum_{j=k_0}^{k} R_{i,j}
\]

and because \( a_{i,k}^0 = d_{i,k}^0 - c_0 \) when \( i \) is not controlled, from (4.3.7) we have

\[
(4.3.13) \quad a_{i,k}^0 = d_{i,k_0-1} + \sum_{j=k_0}^{k} R_{i,j} + (k - k_0 + 1)(2\delta + c_0) - c_0
\]

Subtracting (4.3.12) from (4.3.13) yield

\[
(4.3.12') \quad d_{i,k} = a_{i,k} = a_{i,k}^0 - (k-k_0+1)(2\delta + c_0) + \delta + c_0 = a_{i,k}^0 - (k-k_0)(2\delta + c_0) - \delta, \quad k_0 \leq k < k_e
\]

Since \( h_{i,k}^0 \geq h_0 + c_0 \) by the dispatching condition, \( a_{i,k}^0 \geq d_{i,k}^0 \). Thus from (4.3.12') we have \( a_{i,k}^0 - d_{i-1,k}^0 = a_{i,k}^0 - d_{i-1,k}^0 - (k-k_0)(2\delta + c_0) - \delta = h_{i,k}^0 - (k-k_0)(2\delta + c_0) + \delta, \quad k_0 \leq k < k_e \). Since \( h_{i,k}^0, c_0, \delta \) and \( k_0 \) are all constants, \( h_{i,k} \) is monotonically decreasing in \( k: k_0 \leq k < k_e \) (see Fig. 4.3.1).
Furthermore, by (4.3.11)\( a_{i,k_r} - d_{i-1,k_r}^0 = a_{i,k_r}^0 - d_{i-1,k_r}^0 - (k_r - k_0)(c_0 + 2\delta) = h_i^0 - c_0 \)
\( (k_r - k_0)(2\delta + c_0) < a_{i,k_r-1} - d_{i-1,k_r-1}^0 = h_{i-1}^0 - (k_r - k_0)(2\delta + c_0) + \delta \), it follows that

\[ a_{i,k_r} - d_{i-1,k_r}^0 = \min(a_{i,k} - d_{i-1,k}^0, k \in [k_0, k_r]) \]

Hence \( a_{i,k_r} - d_{i-1,k_r}^0 \geq h_0 \) implies \( a_{i,k} - d_{i-1,k}^0 \geq h_0 \) for all \( k: k_1 \leq k \leq k_2 \), so (4.3.8) also implies (4.2.3). This completes the proof. \( Q.E.D \)

![Fig. 4.3.1: Trajectory of a Deadheaded Vehicle](image)

In Lemma 4.2, (4.3.8) says that the minimal safe headway should be maintained at the end of a deadhead segment, and (4.3.9) says that a deadhead segment should not reverse directions. As an example, Fig. 4.3.1 shows the trajectory change of a deadhead vehicle \( i \) at each station. Here the minimal safe headway \( h_0 = 2 \) minutes, the no-control headway \( h_i^0 = 12 \) minutes, \( \delta = 0.25 \) minutes and \( c_0 = 0.5 \) minutes. After \( i \) skipped stations 1 to 7, it has advanced \( 7 \times (2 \times 0.25 + 0.5) = 7 \) minutes, and the new headway at stations 8 to 22 is 5 minutes. The deadhead segment [1,8] is feasible because \( a_{i,8} - d_{i,1,8} = 4.5 > 2 \), and \( 1 < 8 < N/2 = 11 \).
4.3.3 Subproblem 1-DH and Solution Properties

We derive the mathematical formulation of 1-DH from [RTDPF] as follows.

First, observe that, when vehicles $i$, $i-1$ and $i+1$ are not controlled, the cost of a 1-DH problem given by (4.3.4) is

$$
(4.3.14) \quad f_i^0 = \sum_{j=i}^{i+1} \left \{ \sum_{k \in K'_c} (d_{j,k}^0 - d_{j-1,k}^0)^2 + u_c |K'_c| (d_{j,k_i}^0 - d_{j-1,k_i}^0)^2 \right \} / 2
$$

which is a constant. The 1-DH problem is to find an optimal $k_e$ for $i$ that minimizes the 1-DH cost

$$
(4.3.15) \quad f_i(y_i) = \sum_{j=i}^{i+1} \left \{ \sum_{k \in K'_c} (d_{j,k} - d_{j-1,k})^2 + u_c |K'_c| (d_{j,k_i} - d_{j-1,k_i})^2 \right \} / 2 + \sum_{k=k_0}^{k-1} (d_{i,k} - d_{i-1,k})(d_{i+1,k}^0 - d_{i,k})
$$

where the first summation computes waiting time of randomly arrived passengers with the new trajectory of vehicle $i$ (the trajectories of $i-1$ and $i+1$ are unchanged); and the second part is the additional waiting time of leftover passengers at stations skipped by $i$.

What we did here is that we have simply substituted $y_{i,k} = 1 \quad \forall k: k_e \leq k \leq k_i$, $y_{i,k} = 0 \quad \forall k: k_0 \leq k < k_e$, and $y_{i-1,k} = y_{i+1,k} = 1 \quad \forall k \in K'_c$, which incorporates constraints (4.2.5) and (4.3.5), into (4.3.4) and obtained (4.3.15). In this way, we have transformed the decision variables $y_{i,k}$ into a single decision variable $k_e$.

To simplify the objective function (4.3.15), we note that to minimize $f_i(y_i)$ is equivalent to the minimization of $f_i(y_i) - f_i^0$. Then subtracting (4.3.14) from (4.3.15), when $i$ is deadheaded on $[k_0, k_e]$ we have the following new cost function

$$
f_i(k_e) = f_i(y_i) - f_i^0 = \sum_{k=k_0}^{k-1} \left \{ (d_{i,k} - d_{i-1,k})(d_{i+1,k}^0 - d_{i,k}) \right \} + \sum_{k=k_0}^{k-1} \left \{ (d_{i,k} - d_{i-1,k})^2 + (d_{i+1,k}^0 - d_{i,k})^2 - [(d_{i,k}^0 - d_{i-1,k})^2 + (d_{i+1,k}^0 - d_{i,k})^2] / 2 + u_c |K'_c| \{ (d_{i,k_i} - d_{i-1,k_i})^2 + (d_{i+1,k_i}^0 - d_{i,k_i})^2 - [(d_{i,k_i}^0 - d_{i-1,k_i})^2 + (d_{i+1,k_i}^0 - d_{i,k_i})^2] / 2
$$

76
Since the no-control headways are constant across all \( k \in K^* \), we can omit the subscript \( k \) and simply denote them as \( h^0_j = d^0_{j,k} - d^0_{i-1,k} \), \( j = i, i + 1 \). Also, letting \( n = k - k_0 \) (i.e., the number of skipped stations), after deadheading vehicle \( i \) on \([k_0, k_1]\), at every benefit station including \( k_1 \) the headway \( h^0_i \) has decreased by \( \Delta h_i = n(c_0 + 2\delta) = n\Delta \), \( k \geq k_1 \) (refer to the proof of Lemma 4.2). Letting \( h_{i,k} = d_{i,k} - d^0_{i-1,k} \) and \( h_{i+1,k} = d^0_{i+1,k} - d_{i,k} \), we obtain from the above

\[
(4.3.16) f_i(k) = \sum_{k = k_0}^{k_1 - 1} \left\{ h_{i,k}h_{i+1,k} + \left[ h^2_{i,k} + h^2_{i+1,k} - (h^0_{i,k})^2 - (h^0_{i+1,k})^2 \right]/2 \right\} \]

\[
+ \sum_{k = k_0}^{k_1 - 1} \left[ (h^0_{i,k} - \Delta h_i)^2 + (h^0_{i+1,k} + \Delta h_i)^2 - (h^0_{i,k})^2 - (h^0_{i+1,k})^2 \right]/2 \]

\[
+ u_c(K_c') \left[ \left( (h^0_{i,k} - \Delta h_i)^2 + (h^0_{i+1,k} + \Delta h_i)^2 - (h^0_{i,k})^2 - (h^0_{i+1,k})^2 \right)/2 \right] \]

\[
= \sum_{k = k_0}^{k_1 - 1} h^0_{i,k}h^0_{i+1,k} + \sum_{k = k_0}^{k_1 - 1} \left( h^0_{i+1,k} - h^0_{i,k} + \Delta h_i \right)h^0_i \]

\[
= n(h^0_i - h^0_i + \Delta h_i)\]n\Delta

Using this, we have the following formulation of 1-DH:

[1-DH] Minimize

\[
(4.3.17) f_i(n) = n(h^0_i - h^0_i + \Delta h_i)\]n\Delta

Subject to:

\[
(4.3.18) \quad n\Delta \leq h^0_i - c_0 - h_0
\]

\[
(4.3.19) \quad n \leq N/2 + 1 \quad n \in \mathbb{Z}^+
\]

Here (4.3.18) is (4.3.8) with rearranged terms, and with \( a^0_{i,k} - d^0_{i-1,k} \) replaced by \( h^0_i - c_0 \) and \( k_1 - k_0 \) replaced by \( n \). Also, (4.3.19) is (4.3.9) with rearranged terms and with \( k_1 - k_0 \) replaced by \( n \). Constraints (4.2.5) and (4.3.5) in program [RTDPF] have been substituted into the objective function (see the derivation of objective function (4.3.4)).

The interpretation of the objective function (4.3.17) is as follows. When \( n = 0 \), there is no control action and the total passenger waiting time does not change. Thus \( f_i(n) = 0 \). When \( n > 0 \), the cost is composed of two parts. One is the additional waiting time of passengers who arrived at the \( n \) skipped stations during \( h_{i,k} \). This is the first term in (4.3.17). The
other is the change in passenger waiting time at the $|K'_c|-n$ benefit stations including the artificial cost (or weight) at station $k_i$ (the second term in (4.3.17)). Since [1-DH] allows a no-control solution with $n=0$ and $f_i(n)=0$, the optimal value of $f_i(n)$ is always non-positive.

We now give another definition and a number of solution properties of [1-DH].

**Definition 4.4**

A beneficial deadheading policy is a deadheading policy that results in lower cost than in the no-control case. That is, a policy in which $f_i(n) < 0$.

**Lemma 4.3**

A beneficial deadheading policy for [1-DH] exists if and only if the following holds.

\[(4.3.20) \quad (h_i^0 - h_{i+1}^0 - n\Delta)((1 + u_c)|K'_c| - n)\Delta > h_i^0 h_{t+1}^0 \quad n>0\]

where $\Delta = c_0 + 2\delta > 0$.

**Proof:** Deadheading is beneficial if and only if $f_i(n)$ in [1-DH] is negative, which implies $n>0$. Since the first term of $f_i(n)$ in (4.3.17) is always positive, $f_i(n)$ is negative if and only if the second term in the right hand side of (4.3.17) is negative and its absolute value is larger than the first term. That is,

\[-((1 + u_c)|K'_c| - n)(h_{i+1}^0 - h_i^0 + n\Delta)n\Delta = ((1 + u_c)|K'_c| - n)(h_i^0 - h_{i+1}^0 - n\Delta)n\Delta > nh_i^0 h_{t+1}^0\]

Since $n>0$ and $h_i^0, h_{i+1}^0 > 0$, the above inequality is equivalent to (4.3.20). Q.E.D

Although (4.3.20) is a necessary and sufficient condition for a beneficial policy, it is difficult to verify unless the value of $n$ is known first. However, it is easy to see that the left-hand side of (4.3.20) is a monotonically decreasing function of $n$. If a given value of $n$ does not satisfy (4.3.20), neither do all $n$'s which have larger values. This is a useful property for checking whether a 1-DH policy is beneficial.

78
Because (4.3.20) is difficult to verify, we develop the following more useful results.

**Proposition 4.1**

A beneficial deadheading policy for [1-DH] exists only if:

\[(4.3.21)\quad 0 < n < (h_i^0 - h_{i+1}^0)/\Delta\]

and hence, only if:

\[(4.3.22)\quad h_i^0 - h_{i+1}^0 \geq \Delta\]

Furthermore, a beneficial policy is also a feasible policy.

**Proof:** First observe that \(n, \Delta, h_i^0\) and \(h_{i+1}^0\) are all positive (there would be no benefit if \(n=0\)). Also, because \(n \leq |\mathbf{K}_i^+| = N/2 - 1\), and \(0 \leq u_c \leq 1\), \((1+u_c)|\mathbf{K}_i^+| - n\) is positive. Thus (4.3.20) implies \(h_i^0 - h_{i+1}^0 - n\Delta > 0\). Rearranging the terms we have \(n < (h_i^0 - h_{i+1}^0)/\Delta\), which is precisely (4.3.21). And since \(n\Delta\) is positive, (4.3.21) implies (4.3.22). Hence (4.3.21) and (4.3.22) are necessary conditions for a beneficial policy. Furthermore, \(n\Delta < (h_i^0 - h_{i+1}^0) = h_i^0 - (d_{i+1}^0 - d_i^0) = h_i^0 - (a_{i+1}^0 + c_0 - d_i^0) < h_i^0 - (h_0 + c_0), \) since \(a_{i+1}^0 - d_i^0 \geq h_0\) by (4.3.8). Thus, \(h_i^0 - (h_0 + c_0) = d_i^0 - d_{i+1}^0 - c_0 - h_0 = a_i^0 - d_{i+1}^0 - h_0 > n\Delta, \) or \(a_i^0 - n\Delta > h_0 + d_{i+1}^0\). It thus follows from Lemmas 4.2 and 4.3 that deadheading feasibility is attained at all \(n\) stations. Thus a beneficial policy is also a feasible deadheading policy. \(Q.E.D.\)

To simplify the notation, hereafter we will let \(a = h_i^0 - h_{i+1}^0, b = h_i^0 - h_{i+1}^0, \) and \(K = (1 + u_c)|\mathbf{K}_i^+|\). \(a\) must be positive for a beneficial deadheading policy by Proposition 4.1. Thus we will only look at the 1-DH problems where \(a>0\). Using this notation, (4.3.17) is rewritten in a much simpler form as:

\[(4.3.17)\quad f_i(n) = nb - (K-n)(a-n\Delta)n\Delta\]
Proposition 4.2

In an optimal deadheading policy of [1-DH], the number of skipped stations $n$ is bounded from above by $(a+\Delta)/(2\Delta)$. That is

$(4.3.23) \quad n^* \leq (a+\Delta)/(2\Delta).$

where $n^*$ is the optimal solution to [1-DH].

Proof: $(4.3.23)$ holds trivially when $n=0$. We prove this proposition by showing that $n \geq (a+\Delta, \lambda, \Delta)$ implies $f(n) > f(n-1)$ when $n \geq 1$.

$$f(n) - f(n-1) = \left[ nb - (K-n)(a-n\Delta)n\Delta\right] - \left((n-1)b-(K-n+1)(a-(n-1)\Delta)(n-1)\Delta\right)$$

$$= [nb-(n-1)b+(a-(n-1)\Delta)(n-1)\Delta]-(K-n)((a-n\Delta)n\Delta-(a-(n-1)\Delta)(n-1)\Delta]$$

It is obvious that the first term $[nb-(n-1)b+(a-(n-1)\Delta)(n-1)\Delta]>0$ since both a no-control (i.e. $n=0$) and a beneficial solution satisfy $a>(n-1)\Delta$ by Proposition 4.1. So all we have to do is to show that the second term is also positive. Or that $n > (a+1)/(2\Delta)$ implies

$$-(a-n\Delta)n\Delta-(a-(n-1)\Delta)(n-1)\Delta > 0,$$

or equivalently, $(a-n\Delta)n\Delta-(a-(n-1)\Delta)(n-1)\Delta < 0$.

$$(a-n\Delta)n\Delta-(a-(n-1)\Delta)(n-1)\Delta$$

$$= (a-n\Delta)n\Delta-(a-(n-1)\Delta)n\Delta+(a-(n-1)\Delta)\Delta$$

$$= [a-n\Delta-(a-(n-1)\Delta)]n\Delta+a\Delta-(n-1)\Delta^2$$

$$= -n\Delta^2+a\Delta-(n-1)\Delta^2$$

$$= a\Delta-(2n-1)\Delta^2$$

$$= [a-(2n-1)\Delta]\Delta$$

Since $\Delta>0$ and $n \geq 1$, it follows trivially that $n > (a+\Delta)/(2\Delta) \iff [a-(2n-1)\Delta]\Delta < 0 \implies f(n)-f(n-1) > 0$. This completes the proof. Q.E.D.

The intuitive explanation of this proposition is that there exists an ideal headway $h^*$ at station $k=k_e$, once $n$ reaches the point such that $h_{ik}<h^*$ (and hence $h_{i,k+1}>h^*$), the cost at the beneficial stations will start to increase due to the unbalanced headways. Since at the
skipped stations the cost is monotonically increasing with \( n \), the total cost along the whole direction is increasing. Such a point is when \( n = (a+\Delta)/(2\Delta) \). However, this point is not necessarily the lowest point on the cost curve, because total cost may have already started increasing before this point, due to the increase of \( nb \), the cost at the skipped stations. Thus \((a+\Delta)/(2\Delta)\) is only an upper bound, although it may be a tight upper bound.

Obviously (4.3.21) is a tighter constraint than (4.3.18), and (4.3.23) is even tighter. Since (4.3.23) is an upper bound for an optimal policy in [1-DH], replacing (4.3.18) with (4.3.23) in [1-DH] does not change the optimal solutions. Furthermore, since \( n=0, f_i(n)=0 \) is a feasible solution of [1-DH], the optimal solution of [1-DH] will be either a beneficial deadheading policy or a no-control policy.

In order to analyze the solution properties of [1-DH] further, we examine its real relaxation [1-DH'], in which \( n \in \mathbb{Z}_- \) is replaced by \( n \in \mathbb{R}_- \), and the constraints are tightened up using the above results. That is,

[1-DH'] Minimize (4.3.17')
Subject to (4.3.19) and (4.3.23), \( n \in \mathbb{R}_- \)

We next present a result on the property of optimal solution to [1-DH'], and then show that the global optimum to [1-DH'] also leads to a global optima to [1-DH].

*Proposition 4.3*

[1-DH'] is a strictly convex program if \( n \leq (1+u_i)\|K'_i\|/2 \) for all beneficial solutions of \( n \).

*Proof:* We prove this proposition by showing that \( n \leq (1+u_i)\|K'_i\|/2 \) is a sufficient condition for [1-DH'] to be a strictly convex program. It is easy to see that the constraint set of [1-DH'] is a convex set because it is a polyhedral set. So, all that remains is to show that the objective function in [1-DH'] is strictly convex over the constraint set when \( n \leq (1+u_i)\|K'_i\|/2 \).

Expanding (4.3.17) yields
(4.3.24) \( f(n) = -\Delta^2 n^3 + [a\Delta + K\Delta^3]n^2 + [b-a\Delta K]n \)

The first order derivative of (4.3.24) is

(4.3.25) \( \frac{df}{dn} = -3\Delta^2 n^2 + 2[a\Delta + K\Delta^3]n + b-a\Delta K \)

and the second order derivative is

(4.3.26) \( \frac{d^2f}{dn^2} = -6\Delta n + 2[a\Delta + K\Delta^3] = 2\Delta(a\Delta + 3K - 3\Delta n) = 2\Delta((a - \Delta n) + (K - 2n)\Delta) \)

Now \( f \) is strictly convex if and only if (4.3.26) is positive. Obviously when \( n=0 \), (4.3.26) is positive. Since \( a > \Delta n \) and \( \Delta, b, K > 0 \) from the constraints, it follows that when \( 0 \leq n \leq K/2 = (1+u_c)|K'|/2 \), (4.3.26) is positive. This completes the proof. \textit{Q.E.D.}

\textit{Corollary 4.1}

\([1-DH']\) is a convex program if and only if all of its feasible solutions satisfy

(4.3.27) \( n \leq (a/\Delta + (1+u_c)|K'|)/3 \)

The proof of this corollary follows trivially from equation (4.3.26).

It is easy to obtain an integer global optimal solution to \([1-DH]\) from its real relaxation \([1-DH']\). This is trivial when the real-valued global optimum \( n^* = 0 \). When \( n^* > 0 \), because the objective function of \([1-DH']\) is convex over the feasible region, and is a function of one variable, an integer global optimum must be either the round-up or the round-down, within the feasible region, of the real valued optimal solution. To see this, let \( \Omega \) denote the feasible region of \([1-DH]\) and let \([n^*]\) denote the integer part of \( n^* \). Then the set of possible integer solutions is \( \{[n^*], [n^*]+1\} \cap \Omega \). Note that \([n^*]\) is always an element of \( \Omega \), while \([n^*]+1\) may be or may not be. Since this set contains at least one and at most two elements, an integer global optimal solution to \([1-DH]\) is completely determined by and easily obtained from a real valued optimal solution to \([1-DH]\).
The sufficient condition \( n \leq (a/\Delta+K)/3 \) and the upper bound of \( n \leq (a+\Delta)/(2\Delta) \) implies that \([1-DH']\) is always a convex program when \( (a+\Delta)/(2\Delta) \leq (a/\Delta+K)/3 \Rightarrow a \leq (2K-3)\Delta \). This is not a restrictive condition in reality. To see this, let us look at two extreme cases: \( u_i=1 \) and \( u_i=0 \). When \( K = 2|K'| = 2(N/2-1) \) due to \( u_i=1 \), \([1-DH']\) is certainly convex by Proposition 4.3, because all feasible \( n \leq N/2-1 \), and hence \( n \leq K/2 \). When \( u_i=0 \), \([1-DH']\) is convex if \( a \leq (2K-3)\Delta \). This is easily satisfied in reality because \( a \), the difference between the preceding and following headways of a vehicle, is rarely larger than twice the scheduled headway when there is no major disruption in service. For example, if \( K = 25 \) stations (as in the Green Line B line), and \( \Delta = 0.65 \) minutes, the condition \( a/\Delta \leq 2K-3 \) is violated only when \( a > 30.55 \) minutes! Such a situation hardly ever occurs in a service with scheduled headway of only 5 minutes. Therefore \([1-DH']\) will almost always be convex.

Also note that although a beneficial deadheading policy, if it exists, is always feasible policy, a feasible policy may not be beneficial. This is because the necessary and sufficient condition (4.3.20) for a beneficial policy is not necessary for either feasible or optimal solutions to \([1-DH']\). On the other hand, an optimal solution to \([1-DH]\) must be either a beneficial policy or a no-control policy. Thus, if there exists a beneficial policy, then the optimal solution is obviously also beneficial.

The weight \( u_i \) favors longer term optimization when larger than zero. When there is no layover time at \( k_i \), the benefits of deadheading may be underestimated if \( u_i \) is too small, and we may lose a chance to improve service quality. This can be illustrated using the data in Example 4.3.1 in section 4.3.7, where \( k_0 = 1 \) and \( k_i = N/2 \), and \( u_i = (N/2-1)r \). The optimal solution is to skip 2 stations. If we set \( u_i = 0 \), the optimal policy from \([1-DH]\) is to skip 1 station, then at \( k_0 = N/2+1 \) the optimal policy from \([1-DH]\) with \( k_i = N \) is no control. In this case the total passenger waiting time saved over the round trip of \( i \) is \( 0.8r \), which is 24.4% less than the benefit from skipping 2 stations.
4.3.4 Exact Solution Algorithm for RTDPF

Set Optimal Solution

In the above we have studied solution properties of a 1-DH subproblem of RTDPF, which considers a one-vehicle deadheading problem in which the adjacent vehicles are not controlled. If the rolling horizon is chosen as \( m=1 \), then we are done for that \( I_m \) set after we solve the 1-DH! problem. When \( m>1 \), however, we may find that adjacent vehicles have beneficial deadhead cost after solving each 1-DH problem for the \( m \) vehicles in set \( I_m \). In this case it still must be determined which vehicles are to be deadheaded such that the total cost over the \( m \) vehicle trips is minimized and no adjacent vehicles are deadheaded. This problem is trivial when \( m \) is small (e.g., \( m \leq 3 \)). In the following we discuss a general case when \( m \) is large, which may be mainly of theoretical interest.

Definition 4.5

If a subset of \( v \leq m \) vehicles with contiguous indices, i.e., \( \{j, j+1, \ldots, j+v-1\} \subset I_m \), all have beneficial 1-DH costs, while \( j-1 \) and \( j+v \) do not, such a subset is called a \( v \)-opt subset.

With this definition, the largest possible \( v \)-opt subset has size \( m \), and the smallest possible \( v \)-opt subset has size 0. A single vehicle with beneficial 1-DH cost is then regarded as a 1-opt subset if its preceding and following vehicles both have zero optimal 1-DH costs. In reality, there are few cases where \( v>2 \). This is because a beneficial deadheading policy usually requires a vehicle's preceding headway to be larger than its following headway, and such a headway pattern rarely occurs repeatedly among more than three adjacent vehicles. If the largest \( v \)-opt subset in a RTDPF has size 2, the problem of finding an optimal set of deadhead vehicles for \( I_n \) is simple. We just choose from each of the 2-opt subsets the vehicle that has the lower 1-DH cost, and include them and all 1-opt subsets in the set optimal solution. See Example 4.3.4 for details.

When \( v>3 \), the problem of finding optimal deadhead set becomes more complicated, and though it will not often occur, it is helpful to have a discussion of the general problem. When \( m \) is large, there can be many feasible combinations of optimal 1-DH solutions.
such that no adjacent deadhead vehicles are in the optimal solution set. For example, with a total of \( m=10 \) vehicles, in which the vehicle trips 1 and 11 are not controlled, the feasible combinations include \{3,5,7,9\}, \{2,4,6,8,10\}, \{3,5,8,10\}, \{2,4,7,9\}...etc. In fact, any ordered pair of vehicles \( \{i,j; j>i+1\} \) can be included in the same feasible solution. Letting \( v_{j} \) denote the \( j \)th member of a feasible set of deadhead vehicles for set solution, and defining \( v_{0} \equiv 1 \), a feasible set can then be expressed as \( V = \{v_{j}; v_{j}>v_{j-1}+1, 1<v_{j}<m\} \). We will call any two feasible sets of deadhead vehicles that have exactly the same total cost *equivalent feasible sets*. A feasible set of deadhead vehicles, denoted by \( V \), is called a *maximal feasible set*, denoted as \( V_{m} \), if the following two conditions are satisfied: (a) The index difference between any two neighboring members in \( V \) is no less than 2 and no more than 3, that is,

\[(4.3.28) \; v_{j} \geq v_{j-1} + 2 \quad \text{and} \quad v_{j} \leq v_{j-1} + 3; \]

and (b) No vehicle \( i: i \in I, i \notin V \) can be inserted into \( V \) without violating condition (4.3.28).

An optimal set of deadhead vehicles for RTDPF must be an equivalent maximal feasible set. This is because the optimal 1-DH cost is always non-positive, and 1-DH costs between non-adjacent vehicles are additive. Thus, if two neighboring members of a feasible set have indices that differ by 4 or more, we can always insert a new member in between which will form a new feasible set that has smaller or equal total cost. For example, for feasible set \{3,8,10\}, we can insert 5 to form another feasible set \{3,5,8,10\}, and the total cost of the latter is no larger than the former. The latter is, in fact, a maximal feasible set for a set of 10 vehicles.

To represent the problem of finding an optimal set of deadhead vehicles in RTDPF, we construct a directed, non-cyclic network. This network contains \( m+2 \) nodes in total, which represent the \( m \) vehicle trips from 2 to \( m+1 \) that are considered for control and the 2 boundary vehicle trips 1 and \( m+2 \). Each link in this network represents a pair of non-adjacent vehicles whose indices differ by no more than three. There are a total of \( 2(m+1)-3 \) such links. This construction ensures that between any two non-adjacent
vehicles there is a directed path, and a path from node 1 to \( m+2 \) represents a maximum feasible set of deadhead vehicles. We denote the optimal 1-DH cost for each vehicle \( i: 2 \leq i \leq m+1 \) as \( w_i \), and set the cost on each link \((i,j: i>1)\) to \( w_i \). The two links that start from node 1 have zero costs. Fig. 4.3.2 depicts such a network for a set of \( m=9 \) vehicles, indexed from 2 to 10, and two boundary vehicles, 1 and 11. Thus there are 11 nodes and \( 2 \times 10 - 3 = 17 \) links in this network.

![Network Diagram](image)

**Fig. 4.3.2.** Network Representing the Problem of Finding Optimal Set of Deadhead Vehicles

With a network so constructed, the set of nodes on any path from node 1 to \( m+2 \) represents the union of a maximal feasible set with \( \{1\} \) and \( \{m+2\} \). Thus the problem of finding an optimal set of deadhead vehicles reduces to the problem of finding a shortest path from node 1 to \( m+2 \). Such a problem can be easily solved using the label-correcting algorithm (see Dial et al, 1977). Note that the network can also be constructed by first dropping out the vehicles with \( w_i = 0 \); but it has little advantage in practice since a transit network is usually not very large, and it is more convenient to have all \( i \) in the network when representing it in a digital computer.\(^5\)

\(^5\) For example, a simple data structure to represent such a network involves two arrays, one has all vehicles with odd indices as its elements, call it ODD\([j]\); and the other has all vehicles with even indices, call it EVEN\([j]\). Then a link in the network is either two neighboring elements in each array, or (ODD\([j]\), EVEN\([j+2]\]), or (EVEN\([j]\), ODD\([j+1]\)).
We now present the exact algorithm that solves the RTDPF when adjacent deadheading is not allowed, regardless whether a 1-DH subproblem is convex. Without lose of generality, we assume the vehicles are indexed from 1 to \( m+2 \), with 1 and \( m+2 \) as boundary vehicles.

**Algorithm Non-Adjacent Deadheading**

**Step 1:** Solve the \( m \) 1-DH problems for each vehicle \( i \) with \( 2 \leq i \leq m+1 \), using algorithm 1-DH below, and obtain an optimal 1-DH cost \( w_i \) for each \( i \).

**Step 2:** Find the largest \( v \)-opt subset, denoting its size \( \max(v) \). If \( \max(v) > 2 \), go to Step 3. Otherwise, the optimal set of deadhead vehicles is composed of all 1-opt subsets and the vehicles with lower \( w_i \) from each 2-opt subsets.

**Step 3:** If \( \max(v) > 2 \), construct the network of feasible deadhead vehicle sets as indicated in Fig. 4.3.2, with \( m+2 \) nodes and \( 2(m+1)-3 \) links. Set the link cost for each link \((i,j)\) to \( w_i \) if \( 1 < i < m+2 \). Set link cost to zero if \( i=1 \). Use the label-correcting algorithm to find a shortest path from node 1 to node \( m+2 \). The optimal set of deadhead vehicles are the nodes in the resulting shortest path excluding 1, \( m+2 \), and the vehicles with \( w_i=0 \).

**END**

**Algorithm 1-DH**

**Stage 1:** Preprocessing: Set all \( y_{ik} = 1 \) and compute \( d_{ik}^o \) and \( h_i^o \) for all vehicles at all stations \( k \in Kc \) with given \( d_{i,ki-1} \), using equations (4.3.1) and (4.3.2).

**Stage 2:** For each vehicle \( i : 2 \leq i \leq m+1 \) DO:

- **Step 1:** Check beneficial condition: If \( h_{i+1}^o \geq h_i^o \), go to vehicle \( i+1 \); Otherwise, set \( a = h_{i+1}^o - h_i^o \), \( b = h_{i+1}^o h_i^o \) and \( \Delta = c_0 + 2 \delta \) with given \( c_0 \) and \( \delta \). Find \( \max(n) \) that satisfies \( n \leq \min(N/2-1, (a+\Delta)/(2\Delta)) \). The feasible deadhead segment of \( i \) is \([k_0, k_0+\max(n)]\).

- **Step 2:** Check convexity: If \( \max(n) \leq (a/\Delta+(1+u_i)|K'c|)/3 \), go to Step 3, otherwise go to Step 4.

- **Step 3:** Compute the root of equation (4.3.25), and obtain the real valued optimal solution \( n \). Let \([n]\) denote the integer part of \( n \). If \([n]+1 > \max(n)\), set \( n' = [n] \).
Otherwise, set \( n' = \arg(\min(f_i([n]), f_i([n]+1])) \). If \( f_i(n') < 0 \), the optimal solution to the 1-DH problem of \( i \) is \( n^* = n' \), and the optimal cost for \( i \) is \( w_i = f_i(n^*) \).

Otherwise, \( n^* = 0 \) and \( w_i = 0 \).

**Step 4:** Enumerate \( n \) from 0 to \([\max(n)]\) and obtain \( n^* = \arg(\min f_i(n)) \).

**END Stage 2**

**END 1-DH**

### 4.3.5 Numerical Examples

In this section we present two numerical examples. One is for 1-DH, the other a complete example of RTDPF using algorithm Non-Adjacent Deadheading.

#### Example 4.3.1 1-DH

Suppose \( h_i = 9 \), \( h_{i+1} = 3 \), \( \Delta = c_0 + 2\delta = 1 \), \( k_0 = 1 \), \( k = N/2 = 11 \), no layover time at \( N/2 \) and \( u_c = lK'_clr \), we have \( |K_c| = 11 \), \( |K'_c| = 10 \), \( u_c = 1 \), \( (1 + u_c)|K'_c| = 20 \), \( b = h_{i+1}h_0 = 27 \) and \( a = h_i - h_{i+1} = 9-3 = 6 > \Delta = 1 \). Since \( a/\Delta = 6 < (1 + u_c)|K'_c|/2 = 10 \), it follows from Proposition 4.3 that this problem is convex. We have \( (a+\Delta)/(2\Delta) = (6+1)/2 = 3.5 \) and the feasible region of \( n \) is \( 0 \leq n \leq 3.5 \) in \([1-DH']\), and we know \( n^* \leq 3 \) by Proposition 4.2. The optimal solution can be easily obtained by enumerating \( n \) from 0 to 3. For illustrative purposes, however, we solve it analytically. Substituting the data into (4.3.24) and (4.3.25) we have

\[
(4.3.29) \quad f_i(n) = -n^3 + 26n^2 - 93n
\]

\[
(4.3.30) \quad df/dn = -3n^2 + 52n - 93
\]

Note that the constraint \( n \leq N/2-1=10 \) is not binding for \( n^* \) because \( n^* \leq 3 \). If there exists a beneficial policy, then the constraint \( 0 \leq n \leq 3.5 \) is also not binding for \( n^* \). We know from the Karush-Kuhn-Tucker (KKT) conditions that in this case all Lagrange multipliers associated with the constraints are zero. Hence, a beneficial optimal solution must be the root of (4.3.30), denoted by \( n' \), rounded-up or rounded-down within the feasible region. The optimal solution to \([1-DH']\) must be either 0 or \( n' \). It turns out that the root of \( df/dn = 0 \) is \( n = 2.025 < 4.65 \). Since \( f_i(2) = -90 < f_i(3) = -72 \), \( n' = 2 \). And since \( f_i(n') = -90 < 0 \), \( n^* = n' = 2 \) is the integer global optimal solution to \([1-DH]\). In fact, at the
benefit stations (from 3 to 22), the total waiting time reduction is \( n\Delta(a-n\Delta)(N-2-n) \)
\[=2*4*18r =144r \] (see equation (4.3.17)), while at the skipped stations 1 and 2, the total additional waiting time is \( 2*27r=54r \). The net benefit is \( 90r \). The new preceding headway of \( i \) after deadheading becomes 7 minutes.

Fig. 4.3.3 depicts the trajectories of vehicles \( i, i-1, \) and \( i+1 \) before and after \( i \) skips stations 1 and 2, where \( i' \) denotes \( i \) after deadheading.

![Diagram](image)

**Fig. 4.3.3. Example 4.3.1: 1-DH**

**Example 4.3.2 Set Optimization of RTDPF**

In this example we consider a set of \( m=12 \) vehicle trips for control. The input data is given as follows: \( c_0 = 0.4, \delta = 0.15 \) (minutes), \( \Delta = c_0 + 2\delta = 0.7 \), \( k_i = 52 \), \( k_0 = 1 \), \( u_i = 0 \) and \( (1+u_i)\|K_i\|^\| = 51 \). The initial departure time (in minutes) of each vehicle at station 0 (\( N \), \( d_{i,0} \), is given below:

89
Again, vehicles 1 and 14 are the boundary vehicles. Applying algorithm RTDPF in section 4.6 with objective function (4.3.24), the trajectories of all 14 vehicle trips when no control action is taken are first computed, as depicted in Fig. 4.3.4. For each vehicle $i$: $2 \leq i \leq 13$, we first check to see if its preceding headway is larger than its following headway by; and if it is, we proceed to solve the 1-DH problem as in the previous examples. It turns out that vehicles 2, 6, 7, 9, 12 met the necessary headway condition (4.3.22), and all of them have negative optimal 1-DH costs. The optimal 1-DH solutions for these vehicles are listed in Table 4.3.1 below:

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>6</th>
<th>7</th>
<th>9</th>
<th>12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n^*$</td>
<td>7</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>1</td>
</tr>
<tr>
<td>$w_i$</td>
<td>-1,374.31</td>
<td>-526.18</td>
<td>-120.36</td>
<td>-977.18</td>
<td>-13.82</td>
</tr>
</tbody>
</table>

Table 4.3.1 Optimal Solution to 1-DH subproblems in Example 4.3.2

Among the optimal 1-DH solutions, there are three 1-opt subsets, namely \{2\}, \{9\}, \{12\}, and one 2-opt subset, \{6,7\}. Since $v \leq 2$, we do not need the shortest path algorithm. By simply choosing vehicle 6, which has lower $w_i$ than vehicle 7, in the 2-opt subset, and including all 1-opt subsets, we obtain \{2,6,9,12\} as the optimal set of deadhead vehicles. The optimal total cost is 26131.01r. Compared to the no-control total cost 29022.5r, a total of 2891.49r passenger minutes have been saved, or about 10%. Fig. 4.3.5 depicts the trajectories of all 14 vehicle trips after vehicles 2, 6, 9, and 12 are deadheaded on their optimal deadhead segments.
Fig. 4.3.4. Example 4.3.2. Vehicle Trajectories before Deadheading

Fig. 4.3.5. Example 3.3.2. Vehicle Trajectories after Deadheading

4.3.6 Effects of Adjacent Deadheading

In the RTDPF model discussed above we have incorporated assumption A4.1, which precludes adjacent vehicle deadheading (i.e., two or more adjacent vehicles being deadheaded). While A4.1 is a reasonable assumption when any two consecutive vehicle trips both have large headways, it may rule out a better control policy which allows adjacent deadheading in cases where a vehicle with large headway is followed by two or more vehicles with short headways. On the other hand, when adjacent deadheading is allowed, the problem is no longer separable, as the deadheading policy on a vehicle will depend on the controls of its adjacent vehicles. In this case the optimal control policies will also depend on rolling horizon size. To illustrate this, we again use the input data in Example 4.3.2 to compute deadheading policies with rolling horizon $m=1,2,$ and $3$ respectively. For each rolling horizon size $m$, the optimal solution is obtained from completely enumerating all combinations of $\{n_i, n_{i+1}, \ldots n_{i+m-1}\}$ for each vehicle set $I_m = \{i, i+1, \ldots, i+m\}$, $i=2,3,\ldots,13$. The results are listed in Table 4.3.2.

<table>
<thead>
<tr>
<th>$i$</th>
<th>2</th>
<th>3</th>
<th>6</th>
<th>7</th>
<th>9</th>
<th>10</th>
<th>12</th>
<th>total benefit</th>
<th>cost reduction</th>
<th>computation time (sec.)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>2891.50$r$</td>
<td>9.96%</td>
<td>2</td>
</tr>
<tr>
<td>$n^*$</td>
<td>7</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>no adjacent deadheading</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$n^*, m=1$</td>
<td>7</td>
<td>1</td>
<td>4</td>
<td>3</td>
<td>6</td>
<td>3</td>
<td>1</td>
<td>3502.31$r$</td>
<td>12.07%</td>
<td>3</td>
</tr>
<tr>
<td>$n^*, m=2$</td>
<td>8</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>3803.02$r$</td>
<td>13.11%</td>
<td>42</td>
</tr>
<tr>
<td>$n^*, m=3$</td>
<td>8</td>
<td>2</td>
<td>7</td>
<td>5</td>
<td>8</td>
<td>3</td>
<td>1</td>
<td>3803.02$r$</td>
<td>13.11%</td>
<td>135</td>
</tr>
</tbody>
</table>

Table 4.3.2 Results of Example 4.3.2 with Adjacent Deadheading

In Table 4.3.2 we see three pairs of adjacent deadhead vehicles: $\{2,3\}$, $\{6,7\}$, and $\{9,10\}$ when A4.1 is relaxed. There is no case where more than two consecutive vehicles are deadheaded. It turns out in this example the optimal solution for different rolling horizon size does not change between $m=2$ and $m=3$. Compared with the results when A4.1 is in effect, we see that the total benefits are increased by allowing adjacent deadheading. When $m=1$, the impact on total cost is about 2%, and is about 3% for $m>1$. While the complexity and computational burden of the problem increase quickly as $m$ increases, the
marginal benefits are quite small. In all the consecutively deadheaded pairs, the first vehicle skips more stations when \( m > 1 \) than when \( m = 1 \). This shows the impact of deadheading the second vehicle on the control policy for the first in such a pair. On the other hand, if the first vehicle is not deadheaded, the second vehicle also skips fewer, if any, stations.

Table 4.4.3 lists the headway pattern of the three pairs of consecutively deadheaded vehicles. In all the cases the headways follow a large-short-short pattern.

<table>
<thead>
<tr>
<th>( i, i+1 )</th>
<th>2.3</th>
<th>6.7</th>
<th>9.10</th>
</tr>
</thead>
<tbody>
<tr>
<td>( h_i, h_{i+1}, h_{i+2} )</td>
<td>18.87, 4.70, 5.42</td>
<td>13.30, 4.63, 1.30</td>
<td>13.72, 3.03, 2.50</td>
</tr>
<tr>
<td>opt. { ( n_i, n_{i+1} ) }</td>
<td>8.2</td>
<td>7.5</td>
<td>8.3</td>
</tr>
<tr>
<td>benefit</td>
<td>1,445.50</td>
<td>1,085.54</td>
<td>1,258.96</td>
</tr>
</tbody>
</table>

Table 4.3.3 Headway Pattern of Adjacent Deadheading Pairs in Example 4.3.2

Since allowing adjacent deadheading shows more significant improvement with the above headway pattern, we next consider relaxing assumption A4.1 in the RTDPF model. When assumption A4.1 is relaxed, even with rolling horizon \( m = 1 \) the general 1-DH problem, for which any number of consecutive vehicles may be deadheaded, becomes very complicated. This is because it now requires consideration of headway change combinations that result from previous adjacent deadheading policies. Even more important there will be leftover passengers at each station skipped as a result of previous deadheading, and the number of passenger left is a complicated function of the previous deadheading pattern. In addition to these complexities, deadheading of more than two consecutive vehicles is unlikely to be desirable in practice.

When at most two consecutive deadhead vehicles are allowed, however, the problem is tractable with \( m = 1 \). If the preceding vehicle was not controlled, we just need to solve the 1-DH subproblem presented earlier. When the preceding vehicle is deadheaded, the general 1-DH subproblem reduces to one of the subproblems below. Let \( n_i \) and \( n_{i-1} \) denote the number of skipped stations by vehicle \( i \) and \( i-1 \) respectively. Given \( n^{*}_{i-1} \), there are
only two possibilities for \( n_i \): either \( n_i > n_{*i-1} \), or \( n_i \leq n_{*i-1} \). In light of the 1-DH models presented in subsection 4.3.3, the two subproblems are formulated as follows:

[1-DHa] Minimize

\[
(4.3.31) \quad f(n_i) = n_i \beta - (K-n_i)(a' - n_i \Delta) n_i \Delta + n_i \ h_{i-1}^0 \ h_{i+1}^0 - (n_{*i-1} - n_i)(h_{i-1}^0 - n_{*i-1} \Delta)n_i \Delta
\]

Subject to

\[
(4.3.32) \quad n_i \leq n_{*i-1}
\]

where \( n_{*i-1} \geq 0, n_{*i-2} = 0, a' = a + n_{*i-1} \Delta. \)

[1-DHb] Minimize

\[
(4.3.33) \quad f(n_i) = n_i \beta - (K-n_i)(a' - n_i \Delta) n_i \Delta + n_{*i-1} \ h_{i-1}^0 \ h_{i+1}^0
\]

Subject to (4.3.19), (4.3.23) and

\[
(4.3.34) \quad n_i > n_{*i-1}
\]

where \( n_{*i-1} \geq 0, n_{*i-2} = 0, a' = a + n_{*i-1} \Delta. \)

In both programs \( n_{*i-1} \) is known and positive, and \( i-2 \) is not controlled. Let us first look at program [1-DHb] where \( n_i > n_{*i-1} \). In this case all leftover passengers from \( i-1 \) would have to wait for vehicle \( i+1 \), and this results in the additional cost \( n_{*i-1} \ h_{i-1}^0 \ h_{i+1}^0 \). But since this term is a constant, the objective function reduces to (4.3.17), with \( a \) replaced by \( a' \) due to the headway change of \( i-1 \). Hence all necessary conditions for [1-DH'] also apply to [1-DHb]. If the optimal solution to [1-DH] satisfies constraint (4.3.34), it is also the optimal solution to [1-DHb]. Otherwise there is no feasible solution to [1-DHb].

In program [1-DHa], since \( n_i \leq n_{*i-1} \), all passengers left behind by \( i-1 \) (amount of \( h_{i-1}^0 \) at a skipped station of \( i-1 \)) at the \( n_i \) skipped stations of \( i \) have to wait an additional time \( h_{i+1}^0 \), and this results in the term \( n_i \ h_{i-1}^0 \ h_{i+1}^0 \). But for the same reason, some leftover passengers from \( i-1 \) will also benefit from the deadheading of \( i \) by waiting \( n_i \Delta \) shorter time, and this results in one more additional term \(-(n_{*i-1} - n_i)(h_{i-1}^0 - n_{*i-1} \Delta)n_i \Delta\). Here \( n_{*i-1} - n_i \) is the additional number of skipped stations by \( i-1 \) over \( i \), and at each station between \( k_0 + n_i \) and \( k_0 + n_{*i-1} \) the number of benefitted leftover passengers from \( i-1 \) is \( h_{i-1}^0 - n_{*i-1} \Delta \) instead of \( h_{i-1}^0 \).
because the other \( n^*_{i-1} \Delta \) passengers have been counted by \( a' \) in the second term of (4.3.31). The more formal derivation of the objective function (4.3.31) is given as follows.

After vehicle \( i-1 \) skipped \( n^*_{i-1} \) stations, if \( i \) were not controlled we have the cost over the skipped segment of \( i-1 \) as:

\[
(4.3.35a) \quad f_1^0 = n^*_{i-1} \left[ (h^0_{i-1} + h^0_i)^2 + (h^0_{i+1})^2 \right]/2
\]

and the cost for the station \( k_0 + n^*_{i-1} \) and beyond:

\[
(4.3.35b) \quad f_2^0 = (K - n^*_{i-1}) \left[ (h^0_i + n^*_{i-1} \Delta)^2 + (h^0_{i+1})^2 \right]/2
\]

Now suppose vehicle \( i \) skips \( n_i \leq n^*_{i-1} \) stations. Over the skipped segment of \( i \), the cost is

\[
(4.3.36a) \quad f_1 = n_i (h^0_{i-1} + h^0_i + h^0_{i+1})^2/2
\]

From station \( k_0 + n_i \) to \( k_0 + n^*_{i-1} \) the cost is

\[
(4.3.36b) \quad f_2 = (n^*_{i-1} - n_i) \left[ (h^0_{i-1} + h^0_i - n_i \Delta)^2 + (h^0_{i+1} + n_i \Delta)^2 \right]/2
\]

For station \( k_0 + n^*_{i-1} \) and beyond the cost is

\[
(4.3.36c) \quad f_3 = (K - n^*_{i-1}) \left[ (h^0_i + n^*_{i-1} \Delta - n_i \Delta)^2 + (h^0_{i+1} + n_i \Delta)^2 \right]/2
\]

Now the difference in cost with and without deadheading \( i \) is

\[
(4.3.37) \quad f_i(n_i) = (f_1 + f_2 + f_3) - (f_1^0 + f_2^0) \\
= n_i[(h^0_{i-1} + h^0_i + h^0_{i+1})^2 - (h^0_{i-1} + h^0_i)^2 - (h^0_{i+1})^2]/2 \quad \text{(the skipped segment)} \\
+ (n^*_{i-1} - n_i)[(h^0_{i-1} + h^0_i - n_i \Delta)^2 + (h^0_{i+1} + n_i \Delta)^2 - (h^0_{i-1} + h^0_i)^2 - (h^0_{i+1})^2]/2 \\
+ (K - n^*_{i-1})[(h^0_i + n^*_{i-1} \Delta - n_i \Delta)^2 + (h^0_{i+1} + n_i \Delta)^2 - (h^0_{i-1} + n^*_{i-1} \Delta)^2 - (h^0_{i+1})^2]/2 \\
= n_i(h^0_{i-1} + h^0_i + h^0_{i+1}) + (n^*_{i-1} - n_i)[(-h^0_{i-1} - h^0_i + h^0_{i+1})n_i \Delta + (n_i \Delta)^2] \\
+ (K - n^*_{i-1})[(-h^0_{i-1} - n^*_{i-1} \Delta + h^0_{i+1})n_i \Delta + (n_i \Delta)^2]
\]
\begin{align*}
&= n h^0_{i-1} h^0_{i+1} + n h^0_{i+1} h^0_{i+1} + (K - n^*_{i+1} + n^*_{i-1} - n_i) [(-h^0_{i+1} + h^0_{i+1}) n_i \Delta + (n_i \Delta)^2] \\
&\quad - (n^*_{i-1} - n_i) h^0_{i+1} n_i \Delta - (K - n^*_{i-1}) n^*_{i-1} n_i \Delta \\
&= n h^0_{i-1} h^0_{i+1} + n b - (K - n_i) (a - n_i \Delta) n_i \Delta \\
&\quad - (n^*_{i-1} - n_i) h^0_{i+1} n_i \Delta - (K - n_i) n^*_{i-1} n_i \Delta + (n^*_{i-1} - n_i) n^*_{i-1} n_i \Delta \\
&= n b - (K - n) (a - n_i \Delta) n_i \Delta + n_i h^0_{i-1} h^0_{i+1} - (n^*_{i-1} - n_i) (h^0_{i+1} - n^*_{i-1} \Delta) n_i \Delta
\end{align*}

which is precisely (4.3.31).

