USE OF THESES

This copy is supplied for purposes of private study and research only. Passages from the thesis may not be copied or closely paraphrased without the written consent of the author.
Addendum

Chapter 2:

In this Chapter, we have assumed that the control weighting matrix \( R \) is strictly positive definite. From the work of Chen, Li and Zhou [58] as well as the results in Lim and Zhou [32] (as discussed in Chapter 4), it is clear that this restriction may not be necessary when the diffusion term in the state equation depends on the control. We shall briefly comment on this issue in relation to the results in Chapter 2. Consider the system

\[
\begin{align*}
\frac{dx(t)}{dt} &= (A(t)x(t) + B(t)u(t) + g(t)) \ dt + (h(t) + C(t)x(t) + D(t)u(t)) \ dW(t), \\
x(s) &= y
\end{align*}
\]

(1)

It is easily seen that (1) defines an affine subspace of \( \mathcal{X} \times \mathcal{U} \). Using the techniques described in Chapter 2.2.3, together with the Ito differential rule in conjunction with the 'completion of square' technique (see [58]), it can be shown that the problem,

\[
E \left[ \frac{1}{2} \int_0^T \left[ x'(t)Q(t)x(t) + u(t)'R(t)u(t) \right] \ dt + \frac{1}{2}x'(T)Hx(T) \right] \rightarrow \min
\]

subject to:

\[
E \left[ \int_0^T \left[ a'_i(t)x(t) + b'_i(t)u(t) \right] \ dt + h'_ix(T) \right] \leq c_i
\]

is equivalent to

\[
\max \left\{ \frac{1}{2} \xi'P(0)\xi + d'(\lambda,0)\xi + \frac{1}{2}p(\lambda,0) - \lambda'c \right\}
\]

subject to:

\[
\begin{cases}
\dot{P} + PA + A'P + C'PC - (PB + C'PD)R^{-1}(B'P + D'PC) + Q = 0 \\
P(T) = H \\
K = R + D'PD > 0
\end{cases}
\]

(5)

\[
\begin{cases}
\dot{d} + [A - BR^{-1}B']d + a(\lambda) - PBR^{-1}b(\lambda) = 0 \\
d(T,\lambda) = h(\lambda)
\end{cases}
\]

(6)

\[
\begin{cases}
\dot{p} = [B'd + b(\lambda)]'R^{-1}[B'd + b(\lambda)] \\
p(\lambda,T) = 0
\end{cases}
\]

(7)

with the optimal control of (1)-(3) given by

\[
u^*(t) = -(R + D'PD)^{-1} \left[ (B'P + D'PC)x + B'd(\lambda^*) + b(\lambda^*) \right]
\]

(8)

where \( \lambda^* \) is the optimal solution of (4)-(7).
There are several important issues associated with (4)-(7). First, it is unclear whether the Riccati equation (5) is solvable. However, if it is solvable, then the problem (1)-(3) is well posed and (8) is the optimal control.

We have assumed that $R$ is positive definite. This assumption guarantees that the cost functional is strictly convex and positive. Convexity is required to invoke the Lagrange Duality Theorem; strict positivity together with strict convexity guarantee the existence of a unique optimal control. (In fact, strict positivity is used to prove boundedness of the feasible region of a related optimization problem which together with the convexity property, implies existence of a unique optimal control). Under this assumption on $R$, the solvability of (5) implies that (8) is the unique optimal control for (2)-(3). In the case $C(t) = 0$, (5) becomes

$$
\begin{align*}
\dot{P} + PA + A'P + C'PC - (PB + C'PD)R^{-1}(B'P + D'PC) + Q &= 0 \\
P(T) &= H \\
K &= R + D'PD > 0
\end{align*}
$$

(9)

In [58], necessary and sufficient conditions for the solvability of (9) are derived. (This is outlined in Chapter 4.2). In particular, it is shown that $R$ may have negative eigenvalues but the Riccati equation (9) still be solvable. Moreover, it is shown in [32] and in Proposition 4.1 (in Chapter 4.3.1) that these conditions for solvability of (9) guarantee strict convexity of the cost functional (2). Furthermore, these conditions guarantee existence of a unique optimal control since they imply that the feasible region of a related optimization problem is bounded (refer to Lemma 1 below for a proof of this).