The first two terms in (4.3.31) have the same form as (4.3.17), except that $a$ is replaced by $a' = a + n^*_{i-1} \Delta$ because the headway of $i-1$ has reduced by $n^*_{i-1} \Delta$ after deadheading, and hence the no-control headway of $i$ is increased by the same amount at station $k_0 + n^*_{i-1} \Delta$ and beyond. The second order derivative for the remaining terms in (4.3.31) with respect to $n_i$ is non-negative. Hence the real relaxation of [1-DHa] is also convex if the corresponding [1-DH'] (with $a'$) is convex, and the complexity of solving [1-DHa] is no greater than solving [1-DH]. Due to constraint (4.3.32), constraints (4.3.19) and (4.3.23) are no longer needed. The upper bound for $n_i$, naturally, is $n^*_{i-1}$. Also, note that $h^0_i$ is always smaller than $h^0_{i+1}$, otherwise we would not have $n^*_{i-1} > 0$. An important implication of (4.3.31) is that there is an incentive for $i$ to skip fewer stations if there are leftover passengers from $i-1$, so that some of them will benefit from deadheading $i$. Therefore, the optimal $n$ for (4.3.31) will never be larger than the optimal $n$ for (4.3.17') with $a$ replaced by $a'$.

With the 1-DHa model, we present another algorithm for the RTDPF which allows adjacent deadheading.

**Algorithm Adjacent Deadheading**

For each vehicle $i=2,3,...,M-1$

Step 1: If $n^*_{i-1} > 0$ and $n^*_{i-2} > 0$, set $n^*_{i-1} = 0$ (i is not controlled). Go to $i+1$.

Step 2: Else if $n^*_{i-1} = 0$, use algorithm 1-DH to solve the 1-DH subproblem for $i$.

Step 3: Else if $n^*_{i-1} > 0$ and $n^*_{i-2} = 0$, solve both the subproblems 1-DHa and 1-DHb without constraint (4.3.34). If 1-DHa is non-convex, simply enumerate $n$ from 0
to \( n^*_{i-1} \). 1-DHb without constraint (4.3.34) can be solved using the 1-DH algorithm with \( a \) replaced by \( a' \). Let \( n_u \) and \( n_v \) denote the optimal solutions to 1-DHa and 1-DHb without constraint (4.3.34) respectively. If there \( n_v > n^*_{i-1} \) and \( f(n_v) < f(n_u) \), \( n^*_i = n_v \). Otherwise, \( n^*_i = n_u \).

Step 4: Go to \( i+1 \).

The above algorithm solves each 1-DH subproblem optimally whether the preceding vehicle is deadheaded or not. However, when rolling horizon \( m > 1 \) is chosen, the optimal solution may be different. In general, we can call a subproblem of the RTDP the \( m \)-DH subproblem if \( m \) consecutive vehicles are simultaneously considered for control, and adjacent deadheading is not restricted. When \( m > 1 \), the \( m \)-DH subproblem is non-convex and very difficult to solve analytically. On the other hand, there is seldom a need to deadhead more than two consecutive vehicles. While complete enumeration may be feasible for the 2-DH subproblem when \( K \) is small, it is not practical when \( K \) is large or when an iterative search, such as in the case of combined control studied in Chapter 7, is needed. In addition, when assumption A4.1 is relaxed, the system improvement with \( m = 2 \) may not be significantly larger than when \( m = 1 \). Thus, algorithm Adjacent Deadheading may be sufficient. This issue will be investigated further in the next subsection.

4.3.7 Computational Tests with Green Line Data

In this section we test the overall effectiveness of deadheading in System F using the Green Line data sets. The network contains 26 stations in each direction. Because there is layover time at station 52 but not at station 26, we set \( u_1 = 1 \) for control direction 1, and \( u_2 = 0 \) for control direction 2. Letting \( c_0 = 0.37, \delta = 0.15, h_0 = 0.5 \) (all in minutes), two different tests were performed using the Green Line data sets:

Test 1. Use Algorithm Non-Adjacent Deadheading to solve the RTDPF.

Test 2. Allow adjacent deadheading, and use complete enumeration to obtain the optimal solution for each \( I_m \) set with \( m = 1, 2, 3 \) respectively.
The main purposes of these tests are to:

(i) Investigate effectiveness and characteristics of deadheading policies in general.
(ii) Further investigate the effects of adjacent deadheading.
(iii) Investigate the effects of the impact set size $m$ when adjacent deadheading is not restricted.

The computational results for these tests are listed in Tables 4.3.4 and 4.3.5.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>M</th>
<th>No-control Cost</th>
<th>#skipped stations</th>
<th># deadheaded vehicle trips</th>
<th>Change in Cost</th>
<th>% Change in Cost</th>
<th>StdDev of Headway</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>36</td>
<td>49,114.70</td>
<td>11</td>
<td>10</td>
<td>-610.50</td>
<td>-1.24</td>
<td>1.67</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>50,409.10</td>
<td>14</td>
<td>10</td>
<td>-1,019.47</td>
<td>-2.02</td>
<td>1.08</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>55,021.20</td>
<td>12</td>
<td>9</td>
<td>-1,098.12</td>
<td>-2.00</td>
<td>0.63</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>55,002.80</td>
<td>15</td>
<td>8</td>
<td>-1,792.36</td>
<td>-3.26</td>
<td>1.30</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>43,244.70</td>
<td>9</td>
<td>8</td>
<td>-589.08</td>
<td>-1.36</td>
<td>0.85</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>38,361.50</td>
<td>47</td>
<td>12</td>
<td>-7,214.02</td>
<td>-18.81</td>
<td>0.63</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>37,232.50</td>
<td>51</td>
<td>14</td>
<td>-6,448.33</td>
<td>-17.32</td>
<td>2.26</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>39,846.70</td>
<td>37</td>
<td>9</td>
<td>-5,848.65</td>
<td>-14.68</td>
<td>1.68</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>38,018.00</td>
<td>41</td>
<td>12</td>
<td>-5,445.93</td>
<td>-14.32</td>
<td>2.64</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>30,820.70</td>
<td>41</td>
<td>12</td>
<td>-4,880.34</td>
<td>-15.83</td>
<td>2.16</td>
</tr>
<tr>
<td>Total</td>
<td>346</td>
<td>437,071.90</td>
<td>278</td>
<td>104</td>
<td>-34,946.80</td>
<td>-8.00</td>
<td></td>
</tr>
<tr>
<td>Aver.</td>
<td></td>
<td>87,414.38</td>
<td></td>
<td></td>
<td>-6,989.36</td>
<td></td>
<td>1.49</td>
</tr>
</tbody>
</table>

Table 4.3.4 System F: Test 1 Results (without Adjacent Deadheading)

From Tables 4.3.4 and 3.4.5 we can see that the total costs are reduced for all datasets with the reduction ranging from 1% to 19% with an average of 8%. This is not bad considering the high absolute value of cost. Using the constant passenger arrival rate $r=2.39$ (the average for the Green Line B line), average total waiting time is reduced by about 7000 passenger-minutes per morning peak. Also, compared to Table 3.1 in Chapter 3, the average standard deviation of headways is reduced from 2.47 to 1.49 minutes, a 40% reduction. In general, if the variance of the headways is higher, the percentage cost reduction is also higher. A higher percentage cost reduction is always associated with a
higher number of skipped stations. This is the case for all data sets in direction 2, which have higher headway variances because there is no layover time after trips in direction 1.

<table>
<thead>
<tr>
<th>rolling horizon</th>
<th>m=1</th>
<th></th>
<th>m=2</th>
<th></th>
<th>m=3</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>cost change</td>
<td>total n</td>
<td>cost change</td>
<td>total n</td>
<td>cost change</td>
</tr>
<tr>
<td>data set</td>
<td>M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m1</td>
<td>36</td>
<td>-1.40</td>
<td>12</td>
<td>-1.53</td>
<td>16</td>
<td>-1.53</td>
</tr>
<tr>
<td>t1</td>
<td>35</td>
<td>-2.02</td>
<td>14</td>
<td>-2.02</td>
<td>14</td>
<td>-2.02</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>-2.01</td>
<td>13</td>
<td>-2.01</td>
<td>13</td>
<td>-2.01</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>-3.39</td>
<td>16</td>
<td>-3.39</td>
<td>16</td>
<td>-3.39</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>-1.36</td>
<td>9</td>
<td>-1.36</td>
<td>9</td>
<td>-1.36</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>-20.14</td>
<td>55</td>
<td>-20.72</td>
<td>65</td>
<td>-20.72</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>-17.36</td>
<td>52</td>
<td>-17.47</td>
<td>57</td>
<td>-17.47</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>-16.28</td>
<td>42</td>
<td>-17.17</td>
<td>50</td>
<td>-17.25</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>-15.23</td>
<td>47</td>
<td>-15.46</td>
<td>54</td>
<td>-15.46</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>-15.83</td>
<td>41</td>
<td>-15.89</td>
<td>42</td>
<td>-15.89</td>
</tr>
<tr>
<td>week total</td>
<td>346</td>
<td>-8.38</td>
<td>301</td>
<td>-8.56</td>
<td>336</td>
<td>-8.56</td>
</tr>
</tbody>
</table>

Table 4.3.5 System F: Test 2 Results (with Adjacent Deadheading)

A number of remarks on the test results are in order.

(1) When adjacent deadheading is allowed, 8 out of 10 data sets show more improvements in cost reductions compared to Table 4.3.4, though the marginal benefits are small. Such improvements are somewhat higher in direction 2, where more adjacent vehicles can be deadheaded with benefits than direction 1 due to their headway patterns (see Table 4.3.7 below). The differences in cost reduction between m=1 and m>1 are insignificant even though 35 or more (>10%) additional stations would be skipped when m>1. This indicates that solving m-DH subproblems with m>1 is probably not worth the effort. In Test 1, there are only 4 cases in which two adjacent vehicles each have beneficial deadheading cost, and the ones with lower benefits were not deadheaded (see Table 4.3.6). In Test 2, on the other hand, there are 15 adjacent deadheading cases when m=1, and 20 when m>1. Although the number of adjacent deadheading cases increased when m>1, with either m=1 or m>1 there are only 2 cases where 3 consecutive vehicles were deadheaded (data set "m2", vehicles 15, 16, 17 and "w2", 4,5,6). All other adjacent deadheading cases involve just 2 vehicles. This shows that to assume no more than 2
consecutive deadhead vehicles is very reasonable. Since the effectiveness of control with and without adjacent deadheading is very similar, assumption A4.1 (no adjacent deadheading at all) also seems reasonable. Therefore, with similar computational effort, either algorithm Non-Adjacent Deadheading or Adjacent Deadheading can be used depending on whether adjacent deadheading is acceptable.

<table>
<thead>
<tr>
<th>data set</th>
<th>w1</th>
<th>th1</th>
<th>m2</th>
<th>th2</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td>i</td>
<td>18</td>
<td>15</td>
<td>20</td>
<td>18</td>
<td></td>
</tr>
<tr>
<td>n*</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>w*</td>
<td>-5.60</td>
<td>-1.70</td>
<td>-20.88</td>
<td>-10.03</td>
<td>-38.24</td>
</tr>
</tbody>
</table>

Table 4.3.6 System F: Discarded Adjacent Deadheading in Test 1

(2) The headway sequences of all consecutively deadheaded vehicle pairs fall in one of the following two patterns:

**Headway Pattern 1:** \( h^0_i > h^0_{i+1} > h^0_{i+2} \) and \( h^0_i / h^0_{i+1} < h^0_{i+1} / h^0_{i+2} \) \( \iff \) \((h^0_{i+1})^2 > h^0_i h^0_{i+2}\). In words, the headway of \( i \) is large, the headway of \( i+2 \) is small, but the middle one is not too small. An example is the headway sequence of vehicles 8,9,10 in data set "m1", see Appendix C. When adjacent deadheading is beneficial with such a headway pattern, there is a good chance the second vehicle will skip more stations than the first.

**Headway Pattern 2:** \( h^0_{i+1}, h^0_{i+2} \ll h^0_i, h^0_{i+2} \) may or may not be smaller than \( h^0_{i+1} \). This is the typical "long-short-short" headway pattern for candidate adjacent deadheading. An example is the headway sequence of vehicles 8,9,10 in data set "m2" (see Appendix C). With this headway pattern, the second vehicle always skips fewer, if any, stations.

In the Test 2 results, when \( m > 1 \), each headway pattern claims half of the adjacent deadheading pairs. A pair \( \{i, i+1\} \) with headway pattern 2 always have \( n^*_i < n^*_{i+1} \), and the reverse \((n^*_i \geq n^*_j)\) occurs in 8 of the 9 pairs with headway pattern 1. Table 4.3.7a lists the pairs when \( m = 3 \). In this table headway patterns 1 and 2 each occurs 50% of the time.
Adjacent deadheading with headway pattern 1 (highlighted in Table 4.3.7a) is particularly undesirable by passengers, because the headways of both deadhead vehicles are large. In Table 4.3.7a, passengers left by vehicle $i$ have to wait between 11-20 minutes. For a 5 minute headway service, this waiting time is very large. Therefore, such adjacent deadheading should be avoided. In this case 1-DHb subproblem can be dropped out from algorithm Adjacent Deadheading.

(3) In Test 2, when $m=1$, there are 13 adjacent deadheading pairs (see Table 4.3.7b), a subset of the 18 for $m=1$ (see Table 4.3.7a). Compared to the same pairs in Table 4.4.7a, we found that 10 of the first vehicles ($i$) skipped fewer stations, 5 of the second vehicles ($i+1$) skipped fewer stations, and the remainder skipped the same number of stations. Interestingly, only 3 of the pairs (23%) have headway pattern 1 (highlighted). This shows that when the impact of future deadheading is not considered, the resulting deadheading
policies are less severe, and the undesirable headway pattern is less likely to occur in adjacent deadheading cases.

The following points emerge from Test 1 results, which are for 1-DH subproblems where the preceding vehicle is not controlled. These results are important for both non-adjacent and adjacent deadheading cases, because the 1-DHa and 1-DHb models were developed based upon the 1-DH model, and algorithm "Adjacent Deadheading" is an extension to the 1-DH algorithm.

(4) In Test 1, there are only 4.6% (16 out of 346) vehicle trips, all in direction 2 (where \( u_2 = 0 \) and \( K/N/2 - 1 = 25 \)), with \( a/\Delta > K/2 = 25/2 \). That is, \( h_i^0 - h_{i+1}^0 > K\Delta/2 = 25\times0.67/2 = 8.375 \), since \( \Delta = 0.37 + 2\times0.15 = 0.67 \). The highest value of \( a \) is 14.72, which satisfies \( (a+\Delta)/(2\Delta) = 11.485 < (a/\Delta + K)/3 = 15.65 \). Thus all 1-DH subproblems, including those with \( u_2 = 0 \) and \( a/\Delta > K/2 \), have a convex real relaxation and a unique minimum which can be easily obtained by applying Algorithm 1-DH. Though not definitive, this result is likely to hold in a variety of circumstances.

(5) \( h_i^0/h_{i+1}^0 \) can be used as rough indicator of how large the benefits are when deadheading vehicle \( i \). Figures 4.3.6 shows that in Test 1, the 1-DH costs decrease as \( h_i^0/h_{i+1}^0 \) increases (although not monotonically). In Direction 1, with \( u_1 = 1 \), \( h_i^0/h_{i+1}^0 \) must be at least 1.34 for deadheading to be beneficial. In Direction 2, with \( u_2 = 0 \), \( h_i^0/h_{i+1}^0 \) must be at least 1.5 for deadheading to be beneficial.

(6) It is interesting to observe that, although \( a \) appears in all sorts of conditions to bound the optimal \( n^* \), \( n^* \) does not parallel \( a \). Instead, figures 4.3.7 and 4.3.8 show that in Test 1 as \( a \) and \( n^* \) increase the gap between them seems to grow exponentially. On the other hand, the upper bound from inequality (4.3.23) is quite tight in both directions. When the headway ratio \( h_i^0/h_{i+1}^0 \) is small (\( \leq 2H \)), it almost entirely overlaps with \( n^* \) (direction 1). However, in direction 2 when the deadhead vehicles are tailed by their following vehicles, the offsets between \( h_i^0/h_{i+1}^0 \) and \( n^* \) become larger and larger. This seems to indicate that, when the headway ratio \( h_i^0/h_{i+1}^0 \) is smaller than \( 2H \), it can be used to
approximate the optimal number of skipped stations. Otherwise, the upper bound is a better approximation.

Fig. 4.3.6  System F: Cost Change vs. Headway Ratio

Fig. 4.3.7. System F: Optimal 1-DH Solution vs. Headway Pattern (Dir 1)
(7) Most optimal $n^*$s are very close to the integer upper bound $[(a+\Delta)/(2\Delta)]$, where the square brackets denotes the integer part of a real value. Among the 108 deadheading solutions in Test 1, 43 of them (40%) have $n^*=[(a+\Delta)/(2\Delta)]$, 44 (41%) differ by only one station, and only 21 (19%) differ by two or more stations. This shows that $(a+\Delta)/(2\Delta)$ is a quite tight upper bound, and when a 1-DH subproblem is not convex the enumeration algorithm from 0 to the upper bound is efficient.

(8) The 1-DH subproblems in which the difference between $(a+\Delta)/(2\Delta)$ and $n^*$ is larger appear to have larger than usual preceding headways $h_i^0$. This clearly shows that a large preceding headway does not necessarily mean a large number of skipped stations, as might have being a "rule of thumb" in practice.

4.4 Deadheading in System G

System G is a more general and more complicated system type than System F (See Chapter 3 for a detailed description). In this section we study the RTDP for system G, called RTDPG, in which we do not restrict adjacent deadheading. In this more general problem, we are interested in whether the objective function (4.2.2) is still well behaved,
what are the impacts of the variable dwell time, variable demand, adjacent deadheading, and forced interstation stopping on the solutions, and how effective deadheading policies are in this type of system.

This section is organized as follows: in section 4.4.1, we present the formulation for RTDPG and discuss the solution methods. Section 4.4.2 analyzes the appropriate size of an "impact set". Section 4.4.3 analyzes cost characteristics of the RTDPG. Section 4.4.4 studies the effectiveness of deadheading in System G and the influences of various factors. Section 4.4.5 discusses conditions for a beneficial deadheading policy. Section 4.4.6 discusses differences between System G and System F. Section 4.4.7 concludes with an efficient heuristic algorithm for RTDPG.

In the following discussion, Definitions 3.1, 3.2 and 4.1-4.4 apply.

4.4.1 The RTDPG Formulation

The formulation of the deadheading model for System G is no different from RTDP, presented in section 4.2.3, except for how variables $a_{i,k}$, $d_{i,k}$, $s_{i,k}$, and $P_{i,k}$ are defined. We give these variable definitions for System G below.

\[(4.4.1)\quad a_{i,k} = (\max(d_{i-1,k} + h_{0}, d_{i,k-1} + R_{i,k} + 2\delta) y_{i-1,k} y_{i,k} + (d_{i-1,k} + R_{i,k} + (y_{i,k} + y_{i,k-1}) \delta) (1-y_{i,k-1} y_{i,k}), \forall i,k \]

\[(4.4.2)\quad d_{i,k} = a_{i,k} + s_{i,k} \quad \forall i,k \]

\[(4.4.3)\quad s_{i,k} = [c_0 + c_1(P_{i-1,k} + r_i h_{i,k}) + c_2 q_k L_{i,k-1}] y_{i,k}, \forall i,k \]

\[(4.4.4)\quad P_{i,k} = (P_{i-1,k} + h_{i,k} r_{i,k}) (1-y_{i,k}) ; \]

\[(4.4.5)\quad L_{i,k} = L_{i,k-1} + y_{i,k} (P_{i-1,k} + h_{i,k} r_{i,k} - L_{i,k-1} q_k), \forall i,k \]

These variable definition equations are much more complicated than in [RTDPF]. (4.4.1) describes the nonlinear behavior of vehicle arrival times. The first half of its right-hand side gives the arrival time at station $k$ when neither $k-1$ nor $k$ is a control station, which is exactly the same as (2.1), and the second half gives the arrival time at $k$ when either $k-1$ or $k$ is a control station. The logic here is that the no-control movement of
a vehicle allows interstation stopping when the vehicle is blocked by its preceding one, but deadheading does not. It is simply not feasible to skip a station when the vehicle is blocked at that station. If \( y_{i,k-1} = y_{i,k} = 1 \), (4.4.1) is exactly the same as (2.1); otherwise if one or both of them is 0, the arrival time will be different. (4.4.2) has the same form as (2.6), but the dwell time, \( s_{i,k} \), has a particular definition given by (4.4.3). Here the number of alighting passengers is \( A_{i,k} = q_k L_{i,k-1} \) as defined in (2.3). The number of boarding passengers is \( B_{i,k} = P_{i+1,k} + r_k h_{i,k} \) when station \( k \) is not skipped, which is the leftover passengers from the previous vehicle \( i-1 \) plus random passenger arrivals before the departure of the present vehicle \( i \). This is also the number of leftover passengers from \( i \) if \( i \) skips \( k \), as defined in (4.4.4). By definition, at a control station we should have

\[
(4.4.6) \quad s_{i,k} = 0
\]

\[
(4.4.7) \quad P_{i,k} = P_{i+1,k} + h_{i,k} r_k
\]

Otherwise, if vehicle \( i \) stops at station \( k \) normally, we should have

\[
(4.4.6') \quad s_{i,k} = c_0 + c_1 (P_{i+1,k} + r_k h_{i,k}) + c_2 q_k L_{i,k-1}
\]

\[
(4.4.7') \quad P_{i,k} = 0
\]

It is easy to verify that when \( y_{i,k} = 0 \) in (4.4.3) and (4.4.4) we will obtain (4.4.6) and (4.4.7), otherwise we have (4.4.6') and (4.4.7'). Vehicle load information is required by (4.4.3) and this is defined by (4.4.5). If station \( k \) is not skipped by vehicle \( i \) (i.e., \( y_{i,k} = 1 \)), the departure load of vehicle \( i \) at station \( k \) as its arrival load at \( k \), which is equal to its departure load at \( k-1 \), plus the number of boarding passengers minus the number of alighting passengers at \( k \). Otherwise if \( k \) is skipped by \( i \) (i.e., \( y_{i,k} = 0 \)), the load does not change.

The formulation for RTDPG is then given by:

**[RTDPG]:**

Minimize (4.2.2)

Subject to (4.2.4)-(4.2.8)
where \( a_{i,k}, d_{i,k}, P_{i,k}, s_{i,k}, \) and \( L_{i,k} \) are given by (4.4.1), (4.4.2), (4.4.4), (4.4.3), and (4.4.5) respectively.

[RTDPG] is a complicated nonlinear integer program with nonlinear constraints. Although we can slightly simplify the model by using the technical trick we used in the RTDPF model (i.e., using the difference between before and after deadheading as the cost), it does not provide much help in this case. The main difficulty is that even with no adjacent deadheading the problem is no longer separable. An uncontrolled vehicle’s trajectory may be significantly changed by control on its preceding vehicle. This is because, first, a vehicle’s dwell time is a function of its preceding headway; second, an uncontrolled vehicle may be blocked by its preceding vehicle; neither phenomena exist in System \( F \). Furthermore, adjacent deadheading becomes a much more complicated problem. In System \( F \), while the \( m \)-DH subproblems are non-convex with \( m>1 \), the \( 1 \)-DH subproblem is generally convex and tractable as we have shown in the previous section. In System \( G \), the \( m \)-DH subproblems will not be as well behaved. Even with rolling horizon \( m=1 \), [RTDPG] is mathematically intractable. Therefore, we need alternative methods to study RTDPG.

In particular, we will conduct various computational tests to investigate the properties of RTDPG, the similarities and differences between RTDPG and RTDPF, and the effectiveness of deadheading policies and influences of various factors in System \( G \). Because the complexity and computational burden for exact \( m \)-DH solutions are too high when \( m>3 \) (the computation time is in the order of \( K^m \)), we consider an alternative method referred to as "control-the-first". In this method, we control only the first vehicle \( i \) in each \( I_m \) set, and instead of evaluating the impact on the control policy for \( i \) from controlling the following \( m-1 \) vehicles, we evaluate the impact based on their no-control trajectories. Due to the non-separability of System \( G \), even with only one vehicle to be controlled, the optimal control policy can be different when different numbers of following vehicles are considered in the cost evaluation. This is a major difference between Systems \( G \) and \( F \).
The main advantage of this method is its simplicity and feasibility for real-time implementation. The computation time increases only linearly instead of exponentially as \( m \) increases. The method does not prevent adjacent control, due to the rolling scheme. A concern with this method is, however, whether the resulting control policies will be significantly less effective than those obtained by solving each \( m \)-DH problem exactly. For System \( F \), we have shown that the results are different, although these differences may not be viewed as significant. While System \( G \) is different, we again expect that such differences would be small, especially when deadheading more than two consecutive vehicles does not frequently occur. On the other hand, a small marginal improvement in total cost reduction with higher price paid by skipped passengers may be a problem with exact \( m \)-DH solutions. What is not taken into account in the "control-the-first" method is the impact of controlling future vehicles on the current control policy. When station skipping strategies are used in isolation, including the impact of a future control in cost evaluation is likely to result in more severe current control policies. That is, the current vehicle is likely to skip more stations, as we have seen in the System \( F \) test results. In this sense, it may not be a disadvantage that current control policies do not consider future control. In fact, considering the dramatic nature of station skipping strategies, the independence of future control may be an advantage in a stochastic operating environment for which forecast information is not accurate.

To verify the above points, we investigate the differences between the exact \( m \)-DH solutions and the "control-the-first" solutions by again conducting two computational tests with the Green Line data:

**Test 1:** Solve the RTDPG with "control-the-first" method, using different "impact set" (defined below) size \( m \).

**Test 2:** Solve each \( m \)-DH subproblem exactly, with \( m=2 \) and 3.

For both tests, we perform complete enumeration using the model RTDPG. A more efficient heuristic will be developed later in this Chapter. For Test 2, the 1-DH results are not different from Test 1 with \( m=1 \), and hence are omitted.
We now define the concept of an "impact set" for the "control-the-first" problem.

**Definition 4.6**

An impact set for vehicle $i$ is a set of adjacent vehicle trips $I_m^i = \{i, i+1, \ldots, i+m\} \subseteq I$, where $i (>1)$ is considered for control. $m$ is called the size of the impact set. Let

$$f_j = \sum_{k=k_0}^{k_i} r_k (h_{j,k}^2 + P_{j-1,k}x_{j,k})$$

then the impact set cost for vehicle $i$ is defined by $w_m^i(i) = \sum_{j=i}^{i+m} f_j$ if $i+m < M$, and $\sum_{j=i}^{M} f_j$ otherwise.

Defining the total number of vehicles active on the system as a system of vehicles and denoting it $M_s$, we first examine the choice of a good impact set size, $m$, for the "control-the-first" problem in Test 1. We then compare the results with chosen $m$ to the results from Test 2. In the discussion that follows we refer to the total cost over the $M_s$ vehicle trips as the system cost.

**4.4.2 Effects of Impact Set Size**

In System $F$, control on vehicle $i$ does not impact other vehicle trajectories, and the cost change does not involve any vehicle beyond $i+1$. Hence, there is no need to consider uncontrolled vehicle trips beyond $i+1$ in the cost evaluation. In System $G$, however, control on a vehicle will have impact on several following vehicles due to trajectory interactions. In Chapter 3 we have shown that such interactions are mainly due to demand variability via vehicle dwell time variability across stations. Trajectory change of one vehicle will change the demand for next vehicle, that vehicle's dwell time change will again influence the demand for the next vehicle, and so on, until the effect diminishes after several vehicle trips. Therefore, it is necessary to consider the impact of a control policy on several following vehicles.

A necessary question in this context is what would be an appropriate impact set size, or rolling horizon size. A first and natural answer may be to consider the whole system of $M_s$ vehicles when examining each individual vehicle. However, considerations of future information requirements and computational efficiency make this impractical. Another
extreme is to consider solely a 1-DH set for each vehicle $i$, where the impact of deadheading $i$ is computed in the cost function only over vehicle trips $i$ and $i+1$. Although this approach requires the least future information, it may be nearsighted and result in "bad" deadheading policy from the system point of view. Since we have shown in Chapter 3 the "diminishing trajectory change effect" in System $G$, we could also consider a small number $m$, $1 < m < M$, for evaluation of the impact of a control.

Let us consider first how the following vehicle trajectories change after deadheading vehicle $i$. Here both the "diminishing trajectory change effects" and the "alternating sign effect" along the vehicle dimension, and the "dispatching headway effects" along the station dimension play a role. In general $\sum_k \Delta d_{ik}$ for $k \in K$, will have a negative sign, and in turn cause positive sign of $\sum_k \Delta d_{i+1,k}$. While $\sum_k \Delta d_{i+1,k}$ can be quite significant, following the proof of diminishing effects in Chapter 3, after a few vehicles the impact of such a change becomes insignificant. The alternating sign effect, on the other hand, will have a particular impact on an impact set cost depending on how the set size is chosen and whether a vehicle in the set is blocked. In particular, consider $m$ vehicles after $i$ to examine the impact of deadheading $i$. If $m$ is chosen as 1, then the decreased value of $h_{i+2,k}$ and the associated cost reduction is overlooked; if $m$ is chosen as 2, then the increased value of $h_{i+3,k}$ and the associated cost increase is overlooked, and so on. But it will not be a big problem when $m > 3$, because the impact of controlling vehicle $i$ would be insignificant on the vehicles $j > i + 3$. To further explore this issue, we perform a computational test again using the Green Line data.

**Computational Results for Test 1**

We perform a number of simulation runs using model RTDPG with different rolling horizon sizes $m$ on the Green Line data sets. $M$ is set to 16 vehicles, which is the average number of operating vehicles in the Green Line data sets. The simulation model was coded in C and was run on a PC 486. In the simulation we use complete enumeration to find an optimal deadhead segment for each vehicle $i$ within the boundary of its impact set. In particular, we deadhead each vehicle on its feasible deadhead segments with increasing...
number of skipped stations, and record the segment that results in the lowest set cost. We simulate control actions in a rolling horizon scheme: for each vehicle $i$, if the optimal impact set cost for deadheading $i$ found is beneficial, the deadheading policy is set permanently for $i$ and the trajectories of all vehicles from $i$ to $M_i$ are updated before examining the next vehicle $i+1$. This also gives us a chance to examine the effects of adjacent deadheading. We set $c_0=0.37$, $c_1=0.007$, $c_2=0.008$, $\delta=0.15$, $h_0=0.5$, and $H=5$ (all in minutes). The differences between input data in RTDPF and RTDPG are the dwell function parameters and the variable demand (see Appendix B for the demand profile of the Green Line B line).

We examine the value of $m$ after which the impact of a control policy becomes insignificant by comparing the differences between optimal solutions by rolling horizon size. The reason for using this indicator is that if an impact set is more or less "independent", then there would be no significant difference between cost reductions after the optimal deadheading.

Table 4.4.1 compares the resulting total cost reductions in each data set by rolling horizon size, where $\Delta f\% = (f(h(y)) - f^0(h^0))/f^0(h^0) * 100\%$, and $f(h(y)) = \sum_{i=2}^{M_i} f_i$. We can see from Table 4.4.1 and Fig. 4.4.1, in general, total cost decreases as $m$ increases. From Fig. 4.4.1, it is obvious that the total cost decreases sharply when $m=2$ and $m=3$. When $m>3$, however, the overall cost shows little change.

We should especially note that different rolling horizon sizes bring different biases into cost estimation due to "alternating sign effects" discussed in Chapter 3. When $m=2$ is chosen as the rolling horizon, there is a bias toward overestimating the benefits of deadheading, because of both the significant $\Delta f_{i+2}$ and its negative sign. On the other hand, because of $\Delta f_{i+1}$ 's positive sign, the use of $m=1$ tends to underestimate deadheading benefits. Both of these biases lead to lower systematic improvement, because "over-deadheading" (when $m=2$) will cause larger headways for following vehicles, while "under-deadheading" (when $m=1$) may not be fully effective. Such biases are largely
eliminated when \( m \geq 3 \) is chosen. One can observe from Table 4.4.1 that when \( m=2 \) the number of deadheaded vehicles is the largest and when \( m=1 \) the number of deadheaded vehicles is the smallest. As \( m \) increases, the total number of deadheaded vehicles stabilizes.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>( M )</th>
<th>( m=1 )</th>
<th>( m=2 )</th>
<th>( m=3 )</th>
<th>( m=4 )</th>
<th>( m=5 )</th>
<th>( m=16 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>36</td>
<td>-14.08</td>
<td>-15.28</td>
<td>-18.82</td>
<td>-19.68</td>
<td>-19.68</td>
<td>-19.68</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>-14.20</td>
<td>-15.35</td>
<td>-16.35</td>
<td>-15.95</td>
<td>-16.35</td>
<td>-16.35</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>-11.36</td>
<td>-15.69</td>
<td>-14.02</td>
<td>-16.76</td>
<td>-16.46</td>
<td>-15.65</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>-6.30</td>
<td>-10.75</td>
<td>-8.84</td>
<td>-8.84</td>
<td>-8.84</td>
<td>-9.01</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>-5.19</td>
<td>-6.12</td>
<td>-6.21</td>
<td>-6.20</td>
<td>-6.21</td>
<td>-6.21</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>-5.84</td>
<td>-9.22</td>
<td>-10.97</td>
<td>-11.32</td>
<td>-11.32</td>
<td>-11.48</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>-0.88</td>
<td>-7.33</td>
<td>-7.31</td>
<td>-7.80</td>
<td>-7.80</td>
<td>-7.80</td>
</tr>
<tr>
<td>week total</td>
<td>346</td>
<td>-10.25</td>
<td>-12.94</td>
<td>-13.70</td>
<td>-14.45</td>
<td>-14.49</td>
<td>-14.40</td>
</tr>
</tbody>
</table>

| #deadheaded veh. | 91 | 125 | 118 | 119 | 115 | 117 |

Table 4.4.1 RTDPG: Total Cost Reduction by Rolling Horizon Size

One can also observe that there are a few cases where total cost may sometimes increase slightly as \( m \) increases, For example when \( m=16 \) the total cost reduction is slightly lower than when \( m=5 \) in Table 4.4.1. This phenomena is due to the fact that a dynamic optimization problem with a rolling horizon scheme is different from static optimization over the entire period. In general, larger rolling horizon sizes support longer term optimization. While this trend can be clearly seen from Table 4.4.1 and Fig. 4.4.1, because we are not considering controlling the entire set of vehicles simultaneously, any optimization results in a rolling scheme are generally not equal to the "true" global optimal result over the entire period. Small difference between system costs resulting from using different rolling horizon sizes can occur in either direction at some point during the study period. Nonetheless, overall the effectiveness of control increases with \( m \), and such "rolling horizon errors" are usually very small (under 1% in the above computational results).
It is evident from the simulation results shown above that when \( m > 3 \), the impact of deadheading \( i \) is insignificant on vehicles \( j > i + m \). This is consistent with the theoretical analysis given in Chapter 3. In general, if the average passenger arrival rate is smaller (larger), \( m \) can also be smaller (larger).

In the next subsection, we will compare the Test 1 Results when \( m = 3 \) with the Test 2 results.

### 4.4.3 Comparison between Different Solutions Methods

To investigate the differences between solutions from the "control-the-first" method and the exact \( m \)-DH solutions, we also performed complete enumeration for the 2-DH and 3-DH problems. The results are listed in Table 4.4.2 below. For direct comparison, we also included results from Test 1 with \( m = 3 \) in this table.

Table 4.4.2 shows that, the results from the "control-the-first" method are very similar to 2-DH solutions. The 3-DH solutions seem slightly better: the overall cost reduction is about 1.2% higher than the "control-the-first" solution. Such a small improvement however is obtained as a result of 40% more skipped stations and 38% more deadheaded vehicles. Furthermore, adjacent deadheading is very heavily used to obtain optimal solutions. Table 4.4.3 lists the adjacent deadheading cases in the 3-DH solutions, and Table 4.4.4 lists the adjacent deadheading cases in the "control-the-first" solutions. There are 36 subsets of consecutively deadheaded vehicles in the 3-DH solutions. Among them,
1 case of 6 consecutive vehicles, 3 cases of 5 consecutive vehicles, 5 cases of 4 consecutive deadhead vehicles, 11 cases of 3 consecutive deadhead vehicles, and 16 cases of 2 adjacent vehicles were deadheaded. The headways sum up to 10-37 minutes at station \( k_0 \) in each case. Most of these deadheading policies would almost certainly not be acceptable in practice.

<table>
<thead>
<tr>
<th>Method</th>
<th>Control-the-First</th>
<th>Exact ( m )-DH</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( m=3 )</td>
<td>( m=2 )</td>
</tr>
<tr>
<td>Rolling horizon</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Data Set</td>
<td>( M )</td>
<td>cost change%</td>
</tr>
<tr>
<td>m1</td>
<td>36</td>
<td>-18.82</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>-18.86</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>-16.35</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>-14.02</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>-15.09</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>-9.81</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>-8.84</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>-6.21</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>-10.97</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>-7.31</td>
</tr>
<tr>
<td>total</td>
<td>346</td>
<td>-13.70</td>
</tr>
</tbody>
</table>

Table 4.4.2 Comparison between Different Solution Methods for RTDPG

<table>
<thead>
<tr>
<th>data set</th>
<th>adj. veh.</th>
<th>( n )</th>
<th>data set</th>
<th>adj. veh.</th>
<th>( n )</th>
<th>data set</th>
<th>adj. veh.</th>
<th>( n )</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>3-6</td>
<td>{1,1,1,3}</td>
<td>w1</td>
<td>4-6</td>
<td>{6,3,3}</td>
<td>th1</td>
<td>29-34</td>
<td>{1,2,1,1,1,1}</td>
</tr>
<tr>
<td></td>
<td>8,9</td>
<td>{3,4}</td>
<td></td>
<td>8-10</td>
<td>{1,3,1,1}</td>
<td>f1</td>
<td>3-5</td>
<td>{2,1,3}</td>
</tr>
<tr>
<td></td>
<td>11-13</td>
<td>{1,2,3}</td>
<td></td>
<td>13-16</td>
<td>{2,1,2,1}</td>
<td></td>
<td>7-9</td>
<td>{1,1,1}</td>
</tr>
<tr>
<td></td>
<td>15-17</td>
<td>{2,1,2}</td>
<td></td>
<td>18-21</td>
<td>{2,3,2,1}</td>
<td></td>
<td>11-15</td>
<td>{1,1,2,1,3}</td>
</tr>
<tr>
<td></td>
<td>20-22</td>
<td>{2,2,1}</td>
<td></td>
<td>25-26</td>
<td>{1,1}</td>
<td></td>
<td>19,20</td>
<td>{2,4}</td>
</tr>
<tr>
<td></td>
<td>28-32</td>
<td>{3,3,5,4,1}</td>
<td></td>
<td>28,29</td>
<td>{2,1}</td>
<td></td>
<td>28,29</td>
<td>{2,3}</td>
</tr>
<tr>
<td>tu1</td>
<td>2-4</td>
<td>{2,1,4}</td>
<td>th1</td>
<td>4-6</td>
<td>{2,1,3}</td>
<td>m2</td>
<td>3,4</td>
<td>{1,1}</td>
</tr>
<tr>
<td></td>
<td>8-11</td>
<td>{1,3,1,1}</td>
<td></td>
<td>8-10</td>
<td>{3,2,1}</td>
<td>tu2</td>
<td>20,21</td>
<td>{1,2}</td>
</tr>
<tr>
<td></td>
<td>13-16</td>
<td>{2,1,1,2}</td>
<td></td>
<td>12,13</td>
<td>{4,1}</td>
<td></td>
<td>31,32</td>
<td>{1,1}</td>
</tr>
<tr>
<td></td>
<td>20,21</td>
<td>{1,2}</td>
<td></td>
<td>15,16</td>
<td>{4,5}</td>
<td>w2</td>
<td>4,5</td>
<td>{2,1}</td>
</tr>
<tr>
<td></td>
<td>27-29</td>
<td>{1,1,1}</td>
<td></td>
<td>18,19</td>
<td>{1,5}</td>
<td>th2</td>
<td>12,13</td>
<td>{2,1}</td>
</tr>
<tr>
<td></td>
<td>31,32</td>
<td>{2,3}</td>
<td></td>
<td>21-25</td>
<td>{2,2,1,1,1}</td>
<td></td>
<td>18,19</td>
<td>{1,1}</td>
</tr>
</tbody>
</table>

Table 4.4.3 System G: Adjacent Deadheading Cases in 3-DH Solutions
In contrast, there are only 12 adjacent deadheading cases in the "control-the-first" solutions with impact set size \( m=3 \), and only two of them contain three adjacent vehicles. These results suggest that the additional cost reductions in the 3-DH solutions are mainly due to extremely heavy use of adjacent deadheading. Therefore, the "control-the-first" method is justified in terms of real-time implementability, practical control policies and overall effectiveness.

In the remaining subsections, we will base our analysis on the "control-the-first" method solely.

4.4.4 Cost Characteristics

In section 4.3 we showed that a 1-DH cost function in RTDPF, either with or without the restriction on adjacent deadheading, has a convex real relaxation, which made efficient solution algorithms possible. In this subsection we investigate the characteristics of the cost function in RTDPG. In particular, we are interested in whether the cost function is still well behaved. Because an analytic method is not viable, we study the cost characteristics of RTDPG through computational experiments. Since the cost data points are discrete due to the integer station indices, we connect each data point to form a cost curve that is more meaningful for our analysis, and in the following we will use the term "cost curve" to refer to the connected cost data points by station. A cost curve is associated with each deadheadable vehicle, and its length is equal to the length of the maximal feasible deadhead segment for that vehicle.

In particular, we are interested in the quasiconvexity of the cost curves. A cost curve \( f \) is quasiconvex if, whenever \( f(x_2) \geq f(x_1) \), \( f(x_2) \) is greater than or equal to \( f(x_1) \) at all convex}

115
combinations of \( x_1 \) and \( x_2 \). We placed a quasiconvexity check in all the simulation runs for RTDPG. It turns out that, all cost curves are quasiconvex when \( m=1 \). With \( m>1 \), only a small portion (at most 28, or 8%, with 20 deadhead trips) of the total 346 vehicle trips have non-quasiconvex deadheading cost curves. In the following discussion, we look at cost curves when \( m=1 \) and \( m=3 \).

**Quasiconvexity of 1-DH Cost**

The first observation is that all cost curves are quasiconvex when \( m=1 \). This is not a coincidence. Although it is difficult to give an exact proof for the quasiconvexity of 1-DH cost, we can show that the quasiconvexity of the net benefit portion of the cost curve is quite independent of demand and headway patterns.

When there exists a feasible deadhead segment for vehicle \( i \), the associated 1-DH cost has two portions. The first portion is the sum of passenger waiting times at the skipped stations, called the *skipped portion* of the cost and denoted as \( f^1(n) \), where \( n \) is the number of skipped stations. Such a cost associated with a deadhead vehicle \( i \) can be expressed as follows:

\[
(4.4.8) \quad f^1_i(n) = \sum_{k:k_{da}+n \leq k \leq k_{uf}} [r_k(h_{i,k} + h_{i+1,k})^2/2 + P_{i-1,k}(h_{i,k} + h_{i+1,k})]
\]

The first term is for passengers randomly arriving between vehicles \( i-1 \) and \( i+1 \), and the second term is waiting time of passengers who were left by \( i-1 \) if it also skipped station \( k \). Since all terms are positive, the summation is obviously increasing with \( n \). In fact, \( f^1(n) \) is quasimonotone in \( n \).

The second component of the 1-DH cost is the sum of cost at each station in the benefit segment, i.e., from station \( k_r \) to \( k_i \). This portion is called the *benefit portion* of the cost and denoted as \( f^2(n) \). For notation simplicity let us set \( r'_k = \mu_r \sum_{k \in K_r} r_k \), and \( r'_k = r_k \) for \( k<k_r \).

Thus \( f^2(n) \) is expressed as follows:
(4.4.9) \( f_i^2(n) = \sum_{k=\text{min}+n}^{k_i} \left[ r_i'(h_{i,k}^2 + h_{i+1,k}^2)/2 + P_{i-1,k}h_{i,k} \right] \)

The first term is the waiting time of randomly arriving passengers who board vehicles \( i \) and \( i+1 \), and the second term is the waiting time of leftover passengers (if any) from vehicle \( i-1 \). Note that in the benefit segment of \( i \) (from station \( k_e = k_0+n \) to \( k_i \)), both vehicles \( i \) and \( i+1 \) are uncontrolled and hence \( P_{i,k} = P_{i+1,k} = 0 \).

Suppose there exists a beneficial deadhead segment for vehicle \( i \), in which the values of \( n \) are taken from the feasible region \( [0, n_{\text{max}}] \), where \( n_{\text{max}} < N/2 \). Let us now look at how \( f^2 \) changes with \( n \) in the beneficial region. When \( n \) is increased by 1, we have

\[
(4.4.9') f_i^2(n + 1) = \sum_{k=\text{min}+n+1}^{k_i} \left[ r_i'(h_{i,k}^2 + h_{i+1,k}^2)/2 + P_{i-1,k}h_{i,k}' \right]
\]

and subtracting (4.4.9') from (4.4.9) we obtain the change of \( f^2 \), denoted as \( \Delta f^2 \):

\[
\Delta f_i^2 = f_i^2(n) - f_i^2(n + 1) = r_{k_0+n}(h_{i,k_0+n}^2 + h_{i+1,k_0+n}^2)/2 + P_{i-1,k_0+n}h_{i,k_0+n} + \sum_{k=\text{min}+n+1}^{k_i} \left[ r_i'(h_{i,k}^2 - h_{i,k}'^2 + h_{i+1,k}^2 - h_{i+1,k}'^2)/2 + P_{i-1,k}(h_{i,k} - h_{i,k}') \right].
\]

The first two terms always have positive value because they are the cost at the extra station of the longer benefit segment. The possible negative contribution to \( \Delta f^2 \) can only come from the third term, whose sign depends on whether and how much \( h_{i+1,k} \) is less than \( h_{i+1,k}' \) at each \( k \geq k_0 + n+1 \), since \( h_{i,k} \geq h_{i,k}' \) at all \( k \) due to the headway reduction of \( i \) when \( n \) increases.

The change in \( h_{i+1,k} \) can be indicated by the difference between \( h_{i+1,k} \) and the average \( h_{i,k} \) over all \( k \geq k_e \) (denoted as \( \bar{h}_i \)). As shown in the "dispatching headway effect" section in Chapter 3, when \( h_{i+1,k} \) is smaller than \( \bar{h}_i \), \( h_{i+1,k} \) will keep decreasing along the benefit segment, and the increase of \( h_{i+1,k} \) caused by the increase of \( n \) will be quite small when \( n \) is small. This is why a smaller dispatching headway of \( i+1 \) at \( k_0 \) is important for having a beneficial deadheading policy. In this case the increase in \( h_{i,k} \) may not be sufficient to alter the positive sign of \( \Delta f^2 \) and \( f^2 \) is decreasing as \( n \) increase. When \( n \) is large enough, however, \( \bar{h}_i - h_{i+1,k} \), may become negative and \( h_{i+1,k} \) will increase along the benefit segment. In this case the overall increase of \( h_{i+1,k}' \) over \( h_{i+1,k} \) will be much more and \( \Delta f^2 \).
may become negative, i.e., $f^2$ starts to increase with $n$. This indicates that the curve of $f_i^2(n)$ is quasiconvex over the beneficial region of $n$.

Thus $f_i^1(n) + f_i^2(n)$ as the sum of a quasimonotone and a quasiconvex function, is also quasiconvex. This is illustrated by the example shown in Fig. 4.4.2 below, where $n_{r_{ax}} = 10$.

It turns out that the demand variability across stations does not change quasiconvexity of the net benefit portion of a 1-DH cost curve. For example, the second station in Direction 2 (station 28) of the Green Line B line has the highest passenger arrival rate in morning peak (see Appendix B), but all 1-DH cost curves crossing that station are quasiconvex. Not even such a dramatically different passenger arrival rate results in non-quasiconvexity, this shows the conclusion is robust.

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure}
\caption{Quasiconvex Curves of 1-DH cost}
\end{figure}

\textbf{Non-quasiconvexity When $m>1$}

On the other hand, not all the cost curves are quasiconvex when $m>1$. When $m$ is chosen in the range of 2 to 16, at most 20 deadhead vehicle trips have non-quasiconvex cost curves, and different $m$ in this range results in slightly different subsets of the 20 vehicles. Close examination of all these cost curves reveals a common phenomenon: each of them
has one local maximum at an early point of the curve, and a global minimum at a later point (see Fig. 4.4.3a below for an example). Furthermore, we found that without control, all these vehicles, without exception, block their following vehicles before the end of the control direction. Table 4.4.5 lists the impact sets for these vehicles. As an example of them, Fig. 4.4.3a plots the cost curve when $m=3$ for the vehicle set in which $i=20$, in data set "f1". The cost curve has a maximum at $n=2$ and the global minimum at $n=5$. For comparison, Fig. 4.4.3b shows the cost curve when $m=1$ for the same vehicle set- it looks perfectly well behaved. In this case the optimal number of skipped stations is 2 (3 stations less).