Chapter 4:

When the control weighting matrix $R$ is positive definite, the resulting cost functional is strictly positive and strictly convex. Strict convexity is required for the Lagrange Duality Theorem to be applicable. Positivity guarantees boundedness of the feasible set (defined by the constraints) which together with strict convexity, guarantees the existence of a unique optimal control. In Proposition 4.1, a condition that guarantees strict convexity of the cost (and constraint) functionals is derived. However, a proof of the boundedness of the feasible set was not given. The proof of this result is as follows.

**Lemma 1** Suppose that for every $i \in \{1, \ldots, m\}$, there exists $K_i \in K$ such that $R_i + D'\Psi_i(K_i)D \geq K_i$. Then the set of all $(x(\cdot), u(\cdot))$ that satisfy the constraints (4.7) is a bounded set in $L^2_x(0, T; \mathbb{R}^n) \times L^2_x(0, T; \mathbb{R}^k)$.

**Proof:** Due to the decomposition $X = z(\cdot) + X_0$, we may assume without loss of generality that $g = 0$, $h = 0$ and $y = 0$ in the state equation (1), namely, that $(x(\cdot), u(\cdot)) \in X_0$. By the assumption (4.6), there is a constant $c > 0$ such that every admissible pair $(x(\cdot), u(\cdot))$ must satisfy $E|x(t)|^2 \leq cE \int_0^T |u(t)|^2 dt$ for all $t \in [0, T]$. Therefore to prove the lemma it suffices to show that the set of all $u(\cdot)$ satisfying the constraints in (4.7) is bounded in $L^2_x(0, T; \mathbb{R}^k)$. 

2
To this end, we first show that for any \( i = 1, \cdots, m \) and any \((x(\cdot), u(\cdot))\) that satisfies the constraints in (4.7), it holds

\[
E\left\{ \frac{1}{2} \int_0^t \left( x'(t) Q_i(t) x(t) + u'(t) R_i(t) u(t) \right) dt + \frac{1}{2} x'(t) P_i(t) x(t) \right\} \leq c_i, \quad \forall t \in [0, T]. \tag{10}
\]

Indeed, fix an \( i \in \{1, \cdots, m\} \) and \( t \in [0, T] \). Under the condition of the lemma, the stochastic Riccati equation (9) admits a unique solution \( P_i \in C(0, T; S^+_n) \). Put \( K_i = R_i + D' P_i D > 0 \). Applying Itô’s formula to \( x'(s) P_i(s) x(s) \), then integrating from \( t \) to \( T \) and taking expectation, we have

\[
E\left\{ \frac{1}{2} \int_t^T \left( x'(s) P_i(s) x(s) + 2 x'(s) P_i(s) B(s) u(s) + u'(s) D' P_i(s) D(s) u(s) \right) ds \right\}.
\]

Hence,

\[
E\left\{ \frac{1}{2} \int_t^T \left( x'(s) Q_i(s) x(s) + u'(s) R_i(s) u(s) \right) ds + \frac{1}{2} x'(t) P_i(t) x(t) \right\} dt \geq 0.
\]

This leads to

\[
E\left\{ \frac{1}{2} \int_t^T \left( x'(s) Q_i(s) x(s) + u'(s) R_i(s) u(s) \right) ds + \frac{1}{2} x'(t) P_i(t) x(t) \right\} \geq c_i,
\]

which proves (10). Next, by applying Itô’s formula and completion of squares again, we obtain

\[
c_i \geq \frac{\delta}{2} E\left\{ \frac{1}{2} \int_0^t \left( x'(s) Q_i(s) x(s) + u'(s) R_i(s) u(s) \right) ds + \frac{1}{2} x'(t) P_i(t) x(t) \right\} \geq C \int_0^t |u(s)|^2 ds + \frac{C^2}{\delta} E\int_0^t |x(s)|^2 ds.
\]

In the above we have used the well-known inequality \( ab \leq \frac{b^2}{2\delta} a^2 + \frac{\delta}{2} b^2 \). Then (12) gives

\[
c_i \geq \frac{\delta}{2} E\int_0^t |u(s)|^2 ds - \frac{C^2}{\delta} E\int_0^t |x(s)|^2 ds \geq \frac{\delta}{2} E\int_0^t |u(s)|^2 ds - C' \int_0^t |u(r)|^2 dr ds, \quad \forall t \in [0, T].
\]

By Grownall’s inequality, we conclude that

\[
E\int_0^t |u(s)|^2 ds \leq \frac{2c_i}{\delta} e^{\frac{2C'}{\delta} t}, \quad \forall t \in [0, T].
\]

This completes the proof. \( \blacksquare \)
Optimal Control of Systems with Constraints

Andrew E.B. Lim
Bachelor of Science

August 1997

A thesis submitted for the degree of Doctor of Philosophy
of the Australian National University

Department of Systems Engineering
Research School of Information Sciences and Engineering
The Australian National University
Statement of Originality

These doctoral studies were conducted under the supervision of Professor John B. Moore, with Dr Peter Bartlett and Dr Subhrakanti Dey as advisors.

The work submitted in this thesis is the result of original research carried out by myself, in collaboration with others, while enrolled as a PhD student in the Department of Systems Engineering, Australian National University. It has not been submitted for any other degree or award in any other university or educational institution.

Andrew Lim
August 1997.


[Text not legible due to image quality]
Acknowledgements

I have completed my PhD under the supervision of Professor John Moore. I wish to thank him for this opportunity to work with him. His insight and vision are inspirational and the personal interest he has for his students is always an encouragement. I would also like to thank Dr Subhrakanti Dey and Dr Peter Bartlett to their guidance as members of my supervisory panel. Furthermore, I offer my sincere thanks to Professor K.L. Teo at Curtin University of Technology for his continual support, advice and friendship over the years.

I shall always look back at my time in Canberra with joy. It has been a time of learning, contemplation, frivolity and maturing; one that I shall never forget. I would like to thank the staff and and students of the Department of Systems of Engineering at the Australian National University. In particular, to Jason Ford, Leonardo Kammer, Llewellyn Mason, Peter Dower, Allan Connolly, Gordon Sutton, David Jung, Gordon Cheng, Jochen Heinze, Natasha Linard, Tanya Conroy, Sang Heon Lee. Furthermore, I would especially like to thank former students Dr Danchi Jiang and Dr Subhrakanti Dey for their friendship. It's been a pleasure working with and learning from them. Special thanks also to Ms Marita Rendina, Ms Vicky Bass Becking and Ms Catrina Watson for their cheerful, professional (and patient!) secretarial support, and to Mr James Ashton for his computer support. Finally, I wish to acknowledge the funding of the activities of the Cooperative Research Centre for Robust and Adaptive Systems by the Commonwealth Government under the Cooperative Research Centre Programme.

I've been fortunate to spend time at various universities during my PhD studies. In particular, I would like to thank Professor Robert Elliott from the University of Alberta, and Mrs Anne Elliott for their generous hospitality during my stay in Canada. I would also like to thank Professor K.L. Teo for his continual support and encouragement over the years as well as for providing opportunities to undertake joint work with him during my doctoral studies, both in Canberra as well as in
Perth. Thanks also to Dr T.T. Tay for his hospitality during my stay in Singapore. Finally, I would like to thank the Department of Systems Engineering and Engineering Management at the Chinese University of Hong Kong. In particular, I would like to thank Professor X.Q. Cai for providing funding for the trip and Professor X.Y. Zhou for many interesting discussions and ideas. The time I spent in Hong Kong was an exciting and rewarding experience.

I have been fortunate to cross paths with a large number of wonderful people during my time in Canberra. Thanks to Simon Wallace-Pannell, for his friendship and support, not to mention the discussions relating to matters of all kinds of wonder, majesty and absurdity! Thanks also to Mei Ching Wong, for her encouragement, patience and friendship. Thanks to Siew Cheun Cho for his friendship, and his interesting sense of humour that has been the source of many light hearted moments over the years, and to Kathryn Carr, for her friendship, chocolate cake and camomile tea! Susannah Hill, for her friendship as well as in her capacity as the owner of a Volvo motor vehicle, must also not be forgotten! Many others have contributed in various ways to make this experience an enriching one: Janice Koh, Ai Lin Chay, Hatta, Alan Siu, members of the Crossroads Christian Church and the Overseas Christian Fellowship. Thanks also to Irene, Sim, Chenyce and Michelle for their love and support.