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$i$</th>
<th>headway at $k_i$</th>
<th>$i+1$ blocked at</th>
<th>Dataset</th>
<th>$i$</th>
<th>headway at $k_i$</th>
<th>$i+1$ blocked at</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>30</td>
<td>5.49, 2.42, 1.37</td>
<td>3.57</td>
<td>tu2</td>
<td>2</td>
<td>8.03, 0.76, 11.6</td>
<td>28</td>
</tr>
<tr>
<td></td>
<td>31</td>
<td>2.42, 1.37, 3.57</td>
<td>22</td>
<td></td>
<td>4</td>
<td>11.6, 0.73, 9.97</td>
<td>27</td>
</tr>
<tr>
<td>tu1</td>
<td>4</td>
<td>6.7, 3.1, 7.17</td>
<td>23</td>
<td></td>
<td>21</td>
<td>4.31, 1.00, 9.41</td>
<td>28</td>
</tr>
<tr>
<td>th1</td>
<td>6</td>
<td>5.7, 2.7, 7.06</td>
<td>23</td>
<td>31</td>
<td>8.51, 2.51, 2.36</td>
<td>13.79</td>
<td></td>
</tr>
<tr>
<td></td>
<td>16</td>
<td>4.95, 2.13, 6.42</td>
<td>23</td>
<td></td>
<td>32</td>
<td>3.63, 2.36, 13.26</td>
<td>31</td>
</tr>
<tr>
<td></td>
<td>19</td>
<td>7.6, 1.82, 5.19</td>
<td>14</td>
<td>w2</td>
<td>23</td>
<td>7.42, 0.74, 8.75</td>
<td>28</td>
</tr>
<tr>
<td>f1</td>
<td>20</td>
<td>5.27, 1.75, 5.5</td>
<td>20</td>
<td>th2</td>
<td>6</td>
<td>7.34, 0.73, 11.77</td>
<td>27</td>
</tr>
<tr>
<td>m2</td>
<td>6</td>
<td>7.88, 0.8, 13.04</td>
<td>28</td>
<td>15</td>
<td>9.28, 1.00, 0.93</td>
<td>12.84</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>5.18, 0.87, 15.59</td>
<td>28</td>
<td>f2</td>
<td>3</td>
<td>7.34, 0.89, 11.7</td>
<td>28</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Dataset</th>
<th>$i$</th>
<th>headway at $k_i$</th>
<th>$i+1$ blocked at</th>
<th>Dataset</th>
<th>$i$</th>
<th>headway at $k_i$</th>
<th>$i+1$ blocked at</th>
</tr>
</thead>
<tbody>
<tr>
<td>m2</td>
<td>6</td>
<td>7.88, 0.8, 13.04</td>
<td>28</td>
<td>15</td>
<td>9.28, 1.00, 0.93</td>
<td>12.84</td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>13</td>
<td>5.18, 0.87, 15.59</td>
<td>28</td>
<td>f2</td>
<td>3</td>
<td>7.34, 0.89, 11.7</td>
<td>28</td>
</tr>
</tbody>
</table>

Table 4.4.5 Vehicle Sets with Non-quasiconvex Set Cost when $m=3$

Why does a maximum occur before the global minimum in a cost curve when $m>1$? It has to do with certain headway pattern of vehicles in the impact set and the role of deadheading under these headway patterns. To show this, Fig. 4.4.4 plots trajectories of vehicles 19, 20, 21, 22 and 23 in data set "f1" in three scenarios: no control; vehicle $i$ ($=20$) skips 2 stations (the maximal cost point), and $i$ skips 5 stations (the minimal cost point).
In this example, in the no-control scenario, vehicle \( i+1 \) is blocked by \( i \) at station 20, and \( i \)’s headways increase while \( i+1 \)’s headways decrease along the route. Before station 20, because \( i+1 \)’s headways decrease faster than \( i \)’s headways increase, the sum of \( h_{i,k} \) and \( h_{i+1,k} \) decreases as \( k \) increases. From station 20 on, \( i+1 \) tails \( i \) and the sum of their headways increase with \( h_{i,k} \).
In the second scenario when \( i \) skips 2 stations, its headways are decreasing along the route, but its minimal headway is still larger than \( i+1 \)'s initial headway. It follows from the "dispatching headway effects" shown in Chapter 3 that \( i+1 \)'s headway again decreases along the route. This time \( i+1 \) is no longer blocked by \( i \) and the sum of their headways is also decreasing. From \( k=20 \) on, \( h_{i,k} + h_{i+1,k} \) becomes less than in the no-control scenario.

In both the scenarios \( i+2 \)'s headways keep increasing, since \( i+2 \) has a much larger dispatching headway than \( i+1 \). But in the second scenario \( i+2 \)'s headways become larger than in the no-control scenario from station 20 on, due to the decrease of \( h_{i,k} + h_{i+1,k} \). As \( i+2 \) has the largest headways in the vehicle set, the cost increase associated with it leads to the increase of the set cost.

In the third scenario, \( i \) is apparently "over-deadheaded" and the larger number of skipped stations switched the headways of \( i+1 \) from decreasing to increasing. It can be observed from Fig. 4.4.4 that at the end of the deadhead segment (\( k=6 \)), \( h_{i+1,k} \) becomes almost as
large as $h_{i+2,k}$. From station 6 on, headways of $i+1$ and $i+2$ become quite even, and this makes the greatest contribution to the set cost reduction.

All vehicle sets listed in Table 4.4.2 have a headway pattern similar to the above example. Specifically, either vehicle $i+1$ is bunched with $i$ and vehicle $i+2$ has the largest headway in the set, or vehicles $i$, $i+1$, and $i+2$ are bunched together and $i+3$ has the largest headway in the set (as for the three highlighted vehicle sets in Table 4.4.5). At the global minimal point, the role of deadheading is no longer to even out the preceding or following headways of vehicle $i$, but to even out the larger headways of some later vehicle. The maximum point in their cost curve is due to a short deadheading that decreases the cumulative headways between $i$ and the bunched vehicles but increases the largest headway after them. If vehicle $i$ did not cause interstation stopping for the following vehicles when not controlled, such a maximum point is unlikely to occur after deadheading $i$ and the cost curve will still be quasiconvex.

Because 92% or more of the cost curves are quasiconvex with any choice of $m$, it leads to an important algorithmic improvement. That is, we can terminate a search for the optimal number of stations whenever the cost starts to increase. We will present an efficient solution algorithm for RTDPG which utilizes this property in subsection 4.4.8.

We should note that the different cost characteristics bring both advantages and disadvantages to either choice of $m=1$ or $m>1$. We further discuss this issue in the next subsection.

4.4.5 Effectiveness of Deadheading vs. Choice of $m$

From Table 4.4.1 and Fig. 4.4.1 we saw that the systematic effectiveness of deadheading increases with $m$ and becomes stable when $m \geq 3$. Also, as discussed in Chapter 3, $m=2$ is a biased choice which will result in overestimation of deadheading benefits. On the other hand, $m=1$ is a biased choice which will underestimate deadheading benefits. Hence $m =3$ which has little bias is a better choice from the viewpoint of the effectiveness.

However, there are other advantages and disadvantages with either choice of $m=1$ or $m=3$. 

122
Present Cost vs. Future Cost

Table 4.4.1 shows that when \( m=3 \), the week total cost reduction is about 2.75\% higher than when \( m=1 \). The systematic effectiveness when \( m>1 \) comes from the very fact that a deadheading policy achieves future benefits but often at a price of "over-deadheading" the present vehicle and more often producing adjacent vehicle deadheading. Especially in situations with non-quasiconvex cost curves, due to the cost characteristics, an impact set with \( m>1 \) results in more skipped stations than with \( m=1 \), and likely causes "over-deadheading" - in the sense that it results in a much shorter preceding headway than following headway of the deadhead vehicle. As we saw from the cost curve in Fig. 4.4.3, the optimal number of skipped stations for vehicle 20 in data set "f1" is 5 when \( m=3 \). The third scenario in Fig. 4.4.4 shows that skipping 5 stations makes vehicle 20's preceding headways smaller than its following headways from station 6 on, a phenomenon of "over deadheading". In this way, the trajectory of vehicle 21 is pushed back, which in turn reduces the large headways of vehicle 22. On the other hand, when vehicle 20 skips only 2 stations, while it is the worst control policy when \( m=3 \) is used for cost evaluation, it is the optimal control policy when \( m=1 \) is used. The second scenario in Fig. 4.4.4 shows that, after skipping 2 stations, the preceding and following headways of vehicle 20 \((i)\) are quite even along the entire direction. Although skipping 5 stations results in more regular headways between later vehicles, in a very dynamic situation, to weigh future cost more than present cost may be risky.

When \( m=1 \) is chosen as the impact set size, it weighs present cost more than future cost. It always results in more conservative deadheading policy than when \( m>1 \), and hence may be safer.

Adjacent Vehicle Deadheading

Another consideration in the choice of \( m \) is the frequency of "adjacent vehicle deadheading", the phenomenon of two or more adjacent vehicles being deadheaded. Because the use of \( m>1 \) as impact set size sometimes results in over-deadheading, it also causes adjacent vehicles to be deadheaded more often.
In the simulation results, when \( m=1 \), only three pairs of adjacent vehicles are deadheaded (vehicles 8,9, 30,31 in data set "m1", and 15,16 in "t1"). In contrast, when \( m=3 \), there are 12 cases of two or three adjacent vehicles deadheaded, as shown in Table 4.4.4.

While adjacent deadheading results in lower system cost when \( m=3 \) than when \( m=1 \), it may leave a large headway at the skipped stations, which may be intolerable to waiting passengers. One situation in which adjacent deadheading may be beneficial is, as we discussed for System \( F \), when the headway sequence is of headway pattern 2, i.e., the following vehicles are bunched together. In this case the skipped passengers do not have to wait a long time. Unfortunately, only 4 of the 12 adjacent deadheading cases have headway pattern 2.

To summarize, if more conservative control policies are desired and the transit system performance is highly stochastic, \( m=1 \) may be a better choice than \( m>1 \). The choice of \( m=1 \) also has the advantage of well behaved cost curves and tractable headway conditions for beneficial deadheading, as we will see in the next subsection. Due to its simplicity, \( m=1 \) is the most attractive choice for a manual control system, where any \( m>1 \) may be too complicated to be practical.

It should be noted that the above discussion is most applicable when station skipping strategies are used alone. When station skipping strategies are used with holding, adjacent deadheading and non-quasiconvex cost curves are much less likely to occur. In this case \( m>3 \) has clear advantages over \( m=1 \). Again, in this case a manual control system will not work well and computerized control system becomes necessary.

### 4.4.6 Conditions for a Beneficial Deadheading Policy

**Headway Condition**

In RTDPF we showed that \( h_{i,k}/h_{i+1,k}>1 \) at \( k_0 \) is a necessary condition for a beneficial deadheading policy for vehicle \( i \). In RTDPG, this condition is again satisfied by all deadhead vehicles when \( m=1 \). When \( m=3 \), however, there are 6 cases in which a vehicle is beneficially deadheaded while this condition is violated as shown in Table 4.4.5.
<table>
<thead>
<tr>
<th>Data Set</th>
<th>i</th>
<th>$h_{i,k_0}$</th>
<th>$h_{i,k_0}/h_{i+1,k_0}$</th>
<th>$h_{i,k_i}/h_{i+1,k_i}$</th>
<th>#Skipped Station</th>
<th>Cost* Reduction (psg.min.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>26</td>
<td>5.06</td>
<td>0.92</td>
<td>0.82</td>
<td>1</td>
<td>34.78</td>
</tr>
<tr>
<td>w1</td>
<td>2</td>
<td>1.35</td>
<td>0.26</td>
<td>0.13</td>
<td>1</td>
<td>-308.92</td>
</tr>
<tr>
<td>th1</td>
<td>5</td>
<td>4.80</td>
<td>0.85</td>
<td>1.80</td>
<td>1</td>
<td>-762.45</td>
</tr>
<tr>
<td>th1</td>
<td>10</td>
<td>4.17</td>
<td>0.90</td>
<td>0.38</td>
<td>1</td>
<td>143.68</td>
</tr>
<tr>
<td>w2</td>
<td>10</td>
<td>4.28</td>
<td>0.82</td>
<td>0.76</td>
<td>1</td>
<td>64.91</td>
</tr>
<tr>
<td>th2</td>
<td>13</td>
<td>2.18</td>
<td>0.91</td>
<td>15.16</td>
<td>1</td>
<td>-71.43</td>
</tr>
</tbody>
</table>

Note: *Impact set cost.

Table 4.4.6 System G: Deadhead Vehicles with $h_{i,k_i}/h_{i+1,k_i} < 1$ at $k=k_0$, $m=3$

In all these cases, the preceding-following headway ratio of the deadhead vehicle $i$ at $k_0$ is less than 1, and the preceding headway itself is large enough to allow deadheading. The number of skipped stations is 1 in all cases, indicating a rather small contribution to systematic benefits. The reason such "abnormal" headway conditions occur is similar to the causes of non-quasiconvex cost curves, only to a lesser degree. That is, in each of the impact sets associated with the deadhead vehicle $i$, there is a vehicle trip $j > i+1$ which has the largest headways in the impact set. The function of deadheading may or may not serve to reduce the current cost associated with $i$, but reduces the larger future cost associated with $j$, thus minimizing the impact set cost. As an example, Fig. 4.4.5 plots vehicle trajectories in the impact set of vehicle 10 in data set "th1", before and after deadheading vehicle 10.

From Fig 4.4.5 we can see that, after deadheading $i$, the short preceding headway of vehicle $i$ becomes even shorter, apparently a case of over-deadheading. However, the preceding and following headways of $i+1$ are more even than with no control, and this makes a larger contribution to the set cost reduction. The cost curve in this case is still quasiconvex because vehicle $i+1$ is not blocked by $i$ in the no-control case.
Fig. 4.4.5 Example of Abnormal Headway Condition (m=3)

To summarize the above discussion, we can still use headway condition $h_{ij} / h_{i+1,k} > 1$ at $k_0$ to speed up the solution algorithm for RTDPG for any choice of $m$. When $m > 1$, the consequence of this is that we may end up taking slightly more conservative control actions. In the above simulation, this means about 3% (i.e., 5 out of 167) less skipped stations. This may result in a slightly lower system cost reduction. However, it may not be a bad thing to weigh present cost more than future cost.

**Length of Deadhead Segment**

We discuss the length of deadhead segment for the case where high demand stations are not at the beginning of the control direction, so that demand is not a major restriction on it. Direction 1 of the Green Line B line thus serves as a good example.

For System $F$ we showed that the optimal number of skipped stations is bounded from above by $(a + \Delta)/(2\Delta)$, where $a = h^0_{i,k_0} - h^0_{i+1,k_0}$, $\Delta = s + 2\delta$, and $s = c_0$ is the constant dwell time.
in System $F$. For System $G$, we are interested whether a similar upper bound exists. As an experiment, we compute a similar "upper bound" $(a + \Delta)/(2\Delta)$ for system $G$, except that we use the mean dwell time $\bar{\delta}$ instead of $c_0$ to compute $\Delta$. In practice, the value of the mean dwell time on a transit route during a given period can be estimated either from field survey or simulation. In our simulation, the mean dwell time is found to be 0.37 minutes (which is the same value we used in System $F$ for $c_0$). Fig. 4.4.6 plots $n$, $h_{i,k_0}/h_{i+1,k_0}$, $h_{i,k_0}$, and the "upper bound" for the Green Line data sets in direction 1 for $m=1$.

![Graph showing 1-DH Subproblems (Sorted by Upper Bound)]

Fig. 4.4.6. System G: Deadheading Results and Conditions (Dir. 1)

A few observations from Fig. 4.4.6 are in order. First, almost all deadhead vehicle $i$ have a preceding headway larger than $H$ (=5 min.). Second, the "upper bound" is valid and tight in all cases for direction 1, and there is a strong similarity between Fig.4.4.6 and Fig.4.3.7. Third, the optimal number of skipped stations $n^*$ closely follows the preceding/following headway ratio of $i$ at $k_0$. Finally, the maximal optimal $n$ is quite small (5 in the above cases).

When $m=3$, we first note that the "upper bound" is meaningless if $h_{i,k_0}^0/h_{i+1,k_0}^0<1$. Hence we exclude such vehicle trips (see Table 4.4.6) and plot the same data items for $m=3$ in Fig. 4.4.7. It can be seen that quite similar patterns exist for $m=3$ as for $m=1$, although
they are not so clear. Once again almost all deadhead vehicles have preceding headway larger than $H$, and the maximal optimal $n$ is the same small number, 5. The "upper bound" of $n^*$ is valid for most deadhead vehicles, but violated in 3 cases. $n^*$ is again very close to $h_{i,k_0}^0/h_{i+1,k_0}^0$, but is has a tendency to be larger. It is not surprising the conditions for $m=3$ are not as neatly established as for $m=1$, due to more complicated headway patterns of vehicles in the impact set.

![Graph showing headway, $n$](image)

**Fig. 4.4.7. System G: Deadheading Results and Condition, $m=3$ (Dir 1)**

The above observations in Direction 1 and the demand pattern effect in Direction 2 as discussed in the previous subsection both suggest that, in general, either $(a+\Delta)/(2\Delta)$ or the number of stations between $k_0$ and the first high demand station, whichever is smaller, may be used as an upper bound on $n$, regardless of the choice of $m$. In a manual control system, $m=1$ should be chosen and $n$ can be approximated by the rounded-up headway ratio $h_{i,k_0}^0/h_{i+1,k_0}^0$. Such an approximation is likely to result in more conservative control actions and will not cause over-deadheading.

### 4.4.7 Comparison between System G and System F Results

Fig. 4.4.8a compares the effectiveness of deadheading (indicated by percent cost reduction), Fig. 4.4.8b compares the number of deadheaded vehicle trips, and Fig. 4.4.8c
compares the number of skipped stations, between System G and System F by data set. The 1-DH solutions (with adjacent deadheading allowed in both Systems) are used for these comparisons. They show very interesting differences between the two Systems.

![Graph 1](image1.png)

*Fig. 4.4.8a Comparison between System F and System G (1)*

![Graph 2](image2.png)

*Fig 4.4.8b Comparison between System F and System G (2)*
First, Fig. 4.4.8a shows that deadheading is much more effective in direction 1 than in direction 2 for System G; while the opposite is true for System F. Second, the number of deadhead vehicles is larger in System F in all but one data set. Third, the number of skipped stations per deadhead vehicle in Direction 1 is slightly larger in System G than in System F, while in Direction 2 it is much smaller.

Since both systems used exactly the same input data sets which contain identical dispatching time data for the same direction in the same morning peak, the same impact set size ($m=1$), and both allowing adjacent deadheading, why are these results so dramatically different? It is because there are a number of important factors that affect System G but not System F, namely the dwell time effects, the dispatching headway effects and the demand effects.

**Dispatching Headway and Dwell Time Effects**

In direction 1, the standard deviation of the dispatching headway is much smaller (an average of 1.3 minutes) than in direction 2 (an average of 3.65 minutes). Since in System F such variance does not increase along the route, only minor control is needed for direction 1. Consequently, the cost reduction and the number of stations skipped are smaller in this direction for System F. In System G, however, the small dispatching headway variance is amplified along the route due to the dwell time effects, as illustrated in Chapter 3, Fig.3.3. At the end of direction 1, this variance has grown to the dispatching headway variance of direction 2. Hence, stronger control is needed for
System $G$. On the other hand, due to the dispatching headway effects, an effort to even out headways across vehicles at the beginning of the direction has great effectiveness to reduce headway variance along the route. This explains why in direction 1 the control benefits are much greater for System $G$ than for System $F$, and why more stations were skipped in System $G$.

The reason for a smaller number of beneficial deadhead vehicle trips in System $G$ is also due to the dispatching headway effects. For example, suppose we have $h_{i-1} > h_i > h_{i+1}$ at $k_0$, such that deadheading vehicle $i$ is feasible. While in System $F$ this deadheading may be beneficial, in System $G$, $i$'s headway will decrease along the route even in the absence of control. In this case deadheading $i$ may not be beneficial if it causes large increase in $h_{i+1}$. Such a headway pattern is exactly the case of vehicle 15 in data set "w1" (see Appendix C), with the consequence that it is deadheaded in System $F$ but not in System $G$.

**Demand Pattern Effect**

Now the question is, why are the benefits of control in direction 2 much smaller in System $G$? This is mainly due to the demand pattern of direction 2. While variability of demand across stations has little impact on the quasiconvexity of the cost curves, it has direct impact on the length of deadhead segments. In Direction 2 all deadhead segments are much shorter in System $G$, in fact most of them have only 1 skipped station. In sharp contrast, an optimal deadhead segment in System $F$ can have as many as 8 skipped stations in direction 2. Because demand is invariant across stations in System $F$, the larger the headway ratio, the more likely the optimal deadhead segment will be longer. Since direction 2 has many paired or bunched vehicle trips and much larger headway ratios (again see Appendix C), the deadhead segment lengths are often longer than in direction 1 in System $F$.

In System $G$, while Direction 1 has low passenger arrival rates from station 1 to station 12 ($r_x < 3$ passengers/minute), the second station in Direction 2, station 28, has an extremely
high passenger arrival rate of 30 passengers/minute because it is a transfer station. To skip this station would definitely result in a very high cost. This limits almost all deadhead segments in Direction 2 to 1 skipped station (27) only. This shows that if a direction has high demand at its beginning, it is not effective to use a deadheading strategy, even when the direction has a higher headway variance.

The demand pattern effect makes a major difference between System F and System G. In System F, the higher headway variation (as in direction 2 of the Green Line data sets) leads to more severe and effective deadheading policies (i.e., longer deadhead segments) in the optimal RTDP solution. In System G, however, a certain demand pattern may prevent such effective deadheading policies in the optimal RTDP solution. The large improvement (15.23%-24.14% cost reduction) in System F compared to the insignificant improvement (0.88%-6.4%) in System G for the Green Line B line direction 2 is the result.

One implication of the above comparisons is that, when high demand stations are late in a direction and dispatching headway variation is not too high (e.g., C.O.V.<0.3), as in the direction 1 cases above, RTDPF and RTDPG have more in common. Although the effectiveness of control are very different, the optimal control policies are quite similar for both Systems in the direction 1 results shown in Fig. 4.4.8b and c. They also have a similar upper bound on deadhead segment lengths, as shown in Fig. 4.4.7. In this case RTDPF may be used to infer control policies for System G. However, when high demand stations are at the beginning of a direction, and/or dispatching headway variation is very high (e.g., frequent vehicle pairing and bunching as in the direction 2 cases above), RTDPF and RTDPG are totally different. In this case RTDPF can not replace RTDPG at all.

4.4.8 An Efficient Algorithm for RTDPG

Based on the above discussions, two properties leads to an efficient algorithm for RTDPG: the quasiconvexity of the deadheading cost curves, and the headway condition
\( h_{i,k}h_{i+1,k} > 1 \) at \( k_0 \) for a beneficial deadheading policy. Thus we propose an efficient algorithm below.

**Algorithm RTDPG**

For each vehicle set \( I_m \) in System \( G \): Let \( i \) denote the first vehicle in \( I_m \).

1. Set \( j = i + m \) with given impact set size \( m \), and compute no-control trajectories in the control direction for vehicles \( i \) to \( j \) using equations (4.4.1) to (4.4.4) and setting all \( y = 1 \). Because \( d_{i,k} \) and \( s_{i,k} \) defined in (4.4.2) and (4.4.3) are interdependent, the following equation sequence should be used to compute \( d_{i,k} \) and \( s_{i,k} \)

\[
\begin{align*}
(4.4.10) & \quad d_{i,k} = (a_{i,k} + c_0 - c_1 r_k d_{i-1,k} + c_2 q_k L_{i,k})/(1 - c_1 r_k) \\
(4.4.11) & \quad s_{i,k} = d_{i,k} - a_{i,k}
\end{align*}
\]

\( a_{i,k} \), \( P_{i,k} \), and \( L_{i,k} \) can be computed directly from equations (4.4.1), (4.4.4) and (4.4.5) respectively.

2. If \( h_{i,k}h_{i+1,k} > 1 \) at either \( k_0 \) or \( k_e \), compute the no-control cost \( f_i(0) \) using the objective function (4.2.2). Set \( \min\{f_i\} = f_i(0) \) and \( \arg\min\{f_i\} = k_0 \). Set \( k_e = k_0 + 1 \) and \( n = 1 \).

Otherwise, terminate and \( i \) is not deadheaded.

3. Set \( y_{i,k} = 0 \) for each \( k : k_0 \leq k \leq k_e \) and \( y_{i,k} = 1 \) otherwise. Compute the trajectories of vehicles \( i \) to \( j \) using equations (4.4.1), (4.4.4),(4.4.5) (4.4.10) and (4.4.11) for where \( y_{i,k} = 1 \). For where \( y_{i,k} = 0 \), simply set \( d_{i,k} = a_{i,k} \) while using (4.4.1), (4.4.4) and (4.4.11) to compute \( a_{i,k} \), \( P_{i,k} \) and \( s_{i,k} \) respectively. In the mean time, check the feasibility of the deadhead segment \([k_0, k_e]\) using constraint (4.2.4).

4. If \([k_0, k_e]\) is not a feasible deadhead segment, or \( k_e = k_0 + N/4 \), or \( k_e \) is a higher than average demand station, terminate. Otherwise compute \( f_i^{(n)} \). If \( f_i^{(n)} > f_i^{(n-1)} \), terminate;

Otherwise, if \( f_i^{(n)} < \min\{f_i\} \), set \( \min\{f_i\} = f_i^{(n)} \) and \( \arg\min\{f_i\} = k_e \); increase \( n \) and \( k_e \) by 1 respectively, and go back to Step 3.

5. The optimal deadhead segment is \( k_e^* = \arg\min\{f_i^{(n)}\} \).

**End**

---

\( \text{Equation (4.4.10)} \) is derived by substituting (4.4.2) into (4.4.3) and collecting terms.
We should note that because the computation time of a complete enumeration algorithm for program [RTDPG] is in the order of $N/2$, and in general $N$ is not a very large number in the U.S. transit systems, the advantage of Algorithm RTDPG over complete enumeration may not be obvious in a single iteration. However, when deadheading is used with other control strategies, many more iterations will be required to obtain an optimal or near-optimal control decision, as we will see later in Chapter 7 for the combined control. In this case the advantage of Algorithm RTDPG will be more obvious.
CHAPTER 5

THE REAL-TIME EXPRESSING PROBLEM

5.1 Introduction

This chapter studies real-time expressing, another commonly used but little studied control strategy in transit operations. Like deadheading, when a vehicle is expressed, it skips a number of stations in order to gain time and reduce the preceding headway. Unlike deadheading, a vehicle can be expressed from either a dispatching terminal or an intermediate station, and it does not have to run empty. The decision to express a vehicle is announced to passengers at the express initiation station, where passengers traveling to stations to be skipped must alight (or not board). On the other hand, passengers traveling to stations beyond the express segment will still board or stay on-board. While deadheading can save more time at the initiation station and bring less frustration to passengers, expressing benefits from partially (at least) utilizing vehicle capacity in the express segment and the flexibility of choosing the best initiation station. The real-time expressing problem (RTEP) is to decide which vehicle should be expressed from which station to which other station, such that the total passenger waiting time is minimized within a given boundary. Although the RTEP is in many respects similar to the RTDP, it is much more difficult to formulate and solve, due to the higher dimension of decision
variables (both initiation and ending station of a express segment), and the dependency of a decision on more intermediate variables such as passenger alightings.

We should note that the RTEP is different from the pre-planned expressing problem. In the pre-planned expressing problem, the expressing schedule and segments are planned off-line, and stay fixed for the planned period. In the RTEP, the express vehicles, segments, and schedule are decided in real time, depending on the dynamics of demand and vehicle operations.

In this study of RTEP, we are mainly interested in the advantages and disadvantages of expressing over deadheading. This chapter is organized as follows. In section 5.2, we develop a general model for the RTEP. In section 5.3, we study the RTEP in System F and compare it with RTDPF. In section 5.4, we study RTEP for System G and compare it with RTDPG. Computer algorithm and empirical guidelines for expressing are also presented.

5.2 The general Model of RTEP

Throughout the following discussion we assume:

A5.1 A vehicle can be expressed at most once in a direction.

This assumption is both realistic and helpful in model and algorithm development. As we discussed in previous chapters, passengers do not like to see a vehicle passing by without stopping, and we should use station skipping strategies with caution. In practice, it is rare that a vehicle is expressed more than once on a round trip, although here we only restrict it in a single direction trip.

Due to the similarity between the RTEP and the RTDP, the decision variables and objective functions are defined exactly the same as in equations (4.2.1) and (4.2.2) respectively. Hence we do not repeat them here. In this chapter we will use the concepts
of *express segment* \([k_i, k_j]\), which is defined similarly to the deadhead segment \([k_0, k_e]\) (see Definition 4.1 in Chapter 4), except that for the express initiation station we replace \(k_0\) by \(k_s\), where \(k_s\) can take any value in the range \(k_0\) to \(k_e\). Since a meaningful express segment contains at least one skipped station, and \(k_s\) is not skipped, we must have \(k_s < k_e - 2\). The largest \(k_s\) can be station \(k_i\) (the ending terminal in the control direction), thus the largest \(k_s\) can only be \(k_e - 2\).

The definition of an *expressible vehicle*, a *feasible express segment*, and a *feasible express policy* are also defined the same as in the RTDP (see Definition 4.2 in Chapter 4), except we replace \([k_0, k_e]\) by \([k_s, k_j]\), and replace \(k_0 < k_e\) by \(k_0 \leq k_s < k_e - 1\).

We now present a general formulation of the RTEP. This general model represents the RTEP in both System \(F\) and System \(G\), given appropriate variable definitions for each system.

[RTEP] Minimize (4.2.2)

Subject to:

(5.2.1) \(a_{i,k}(y) - d_{i,1,k}(y) \geq h_0, \forall (i,k) \in I_m \times K_e\)

(5.2.2) \(\sum_{k=k_0+1}^{k_i} (y_{i,k} - y_{i,k}y_{i,k-1}) \leq 1, \forall i\)

(5.2.3) \(y_{i,k_0} = y_{i,k_i} = 1, \forall i \in I_m\)

(5.2.4) \(y_{j,m,k} = 1, \forall k \in K_e\)

(5.2.5) \(y_{i,k} \in \{0,1\}, \forall (i,k) \in I_m \times K_e\)

where \(j\) is the first vehicle in \(I_m\); \(y_{i,k}\) is defined by equation (4.2.1); and, \(P_{j-1}, a_{j-1,k} \forall k \in K_e\), \(a_{i,k_0} \forall i \in I_m\), as well as \(k_0\), are given.

While this program looks a lot like [RTDP] in the previous chapter, there are a couple of major differences between them. In [RTDP], constraint (4.2.7) prevents any intermediate station from being an initiation station of a deadhead segment; this constraint is not present in [RTEP]. Instead, constraint (5.2.2) requires that there be at most one express segment for each vehicle in a control direction, and constraint (5.2.3) rules out the possibility that either terminal station in the control direction will be skipped.
We next show that constraint (5.2.2) indeed allows there to be at most one express segment.

There are four possible pairing of \( y_{i,k} \) and \( y_{i,k-1} \): (0,0), (0,1), (1,1) and (1,0). The only case that \( y_{i,k} \) and \( y_{i,k} y_{i,k-1} \) are different is \( y_{i,k} = 1 \) and \( y_{i,k-1} = 0 \). Since \( y_{i,k} = 1 \) by (5.2.3), if \( i \) is not expressed at all, we have \( y_{i,k} = y_{i,k} y_{i,k-1} = 1 \) for all \( k \). In this case the result of the left-hand side of (5.2.2) is 0 and (5.2.2) holds. For any single express segment of \( i \), the only difference between \( y_{i,k} \) and \( y_{i,k} y_{i,k-1} \) is at the ending station \( k_e \), where \( y_{i,k} = 1 \), \( y_{i,k-1} = 0 \) so \( y_{i,k} y_{i,k-1} = 0 \), and the left-hand side of (5.2.2) equals 1. Hence if \( i \) is expressed once (5.2.2) also holds. Now if \( i \) is expressed \( t > 1 \) times, the left-hand side of (5.2.2) will be \( t > 1 \), and in this case (5.2.2) is violated. Thus inequality (5.2.2) holds if and only if \( i \) is expressed no more than once.

We now give a more formal proof of this. What we want is that the number of changes in \( y_i \) is less than or equal to 2. That is, at most one change from 1 to 0 with another one change back from 0 to 1. This can be written as follows:

\[
(5.2.6) \sum_{k=k_0+1}^{k_i} |y_{i,k} - y_{i,k-1}| \leq 2
\]

Now, since \( y_{i,k} \in \{0,1\} \ \forall (i,k) \), it follows that \( |y_{i,k} - y_{i,k-1}| = (y_{i,k} - y_{i,k-1})^2 \). So (5.2.6) can be re-written as:

\[
(5.2.7) \sum_{k=k_0+1}^{k_i} (y_{i,k} - y_{i,k-1}) \leq 2
\]

\[
\leq 2 \sum_{k=k_0+1}^{k_i} \left( y_{i,k}^2 - 2y_{i,k}y_{i,k-1} + y_{i,k-1}^2 \right) \leq 2
\]

\[
\leq 2 \sum_{k=k_0+1}^{k_i} (y_{i,k} - 2y_{i,k}y_{i,k-1} + y_{i,k-1}) \leq 2, \text{ since } y_{i,k} = y_{i,k'}.
\]

Expanding the above yields

\[
(5.2.8) (y_{i,k_0+1} - 2y_{i,k_0+1}y_{i,k_0} + y_{i,k_0}) + (y_{i,k_0+2} - 2y_{i,k_0+2}y_{i,k_0+1} + y_{i,k_0+1}) + \ldots
\]

\[
\ldots + (y_{i,k} - 2y_{i,k}y_{i,k-1} + y_{i,k-1}) \leq 2, \text{ or }
\]

138
(5.2.9) $y_{i,k_0} - y_{i,k_t} + \sum_{k=k_0+1}^{k_t} (2y_{i,k} - 2y_{i,k}y_{i,k-1}) \leq 2$

Now it follows from constraint (5.2.3) that $y_{i,k_0} - y_{i,k_t} = 0$, and hence (5.2.9) yields (5.2.2).

In the discussion that follows, the definition of the 1-EXP subproblem is the same as the definition for 1-DH except that "deadheading" is replaced by "expressing" (see Definition 4.3 in Chapter 4). Next we look at the RTEP in System $F$.

### 5.3 Expressing in System $F$

In Chapter 4 we showed that the RTDPF has a closed form solution when rolling horizon size $m=1$. In this section we are interested in whether such a solution exists for RTEPF, and what are the differences and similarity between them. Because the 1-DH subproblem without adjacent deadheading is the most fundamental problem in RTDPF, and likewise the 1-EXP subproblem without adjacent expressing is also the most fundamental problem in RTEPF, our comparison will focus on this basic problem. Therefore for the 1-EXP subproblem in System $F$ analyzed below we also assume that, similar to assumption A4.1, no adjacent vehicles can be expressed in the same direction.

Because passenger alighting information is needed to solve RTEP, in keeping with the fixed parameter characteristics of System $F$, we also assume a constant passenger alighting ratio $q$, $0 < q < 1$, so that $q_k = q$ at all $k$ except at station $k_0$, where $q_{k_0} = 0$, and station $k_r$, where $q_{k_r} = 1$. The constant alighting ratio assumption is probably not very realistic, however, solving such a special case may provide interesting insights for the more general case.

#### 5.3.1 The 1-EXP Model and Properties

The definitions of $a_{i,k}$ and $d_{i,k}$ are exactly the same in RTEPF as in RTDPF. The definition of leftover passengers $P_{i,k}$ is also the same if $k$ is not the express initiation station.

However, one piece of information required for making an expressing decision which is
not required for deadheading is the number of passengers who can not board or are forced off a vehicle at the express initiation station \(k\). Suppose vehicle \(i\) is expressed from station \(k\) to station \(k+n+1\). If there were no control, the normal vehicle departure load at station \(k\) is \(L_{i,k}\). Given the passenger alighting function \(A_{i,k} = q_k L_{i,k}\) with no control, there would be \(L_{i,k} q_k + 1\) passengers from station \(k\) going to station \(k+1\), \(L_{i,k} (1-q_k) q_k + 2\) going to station \(k+2\), and \(L_{i,k} (1-q_k) q_k + 2\) going to station \(k+n\). Hence when the vehicle arrives at station \(k = k+n+1\), the number of passengers from station \(k\) still on board is \(L_{i,k} \Pi_{j=k+1}^{k+n+1} (1-q_j)\). Thus, the total number of normal alightings at the skipped stations from \(k+n+1\) to \(k+n\) is \(L_{i,k} [1 - \Pi_{j=k+1}^{n} (1-q_j)]\). We define that

\[
(5.3.1) \quad \gamma_{k,k'} = \Pi_{j=k+1}^{k'-1} (1-q_j) \quad \text{if} \quad k' > k+1, \quad \text{and} \quad \gamma_{k,k} = 1 \quad \text{otherwise}
\]

It then follows that the portion of \(L_{i,k}\) going to \(k = k+n+1\) and beyond is \(\gamma_{k,k}\).

The departure load \(L_{i,k}\), when \(i\) is expressed on \([k,s]\) then is \((L_{i,k} + r_k h_{i,k}) \gamma_{l_{k,k}}\). It follows that given \(q \equiv q\) for \(k_0 < j < k_n\), \(\gamma_{l_{k,k}} = (1-q)^n\), where \(n = k - k_0 - 1 \geq 0\), which is the number of skipped stations. Note that the \(n\) skipped stations will never include \(k_0\) and \(k_n\) due to constraint (5.2.3). Furthermore, when not controlled the departure load of \(i\) at \(k\) is

\[
(5.3.2) \quad L_{i,k}^0 = r h_i^0 [1+(1-q)+\ldots+(1-q)^{n-1}]/q, \quad n'=k-k_0.
\]

Thus for System \(F\) we can write the number of leftover passengers at any initiation station \(k\) as

\[
(5.3.3) \quad P_{i,k} = (1-\gamma_{k,k+n+1}) L_{i,k}^0 = (1-q)^n (r h_i^0 [1-(1-q)^{n+1}]/q) = r h_i^0 [1-(1-q)^{n+1}][1-(1-q)^n]/q, \quad k=k_s
\]

If \(k = k_s = k_0\), we have \(n'=k-k_0=0\), and (5.3.3) reduces to

\[
(5.3.4) \quad P_{i,k} = r h_i^0 [1-(1-q)^n], \quad k=k_0
\]

Because of the separability of System \(F\), we can first look at each individual vehicle to solve the expressing problem. As for the 1-DH problem, we define the 1-EXP problem as the RTEP for an \(I_m\) group where \(m=1\). In light of [1-DH] (see section 4.3) and equation (5.3.3), we can write the model for 1-EXP as follows:
[1-EXP] Minimize

\[(5.3.5) \quad f(n', n) = P_{i,k} h_{i+1}^0 / r + n h_i^0 h_{i+1}^0 - ((1+u_c)lK'_{i-l-n'2-1})(h_i^0 h_{i+1}^0 - n\Delta) n\Delta \]

Subject to: (4.3.18) and

\[(5.3.6) \quad n' + n \leq N/2-2 \]

\[n', n \in \mathbb{Z}^* \]

The only difference between (5.3.5) and the [1-DH] objective function (4.3.17) is that (5.3.5) has an additional non-negative term \(P_{i,k} h_{i+1}^0 / r\) and the number of beneficial stations is \(n' - 1\) less. Here the first term is the portion of 1-EXP cost at station \(k_x\), the second term is the cost at the skipped stations, and the third term is the cost on the benefit segment. It is obvious from (5.3.5) and (5.3.2) that \(f(n', n)\) is increasing in \(n'\) regardless of the value of \(n\), since both \(P_{i,k}\) is increasing in \(n'\) and the absolute value of the only negative term, \((lK'_{i-l} + u_c - n'2-1)\), is decreasing in \(n'\). This means the costs in both the express segment and the benefit segment are increasing with \(n'\) increasing. This is quite intuitive because \(n'\) is the number of stations before the express segment and hence a larger \(n'\) means a later \(k_x\). In System \(F\), the cost portion at skipped stations depends solely on \(n\) regardless of \(k_x\), as the second term in \(f()\) reflects. However, the later is \(k_x\), the shorter is the benefit segment, and the larger the vehicle load at \(k_x\). Thus fewer passengers could save waiting time while more passengers would be left behind by the express vehicle at \(k_x\). Therefore, the best initiation station for expressing in System \(F\) is the first station, \(k_0\), and we can drop \(n'\) from [1-EXP] and rewrite it as

[1-EXP'] Minimize

\[(5.3.7) \quad f(n) = b(n+1-(1-q)^n)-(K-n-1)(a-n\Delta)n\Delta \quad n \in \mathbb{Z}^* \]

Subject to: (4.3.18) and (4.3.19)

where \(a = h_i^0 - h_{i+1}^0\), \(b = h_i^0 h_{i+1}^0\) and \(K = (1+u_c)lK'_{i-l}\).

In equation (5.3.7) we have incorporated equation (5.3.4). Equation (5.3.7) has an additional positive term \(b[1-(1-q)^n]\) compared to (4.3.17). Hence the necessary
conditions (4.3.21) and (4.3.22) for a beneficial policy stated in Proposition 4.1 also hold for [1-EXP]. Expanding (5.3.7) and relaxing the integer constraint yields:

\[(5.3.7') f(n) = b[1-(1-q)^n]- \Delta^2 n^3 + [a\Delta+(K-1)\Delta^2]n^2 + [b-a\Delta(K-1)]n, \ n \in \mathbb{R}^+\]

The first order derivative of (5.3.7') is

\[(5.3.8) \quad df/dn = -b(1-q)^n \ln(1-q)-3\Delta^2 n^2 + 2[a\Delta+(K-1)\Delta^2]n + b-a\Delta(K-1)\]

And the second order derivative of (5.3.7') is

\[(5.3.9) \quad d^2f/dn^2 = -b(1-q)^n[(\ln(1-q))^2-6\Delta^2 n + 2[a\Delta+(K-1)\Delta^2]]\]

\[\hspace{1cm} = -b(1-q)^n[(\ln(1-q))^2+2\Delta(a+(K-1)\Delta-3n)]\]

It follows immediately that:

**Proposition 5.1**

The real relaxation of [1-EXP] in System F is a convex program if and only if

\[(5.3.10) \quad -b(1-q)^n[(\ln(1-q))^2+2\Delta(a+(K-1)\Delta-3n)] \geq 0.\]

How strict is condition (5.3.10) in reality? While the second part of the left-hand side of (5.3.10) will be positive when \(n<(a/(\Delta+K-1))/3\), which is easily satisfied in reality as we have shown in Chapter 4, the first part is always negative. But because the negative term usually has a very small absolute value, the real relaxation of [1-EXP'] is still likely to be convex. For example, if \(1-q=0.5\) and \(n=4\), \(-(1-q)^n[(\ln(1-q))^2]=-0.03.\) Such a small value is unlikely to result in a negative \(d^2f/dn^2\). Hence, we expect that (5.3.10) will be easily satisfied in reality. Whether [1-EXP'] is convex or not, an optimal \(n\) can be easily found using Algorithm RTDPF with slight modifications.
We next compare (5.3.9) with (4.3.17) to explore the advantages or disadvantages of expressing over deadheading.

5.3.2 Comparison with Deadheading

We rewrite the 1-DH objective function (4.3.17) as follows:

\[ f_i^d(n) = bn - (K-n)(a-n\Delta)n\Delta \]

where the superscript 'd' stands for 'deadheading'. Below we will also use the superscript 'e' to indicate 'expressing'. When \( n \) is the same in both (5.3.7) and (5.3.11), subtracting (5.3.7) from (5.3.11) yields

\[ f_i^e(n) - f_i^d(n) = b[n+1-(1-q)^n] - (K-n-1)(a-n\Delta)n\Delta - (bn-(K-n)(a-n\Delta)n\Delta) \]

\[ = 1-(1-q)^n + (a-n\Delta)n\Delta > 0 \]

This shows that the closer \( q \) is to zero, the smaller the difference. Also, the closer \( q \) is to 1, the larger the difference is, although the difference is bounded from above by \( 1 + (a-n\Delta)n\Delta \) and from below by \( (a-n\Delta)n\Delta \). This shows that there is a limited difference between deadheading and expressing, and deadheading has advantages over expressing in System F, provided they have the same optimal number of skipped stations \( n \). This is not surprising. First, deadheading has one more benefit station because its first skipped station is \( k_0 \), while for expressing it is \( k_0+1 \). This results in the positive term \( (a-n\Delta)n\Delta \).

Second, because \( k_0 \) is not a skipped station for expressing and there are non-negative leftover passengers at \( k_0 \), passenger waiting time at \( k_0 \) is larger than in the no-control case. On the other hand, in the deadheading case, \( k_0 \) is one of the \( n \) skipped station. When the optimal \( n \) is the same for both expressing and deadheading, costs for the skipped segment are also the same. This leaves an extra non-negative cost \( 1-(1-q)^n \) at \( k_0 \) in the expressing case. Thus in both segments \([k_0, k_0+n]\) and \([k_0+n+1, k]\) expressing results in higher cost.

Of course the above comparison will not be valid if the optimal \( n \)'s are different for expressing and deadheading (under the same no-control conditions). Because expressing
has one less benefit station, [1-EXP] will not result in a larger \( n \) than [1-DH] other things being equal. Hence, the upper bound of \( n \) (4.2.23) stated in Proposition 4.2 also holds for [1-EXP']. It is possible though, that \( n_c = n_d - 1 \). In this case they both have exactly the same number of benefit stations. If the optimal \( n_c \) is equal to \( n_d \), which is very likely, then we have \( f_i^c(n_d+1) > f_i^d(n_c) > f_i^d(n_d) \). Now suppose the optimal \( n_c = n_d - 1 \), subtracting (5.3.7) from (5.3.11) yields:

\[
\begin{align*}
(5.3.13) \quad f_i^c(n_c) - f_i^d(n_d) &= b[n_c+1-(1-q)^n] - (K-n_c)(a-n_c\Delta)n_c\Delta - b(n_d-K-n_d)(a-n_d\Delta)n_d\Delta \\
&= b[n_d-(1-q)^n] - (K-n_d)(a-(n_d-1)\Delta)(n_d-1)\Delta - b(n_d-K-n_d)(a-n_d\Delta)n_d\Delta \\
&= -b(1-q)^n - (K-n_d)((a-(n_d-1)\Delta)(n_d-1)\Delta - (a-n_d\Delta)n_d\Delta) \\
&= -b(1-q)^n - (K-n_d)((2n_d-1)\Delta^2 - a\Delta)
\end{align*}
\]

To see whether \( f_i^c(n_c) < f_i^d(n_d) \), we first look at whether \((2n_d-1)\Delta > a\) since \( \Delta > 0 \). Since \( n_d \) is optimal, by Proposition 4.2, we have \((2n_d-1)\Delta < a\), hence the second term \(- (K-n_d)((2n_d-1)\Delta^2 - a\Delta) \) is positive. Now whether \( f_i^c(n) < f_i^d(n) \) depends on the difference between \( b(1-q)^n \) and \(- (K-n_d)((2n_d-1)\Delta^2 - a\Delta) \). Because \( 0 < 1 - q < 1 \), \( (1-q)^n \) gets very small as \( n \) becomes large. On the other hand, if \( n \) is small, the multiplier of the second term \( K-n_d = K-n_c + 1 \) becomes large. Thus in either case it is very unlikely that \( b(1-q)^n \) would have a larger value than \(- (K-n_d)((2n_d-1)\Delta^2 - a\Delta) \). This shows that \( f_i^c(n_c) - f_i^d(n_d) \) is likely to be positive and hence that deadheading is likely to result in lower cost than expressing.

To illustrate this point further, we next solve a small 1-EXP problem using the same input data as in Example 4.3.1 (See Chapter 4).

**Example 5.3.1**

Suppose \( h_i = 9, \ h_{i+1} = 3, \ \Delta = c_0 + 2\delta = 1, \ k_x = k_y = 1, \ k_r = N/2 = 11 \), no layover time at \( N/2 \) and \( u_c = 1 \), we have \( |K| = 11, \ |K'| = 10, \ K = (1 + u_c)|K'| = 20, \ b = h_{i+1}^0 = 27 \) and \( a = h_i - h_{i+1} = 9 - 3 = 6 > 0 \). To see the impact of passenger alighting ratio on expressing results, we assume two different values: \( q = 0.6 \) and \( q = 0.2 \). Following Example 4.3.1, we know \( n < 3.5 \). With those input data we have
\[ f(n) = b[n+1-(1-q)n] - (K-n)(a-n\Delta)n\Delta = 27[n+1-0.4^n] - (19-n)(6-n)n \]

It can be easily verified that \( d^2f/dn^2 > 0 \) when \( n < 4 \). Thus \( f(n) \) is convex over the feasible region. For illustration purpose, we enumerate \( n \) from 1 to 3 to find the \( n^* \) that results in the minimal \( f(n) \). Table 5.3.1 shows the computed values of \( f(n) \):

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>q</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-46.80</td>
<td>-50.68</td>
<td>-34.27</td>
<td>0.60</td>
</tr>
<tr>
<td>2</td>
<td>-57.60</td>
<td>-72.28</td>
<td>-49.82</td>
<td>0.20</td>
</tr>
<tr>
<td>3</td>
<td>-68.00</td>
<td>-90.00</td>
<td>-72.00</td>
<td>N/P</td>
</tr>
</tbody>
</table>

Table 5.3.1. Example 5.3.1: Comparison between Deadheading and Expressing in System F

The optimal number of skipped stations is 2 for both values of \( q \), which is the same as in the deadheading solution in Example 4.3.1, but the resulting cost reduction from deadheading was the highest. This is shown in the last row of Table 5.3.1. Table 5.3.1 also shows that, when \( q \) is smaller, \( f^c(n) \) is closer to \( f^d(n) \).

To summarize, there seems to be little advantage to expressing over deadheading, but the differences between them are very small. Because of such small differences, one would also expect similarities between 1-EXP and 1-DH subproblem with adjacent deadheading and adjacent expressing, respectively, and between \( m \)-EXP and \( m \)-DH subproblems with \( m > 1 \). Hence we will not repeat those similar analyses we did in Chapter 4. However, there is another issue, the riding time saving issue, which was not raised in Chapter 4 that needs to be addressed here. We discuss the riding time saving issue below before we proceed to test the above results by solving all 1-EXP subproblems using the Green Line data.