A special mention must go to my parents and brother. Thank you for your love, and your support, through fun times as well as times of struggle. You are never far from my thoughts. To you, this thesis is dedicated.

Finally, I wish to acknowledge Jesus Christ, my lord and savior.
Abstract

In this thesis, we present new developments resulting from our work on constrained LQG control. Our work can be divided into two broad areas: Generalizations of well known results associated with unconstrained LQG control such as the Separation Theorem, and development of new computational algorithms for solving these problems. A summary of the topics we are presenting is as follows:

Linearly constrained LQG control

In this chapter, we study the LQG control problem with finitely many and infinitely many linear inequality constraints. We derive the optimal control for these problems, and prove the Separation Theorem. We show (using duality theory) that when there are finitely many constraints, the optimal control can be calculated by solving a finite dimensional optimization problem. When there are infinitely many constraints, the optimal control is determined by solving an infinite programming problem.

LQG control with IQ constraints

We consider the LQG control problem with finitely many integral quadratic constraints. Using duality theory, we derive the optimal control, and show that it can be calculated by solving a finite dimensional optimization problem. Relevant gradient formulae pertaining to this finite dimensional problem are derived. We prove that the Separation Theorem does not hold. Rather, a result we call a Quasi-separation Theorem is proven.

Indefinite LQG control with IQ constraints

We extend recently discovered results for full observation LQG control with an indefinite control weight to the constrained case. We derive conditions under which the optimal control can be explicitly derived, and calculated by solving a finite dimensional optimization problem. We also derive relevant gradient formulae so that
algorithms for nonlinear optimization problems can be used to solve this problem.

Infinite quadratic programming

In this chapter, we derive an alternative method for solving the linearly constrained LQG problem. Drawing inspiration from the field of interior point methods, we derive a path following interior point method for linearly constrained quadratic programming to infinite dimensions. In this way, an interior point method for linearly constrained LQG problems is derived. We also prove global convergence of this algorithm.

Infinite linear programming

We generalize the potential reduction interior point method for finite dimensional linear programming to the infinite linear programming case. We show how this algorithm can be used to solve linear optimal control problems with continuous state constraints, as well as continuous linear programming problems. In this way, we derive new methods for solving these problems. We also examine some convergence issues.
List of publications

Journal papers


• A.E.B. Lim and J.B. Moore. A quasi-Separation theorem for LQG optimal control with integral quadratic constraints. (To appear in *Systems & Control Letters*).

• A.E.B. Lim and J.B. Moore. A path following algorithm for infinite quadratic programming on a Hilbert space. (To appear in *Discrete and Continuous Dynamical Systems*).

• A.E.B. Lim, Y.Q. Liu, K.L. Teo and J.B. Moore. Linear-quadratic optimal control with integral quadratic constraints. (Submitted to *Optimal Control, Applications and Methods*).


Conference proceedings


Contents

1 Introduction

2 Linearly constrained LQG control
   2.1 Mathematical preliminary ............................................. 12
   2.2 LQG control with finitely many linear constraints .............. 15
      2.2.1 Mathematical Preliminary ....................................... 15
      2.2.2 Deterministic case ............................................. 19
      2.2.3 Full observation case ......................................... 26
      2.2.4 Partial observation case ..................................... 30
   2.3 LQG control with infinitely many linear constraints .......... 34
      2.3.1 Mathematical preliminary ..................................... 35
      2.3.2 Deterministic case ........................................... 39
      2.3.3 Full observation case ......................................... 45
      2.3.4 Partial observation case .................................... 48
   2.4 Conclusion .............................................................. 51

3 LQG control with IQ constraints
   3.1 Deterministic Case ................................................... 54
   3.2 Full observation stochastic case .................................... 59
   3.3 Partial observation stochastic case ................................ 62
   3.4 Optimal parameter selection problems ............................ 65
   3.5 Conclusion ............................................................... 71

4 Indefinite constrained LQG control
   4.1 Problem statement ..................................................... 74
   4.2 Preliminary results ................................................... 76
   4.3 Main results ............................................................. 78
      4.3.1 Conditions for convexity ...................................... 79
<table>
<thead>
<tr>
<th>4.3.2 Optimal control</th>
<th>82</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.4 Conclusion</td>
<td>93</td>
</tr>
</tbody>
</table>