**Riding Time Saving in Expressing**

We note that the above analysis may be biased because we did not include in the objective function the benefit of reduced passenger riding time from expressing. In System \( F \), such a benefit is \(-L^0_{i,k}(1-q)^n\Delta\), where \( k \) is the express initiation station, and \( L^0_{i,k} \).
is the no-control departure load at \( k \) given by equation (5.3.2). When riding time cost is included in the minimization, the objective function for a 1-EXP problem is in the following form:

\[
(5.3.12) \quad f(n', n) = -h_i[1-(1-q)^{n+1}](1-q)^n n\Delta/q + b[n+1-(1-q)^{n}-(K-n-1)(a-n)\Delta]n\Delta
\]

In this objective function the constant arrival rate \( r \) has been dropped out. The first term is the riding time saving of passengers on board through the express segment, and the remaining terms are the same as in objective function (5.3.7). Here the other variable \( n' \) comes into play, which represents the location of the express initiation station \( k_s \).

Obviously the optimal value of \( f \) depends on both \( n' \), the number of stations before \( k_s \), and the number of skipped stations \( n \). From (5.3.12) we can see that while \( L_{l,k} \) gets bigger, and hence the number of riding passengers through the express segment gets larger as \( n' \) increases, the number of benefit stations (in terms of waiting time) \( K-n-n'-1 \) gets smaller. This means there is a tradeoff between passenger waiting time benefits and riding time benefits in this objective function. Such a tradeoff may alter the location of \( k_s \), and the contribution of each cost component (i.e., riding time saving vs. waiting time saving). To examine these issues, let us call the expressing problem with riding time included in its objective function RTEPF\(_2\). The 1-EXP subproblem in RTEPF\(_2\) is defined as follows:

[1-EXP\(_2\)] Minimize (5.3.12)
Subject to: (4.3.18) and (5.3.6)

In the next subsection, we will solve both problems RTEPF and RTEPF\(_2\), and compare benefits between deadheading and the two expressing problems.

5.3.3 Computational Results with Green Line Data

The Green Line data sets are used to obtain the following computational results. All inputs are the same as in the deadheading case, except we add the constant alighting ratio \( q=0.17 \), which is the average alighting ratio of the Green Line B line in morning peak hours over stations other than terminals. The algorithm "Non-Adjacent Deadheading"
presented in Chapter 4 for is modified (with 1-DH replaced by 1-EXP) and used here for system optimization.

**Results for RTEPF**

The computational results for RTEPF, where objective function (5.3.7) is used for each 1-EXP subproblem and hence only passenger waiting time is minimized, are shown in Table 5.3.2. Table 5.3.3 compares these results with the deadheading results given in Chapter 4.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>M</th>
<th>Cost (psg.min)</th>
<th>StDev of h at k₀</th>
<th>#skipped stations</th>
<th># exp. veh.</th>
<th>Change in Cost</th>
<th>% Change in Cost</th>
<th>StDev of h at k₀</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>36</td>
<td>49,114.70</td>
<td>1.32</td>
<td>11</td>
<td>10</td>
<td>-492.76</td>
<td>-1.00</td>
<td>1.08</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>50,409.10</td>
<td>1.13</td>
<td>11</td>
<td>9</td>
<td>-876.39</td>
<td>-1.74</td>
<td>0.70</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>55,021.20</td>
<td>1.66</td>
<td>12</td>
<td>9</td>
<td>-928.13</td>
<td>-1.69</td>
<td>1.30</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>55,002.80</td>
<td>1.39</td>
<td>15</td>
<td>8</td>
<td>-1,643.30</td>
<td>-2.99</td>
<td>0.85</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>43,244.70</td>
<td>0.99</td>
<td>7</td>
<td>6</td>
<td>-513.29</td>
<td>-1.19</td>
<td>0.69</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>38,361.50</td>
<td>3.94</td>
<td>46</td>
<td>12</td>
<td>-6,659.35</td>
<td>-17.36</td>
<td>2.29</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>37,232.50</td>
<td>3.53</td>
<td>47</td>
<td>13</td>
<td>-5,905.11</td>
<td>-15.86</td>
<td>1.80</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>39,846.70</td>
<td>4.00</td>
<td>36</td>
<td>9</td>
<td>-5,388.60</td>
<td>-13.52</td>
<td>2.66</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>38,018.00</td>
<td>3.49</td>
<td>39</td>
<td>11</td>
<td>-5,006.76</td>
<td>-13.17</td>
<td>2.18</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>30,820.70</td>
<td>3.27</td>
<td>40</td>
<td>11</td>
<td>-4,481.95</td>
<td>-14.54</td>
<td>1.71</td>
</tr>
<tr>
<td>Total</td>
<td>346</td>
<td>437,071.90</td>
<td>2.64</td>
<td>264</td>
<td>98</td>
<td>-31,895.64</td>
<td>-7.30</td>
<td></td>
</tr>
<tr>
<td>Avg.</td>
<td></td>
<td>87,414.38</td>
<td>2.47</td>
<td></td>
<td></td>
<td>-6,379.13</td>
<td>1.53</td>
<td></td>
</tr>
</tbody>
</table>

**Note** $r = 2.39$, $q = 1.7$, $H = 5$ min. Avg. cost is per morning peak.

**Table 5.3.2 System F: Expressing Results from Green Line Data**

The following observations from the computational results confirm our analysis in the previous section:

(1) Slightly less vehicles are controlled in RTEPF. Compared to deadheading, 6 fewer vehicles are expressed and 14 fewer stations are skipped over 5 weekday mornings. This is because expressing results in slightly higher optimal waiting time than deadheading. Otherwise, both expressing and deadheading cases have in common 98 controlled vehicle trips.
Table 5.3.3 System F: Comparison between Expressing and Deadheading (1)

(2) In all 1-EXP subproblems, the total number of skipped stations, \( n \), is either the same or less than when deadheading is performed. In fact, among the common subset of 98 controlled vehicles, only 6 expressed vehicles skipped one less stations than if they are deadheaded, and all the others skipped the same number of stations in both cases. All 1-EXP subproblems result in slightly higher cost than 1-DH.

(3) As expected, all real relaxations of [1-EXP] are convex.

Results for RTEPF\(_2\)

To investigate the impact of including riding time in the expressing problem, we also computed the results for RTEPF\(_2\), where the objective function is to minimize both passenger waiting time and riding time. We compare the results with and without the riding time term in the objective function to deadheading results in Table 5.3.4.

The RTEPF\(_2\) results have the following interesting structure:

1. As shown in Table 5.3.5, in most 1-EXP\(_2\) subproblems, the express initiation station is not the dispatching station. This is a consequence of the tradeoff between waiting and
riding time savings as we analyzed earlier. Such a tradeoff is heavily influenced by the dispatching headway pattern. As seen from Table 5.3.5, in Direction 1 station 8 is the most "popular" initiation station, while station 1 is chosen only in one case. In Direction 2 the first station, 27, is the initiation station in a majority of cases. This is because the dispatching headway variance in Direction 1 is much smaller, in other words, most of the dispatching headways are not too large. Thus starting to express at a later station will produce more riding passengers who benefit from riding time savings, while not losing much waiting time savings.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>RTEPF</th>
<th>RTEPF₂</th>
<th>Dead- heading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wait time reduction</td>
<td>ride time reduction</td>
<td>total reduction</td>
</tr>
<tr>
<td>m1</td>
<td>492.76</td>
<td>82.48</td>
<td>575.24</td>
</tr>
<tr>
<td>tu1</td>
<td>876.39</td>
<td>84.43</td>
<td>960.81</td>
</tr>
<tr>
<td>w1</td>
<td>928.13</td>
<td>95.26</td>
<td>1,023.39</td>
</tr>
<tr>
<td>th1</td>
<td>1,643.30</td>
<td>100.72</td>
<td>1,744.02</td>
</tr>
<tr>
<td>f1</td>
<td>513.29</td>
<td>50.07</td>
<td>563.35</td>
</tr>
<tr>
<td>m2</td>
<td>6,659.35</td>
<td>304.32</td>
<td>6,963.67</td>
</tr>
<tr>
<td>tu2</td>
<td>5,905.11</td>
<td>296.75</td>
<td>6,201.86</td>
</tr>
<tr>
<td>w2</td>
<td>5,388.60</td>
<td>226.29</td>
<td>5,614.89</td>
</tr>
<tr>
<td>th2</td>
<td>5,006.76</td>
<td>244.92</td>
<td>5,251.69</td>
</tr>
<tr>
<td>f2</td>
<td>4,481.95</td>
<td>237.27</td>
<td>4,719.22</td>
</tr>
<tr>
<td>Total</td>
<td>31,895.64</td>
<td>1,722.51</td>
<td>33,618.14</td>
</tr>
<tr>
<td>%</td>
<td>94.88</td>
<td>5.12</td>
<td>100.00</td>
</tr>
</tbody>
</table>

Table 5.3.4 System F: Comparison between Expressing and Deadheading (2)

<table>
<thead>
<tr>
<th>Direction 1</th>
<th>kᵦ</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
<th>13</th>
<th>14</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td># exp. trips</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>4</td>
<td>2</td>
<td>6</td>
<td>5</td>
<td>15</td>
<td>2</td>
<td>5</td>
<td>4</td>
<td>2</td>
<td>1</td>
<td>49</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Direction 2</th>
<th>kᵦ</th>
<th>27</th>
<th>28</th>
<th>29</th>
<th>30</th>
<th>31</th>
<th>32</th>
<th>34</th>
<th>35</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td># exp. trips</td>
<td>35</td>
<td>8</td>
<td>3</td>
<td>5</td>
<td>6</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>62</td>
</tr>
</tbody>
</table>

Table 5.3.5 Optimal Initiation Station in 1-EXP₂ Subproblems
2. As seen from Table 5.3.4, including riding time in the objective function of RTEPF\textsubscript{2} changes the cost structure. Although in both RTEPF and RTEPF\textsubscript{2} solutions waiting time benefits dominate, the riding time portion is doubled as waiting time benefits decrease in the latter case. This raises the question whether or not it is appropriate to include riding time in the minimization. Previous research on passenger demand has shown that passengers always weigh waiting time much more heavily than riding time, thus it may not be justified to decrease waiting time benefits in order to increase riding time benefits. One possible modification of RTEPF\textsubscript{2} is to put a smaller weight on the riding time term. However, since riding time saving is already such a small portion of the total cost, this modification would result in little difference compared with minimizing passenger waiting time alone.

3. Both results of RTEPF and RTEPF\textsubscript{2} show lower total costs than deadheading. The total cost reduction in RTEPF\textsubscript{2} is higher than RTEPF, but not much: only about 1% higher. The inclusion of riding time benefits did not eliminate the advantage of deadheading, mainly because the riding time saving benefits passengers from one station only, while the waiting time saving is for passengers over many more stations. This is why in both RTEPF and RTEPF\textsubscript{2} waiting time benefits are dominant.

The above discussion also has implications for System G. Although very different systems, there are a couple of points in common for both Systems F and G. There is no question that including riding time term in the objective function means making a tradeoff between waiting and riding time benefits. If such a tradeoff is significant, it may not be justified to sacrifice much waiting time benefits. On the other hand, if the difference is small, there is little point to include the riding time term. This issue will be investigated further in the next section.
5.4 Expressing in System $G$

In System $F$ we have shown that the best initiation station for expressing is $k_0$ in terms of passenger waiting time minimization. This may not be true in System $G$ due to the variability of demand across stations. Thus in System $G$ we need to find both the optimal express initiation and ending stations. Although the only difference between deadheading and expressing is in the initiation station - both the location of it and passenger activities at it, this difference adds great complexity in the variable definitions of RTEPG (stands for "the RTEP in System $G$")

From the discussion on rolling horizon size in Chapter 4, we saw the choices of $m=1$ and $m>1$ both have their own advantages and disadvantages for a station skipping strategy, and for $m>1$, $m=3$ is a better choice than others. We also saw that deadheading solution structure with $m=3$ is similar to $m=1$ with a few exceptions. Since we had a very detailed discussion in Chapter 4 of different rolling horizon sizes, due to the similarity between expressing and deadheading, in this section we will not go through a similar set of analyses. Because 1-EXP is the most fundamental subproblem in the RTEP, and also because $m=1$ is the most attractive choice for a manual control system, in this section we focus on the 1-EXP subproblem. However, computational tests for $m=3$ will also be conducted later in this section to compare with the deadheading results. In this case the "control-the-first" method as discussed in Chapter 4 is used for both expressing and deadheading.

5.4.1 Variable Definitions

In the RTEPG we distinguish between two types of control stations in the express segment: the initiation station $k_i$ and a skipped station $k, k_i<k<k_e$. While the definitions of vehicle arrival time $a_{ik}$ and departure time $d_{ik}$ are the same as defined for RTDPG in Chapter 4, other variables $s_{ik}, L_{ik}$, and $P_{ik}$ may have different values at different stations. We give the new variable definitions below.

\[(5.4.1) s_{ik} = [c_0 + c_1 (P_{i-1,k} + r_{i,k}) y_{ik}] + c_2 A_{ik} \text{ (vehicle dwell time)}\]
(5.4.2) \( L_{i,k} = L_{i,k-1} - A_{i,k} + (P_{i-1,k} + r_k h_{i,k}) \gamma_{i,k} \) \( \gamma_{i,k} \) (vehicle departure load)

(5.4.3) \( P_{i,k} = y_{i,k}(1-y_{i,k})(1-q_k)L_{i,k-1} + (1-y_{i,k})y_{i,k}(P_{i-1,k} + r_k h_{i,k}) \) \( y_{i,k} \) (leftover passengers)

(5.4.4) \( y_{i,k} = \gamma_{k,k} \) if \( k = k_s \), \( y_{i,k} = 1 \) otherwise.

The explanation of the above expressions are as follows. A new variable \( \gamma_{k,k} \) (\( \leq 1 \) and \( \geq 0 \)), defined by equation (5.3.1) as the fraction of passengers at \( k_s \) going to stations \( k_e \) and beyond, is introduced. With this variable, the actual boardings at \( k_e \) may be less than normal because \( \gamma_{k_e,k_e} \leq 1 \). Furthermore, among the passengers aboard when the vehicle arrives at \( k_s \), those who wanted to go to the stations to be skipped must get off, thus the total alightings is \( L_{i,k-1} q_k + L_{i,k-1} (1-q_k)(1-\gamma_{k,k}) = L_{i,k-1} [q_k + (1-q_k)(1-\gamma_{k,k})] \). This explains (5.4.1) when \( k = k_s \). When \( k \neq k_s \), \( y_{i,k} = 1 \) and (5.4.1) yields \( s_{i,k} = 0 \) if \( y_{i,k} = 0 \), or \( s_{i,k} = c_0 + c_1 (P_{i-1,k} + r_k h_{i,k}) + c_2 A_{i,k} \) (the normal dwell time) if \( y_{i,k} = 1 \).

More about the variable definition for passenger alightings, \( A_{i,k} \), will be discussed shortly.

The meanings of (5.4.2) and (5.4.3) at different \( k \) are shown as follows.

When \( k = k_s \), we have \( \gamma_{i,k} = \gamma_{k,k} \) and \( y_{i,k} = 1 \). Thus (5.4.2) yields \( L_{i,k} = L_{i,k-1} - A_{i,k} + \gamma_{k,k} * (P_{i-1,k} + r_k h_{i,k}) \), which is the fraction of normal departure load with only passengers going beyond the skipped segment. (5.4.3) yields \( P_{i,k} = (1-y_{i,k})[(1-q_k) L_{i,k-1} + (P_{i-1,k} + r_k h_{i,k})] \), which is the fraction of normal departure load with passengers going to the skipped segment.

When \( k \) is a skipped station, we have \( y_{i,k} = 1 \) and \( y_{i,k} = 0 \). In this case (5.4.2) and (5.4.3) yield \( L_{i,k} = L_{i,k-1} \) and \( P_{i,k} = P_{i-1,k} + r_k h_{i,k} \) respectively, which means vehicle load does not change at a skipped station, and all waiting passengers are left behind by the express vehicle.

When \( k \) is not a control station, we have \( \gamma_{i,k} = 1 \) and \( y_{i,k} = 1 \). In this case (5.4.2) and (5.4.3) yield \( L_{i,k} = L_{i,k-1} - A_{i,k} + P_{i-1,k} + r_k h_{i,k} \) and \( P_{i,k} = 0 \) respectively, which means that vehicle departure load is the arrival load minus normal alightings plus all waiting passengers, and there is no leftover passengers at that station.
We note that $\gamma_{ik}$, like other intermediate variables such as $L_{ik}$, $P_{ik}$ and $s_{ik}$, is not an independent variable. It is completely determined by $\gamma_{ik}$'s. More precisely, we can define that

\[(5.4.5) \gamma_{ik} = 1 - (1 - \gamma_{k,i+1})(y_{ik} - y_{ik+1}), \text{ and} \]
\[(5.4.6) n = k_i - k - 1 - \sum_{k'=k+1}^{k_l} y_{ik} \]

$n$ is always non-negative by (5.4.6). For example, if we have $k = 3$ and $k_i = 10$, then $(y_{i,4} + \ldots + y_{i,10}) \leq 6 = 10-3-1$. It follows from equation (5.4.5) that, either $y_{ik} = y_{ik+1}$ or $n = 0$ will yield $\gamma_{ik} = 1$ because $\gamma_{kk} = 1$ by definition. If both $k$ and $k-1$ are skipped stations, or neither of them is a control station, we have $y_{ik} = y_{ik+1}$ and $\gamma_{ik} = 1$. If $k$ is a skipped station but $k+1$ is not, we have $n = 0$ and $\gamma_{ik} = 1$. Only if $y_{ik} > y_{ik+1}$, which means $k$ is the express initiation station, we have $y_{ik} - y_{ik+1} = 1$, $n > 0$ and $\gamma_{ik} = \gamma_{k,k+1}$. For example, suppose a vehicle starts expressing at station 3 and skips 3 stations 4,5, and 6. For $k = 3$ we have $y_{i,3} - y_{i,4} = 1$, $n = 3$ by (5.4.6), and $\gamma_{i,3} = 1 - (1 - \gamma_{i,6})(y_{i,5} - y_{i,4}) = 1 - (1 - \gamma_{i,6}) = \gamma_{i,6}$ by (5.4.5). For $k = 6$ we have $y_{i,6} - y_{i,7} = -1$, $n = 0$, and $\gamma_{i,6} = 1 - (1 - \gamma_{i,6})(y_{i,5} - y_{i,7}) = 1 - (1 - 1)(-1) = 1$. For any $k$ other than 3 and 6 we have $y_{ik} - y_{ik+1} = 0$ and hence $\gamma_{ik} = 1$.

**Passenger Alightings in RTEPG**

The number of passengers alighting from vehicle $i$ at station $k$, denoted as $A_{ik}$, can be simply defined by equation (5.4.7) below if the preceding vehicle $i - 1$ is not expressed:

\[(5.4.7) A_{ik} = L_{ik-1} q_k y_{ik} \gamma_{ik} \]

What (5.4.7) says is that, if $k = k_i$ (i.e., $y_{ik} > y_{ik+1}$), $\gamma_{ik} y_{ik} = \gamma_{ik,k}$, and $A_{ik} = L_{ik-1} q_k \gamma_{ik,k}$. If $k \neq k_i$, $\gamma_{ik} = 1$, we have $A_{ik} = 0$ if $y_{ik} = 0$, and $A_{ik} = L_{ik-1} q_k$ (i.e., normal alightings) otherwise.

When vehicle $i - 1$ is expressed, however, the definition (or the computation formulæ) for $A_{ik}$ becomes quite complicated. It gets even more complicated when more than one preceding vehicle of $i$ consecutively expressed. This is because after the leftover
passengers at the express initiation station of each express vehicle \( j < i \) board \( i \), they no longer follow the normal alighting function defined by equation (2.3), instead they all must alight \( i \) in the segment skipped by \( j \). The correct computation for \( A_{i,k} \) requires tracking all leftover passengers at each express initiation station and their destination station. Hence we extend the notation for express initiation and ending stations \( k_s \) and \( k_e \) to \( k_s(j) \) and \( k_e(j) \), respectively, to represent vehicle specific express segment. With this notation we express the general \( A_{i,k} \) with the following equations.

\[
A_{i,k} = \sum_{j=1}^{i-1} A_{j,i,k} + (L_{i,k-1} - L_{i,k-1}') q_k \gamma_{i,k} y_{i,k} \ \forall (i,k) \quad \text{(total alightings from \( i \) at \( k \))}
\]

\[
L_{i,k}' = \sum_{j=1}^{i-1} (e_{j,i,k} - \sum_{k'=k_s(j)}^{k_e(j)} A_{j,i,k'}) \ \forall (i,k) \quad \text{(total load of initial leftover passengers)}
\]

\[
e_{j,i,k} = y_{i,k_s(j)} \sum_{k'=k_s(j)}^{k_e(j)} A_{j,i,k'} \ \forall (j,i,k; j<i, k_k \geq k_s(j))
\]

\[
e_{j,i,k} = 0 \text{ otherwise} \quad \text{(initial leftover passengers from \( j \) who boarded \( i \) up to \( k \))}
\]

\[
A_{j,i,k} = P_{j,k_s(j)} y_{i,k} \prod_{j' \neq j} (1 - y_{i,j'} y_{j',k} (1 - \gamma_{j',k})/ (1 - \gamma_{j'k_s(j)})) q_k (1 - \gamma_{i,k}) \ \forall (j,i,k; j < i, k_s(j) < k < k_e(j))
\]

\[
A_{j,i,k} = 0 \text{ otherwise} \quad \text{(initial leftover passengers from \( j \) who board \( i \) and must alight at \( k \))}
\]

where \( k_1 = k_s(j) \), \( k_2 = k_e(j) \).

Equation (5.4.8) says that total alightings from vehicle \( i \) at station \( k \) are composed of two parts: the normal alighting passengers \((L_{i,k-1} - L_{i,k-1}')\), and the "initial leftover" passengers \( A_{j,i,k} \) who were forced off (or not let on) vehicle \( j \) at its express initiation station, then board \( i \) and must alight at \( k \) because \( k \) is a skipped station of \( j \). In the following we refer to \( A_{j,i,k} \) as "initial leftover" passengers. There may be such passengers from multiple preceding vehicles consecutively expressed. Thus \( L_{i,k-1}' \) represents all such passengers onboard \( i \) when it arrives at \( k \). Note that if \( y_{i,k} = 0 \), we also have \( A_{j,i,k} = 0 \) by equation (5.4.11).

Equation (5.4.9) defines \( L_{i,k}' \), the "initial leftover" passengers onboard \( i \) when it departs station \( k \), as the total boardings of such passengers from all preceding express vehicles minus their alightings up to station \( k \). Equation (5.4.10) defines \( e_{j,i,k} \), the total number of "initial leftover" passengers from vehicle \( j \) who have boarded \( i \) up to station \( k \). These
passengers could have boarded $i$ at a station $k' \leq k$ if and only if all of the following conditions are satisfied:

(i) $k'$ is the express initiation station of $j$, $k(j)$, and $k' \leq k$;
(ii) Both $k'$ and the destination stations of the passengers are not skipped by $i$.
(iii) There was no vehicle between $j$ and $i$ that has stopped at both $k'$ and $k$.

The first condition is defined in equation (5.4.10), and the other two are defined in equation (5.4.11). For the derivation of (5.4.11) and the definition of $\gamma_{i,k}$ refer to equations (5.4.4) and (5.3.1).

We illustrate the above formula with a simple example. Suppose vehicle $i$-1 expressed on the segment [1,4], i.e., it skips stations 2 and 3, and the number of passengers left by $i$-1 at station 1 is $P_{i-1,1} = 24$. Also assume $q_1 = 0$, $q_2 = 0.2$, $q_3 = 0.5$. Thus $\gamma_{k,\ell} = \gamma(1-0.2)(1-0.5) = 0.4$.

Now $i$ is not skipping stations 1, 2 and 3 and it follows from equations (5.4.10) and (5.4.11) that all the 24 passengers board $i$ at station 1. That is, for station $k=1$ and $k \geq 4$, we have $A_{i-1,k} = 0$, for station 2 we have $A_{i-1,2} = 24*1^*(1-0)*0.2*1/(1-0.4) = 8$ and for station 3 we have $A_{i-1,3} = 24*1^*(1-0)*0.5^*(1-0.2)/(1-0.4) = 16$. Hence by equation (5.4.10) we have $e_{i,k} = A_{i-1,2} + A_{i-1,3} = 8 + 16 = 24 = P_{i-1,1}$ for all $k \geq 1$.

We can verify the values of $A_{i-1,2}$ and $A_{i-1,3}$ as follows. Since $P_{i-1,1} = L_{i-1,1} (1-\gamma_{2,3}) = 0.6L_{i-1,1} = 24$, we know that $L_{i-1,1} = 40$. Hence $40*0.2=8$ passengers should alight at station 2, and $40-8)*0.5 = 16$ should alight at station 3. This is the same as we computed above.

Now let us assume that at station 1, the total number of passengers boarding vehicle $i$ is $L_{i,1} = 30 + P_{i-1,1} = 54$, including the 24 leftover passengers. At station 2, $L_{i,2} = P_{i-1,1} = 24$ and $A_{i-1,1} = 0$ by (5.4.9) and (5.4.11). At station 2, $A_{i-1,2} = 8$, $L_{i,2} = 24-8=16$, $A_{i,2} = 8+(54-24)*0.2 = 14$ by (5.4.8), and $L_{i,2} = 54-14=40$. At station 3, $A_{i-1,3} = 16$, $L_{i,3} = 16-16=0$, and $A_{i,3} = 16+(40-16)*0.5 = 28$. When $i$ arrives station 4, there are
40-28=12 passengers still onboard, and all 24 "initial leftover" passengers from the express initiation station of \(i-1\) are already off.

If equation (5.4.7) is used instead of equations (5.4.8) - (5.4.11) to calculate the alightings, we would end up having 11 alightings at station 2 and 22 at station 3 instead of 14 and 28 respectively. These wrong results may have insignificant impact on dwell times, but its impact on passenger riding times may be large because more passengers ride longer than they should. When only waiting time is concerned, one may use equation (5.4.7) to approximately compute vehicle movements, since the computation burden for the correct number of alightings is much higher in terms of both coding effort and variable value tracking. When riding time is also of a concern, however, equation (5.4.7) may give very inaccurate results.

The model formulation for RTEPG is the same as the general RTEP model with the above variable definitions. Next we investigate the 1-EXP subproblem in System \(G\).

### 5.4.2 Solution Structure of 1-EXP subproblem in RTEPG

Although a 1-EXP subproblem is more complicated than a 1-DH problem in System \(G\), essentially the only differences are in the cost at the initiation station \(k_s\) and the length of the benefit segment, \(k_s-k_s+1\). To see this, let us look at what a beneficial 1-EXP policy means when vehicle \(i\) is considered for expressing from a station \(k_s\). Let us divide the control direction into three segments: station \(k_s\) segment \([k_s+1,k_s-1]\), and segment \([k_s,k_s]\), which are the express initiation station, the skipped segment, and the benefit segment respectively. Letting \(\Delta w_i^j, j=1,2,3\), denote the cost difference with and without control for each of these three segments, we have

\[
\begin{align*}
(5.4.12a) \quad &\Delta w_i^1 = (0.5r_k h_{i,k_s} + P_{i-1,k_s})[h_{i,k_s}^0 - \Delta h_{i,k_s} + (1 - \gamma_{k_s,k_s})(h_{i+1,k_s}^0 + \Delta h_{i+1,k_s})] \\
&- (0.5r_k h_{i,k_s}^0 + P_{i-1,k_s})h_{i,k_s}^0 + 0.5r_k [(h_{i+1,k_s}^0 + \Delta h_{i+1,k_s})^2 - (h_{i+1,k_s}^0)^2] \\
&= (0.5r_k h_{i,k_s} + P_{i-1,k_s})[-\Delta h_{i,k_s} + (1 - \gamma_{k_s,k_s})(h_{i+1,k_s}^0 + \Delta h_{i+1,k_s})] \\
&+ 0.5r_k (2h_{i+1,k_s}^0 + \Delta h_{i+1,k_s})\Delta h_{i+1,k_s}
\end{align*}
\]
\[ \Delta w_i^2 = \sum_{k=\kappa_k+1}^{k-1} \{0.5r_k[(h_{i,k}^0 - \Delta h_{i,k} + h_{i+1,k}^0 + \Delta h_{i+1,k})^2 - (h_{i,k}^0)^2 + (h_{i+1,k}^0)^2]\] 
\[ + P_{i-1,k}(h_{i+1,k}^0 + \Delta h_{i+1,k} - \Delta h_{i,k})\} \]

\[ \Delta w_i^3 = \sum_{k=\kappa_k}^{k} \{0.5r_k[(h_{i,k}^0 - \Delta h_{i,k})^2 + (h_{i+1,k}^0 + \Delta h_{i+1,k})^2 - (h_{i,k}^0)^2 + (h_{i+1,k}^0)^2]\] 
\[ - P_{i-1,k}\Delta h_{i,k}\} \]

where \(\Delta h_{i,k} > 0, \Delta h_{i+1,k} > 0\) are the absolute values of headway changes.

Given \(\Delta w_i^1, \Delta w_i^2,\) and \(\Delta w_i^3,\) as defined above, a 1-EXP policy is beneficial if and only if \(\Delta w_i = \Delta w_i^1 + \Delta w_i^2 + \Delta w_i^3 < 0.\) From equation (5.4.12b), \(\Delta w_i^2\) is always positive and increasing with \(k - k_s.\) \(\Delta w_i^1\) is also likely to be positive, and its absolute value is small.

\(\Delta w_i^3\) is the only term which may have a significant negative value. Hence, when beneficial policies exist, one would expect that the optimal benefit segment contains large \(r_k's\) so as to result in a large negative \(\Delta w_i^3,\) and that the skipped segment contains small \(r_k's\) and \(q_k's\) so as to have a smaller positive \(\Delta w_i^2\) and \(\Delta w_i^1.\)

It is obvious that the forms of \(\Delta w_i^2\) and \(\Delta w_i^3\) are the same as those in a 1-DH subproblem, though \(k - k_s\) can be a later station than in 1-DH. From a waiting time point of view, the cost difference between \(k_s\) and a non-control station \(k\) is just the different number of waiting passengers, or different passenger demand. We have seen from Chapter 4 that neither variability of demand across stations nor the length of benefit segment changes the quasiconvexity of the 1-DH cost curves, and a 1-EXP subproblem with a given initiation station is expected to behave similarly with different size of \(K_c.\) Therefore, a 1-EXP subproblem can be viewed similarly as a \(\min\{1-DH(k_c)\}\) over all feasible stations \(k_c.\)

Hence, we expect that a 1-EXP subproblem with a given initiation station has the following characteristics similar to a 1-DH subproblem:

1. There exists an upper bound on the optimal number of skipped stations.
2. A cost curve of a vehicle over a segment is likely to be quasiconvex when the vehicle is not blocked in the segment.
3. The stations with high demand are not likely to be skipped.
These characteristics, if true, are much more useful in a 1-EXP subproblem than in a 1-DH subproblem. Because to solve a 1-EXP subproblem requires repeatedly searching for an optimal $k_s$ and a corresponding optimal $k_e$, an efficient search time for $k_e$ associated with each $k_s$ has a much more significant impact on the efficiency of the 1-EXP algorithm as a whole, and the first two characteristics make such an efficient search possible. Furthermore, the third characteristic allows us to reduce the range from which to choose a candidate $k_e$.

To verify the above characteristics, we again perform computational experiments on the Green Line data sets, where the vehicle trajectories are computed using the variable definitions given by equations (5.4.1)-(5.4.3) for $k=k_s$, together with the variable definitions given by equations (4.4.1)-(4.4.4) and equations (4.5.1) and (4.5.2) when $k \neq k_s$. To find the optimal express segment for each 1-EXP subproblem, we completely enumerate the initiation and ending stations for each vehicle trip, beginning with $k_s=1$ and $k_e=3$ and increment $k_s$ and $k_e$ as far as feasible.

We discuss the computational results below. Before interpreting the results for System $G$, we first want to note that one should not take a small difference (e.g., < 1%) in the overall costs as definitely meaningful, because such a small difference may be due to a "rolling optimization error" which may have nothing to do with control strategy types or something else. Such an error does not exist in System $F$ but does in System $G$. It is due to interaction between adjacent vehicle trajectories in System $G$. Optimization over a set of rolling horizons is not equal to exact optimization over the entire analysis period. That is to say, suppose we have found a set of "rolling optimal" control policies $\{u_1, u_2, ..., u_M\}$ for an entire set of $M$ vehicle trips, in which each $u_i$ is the optimal solution evaluated over a rolling horizon of $m<M$ vehicle trips, it is possible to obtain a better overall result for the entire set of $M$ vehicle trips by altering the value of some $u_i$. In general, there exists some "rolling horizon optimization error" from the "true" optimal systematic solution. But because of the smoothing effect of rolling optimization and the "trajectory change diminishing effect", such an error should be small. In our computational results for the
Green Line data sets, if cost difference is smaller than 1% in a data set, it may or may not be due to these rolling optimization errors. But if the difference is larger than 1%, it is likely to be due to more than just this type of error.

**Location of Express Initiation Station**

Among 346 vehicle trips (each direction represents a separate trip) on the five weekday mornings, 117 (about 1/3) would have an optimal 1-EXP solution involving some form of expressing. The optimal express segments are listed below by direction.

<table>
<thead>
<tr>
<th>Direction 1</th>
<th>1-3</th>
<th>1-4</th>
<th>1-5</th>
<th>1-7</th>
<th>2-4</th>
<th>6-8</th>
<th>13-18</th>
<th>15-18</th>
<th>16-18</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td># exp.trips</td>
<td>7</td>
<td>8</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>10</td>
<td>14</td>
<td>47</td>
</tr>
<tr>
<td>share of total</td>
<td>0.15</td>
<td>0.17</td>
<td>0.09</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.21</td>
<td>0.30</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Direction 2</th>
<th>27-29</th>
<th>28-30</th>
<th>28-31</th>
<th>37-39</th>
<th>37-40</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td># exp. trips</td>
<td>2</td>
<td>34</td>
<td>27</td>
<td>4</td>
<td>3</td>
<td>70</td>
</tr>
<tr>
<td>share of total</td>
<td>0.03</td>
<td>0.49</td>
<td>0.39</td>
<td>0.06</td>
<td>0.04</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.4.1 System G: Optimal Express Segments, \( m=1 \)

Table 5.4.1 shows that the optimal express segments are quite concentrated: in direction 1, 43% (20 out of 47) of the express trips start from station 1, and another 51% (24 out of 47) skip station 17. It is not coincidental that these two segments have the lowest demand (alightings + boardings) in that direction. In direction 2, where the highest demand is at the second station (28), 87% (61 out of 70) of the express segments start with station 28. The empirical implications of this are:

(a) An optimal express segment is likely to start as early as possible along the route, provided that high demand stations are not skipped. This means it is probably best to start an express segment at the first station in the direction if high demand stations are not at the beginning.
(b) If starting an express segment at the first station of the direction results in skipping a high demand station, such as in the case of Green Line B line direction 2, where the second station (i.e., 28) has the highest passenger demand (see Appendix B), it is probably better to start the express segment at that high demand station. Here we mean by "high demand station" one which has significantly higher than average demand (in terms of passenger arrivals). In the Green Line B line, stations 20, 21, 25, and 28 may be regarded as such high demand stations in the morning peak hours.

The above implications need some explanation: what about the other express trips that start from a station after 1 in direction 1 or a station other than 28 in direction 2? Let us look at direction 1 first.

![Cost Curve Family for Vehicle 11, m1](image)

**Fig. 5.1a Cost Curve Family for Vehicle 11, m1**

Fig. 5.1a shows the cost curve family for vehicle 11 in data set "m1", whose optimal express segment is 15-18. Unlike in deadheading, where there is only a single cost curve associated with each vehicle trip, in expressing there is a family of cost curves for each vehicle trip, each cost curve starts with a different express initiation station. What Fig. 5.1a tells us is that starting to express this vehicle at the optimal initiation station 15
makes little difference from starting at station 1, as long as the number of skipped stations are the same. In fact, the difference in cost reduction between expressing from 1-4 and from 15-18 for this vehicle is only 31 passenger minutes, in contrast to the net benefit of 1004 passenger minutes. As it turns out, an investigation of all the cost curve families for the express trips in direction 1 with $k_1 >1$ showed the same pattern. Among them there are 5 adjacent expressing pairs, in which the first vehicle starts expressing at station 1, and the second starts at some later station. Fig. 5.1b shows another example of a vehicle trip whose optimal express segment is 2-4.

![Cost Curve Family](image)

Fig. 5.1b Cost Curve Family for Vehicle 9, m1

The reason for the second vehicle in an adjacent expressing pair to initiate expressing at a later station is obvious: in this way the two express segments will not overlap which may contribute to additional cost reduction. For the other cases, the "flat" envelope of the minimal points associated with different initiation stations is probably due to the fact that the high demand stations are near, or at, the end of direction 1 ($k>19$), and the large weight we placed at $k_1(26)$ in the cost function. This pattern of express segments in direction 1 suggests that, when high demand stations are located late in a direction, if
there exists a beneficial express segment at all, then starting from the first station is likely to be the optimal or near optimal choice. This is simply due to the fact that a beneficial control action taken earlier in the direction benefits more stations down the line, especially when demand is high at the end.

While implication (2) is intuitive, there are a small number of exceptions, however, from this majority case. Specifically 2 express segments are initiated at station 27 (the first station in direction 2) and 7 are initiated at station 37. None of them is the second vehicle in an adjacent expressing pair. A close look at these exceptional cases reveals their unusual conditions as shown in Table 5.4.2.

<table>
<thead>
<tr>
<th>Case</th>
<th>headway ratio $h_i/h_{i+1}$</th>
<th>$h_i$ at $k_s$</th>
<th>$h_{i+1}$ at $k_{s+1}$</th>
<th>$k_s$</th>
<th>$k_{s+1}$</th>
<th>Cost Reduced (psg. min.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set</td>
<td>i</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>m2</td>
<td>15</td>
<td>27-29</td>
<td>21.07</td>
<td>21.07</td>
<td>27.21</td>
<td></td>
</tr>
<tr>
<td>w2</td>
<td>4</td>
<td>27-29</td>
<td>25.81</td>
<td>25.81</td>
<td>31.46</td>
<td>18.91</td>
</tr>
<tr>
<td>f2</td>
<td>27</td>
<td>37-40</td>
<td>1.78</td>
<td>17.30</td>
<td>19.04</td>
<td>13.54</td>
</tr>
<tr>
<td>th2</td>
<td>18</td>
<td>37-40</td>
<td>2.00</td>
<td>22.64</td>
<td>25.34</td>
<td>18.05</td>
</tr>
<tr>
<td>f2</td>
<td>17</td>
<td>37-40</td>
<td>2.74</td>
<td>6.07</td>
<td>6.36</td>
<td>4.53</td>
</tr>
<tr>
<td>m2</td>
<td>3</td>
<td>37-39</td>
<td>1.38</td>
<td>4.77</td>
<td>4.94</td>
<td>10.86</td>
</tr>
<tr>
<td>tu2</td>
<td>20</td>
<td>37-39</td>
<td>1.41</td>
<td>3.74</td>
<td>3.84</td>
<td>8.30</td>
</tr>
<tr>
<td>th2</td>
<td>29</td>
<td>37-39</td>
<td>1.48</td>
<td>4.37</td>
<td>4.51</td>
<td>8.17</td>
</tr>
<tr>
<td>m2</td>
<td>26</td>
<td>37-39</td>
<td>1.80</td>
<td>5.15</td>
<td>5.35</td>
<td>9.57</td>
</tr>
</tbody>
</table>

Table 5.4.2 Exceptional Express Segments in Dir. 2: Conditions before Expressing

In the first two cases, a vehicle is expressed from station 27, and hence skips the highest demand station 28. This happens because they have the largest headway ratios (among all 1-EXP subproblems) $h_i/h_{i+1}$ with unusually high preceding headways $>3H$ at the initiation station $k_s$, which are still increasing down the line. Furthermore, their following vehicles are all blocked before or in the express segment. The unusually large and increasing preceding headway makes the cost at the high demand station 28 smaller than the total passenger cost at later stations, and the closely following vehicle makes the additional waiting time at 28 negligible. This justifies skipping station 28.
In all the remaining cases, the vehicles start expressing at station 37 rather than 28, and skip stations 38 which has the lowest passenger demand in terms of both passenger alightings and boardings (see Appendix B). One interesting thing in common for all of them is that the express benefit is very low. If a minimal benefit is required (250 passenger minutes as suggested by Macchi, 1990), none of these vehicles would be expressed at all. Once again, their cost curves show that there is not much difference between expressing from station 28 or from station 37, which is a similar phenomenon discussed for direction 1. Namely, when a vehicle can be beneficially expressed from an early station in the direction, it may be slightly better to start expressing at a later station. This happens because, while the preceding headways of the vehicles are larger, their following headways are not small enough at earlier stations, as reflected by the relatively low headway ratio at \( k_0 \) (station 27) in Table 5.4.2. If at a later station with low demand, the following headway is also the smallest, the skipping cost can be the lowest. However, due to the loss of some benefit stations, the result is unlikely to be significantly better.

To verify the above points, we let \( k_s = 1 \) for all 1-EXP subproblems in direction 1, and \( k_s = 28 \) for all 1-EXP subproblems in direction 2, and search for the optimal length of express segments given \( k_s \). The results are shown in Table 5.4.3.

In Table 5.4.3, it is clear that the results from the two sets of \( k_s \) are extremely similar. This verifies the conclusions we drew above. On the other hand, when two adjacent vehicles both are expressed, using the set of \( k_s \) with different values has an advantage that people who are negatively affected by expressing are not always the same ones. This is an advantage of expressing over deadheading. We will discuss the adjacent expressing issue further in a later subsection.
### Table 5.4.3 System G: Expressing Results for Different Sets of $k$,

<table>
<thead>
<tr>
<th>Data Set</th>
<th>M</th>
<th>% Change in Cost</th>
<th>StdDev of hdw</th>
<th>#stations skipped</th>
<th>#vehicle trips expressed</th>
<th>% Change in Cost</th>
<th>StdDev of hdw</th>
<th>#stations skipped</th>
<th># vehicle trips deadheaded</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>36</td>
<td>-13.44</td>
<td>1.44</td>
<td>15</td>
<td>10</td>
<td>-13.11</td>
<td>1.42</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>-15.60</td>
<td>1.08</td>
<td>19</td>
<td>11</td>
<td>-15.77</td>
<td>0.97</td>
<td>15</td>
<td>9</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>-14.54</td>
<td>1.64</td>
<td>14</td>
<td>8</td>
<td>-13.81</td>
<td>1.63</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>-9.88</td>
<td>1.33</td>
<td>20</td>
<td>10</td>
<td>-10.88</td>
<td>1.27</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>-11.13</td>
<td>0.98</td>
<td>12</td>
<td>8</td>
<td>-11.95</td>
<td>0.92</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>-7.46</td>
<td>4.39</td>
<td>21</td>
<td>15</td>
<td>-7.43</td>
<td>4.43</td>
<td>19</td>
<td>13</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>-6.15</td>
<td>4.15</td>
<td>22</td>
<td>15</td>
<td>-6.13</td>
<td>4.16</td>
<td>21</td>
<td>14</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>-6.93</td>
<td>4.09</td>
<td>16</td>
<td>12</td>
<td>-6.93</td>
<td>4.09</td>
<td>16</td>
<td>12</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>-4.47</td>
<td>4.61</td>
<td>22</td>
<td>16</td>
<td>-4.51</td>
<td>4.60</td>
<td>20</td>
<td>15</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>-4.07</td>
<td>3.96</td>
<td>13</td>
<td>12</td>
<td>-4.19</td>
<td>3.94</td>
<td>16</td>
<td>11</td>
</tr>
<tr>
<td>total</td>
<td>346</td>
<td>-10.35</td>
<td>180</td>
<td>117</td>
<td>10.45</td>
<td>157</td>
<td>106</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Avg.</td>
<td></td>
<td>2.77</td>
<td></td>
<td></td>
<td>2.74</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Note:** H=5min. % Change in Cost = (Change in Cost/Cost) * 100%

### Length of Express Segment

Let us now look at the length of express segment given the optimal initiation station. Here we compute the "upper bound" on the number of skipped station in the same way as for deadheading in Chapter 4, section 4.4.6. The optimal number of skipped stations seem to fall into two different patterns:

**Pattern 1:** At the express initiation station $k_s$, if the preceding/following headway ratio of the express vehicles is smaller than the upper bound, the optimal number of skipped stations apparently closely follows the headway ratio $h_{i,k}/h_{i+1,k}$. This can be observed from both Fig. 5.2 and 5.3, and is consistent with the similar observation for the deadheading problem (see Chapter 4).

**Pattern 2:** When the preceding/following headway ratio of the express vehicles is larger than the upper bound, the optimal number of skipped stations no longer increases with the headway ratio $h_{i,k}/h_{i+1,k}$. This can be observed from Fig 5.3. We were not able to...
confirm this pattern in the deadheading cases, due to the constraint by the high demand station 28 on deadhead segment lengths in direction 2.

Fig. 5.2. Expressing Results and Conditions (dir. 1)

Fig. 5.3 Expressing Results and Conditions (dir. 2)
The computed "upper bound" is valid in all but one case. It is tight in direction 1 but not in direction 2. This is probably due to those unusually large preceding headways and minimal following headways in direction 2. This again suggests that, when the headway ratio at station $k$, is smaller than the "upper bound", it better approximates the optimal number of skipped stations. Otherwise, only a small number (e.g., 1 or 2) should be taken, because for already large headways the skipping cost is too high to skip more stations.

Although all beneficial expressing cases have a headway ratio $h_{i,k}/h_{i+1,k}>1.25$, not all vehicle trips satisfying this condition can be beneficially expressed. In fact, there are 9 vehicle trips in direction 1 whose headway ratios at station 1 range from 1.26 to 1.52 which do not have beneficial expressing costs, and 12 in direction 2 whose headway ratios at station 28 range from 1.26 to 1.90 which are not expressed. These vehicles can not be beneficially expressed for one of the following reasons:

(a) After a vehicle skips any station, its following headway tends to grow larger and larger thereby increasing the total (preceding + following) headways. This can happen when the preceding headway is relatively small (e.g., around $H$).

(b) The vehicle has a large preceding headway (e.g., around $2H$), but the following headway is not small enough. In this case the skipping cost would be too high to justify expressing. This also explains why in Fig.5.3 almost all the following headways are minimal ($h_0$).

**Cost Characteristics**

Interestingly enough, although not all expressing cost curves with a given initiation station are quasiconvex, the net beneficial portions always are. The reasons for this are similar to those discussed in Chapter 4 for deadheading cost curves and will not be repeated here. What did not arise in Chapter 4 is the characteristic of the cost curve family associated with each vehicle trip for expressing. We note that the envelope of the minimal point of each curve in such a family is not quasiconvex. However, when the
curves are in the order of the initiation station, the minimal point always occurs no earlier than that in the previous curve (see Fig. 5.1a and b). This characteristic exists also in all the cost curve families with \( m > 1 \), where each single cost curve shows very similar properties as in deadheading case, as one can see from Fig. 5.1a and b. These characteristics lead to an efficient search algorithm for RTEPG, which is presented at the end of this chapter.

**Effects of Impact Set Size**

As in the deadheading case, when the impact set size \( m \) increases, cost reduction from expressing also increases. As shown in Table 5.4.4, the weekly total cost reduction with \( m = 3 \) is 3.3% higher than with \( m = 1 \). On the other hand, the number of expressed vehicles is increased 32%, and the number of skipped stations is increased 48%. The number of adjacent expressing cases also increases from 11 to 26. In direction 1, this number increases from 5 to 13, and in direction 2 from 6 to 13. With \( m = 1 \), there are two cases where three consecutive vehicles are expressed (vehicles 15-17 in "m2", 4-6 in "w2"), while with \( m = 3 \), there are 5 such cases and one case with 4 consecutive vehicles expressed. In direction 1, all adjacent express segments when \( m = 1 \) and most when \( m = 3 \) are not overlapping. But in direction 2 most of them are overlapping with either \( m = 1 \) or \( m = 3 \). As a result, we see greater increase in cost reduction in direction 1 than in direction 2, because the non-overlapping adjacent express segments exact a lower price in direction 1 than in direction 2. The large portion of overlapping adjacent express segments in direction 2 is due to the demand pattern which gives little incentive to start expressing later. The implication from this is that, there are greater benefits for choosing \( m > 1 \) for a control direction with high demand later on the route. This is a major difference between expressing and deadheading, which we will discuss further in the next subsection.
### Impact of Threshold Benefit

In our analysis we have defined a beneficial control policy as one which results in non-zero net benefit. In practice, however, a threshold benefit may be used to decide whether a control policy should be applied. For example, for the Green Line such a threshold value may be 250 passenger-minutes (Macchi, 1990). That is, if the resulting net benefit from a control policy is lower than this value, the vehicle should not be controlled. To see the impact of such threshold, we again conducted computational tests with both impact set sizes \( m=1 \) and \( m=3 \). The results are shown in Table 5.4.5.