5 **Infinite quadratic programming** 95

<table>
<thead>
<tr>
<th>5.1 Central trajectory analysis</th>
<th>96</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.2 Example</td>
<td>113</td>
</tr>
<tr>
<td>5.3 Conclusion</td>
<td>118</td>
</tr>
</tbody>
</table>

6 **Infinite linear programming** 119

<table>
<thead>
<tr>
<th>6.1 Infinite LP and the potential function</th>
<th>120</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2 Algorithm</td>
<td>123</td>
</tr>
<tr>
<td>6.3 Convergence results</td>
<td>131</td>
</tr>
<tr>
<td>6.4 Applications and examples</td>
<td>139</td>
</tr>
<tr>
<td>6.5 Conclusions</td>
<td>143</td>
</tr>
<tr>
<td>6.6 Appendix</td>
<td>144</td>
</tr>
<tr>
<td>6.6.1 Relegated proofs from Section 6.3</td>
<td>144</td>
</tr>
<tr>
<td>6.6.2 Generalizations</td>
<td>152</td>
</tr>
</tbody>
</table>

7 **Concluding remarks and future directions** 157

8 **Appendix** 161

<table>
<thead>
<tr>
<th>8.1 Definitions and theorems from analysis</th>
<th>161</th>
</tr>
</thead>
<tbody>
<tr>
<td>8.2 On the baking of almond biscuits</td>
<td>161</td>
</tr>
<tr>
<td>8.2.1 Production procedure</td>
<td>162</td>
</tr>
<tr>
<td>8.2.2 Concluding remarks</td>
<td>162</td>
</tr>
</tbody>
</table>
Chapter 1

Introduction

Optimization is the underlying issue in many applications and problems in systems theory. Commonly, such problems involve making tradeoffs between several competing variables, or finding a solution for which constraints on certain variables are satisfied; for example, minimal productivity constraints in a manufacturing system, position constraints on a robot arm, or energy constraints in a power system. In general, these constraints must be accounted for if a useful solution is to be found by an algorithm.

However, despite the large range of efficient optimization algorithms which are available, existing algorithms must be constantly refined and new algorithms developed in order to keep in step with the high demands of many modern applications for which the earlier algorithms are inadequate. On the other hand, ever increasing computer power makes implementation of complex, sophisticated algorithms more viable. The challenge is clear: Design new algorithms which meet the requirements of the application by exploiting the available computer power to carry out the complex tasks.

A major focus in optimal control theory has been in the study of unconstrained problems. Moreover, the (unconstrained) linear quadratic Gaussian (LQG) optimal control methodology is perhaps the most popular and widely used modern control methodology. There are several reasons for this. First, the optimal control for a well posed LQG problem is straightforward to calculate, especially with the aid of modern computer technology and numerical techniques. Second, the optimal control is elegant and easy to implement. Third, irrespective of the particular quadratic performance
CHAPTER 1. INTRODUCTION

index chosen, there are attractive guaranteed robustness properties to input uncertainties. Forth, classical design insights can be exploited to advantage in selecting the performance index, perhaps frequency shaped. Fifth, state controllability and state observability issues can be considered separately.

In deterministic problems, the optimal linear quadratic (LQ) control is a linear state feedback. In the stochastic partially observed case, the optimal control is determined by solving a minimum error variance (Kalman) filtering problem and an optimal LQ deterministic control problem. In particular, a result known as the Separation Theorem states that the problems of filtering and control are independent and the solutions of each, when put together using state estimates in lieu of states, achieves optimality in the stochastic case. This leads to an easily derived and implemented optimal control law. In the case when the cost functional is nonlinear and the problem unconstrained, LQG control still can play an important role. Sequential quadratic programming (SQP) is one of the most efficient techniques for solving general unconstrained nonlinear optimal control problems. In SQP, a sequence of optimal control problems are obtained when some form of online linearization of the signal model together with a quadraticization of the cost functional are carried out. The optimal control of the nonlinear problem is determined by solving this sequence of approximations. The ease with which unconstrained LQG can be solved is paramount in its success in this technique.

One reason for the successful use of unconstrained LQG in many applications is its ability to take competing objectives into consideration. By varying the weights in the cost functional, an optimal control that achieves the objectives by making the necessary tradeoffs between performance objectives and control energy costs can be found. However, this is an indirect method of satisfying hard state or control constraints, and the problem of selecting the appropriate weights in the cost functional may involve quite a degree of trial and error. Furthermore, this indirect approach for solving constrained problems means that traditional LQG control is not easy to use in the context of SQP for more complicated constrained nonlinear optimal control problems. Moreover, attractive properties of the SQP algorithm such as quadratic convergence of the iterates near the optimal solution is unlikely to hold.

As already stated, many potential applications of optimal control theory are constrained problems. For example, in recent study we have considered a problem known as the unit power commitment problem. Given the energy demands of a country which
has many power stations, the problem is to determine a systematic method for choos-
ing which power stations to use (and which ones to deactivate) so as to minimize the
overall cost of running the power stations while meeting the power demands. Apart
from the obvious constraint of 'power demand', other constraints in this problem in-
clude constraints on the minimal 'on time' of a power station (once turned on) must
run before it can be deactivated, the minimal 'off time' before it can be activated
again, and the finite power supplying capacity of each power station. In fact, in one
approach, solving a constrained optimal control problem appears fundamental to tack-
ling this task in an elegant manner, avoiding the curse of dimensionality in the more
direct dynamic programming approaches. In other applications, constraints on the
available energy as well as state constraints are commonly encountered. Determining
an efficient method for solving such control problems would be useful.

However, it is fair to say that as of yet, there is no control methodology for
constrained optimal control problems that has captured the same attention as the
unconstrained LQG technique. In our knowledge, it seems that the numerical pack-
age MISER 3.1 [35] is able to solve the most general class of constrained optimal
control problems. Based on a technique known as control parametrization, the op-
timal control is determined by partitioning the time horizon of the problem into a
finite number of intervals. In this way, the optimal control problem is approximated
by a finite dimensional nonlinear optimization problem, with better approximations
corresponding to finer partitions of the time horizon. Convergence properties of this
algorithm, applied to a large class of problems have been obtained; for example, prob-
lems with linear constraints [47], continuous state inequality constraints [50], more
general classes on nonlinear constraints [16], constrained time-lagged systems [51] and
constrained discrete time systems [49], just to name a few. Apart from a few cases
such as [48], the major focus is on deterministic problems. A summary of these results
together with the associated convergence analysis can be found in [46]. However, it is
important to note that this technique is in essence an open-loop control methodology,
and unless there is some built in feedback control structure, there are no guaranteed
robustness properties to plant model uncertainty.

Thus, the challenge to find an elegant, easily computable feedback control method-
ology for constrained optimal control problems remains. In unconstrained optimal
control problems that are more complicated than for the standard LQG setting, it
is generally not possible to find such an expression for the optimal control. There-
Therefore, it seems reasonable that if one wants to find a class of constrained problems for which an elegant, easily calculatable solution can be found, it would be necessary to restrict the class of problems to constrained linear systems with a convex quadratic cost. In fact, a major focus of our study is on LQG control with integral linear and integral quadratic constraints. Keeping all this in mind, we have chosen to focus on the issues of deriving an optimal control, separation theorems and computational algorithms for solving this problem. Moreover, deriving an efficient algorithm for solving this problem opens the door for developing new techniques for solving more general constrained optimal control problems which can be obtained by generalizing SQP to infinite dimensions.

In both the LQG problem with integral linear and integral quadratic constraints, duality theory for convex optimization problems plays an important role. In fact, the optimal controls for the deterministic case are obtained by solving a finite dimensional optimization problem that 'collapses' to the well known method for solving unconstrained LQG when the problem is unconstrained. In the constrained linear case, we prove that the Separation Theorem holds while in the quadratic constrained case, a result that we call a 'Quasi-separation Theorem' holds.