The results show a sharp reduction in control actions, but only a moderate decline in benefits for both \( m=1 \) and \( m=3 \). The overall cost reductions are about 5-10% lower than with 0 threshold, but both the number of skipped stations and number of express vehicles have decreased by about 40-50%. The impact of the threshold value with \( m=3 \) is slightly smaller than with \( m=1 \). The results show that an appropriate threshold can retain 90% or
more of the benefits while requiring only 40-50\% of the control actions. This is very attractive, especially for impact set size $m>1$.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>M</th>
<th>% Change in Cost</th>
<th>StDev of hdw</th>
<th>#stations skipped</th>
<th>#vehicle trips expressed</th>
<th>% Change in Cost</th>
<th>StDev of hdw</th>
<th>#stations skipped</th>
<th>#vehicle trips deadheaded</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>36</td>
<td>-11.98</td>
<td>1.50</td>
<td>13</td>
<td>7</td>
<td>-19.11</td>
<td>1.27</td>
<td>22</td>
<td>11</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>-13.80</td>
<td>1.14</td>
<td>15</td>
<td>7</td>
<td>-19.24</td>
<td>0.81</td>
<td>27</td>
<td>13</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>-13.63</td>
<td>1.68</td>
<td>11</td>
<td>5</td>
<td>-16.31</td>
<td>1.57</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>-9.37</td>
<td>1.35</td>
<td>19</td>
<td>9</td>
<td>-16.80</td>
<td>1.08</td>
<td>26</td>
<td>12</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>-7.43</td>
<td>1.10</td>
<td>8</td>
<td>4</td>
<td>-14.36</td>
<td>0.85</td>
<td>18</td>
<td>9</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>-6.64</td>
<td>4.53</td>
<td>9</td>
<td>5</td>
<td>-6.15</td>
<td>4.58</td>
<td>13</td>
<td>6</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>-5.28</td>
<td>4.32</td>
<td>8</td>
<td>4</td>
<td>-7.14</td>
<td>3.58</td>
<td>12</td>
<td>7</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>-5.98</td>
<td>4.26</td>
<td>3</td>
<td>2</td>
<td>-5.98</td>
<td>4.26</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>-3.86</td>
<td>4.65</td>
<td>6</td>
<td>3</td>
<td>-6.62</td>
<td>4.01</td>
<td>13</td>
<td>7</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>-2.91</td>
<td>4.21</td>
<td>4</td>
<td>2</td>
<td>-5.26</td>
<td>3.86</td>
<td>6</td>
<td>4</td>
</tr>
<tr>
<td>total</td>
<td>346</td>
<td>-9.03</td>
<td>2.87</td>
<td>96</td>
<td>48</td>
<td>-13.17</td>
<td>2.59</td>
<td>158</td>
<td>80</td>
</tr>
<tr>
<td>Avg.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4.5 System G: Impacts of Threshold Benefit on Expressing Policies

5.4.3 Riding Time Savings

For System $F$ we showed that including riding time in the minimization does not seem to be justified. But System $G$ may be different, especially when high demand stations are at the beginning of the control direction as in the Green Line direction 2. At such stations the vehicle load will be very high, and expressing may result in significant riding time saving. In this situation, including riding time in the objective function may be justified.

In order to test this, we add a riding time term with a weight in the objective function, and write program $\text{RTEPG}_2$ as follows:

$\text{[RTEPG}_2\text{]}$ Minimize

\[(5.4.13)\]

\[f(y) = \sum_{i \in I_m} \sum_{k \in K_r} [r_k h_{i,k}^2/2 + P_{i-1,k} h_{i,k} + \theta(L_{i,k-1}(a_{i,k} - d_{i,k-1}) + (L_{i,k} - B_{i,k})s_{i,k})] + r_k h_{i,k}^2/2]\]
Subject to (4.2.4-4.2.8)

where $\theta \leq 1$ is a weight for riding time, $B_{i,k} = \gamma_{i,k}(P_{i,k} + r_{i,k} h_{i,k})$, and $\gamma_{i,k} = \gamma_{i,k,r}$ (see equation (5.3.1)) if $k = k_j$ and $\gamma_{i,k} = 1$ otherwise. $s_{i,k}, L_{i,k}, P_{i,k}$ were defined by equations (5.4.1)-(5.4.3) respectively.

The difference between objective function (5.4.13) and (4.2.2) is the additional riding time term, which is defined as vehicle load between two consecutive stations multiplied by interstation running time plus the remaining onboard passengers (after alighting and before new boarding) multiplied by dwell time. We set $\theta = 1$ to test the extreme case. The results from both [RTEPG] and [RTEPG] with $m=3$ are shown in Table 5.4.6.

The results show that the solution structure is significantly different with and without the riding time term in the objective function. There are already quite significant reduction in riding times after expressing, even when waiting time alone is minimized. This is mainly due to the decrease in interstation stopping and more even dwell times as consequences of reduced headway variance. But the riding time saving takes a smaller portion (37%) in the total cost (waiting + riding) reduction. In the new solution after including riding time in the objective function, the riding time portion has increased to 46%. The absolute increase in riding time savings is about twice as much as the absolute decrease in waiting time savings. However, the total cost reduction is improved only slightly (about 6%) due to the larger waiting time fraction in the solution structure. This is quite consistent with System F results. If the riding time weight is set to $\theta < 1$, the increase in total cost reduction would be even smaller. While the justification for trading off waiting time savings against riding time savings is questionable, the improvement in overall cost reduction seems not significant. Note that for two data set ("m1" and "th1") the total (waiting + riding) costs reduction are not improved at all.
When vehicle capacity is likely to be binding, however, there may be a smaller loss in waiting time saving at the skipped stations. In this case the results may be different.

Next we investigate such a case.

5.4.4 Impact of Capacity Constraint

So far we have assumed that capacity constraint is not binding. While this is often true with the U.S. transit systems, in peak hours with high headway variance it may be violated. In order to test the impact of such violation, we add a capacity constraint and rewrite RTEPG as RTEPG$_{c}$:

$$[\text{RTEPG}_{c}] \text{ Minimize}$$

$$(5.4.14) \ f(y) = \sum_{i \in I_{m}} \sum_{k \in K_{c}} \left\{ r_{k}h_{i,k}^2/2 + P_{i-1,k}h_{i,k} + r_{k}h_{i,k}^2/2 \right\}$$

Subject to (4.2.4) - (4.2.8) and

$$L_{c_{i,k}} = \min\{L_{i,k}, \Lambda\} \quad \text{(capacity constraint)}$$

where

$$(5.4.16) \ B_{c_{i,k}} = \min\{ y_{i,k}(P_{i,i,k} + r_{k}h_{i,k}), \Lambda - (L_{c_{i,k}} - A_{i,k})\}y_{i,k} \quad \text{(passenger boardings)}$$

$$(5.4.17) \ s_{c_{i,k}} = (c_{0} + c_{1}B_{c_{i,k}} + c_{2}A_{i,k})y_{i,k} \quad \text{(vehicle dwell time)}$$

171
\[(5.4.18) \quad P_{i,k}^c = P_{i-1,k} + r_e h_{i,k} - B_{i,k}^c \quad \text{ (leftover passengers)}\]

where \( A \) is vehicle capacity (measured by number of passengers); and \( s_{i,k}, L_{i,k}, P_{i,k} \) and \( A_{i,k} \) were defined by equations (5.4.1)-(5.4.3) and (5.4.8) respectively. The superscript "c" in the variables denotes "(under) capacity constraint".

If riding time is included, the objective function is written as

\[(5.4.19) \quad f(y) = \sum_{i=1}^{n} \sum_{k=K_c}^{n} \left[ r_k h_{i,k}^2 / 2 + P_{i-1,k}^c h_{i,k} + L_{i,k-1}^c (a_{i,k} - d_{i,k-1}) + (L_{i,k}^c - B_{i,k}^c) s_{i,k}^c \right] + r_k h_{i,k}^2 / 2\]

To test this model, we set capacity \( A = 300 \) passengers for a Green Line train, assuming that each train has two cars, and the capacity of each car is 150 passengers. The computational results from this new model (with \( m = 3 \)) are shown in Table 5.4.7 below. For comparison purpose, we include the results with and without the riding term in the objective function.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>minimize waiting time (m=3)</th>
<th>minimize waiting+riding time (m=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wait time change</td>
<td>ride time change</td>
</tr>
<tr>
<td>m1</td>
<td>-12,056.16</td>
<td>-1,610.24</td>
</tr>
<tr>
<td>tu1</td>
<td>-12,104.98</td>
<td>-2,095.60</td>
</tr>
<tr>
<td>w1</td>
<td>-8,590.39</td>
<td>-1,390.19</td>
</tr>
<tr>
<td>th1</td>
<td>-10,744.57</td>
<td>-2,765.87</td>
</tr>
<tr>
<td>f1</td>
<td>-8,370.29</td>
<td>-1,611.77</td>
</tr>
<tr>
<td>m2</td>
<td>-1,296.58</td>
<td>-4,001.67</td>
</tr>
<tr>
<td>tu2</td>
<td>-1,340.36</td>
<td>-4,892.05</td>
</tr>
<tr>
<td>w2</td>
<td>-931.65</td>
<td>-3,188.93</td>
</tr>
<tr>
<td>th2</td>
<td>-1,451.71</td>
<td>-4,543.91</td>
</tr>
<tr>
<td>f2</td>
<td>-1,421.12</td>
<td>-3,806.11</td>
</tr>
<tr>
<td>Total</td>
<td>-58,307.81</td>
<td>-29,906.34</td>
</tr>
<tr>
<td>%</td>
<td>66.10</td>
<td>33.90</td>
</tr>
</tbody>
</table>

Table 5.4.7 System G: Expressing under Capacity Constraint (m=3)
Comparing Table 5.4.7 with Table 5.4.6, we see that for either objective function, introducing the capacity constraint lowers both the waiting and riding time savings (by 8-9%). It can also be seen that the waiting time portion of the total cost reduction (waiting vs. riding time savings) is higher. This is why minimizing both waiting and riding times resulted in a lower improvement (4.5%) in the total cost reduction. In fact, minimizing writing + riding time with contrained capacity actually resulted in an increase in waiting time for "w2". This again shows that the "composite minimization" does not seem to be more advantageous. One implication of these results is that the justification for including riding time in the objective function depends on the contribution of riding time saving reduction to the total cost reduction.

5.4.5 Comparison between Expressing and Deadheading

In this subsection we compare two aspects of expressing with deadheading:

(1) Waiting time reduction for both impact set sizes \( m=1 \) and \( m=3 \), without capacity constraint.

(2) Waiting vs. riding time savings with and without capacity constraint for \( m=3 \).

Waiting Time Savings by Impact Set Size

Table 5.4.8 lists both expressing and deadheading results for \( m=1 \). While the overall cost reductions are very similar for both of these strategies, we see that there are important differences between the two directions. Specifically, in direction 1 the two strategies perform very similarly, with little difference for the first three data sets and about 1.6% higher cost reduction in deadheading for the other two data sets. In direction 2, however, the differences are greater with expressing giving about 3% higher cost reduction in three of the five data sets. These different results are mainly due to the demand pattern effect that does not exist in System F.
<table>
<thead>
<tr>
<th>data set</th>
<th>M</th>
<th>% change in waiting cost</th>
<th>StDev of hdw</th>
<th>#stations skipped</th>
<th>#vehicles expressed</th>
<th>% change in waiting cost</th>
<th>StDev of hdw</th>
<th>#stations skipped</th>
<th># vehicles deadheaded</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>36</td>
<td>-13.44</td>
<td>1.44</td>
<td>15</td>
<td>10</td>
<td>-14.08</td>
<td>1.38</td>
<td>12</td>
<td>9</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>-15.60</td>
<td>1.08</td>
<td>19</td>
<td>11</td>
<td>-14.92</td>
<td>0.99</td>
<td>13</td>
<td>8</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>-14.54</td>
<td>1.64</td>
<td>14</td>
<td>8</td>
<td>-14.20</td>
<td>1.60</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>-9.88</td>
<td>1.33</td>
<td>20</td>
<td>10</td>
<td>-11.36</td>
<td>1.22</td>
<td>16</td>
<td>9</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>-11.13</td>
<td>0.98</td>
<td>12</td>
<td>8</td>
<td>-12.74</td>
<td>0.87</td>
<td>11</td>
<td>7</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>-7.46</td>
<td>4.39</td>
<td>21</td>
<td>15</td>
<td>-4.83</td>
<td>5.36</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>-6.15</td>
<td>4.15</td>
<td>22</td>
<td>15</td>
<td>-6.30</td>
<td>4.84</td>
<td>13</td>
<td>13</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>-6.93</td>
<td>4.09</td>
<td>16</td>
<td>12</td>
<td>-5.19</td>
<td>5.07</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>-4.47</td>
<td>4.61</td>
<td>22</td>
<td>16</td>
<td>-5.84</td>
<td>4.94</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>-4.07</td>
<td>3.96</td>
<td>19</td>
<td>12</td>
<td>-0.88</td>
<td>4.71</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>total.</td>
<td>346</td>
<td>-10.35</td>
<td>2.77</td>
<td>180</td>
<td>117</td>
<td>-10.25</td>
<td>3.10</td>
<td>114</td>
<td>91</td>
</tr>
</tbody>
</table>

Avg. 2.77 3.10

Note: H=5min. % Change in Cost = Change in Cost/Cost*100%

Table 5.4.8 System G: Comparison between Expressing and Deadheading (m=1)

Expressing may result in lower passenger waiting times than deadheading when employed to remedy the same no-control situation. This is because expressing is more flexible than deadheading by virtue of the ability to choose initiation stations. However, when the optimal express initiation station is $k_0$, expressing can be less advantageous than deadheading for two reasons. First because the vehicle stops at $k_0$ and hence is not subject to control as early as in deadheading, the headway ratio is larger at the departure from $k_0$, which implies potentially higher passenger waiting times at the next station. Second, when the number of skipped stations is the same, expressing results in one less benefit station than deadheading which may cause the total cost to be higher. This explains why the resulting passenger waiting time is higher in expressing in some of the direction 1 data sets, but such cost difference between expressing and deadheading tend to be quite small.

Direction 2 data sets are quite different. Because the highest demand station is the second station $k_0+1 (=28)$ in this direction, and the first station a deadhead vehicle skips must be $k_0$, it turns out most optimal deadhead segments in the direction skips only $k_0$, because the...
cost would be too high to skip $k_0+1$. This limits the time saving and hence effectiveness of deadheading in this situation. On the other hand, due to the flexibility of choosing an express initiation station, 97% of the 1-EXP subproblems with beneficial cost have an optimal initiation station other than $k_0$ for reasons discussed in previous subsections. Because passenger arrivals at stations after $k_0+1$ are much lower, more stations may be skipped with low additional cost when needed. In fact, about 42% of the expressed vehicles skip two stations (see Table 5.4.1). This makes expressing a more effective strategy than deadheading in direction 2. The reason why the number of skipped stations is not more than two in direction 2 is due to both the dispatching headway conditions as we discussed earlier, and the demands, hence control benefits, are low at the remaining stations.

Table 5.4.9 compares the two strategies with impact set size $m=3$. Interestingly, with the change in rolling size the direction in which deadheading and expressing are preferred are reversed. In direction 1, this time expressing has slightly better performance than deadheading, while in direction 2 it seems less effective than deadheading. The reasons for such a performance switch are due to multiple factors, including interactions between demand pattern effects, rolling size effects, dispatching headway patterns and control strategy characteristics. As we discussed before, the increase in rolling size produces more adjacent vehicle deadheading or expressing cases. In direction 1, because the high demand stations are at the end, for a vehicle to start skipping stations somewhat later makes a smaller difference than having overlapped skipped segments between adjacent vehicle trips. However, for deadheading there are no other options for the starting station, while for expressing the choice is flexible. As a result, with increased adjacent vehicle control, there are fewer overlapping skipped segments in expressing. This offsets the performance between the two strategies and again supports the point we made earlier that under conditions such as direction 1’s, $m=3$ is a better choice than $m=1$ for expressing.
<table>
<thead>
<tr>
<th>dataset</th>
<th>M</th>
<th>% change in waiting cost</th>
<th>StDev of hdw</th>
<th>#stations skipped</th>
<th>#vehicle expressed</th>
<th>% change in waiting cost</th>
<th>StDev of hdw</th>
<th>#stations skipped</th>
<th># vehicle deadheaded</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>36</td>
<td>-19.52</td>
<td>1.17</td>
<td>32</td>
<td>17</td>
<td>-18.82</td>
<td>1.16</td>
<td>29</td>
<td>15</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>-19.31</td>
<td>0.80</td>
<td>30</td>
<td>16</td>
<td>-18.86</td>
<td>0.78</td>
<td>28</td>
<td>14</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>-16.42</td>
<td>1.56</td>
<td>21</td>
<td>12</td>
<td>-16.35</td>
<td>1.52</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>-16.26</td>
<td>1.03</td>
<td>29</td>
<td>16</td>
<td>-14.02</td>
<td>1.06</td>
<td>32</td>
<td>16</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>-15.97</td>
<td>0.70</td>
<td>22</td>
<td>13</td>
<td>-15.09</td>
<td>0.79</td>
<td>18</td>
<td>10</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>-7.58</td>
<td>4.09</td>
<td>32</td>
<td>19</td>
<td>-8.84</td>
<td>4.32</td>
<td>15</td>
<td>12</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>-8.16</td>
<td>2.69</td>
<td>31</td>
<td>17</td>
<td>-8.84</td>
<td>3.23</td>
<td>17</td>
<td>12</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>-5.35</td>
<td>4.54</td>
<td>18</td>
<td>12</td>
<td>-6.21</td>
<td>4.75</td>
<td>10</td>
<td>9</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>-7.49</td>
<td>3.69</td>
<td>31</td>
<td>19</td>
<td>-10.97</td>
<td>3.62</td>
<td>16</td>
<td>13</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>-7.50</td>
<td>3.02</td>
<td>21</td>
<td>13</td>
<td>-7.31</td>
<td>3.56</td>
<td>10</td>
<td>7</td>
</tr>
<tr>
<td>total</td>
<td>346</td>
<td>-13.69</td>
<td>267</td>
<td>154</td>
<td>134</td>
<td>-13.70</td>
<td>193</td>
<td>118</td>
<td></td>
</tr>
<tr>
<td>Avg.</td>
<td></td>
<td>2.33</td>
<td></td>
<td></td>
<td></td>
<td>2.48</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 5.4.9 System G: Comparison between Expressing and Deadheading (m=3)

In direction 2, the demands at later stations are low. This gives little incentive for a vehicle to start skipping stations later. Hence most of the consecutively expressed or deadheaded vehicles have overlapped skipped segments at early stations. The difference is that deadheaded vehicles started at station 27 while the expressed ones started at station 28. In addition, because the dispatching headways are usually very large for a controlled vehicle, the expressed vehicles do not benefit from choosing longer skipped segments. In this case deadheading seems slightly more effective in terms of waiting time reduction, with fewer skipped stations. This is probably a result of earlier the deadheading control provided.

**Waiting vs. Riding Time Savings**

Expressing has another potential advantage over deadheading, which is that it partially utilizes capacity at the express initiation station, and thereby reduces the riding times of passengers on the express vehicle. To investigate this issue we lift the capacity constraint
as in the RTEPG \(_3\) model while minimizing just waiting times. The results for \(m=3\) are shown in Table 5.4.10.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Expressing ((m=3))</th>
<th>Deadheading ((m=3))</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wait time change</td>
<td>ride time change</td>
</tr>
<tr>
<td>m1</td>
<td>-12,190.63</td>
<td>-1,675.91</td>
</tr>
<tr>
<td>tu1</td>
<td>-12,104.98</td>
<td>-2,095.60</td>
</tr>
<tr>
<td>w1</td>
<td>-11,123.79</td>
<td>-2,105.84</td>
</tr>
<tr>
<td>th1</td>
<td>-10,744.57</td>
<td>-2,765.87</td>
</tr>
<tr>
<td>f1</td>
<td>-8,370.29</td>
<td>-1,611.77</td>
</tr>
<tr>
<td>m2</td>
<td>-2,941.31</td>
<td>-6,978.75</td>
</tr>
<tr>
<td>tu2</td>
<td>-3,085.04</td>
<td>-7,299.27</td>
</tr>
<tr>
<td>w2</td>
<td>-2,040.53</td>
<td>-4,393.69</td>
</tr>
<tr>
<td>th2</td>
<td>-2,895.42</td>
<td>-5,427.09</td>
</tr>
<tr>
<td>f2</td>
<td>-2,244.84</td>
<td>-4,872.23</td>
</tr>
<tr>
<td>Total</td>
<td>-67,741.40</td>
<td>-39,226.02</td>
</tr>
<tr>
<td>%</td>
<td>63.33</td>
<td>36.67</td>
</tr>
</tbody>
</table>

Table 5.4.10 System G: Waiting Time vs. Riding Time Savings

We first note that both expressing and deadheading have resulted in quite significant riding time reductions in addition to waiting time reductions. Their waiting time savings are extremely similar as we already discussed. However, the aggregate riding time saving is almost doubled in expressing case. The riding time reductions from expressing are particular high in direction 2, where the express initiation station is often the highest demand station and more passengers were riding through the express segment. This shows that expressing is more advantageous than deadheading when the optimal express initiation station is a high demand station.

The results of expressing and deadheading under capacity constraints are also consistent with the above observations, as shown in Table 5.4.11, although the waiting time portion in the total benefits are higher than without capacity constraint.
### Table 5.4.11 System G: Waiting Time vs. Riding Time Savings under Capacity Constraint

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Expressing (m=3)</th>
<th>Deadheading (m=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>wait time change</td>
<td>ride time change</td>
</tr>
<tr>
<td>m1</td>
<td>-12,056.16</td>
<td>-1,610.24</td>
</tr>
<tr>
<td>tu1</td>
<td>-12,104.98</td>
<td>-2,095.60</td>
</tr>
<tr>
<td>w1</td>
<td>-8,590.39</td>
<td>-1,390.19</td>
</tr>
<tr>
<td>th1</td>
<td>-10,744.57</td>
<td>-2,765.87</td>
</tr>
<tr>
<td>fl</td>
<td>-8,370.29</td>
<td>-1,611.77</td>
</tr>
<tr>
<td>m2</td>
<td>-1,296.58</td>
<td>-4,001.67</td>
</tr>
<tr>
<td>tu2</td>
<td>-1,340.36</td>
<td>-4,892.05</td>
</tr>
<tr>
<td>w2</td>
<td>-931.65</td>
<td>-3,188.93</td>
</tr>
<tr>
<td>f2</td>
<td>-1,421.12</td>
<td>-3,806.11</td>
</tr>
<tr>
<td>Total</td>
<td>-58,307.81</td>
<td>-29,906.34</td>
</tr>
</tbody>
</table>

**Comparison Summary**

To summarize the above comparisons, expressing and deadheading have different advantages depending on headway and demand condition. In terms of waiting time reduction alone, when controlled vehicles have large preceding headways, deadheading is often more effective. While both strategies also result in significant riding time reductions, expressing is more effective in this aspect. Especially when the optimal express initiation station is a high demand station, expressing can result in much larger riding time savings. Another important advantage of expressing is its non-overlapping skipped segments. However, in terms of modeling and computation efforts, expressing is a lot more costly than deadheading.

#### 5.4.6 Computer Algorithm and Empirical Guidelines for Expressing

Based on the above analysis, the following heuristic is proposed for model RTEPG:

**Algorithm RTEPG**

For each vehicle $i$, start with $k_x = k_0$, and $k_x = k_0 + 2$. 

178
While \( k_s \leq k_{i-2} \) Do:

**Step 1.** Compute the no-control cost of \( i \) and \( i+1 \);

**Step 2.** Express \( i \) from \( k_s \) to \( k_e \), and compute the trajectories of \( i \) and \( i+1 \) using equations (5.4.1)-(5.4.3) and equation (5.3.1).

**Step 3.** If the safe headway constraint is violated at any station from \( k_s \) to \( k_e \), go to step 5. Otherwise, compute the expressing costs of vehicle \( i \) and \( i+1 \), denoted \( f_{i,k_e}^{(n)} \), where \( n \) is the iteration number.

**Step 4.** If \( f_{i,k_e}^{(n)} < f_{i,k_{e-1}}^{(n)} \) and \( k_e < k_i \), set \( k_e = k_e + 1 \) and go to Step 1. Otherwise go to step 5.

**Step 5.** Set \( k_e = k_e + 1 \), \( k_e = \arg \{ \min_{k_e} f_{i,k_e}^{(n)} \} \), and \( n = n+1 \), go to next iteration of the While loop.

End

This algorithm will be used later in Chapter 7 for the combined strategy.

**Empirical Guidelines for Expressing**

The above analysis clearly shows that the choice of optimal express segment depends on the interaction between the vehicle headway conditions and the demand pattern in the control direction. Such interaction is quite complicated and can be best dealt with by computer. However, in decentralized control systems where a computer is not available to every inspector, the following guidelines may help under the condition the inspector is familiar with the demand pattern and is provided information on the following headway of each vehicle. The guidelines are based on the analysis for the 1-EXP subproblems (or \( m=1 \)).

1. **Choice of express initiation station:** if the demand peak is in the middle or later in the control direction, choose the first station as the express initiation station; If the demand peak is within the first couple of stations, choose the peak demand station as the express initiation station.

2. **Decision on whether a vehicle should be expressed:** if the vehicle's headway ratio at the given initiation station is smaller than \( 1+h_0 \), do not express. If the vehicle's headway
ratio is larger than $1+h_0$ but the preceding headway is around $H$, do not express. If the vehicle's preceding headway is around or larger than $2H$ but the next vehicle does not follow immediately, do not express.

3. **Choice of the number of skipped stations:** if the vehicle's headway ratio at the given initiation station is larger than $1+h_0$ but smaller than the upper bound (i.e., $(\text{preceding headway-following headways + time saving per station})/(2*\text{time saving per station})$), round the ratio down to estimate the number of skipped stations. Otherwise, only skip one or two stations.
CHAPTER 6

THE HOLDING PROBLEM

6.1 Introduction

This Chapter investigates the holding problem. When a vehicle is ready to depart from a station after its normal loading and unloading, it may be held for an amount of time in order to even out the headways between its preceding and following vehicles. The holding problem is to decide at a given time at a control station, which vehicle is to be held and for how long, such that the total passenger cost is minimized. In contrast to deadheading and expressing, holding is usually applied where a vehicle's preceding headway is smaller than its following headway. In other words, while deadheading and expressing reduces a vehicle's preceding headway, holding increases it. Another important property of the holding strategy is that it does not result in any station being skipped, and therefore is likely to be less frustrating to passengers. Probably because holding is a much less costly strategy, and is easy to implement, it has attracted most of the attention in previous research (see Chapter 1 for a literature review). As we have pointed out, however, due to modeling difficulties raised by complex probability functions, all the researchers have assumed the optimal holding policy is in the form of "threshold" values, which may not be true when real-time information is used instead. Another problem is that, no one has explicitly incorporated the constraints of terminal
departure times and layover time. These constraints are very important if a holding model is to be used in practice. Furthermore, the effectiveness of holding has not been evaluated systematically over a period of time.

There may be a concern with the waiting time objective function in the holding problem. Since holding means delaying some vehicle departure times, it may increase passenger riding time in order to decrease passenger waiting time. Therefore, one may suggest an alternative objective function, such as minimizing both passenger waiting time and riding time. Here we do not regard such a substitution as being either necessary or preferred. First, there is no such a need in System F as the best holding station is $k_0$ and headways do not change along the route. The delay of passengers at $k_0$ is already included in the waiting time cost. Second, in System G, while holding delay of vehicle $i$ at the control station $k$ may cause a longer travel time beyond $k$, such en-route delay is counted toward reduction of passenger waiting time. An objective function containing a riding time term would result in tradeoff between waiting and riding time reduction, in the favor of the latter. As we discussed in both Chapters 1 and 5, such a tradeoff may not be justified due to the fact passengers are more sensitive to waiting time cost. Further, a vehicle that will be held is usually one with a shorter preceding headway, which means its trip time is likely to be shorter than normally scheduled when not held. In this case if its normal trip time is restored by holding, it should not be regarded as a delay at all. In addition, holding delay at a single station is likely to be very small, as we will show later in this Chapter; hence its impact on passengers is also very small compared to a much larger waiting time saved along the entire route. Finally, holding delay of one vehicle can reduce trip times for more vehicles. In this case, a small delay of that vehicle can be justified from the system point of view. Such effects of holding will be investigated in this chapter.

In this chapter we model the holding problem in both System F and System G, and systematically evaluate effectiveness of the holding strategy in both systems.
6.2 Optimal Holding in System $F$

6.2.1 Model Formulation of HPF

Due to the invariant headway along the route in System $F$, a vehicle’s headway at all stations depends solely on the departure headway at $k_0$, consequently for each vehicle we only need to consider holding at $k_0$, and it makes no difference in the optimal solution how many stations are considered. We omit the subscript $k$ and define the decision variable as $h_j$ for vehicle $j$. The formulation for the HP in System $F$, called HPF, is given as follows.

[HPF] Minimize

\[(6.2.1) \, f(h) = \sum_{j \in I_m} h_j^2 \]

Subject to:

\[(6.2.2) \, \sum_{j' \in j} h_{j'} \geq d_{j,k_0}^0 - d_{i-1,k_0}^0, \, \forall j \in I_m \]

\[(6.2.3) \, \sum_{j \in I_m} h_i = d_{i+m,k_0}^0 - d_{i-1,k_0}^0 \]

\[(6.2.4) \, h_j \geq c_0 + h_0, \, \forall j \in I_m \]

\[(6.2.5) \, \sum_{j' \in j} h_{j'} \leq \max(t_{j,c} - T_c, d_{j,k_0}^0) - d_{i-1,k_0}^0, \, \forall j \in I_n \]

where $h_j = d_{j,k_0} - d_{j-1,k_0}, c_0 > 0, h_0 > 0$; $i$ is the first vehicle in $I_m$ which is to be controlled using the solution to [HPF], $T_c$ is the free trip time from $k_0$ to $k_{c+1}$ plus minimal layover time.

Constraint (6.2.2) says that the departure time of a vehicle with holding is no earlier than without holding. Constraint (6.2.3) says that the sum of all vehicle headways in the $I_m$ set is fixed, because the trajectories of the two boundary vehicles, $i-1$ and $i+m$, are fixed.

Constraint (6.2.4) is the safe headway constraint. Constraint (6.2.5) is the scheduled next trip dispatching time constraint (referred to as "terminal schedule constraint" later).

Incorporating the schedule constraint eliminates any further delay in scheduled dispatching time for the next trip. In other words, if a vehicle is already late it will not be held. This constraint may be of practical importance to most transit agencies in the U.S., because vehicle operators do not like to delay their relief time. When the scheduled last
trip in an operator's piece of work is late, the operator may skip the last trip, resulting in a large gap between two vehicle trips. Hence, holding is less attractive than station skipping strategies in that the latter do not delay the controlled vehicle. On the other hand, depending on how tight the schedule is, constraint (6.2.5) can greatly decrease the effectiveness of holding, because once a vehicle is late, vehicles behind it are also more likely to be running late.

When vehicle \( i \) is already late, i.e., when its no-control departure time at \( k_0 \) plus free trip time will result in a departure time at \( k_c+1 \) later than \( t_{i,c}, d_{i,k_0}^0 + T_c > t_{i,c} \), HPF is trivial due to constraints (6.2.2) and (6.2.5). In this case the only solution for \( i \) is \( h_i = h_i^0 \), regardless of the headways of following vehicles. Hence we next consider the case \( d_{i,k_0}^0 + T_c < t_{i,c} \) only, where for vehicle \( i \) constraint (6.2.5) reduces to

\[(6.2.5') h_i \leq t_{i,c} - T_c - d_{i-1,k_0}^0 \]

Let \( I_m = I_m - \{i+m\} \). By substituting (6.2.3) into the objective function (6.2.1), [HPF] becomes

\[(6.2.6) \quad f(h) = \sum_{j \in I_m'} h_j^2 + (l - \sum_{j \in I_m'} h_j)^2 \]

Subject to (6.2.2), (6.2.4) and (6.2.5) \( \forall j \in I_m' \), where \( l = d_{i+m,k_0}^0 - d_{i-1,k_0}^0 \).

This objective function is convex as we show below. For each \( h_i, i \in I_m' \), the partial derivative of (6.2.6) is

\[(6.2.7) \quad \frac{\partial f(h)}{\partial h_j} = 2h_j - 2(l - \sum_{j' \in I_m'} h_{j'}) \]

and the second order partial derivative with respect to \( h_i \) is

\[(6.2.8) \quad \frac{\partial^2 f(h)}{\partial h_j^2} = 2(1 + 1) = 4, \quad \frac{\partial^2 f(h)}{\partial h_j \partial h_{j'}} = \frac{\partial^2 f(h)}{\partial h_{j'} \partial h_j} = 2, j \neq j' \]

The Hessian of \( f(h) \) with elements given by (6.2.8) is an \( m-1 \) by \( m-1 \) square matrix given by

184
\[(6.2.9) \quad \nabla^2 f(h) = 2 \begin{pmatrix} 2 & 1 & \ldots & 1 \\ 1 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \ldots & 1 & 2 \end{pmatrix} \]

Let \( x = (x_1, x_2, \ldots, x_m)' \), then

\[(6.2.10) \quad x'[\nabla^2 f(h)]x = 2 \sum_{i=1}^{m-1} (2x_i^2 + x_i \sum_{j \neq i} x_j) = 2 \sum_{i \neq j} (x_i + x_j)^2 \geq 0. \]

This clearly shows that [HPF] is a convex quadratic program, because the Hessian of \( f(h) \) is positive semidefinite and the constraint set is polyhedral. Thus there exists a global optimal solution, and a local optimal solution is also a global optimum.

### 6.2.2 Solution Algorithm for HPF

Without constraints (6.2.2), (6.2.4) and (6.2.5), the solution to [HPF] is (trivially) equal headways between vehicles in the \( I_m \) set, and such a solution is the lower bound of the optimal solution to [HPF]. But with all constraints, [HPF] is no longer so trivial. While it can be solved using any standard convex quadratic program solver, here we present a simple and efficient algorithm without using a nonlinear program solver. This algorithm is based on two ideas: first, if a vehicle \( j \) is already late for its next departure, it is not held beyond its current no-control departure time. In this case \( j \) serves as a boundary vehicle of an independent subset, since control policies on vehicles in this subset will have no impact on vehicle trips after \( j \). Subsequently we need only consider this subset instead of \( I_m \), decreasing the problem size. The second idea utilizes the fact that the most even headway (under all constraints), results in the least passenger waiting time which is lower bounded by equal headways in the set. Hence the natural idea is, while maintaining feasibility of the solution, we are to even out the headways as much as we can.

For notation simplicity, in the following algorithm we sometimes omit the subscript \( k_0 \) on departure times at \( l' \),
**Algorithm HPF**

Data structure: (1) an integer vector \( \{p^n\} = (p^0, p^1, ..., p^n) \). \( p^n \) represents a vehicle whose departure time is binding on either constraint (6.2.2) or (6.2.5). (2) two logical vectors \( \{flag_1(p^n)\} \) and \( \{flag_2(p^n)\} \), each with the same dimension as \( \{p^n\} \). \( flag_1(p^n) = 1 \) if (6.2.2) is binding at \( p^n \), and 0 otherwise. \( flag_2(p^n) = 1 \) if (6.2.5) is binding at \( p^n \), and 0 otherwise.

Start with \( i = i \) the first vehicle in the \( I_m \) set.

**While** \((i' < i+m)\) Do\(^7\):

Case 1: If \( d_{i,k_0}^0 + T_c \geq t_{i,c} \) (i.e., (6.2.5) is binding), \( d_i = d_i^0 \) and no control action taken.

Set \( i' = i+1 \), and go to next iteration of the "While" loop.

Case 2:

Iteration 0. Identify a new boundary: Start from \( p^0 = i' - 1 \). Set \( flag_1(p^0) = 1, flag_2(p^0) = 1 \).

Set \( last = i'+m \). If any vehicle \( i < j < last \) and \( d_{j,k_0}^0 + T_c \geq t_{j,c} \), set \( last = j \). Set \( p^i = last, h^0 = \infty, n = 1, d_{last} = d_{last}^0 \).

Iteration \( n \). If \( p^{n-1} < last \), Do:

**Step 1:** Compute the equal headway \( h^* = \frac{d_{p^n} - d_{p^{n-1}}}{(p^n - p^{n-1})} \) between \( p^{n-1} \) and \( p^n \).

**Step 2:** For each vehicle \( j \): \( p^{n-1} < j \leq p^n \),

Compute \( d_j = d_{j-1} + h^* \);

Check violation of constraints (6.2.2) or (6.2.5);

Set \( flag_1(p^n) \) and \( flag_2(p^n) \) to 1 or 0 correspondingly.

Stop at the first \( j : d_j \leq d_0^i \) or \( d_j \geq \max(t_i - T_c, d_0^i) \); If \( j < p^n \), go to Step 3, otherwise go to Step 4.

**Step 3:** Set \( p^n = j \). If \( d_j > \max(t_i - T_c, d_0^i) \), set \( d_j = \max(t_i - T_c, d_0^i) \); else if \( d_j < d_0^j \), set \( d_j = d_0^j \). If \( p^n - p^{n-1} = 1 \), go to Step 4, otherwise go to next iteration without change \( n \).

**Step 4:** If \( h^* \geq h^{n-1} \) and \( flag_2(p^{n-1}) = 1 \), or \( h^* \leq h^{n-1} \) and \( flag_1(p^{n-1}) = 1 \), go to step 5.

Otherwise, if \( h^* > h^{n-1} \) and \( flag_1(p^{n-1}) = 0 \) or if \( h^* > h^{n-1} \) and \( flag_2(p^{n-1}) = 0 \),

set \( p^{n-1} = p^n, n = n - 1 \), go to next iteration.

**Step 5:** Set \( h^* = h^* \), \( n = n + 1 \), \( p^n = last \) and go to next iteration.

**End** Iteration \( n \)

---

\(^7\) This outer "while" loop is not needed if the algorithm is used to obtain control policy for the first vehicle only. It is included here in case control policies for all vehicles in the set are wanted.
For each $n$ and $j$ such that $i^* < j \leq \text{last}$, $i^* < p^n \leq p^a$, set $h_j = h^n$, if $p^{n-1} < j \leq p^n$.

**End Case 2**

Set $i^* = \text{last} + 1$, $\text{last} = i + m$. Go to next iteration of the "While" loop.

**End "While" loop**

The solution from this algorithm is $h_j$ for each vehicle $j$. Given $h_j$ and $d_{i-1,k_0}$, all $d_{j,k_0}$ for each $j$ in $I_m$ are easily obtained.

Before giving the complete proof of optimality, we first explain the algorithm logic and resulting solution properties.

**Property 1.** For any vehicle $j$, if one or both of the constraints (6.2.2) and (6.2.5) are binding, we refer to such a vehicle as a "turning point" vehicle. These are exactly the $p^n$'s in the algorithm. While headways at any two "turning points" may be different, the headway at any non-"turning point" $j$ is equal to the headway at the next "turning point" vehicle. That is, $h_j = h_{p^n}$ for all $j$: $p^{n-1} < j \leq p^n$. The basic logic here is: if neither constraint is binding, the optimal headways are even headways.

**Property 2.** If, at a turning point $j=p^n<\text{last}$, constraint (6.2.2) is binding but not (6.2.5), then we have $h_j \geq h_{j+1}$ in the solution. This property follows from Step 3 and Step 4. In general, if constraint (6.2.5) is not binding on any vehicle, the entire solution vector will be characterized by $h_j \geq h_{j+1}$ for each $j$. The logic behind this is that, because (6.2.5) gives the upper bound for holding, if the headway $h_j$ is less than $h_{j+1}$ and the upper bound is not binding at $j$, one could increase $h_j$ to get more even with $h_{j+1}$, while leaving other things unchanged.

**Property 3.** If constraint (6.2.5) is binding at a turning point $j=p^n<\text{last}$ while constraint (6.2.2) is not, then we have $h_j \leq h_{j+1}$ in the solution. This is also guaranteed by Step 3 and Step 4. The logic behind this is that, because (6.2.2) gives the lower bound for holding, if the headway $h_j$ is larger than $h_{j+1}$ and the lower bound is not binding, one could decrease $h_j$ to make it more even with $h_{j+1}$, while leaving other things unchanged.

We now prove that the above properties represent an optimal solution.
Proposition 6.1

Algorithm HPF solves [HPF] to optimality.

Proof: We first show that the solution from Algorithm HPF is feasible for Case 2 (Case 1 is trivial). First, for each \( n \), because \( h^* \leq h^* > \min(h_i^0) \geq c_0 + h_0 \) in Step 1, constraint (6.2.4) will never be violated. Furthermore, for each \( j < \text{last} \), constraints (6.2.2) and (6.2.5) are always satisfied by Steps 2 and 3. It is also obvious that by the construction of the algorithm we always have \( d_{i+m,k} = d_{i+m,k}^0 \), hence \( h_{i+m} = l - \sum_{j \in I_{m'}} h_j \) and constraint (6.2.3) is always satisfied. The algorithm is obviously finite due to the finite number of iterations given to each loop.

Having proved that [HPF] is a convex program, to prove that the solution \( h = \{ h_j : j \in I_m \} \) from Algorithm HPF is optimal, we only need to find a set of \( u = \{ u_{1j}, u_{2j}, u_{3j} : j \in I_m \} \), where \( u_{1j}, u_{2j}, u_{3j} \) are the Lagrange multipliers associated with constraints (6.2.2), (6.2.4) and (6.2.5) respectively, such that \( h \) and \( u \) satisfies the KKT optimality conditions below:

\[
\frac{\partial f(h)}{\partial h_j} - \sum_{j' \in I_{m'}} u_{1j'} \frac{\partial g_{1j'}(h)}{\partial h_j} - \sum_{j' \in I_{m'}} u_{2j'} \frac{\partial g_{2j'}(h)}{\partial h_j} - \sum_{j' \in I_{m'}} u_{3j'} \frac{\partial g_{3j'}(h)}{\partial h_j} = 0, \quad j \in I_m, \tag{6.2.11a}
\]

\[
\begin{align*}
(6.2.11b) \quad & u_{1j}[b_{1j} - g_{1j}(h)] = 0, \quad u_{2j}[b_{2j} - g_{2j}(h)] = 0, \quad u_{3j}[b_{3j} - g_{3j}(h)] = 0, \quad j \in I_m, \\
(6.2.11c) \quad & u_{1j} \geq 0, \quad u_{2j} \geq 0, \quad u_{3j} \geq 0, \quad j \in I_m,
\end{align*}
\]

where

\[
\begin{align*}
g_{1j} &= \sum_{j' \neq i} h_{j'}, b_{1j} = d_{ik_0}^0 - d_{i,k_i}^0, \\
g_{2j} &= h_{j}, \quad b_{2j} = c_0 + h_0, \quad \text{and} \\
g_{3j} &= -g_{1j}, \quad b_{3j} = \max(t_{jk,i} - T_{k_0,k_i+1}, d_{j,k_0}^0) - d_{i-1,k_0}^0.
\end{align*}
\]

Now, by (6.2.7) we have \( \frac{\partial f(h)}{\partial h_j} = 2(h_j - h_{i+m}) \). Also we have

\[
\begin{align*}
\frac{\partial g_{1j'}(h)}{\partial h_j} &= 1, j' \geq j, \quad \frac{\partial g_{1j'}(h)}{\partial h_j} = 0, j' < j, \\
\frac{\partial g_{3j'}(h)}{\partial h_j} &= -1, j' \geq j, \quad \frac{\partial g_{3j'}(h)}{\partial h_j} = 0, j' < j, \quad \text{and} \\
\frac{\partial g_{2j'}(h)}{\partial h_j} &= 0, \quad j, j' \in I_m. \quad \text{Thus (6.2.11a) yields}
\end{align*}
\]
(6.2.12a) \[ 2(h_j - h_{i+m}) - \sum_{j' \neq j}^{i+m-1} (u_{1j'} - u_{3j'}) = 0, \quad j \in I_m. \]

There may be many feasible solutions to (6.2.12a). Observe that each vehicle \( j \) is either a turning point, at which constraints (6.2.2) and/or (6.2.5) are binding; or \( j \) is a point between two turning points whose headway is equal to the next turning point by Property 1 above. We next construct a feasible solution to (6.2.12a) based on this.

(a) If \( j \) is a turning point \( p^a \) and both constraints (6.2.2) and (6.2.5) are binding, \( u_{1j} = 2h_j \) and \( u_{3j} = 2h_{j+1} \).

(b) If \( j \) is a turning point \( p^a \) but only constraint (6.2.2) is binding, \( u_{1j} = 2(h_j - h_{j+1}) \), and \( u_{3j} = 0 \).

(c) If \( j \) is a turning point \( p^a \) but only constraint (6.2.5) is binding, \( u_{3j} = -2(h_j - h_{j+1}) \), and \( u_{1j} = 0 \).

(d) Otherwise if \( j \) is not a turning point, \( u_{1j} = u_{3j} = 0 \).

With this construction, terms in \( \sum_{j' \neq j}^{i+m-1} (u_{1j'} - u_{3j'}) \) all cancel each other except \( 2h_j \) and \(-2h_{i+m} \), and hence \( u_{1j} \) and \( u_{3j} \) solve equation (6.2.12a). This is better illustrated in a tabular form as follows (where 'x' indicates "binding", and 'o' indicates "not binding").

<table>
<thead>
<tr>
<th>( j )</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>( d^p_j )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>x</td>
<td>o</td>
<td>o</td>
<td>x</td>
<td>o</td>
<td>o</td>
<td>x</td>
</tr>
<tr>
<td>( \text{max}(t_j \cdot T, d^p_j) )</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>x</td>
<td>o</td>
<td>o</td>
<td>o</td>
<td>x</td>
<td>x</td>
<td>o</td>
</tr>
<tr>
<td>( h_j )</td>
<td>( h_2 )</td>
<td>( h_3 )</td>
<td>( h_4 )</td>
<td>( h_5 )</td>
<td>( h_6 )</td>
<td>( h_7 )</td>
<td>( h_8 )</td>
<td>( h_9 )</td>
<td>( h_{10} )</td>
<td>( h_{11} )</td>
</tr>
<tr>
<td>( u_{1j} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2( h_5 )</td>
<td>0</td>
<td>2(h_7-h_8)</td>
<td>0</td>
<td>2( h_9 )</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>( u_{3j} )</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2( h_6 )</td>
<td>0</td>
<td>0</td>
<td>2(h_9-h_{10})</td>
<td>2( h_{10} )</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

In this case \( i=2 \) and \( m=9 \). The turning points here are 5,7,8,9 with the last one \( p=9 \). It is easy to show that for any \( j=2,3,\ldots,10 \), \[ \sum_{j' \neq j}^{i+m-1} (u_{1j'} - u_{3j'}) = \sum_{j' \neq j}^{10} (u_{1j'} - u_{3j'}) = 2(h_j - h_{i+m}) = 2(h_j - h_{11}). \] Take the example \( j=5 \). We have \( 2 \sum_{j'=5}^{10} (u_{1j'} - u_{3j'}) = 2[(h_5 + h_7 + h_8 + h_9) - (h_6 - h_5 + h_7 + h_8 + h_9)] = 2(h_{11} - h_5) \), since \( h_7 = h_6 \) and \( h_{11} = h_{10} \) by Property 1 above. Hence equation (6.2.12a) is satisfied for \( j=5 \). The same is true for all \( j = 2,3,\ldots,10 \).
Because equation (6.2.12a) is satisfied, the KKT condition (6.2.11a) is satisfied. We now show that all \( u_{ij} \) and \( u_{ij} \) are non-negative. This is obvious for \( u_{ij} = 2h_j \) and \( u_{ij} = 2h_{j+1} \), since all headways are positive. When \( u_{ij} = 2(h_j - h_{j+1}) \), constraint (6.2.5) is not binding and hence \( h_j \geq h_{j+1} \) by Property 2. In this case \( u_{ij} \) is also non-negative. Similarly by Property 3 we can show \( u_{ij} = -2(h_j - h_{j+1}) \geq 0 \) when constraint (6.2.2) is not binding.

It is also obvious that \( u_{ij} [b_{ij} - g_{ij}(h)] = 0 \) and \( u_{ij} [b_{ij} - g_{ij}(h)] = 0 \) always holds due to the construction of \( u_{ij} \) and \( u_{ij} \). Finally, we can set all \( u_{ij} = 0 \) and all KKT conditions given in (6.2.11) are satisfied. **Q.E.D.**

We also want to point out that if constraint (6.2.5) is not present or not binding anywhere, we can simply set \( u_{ij} = 2(h_j - h_{j+1}) \) for all \( j \) and prove optimality. This is because in this case we always have, in the optimal solution, \( h_j \geq h_{j+1} \) at a turning point, and \( h_j = h_{j+1} \) otherwise.

### 6.2.3 Numerical Examples

We present two examples with the same input data, one with and one without constraint (6.2.5).

The input data is given as follows:

\[
\begin{array}{cccccccccccc}
  i & 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
  d_i & 0 & 5.0 & 10.0 & 15.0 & 22.0 & 30.5 & 35.0 & 40.0 & 43.5 & 45.0 & 50.0 \\
  t_i - T & 2 & 7 & 12 & 17.5 & 23 & 27 & 32 & 40.5 & 44 & 47 & 52 \\
\end{array}
\]

The first vehicle considered for control is \( i=2 \). In both examples we illustrate the algorithm for the vehicle set \( I_m=(2,3,...,11) \) where \( m=9 \).

**Example 6.1**

In this example constraint (2.2.5) is relaxed. That is, scheduled departure time is not a constraint. The solution in this case gives the lower bound of the objective function value of [HPF]. In the following we illustrate algorithm HPF for the vehicle set \( I_m=(2,3,...,11) \).
Iteration 1: starting with $p_0 = 1$.

Step 1: $h^* = (d^0_{M+1,1} - d^0_{1,1}) / (M+1-1) = 50.0 / 10 = 5.0$.

Step 2: $d_{5,1} = 20.0 < d^0_{5,1} = 22.0$, stop. We have $p_1 = 5$.

Step 3: $h_1 = (d^0_{5,1} - d^0_{1,1}) / 1 = 22.0 / 4 = 5.5$, go to iteration 2.

Iteration 2:

Step 1: $h^* = (d^0_{11,1} - d^0_{6,1}) / (11-5) = 28.0 / 6 = 4.67$.

Step 2: $d_{6,1} = 22.0 + 4.67 = 26.67 < d^0_{6,1} = 30.5$, stop. We have $p_2 = 6$.

Step 3: $h_2 = (d^0_{6,1} - d^0_{5,1}) / 1 = 8.5 > h_1 = 5.5$, set $p_1 = 6$. Recompute $h_1 = (d^0_{6,1} - d^0_{1,1}) / 5 = 30.5 / 5 = 6.1$. Go to next iteration.

Iteration 2*:

Step 1: $h^* = (d^0_{11,1} - d^0_{6,1}) / (11-6) = 19.5 / 5 = 3.9$.

Step 2: $d_{7,1} = 30.5 + 3.9 = 34.4 < 35.0$, stop. We have $p_2 = 7$.

Step 3: $h_2 = (d^0_{7,1} - d^0_{6,1}) / 1 = 4.5 < h_1 = 6.1$, go to next iteration.

Iteration 3:

Step 1: $h^* = (d^0_{11,1} - d^0_{7,1}) / (11-7) = 15 / 4 = 3.75$.

Step 2: $d_{8,1} = 35.0 + 3.75 = 38.75 < 40.0$, stop. We have $p_3 = 8$.

Step 3: $h_3 = (d^0_{8,1} - d^0_{7,1}) / 1 = 5.0 > h_2 = 4.5$. Set $p_2 = 8$, recompute $h_2 = (d^0_{8,1} - d^0_{6,1}) / 2 = 9.5 / 2 = 4.75 < h_1 = 6.1$, go to next iteration.

Iteration 3*:

Step 1: $h^* = (d^0_{11,1} - d^0_{8,1}) / (11-8) = (50-40) / 3 = 3.33$.

Step 2: $d_{9,1} = 40.0 + 3.33 = 43.33 < 43.5$, stop. We have $p_3 = 9$.

Step 3: $h_3 = (d^0_{9,1} - d^0_{8,1}) / 1 = 3.5 < 4.75$. go to next iteration.

Iteration 4:

Step 1: $h^* = (d^0_{11,1} - d^0_{9,1}) / (11-9) = (50-43.5) / 2 = 3.25$.

Step 2: $d_{10,1} = 43.5 + 3.25 = 46.75 > 45$, $d_{11,1} = 46.75 + 3.25 = 50.0$. We have $p_4 = 11$.  

191
Step 3. $h_3 = 3.25 < h_3 = 3.5$. The algorithm terminates for vehicle set (2,3,...11).

The complete solution is given below:

<table>
<thead>
<tr>
<th></th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$h_i$</td>
<td>6.1</td>
<td>6.1</td>
<td>6.1</td>
<td>6.1</td>
<td>6.1</td>
<td>4.75</td>
<td>4.75</td>
<td>3.5</td>
<td>3.25</td>
<td>3.25</td>
</tr>
<tr>
<td>$d_j$</td>
<td>6.1</td>
<td>12.2</td>
<td>18.3</td>
<td>24.4</td>
<td>30.5</td>
<td>35.25</td>
<td>40.0</td>
<td>43.5</td>
<td>46.75</td>
<td>50.0</td>
</tr>
<tr>
<td>$\Delta d_i$</td>
<td>1.1</td>
<td>2.2</td>
<td>3.3</td>
<td>2.4</td>
<td>0.0</td>
<td>0.25</td>
<td>0.0</td>
<td>0.0</td>
<td>1.75</td>
<td>0.0</td>
</tr>
<tr>
<td>$u_i$</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1.35</td>
<td>0</td>
<td>1.25</td>
<td>0.25</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

The total cost is reduced from 138.3$rN$ to 132.275$rN$, about 4.4%. Total delay is 11 min.

**Example 6.2**

In this example we incorporate constraint (6.2.5). That is, each $d_j$ must be no larger than $\max(t_i - T, d_i^0)$, as stated in constraint (6.2.5). This value is calculated as follows.

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
<th>11</th>
</tr>
</thead>
<tbody>
<tr>
<td>$d_i^0$</td>
<td>0</td>
<td>5.0</td>
<td>10.0</td>
<td>15.0</td>
<td>22.0</td>
<td>30.5</td>
<td>35.0</td>
<td>40.0</td>
<td>45.0</td>
<td>45.0</td>
<td>50.0</td>
</tr>
<tr>
<td>$\max(t_i - T, d_i^0)$</td>
<td>2</td>
<td>7</td>
<td>12</td>
<td>17.5</td>
<td>23</td>
<td>30.5</td>
<td>35</td>
<td>40.5</td>
<td>44</td>
<td>47</td>
<td>52</td>
</tr>
</tbody>
</table>

Obviously, the first vehicle whose no-control departure time is bound by constraint (6.2.5) is vehicle 6. Thus we first solve [HPF] for the subset of vehicles (2,3,4,5,6).

**Stage 1 of the "While" loop: i'=2, Case 2:**

Iteration 0: last=6, $h^0 = \infty$, $p^0 = 1$, $p^0 = 6$, $d_i = 0$, $d_6 = d_6^0$. flag$_1(1)$=flag$_2(1)$=1.

Iteration 1: n=1, $p^0 = 1 < last=6$.

Step 1: $h^* = (d_6 - d_1)/(6-1) = 22/4 = 6.1$.

Step 2: Compute $d_j = d_{j+1} + h^*$ for 2$\leq j \leq 6$. $d_j = 12.2 > 12$, violating constraint (6.2.5) but not (6.2.2). Set flag$_1(2)$=flag$_2(2)$=0, flag$_1(3)$=0, and flag$_2(3)$=1. $j=3 < p^1 = 6$, set $p^1 = 3$.

Step 3: Set $d_3 = \max(t_3-T, d_3^0) = 12$. Go to iteration 1*.
Iteration 1*: $n=1, p^0 = 1, p^1 = 3$

Step 1: $h^* = (d_1 - d_0)/(3-1)=6$.

Step 2: Compute $d_j = d_j^0 + h^*$ for $2 \leq j \leq 3$. No other violation except $d_3 = 12$, binding constraint (6.2.5) but not (6.2.2). Flags are unchanged. $j=3= p^1$, go to step 4.

Step 4: $h^1 < h^0$ and $\text{flag}_i(p^0)=1$. No violation.

Step 5. Set $h^1 = h^* = 6$, $n=2$, $p^2=\text{last}=6$.

Iteration 2: $p^1 = 3$, $d_3=12$, $p^2 = 6$, $d_6 = 30.5$.

Step 1: $h^* = (d_6 - d_3)/(6-3) = 18.5/3 = 6.17$.

Step 2. $d_4 = 18.17 > 17.5$, violating constraint (6.2.5) but not (6.2.2). Set $\text{flag}_1(4)=0$, and $\text{flag}_2(4)=1$. $j=4 < p^2 = 6$, set $p^3 = 4$.

Step 3. Set $d_4 = \max(t_4-T, d_4^0) = 17.5$. $p^2-p^1 = 4-3 = 1$, $h^* = d_4 - d_3 = 17.5 - 12 = 5.5$.

Step 4. $h^2 = 5.5 < h^1 = 6$ while $\text{flag}_i(p^1) = \text{flag}_i(3)=0$, violation. Set $p^1 = p^2 = 4$, $n=1$. Go to Iteration 1**.

Iteration 1**: $p^1 = 4$, $d_4=17.5$.

Step 1: $h^* = (d_4 - d_1)/(4-1) = 17.5/3 = 5.833$.

Step 2: Compute $d_j = d_j^0 + h^*$ for $2 \leq j \leq 4$. No other violation except $d_4 = 17.5$ binding constraint (6.2.5) but not (6.2.2). Flags are unchanged. $j=4= p^1$, go to step 4.

Step 4: No violation.

Step 5. Set $h^1 = h^* = 5.833$, $n=2$, $p^2=\text{last}=6$.

Iteration 2*: $n=2$, $p^1 = 4$, $d_4=17.5$, $p^2=6$, $d_6 = 30.5$.

Step 1: $h^* = (d_6 - d_4)/(6-4) = 13/2 = 6.5$.

Step 2: $d_5 = 17.5 + 6.5 = 24 > 23$, violating constraints (6.2.5) but not (6.2.2). Set $\text{flag}_1(5)=0$, and $\text{flag}_2(5)=1$. $j=5 < p^2 = 6$, set $p^3 = 5$.

Step 3. Set $d_5 = \max(t_5-T, d_5^0) = 23$. $p^2-p^1 = 5 - 4 = 1$, $h^* = 23 - 17.5 = 5.5$ and go to Step 4.

Step 4. $h^* = 5.5 < h^1 = 5.833$ while $\text{flag}_i(p^1) = \text{flag}_i(4)=0$, violation. Set $p^1 = p^2 = 5$, $n=1$. Go to Iteration 1***.

Iteration 1***: $p^1 = 5$, $d_5=23$. 193
Step 1: \( h^* = (d_s - d_i)/(5-1) = 23/4 = 5.75 \).

Step 2: Compute \( d_j = d_{p^1} - h^* \) for \( 2 \leq j \leq 5 \). No other violation except \( d_5 = 23 \) binding constraint (6.2.5) but not (6.2.2). Flags are unchanged. \( j = 5 = p^1 \), go to step 4.

Step 4: No violation.

Step 5. Set \( h^1 = h^* = 5.75 \), \( n = 2 \), \( p^2 = \text{last} = 6 \). Go to iteration 2**.

Iteration 2**: \( n = 2 \), \( p^1 = 5 \), \( d_5 = 23 \), \( p^2 = 6 \), \( d_6 = 30.5 \).

Step 1: \( h^* = (d_6 - d_5)/(6-5) = 7.5 \).

Step 2: Compute \( d_6 = d_5 + h^* = 30.5 \). No other violation except \( d_6 = 30.5 \) binding both constraints (6.2.5) and (6.2.2). Set \( \text{flag}_1(6) = 1 \), and \( \text{flag}_2(6) = 1 \). \( j = 6 = p^2 \), go to step 4.

Step 4: \( h^2 = 7.5 > h^1 = 5.75 \), and \( \text{flag}_2(p^0) = 1 \). No violation.

Step 5. Set \( h^2 = h^* = 7.5 \), \( n = 3 \), \( p^1 = \text{last} = 6 \).

Iteration 3: \( p^2 = \text{last} = 6 \), end.

For \( 2 \leq j \leq 6 \), \( 1 < n \leq 3 \), set \( h_2 = h_3 = h_4 = h_6 = h^1 = 5.75 \), since \( p^0 = 1 < 2, 3, 4, 5 \leq p^1 = 5 \); \( h_6 = h^2 = 7.5 \) since \( p^1 = 5 < 6 = p^2 \).

The algorithm is repeated for \( i = 7, 8, 9, 10 \) and the complete solution is given below:

\[
\begin{array}{cccccccccccc}
  i & 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\
  d^0_i & 5.0 & 10.0 & 15.0 & 22.0 & 30.5 & 35 & 40.0 & 43.5 & 45.0 & 50.0 \\
\text{max}(t_i - T, d^0_i) & 7 & 12 & 17.5 & 23 & 30.5 & 35 & 40 & 44 & 47 & 52 \\
  d_i & 5.75 & 11.5 & 17.25 & 23.0 & 30.5 & 35 & 40.0 & 43.5 & 46.75 & 50.0 \\
\Delta d_i & 0.75 & 1.5 & 2.25 & 1.0 & 0.0 & 0.0 & 0.0 & 0.0 & 1.75 & 0.0 \\
  h_i & 5.75 & 5.75 & 5.75 & 5.75 & 7.5 & 4.5 & 5 & 3.5 & 3.25 & 3.25 \\
  u_{ij} & 0 & 0 & 0 & 0 & 7.5 & 4.5 & 1.5 & 0.25 & 0 & \\
  u_{ij} & 0 & 0 & 0 & 1.75 & 4.5 & 5 & 0 & 0 & 0 & \\
\end{array}
\]

Total cost after holding is 133.56rN, around 3.4% reduction. Total delay is 7.75 min.

The solutions in the two examples are different. This shows that if we first solve the relaxed HPF without constraint (6.2.5), then check for violation on this constraint, the
resulting feasible control policy may not be optimal to [HPF]. For example, in Example 6.1 the optimal dispatching headway is 6.1 min. for vehicle 2, which does not violate constraint (6.2.5). But the true optimal dispatching headway should be 5.75 min. from Example 6.2.

Comparison between the two examples also shows that constraint (6.2.5) decreases the effectiveness of holding in terms of passenger waiting time. On the other hand, it also reduces total holding delay. There is clearly a trade off between them.

6.2.4 Computational Results

We applied Algorithm HPF to the Green Line data sets, where the starting headways in both directions are shown in Appendix C. Other input parameters are: \( h_0 = 0.5 \) min., \( c_0 = 0.37 \) min., \( \delta = 0.15 \) min. and \( r = 2.39 \) passenger/min. We present the main results below.

Comparison Between Results with and without the Terminal Schedule Constraint

The computational results show that, the schedule constraint (6.2.5) greatly restricts the effectiveness of holding in terms of passenger waiting time reductions. Table 6.2.1 shows the results with the constraint (where the minimal layover time is 3 min.), and Table 6.2.2 without the constraint. In all the five weekdays, cost reduction is much larger without constraint (6.2.5) in both directions, especially in Direction 2. As we can observe from the StDev column (standard deviation of starting headways), in Direction 2 the starting headway variance is much higher than in Direction 1. When constraint (6.2.5) is present, this means far fewer vehicles can be held in Direction 2 because they are already late. On the other hand, when constraint (6.2.5) is absent, holding can be more effective. The cost reduction difference is significant with and without this constraint: on average per morning peak between around 7% and 17% of the total passenger waiting time (or between 6,250 and 15,000 passenger minutes)!
<table>
<thead>
<tr>
<th>Data Set</th>
<th>M</th>
<th>No Control</th>
<th>with Holding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cost (Psg.min)</td>
<td>Avg. hdw (min.)</td>
</tr>
<tr>
<td>m1</td>
<td>36</td>
<td>49,114.70</td>
<td>4.66</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>50,409.10</td>
<td>4.85</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>55,021.20</td>
<td>5.02</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>55,002.80</td>
<td>4.87</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>43,244.70</td>
<td>4.81</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>38,361.50</td>
<td>4.60</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>37,232.50</td>
<td>4.92</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>39,846.70</td>
<td>4.95</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>38,018.00</td>
<td>4.82</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>30,820.70</td>
<td>4.87</td>
</tr>
<tr>
<td>Avg.</td>
<td></td>
<td>87,414.38</td>
<td>4.84</td>
</tr>
<tr>
<td>Total</td>
<td>346</td>
<td>437,071.90</td>
<td>4.84</td>
</tr>
</tbody>
</table>

Note: H=5min. % Change in Cost = Change in Cost/Cost*100%. Avg. cost is per morning peak. Avg. holding time is per held vehicle.

Table 6.2.1 HPF Results with Terminal Schedule Constraint (m=3)

<table>
<thead>
<tr>
<th>Data Set</th>
<th>M</th>
<th>No Control</th>
<th>with Holding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cost (Psg.min)</td>
<td>Avg. hdw (min.)</td>
</tr>
<tr>
<td>m1</td>
<td>36</td>
<td>49,114.70</td>
<td>4.66</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>50,409.10</td>
<td>4.85</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>55,021.20</td>
<td>5.02</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>55,002.80</td>
<td>4.87</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>43,244.70</td>
<td>4.81</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>38,361.50</td>
<td>4.60</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>37,232.50</td>
<td>4.92</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>39,846.70</td>
<td>4.95</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>38,018.00</td>
<td>4.82</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>30,820.70</td>
<td>4.87</td>
</tr>
<tr>
<td>Avg.</td>
<td></td>
<td>87,414.38</td>
<td>4.84</td>
</tr>
</tbody>
</table>

Note: H=5min. % Change in Cost = Change in Cost/Cost*100%. Avg. cost is per morning peak. Avg. holding time is per held vehicle.

Table 6.2.2 HPF Results without Terminal Schedule Constraint (m=3)
On the other hand, holding delay is also larger in the latter case: it is in an average of 114 min. per morning peak, which is three times as much as in the former case. Clearly there is a tradeoff between passenger waiting time reduction and vehicle/crew schedule delay. However, the week total delay is over 260 held vehicle trips (single-direction), which makes about 2.2 minutes per held vehicle trip. Such vehicle/crew schedule delay is quite small compared to the passenger waiting time saved. The implication from this comparison is that, with a small delay in operators time, a significantly larger amount of passengers waiting time can be saved. This suggests that transit agencies in the U.S. may want to reconsider relaxing the terminal schedule constraint in control practice. In System $F$, a consequence of such a relaxation is that the delay in current trip may also cause delay in the next round trip for the same vehicle/crew, hence the actual delay of that vehicle/crew accumulates over time. This is not a big problem in Green Line case, since the peak period contains only two round trips, or four single-direction trips, for each vehicle. This means the cumulative delay averages only about 8 minutes per held vehicle. In general, a transit agency can have a threshold cumulative delay for the period considered; if relaxing the terminal schedule constraint means an average cumulative delay per held vehicle larger than the threshold value, then the constraint should not be relaxed. The consequence of relaxing this constraint in System $G$ is much more complicated, as will be discussed in section 6.3.

**Comparison Between Different Rolling Horizon Sizes**

Unlike station skipping strategies which do not change vehicle dispatching time, holding at the dispatching station always does. Therefore, separability of System $F$ does not apply to the HP, which means rolling size has an impact on systematic cost. We thus computed the HPF results with different rolling sizes for comparison. The results were computed without constraint (6.2.5) in order to examine the rolling horizon size effects alone. Table 6.2.3 and Fig. 6.1 show the results.
<table>
<thead>
<tr>
<th>Rolling horizon Size</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost Reduction (%)</td>
<td>-14.24</td>
<td>-16.56</td>
<td>-17.23</td>
<td>-17.46</td>
<td>-17.61</td>
<td>-17.85</td>
</tr>
<tr>
<td>Holding Delay/Morn. (min.)</td>
<td>60.75</td>
<td>92.02</td>
<td>114.20</td>
<td>128.80</td>
<td>141.00</td>
<td>184.87</td>
</tr>
</tbody>
</table>

Table 6.2.3 System F: Comparison between Rolling Sizes

Fig. 6.1 (a). Comparison Between Rolling Horizon Sizes

The above shows that, the cost reduction becomes stable when \( m \geq 3 \). On the other hand, holding delay keeps increasing. The small additional reduction in passenger cost after \( m > 3 \) does not seem worth the price of significant increases of vehicle/crew schedule delay.
**Comparison between Deadheading and Holding**

In general, holding seems not to be as effective as deadheading when terminal schedule constraint (6.2.5) is present. On the other hand, holding is no doubt much more effective than deadheading when constraint (6.2.5) is relaxed, as is obvious in Table 6.2.4. This is not hard to understand, because the effectiveness of holding depends on the tightness of constraint (6.2.5). When there is a large headway between vehicles $i-1$ and $i$, it can be reduced either by holding $i-1$, or by deadheading $i$. When $i-1$ is late, it can not be held long, if at all. In such a situation deadheading may be more effective. Data set "w2" (i.e., Wednesday, Direction 2) is a typical example of this. In this data set, there is high variance in starting headways (see Table 6.2.1 and Appendix C), which indicates some unusually large headways. In the mean while, vehicles are also frequently late: only 3 vehicles (other than 1 and $M$) are not late. Consequently, only these vehicles can be held. In contrast, 9 vehicles could be deadheaded with benefits. This explains why in this case the overall cost reduction is much better with deadheading alone than with holding alone.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Holding with (6.2.5)</th>
<th>Holding without (6.2.5)</th>
<th>Deadheading</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>% Change in Cost</td>
<td># vehicles held</td>
<td>% Change in Cost</td>
</tr>
<tr>
<td>m1</td>
<td>-3.55</td>
<td>18</td>
<td>-5.33</td>
</tr>
<tr>
<td>tu1</td>
<td>-2.05</td>
<td>15</td>
<td>-4.34</td>
</tr>
<tr>
<td>w1</td>
<td>-5.27</td>
<td>5</td>
<td>-8.51</td>
</tr>
<tr>
<td>th1</td>
<td>-2.66</td>
<td>19</td>
<td>-6.29</td>
</tr>
<tr>
<td>f1</td>
<td>-2.22</td>
<td>12</td>
<td>-3.74</td>
</tr>
<tr>
<td>m2</td>
<td>-18.56</td>
<td>17</td>
<td>-38.77</td>
</tr>
<tr>
<td>tu2</td>
<td>-15.22</td>
<td>11</td>
<td>-31.09</td>
</tr>
<tr>
<td>w2</td>
<td>-7.50</td>
<td>3</td>
<td>-32.91</td>
</tr>
<tr>
<td>th2</td>
<td>-9.11</td>
<td>15</td>
<td>-31.23</td>
</tr>
<tr>
<td>f2</td>
<td>-12.79</td>
<td>12</td>
<td>-30.21</td>
</tr>
<tr>
<td>total</td>
<td>-7.16</td>
<td>127</td>
<td>-17.23</td>
</tr>
</tbody>
</table>

Table 6.2.4 System F: Comparison between Holding and Deadheading
6.3 Holding in System G

In System G, the effectiveness of holding may vary by the location of holding stations, due to uneven passenger demand along the route. Unlike in System F, here the best holding station is not necessarily the first in a direction. That is, there may exist stations other than $k_0$ which are more effective holding stations. Some previous research has suggested that the best holding point should be located just before high demand stations (see for example, Engelstein, 1983). In this section we first develop a model and algorithm for holding in System G (called HPG) at any control station $k$, and then perform computational experiments using the algorithm for holding at station $k_0$ and analyze the results. Finally, we discuss the choice of holding station.

6.3.1 Model Formulation

The model for a general HP in System G, called HPG, at a given holding station $k$, is formulated below:

[HPG] Minimize

$$f(d) = \sum_{j=1}^{k_0} \left[ \sum_{k'=k}^{k_0} r_{k'}(d_{j,k'} - d_{j-1,k'})^2 + u_c \left( \sum_{k'\in K_c} r_{k'}(d_{j,k'} - d_{j-1,k'})^2 \right) \right]$$

Subject to:

(6.3.2) $d_{j,k} - a_{j,k} - s_{j,k} \geq 0, \forall j \in I_m$

(6.3.3) $d_{i0m,k} - a_{i0m,k} - s_{i0m,k} = 0$

(6.3.4) $a_{j,k'} - d_{j-1,k'} \geq h_0, \forall j, k'$

(6.3.5) $d_{j,k_c} \leq \max(t_{j,c}, d_{j,k_c}^0), \forall j \in I_m$

where $i$ is the first vehicle in the $I_m$ set; $k$ is the holding station; $d_{i,k'} \forall k' \geq k$ and $a_{j,k} \forall j \in I_m$ are given; $t_{j,c}$ is the scheduled departure time at the next dispatching terminal $k_c$ minus minimal layover time.

Constraint (6.3.2) says that each vehicle $j$ can leave the holding station $k$ either when finished normal loading and unloading or later, while at other stations it leaves right after the last passenger boards. Constraint (6.3.3) is the boundary condition, which says the last
vehicle, $i+m$, in the set is not held at the control station $k$ (though its departure time, $d_{i+m,k}$, may change due to control on preceding vehicles). (6.3.4) is the safe headway constraint. (6.3.5) is the dispatching terminal schedule constraint. The decision variables are the departure times $d_{i,k}$. The intermediate variable definitions are given below.

(6.3.6) $a_{i,k'} = \max( d_{j,k} + R_k + 2\delta, d_{j-1,k} + h_0 )$, $\forall j,k': k' \geq k$

(6.3.7) $s_{i,k'} = c_0 + c_1 r_i h_{i,k} + c_2 q_k L_{i,k',1}$, $\forall j,k': k' \geq k$

(6.3.8) $L_{i,k'} = L_{i,k',1} + r_k h_{i,k} - q_k L_{i,k',1} = r_k h_{i,k'} + (1 - q_k) L_{i,k',1}$, $\forall j,k': k' \geq k$

6.3.2 Solution Algorithm

In the constraints of program [HPG], a nonlinear factor may be imposed by equation (6.3.6) for vehicle arrival time at a station. In other words, if no interstation stopping occurs for any vehicle anywhere along the route, [HPG] is likely to be convex; however this will not always be true in reality. In fact, with uneven dispatching headways and demand, vehicles often pair or bunch together. In this case [HPG] is most likely to be quasiconvex, as in the case of deadheading. In either case, one may employ a commercial nonlinear program solver to search for a local optimal solution of [HPG], but such a general search algorithm may be inefficient for our particular problem and its performance will not be guaranteed. In this subsection we develop an efficient and effective heuristic based on the properties of the problem.

In Chapter 3 we proved that, if the dispatching headway of vehicle $i$ is larger (smaller) than the maximal (minimal) headway in the non-blocked trip segment of vehicle $i-1$, then headways of vehicle trip $i$ in the same segment will keep increasing (decreasing), regardless of the demand pattern. We have also shown that, when the dispatching headway of $i$ is about in the mid range of $i-1$'s headways (i.e., $\max_k \{ h_{i,k'} \} + \min_k \{ h_{i,k'} \} / 2$, $k \leq k' \leq k$), the headway variation of $i$ along the route will likely be minimal. The main difference in the holon problem between System $G$ and System $F$ is that, here headways of $i$ along the route depend on both its dispatching headway and the headway of the preceding vehicle $i-1$. Therefore, in order to minimize total passenger waiting time in the entire system, we should not only consider evening
out dispatching headways across vehicles, but also take care of headways across stations for each vehicle. There is clearly a trade off between the two dimensions. For example, if we have no-control dispatching headways \((h_{1,k}, h_{2,k}, h_{3,k}) = (4,5,6)\), suppose only vehicle 2 is to be controlled, from algorithm HPF the best dispatching headway would be \((5+6)/2 = 5.5\) for vehicle 2 (if none of the constraints are binding). However, if \(\max_k\{h_{1,k}\} = 4.5\), \(\min_k\{h_{1,k}\} = 4\), and the middle range is \((4.5+4)/2 = 4.25\), then \(h_{2,k} = 5.5\) would no longer be the optimal choice to minimize total waiting times in the system. A better choice for \(h_{2,k}\) would probably be some number between 4.25 and 5.5 in this case. In general, maintaining small headway variation across stations for each vehicle will also help even out dispatching headways for future vehicle trips.

By analogy the same argument also applies to any control station other than the dispatching station: we just need to include the consideration of vehicle load differences at the control station.

The solution algorithm for HPG thus makes use of the similarity between HPG and HPF, and treats their difference by considering the impact of vehicle \(i\)'s headways. In particular, let us first consider the dispatching headways at station \(k\), \(h^*_{j,k}, i \leq j \leq i+m\), resulting from a slightly modified HPF algorithm (to take into account variable trip times). These \(h^*_{j,k}\)'s are as even as possible under all constraints for the vehicles in \(I_m\), but \(h^*_{i,k}\) may be larger or smaller than \(\{\max_k\{h_{i,k}\} + \min_k\{h_{i,k}\}\}/2\), \(k \leq k' \leq k_p\), denoted by \(h'\), which will increase headway variation of \(i\) along the route. Hence, we may improve the overall set cost by reconciling the headway evenness at station \(k\) across vehicles and the headway variation of vehicle \(i\) across stations. For this reason we search for a better \(h_{i,k}\) between \(h^*_{i,k}\) and \(h'\), while trying to maintain the evenness between other vehicles' starting headways. We present this algorithm below at any holding station \(k\).

**Algorithm HPG**

**Iteration 0.** Start with current no-control departure times of each vehicle \(j\) in the set \(I_m\) at \(k\). If constraint (6.3.5) is binding for \(i\), terminate the algorithm. Otherwise, run Algorithm HPF for \(I_m\) at station \(k\). The resulting departure times are denoted as
\(d_{j,k}^{(0)}\) and the departure headway \(h_{j,k}^{(0)}\) for each \(j\) in \(I_m\). Set a sufficiently small step size \(\omega\) (e.g., \(\omega = 0.05-0.25 \text{ min}\)). Compute \(h' = \left[\min_k\{h_{i-1,k}\} + \max_k\{h_{i-1,k}\}\right]/2, k \leq k' < k_i\), where \(k_i\) denotes the first station at which \(i-1\) was blocked by \(i-2\).

**Iteration \(n\).**

**Step 1.** Use the step size \(\omega\) to search for the optimal departure headway of \(i\) at \(k\), denoted as \(h^*\), where the set cost (6.3.1) is computed based on the no-control vehicle trajectories from station \(k_0\) to \(k\) and the new trajectories start with \(d_{i-1,k} + h^*\) and \(d_{j,k}^{(n-1)}, j > i\) at \(k\). Start the search with \(h^* = h_{i,k}^{(n-1)}\). If \(h_{i,k}^{(n-1)} > h'\), decrement \(h^*\) by \(\omega\); otherwise if \(h^* < h'\), increment \(h^*\) by \(\omega\). Stop the search when the set cost starts increasing or any constraint is binding and then stop. Compute the new trajectory of \(i\).

**Step 2.** Compute \(h'' = \left[\min_k\{h_{i,k}\} + \max_k\{h_{i,k}\}\right]/2, k \leq k' < k_i\). If \(h^* < h''\), increment \(h^*\) by \(\omega\), again stop when the set cost starts increasing or any constraint is binding.

**Step 3.** Set \(d_{i,k}^{(n)} = d_{i-1,k} + h^*\). Terminate the algorithm if the difference between \(d_{i,k}^{(n)}\) and \(d_{i,k}^{(n-1)}\) is smaller than a given criteria. Otherwise, run Algorithm HPF for vehicle set \(I_m - \{i\}\) at station \(k\) with \(d_{i,k}^{(n)}\) fixed, and using \(d_{j,k}^{(n-1)}, j > i\) as input. The resulted departure time for each vehicle \(j, i < j \leq i + m_i\), is \(d_{j,k}^{(n)}\), and the departure headway for each \(j\) is \(h_{j,k}^{(n)}\).

**Step 4.** Set \(n = n + 1\), and go to next iteration.

### 6.3.3 Computational Results for Holding at Station \(k_0\)

We implemented Algorithm HPG on the same Green Line data sets, where the actual starting headways in both directions are given in Appendix C. Other input parameters are: \(h_0 = 0.5\ \text{min.}\), \(c_0 = 0.2\ \text{min.}\), \(c_1 = 0.007\ \text{min.}\), \(c_2 = 0.008\ \text{min.}\), \(\delta = 0.15\ \text{min.}\), and \(\omega = 0.05\ \text{min.}\). For direction 1, \(\mu_c\) is set to 1, and for direction 2, \(\mu_c\) is set to 0. The morning peak demand profile of Green Line B line is shown in Appendix B. For the computational results discussed below, we consider two cases: with and without the terminal schedule constraint (6.3.5). When (6.3.5) is present, the data for dispatching times at station 1 is taken from the Green Line B line schedule.
**Rolling Horizon Size Effects**

To examine the impact of rolling size on holding in System G, we again set $m=1, 2, \ldots, 5$ for computational experiments. The overall results (total from all data sets for each setting) are shown in Table 6.3.1. Once again we see that at $m=3$ the cost reduction becomes stable in both cases with and without terminal schedule constraint (6.3.5) (see Fig. 6.2 below).

<table>
<thead>
<tr>
<th>Rolling Horizon Size</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>16</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Cost Reduction</td>
<td>89,711.55</td>
<td>94,777.80</td>
<td>97,127.95</td>
<td>97,784.11</td>
<td>98,219.20</td>
<td>98,893.91</td>
</tr>
<tr>
<td>Total Delay at $N$</td>
<td>-277.38</td>
<td>-260.30</td>
<td>-214.43</td>
<td>-189.71</td>
<td>-166.86</td>
<td>-111.99</td>
</tr>
<tr>
<td>Total Holding Time</td>
<td>197.89</td>
<td>216.49</td>
<td>266.62</td>
<td>289.89</td>
<td>311.09</td>
<td>359.13</td>
</tr>
</tbody>
</table>

without terminal schedule constraint (6.3.5)

| Total Cost Reduction | 133,806.46 | 148,180.56 | 153,148.64 | 154,598.86 | 155,472.62 | 157,117.49 |
| Cost Reduction (%)    | 27.05     | 29.95     | 30.96     | 31.25     | 31.42     | 31.76     |
| Total Delay at $N$    | -461.93  | -389.08  | -269.54  | -199.36  | -139.58  | -72.23   |
| Total Holding Time    | 365.08   | 431.53   | 536.26   | 589.10   | 632.82   | 748.52   |

Note: cost unit is pass.min., and time unit is minute.

Table 6.3.1 HPG: Overall Results by Rolling Horizon Size

![Fig. 6.2](image)

Unlike station skipping strategies, where $m=1$ may sometimes be a better choice than $m>1$ as discussed in Chapter 4, in the holding problem there is no negative impact from
choosing $m>1$. From the systematic effectiveness point of view, it is actually better to take into account more than one following vehicle trip when considering holding vehicle $i$. Since cost reduction becomes stable at $m=3$, 3 is probably the best choice of rolling horizon size for HPG in the Green Line case. This is consistent with the analysis presented in Chapter 3.

It is interesting to note that, although total holding time at station $k_0$ and total delay (computed as vehicle arrival time after holding minus before holding) at the terminal station $N$ tends to go up as rolling horizon size increases, overall vehicle arrival times at $N$ are all earlier than in no-control case (as the "--" sign indicates). This consequence of holding may be somewhat counter-intuitive, but it is simply the positive systematic impact of holding as will be discussed further in the next subsection.

**Effectiveness of Holding**

Tables 6.3.2a and 6.3.2b show more detail of the computational results for rolling horizon size $m=3$. As in System $F$, again we see the terminal schedule constraint is the bottleneck constraint. Without it, the overall improvement in passenger waiting times in a week of morning peak hours is as high as 31%, one third higher than with the constraint. Compared to deadheading (see Table 6.3.3), holding at $k_0$ is more effective without constraint (6.3.5), but may be less effective when (6.3.5) is in effect, depending on the tightness of the constraint. The limit of holding alone is in situations where the total headway in the vehicle set is large, causing a longer holding time than later vehicles. This implies a combined strategy may do better in such a situation.

Another interesting observation is that a higher cost reduction does not necessarily mean lower headway variation. For example, in data set "wl", deadheading results in about a 1% higher cost reduction than holding, but the standard deviation of headway is higher in the deadheading case (1.52 vs. 1.45, see Table 6.3.3). This is not hard to understand, because at a skipped segment while the evenness of vehicle headways may be improved,
passenger’s waiting time are increased. This shows that minimizing vehicle headway variation does not necessarily minimize passenger waiting time.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>M</th>
<th>Cost (Psg.min)</th>
<th>StDev of hw</th>
<th>Change in Cost</th>
<th>% Change in Cost</th>
<th>StDev of hw</th>
<th>Delay at N (min.)</th>
<th>Holding Time (min.)</th>
<th># veh. held</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>36</td>
<td>62,466.84</td>
<td>1.84</td>
<td>-15,385.07</td>
<td>-24.63</td>
<td>0.83</td>
<td>-80.82</td>
<td>27.29</td>
<td>25</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>62,790.51</td>
<td>1.73</td>
<td>-9,895.23</td>
<td>-15.78</td>
<td>1.19</td>
<td>-40.99</td>
<td>20.75</td>
<td>17</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>67,756.57</td>
<td>2.17</td>
<td>-9,436.20</td>
<td>-13.93</td>
<td>1.45</td>
<td>-38.06</td>
<td>11.52</td>
<td>9</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>66,082.73</td>
<td>1.92</td>
<td>-4,554.50</td>
<td>-6.89</td>
<td>1.64</td>
<td>2.05</td>
<td>14.64</td>
<td>17</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>52,427.08</td>
<td>1.52</td>
<td>-10,528.96</td>
<td>-20.08</td>
<td>0.44</td>
<td>-42.23</td>
<td>18.95</td>
<td>23</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>38,820.45</td>
<td>5.74</td>
<td>-12,202.06</td>
<td>-31.43</td>
<td>3.70</td>
<td>-13.41</td>
<td>46.92</td>
<td>19</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>37,792.15</td>
<td>5.30</td>
<td>-12,054.72</td>
<td>-31.90</td>
<td>2.71</td>
<td>-2.01</td>
<td>39.43</td>
<td>16</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>38,118.50</td>
<td>5.55</td>
<td>-5,175.36</td>
<td>-13.58</td>
<td>4.62</td>
<td>-3.33</td>
<td>16.15</td>
<td>7</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>38,640.17</td>
<td>5.40</td>
<td>-9,160.57</td>
<td>-23.71</td>
<td>3.63</td>
<td>2.29</td>
<td>36.83</td>
<td>20</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>29,938.91</td>
<td>4.79</td>
<td>-8,735.28</td>
<td>-29.18</td>
<td>2.65</td>
<td>2.08</td>
<td>34.14</td>
<td>18</td>
</tr>
<tr>
<td>Aver.</td>
<td></td>
<td>98,948.78</td>
<td>3.60</td>
<td>-19,425.59</td>
<td>-19.63</td>
<td>2.29</td>
<td>1.56</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>346</td>
<td>494,743.91</td>
<td>97,127.95</td>
<td>-19,63</td>
<td>-19.63</td>
<td>214.43</td>
<td>266.62</td>
<td>171</td>
<td></td>
</tr>
</tbody>
</table>

Note: H=5min. % Change in Cost = Change in Cost/Cost*100%. Avg. holding time is per held vehicle. Avg. cost is per morning peak.

Table 6.3.2a HPG Results with Terminal Schedule Constraint (m=3)

Unlike in System F, where holding delay causes exactly the same amount of delay at the end of a trip, in System G holding a vehicle can often result in earlier arrival of the following vehicle at the ending terminal by reducing its demand and hence also its total travel time. In both Tables 6.3.2a and 6.3.2b we see that, in more than half of the cases, aggregate arrival times at station N are actually earlier rather than later, even though the holding times at station k₀ are all positive. Furthermore, in Table 6.3.2b, when the terminal schedule constraint is not present, the overall lateness at station N is improved even more! On the other hand, such improvement is mostly in Direction 1, while in Direction 2 we see some increase in the lateness. This is because more vehicles are late in Direction 2 than in Direction 1, and more late vehicles are held in Direction 2 without constraint (6.3.5). This again suggests that, as we discussed earlier for System F, while
there is some advantage to relaxing the terminal schedule constraint in practice, if this is
to be in effect we should also control the cumulative delay for each held vehicle.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>M</th>
<th>No Control</th>
<th>with Holding</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cost (Psg.min)</td>
<td>StDev of hdw</td>
</tr>
<tr>
<td>m1</td>
<td>36</td>
<td>62,466.84</td>
<td>1.84</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>62,700.51</td>
<td>1.73</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>67,756.57</td>
<td>2.17</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>66,082.73</td>
<td>1.92</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>52,427.08</td>
<td>1.52</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>38,820.45</td>
<td>5.74</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>37,792.15</td>
<td>5.30</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>38,118.50</td>
<td>5.55</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>38,640.17</td>
<td>5.40</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>29,938.91</td>
<td>4.79</td>
</tr>
<tr>
<td>Aver.</td>
<td></td>
<td>98,948.78</td>
<td>3.60</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>494,743.91</td>
<td>153,148.64</td>
</tr>
</tbody>
</table>

Note: H=5min. % Change in Cost = Change in Cost/Cost*100%. Avg. holding time is per held vehicle. Avg. cost is per morning peak.

Table 6.3.2b HPG Results without Terminal Schedule Constraint (m=3)

<table>
<thead>
<tr>
<th>Data Set</th>
<th>% Change in Cost</th>
<th>StDev of h</th>
<th># veh. held</th>
<th>% Change in Cost</th>
<th>StDev of h</th>
<th># veh. held</th>
<th>% Change in Cost</th>
<th>StDev of h</th>
<th># vehicles deadheaded</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>-24.63</td>
<td>0.83</td>
<td>25</td>
<td>-25.60</td>
<td>0.76</td>
<td>29</td>
<td>-18.82</td>
<td>1.16</td>
<td>15</td>
</tr>
<tr>
<td>tu1</td>
<td>-15.78</td>
<td>1.19</td>
<td>17</td>
<td>-23.19</td>
<td>0.54</td>
<td>27</td>
<td>-18.86</td>
<td>0.78</td>
<td>14</td>
</tr>
<tr>
<td>w1</td>
<td>-13.93</td>
<td>1.45</td>
<td>9</td>
<td>-25.96</td>
<td>0.68</td>
<td>26</td>
<td>-16.35</td>
<td>1.52</td>
<td>10</td>
</tr>
<tr>
<td>th1</td>
<td>-6.89</td>
<td>1.64</td>
<td>17</td>
<td>-22.05</td>
<td>0.64</td>
<td>29</td>
<td>-14.02</td>
<td>1.06</td>
<td>16</td>
</tr>
<tr>
<td>f1</td>
<td>-20.08</td>
<td>0.44</td>
<td>23</td>
<td>-20.97</td>
<td>0.31</td>
<td>24</td>
<td>-15.09</td>
<td>0.79</td>
<td>10</td>
</tr>
<tr>
<td>m2</td>
<td>-31.43</td>
<td>3.70</td>
<td>19</td>
<td>-49.48</td>
<td>1.27</td>
<td>30</td>
<td>-9.81</td>
<td>4.32</td>
<td>12</td>
</tr>
<tr>
<td>tu2</td>
<td>-31.90</td>
<td>2.71</td>
<td>16</td>
<td>-42.80</td>
<td>1.29</td>
<td>28</td>
<td>-8.84</td>
<td>3.23</td>
<td>12</td>
</tr>
<tr>
<td>w2</td>
<td>-13.58</td>
<td>4.62</td>
<td>7</td>
<td>-40.43</td>
<td>1.97</td>
<td>29</td>
<td>-6.21</td>
<td>4.75</td>
<td>9</td>
</tr>
<tr>
<td>th2</td>
<td>-23.71</td>
<td>3.63</td>
<td>20</td>
<td>-43.26</td>
<td>1.20</td>
<td>33</td>
<td>-10.97</td>
<td>3.62</td>
<td>13</td>
</tr>
<tr>
<td>f2</td>
<td>-29.18</td>
<td>2.65</td>
<td>18</td>
<td>-39.94</td>
<td>0.66</td>
<td>29</td>
<td>-7.31</td>
<td>3.56</td>
<td>7</td>
</tr>
<tr>
<td>total</td>
<td>-19.63</td>
<td>2.29</td>
<td>171</td>
<td>-30.96</td>
<td>0.93</td>
<td>284</td>
<td>-13.70</td>
<td>2.48</td>
<td>118</td>
</tr>
</tbody>
</table>

Table 6.3.3 System G: Comparison between Holding and Deadheading
Fig. 6.3 plots vehicle trajectories after holding computed for Wednesday morning data. It again shows the dispatching headway effects discussed in Chapter 3: evening out the dispatching headways has also served to reduce headway variation along the route.

---

Fig. 6.3a Wednesday: Vehicle Trajectories after Holding with Terminal Schedule Constraint

Fig. 6.3b Wednesday: Vehicle Trajectories after Holding without Terminal Schedule Constraint
Compared to the no-control trajectories (Fig. 3.2), Fig. 6.3a shows clear improvements in overall headway variation, and Fig 6.3b shows the more significant improvement attained without the terminal schedule constraint. Interstation stopping of vehicles when blocked by preceding vehicle is also reduced tremendously after holding. In the case of Wednesday morning, 10 of 34 vehicles were blocked at some station(s) in direction 1. After holding with constraint (6.3.5), there are 3 fewer such cases. After holding without constraint (6.3.5), there is no interstation stopping at all in direction 1!

**Algorithm Performance**

Since the holding problem involves continuous variables, the solution optimality from the heuristics is not clearly. To check the performance of the HPG algorithm, we also implemented an exhaustive search algorithm which searched for the optimal $h_{i,k}$ in a broad region $(h_0, d_{i+1,k} - d_{i-1,k})$. The results turn out to be extremely similar for the two algorithms. Table 6.3.4 lists the cost reduction resulting from both algorithms by data set.

<table>
<thead>
<tr>
<th>data set</th>
<th>with constraint (6.3.5)</th>
<th>without constraint (6.3.5)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HGP</td>
<td>exhaustive search</td>
</tr>
<tr>
<td>m1</td>
<td>-24.63%</td>
<td>-24.75%</td>
</tr>
<tr>
<td>tu1</td>
<td>-15.78%</td>
<td>-15.82%</td>
</tr>
<tr>
<td>w1</td>
<td>-13.93%</td>
<td>-13.93%</td>
</tr>
<tr>
<td>th1</td>
<td>-6.89%</td>
<td>-7.13%</td>
</tr>
<tr>
<td>f1</td>
<td>-20.08%</td>
<td>-20.09%</td>
</tr>
<tr>
<td>m2</td>
<td>-31.43%</td>
<td>-31.52%</td>
</tr>
<tr>
<td>tu2</td>
<td>-31.90%</td>
<td>-31.84%</td>
</tr>
<tr>
<td>w2</td>
<td>-13.58%</td>
<td>-12.63%</td>
</tr>
<tr>
<td>th2</td>
<td>-23.71%</td>
<td>-23.76%</td>
</tr>
<tr>
<td>f2</td>
<td>-29.18%</td>
<td>-29.02%</td>
</tr>
</tbody>
</table>

Table 6.3.4. Cost Reduction Resulted from Two Algorithms (m=3)

The differences between the two sets of results are negligible and are just due to discrete computation error. On the other hand, the computation time is much shorter with the HPG algorithm. For example, a single iteration of each algorithm for the entire data set...
f2 (29 vehicle sets with rolling horizon size m=3) took 8 seconds for HPG 8, but 30 seconds for the exhaustive search algorithm.

6.3.4 Solution Property of HPG

A fundamental difference between holding and station skipping strategies is that holding policies are substantially independent of demand pattern, although their effectiveness is not. That is to say, as long as the dispatching headway of i, and the maximum and minimum headways of i-1 are the same, the optimal holding policy for i will be changed very little by changing demand patterns. This property of holding is due to the dispatching headway effects proved in Chapter 3.

Figs. 6.4a and 6.4b plot, for directions 1 and 2 respectively, the no-control dispatching headway \( h_{i,k_0}^0 \), the new dispatching headway \( h_{i,k_0} \) that results from holding, and the midrange headway of the preceding vehicle. The midrange headway of vehicle i-1 is defined as \( (\max_k \{h_{i-1,k}\} + \min_k \{h_{i-1,k}\})/2 \). For both directions, the new dispatching headway of i corresponds quite closely with the mid range headway of i-1, although the demand patterns in the two directions are totally different.
The correspondence of the optimal dispatching headway with the previous headway midrange is a very attractive property. Because the solutions for the optimal dispatching headways are independent of demand patterns, only vehicle headway information is required as input. This makes "Algorithm HPG" a very general, and easy to implement, algorithm.

6.3.5 Choice of Holding Station

Based on the analysis in Chapter 3, we expect that the most effective holding station in System G will be the first station in the control direction. This is because the most even dispatching headways result in the least headway variation along the route; and the earlier the holding station in the route, the more stations will be benefited down the line. As it turns out, computational results support this \textit{a priori}.

The computational experiment is performed as follows. In each run we select a different station as the holding station and run Algorithm HPG with rolling horizon size \( m=3 \) for an entire morning peak. We repeat this process for every station in directions 1 and 2, respectively. Figs. 6.5a and 6.5b show the computational results for a morning (Wednesday) of the Green Line by direction. In either direction, the total cost reduction gets smaller as the holding station moves down the line.
While the benefit drops gradually in direction 1, the decrease is much more rapid in direction 2. This is because in order to take into account cost in direction 2, in direction 1 we set a large weight in the cost function for station 26 (the last station in the direction, or $N/2$). This weight is actually the cumulative passenger arrival rate in direction 2, which totals about 42 passengers/min., much higher than the highest actual arrival rate.
(17 passengers/min. at station 25) in direction 1. Thus cost computed at station 26 is dominant in this direction, and holding at any station in the direction will have a similarly large benefit. In addition, passenger demand (including both boardings and alightings) is higher at the end of this direction; before station 20 the demand curve is quite flat. Therefore, the curve of reduced cost in Fig 6.5a drops slowly up to station 20 as headway variance increases slowly (see Fig. 3.3). The reduced cost curve drops more rapidly when the holding station is later than station 20-22, because another demand peak occurs in this segment, which raises the cost due to both the increased vehicle headway variation and larger passenger arrivals.

In direction 2, however, the highest passenger arrival rate (30 passengers/min.) occurs at the beginning (station 28), and no weight is set for the ending station 52. Thus the largest benefit is gained when the holding station is no later than station 28. When holding is applied at or after station 29, the overall benefit of holding drops dramatically, as shown in Fig. 6.5b.

To see how headway variation across stations has changed after holding at $k_0$ in each direction, Fig 6.6 plots the standard deviation of headway by station in each direction, before and after holding (without constraint (6.3.5)). For the sake of clarity we plot one morning data set only; the other data sets are all very similar as we have seen from Fig. 3.3.

If Fig. 6.3 were not clear enough about headway variation change after holding, Fig. 6.6 should be sufficient. From Fig. 6.6 we see that, while both the magnitude and increasing speed of headway variance is high without control, after holding at $k_0$ the variance is kept quite low and flat along the route. Though there are still increases at peak demand stations (25,26 in direction 1, and 28, 30 in direction 2), the differences between the minimal and maximal value of the standard deviation are very low: 0.36 (=0.62-0.98) minutes in direction 1, and 0.53 (=1.61-2.14) minutes in direction 2. The results suggests that, the first station in a direction (i.e., $k_0$) is the most effective holding station. Additional holding stations around the high demand stations may, presumably, further
reduce headway variation and passenger waiting time, but the additional improvement is likely to be small.