Related to our work on constrained LQG control is the work of Khargonekar and Rotea [19, 20], and Rotea [43]. Here, they study various multiple objective $H_2$ and $H_{\infty}$ optimal control problems involving deterministic, finite dimensional, linear time invariant systems, and develop synthesis techniques using results from convex optimization theory. The basic idea in their work is as follows: In the state feedback case, they show that a control that achieves a performance that is arbitrarily close to the optimal one can be obtained using static state feedback. In this way, an equivalent finite dimensional optimization problem over a bounded set of matrices is derived and shown to be convex. In the output feedback case, they show how this problem can be reduced to the one studied in the state feedback case and hence, solvable using the techniques developed for that problem. Our work is somewhat different to this in that the emphasis is on stochastic, linear time varying systems rather than deterministic, linear time invariant ones. Indeed, for such problems, the optimal control can not be obtained with a static state feedback. Also, much of our focus is on generalizing the Separation Theorem. Our synthesis procedure is based on solving the related dual problem (as opposed to [19, 20, 43] where the emphasis is on solving the primal problem). It is also appropriate to mention here that convex optimization is also used for
control sythesis in the work of Boyd and Barratt [6]. Here, they reduce controller synthesis problem to a convex optimization problem over the infinite dimensional space of stable transfer functions. By making finite dimensional approximations, (suboptimal) controllers are constructed using finite dimensional optimization algorithms.

More closely related to our work on constrained LQG control are the papers on multiple objective LQG control by Carvalho and Ferreira [8], Li [22, 23], Shtessel [45] and Toivonen and Makila [54]. Also, results on this topic are also published in the book by Boyd, El Ghaoui, Feron and Balakrishnan [7]. As in our study on the LQG problem with integral quadratic constraints, the common topic in these papers is the optimal control of a linear system with several objectives, each described by a quadratic performance index. The papers by Carvalho and Ferreira [8] and Li [22, 23] are concerned with deterministic, multiple objective LQ problems in which the objective is to minimize the maximum cost associated with a finite set of LQ performance indices (a so called min-max cost functional). In Carvalho, the optimal solution is shown to be characterized by a certain infinite dimensional inequality constraint, and a procedure based on convex optimization is used to satisfy this constraint and determine the optimal solution. In the paper by Li [22], a condition characterizing the optimal control is derived, and algorithms designed to satisfy this condition are determined. A generalization of this result to the case when the cost functional is a convex function of several LQG performance indices is presented in [23]. The approaches in [8, 22, 23] differ quite significantly from ours in that we seek to obtain the optimal control by solving the related finite dimensional dual problem. Although our work on min-max LQG control is not included in this thesis, it is similar in spirit to our work on LQG control with integral quadratic constraints, which we have presented here. The interested reader may refer to the paper by Lim, Liu, Teo and Moore [27] for further details. The paper by Toivonen and Makila [54] deals with the discrete time, time invariant, infinite horizon LQG control problem with quadratic constraints. These assumptions reduce the problem to a finite dimensional one. As in our work, they use duality theory to tackle this problem. However, their techniques rely heavily on the finite dimensional nature of the problem and can not be extended to the infinite dimensional case that we consider. A similar comment can be made on the results on the LQG problem with integral quadratic constraints, as appearing in Boyd et al [7]. Here it is shown that the infinite horizon, time invariant LQ optimal control problem with integral quadratic constraints is equivalent to an eigenvalue problem (EVP); that is, an optimization problem with linear cost subject
to a linear matrix inequality (LMI) which can be solved using interior point methods for LMI's. However, his techniques can not be extended to the finite-horizon time varying systems we consider in our work.

Moving from this, we turn our attention to alternative algorithms for solving constrained LQG problems, drawing on recent advances in finite-dimensional optimization theory for our inspiration. In recent years, it has become clear that interior point methods (IPM's) are an efficient tool for solving many classes of constrained optimization problems. Although their recent popularity stems from the work of Karmarkar for linear programming [18], interior point algorithms have spread into the domain of nonlinear programming problems. In particular, generalizations include nonlinear programming problems such as linearly constrained and (convex) quadratically constrained convex quadratic programming, linear matrix inequalities [7], as well as more general convex optimization problems [36].

The primary attraction of IPM's is an attractive theoretical computational complexity which has led (in many cases) to efficient numerical computer packages. In particular, a large class of IPM's (Karmarkar's being the first) have been shown to have a worse case polynomial complexity, meaning that the number of iterations required to obtain the optimal solution is bounded above by a polynomial in the problem size. In the case of linear programming, this is an improvement of the exponential worse case complexity of the Simplex method. In practice moreover, it is becoming accepted that many IPM's are competitive with the Simplex Method for small and medium scale problems, but have superior performance when the number of variables and constraints becomes large.