Fig. 6.6a. Comparison of Headway Standard Deviation by Station (Dir 1)

Fig. 6.6b. Comparison of Headway Standard Deviation by Station (Dir 2)
To verify this, we again perform a computational experiment (without constraint (6.3.5)). For each direction, we set the first holding station as \( k_p \), and then examine all other stations by setting each of them as the second holding station. The differences in percent cost change between different additional holding stations are so small, one can not even tell such difference if we plot them in the same scale as Fig 6.2. We give the optimal results in Table 6.3.5.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>Opt. 2nd holding station</th>
<th>Maximal Cost Reduced</th>
<th>% Cost Reduced</th>
<th>Additional Cost Reduction</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>20</td>
<td>-16,432.42</td>
<td>-26.31%</td>
<td>0.71%</td>
</tr>
<tr>
<td>tu1</td>
<td>2</td>
<td>-14,736.29</td>
<td>-23.50%</td>
<td>0.31%</td>
</tr>
<tr>
<td>w1</td>
<td>5</td>
<td>-17,894.85</td>
<td>-26.41%</td>
<td>0.45%</td>
</tr>
<tr>
<td>th1</td>
<td>2</td>
<td>-14,854.46</td>
<td>-22.48%</td>
<td>0.43%</td>
</tr>
<tr>
<td>f1</td>
<td>20</td>
<td>-11,117.39</td>
<td>-21.21%</td>
<td>0.76%</td>
</tr>
<tr>
<td>m2</td>
<td>28</td>
<td>-19,431.68</td>
<td>-50.06%</td>
<td>0.58%</td>
</tr>
<tr>
<td>tu2</td>
<td>28</td>
<td>-16,290.01</td>
<td>-43.11%</td>
<td>0.31%</td>
</tr>
<tr>
<td>w2</td>
<td>28</td>
<td>-15,898.83</td>
<td>-41.71%</td>
<td>1.28%</td>
</tr>
<tr>
<td>th2</td>
<td>28</td>
<td>-16,916.73</td>
<td>-43.78%</td>
<td>0.52%</td>
</tr>
<tr>
<td>f2</td>
<td>28</td>
<td>-12,032.04</td>
<td>-40.19%</td>
<td>0.25%</td>
</tr>
</tbody>
</table>

Table 6.3.5 Results from Two Holding Stations

From Table 6.3.4, we see the differences with and without the additional holding station are too insignificant to be worth the effort. In all but one data set the additional cost reduction after holding at the additional station is less than 1%.

If the additional holding station is indeed desired, however, the computational results seem to indicate that the best choice for its location is either the second station, or a local peak demand station, at which passenger arrivals is a local maximum and departure headway variance is likely to increase. Fig. 6.7 plots cost curves resulted from holding at the additional station for data sets: "m1", "tu1", "w1" and "m2". The cost curve for data set "th1" is similar to "tu1", "f1" is similar to "m1", and all direction 2 data sets are very similar. A common phenomenon in all the cost curves is that the cost drops down the most at the second station, and afterward the cost change stays small but with an
Fig. 6.7. Choice of the Second Holding Station

increasing trend as the additional holding station is later in the direction. However small
the cost change is, we can still observe that the local minimum in each cost curve occur
around local peak demand stations: for example, station 5, 13, 20, 25 in Direction 1, and
station 28 in Direction 2.
6.3.6 Impact of Holding on Passenger Riding Times

Just as for station skipping strategies, holding reduces not only passenger waiting times but also riding times. Computational results show that the riding time reduction from holding is often higher than that achieved by deadheading. Table 6.3.6 shows these results.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>No Control</th>
<th>Holding (m=3)</th>
<th>Deadheading (m=3)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>with (6.3.5)</td>
<td>without (6.3.5)</td>
<td></td>
</tr>
<tr>
<td>m1</td>
<td>125,962.70</td>
<td>-1,774.02</td>
<td>-1.41</td>
</tr>
<tr>
<td></td>
<td>-1,818.19</td>
<td>-1.44</td>
<td>-1,473.20</td>
</tr>
<tr>
<td>tu1</td>
<td>127,902.65</td>
<td>-1,150.17</td>
<td>-0.90</td>
</tr>
<tr>
<td></td>
<td>-2,220.21</td>
<td>-1.74</td>
<td>-1,866.29</td>
</tr>
<tr>
<td>w1</td>
<td>129,330.85</td>
<td>-1,667.40</td>
<td>-1.29</td>
</tr>
<tr>
<td></td>
<td>-2,748.10</td>
<td>-2.12</td>
<td>-1,647.03</td>
</tr>
<tr>
<td>th1</td>
<td>136,174.07</td>
<td>-714.76</td>
<td>-0.52</td>
</tr>
<tr>
<td></td>
<td>-2,637.19</td>
<td>-1.94</td>
<td>-2,169.76</td>
</tr>
<tr>
<td>f1</td>
<td>111,593.53</td>
<td>-1,617.52</td>
<td>-1.45</td>
</tr>
<tr>
<td></td>
<td>-1,651.62</td>
<td>-1.48</td>
<td>-1,126.76</td>
</tr>
<tr>
<td>m2</td>
<td>70,348.71</td>
<td>-4,388.78</td>
<td>-6.24</td>
</tr>
<tr>
<td></td>
<td>-6,950.25</td>
<td>-9.88</td>
<td>-2,858.39</td>
</tr>
<tr>
<td>tu2</td>
<td>72,993.06</td>
<td>-4,954.33</td>
<td>-6.79</td>
</tr>
<tr>
<td></td>
<td>-6,573.42</td>
<td>-9.01</td>
<td>-2,725.62</td>
</tr>
<tr>
<td>w2</td>
<td>70,784.24</td>
<td>-1,663.51</td>
<td>-2.35</td>
</tr>
<tr>
<td></td>
<td>-5,449.74</td>
<td>-7.70</td>
<td>-1,952.61</td>
</tr>
<tr>
<td>th2</td>
<td>74,768.11</td>
<td>-3,326.52</td>
<td>-4.45</td>
</tr>
<tr>
<td></td>
<td>-6,108.52</td>
<td>-8.17</td>
<td>-3,178.92</td>
</tr>
<tr>
<td>f2</td>
<td>62,308.01</td>
<td>-3,514.41</td>
<td>-5.64</td>
</tr>
<tr>
<td></td>
<td>-4,638.87</td>
<td>-7.45</td>
<td>-1,209.42</td>
</tr>
<tr>
<td>total</td>
<td>982,165.93</td>
<td>-24,771.42</td>
<td>-2.52</td>
</tr>
<tr>
<td></td>
<td>-40,796.11</td>
<td>-4.15</td>
<td>-20,208.00</td>
</tr>
</tbody>
</table>

Table 6.3.6 System G: Riding Time Saving Comparison

When the terminal schedule constraint is present, 8 out of the 10 data sets show higher riding time savings from holding than from deadheading. Without this constraint the riding time savings are all much higher. The riding time savings are due to both reductions in vehicle dwell times and interstation stoppings. For example, in data set "m1", there were 10 vehicles blocked by their preceding vehicles at some stations without control. With holding this number reduces to 3. The riding time savings are even higher in direction 2, because the dwell times were more uneven and more vehicles were blocked without control. Going back to our discussion on objective functions at the beginning of this chapter, we confirm that the small holding delays (<2 min./vehicle) have indeed gained both large waiting time savings and significant riding time savings for the whole system.
CHAPTER 7

THE COMBINED CONTROL PROBLEM

7.1 Introduction

In the previous chapters we have seen that deadheading and expressing work in a totally different way from holding. We have also seen that holding is more smooth and also often more effective systematically than deadheading or expressing when each strategy is used alone. Furthermore, holding is easier to implement and has much less negative effects. This suggests that holding should be used more often than the other strategies. The question now is, if we combine the strategies how much better can we do than using holding alone. One observation from the results of holding in the previous chapter is that some vehicles were held much longer than others, due to large headways of later vehicles. Intuitively, in such a situation deadheading and expressing may play a role to supplement holding. For example, if vehicle $i$ is considered for control and it has a much shorter preceding headway than its following vehicle $i+1$'s, we can either hold $i$ for a time $t$, or deadhead/express $i+1$ to skip a number of stations $n$, or hold $i$ for a time $t' < t$ and also deadhead/express $i+1$ to skip a number of stations $n' < n$. The last strategy may do better than holding alone when the holding time of $i$ is reduced and deadheading/expressing $i+1$ is also beneficial. This motivates the study of combined control strategies. In this chapter
we study combined strategies first in System $F$ and then in System $G$. For the same reason discussed in Chapters 4 and 5, throughout the discussion that follows we assume:

A7.1 A vehicle trip can have at most one skipped segment.

7.2 Combined Control in System $F$

In Chapter 5 we showed that deadheading is more effective than expressing in System $F$. Since at a dispatching station we may apply either deadheading or expressing but not both for a vehicle, here we consider combining deadheading and holding. Furthermore, because holding requires much less restrictive headway conditions and can be used more often, we consider a combined control strategy that supplements holding with deadheading.

7.2.1 Model Formulation and Properties

The results of holding in the previous chapter showed that, constrained by feasible departure times at station $k_0$ (inequality (6.2.2) and future departure times at $k_{c+1}$ (inequality (6.2.5)), in the rolling horizon optimal solutions every so often there is a "turning point" vehicle whose departure time is binding on one or both of the constraints. Such vehicles either can not be held long enough, or can not be held at all. If constraint (6.2.5) is already binding on a vehicle without holding, that vehicle can not be held at all. As a result, the rolling horizon optimal headways are even between two consecutive "turning point" vehicles, but different on either side of such a vehicle. Because a beneficial deadheading policy requires headway condition $h_{i,k} > h_{i+1,k}$, naturally only a "turning point vehicle" which satisfies this condition should be considered for deadheading in a combined strategy. Based on these observations, in this section we consider a combined strategy for an $I_m$ set which contains one permissible deadheading vehicle after optimal holding.
For an I_m set with size m=1, the combined control problem is trivial: if vehicle i's preceding headway is larger than its following headway, consider deadheading; otherwise, consider holding. When m≥2, however, the problem gets more complicated. Let us first consider a general case m≥3, where i_1 is the first vehicle in the set which is to be dispatched at k_0, and there are other m vehicles following i_1, with i_1 + m being the boundary vehicle. Holding alone would be the best strategy if it can result in even headways in the set. Otherwise, when there is a "turning point" vehicle i in the set, the optimal headways resulting from holding alone are h' = h_i = L_i / m_i for i' ≤ i, where L_i = d_{i,k_0}^0 - d_{i'k_0}^0 and m_i = i'(i-1); and h'' = h_i = L_2 / m_2 for i' > i, where L_2 = d_{i+1k_0}^0 - d_{i,k_0}^0 and m_2 = i_1 + m - i. When h' > h'', the set cost may be improved by deadheading i. Since a larger cost reduction in deadheading i is usually associated with a larger value of h_i/h_{i+1} as shown in Chapter 4, we may increase h_i or decrease h_{i+1} within their feasible regions to obtain large deadheading benefits, while using holding to maintain even headways between the remaining vehicles. Thus the combined control problem in System F, called CCPF, has three dimensions in general in its decision variables: h_i, h_{i+1} and n. This problem can also degenerate into two or one dimensions, with either h_i and n, or h_{i+1} and n, or n alone, depending on the size of the I_m set and the position of the turning point vehicle i in the set. We formulate the nondegenerate CCPF first, and then consider its degenerate cases.

Consider a case where a turning point vehicle i with h' > h'' has been identified after trying holding alone, where h' is the optimal headway preceding i, and h'' is the optimal headway following i. Now let c( ) denote the 1-DH cost function for i as defined in Chapter 4 by equation (4.3.24), the CCPF is formulated as follows:

[CCPF] Minimize

\[(7.2.1) \, f(h_i, h_{i+1}, n) = \left\{ h_i^2 + h_{i+1}^2 + (m_1 - 1)[(L_1 - h_i)/(m_1 - 1)]^2 + (m_2 - 1)[(L_2 - h_{i+1})/(m_2 - 1)]^2 \right\} K/2
\]

\[+ c(h_i, h_{i+1}, n) \]

\[= \left\{ h_i^2 + h_{i+1}^2 + (L_1 - h_i)^2/(m_1 - 1) + (L_2 - h_{i+1})^2/(m_2 - 1) \right\} K/2 \]

\[-\Delta^2 n^3 + \left[(h_i - h_{i+1}) \Delta K\right] n^2 + \left[ h_i h_{i+1} - (h_i - h_{i+1}) \Delta K \right] n \]
Subject to:

(7.2.2) \( j(L_i - h_i)/(m_1 - 1) \geq d_{i-1}^0 - j d_{i-1,k_0}^0, 1 \leq j < m_1 \)

(7.2.3) \( h_{i+1} + j(L_i - h_i)/(m_2 - 1) \geq d_{i+1}^0 - j d_{i,k_0}^0, 0 \leq j < m_2 \)

(7.2.4) \( n(h_i - h_{i+1} + \Delta)/(2\Delta) \leq 0, n \in Z^+ \)

(7.2.5a) \( h_i \geq h^r \)

(7.2.5b) \( h_{i+1} \leq h^r \)

where \( i \) is the first vehicle in the \( I_m \) set, \( i \) is the turning point vehicle, \( h^0 > h^0_{i+1} \), \( m_1 = i(i-1)+1 \), \( m_2 = i_i+m - i > 1 \), \( L_i = d_{i,k_0}^0 - d_{i-1,k_0}^0 \), and \( L_2 = d_{i,m,k_0}^0 - d_{i,k_0}^0 \). \( h^r \) and \( h^r \) are the optimal headways preceding and following vehicle \( i \) after applying holding alone.

The first part of the objective function is the sum of passenger waiting times for each vehicle without deadheading, and the second part, \( c(h_i, h_{i+1}, n) \), is the cost reduction from deadheading \( i \). It is obvious that if \( c^*(h_i, h_{i+1}, n) = 0 \), then the optimal headways should be the result of holding alone. It is not possible that an optimal solution to CCPF would have \( c^*(h_i, h_{i+1}, n) > 0 \), since by reducing \( n \) to zero we can reduce \( c(h_i, h_{i+1}, n) \) to zero without changing \( h_i \) and \( h_{i+1} \). Therefore, we always have \( c^*(h_i, h_{i+1}, n) \leq 0 \) in an optimal solution to CCPF. Hence all 1-DH properties stated in Chapter 4 also apply to CCPF.

The first two constraints, (7.2.2) and (7.2.3), are feasible headway constraints, they restrict the resulting departure times from [CCPF] to be no earlier than in the no-control case. Constraint (7.2.4) is the feasible region constraint for the 1-DH problem as discussed in Chapter 4. Constraint (7.2.5) gives a lower bound on \( h_i \) and an upper bound on \( h_{i+1} \), respectively, as their optimal holding headways. Note that the "terminal schedule constraint" (6.2.5) does not appear here, because (7.2.5) ensures that the changes of departure times on both sides of \( i \) can only be earlier but not later than the departure times resulting from holding alone. That is, (6.2.5) will never be violated.

Let us now examine convexity of the real relaxation of [CCPF]. The first order derivatives of the objective function (7.2.1) with real-valued \( n \) are given by:

(7.2.6a) \( \partial f(h_i, h_{i+1}, n)/\partial h_i = [h_i(L_i - h_i)/(m_1 - 1)]K + \Delta n^2 + (h_{i+1} - \Delta K)n \)

(7.2.6b) \( \partial f(h_i, h_{i+1}, n)/\partial h_{i+1} = [h_{i+1}(L_i - h_{i+1})/(m_2 - 1)]K - \Delta n^2 + (h_i + \Delta K)n \)

221
\[ (7.2.6c) \partial f(h, h_{i+1}, n, \partial h) = -3\Delta^2 n^2 + 2[(h_i - h_{i+1})\Delta + K\Delta^2]n + [h_i h_{i+1} - (h_i - h_{i+1})\Delta K] \]

The second order derivatives are given by

\[ (7.2.7a) \partial^2 f(h, h_{i+1}, n, \partial h^2) = \left(1 + 1/(m_i - 1)\right)K = mK/(m-1) \]

\[ (7.2.7b) \partial^2 f(h, h_{i+1}, n, \partial h \partial h_{i+1}) = n \]

\[ (7.2.7c) \partial^2 f(h, h_{i+1}, n, \partial h \partial n) = 2\Delta n + h_{i+1}\Delta K \]

\[ (7.2.7d) \partial^2 f(h, h_{i+1}, n, \partial h^2_{i+1} = \left[1 + 1/(m_i - 1)\right]K = mK/(m-1) \]

\[ (7.2.7e) \partial^2 f(h, h_{i+1}, n, \partial h_{i+1} \partial h_i) = n \]

\[ (7.2.7f) \partial^2 f(h, h_{i+1}, n, \partial h_{i+1} \partial n) = -2\Delta n + h_i + \Delta K \]

\[ (7.2.7g) \partial^2 f(h, h_{i+1}, n, \partial n \partial h) = -6\Delta^2 n + 2[(h_i - h_{i+1})\Delta + K\Delta^2] \]

\[ (7.2.7h) \partial^2 f(h, h_{i+1}, n, \partial n \partial h_{i+1}) = 2\Delta n + h_{i+1}\Delta K \]

\[ (7.2.7i) \partial^2 f(h, h_{i+1}, n, \partial n \partial h_{i+1}) = -2\Delta n + h_i + \Delta K \]

So the Hessian is given by

\[ (7.2.8) \nabla^2 f(h_i, h_{i+1}, n) = \begin{pmatrix}
\frac{m_1}{m_i - 1}K & n & 2\Delta n + h_{i+1} - K\Delta \\
n & \frac{m_2}{m_{i+1} - 1}K & -2\Delta n + h_i + K\Delta \\
2\Delta n + h_{i+1} - K\Delta & -2\Delta n + h_i + K\Delta & 2\Delta[-3\Delta n + h_i - h_{i+1} + K\Delta]
\end{pmatrix} \]

which is a 3 by 3 symmetric matrix.

The determinants of the three submatrices along the major diagonal are given by:

\[ (7.2.9a) D_1 = m_i K/(m_i - 1) > 0 \]

\[ (7.2.9b) D_2 = [m_i K/(m_i - 1)][m_i K/(m_{i+1} - 1)]n^2 > K^2 - n^2 \geq 0 \text{ because } n \leq N/2 - 1 \leq K \]

\[ (7.2.9c) D_3 = [m_i K/(m_i - 1)][[m_{i+1} K/(m_{i+1} - 1)](-6\Delta^3 n + 2[(h_i - h_{i+1})\Delta + K\Delta^2]) + (-2\Delta n + h_i + \Delta K)^2] \]

\[ = 2((m_i m_i'((m_i - 1)(m_{i+1} - 1))K^2 - n^2)(h_i - h_{i+1} - \Delta n + \Delta(K - 2n))) \]

\[ -[m_i/(m_i - 1)]K(h_i + \Delta(K - 2n))^2 \]

\[ -[m_{i+1}/(m_{i+1} - 1)]K(h_i + \Delta(K - 2n))^2 \]
\[ +2n(h_{i+1} - \Delta(K-2n))(h_i + \Delta(K-2n)) \]

While no general conclusion can be made on the sign of \( D_3 \), the following empirical facts may help evaluate the possibility of \( D_3 \) being positive:

- \( K \) is usually much larger than \( h_i \) and \( h_{i+1} \).
- Due to holding vehicles before and in the current \( I_m \) set, the optimal headway difference \( h_i - h_{i+1} \) is small, and as a consequence the feasible region of \( n \) is small.

It can be easily observed from (7.2.9c) that when \( n < \frac{([h_i - h_{i+1}] / \Delta + K) / 3}{3} \), the first term of \( D_3 \) is positive. As discussed in Chapter 4, this condition is not restrictive in reality. The second and the third terms are always negative. Whether the last term is negative or not depends on whether \( h_{i+1} < \Delta(K-2n) \). Although when \( h_{i+1} > \Delta(K-2n) \) the last term will be positive, it always has a much smaller value than the first term. On the other hand, among the two negative terms, the second term in \( D_3 \), \(-[m_i K/(m_i-1)](h_i + \Delta(K-2n))^2\), has the largest absolute value due to \( h_i > h_{i+1} \). Comparing the first term and the second term in \( D_3 \) we see that, when \( h_i - h_{i+1} \) and \( n \) are small, the difference between \((h_i - h_{i+1} - \Delta n + \Delta(K-2n))\) and \((h_i + \Delta(K-2n))\) is also small. Therefore, for the first term to have larger absolute value means \( 2((m_1 m_2 [(m_1-1)(m_2-1)])K^2-n^2)\Delta \) must be larger than \([m_i/(m_i-1)]K(h_i + \Delta(K-2n))\).

When omitting the small value of \( n \) and \( m_2/(m_2-1) \), this implies \( 2K\Delta > h_i + \Delta K \) or \( K > h_i \). In other words, the larger is \( K - h_i \), the larger is \( D_3 \). When taking into consideration \( h_{i+1} \) and \( n \), a smaller \( h_{i+1} \) adds more positive value to the first term in \( D_3 \), and a larger \( h_{i+1} \) adds more negative value to other terms. The change of \( n \) does not have large impact because it changes both the positive and negative terms similarly.

To see the above analysis intuitively, let us look at an empirically "worst case" in a transit service with scheduled headway of \( H=5 \) minutes. Suppose \( h_i = 10 \), \( h_{i+1} = 4.5 \) minutes, \( K=25 > 2h_i \), \( \Delta = 0.62 \), and \( m_1 = m_2 = 2 \). In this case \( n^* \) is bounded by \((10 - 4.5 + 0.62)/1.24 = 4.9\). Then for \( n=4 \) we have \( \Delta(K-2n) = 0.62*17 = 10.54 \) and the first term in \( D_3 \) is

\[
2((m_1 m_2 [(m_1-1)(m_2-1)])K^2-n^2)(h_i-h_{i+1}-\Delta n+\Delta(K-2n))\Delta
\]

\[=2*(4*25^2 - 16)(3.02+10.54)*0.62 = 2*2484*13.56*0.62 = 41767.0.\]
and the sum of the remaining terms is

\[-[m_1K/(m_1-1)](h_i+\Delta(K-2n))^2-[m_2K/(m_2-1)](h_{i+1}-\Delta(K-2n))^2+2n(h_{i+1}-\Delta(K-2n))(h_i+\Delta(K-2n))\]

\[= -50(10+10.54)^2 -50(4.5-10.54)^2 + 8(10+10.54)(4.5-10.54) = -21094-1824-1643.2=-24561\]

In the above example, the positive term has larger value than the absolute value of the sum of the negative terms, hence \(D_1>0\). When \(K\) is fixed, the possibility of \(D_1\) being negative grows when \(h_i\) gets larger. As another example, suppose we change the value of \(h_i\) in the above example to an extreme case of 25 minutes, such that the upper bound of \(n\) is \((25-4.5+.62)/1.24 = 17\), and \(K-h_i = 0\). For \(n=17\), \(\Delta(K-2n) = 0.62*(-9)=-5.58\). In this case, the first term in \(D_1\)

\[2((m_1m_2[(m_1-1)(m_2-1)])K^2-n^2)(h_i-h_{i+1}-\Delta n+\Delta(K-2n))\Delta =
2*2211*(9.96-5.58)*0.62=12008.38\]

and the sum of the remaining terms is

\[-[m_1K/(m_1-1)](h_i+\Delta(K-2n))^2-[m_2K/(m_2-1)](h_{i+1}-\Delta(K-2n))^2+2n(h_{i+1}-\Delta(K-2n))(h_i+\Delta(K-2n))\]

\[= -50(25-5.58)^2 - 50(4.5+5.58)^2 + 34(25-5.58)(4.5+5.58) = -18856.82-5080.32+6655.62 =-17281.52\]

Now \(D_1<0\). In this case the second term has larger absolute value than the first term. Even with the last term also being positive, it did not help. If the difference between \(h_i-h_{i+1}\) is smaller and hence \(h_{i+1}\) is larger, say \(h_{i+1} =17\), \(n\) is bounded by \((8+0.62)/ 1.24=6.95\), and we have \(\Delta(K-2n) = 8.06\) for \(n=6\). In this case the first term is

\[2*2464*(1.05+8.06)*0.62=27834.3\], and the sum of the other terms is

\[-50(25+6.95)^2-50(4.5-6.95)^2+12(25+6.95)(4.5-6.95)=-52279.58\]

\(D_3\) is even more negative!

With holding being used systematically, \(h_i\) is expected to be much smaller than in the no-control case, and hence most often \(D_3\) will be positive. With all the three determinants being positive, the Hessian given by eq. (7.2.8) is positive definite. Because the
constraint set of [CCPF] is a polyhedral set, the real relaxation of [CCPF] is a convex program in this case.

**The Degenerate Cases**

It is possible that the "turning point vehicle" $i$ is the first vehicle $i_1$ in the set, in which case $m_1 = -(i_1-1) = 1$, or $i$ is the second to last vehicle in the set, in which case $m_2 = i_1 + m - (i_1 + m - 1) = 1$. In either case we can not simply use the model [CCPF] because there would be a term in $f()$ divided by zero. We thus consider a formulation for each of the two degenerate cases separately.

**Degenerate Case 1**: $i = i_1$

In this case, we have $h^0_i$ and $L_i = h^0_i$ fixed, $h^0_i > h^0 = L_2 / m_2$, where $L_2 = d_{i_1 + m k_0} - d_{i_1, k_0}$, $m_2 = m > 1$, and $h^0_{i_1} < h^0$. When considering deadheading $i$, we want to move between $h^0_{i_1}$ and $h^0$ to find both optimal $h_{i_1}$ and $n \geq 0$ by adjusting holding time of the remaining vehicles to maintain their even headways. The reason we expect in the optimal solution $h_{i_1} \leq h^0$ is because a larger cost reduction in deadheading $i$ is usually accompanied by a larger ratio of $h / h_{i_1}$, as shown in Chapter 4; and since $h^0_i$ is fixed we can only decrease $h_{i_1}$ to increase $h_i / h_{i_1}$. This degenerate case of CCPF denoted as CCP_1, can then be written as follows:

$$[\text{CCP}_1] \text{ Minimize}$$

$$(7.2.10) \quad f(n, h_{i_1}) = \{(h^0_i)^2 + h^0_{i_1} + (m-1)(L_2 - h_{i_1})/(m-1)\} K/2 + c(n, h_{i_1})$$

$$= \{(h^0_i)^2 + h^0_{i_1} + (L_2 - h_{i_1})^2/(m-1)\} K/2 - \Delta^2 n^3$$

$$+[(h^0_i - h_{i_1})\Delta + K\Delta^2] n^2 + [h^0_i - h_{i_1} - (h^0_i - h_{i_1}) \Delta K] n$$

Subject to (7.2.2), (7.2.4) and (7.2.5b).

We now examine convexity of the real relaxation of this program.

The first order partial derivatives are:

$$(7.2.11a) \quad \partial f(h_{i_1}, n) / \partial h_{i_1} = [h_{i_1} - (L-h_{i_1})/(m-1)]K - \Delta n^2 + [h^0_i + \Delta K] n$$

$$(7.2.11b) \quad \partial f(h_{i_1}, n) / \partial n = -3\Delta^2 n^2 + 2[(h^0_i - h_{i_1})\Delta + K\Delta^2] n + [h^0_i - h_{i_1} - (h^0_i - h_{i_1}) \Delta K]$$

225
The second order partial derivatives are:

\[ (7.2.12a) \frac{\partial^2 f(h_{i+1}, n)}{\partial h_{i+1}^2} = [1+1/(m-1)]K = mK/(m-1) \]

\[ (7.2.12b) \frac{\partial^2 f(h_{i+1}, n)}{\partial h_{i+1} \partial n} = \frac{\partial^2 f(n, h_{i+1})}{\partial n \partial h_{i+1}} = -2\Delta n + h_i^0 + \Delta K \]

\[ (7.2.12c) \frac{\partial^2 f(h_{i+1}, n)}{\partial n^2} = -\Delta^2 n + 2[(h_i^0 - h_{i+1}) + \Delta K] \]

So the Hessian is:

\[ \nabla^2 f(h_{i+1}, n) = \begin{pmatrix}
\frac{m}{m-1} K & -2\Delta n + h_i^0 + \Delta K \\
-2\Delta n + h_i^0 + \Delta K & 2\Delta[-3\Delta n + h_i^0 - h_{i+1} + \Delta K]
\end{pmatrix} \]

It is again a symmetric matrix, which is also a submatrix of the Hessian given by (7.2.8).

The determinants of the submatrices in this Hessian are

\[ (7.2.14a) \quad D_1 = mK/(m-1) > 0 \]

\[ (7.2.14b) \quad D_2 = 2[m/(m-1)]K\Delta(h_i^0 - h_{i+1} - \Delta n + \Delta(K-2n)) \]

Following the discussion in the nondegenerate case, \( D_2 \) will be positive when \( K \) is sufficiently larger than \( h_i^0 \), and this makes the relaxation of \( [\text{CCP}_1] \) also a convex program.

**Degenerate Case 2: \( i = i_1 + m-1 \)**

In this case, we have \( h_{i+1} \) and \( L_2 = h_{i+1} \) fixed, and \( h_{i+1}^0 < h_i^0 = L_i/m_i \), where \( L_i = d_{i,k_0} - d_{i-1,k_0} \) and \( m_i = m-1 > 1 \). Also, we have \( h_i^0 > h_i^0 \). To find the optimal \( n \) and \( h_i \), we move \( h_i \) between \( h_i^0 \) and \( h_i^0 \) by adjusting the holding time of other vehicles. This degenerate case, denoted as \( \text{CCP}_2 \), is formulated as follows:

**[CCP] Minimize**

\[ (7.2.15) \quad f(n, h_i) = \{ h_i^2 + (h_i^0 - h_i)^2 \} + (m_i - 1)[(L_i - h_i)/(m_i - 1)]^2 \] \[ = \{ h_i^2 + (h_i^0 - h_i)^2 + (L_i - h_i)^2/(m_i - 1) \} K/2 - \Delta^2 n^3 \]

\[ +[(h_i - h_i^0) + \Delta K]^2 n^2 + [h_i^0 - (h_i - h_i^0) + \Delta K] n \]

Subject to (7.2.3), (7.2.4) and (7.2.5a).

In line with the first two programs, we have the Hessian for the real relaxation of (7.2.15) as follows
\[ (7.2.16) \nabla^2 f(h_i, n) = \begin{pmatrix} \frac{m_2}{m_1} K & \frac{2\Delta n + h^0_{i+1} - \Delta K}{2\Delta n + h^0_{i+1} - \Delta K} \\ \frac{2\Delta n + h^0_{i+1} - \Delta K}{2\Delta n + h^0_{i+1} - \Delta K} & \frac{2\Delta [3\Delta n + h_i - h^0_{i+1} + \Delta K]}{2\Delta n + h^0_{i+1} - \Delta K} \end{pmatrix} \]

The first element of the matrix is positive. The determinant of the matrix is

\[ (7.2.17) D_2 = 2\Delta [m_1/(m_1-1)]K(-3\Delta n + h_i - h^0_{i+1} + \Delta K) - (2\Delta n + h^0_{i+1} - \Delta K)^2 \]

Under the same conditions for 1-DH, we have \(-3\Delta n + h_i - h^0_{i+1} + \Delta K > 0\). Because \(m_2/(m_2-1) > 1\), we next show that \(D_2 > D_2' = 2\Delta K(-3\Delta n + h_i - h^0_{i+1} + \Delta K) - (2\Delta n + h^0_{i+1} - \Delta K)^2 > 0\).

\[ (7.2.18) D_2' = 2\Delta K(-3\Delta n + h_i - h^0_{i+1} + \Delta K) - (2\Delta n + h^0_{i+1} - \Delta K)^2 \]

\[ = -6\Delta^2 Kn + 2(h_i - h^0_{i+1})\Delta K + 2\Delta^2 K^2 - 4\Delta^2 n^2 - 4\Delta K^2 + 4\Delta^2 Kn - 2\Delta n + 2\Delta K + (h^0_{i+1})^2 \]

\[ = -2\Delta^2 Kn + 2(h_i - h^0_{i+1})\Delta K + 2\Delta^2 K^2 - 4\Delta^2 n^2 - 4\Delta n + 2\Delta K + (h^0_{i+1})^2 \]

\[ = 2\Delta K(h_i - h^0_{i+1} - \Delta n) + 2\Delta (K^2 - 4n^2) + 2\Delta (K - 2n) + (h^0_{i+1})^2 \]

Since \(\Delta n < (h_i - h^0_{i+1})\), when \(n \leq K/2\), we have \(D_2 > 0\) and hence \(D_2 > 0\). In this case the real relaxation of [CCP2] is a convex program. Again, the condition of \(n \leq K/2\) is easily satisfied because of the small headway ratio \(h_i/h^0_{i+1}\) due to the holding of other vehicles.

### 7.2.2 Algorithms for CCPF

The three CCPF programs discussed above are nonlinear mixed integer programs with up to three dimensions in decision variables. Nonlinear mixed integer programs are usually difficult to solve. But because we formulated the problems with only a small number of dimensions where only one integer variable is involved, and their real-relaxations are likely to be convex, we can take advantage of these properties.

The proposed algorithm for solving a CCPF works as follows. We first solve its real-relaxation and obtain the real valued \(n', h_i\) and/or \(h^0_{i+1}\). We then use the optimal solution to the relaxed problem as the lower bound for a Branch-and-Bound method to obtain the integer \(n\) and new \(h_i\) and/or \(h^0_{i+1}\). Since \(n\) is the only integer variable, we need to branch at most twice: once to add the constraint \(n = \lfloor n' \rfloor\), and once to add the constraint \(n = \lceil n' \rceil + 1\), where \(\lfloor n' \rfloor\) denotes the integer part of \(n'\). Branch-and-Bound is a very
effective and easy-to-use method when a lower bound can be determined. Another advantage is that it can be applied to any problem that can be lower bounded, whether or not the cost function or constraints are linear (Papadimitriou & Steiglitz, 1982).

The algorithm we use to solve the real-relaxation of a CCPF is a modified "downhill simplex method" with penalty functions on the constraints. Although it sounds similar, the downhill simplex method has nothing to do with the simplex method for linear programming. It was developed by Nelder and Mead (1965) as a multidimensional search method for minimization problems. This method does not use derivatives, and it is suitable for problems with a small number of dimensions where the computational burden is not too large. The algorithm must be started not just with a single point, but with \( ndim + 1 \) points, where \( ndim \) is the number of dimensions of the problem (see Press et al, 1988, for detailed description). It always converges to an (at least local) minimum. If one of the starting points is close to the optimal, the convergence is very quick. For our CCPF, we can select good starting points based on the properties analyzed. For example, one of the starting points in the nondegenerate case can be \((h, h_{i+1}, n)=(L_1/m_1, L_2/m_2, 0)\), which is the optimal holding-only solution. As we have shown in Chapter 4, the optimal \( n \) for a deadheading problem can be approximated by \( h/h_{i+1} \), so the second starting point can be \((h, h_{i+1}, n)=(L_1/m_1, L_2/m_2, h/h_{i+1})\). The other two points can then be given arbitrarily as long as three of the four form a full rank array.

The downhill simplex method was originally proposed for unconstrained minimization problems. To make use of this method for our constrained CCPF, we use penalty functions to transform the CCPF into an unconstrained problem. Since our constraints can all be written in the form of \( g(x) \leq 0 \), we use an auxiliary function of the form \( f(x) + \Sigma \mu_j \text{maximum}(0, g_j(x)) \), where \( \mu_j \) is a large positive number. See Bazarra et al (1993) for details of the penalty methods. Note that any non-linear optimization search method which converges quickly given good starting point(s) could be used to replace the downhill simplex. The choice of the downhill simplex here is, in large part, due to its ease of coding.
We now give a complete algorithm for CCPF, including nondegenerate and degenerate cases, as follows.

**Algorithm CCPF**

For each vehicle set \( I_m \) in System \( F \):

0. Set \( j = i + m \) with given set size \( m > 0 \). If \( j > M \), set \( j = M \).

1. If \( m = 1 \): Call Algorithm RTDPF if \( h_{i}^{0} - h_{i+1}^{0} > \Delta \). If \( n^* > 0 \), deadhead \( i \) to station \( k_0 + n^* \), otherwise do not control \( i \) and go to \( i = i + 1 \). If \( h_{i}^{0} < h_{i+1}^{0} \), dispatch \( i \) with \( h_i = (h_i^{0} + h_{i+1}^{0})/2 \) and go to \( i = i + 1 \).

2. If \( m > 1 \): Call Algorithm HPF to the vehicle set \( \{ i, i+1, ... j \} \) and obtain the optimal holding-only headways \( \{ h_{i}^{p}, h_{i+1}^{p}, ..., h_{j}^{p} \} \).

3. Find the turning point vehicle \( i_j \) that has \( h_{i_j}^{p} > h_{i_j+1}^{p} \). If \( i_j < j \), search for \( j' > i_j \) that has \( h_{j'}^{p} > h_{j+1}^{p} \). If such a \( j' \) is found, set \( j = j' \). If no such \( i_j \) is found, dispatch \( i \) with \( h_i^{p} \) and go to \( i + 1 \). Otherwise set \( m_1 = i - (i - 1), m_2 = j - i, L_1 = d_{j+k_0}^{0} - d_{i-1+k_0}^{0}, \) and \( L_2 = d_{j+k_0}^{0} - d_{i+k_0}^{0} \).

4. If \( i = i \), set the three starting points of \( (h_{i+1}^{p}, n) \) for the downhill simplex algorithm as \((h_{i+1}^{p}, 0), (h_{i+1}^{p}, h_{i+1}^{p} / h_{i+1}^{p} + 1), (h_{i+1}^{p} - \Delta, h_{i+1}^{p} / h_{i+1}^{p} + 1) \). Write the constraints in \([CCPF]_1\) in the form of \( g_i(h_{i+1}^{p}, n) \leq 0 \), and set a large positive number \( \mu_i \) for each \( t \). Set the auxiliary function as \( f(h_{i+1}^{p}, n) + \sum \mu_i \max(0, g_i) \).

5. Else if \( i = j - 1 \), set the starting points for the downhill simplex algorithm in a similar way, but replace the first element of each point by \( h_{i}^{p}, h_{i}^{p}, h_{i}^{p} + \Delta \) respectively. Also set the auxiliary function as in Step 4 but replacing the objective function \( f \) and constraints by those in \([CCPF]_2\).

6. Else if \( m_1 > 1 \) and \( m_2 > 1 \), set the four starting points for the downhill simplex method as \((h_{i}^{p}, h_{i+1}^{p}, 0), (h_{i}^{p}, h_{i+1}^{p}, h_{i}^{p} / h_{i+1}^{p} + 1), (h_{i}^{p} + \Delta, h_{i+1}^{p} + \Delta, h_{i}^{p} / h_{i+1}^{p} + 2) \). Set the auxiliary function as in Step 4 but replacing the objective function \( f \) and constraints by those in \([CCPF]_2\).

7. Call the downhill simplex algorithm passing appropriate arguments for the corresponding CCPF case.

8. If the real-valued solution \( n = 0 \), dispatch vehicle \( i \) with headway \( h_i^{p} \) and go to \( i = i + 1 \). Otherwise if \( n > 0 \), first add the constraint \( n \leq \lfloor n' \rfloor \) and \( n \geq \lfloor n' \rfloor + 1 \) separately to the auxiliary function and call the downhill simplex algorithm separately. This will give two new sets of values of the original objective function and the real valued variables.
Choose the one with the lower objective function value, and dispatch vehicle \( i \) with corresponding \( h_i \).

\textit{End}

7.2.3 Computational Results

Applying the above algorithm to the Green Line data sets, we obtain the results shown in Table 7.2.1. Below we summarize a few important aspects of the results, where the terminal schedule constraint (6.2.5) is in effect whenever holding is involved.

\textit{Effectiveness of Control}

Fig. 7.1 shows cost reduction by control strategy for all 10 data sets. In System \( F \), headway variation along the route is caused by dispatching headway variance alone. Since the two directions in each data set have different dispatching headway patterns, it is interesting to compare the strategies by direction.

![Cost Reduction %](image)

Legend

1. Deadheading
2. Expressing
3. Holding
4. Combined

Data Sets

Fig. 7.1 Effectiveness of Control Strategies in System \( F \)

In direction 1 (the first 5 data sets), since the dispatching headway standard deviation is low (<2 min.), not much control (especially station skipping) is needed. So the combined control strategy has similar effectiveness to holding alone. In direction 2 (the last 5 data sets), however, the dispatching headway standard deviation is high (>3 min.). In this
direction, a lot more stations are skipped in the optimal station skipping control strategies. Furthermore, due to the terminal schedule constraint, holding alone is often not as effective as deadheading. In such circumstances the combined control strategy shows clear advantages. While each single strategy reduces passenger waiting time by 7.5% to 19% in direction 2, the combined strategy results in cost reductions between 21% to 27%.

<table>
<thead>
<tr>
<th>Data Set</th>
<th>M</th>
<th>Cost (Psg.min)</th>
<th>Avg. hdw (min.)</th>
<th>StDev of hdw</th>
<th># skipped stations</th>
<th># vehicles Controlled Dh. Held</th>
<th>Change in Cost</th>
<th>% Change in Cost</th>
<th>Holding time (min.)</th>
<th>StDev of hdw (min.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>36</td>
<td>49,114.70</td>
<td>4.66</td>
<td>1.37</td>
<td>5</td>
<td>4</td>
<td>14</td>
<td>-1,888.55</td>
<td>12.96</td>
<td>0.93</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>50,409.10</td>
<td>4.85</td>
<td>1.14</td>
<td>6</td>
<td>5</td>
<td>14</td>
<td>-1,314.75</td>
<td>9.12</td>
<td>0.70</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>55,021.20</td>
<td>5.02</td>
<td>1.66</td>
<td>7</td>
<td>7</td>
<td>5</td>
<td>-3,119.08</td>
<td>5.67</td>
<td>0.96</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>55,002.80</td>
<td>4.87</td>
<td>1.39</td>
<td>7</td>
<td>6</td>
<td>16</td>
<td>-1,759.77</td>
<td>7.99</td>
<td>0.96</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>43,244.70</td>
<td>4.81</td>
<td>1.01</td>
<td>2</td>
<td>2</td>
<td>11</td>
<td>-963.35</td>
<td>8.11</td>
<td>0.64</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>38,361.50</td>
<td>4.60</td>
<td>3.96</td>
<td>20</td>
<td>4</td>
<td>14</td>
<td>-10,472.19</td>
<td>28.43</td>
<td>1.98</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>37,232.50</td>
<td>4.92</td>
<td>3.53</td>
<td>20</td>
<td>5</td>
<td>11</td>
<td>-8,310.38</td>
<td>26.56</td>
<td>1.73</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>39,846.70</td>
<td>4.95</td>
<td>4.00</td>
<td>33</td>
<td>8</td>
<td>3</td>
<td>-8,014.37</td>
<td>7.84</td>
<td>2.33</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>38,018.00</td>
<td>4.82</td>
<td>3.49</td>
<td>32</td>
<td>7</td>
<td>14</td>
<td>-7,981.85</td>
<td>20.00</td>
<td>1.66</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>30,820.70</td>
<td>4.87</td>
<td>3.28</td>
<td>17</td>
<td>4</td>
<td>12</td>
<td>-6,405.23</td>
<td>17.40</td>
<td>1.51</td>
</tr>
<tr>
<td>Aver.</td>
<td></td>
<td>87,914.38</td>
<td>4.84</td>
<td>2.48</td>
<td></td>
<td></td>
<td>-10,045.90</td>
<td>1.31</td>
<td>1.34</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>346</td>
<td>437,071.90</td>
<td>149</td>
<td>52</td>
<td>114</td>
<td>-50,229.52</td>
<td>-11.49</td>
<td>148.97</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 7.2.1 CCPF Results with Terminal Schedule Constraint (m=3)

**Magnitude of Control**

When holding and deadheading are combined in CCPF, far fewer stations need to be skipped than under deadheading-only control, and also holding delay is much less than when holding is used alone. Fig. 7.2 shows these differences in the magnitude of control intervention.

Compared to deadheading alone, the total number of stations skipped in combined control is reduced by about half, from 278 to 149. The number of vehicle trips deadheaded is
also halved (from 104 to 52). This decreases the frequency of using deadheading from one in every 3 vehicle trips to one in every 7. Such a reduction will be very welcome by passengers, who do not like to see vehicles pass by without stopping. Fig. 7.2 also shows that, holding time is also significantly reduced under combined control.

Fig. 7.2 Comparison between Single and Combined Control Strategies

Compared to holding alone, total holding time of all data sets is reduced from 184 to 149 minutes; total number of vehicles held is reduced from 127 to 114; and average holding time per vehicle is reduced from 1.45 min. to 1.31 min. (see also Table 6.2.1 and Table 7.2.1). On the other hand, the total number of vehicle trips controlled here is 166, which is higher than under control of either deadheading-only (104) or holding-only (127).
Clearly, there is a tradeoff between the frequency of control actions, and the amount of control in each control action.

The above results show that, combined control not only is systematically more effective, but also much less severe in terms of control actions. In other words, the combined strategy results in a systematic effect that equalizes passengers’ costs and benefits.

7.3 Combined Control in System G

In the previous section, we noted that deadheading is always more advantageous than expressing in System F, and hence the combined control strategy considers only deadheading and holding. In System G, expressing and deadheading each has its own advantages, as shown in Chapters 4 and 5. The combined control in System G must then consider combinations of deadheading, expressing, and holding. The formulation of the combined control problem in System G (referred to as CCPG) thus is the general model, namely the objective function (2.7) with all constraints of RTDPG, RTEPG, and HPG.

Although a much more complicated problem than a single type of control, it can be simplified by the fact that at most one of the three strategies will be actually applied on a particular vehicle. The station skipping strategies and holding strategy are mutually exclusive because of their opposite functionality: the former shortens the preceding headway while the latter lengthens it. The two station skipping strategies, deadheading and expressing, are not both used on the same vehicle trip because it is unnecessary and will increase passenger frustration; that is why we have assumed that a vehicle can have at most one skipped segment in a direction.

Therefore, the combined control problem reduces to the problem of choosing which (if any) strategy is appropriate for each vehicle. This problem is more complicated than any single control problem we have studied so far, because when an impact set contains multiple vehicles, each may use a different control strategy. Fortunately, based on the results from previous chapters, the development of an efficient and effective heuristic for
this problem becomes quite straightforward. Next we develop such a heuristic for the combined control problem in System G (CCPG).

### 7.3.1 Heuristic Algorithm for CCPG

The logic of this algorithm is basically a hierarchical comparison structure. Since holding is a "smooth" control strategy which can be used most often with the least negative effects, for a vehicle set \( I_m \), we first try the holding strategy alone. We then try deadheading and expressing separately for \( i \) while holding other vehicles. The one skipping strategy which gives better set cost will be chosen to compare with holding-only solution. The best solution is then chosen. The efficiency and effectiveness of this hierarchical comparison algorithm is grounded in the efficiency and effectiveness of each individual strategy algorithm.

**Algorithm CCPG**

For vehicle set \( I_m \) in System G:

**Step 1.** Apply algorithm HPG alone to the set of vehicles \( i, i+1, \ldots, i+m \). Set \( d^* = d_{i,k_0} \). The resulted set cost is \( f^* \).

**Step 2.** Check the headway condition of vehicle \( i \) to decide whether station skipping is feasible. If it is feasible, go to step 3; otherwise terminate and \( i \) is held until \( d^* \).

**Step 3.** Try deadheading \( i \), starting from \( n=1, \ k_s = k_0+1 \).

Iteration \( n \): Let \( i \) skip \( n \) stations from \( k_0 \) to \( k_c-1 \). Compute the new trajectories of vehicles \( i \) to \( i+m \), and then apply algorithm HPG to vehicles \( i+1, \ldots, i+m \). Compute cost \( f^{(n)} \). If \( f^{(n)} > f^{(n-1)} \), stop; set \( n = n-1, f^d = f^{(n-1)} \). Otherwise increase \( n \) and \( k_c \) by 1, go to next iteration.

**Step 4.** Try expressing \( i \), starting from \( k_s = k_0, k_c = k_0+2 \).

**While** \( k_s < k_c - 2 \) **Do**

**While** \( k_c < k_i \) **Do**

Let \( i \) skip \( n \) stations from \( k_s+1 \) to \( k_c-1 \).

Compute the new trajectories of vehicles \( i \) to \( i+m \).

Apply HPG to vehicles \( i+1, \ldots, i+m \).
Compute $f^{n_0}$, where $n = k_\varepsilon - k_\varepsilon - 1$.

If $f^{n_0} > f^{n_1}, f^* = f^{n_0}, n^* = n - 1$, break;
else $k_\varepsilon = k_\varepsilon + 1$.

**End** inner while loop

$k_\varepsilon = k_\varepsilon + 1, k_\varepsilon = k_0 + n^*$.

$f^* = \min_{k_\varepsilon} (f^*)$.

**End** outer while loop

**Step 5.** If $\min(f^h, f^e, f^d) = f^h, i$ is held until $d^*$ and then dispatched. If $\min(f^h, f^e, f^d) = f^e$ and $n = k_\varepsilon - k_\varepsilon - 1 > 0$, $i$ is expressed from $k_\varepsilon$ to $k_\varepsilon$. If $\min(f^h, f^e, f^d) = f^d$ and $n^* > 0$, $i$ is deadheaded to $k_\varepsilon$.

### 7.3.2 Computational Results

We first note that the combined control problem is too complicated to solve with a manual control system, and hence a computerized control system is required. Following discussions in previous chapters, in this case the rolling horizon size $m = 3$ is the most appropriate and hence it is chosen for the computational tests. Also note that the terminal schedule constraint (6.2.5) is always in effect whenever holding is involved. We discuss below the computational results from running algorithm CCPG on the Green Line data sets.

**Efficiency of Algorithm CCPG**

The algorithm was coded in C and implemented on a 25 MHZ 486 PC. Table 7.3.1 shows the execution time (including input and output time) by data set. The average execution time of the procedure is about 2.5 seconds for a vehicle set of size $m = 3$ and a transit route of 26 stations per direction. The highest set average is about 5 seconds. Such a small computation time clearly meets the needs of real-time control system.
<table>
<thead>
<tr>
<th>data set</th>
<th>m1</th>
<th>tu1</th>
<th>w1</th>
<th>th1</th>
<th>f1</th>
<th>m2</th>
<th>tu2</th>
<th>w2</th>
<th>th2</th>
<th>f2</th>
<th>total</th>
</tr>
</thead>
<tbody>
<tr>
<td># vehicle sets</td>
<td>34</td>
<td>33</td>
<td>32</td>
<td>35</td>
<td>29</td>
<td>34</td>
<td>33</td>
<td>32</td>
<td>35</td>
<td>29</td>
<td>326</td>
</tr>
<tr>
<td>total execution time (sec.)</td>
<td>35</td>
<td>81</td>
<td>52</td>
<td>71</td>
<td>49</td>
<td>120</td>
<td>142</td>
<td>60</td>
<td>62</td>
<td>135</td>
<td>807</td>
</tr>
<tr>
<td>set avg. (sec., m=3)</td>
<td>1.03</td>
<td>2.45</td>
<td>1.63</td>
<td>2.03</td>
<td>1.69</td>
<td>3.53</td>
<td>4.30</td>
<td>1.88</td>
<td>1.77</td>
<td>4.66</td>
<td>2.48</td>
</tr>
</tbody>
</table>

Table 7.3.1 Execution Time of Algorithm CCPG

**Comparison between Single and Combined Strategies**

Table 7.3.2 summarizes the computational results of CCPG and Fig. 7.3 shows cost reduction by control strategy for all data sets. For the station skipping strategies, the results with no adjacent deadheading/expressing are used and for holding the results with the terminal schedule constraints are used in the comparison. Among the single strategies, the effectiveness of the two station skipping strategies, deadheading and expressing, is quite similar in terms of waiting time reductions, while expressing does slightly better in direction 1 and deadheading slightly better in direction 2. As discussed in Chapters 4 and 5, station skipping strategies are less effective in direction 2 due to the demand pattern. On the other hand, holding is more effective in direction 2 than the station skipping strategies, and is also more effective than holding in direction 1. This is due to the dispatching headway patterns. However, when the terminal schedule constraint is tight, it can be much less effective, as shown in the extreme case in Fig. 7.3 of data set "th1". In all cases, combined control shows clear advantages. The effectiveness of combined control is close to holding-only in direction 2, because the station skipping strategies are not very effective here. The cost reduction by combined control is as high as about 38%. The marginal benefit of combined control, compared to the most effective single strategy in each data set, varies between 0.4% (m1) and 7% (w1), with an average of about 4% (or about 4000 passenger-minutes) per data set. The average standard deviation of vehicle headways is also the lowest in the CCPG solution. Next we will see that this improvement is mainly due to the complementarity between station skipping and holding strategies, which does not exist when a single control strategy is used.

While the marginal benefits of the combined control may not seem very high, it has some structural advantages as we will see next.
<table>
<thead>
<tr>
<th>Data Set</th>
<th>M (Psg.min)</th>
<th>Cost (Avg. hdw (min.))</th>
<th>StdDev of hdw</th>
<th># Skipped stations</th>
<th># vehicles controlled</th>
<th>Dh. Exp.</th>
<th>Held</th>
<th>Change in Cost</th>
<th>% Change in Cost</th>
<th>Holding Time (min.)</th>
<th>StdDev of hdw (min.)</th>
</tr>
</thead>
<tbody>
<tr>
<td>m1</td>
<td>36</td>
<td>62,466.84</td>
<td>4.61</td>
<td>1.84</td>
<td>5</td>
<td>2</td>
<td>2</td>
<td>23</td>
<td>-15,737.97</td>
<td>-25.19</td>
<td>25.02</td>
</tr>
<tr>
<td>tu1</td>
<td>35</td>
<td>62,700.51</td>
<td>4.99</td>
<td>1.73</td>
<td>9</td>
<td>1</td>
<td>6</td>
<td>20</td>
<td>-14,279.90</td>
<td>-22.77</td>
<td>18.66</td>
</tr>
<tr>
<td>w1</td>
<td>34</td>
<td>67,756.57</td>
<td>4.97</td>
<td>2.17</td>
<td>13</td>
<td>2</td>
<td>6</td>
<td>12</td>
<td>-15,889.40</td>
<td>-23.45</td>
<td>15.54</td>
</tr>
<tr>
<td>th1</td>
<td>37</td>
<td>66,082.73</td>
<td>4.86</td>
<td>1.92</td>
<td>11</td>
<td>2</td>
<td>5</td>
<td>25</td>
<td>-13,759.51</td>
<td>-20.82</td>
<td>22.51</td>
</tr>
<tr>
<td>f1</td>
<td>31</td>
<td>52,427.08</td>
<td>4.91</td>
<td>1.52</td>
<td>4</td>
<td>2</td>
<td>2</td>
<td>22</td>
<td>-10,745.20</td>
<td>-20.50</td>
<td>16.15</td>
</tr>
<tr>
<td>m2</td>
<td>36</td>
<td>38,820.45</td>
<td>4.57</td>
<td>5.74</td>
<td>10</td>
<td>5</td>
<td>3</td>
<td>21</td>
<td>-14,493.13</td>
<td>-37.33</td>
<td>48.34</td>
</tr>
<tr>
<td>tu2</td>
<td>35</td>
<td>37,792.15</td>
<td>5.12</td>
<td>5.30</td>
<td>13</td>
<td>3</td>
<td>7</td>
<td>18</td>
<td>-13,839.72</td>
<td>-36.62</td>
<td>43.05</td>
</tr>
<tr>
<td>w2</td>
<td>34</td>
<td>38,118.50</td>
<td>4.94</td>
<td>5.55</td>
<td>11</td>
<td>7</td>
<td>2</td>
<td>9</td>
<td>-7,364.21</td>
<td>-19.32</td>
<td>17.42</td>
</tr>
<tr>
<td>th2</td>
<td>37</td>
<td>38,640.17</td>
<td>4.86</td>
<td>5.40</td>
<td>13</td>
<td>4</td>
<td>4</td>
<td>22</td>
<td>-10,380.67</td>
<td>-26.86</td>
<td>41.15</td>
</tr>
<tr>
<td>f2</td>
<td>31</td>
<td>29,938.91</td>
<td>5.00</td>
<td>4.79</td>
<td>9</td>
<td>3</td>
<td>4</td>
<td>19</td>
<td>-9,458.99</td>
<td>-31.59</td>
<td>33.64</td>
</tr>
<tr>
<td>Aver.</td>
<td></td>
<td>98,948.78</td>
<td>4.88</td>
<td>3.60</td>
<td></td>
<td></td>
<td></td>
<td>-25,189.74</td>
<td></td>
<td>1.47</td>
<td>1.61</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>494,743.91</td>
<td></td>
<td></td>
<td>98</td>
<td>31</td>
<td>41</td>
<td>191</td>
<td>-125,948.70</td>
<td>-25.46</td>
<td>281.48</td>
</tr>
</tbody>
</table>

**Note:** H=5min. % Change in Cost = Change in Cost/Cost*100%. Average cost and change in cost are per morning peak. Average holding time is per held vehicle.

**Table 7.3.2 CCPG Results (m=3)**

![cost reduction (%), m=3](image)

1-Deadheading
2-Expressing
3-Holding
4-Combined

**Fig 7.3 Comparison between Single and Combined Control in System G**
Solution Structure of Combined Control

Considering all data sets together, about 76% of vehicle trips (263 out of 346) are controlled. Among the controlled vehicles, about 27% are subject to station skipping strategies, and 73% are held. Holding is far more frequently used in the solution, and the portion of deadhead vehicle trips (11.8%) is slightly lower than the portion of express vehicle trips (15.6%) among the controlled vehicles. A number of characteristics in the solution structure are discussed below.

(1) The use of station skipping strategies increases the feasibility and effectiveness of holding. It is not surprising that in the combined control solution holding is used most frequently. Its "smooth" and nonrestrictive nature makes it the most feasible and effective strategy. Interestingly though, unlike in the System F solution where both total holding time and the number of vehicles held are reduced in combined control, here total holding time has increased from 267 minutes to 281 minutes, and the number of vehicles held is increased from 171 to 191, compared to holding alone (see Table 6.3.2 and Table 7.3.2). This is because the number of vehicles which are feasible for holding has increased due to station skipping control of other vehicles. For example, to hold vehicle \( i \) may be infeasible in HPG because it is blocked by vehicle \( i-1 \) and therefore \( i \)'s arrival time at the terminal is binding on the terminal schedule constraint. However, after deadheading/expressing vehicle \( i-1, \ i \) may no longer be blocked and may arrive at the terminal earlier than scheduled. In this case holding \( i \) becomes feasible. This shows that in combined control of System G, station skipping strategies can increase the feasibility and effectiveness of holding. Such interesting effects do not exist in System F because

![Graph showing holding delay per held vehicle (min.)](image)

Fig. 7.4 System G: Average Holding Time (m=3)
no vehicle can be blocked there due to the system operating assumptions. The increased frequency of holding also slightly decreased average holding time per held vehicle from 1.6 to 1.5 minutes. Fig. 7.4 shows average holding time per held vehicle for holding alone and combined control by data set.

(2) **The use of holding decreases the frequency and side effects of station skipping.** Compared to deadheading or expressing alone, the frequency of station skipping decreases dramatically in the CCPG solution. While 118 vehicles are deadheaded and 193 stations skipped in the RTDPG solution, and 143 vehicles are expressed and 252 stations skipped in the RTEPG solution, there are only 72 vehicles expressed or deadheaded, and 98 stations skipped in the CCPG solution. That is, both the frequency and magnitude of station skipping control are substantially decreased. This is no doubt good news to passengers, because no one who waits at a station likes to see the bus or train passing-by without stopping. Fig. 7.5 shows the number of skipped stations for station skipping and combined control by data set.

![# skipped stations](image)

**Fig. 7.5 System G: Magnitude of Control (m=3)**

In addition, the side effects of station skipping we discussed in Chapters 4 and 5 almost all disappear in the CCPG solution due to the use of holding. For example, in a RTDPG or RTEPG solution, vehicle \( i \) might "over-skip" stations in order to "push back" the trajectory of a following vehicle. As a result, passengers who wait for vehicle \( i \) and \( i+1 \) may be benefited less or not at all. In combined control, since trajectories of following
vehicles can be pushed back by holding them, over-skipping is no longer needed. An additional consequence of this is that the cost curves are more often quasi-convex than in RTDPG/RTEPG. This, needless to say, contributes to the performance of the solution algorithm.

(3) **Frequent control actions are needed for maximum effectiveness.** As we have seen from the CCPG solution, 76% vehicles are controlled. This implies that for real-time control to be fully effective, it should be a continuous process. On the other hand, less frequent control may result in more severe control actions, and increase passenger frustration induced by control actions.

(4) **CCPF can not replace CCPG.** Comparing Tables 7.3.2 and 7.2.1, we see the solution structures of CCPF and CCPG are very different, except that in both systems holding is more frequently used than station skipping. The main difference lies in separability of System F and non-separability of System G, and the different demand pattern. In other words, in System F control on a vehicle does not impact any other vehicle’s trajectory, while in System G every vehicle’s trajectory is impacted by the previous vehicle’s trajectory. This results in large differences in the solution structures. Also, as discussed in detail in Chapter 4, the demand pattern effect is sharp in direction 2 data sets. In these data sets, many vehicles have very large headway ratios $h_{i,k}/h_{i+1,k}$ at the dispatching station, therefore correspondingly large amounts of control are needed to "restore" normal headways. While these conditions did result in a large number of skipped stations in the CCPF solution (see Table 7.2.1), a much smaller number of skipped stations appears in Table 7.3.2 for the same data sets. This is because the highest demand station is the second in this direction, the cost to skip it is often higher than the benefits to other stations. This greatly restricts the length of a skipped segment.

Fig. 7.6(a-e) show a set of vehicle trajectories without control, with deadheading, with expressing, with holding, and with combined control respectively. The limitations of single control strategies and the effectiveness of combined control can be easily observed from these figures.
Fig. 7.6a. Vehicle Trajectories: Th1, System G, No Control

Fig. 7.6b. Vehicle Trajectories: Th1, System G, after Deadheading

Deadheaded vehicles:
- 4: to station 4
- 6: to station 6
- 8: to station 2
- 10: to station 2
- 12: to station 4
- 15: to station 4
- 16: to station 5
- 19: to station 6
- 21: to station 2
- 25: to station 2
- 34: to station 2
Fig. 7.6c. Vehicle Trajectories: Th1, System G, after Expressing
Expressed vehicles

2: from 4 to 6
4: from 3 to 6
5: from 2 to 4
6: from 1 to 6

8: from 1 to 4
10: from 3 to 5
12: from 1 to 5
13: from 1 to 3

15: from 2 to 5
16: from 1 to 6
19: from 2 to 7
21: from 1 to 3

25: from 1 to 3
30: from 16 to 18
34: from 1 to 3
36: from 16 to 18

Fig. 7.6d. Vehicle Trajectories: Th1, System G, after Holding

242
7.4 Effects of Randomness

As discussed in Chapter 1, in this thesis we have taken the deterministic modeling approach to attack the real-time control problems by assuming (invariant) average speeds across vehicles and expected demand over time at a station. System $G$ problems, such as RTDPG, RTEPG, HPG, or CCPG, also require as input forecast information on future vehicle movement and passenger demand. In the deterministic models developed in the thesis, we have implicitly assumed that such forecast information is available and perfect. However, in reality various degrees of stochastic conditions exist for all transit systems, and these nonstochastic assumptions do not hold. An important question is how well our models do when there are random disturbances and forecast errors. We expect that when vehicle running time and passenger demand variances and forecast errors are small, the deterministic models will still work well. In this section we test the model performance
of CCPG under various stochastic scenarios. We choose to test CCPG because it involves all the control strategies.

7.4.1 Sensitivity Test Design

In Chapter 1 we discussed a number of different sources of randomness for the models we are working with. In order to understand the relationship between these stochastic elements and the control policies we have investigated we will introduce randomness into the CCPG model and test the sensitivity of the results to this randomness. We will introduce different degrees of randomness by changing the extent to which variables will change around their deterministic values. Keeping in mind that a change of extent 0 gives us precisely the deterministic model discussed in the previous section, from this testing we can gain an understanding of the impact of randomness on the models and algorithms we have already considered.

Following on the discussion of Chapter 1, we focus on waiting and travel times as the elements for which random disturbances will be most important. The simplest approach then, would be to add random disturbances to the computed dwell times and interstation travel times. Since both quantities must be positive, and the disturbance can be either positive or negative, we could generate the random disturbances in a given range around the expected values. There are, unfortunately, significant problems in this simple approach. First, simply adding random disturbances to dwell times does not work in station skipping control cases, because dwell time is zero at a skipped station. Second, the dwell time at a station depends on passenger demand and therefore also previous controls that may have been used.

Another important point to note is that the cost function explicitly involves passenger arrival rates, whose randomness is the main contribution to dwell time randomness. The randomness in the passenger arrival rate is time dependent, and is independent of control. Therefore, it is necessary to move to a more fundamental view of the problem and add randomness into passenger arrival rates instead of the computed dwell times.
**Generation of Random Disturbances**

In designing the tests there are a number of important restrictions on how randomness can enter the model. Conceptually, the two most important things are that stochastic realizations not depend on control actions, and that there be substantial room for variation over time, across stations and between vehicles. To capture these ideas, we use the following guiding principles in adding randomness:

1. Any randomness in vehicle movement and passenger demand is independent of control.
2. Passenger arrival rate varies randomly over time and is station dependent.
3. Vehicle interstation travel time is random and depends on the vehicle trip and route segment.
4. At a decision time \( t_0 \), all vehicle trajectories before and at time \( t_0 - \tau \) are known, where \( \tau \) is a time gap.
5. Mean vehicle speed and expected passenger arrival rates (varying across station) are available.

These assumptions are reasonable and realistic. For the last assumption, the mean vehicle speed and expected passenger arrival rates can be estimated from historical data. The Green Line data sets we used for the deterministic models contain such data. An important consequence of these assumptions is that we need to break down time in a way that is completely independent of both train movements and control.

In particular, we divide the entire period (the morning peak period in our tests) into \( T \) short intervals indexed by \( t \). For each interval and each station, we generate different random disturbances within a given range of variance around the expected passenger arrival rates. For example, suppose the morning peak starts at 6:30 am and ends at 10:30 am, and each interval is 5 minutes long, we have 48 time intervals. Further, suppose that the standard deviation of passenger demand at each station is 10% around the mean, and we have 26 stations in a direction. Let \( r_k' \) denote the true (or realized) demand at station \( k \) in time interval \( t \), and \( r_k \) the expected demand in the period at station \( k \), and
\(\phi_{t,k}(-0.1, +0.1)\) the \(t \times k\)'th generation of uniformly distributed random numbers with minimum of -0.1 and maximum of +0.1. We use a computer random number generator to generate \(\phi_{t,k}(-0.1, +0.1)\) and compute \(r'_t = (1 + \phi_{t,k}(-0.1, +0.1))r_k\), for each station \(k\) in each time interval \(t\). The resulting "true" passenger arrival rates form an \(N\) by \(T\) matrix with entries \(r'_t\). In our example, this is a 26 by 48 matrix, with the entries varying by both time interval and station.

In generating disturbances for interstation running time it is not possible to use this same breakdown of time since vehicles will be at different places at different times depending on which controls are used. We can, however, view a vehicle as having an interstation running time for every station in the route. This running time can be stochastic but independent of clock time and control. Thus we generate the "true" free interstation travel times \(R_{i,k}\) by adding \(\phi_{i,k}\) for vehicle \(i\) at station \(k\) (representing the segment between \(k-1\) and \(k\)) to \(R_k\), the mean running time. So generated "true" travel times are independent of control because station skipping strategies involve \(\delta_k\) but not \(R_{i,k}\). The resulting "true" free interstation travel times form an \(M\) by \(N\) matrix with entries \(R_{i,k}\) whose values vary by both vehicle and route segment. For example, at the 0.1 COV we would have:

\[R_{i,k} = (1 + \phi_{i,k}(-0.1, +0.1))R_k\]

where \(\phi_{i,k}(-0.1, +0.1)\) is a uniform distribution with values ranging between -0.1 and 0.1.

**Systematic Biases in Vehicle Running Times**

When considering interstation running time variation it is appropriate to consider the possibility of systematic biases in interstation running time of vehicles. Vehicle speed can vary because of driver behavior and the mechanical condition of vehicles. In this case a slower than normal speed between two stations might be associated with a slower than normal speed between other stations in the route. To accommodate this we added a vehicle specific disturbance term that varies across vehicles but not from station to station. The new realized running time is \(R_{i,k} = (1 + \phi_i + \phi_{i,k})R_k\) where \(\phi_i\) is a uniformly distributed, vehicle dependent random variable. At 0.1 COV level, the parameters for \(\phi_i\)
are (-0.1,0.1), and the mean running time of a given vehicle will vary by 10% around the "global mean". For each vehicle and route segment, $\phi_{i,k}$ is generated in the same way as before. The realized running time of each vehicle between two stations depends on both $\phi_i$ and $\phi_{i,k}$. The total variation in running time possible will be the sum of the variation of each component. The distribution of running time, however, will no longer be uniform.

**Test Scenarios**

With this randomness in place we can now run the simulations, in which the control decisions will be made based on known past information and estimated future information, but will be realized in a stochastic environment. More specifically, we assume the AVI and APC information allows access to realized past vehicle movement and passenger demand, but future demand and vehicle running times are just estimates at the time the control decision is made. In this way, we have a fairly good idea of how sensitive the control policies developed are to the important sources of randomness in the real system.

Since the travel time randomness and passenger demand randomness may have different impacts, we consider different randomness scenarios as given in Table 7.4.1. As for the range of randomness, the Green Line AVI records show the coefficient of variation for composite travel times is around 0.05-0.1. In the interest of "worst case" testing, we chose to look at the range of 10-30% variation around the mean. Because we have no empirical data for the variances of passenger arrival rates, the same range of variances are assumed for the vehicle travel times $R_{i,k}$ and passenger arrival rates $\lambda_k'$. For the systematic biases in vehicle travel times (disturbances $\phi_i$) we test both the 0.05 and 0.1 COV levels with 0.3 COV level of $\phi_{i,k}$. That is, we run two additional tests with $R_{i,k} = (1 + \phi_i(-0.1, +0.1) + \phi_{i,k}(-0.3, 0.3))R_k$ and $R_{i,k} = (1 + \phi_i(-0.05, +0.05) + \phi_{i,k}(-0.3, 0.3))R_k$ respectively. This parameter setting is probably too high to be realistic, but it will give us a good idea about model sensitivities.
Table 7.4.1 Sensitivity Test Scenarios

<table>
<thead>
<tr>
<th>Scenario</th>
<th>1A</th>
<th>1B</th>
<th>2A</th>
<th>2B</th>
<th>3A</th>
<th>3B</th>
<th>4A</th>
<th>4B</th>
</tr>
</thead>
<tbody>
<tr>
<td>C.O.V. for $r'_k$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.10</td>
<td>0.30</td>
<td>0.10</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>C.O.V. for $R_{ik}$</td>
<td>0.10</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>0.10</td>
<td>0.30</td>
<td>0.30</td>
<td>0.30</td>
</tr>
<tr>
<td>C.O.V. for mean $R_{ik}$</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>0.05</td>
<td>0.10</td>
</tr>
</tbody>
</table>

Note: C.O.V. = Coefficient of Variation

We will denote the deterministic scenario as Scenario 0, where both C.O.V.’s for $R_{ik}$ and $r'_k$ are zero, and the expected values $R_k$ and $r_k$ are assumed.

We again use the Green Line data sets for the tests. The deterministic values are regarded as the expected values. Using the random parameters given in Table 7.4.1, 8 random data sets of no-control vehicle trajectories and passenger demand are generated for each direction each morning. This gives a total of 80 data sets for the 5 days.

**Forecast Information**

After generating the "true" free travel times and passenger demand incorporating randomness, we perform the test as follows. For each control problem considered for vehicle set $I_m = \{i, ..., i+m\}$, we set $t_0 = a_{i,k_0}$, the known arrival time of vehicle $i$ at the first station of the control direction. When evaluating the cost function for the vehicle set, all vehicle trajectories and passenger demand up to $t_0$ are computed using their "true" values as generated, since they are already realized. For forecast information on future vehicle trajectories and passenger demand, we consider the following two cases:

Case $m$: All vehicle trajectories and passenger demand after $t_0$ are predicted using the mean interstation travel time across vehicles (varying by segment), and expected passenger arrival rates over time (varying by station). This is the same as in the deterministic model. Here "$m$" stands for "mean".

Case $p$: All vehicle trajectories and passenger demand after $t_0$ are predicted with perfect information. That is, the true random values of $R_{ik}$ and $r'_k$ as we generated. Here "$p$" stands for "perfect".
Here Case $m$ represents a realistic forecasting method. Case $p$, on the other hand, plays two different roles in the tests. By comparing it with Case $m$, we can gain insights on the role of forecasting information. By comparing it with the deterministic case, we can learn about the impact of randomness on model structure. The logic of such comparisons are explained below.

**Impacts to be Tested**

The first impact of randomness to be tested is on the CCPG model properties. The system and operating properties of System $G$ shown in Chapter 3 may be changed by the randomness. The proof of such properties were based strictly on the assumptions that free interstation running times are constant across vehicles, and also that passenger arrival rates do not change with time. The CCPG model is developed based on these system properties. Because the random disturbances violate these assumptions, the difference between the solutions in the deterministic scenario and a random scenario with Case $p$ will reflect the model sensitivity to the property changes. Hence we want to use Case $p$ in the random scenarios to compare with the deterministic scenario.

Second, using Case $p$ to compare with Case $m$ in each random scenario will reflect two types of randomness impacts: in addition to the impact from system property changes there is also the impact from forecast information accuracy. A control policy for vehicle $i$ is generated by the CCPG model uses forecast information on future vehicle trajectories and demand. In a test with Case $m$, this forecast is based on the mean or expected values. The control policy based on this information will then be "performed" on $i$ in the "true" stochastic environment. Since we obtain and perform the control decisions in two different environments, one without and the other with accurate information, such a test will further show the model sensitivity to forecast information errors. By comparing Cases $p$ and $m$ for each random scenario, we can estimate how much impact forecast accuracy has on algorithm performance.
7.4.2 Results of Sensitivity Tests

For 6 unbiased randomness scenarios and 2 forecast information cases, 12 simulations runs are performed for each input data set. Another 4 simulation runs are performed for scenarios 4A and 4B with systematic biases in vehicle travel times. Table 7.4.2 and Fig. 7.7 show the test results (week totals).

A number of important results from the randomness tests are summarized below:

1. **Worse system performance in absence of control.** Compared to the deterministic Scenario 0, all but one (1A) stochastic scenarios have higher costs and all have higher standard deviation of headways in the no-control situation. Also, the costs and variances increase as the randomness increases. Under the same degree of variance, the no-control system performance with a single source of randomness is better than with both sources of randomness. Thus, the system performance is the worst in scenario 3 among the unbiased random scenarios. With the 0.05 COV level of biases in mean vehicle travel times and other things held equal, the system performance in scenario 4A is not worse than 3B. However, with the 0.1 COV level of biases in mean vehicle travel times, the system performance degrades sharply: the total waiting time has increased 10% compared with the deterministic scenario. All this is as expected.

2. **Lower cost reductions after control, but not much lower in unbiased random scenarios.** All cost reductions in the stochastic scenarios are lower, which is also expected. In all "A" unbiased scenarios (where the COV is 10% or 0 for either type of randomness), however, the difference between the highest and the lowest cost reduction is only 1.1%. The lowest cost reduction compared to the deterministic scenario is only 1.2% different. This shows the control models perform very well at the 10% COV level. At the 30% COV level, the model performance is both worse and less consistent. For unbiased scenarios, the difference between the highest and lowest cost reductions is about 5 percent which translates into about 5,000 passenger-minutes per morning. Though this
is a significant amount, the model generated controls are still decreasing the total passenger waiting time by (21%-25%) relative to the no control situation.

<table>
<thead>
<tr>
<th>Scenario</th>
<th>C.O.V. for $R_{ik}$</th>
<th>C.O.V. for $r'_k$</th>
<th>C.O.V. for mean $R_{ik}$</th>
<th>Total Cost</th>
<th>hdw. StDev</th>
<th>% Cost Reduction</th>
<th>hdw. StDev</th>
<th>% Cost Reduction</th>
<th>hdw. StDev</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
<td>494,743.90</td>
<td>3.60</td>
<td>25.46</td>
<td>1.51</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1A</td>
<td>0.10</td>
<td>0.00</td>
<td>0.00</td>
<td>493,985.88</td>
<td>3.59</td>
<td>24.28</td>
<td>1.67</td>
<td>25.17</td>
<td>1.65</td>
</tr>
<tr>
<td>1B</td>
<td>0.30</td>
<td>0.00</td>
<td>0.00</td>
<td>508,250.48</td>
<td>3.63</td>
<td>21.00</td>
<td>1.98</td>
<td>25.01</td>
<td>1.86</td>
</tr>
<tr>
<td>2A</td>
<td>0.00</td>
<td>0.10</td>
<td>0.00</td>
<td>497,374.78</td>
<td>3.62</td>
<td>25.38</td>
<td>1.63</td>
<td>25.35</td>
<td>1.63</td>
</tr>
<tr>
<td>2B</td>
<td>0.00</td>
<td>0.30</td>
<td>0.00</td>
<td>505,283.48</td>
<td>3.68</td>
<td>25.27</td>
<td>1.73</td>
<td>24.60</td>
<td>1.75</td>
</tr>
<tr>
<td>3A</td>
<td>0.10</td>
<td>0.10</td>
<td>0.00</td>
<td>498,880.74</td>
<td>3.62</td>
<td>24.69</td>
<td>1.69</td>
<td>24.30</td>
<td>1.75</td>
</tr>
<tr>
<td>3B</td>
<td>0.30</td>
<td>0.30</td>
<td>0.00</td>
<td>525,391.33</td>
<td>3.74</td>
<td>22.34</td>
<td>2.03</td>
<td>24.39</td>
<td>2.09</td>
</tr>
<tr>
<td>4A</td>
<td>0.30</td>
<td>0.30</td>
<td>0.05</td>
<td>518,278.71</td>
<td>3.72</td>
<td>19.11</td>
<td>2.09</td>
<td>23.32</td>
<td>1.86</td>
</tr>
<tr>
<td>4B</td>
<td>0.30</td>
<td>0.30</td>
<td>0.10</td>
<td>548,137.44</td>
<td>3.88</td>
<td>15.62</td>
<td>2.38</td>
<td>24.08</td>
<td>2.10</td>
</tr>
</tbody>
</table>

Table 7.4.2 Results of Randomness Tests

![Cost Reduction %](image)

Fig. 7.7 Cost Reduction after Control under Randomness

With systematic biases in vehicle travel times, however, the model performance declines with the COV levels of the biases. At the 0.05 COV level of biases, the total cost reduction has gone down to 19%, and to only 15% at the 0.1 COV level of biases. This decline is quite significant. This shows that the models developed in this dissertation are
very sensitive to systematic biases, which is as expected since we did not consider these biases in the modeling.

These results show that the deterministic models have an acceptable performance in a stochastic environment with coefficients of variation of up to 0.3, in either vehicle travel times or passenger arrival rates, or both, and with low (<0.05) systematic biases. Performance degrades as the amount of randomness and systematic bias increases, and this degradation is most pronounced for variation in vehicle running time. We should also note that the 0.3 COV for free interstation travel times and the 0.1 COV for travel time bias are very high, probably too high to be realistic. The tests we have performed indicate that the control models developed in this thesis will probably work well in most U.S. transit systems.

3. System properties insensitive to randomness. Compared to Scenario 0, Case p in all random scenarios show insignificant differences. The range of such differences is from only 0.2% to 1.1% in cost reduction. This means that randomness in interstation trip time and passenger demand has little impact on the properties of System G shown in Chapter 3, which in turn implies the robustness of the control models. This is to say, with accurate forecast information, the control models will perform well under all random scenarios considered here.

4. Larger impact from vehicle travel time variance. Free interstation vehicle travel time variation has a much larger impact than passenger arrival rate variation at any given value for their respective coefficients of variation. This is reflected most clearly by the lower cost reduction in scenario 1 relative to scenario 2. There are two probable explanations for this difference in impact: First, vehicle travel times in many segments are much larger quantity than dwell times (both in minutes), where the latter is influenced by passenger arrival rates (in passenger/minute). Second, the variation of dwell times around mean at each station across vehicles is much smaller than the variation of mean dwell times across stations. Since the mean dwell times are already captured by the deterministic model, the small added randomness apparently does not have a significant
impact. Furthermore, the time dependent randomness in passenger arrival rates may, to some degree, cancel each other over a long period, and this is probably why there is almost no difference between Case m and Case p in Scenario 2.

5. Greater sensitivity to forecast information accuracy for vehicle travel times.
Apparently, forecast information accuracy has larger impact in scenarios 1B, 3B, 4A and 4B, where the coefficient of variation of vehicle travel times are the highest (0.3). In all other stochastic scenarios, such impact is quite small. This indicates that forecast accuracy for vehicle travel times is more important than for passenger arrival rates. This is good news since real-time vehicle information is easier to obtain than passenger information. On the other hand, this also shows a limitation of the deterministic models: in an operating environment where vehicle speed variance is high, and forecast information for vehicle interstation travel times is not accurate, the models will have worse performance.
CHAPTER 8

CONCLUSIONS AND FUTURE RESEARCH

In this thesis we have studied the real-time control strategies deadheading, expressing and holding, both singly and in combination. We first investigated these strategies in a simplified transit system, called System F, with constant passenger arrival rates across stations and fixed vehicle dwell time assumed. The second type of transit system studied, called System G, is more general and realistic. In system G passenger demand varies across stations, and vehicle dwell times depend on both passenger boardings and alightings. We were able to obtain analytical results for the control problems in System F. While the control problems in System G are much more difficult, we combined analytical and empirical methods to understand their properties and structures, and then developed efficient and effective solution algorithms based on this understanding. The control problems considered in this thesis were modeled using deterministic formulations with stochastic testing conducted in order to test model sensitivity to various degrees of randomness. In this chapter we first summarize our findings and contributions, and then outline suggestions for future research.

8.1 Conclusions

In this section we summarize the main results from this dissertation.
8.1.1 System Properties

The fundamental property of System $F$ is its "separability". That is, as long as the dispatching time of each vehicle is unchanged, control on a vehicle has no impact on the trajectory of any other vehicle in the system. This property, which was proved in Lemma 3.1, Chapter 3, makes it possible to decompose the control problems into small subproblems without loss of generality, and greatly enhances the tractability of the control models.

System $G$, on the other hand, is non-separable and hence much more complicated. Vehicle trajectories interact with each other so that control on one vehicle impacts the trajectories of the vehicles that follow. In Chapter 3 we illustrated the "dwell time effects" in the absence of control as being the consequence of variability in passenger demand across stations. This leads to variance in vehicle dwell times across stations, increasing vehicle headway variances along the route and, ultimately, the well known "pairing" and "bunching" phenomena. This sharply increases passengers' waiting time, making real-time control more necessary, but also harder. However, we also proved or illustrated two important and useful properties of System $G$.

**Dispatching headway effects.** If a vehicle $i$'s dispatching headway is smaller than the minimal headway of its preceding vehicle $i-1$, $i$'s headway will be monotonically decreasing along the route until it is blocked or reaches the terminal. If $i$'s dispatching headway is larger than the maximal headway of $i-1$, its headway will be monotonically increasing along the route. This is proved in Proposition 3.1, Chapter 3. As a consequence of the dispatching headway effects, when $i$'s dispatching headway is in the midrange of the maximum and minimal headway of $i-1$'s, $i$'s headway variation along the route will be very small. The great significance of this property is its independence of demand pattern and hence its generality. It gives a convenient and accurate way to foresee vehicle headway patterns down the line regardless of demand pattern, and hence helps to suggest good control decisions early on.
**Trajectory change diminishing effect.** In a System $G$ type of transit service with $c_i r_i < 0.5$, the impact of a change in trajectory of vehicle $i$ on its following vehicles diminishes quickly. This was also discussed in Chapter 3. In particular, after a small number of vehicle trips, $m$, the impact of $i$'s trajectory change becomes insignificant. The value of $n$ depends on the value of $c_i r_i$, which can be easily estimated. This property is most useful for determining the length of rolling horizon, which is important for modeling dynamic systems such as the ones in this thesis. For the Green Line case, the appropriate value for $m$ was estimated to be 3.

### 8.1.2 Properties and Effectiveness of Control Strategies in System $F$

We formulated nonlinear integer programming models for the real-time deadheading and expressing problems, nonlinear programming models for the holding problem, and a nonlinear mixed integer programming model for the combined deadheading and holding problem in System $F$. All the programs, or their real relaxations when integer decision variables are involved, have convex cost functions under unrestrictive conditions. This, together with the separability of System $F$, makes all the problems mathematically tractable. Chapters 4, 5, 6 and 7 study these problems in detail, and the convexity of the mathematical programs or their real relaxations proved. The two station skipping strategies, deadheading and expressing, give very similar results. The most effective starting station for expressing is the first station in the control direction. For both deadheading and expressing, the number of skipped stations is bounded from above by $(a+\Delta)/(2\Delta)$, where $a = h_i^0 - h_{i+1}^0$ (the difference between preceding and following no-control headways of $i$), and $\Delta$ is the sum of acceleration and deceleration time plus dwell time. Thus, the magnitude and effectiveness of the station skipping control increase with $a$ for each controlled vehicle.

The holding problem is a convex quadratic problem. In Chapter 6, we developed efficient algorithms for its non-degenerate and degenerate versions. The optimality of these algorithms were proved in Proposition 6.1. The effectiveness of holding mainly depends on the tightness of the terminal schedule constraint. When this constraint is relaxed,
holding is more effective than station skipping strategies, while it can be less effective otherwise.

The algorithm developed for the combined deadheading and holding problem makes use of the properties of deadheading and holding. The computational tests show that the combined control not only is systematically much more effective than any single type of control strategy, but also much less severe in terms of control actions. The latter is reflected in the result that both the number of skipped stations and the amount of holding delays were greatly reduced over the analysis period.

8.1.3 Properties and Effectiveness of Control Strategies in System G

We again formulated mathematical programming models for all the control problems in System G. Though the objective functions and constraints are similar in form to those of System F, the variable definitions in these models are much more complicated than in the System F models, and the problems are mathematically intractable. However, the properties of their counterparts in System F provided useful insights into how to attack them. For example, the real relaxations of the deadheading and expressing problems in System F are convex. Now, essentially the only difference in the objective functions between the problems in the two types of systems is that the vehicle headways and passenger arrival rates in System G vary across stations. While the cost functions can be easily shown as non-convex in System G, would they be quasi-convex? From the analysis in Chapter 4 on the cost function of 1-DH subproblems in RTDPG, this seems very likely. We conducted fairly large computational experiments, and all the computational results show that the net beneficial portion of the cost functions is always quasi-convex for both the 1-DH and the 1-EXP subproblems in System G. For the impact set with size $m>1$, things are more complicated, but over 95% of individual vehicle sets show similar quasi-convexity in their objective functions. This property leads to an efficient search algorithm, which is particularly useful for the expressing and the combined control problems, where repeated search is needed to find the optimal skipped segment.
Chapters 4 and 5 again show that deadheading and expressing have very similar
effectiveness, and their effectiveness largely depends on demand pattern. Expressing has
the flexibility to choose the best starting station. When the high demand stations are later
in the control direction, computational results show that the best starting station is most
likely to be the first station and that even when a later station is chosen the difference in
results is often insignificant. Conversely, when a high demand station occurs in the
beginning of the direction, the best starting station is most likely to be that station.
Nonetheless, to start expressing at different stations for different vehicle trips has the
advantage of not always affecting the same passengers in the same skipped segment.

In Chapter 5 we also investigated the impacts from threshold benefit, capacity constraint
and minimization of both waiting and riding times through computational tests. The
results show that using appropriate threshold benefit can reduce by almost half of the
number of control actions while retaining 90% of the control benefits. This is very
attractive. Under the capacity constraint the effectiveness of control in terms of either
passenger waiting or riding time reductions declined.

We especially note that the minimization of passenger waiting times also resulted in
significant passenger riding time reductions in all tests, due to the more even dwell times
and fewer interstation stops after control. When passenger riding time is also included in
the objective function with the same weight as waiting time, some waiting time savings
are traded off against additional riding time savings. Although the increase in riding
time reduction is somewhat larger than the decrease in passenger waiting time reduction,
the improvement in the total (waiting time + riding time) is not very significant. If the
riding time weight were smaller, the improvement in total cost would be even smaller.

Comparison between the two station skipping strategies, expressing and deadheading,
shows that they have different advantages and disadvantages. The main advantages of
expressing are: first it can have non-overlapping optimal skipped segments even when
adjacent vehicles are expressed; second, when the optimal express initiation station is a
high demand station, in addition to waiting time savings it can also result in much greater
passenger riding time savings than deadheading. However, the modeling complexity and computational burden for expressing is substantially higher than in the deadheading case, mainly due to complicated passenger activities at the express initiation station, and the need to determine both optimal initiation and ending express stations. The computation time required to search for an optimal expressing segment is, in general, $K$ times higher than for deadheading. In addition, to track the alightings of leftover passengers from each express initiation station is cumbersome.

Deadheading is easier to implement and a deadhead vehicle is subject to control earlier than with expressing. This makes deadheading more effective in a high headway variation situation in terms of passenger waiting time reduction. A major disadvantage of deadheading is that vehicles always skip a common set of stations, which negatively affect the same people. Also, the riding time savings realized are lower than those from expressing.

Based on the property of dispatching headway effects, in Chapter 6 we extended the algorithm for holding in System $F$ to System $G$. Computational experiments show that the algorithm is both effective and efficient. The most interesting property of holding is that it is insensitive to demand patterns. Unlike the station skipping control policies which are very demand sensitive, optimal holding policies just reconcile dispatching headways of vehicles considered with the midrange headway of the preceding vehicle. As in System $F$, holding is more effective without the terminal schedule constraint. The limitations of holding occur in situations where the total headway in the vehicle set is large, causing a longer holding time than later vehicles'. An interesting result is that, while holding was expected to have the negative effect of delaying trip time, when it is applied continuously the overall delay at terminal stations is actually decreased. Furthermore, passenger riding time savings resulting from holding are larger than those from deadheading.

Chapter 6 also explored the choice of holding station location. Computational experiments show that the best choice is again the first station in a direction. This is
because the headway pattern down the line largely depends on the dispatching headway pattern, and early control benefits more stations.

In Chapter 7, a heuristic for Combined control in system $G$ was developed based on algorithms for each single type of control problems. Computational results show that the effectiveness of combined control is higher than any single type of control, although the marginal benefits are somewhat limited. Holding seems the most effective single type of control strategy. When holding is used continuously, headway variance decreases, and the chance for beneficial station skipping control also decreases. This indicates that combined control is most effective where holding is tightly restricted by schedule constraints. Computational results again show that combined control leads to less severe control actions. It also increased control feasibility, and as a consequence more vehicle trips were controlled than in any single type of control. This tendency is welcome because it brings, at the same time, more benefits and less frustration to passengers. One important conclusion from the results is that for maximum effectiveness frequent control actions are needed. In the tests using Green Line data sets, over three fourths of vehicle trips were subject to some form of control.

8.1.4 Model Sensitivity to Randomness

Because in reality various degrees of stochastic conditions exist for all transit systems, it is important to test the deterministic models developed in this thesis under randomness. In Chapter 7 we designed tests to evaluate the models when various degrees of randomness exist in the system. The computational results for the combined control model show that, while system performance without control was worse with random disturbances, the change in effectiveness of control was not large when systematic biases were small. The system properties proved in Chapter 3 are insensitive to randomness. The impact from randomness in vehicle interstation travel times is larger than randomness in passenger arrival rates. The effectiveness of control is more sensitive to systematic biases in vehicle travel times. The model performance is also more sensitive to forecast information accuracy on vehicle travel times than on passenger demand. In
conclusion, the control models developed in this thesis will probably work well in most U.S. transit systems, but in an operating environment where vehicle speed variance is high, and forecast information for vehicle interstation travel times is not accurate, the models will have significantly worse performance.

8.1.5 Extensions of the Models

This thesis has investigated control problems under deterministic conditions (except in the sensitivity tests). As a result of the analysis we have concluded that station $k_0$ is most often the best control station, and the algorithms developed in the thesis often assume the control decision is made at $k_0$. This is, of course, a simplification of the stochastic reality. However, the algorithm and analysis presented here can be easily extended to more complicated situations. We briefly list a few possibilities below.

1. If passenger demand or vehicle trajectories change substantially at some station due to an unpredicted disturbance or event, the control decision can be remade by applying the algorithm and guidelines to the subset of the route starting from that station.

2. If there is more than one peak demand station in a direction, and these stations are far apart from each other, the direction can be divided into more than one section and the algorithms and guidelines can then be applied to each section.

3. If necessary, a vehicle can be expressed, deadheaded, or held more than once in a direction by repeatedly using the algorithms and guidelines.

8.2 Suggestions for Future Research

8.2.1 Extensions of Research

Real-time control in transit operations is a rich research area, and this dissertation is in no way a complete coverage of this topic. Possible extensions to the work contained here include:
**Stochastic Version of the Control models.** Although the randomness tests on the CCPG model show that the deterministic models are quite robust with coefficients of variation up to 0.3 in both vehicle travel times and passenger demand with the same mean vehicle speed, the results can probably be improved by stochastic models especially when interstation travel time variance is high across vehicies and forecast information is not accurate. Now with the deterministic models developed and in place, the development of stochastic models based on them should be easier.

**Different Cost Functions.** In this dissertation we focused on passenger waiting time cost. As the cost functions of weighted sum of waiting and riding times, or vehicle headway variances are also quite commonly used in transit research, it would be very useful to compare control policies, their properties, and the pros and cons of using different cost functions. We should note that there is a particular lack of appropriate cost functions for dealing with unusually long headways, such as at stations skipped by multiple consecutive vehicles. In such situation passenger waiting costs may grow exponentially.

**Different Demand Function.** This research has assumed passengers arrive at stations randomly at a constant rate, and also, passenger alightings at a station are a proportion of the vehicle arrival load. These demand characteristics made their contribution to the system properties we showed in Chapter 3, and also the cost function properties we showed in chapters 4-7. However, it would be more realistic to treat passenger arrivals at a transfer station as bulk arrivals plus random arrivals. There may also be alternative passenger alighting functions. The impact of different passenger demand functions on control policies needs to be investigated.

**Other Types of Control Strategies** There are many more control strategies not addressed in this research that deserve investigation. Three other control strategies are of particular interest to us and, we believe, many transit agencies: short-turning, train splitting, and adding vehicles. The last strategy requires consideration of how many extra vehicles and
operators need to stand-by daily. Research on this strategy may not only help to decrease passengers cost but also help to cut down operating cost for the transit agency.

8.2.2 Related Areas

As we have seen from the sensitivity test in Chapter 7, forecast information accuracy has a significant impact on the effectiveness of control policies. Since AVI/AVL detectors may not, in general, exist at every station, a better forecasting technology to handle the missing information at stations which do not have a detector may be data filtering techniques. Further research needs to be conducted in this direction.

Another related research issue is how do we get the best out of the real-time information provided by AVI/AVL and APC systems. While this dissertation does not address any data processing issues, without good data provision, no control models would work properly.

8.3 Concluding Remarks

This dissertation represents one of the first attempts to develop models and algorithms for the real-time control strategies deadheading, expressing and holding, both singly and in combination. Properties and effectiveness of the strategies were analyzed based on both mathematical models and computational results. Although the models are deterministic, sensitivity testing under randomness has shown that the models work well with coefficients of variation up to 0.3 in both vehicle travel times and passenger arrival rates. From the computational results, it is clear that when both station skipping strategies and the holding strategy are used in combination, control is more effective with many fewer "side effects" than any single type of control strategy used alone. Passenger waiting time reduction from the combined control over a three-hour peak period can be as high as 38%. Thus it is recommended that transit agencies use such combined control strategies. If for some reason a single type of control strategy is desired, holding is the easiest to use and is often more effective and less frustrating to passengers. Holding is also largely
demand pattern insensitive, and hence the holding algorithm developed in this thesis has quite general applicability. The two station skipping strategies, deadheading and expressing, can be very effective in situations holding is not, but because they require more restrictive headway conditions and are sensitive to demand pattern, their effective use requires better understanding of the interaction between the transit system, the demand pattern, and the control policies. This dissertation provides detailed analysis to gain such understandings. For transit agencies with computerized control systems, the algorithms developed in this dissertation are efficient in real-time. For a reasonable size vehicle set (e.g., 3-16), the computation time required for any control problem is no more than a few seconds.

The research approach taken in this thesis makes it possible for transit services with either computerized or manual control systems to make use of the results. The significance of the analysis can also go beyond the transit control problems themselves - results from this thesis may apply to other problems with similar settings and properties. We believe, this research has made a significant contribution to both the theory and practice of real-time control in transit operations.
APPENDIX A: SUMMARY OF NOTATION

Parameters and Constants

\( M \): the total number of vehicle trips in the time period considered.
\( M_s \): the total number of vehicles active on the system.
\( N \): the total number of stations in the transit network. An even number.
\( h_0 \): the required minimal safe headway between two vehicles (minutes).
\( r_k \): passenger arrival rate at station \( k \) (passengers/minute).
\( q_k \): normal passenger alighting proportion (%) at station \( k \).
\( R_{i,k} \): minimum running time of vehicle \( i \) from station \( k \) to station \( k \) (minutes).
\( \delta_i \): the acceleration (deceleration) time of vehicle \( i \) at a station (minutes).
\( H \): the scheduled headway (minutes).
\( c_0 \): the constant dwell time.
\( u_c \): weight for terminal condition in the control direction \( c \). \( 0 \leq u_c \leq 1 \)
\( t_{i,c} \): scheduled departure time of vehicle \( i \) at \( k_c \) minus minimal layover time

Sets

\( \mathbb{Z}_+ \): the set of non-negative integers.
\( \mathbb{Z}_{++} \): the set of positive integers.
\( \mathbb{R}_+ \): the set of non-negative reals.
\( \mathbb{R}_{++} \): the set of positive reals.
\( K \): an ordered set of all stations in the network, indexed from 1 to \( N \). \( |K| = N \).
\( K_c \): the ordered set of all stations in the control direction, \( \{k_0, k_0+1, \ldots, k_i\} \). \( |K_c| = N/2 \).
\( K_c' \): \( K_c \) - \{\( k_i \)\}.
\( \tilde{K}_c \): the ordered set of all stations in the direction opposite to the control direction.
\( \Pi \): the set of all vehicle trips in the time period considered \( |\Pi| = M \).
\( \Pi_m \): an impact set of vehicles (vehicle trips) \( i, i+1, \ldots, i+m, 1 < m \leq M \), in which the first vehicle, \( i \), is to be controlled and the impact of the control is to be evaluated over this set of vehicles.

Indices

\( i, j \): index for vehicles
\( k \): index for stations
\( k_0 \): the first and dispatching station in the control direction. \( k_0 = 1 \) or \( N/2 + 1 \).
\( k_i \): the ending terminal station in the control direction. \( k_i = N/2 \) or \( N \).
\( k_c \): the next dispatching terminal

Variables

\( y_{i,k} \): the decision variable: in deadheading and expressing problems. \( y_{i,k} = 0 \) if vehicle \( i \) skips station \( k \), \( y_{i,k} = 1 \) if vehicle \( i \) stops at station \( k \).
\( h_{i,k} \): the departure headway (minutes) at station \( k \) between departure times of vehicle \( i \) and the previous vehicle \( i-1 \).
\( d_{i,k} \): departure time of vehicle \( i \) from station \( k \).
$P_{i,k}$: the number of "leftover" passengers, i.e., the passengers who waited at station $k$ but could not get on vehicle $i$ when $i$ skips $k$, and the passengers who are dumped by $i$ at $k$ when $k$ is a starting station of $i$'s deadheading.

$L_{i,k}$: the departure load of vehicle $i$ from station $k$.

$B_{i,k}$: the number of passenger boardings from station $k$ onto vehicle $i$.

$A_{i,k}$: the number of passenger alightings from vehicle $i$ at station $k$.

$s_{i,k}$: the dwell time (in minutes) of vehicle $i$ at station $k$. 
APPENDIX B: DEMAND PROFILE

Weekday Morning Peak Demand Profile of
MBTA Green Line B Line
(Passengers/minute)

#Passengers

0  10  20  30  40

1  4  6  8  10  12  14  16  18  20  22  24  26  28  30  32

Direction 1

out mid out mid in in out in

Direction 2

Station

AVI detector

in: at entrance of the station, out: at the exit, mid: in the middle

Passenger Boardings

Passenger Alightings

Total
APPENDIX C: DISPATCHING HEADWAY PROFILES

Green Line Data Sets: Vehicle Headways at \( k_0 \)

Monday, Direction 1 (m1)

Tuesday, Direction 1 (tu1)
Thursday, Direction 2 (th2)

Friday, Direction 2 (f2)
REFERENCES