Furthermore, interior point algorithms possess an additional property that is useful in many applications, this being that suboptimal iterates of the algorithm also satisfy the constraints of the problem being solved. Many optimization algorithms produce iterates that may approach the optimal solution from outside the feasible set; for example, the popular Sequential Quadratic Programming algorithm. The drawback of such an algorithm is that it must be run to completion before a solution satisfying all the problem constraints is determined. Suboptimal iterates of IPM's however, do not have this problem and for this reason, such algorithms are well suited for engineering systems that require 'online' optimization. They can be stopped whenever a feasible solution is required and the current suboptimal iterate used with no fear of violating the constraints.
Recent papers by Faybusovich and Moore [11, 12] have seen further generalizations of IPM's which has resulted in new algorithms for solving constrained optimal control problems. By generalizing a certain path following algorithm to the case where the variable belongs to an infinite dimensional Hilbert space, interior point algorithms for solving LQ control problems with finitely many linear constraints [11] and finitely many integral quadratic constraints [12] are derived and studied. Interestingly, despite the set of variables being an infinite dimensional Hilbert space, the algorithms run in polynomial time!

We generalize this path following method further by considering infinite quadratic optimization problems; that is, quadratic optimization problems with infinitely many linear constraints, where the variable may belong to an infinite dimensional Hilbert space. As a special case, this problem contains the LQ control problem with infinitely many linear integral constraints. Although a complexity bound is not proven, we can take heart in the fact that IPM's are known to be highly efficient for problems with many constraints. For this reason, it is not unreasonable to hope that this algorithm will perform well in practice.

Infinite linear programming refers to a class of linear optimization problems where the variable belongs to an infinite dimensional space, and there are infinitely many linear constraints. Special classes of these problems include linear optimal control problems with continuous linear state constraints as well as a class of problems called continuous linear programming, which were introduced in 1953 by Bellman to model problems in production planning, which he referred to as bottleneck processes [4]. We generalize an IPM called the potential reduction method, that was introduced by Ye for linear programming [56], to infinite dimensions as an alternative technique for solving these problems.

Somewhat related to this generalization of Ye's algorithm are the papers by Ferris and Philpott [13, 14], Powell [40] and Todd [53] on semi-infinite programming (that is, when the variable is finite dimensional, but the number of constraints is infinite). Of closest interest is the paper by Todd [53] where a number of IPM's are studied to determine which finite dimensional algorithms converge to a sensible limiting algorithm when the number of constraints goes to infinity. It does not deal with the issue of convergence of the iterates produced by the limiting algorithm. He shows that the potential reduction method of Ye [56] has a sensible generalization to the semi-infinite programming case. In the papers by Ferris and Philpott [13, 14] and
Powell [40] the affine scaling algorithm, and Karmarkar's algorithm respectively are generalized to the semi-infinite setting. In particular, Powell shows that convergence to a non-optimal point may occur if the semi-infinite generalization of Karmarkar's algorithm is used to solve the semi-infinite linear programming problem.

Our focus is on generalizing Ye's algorithm to the infinite linear programming case. In doing so, we present an alternative technique for tackling the continuous linear programming problem, and linear optimal control problem with state constraints. In the papers by Pullan [41], theoretical properties of CLP and numerical algorithms for solving this problem are presented. Essentially, they involve making a discretization of the problem and using finite dimensional linear programming algorithms to solve them. Convergence results are also presented. Accompanying our generalization of Ye's algorithm, we present a convergence analysis of the algorithm and illustrate how it can be used to solve continuous linear programming problems.

In summary, an outline of this thesis is as follows. In Chapter 2, we study the LQG control problem with finitely many, and infinitely many linear integral constraints. Duality theory plays a fundamental role in this study, and allows us to explore the issue of the Separation Theorem as well as to derive computational algorithms for solving this problem. We prove that the Separation Theorem can be generalized to these cases. The computational algorithms are derived from the so called 'dual problem', and can be solved (in the case of finitely many constraints) using standard algorithms for finite dimensional optimization problems. We examine computational issues and in particular, the issue of calculating the gradient of the dual cost functional.

In Chapter 3, we turn our attention to the convex LQG problem with finitely many convex integral quadratic constraints. As in the linearly constrained case considered in Chapter 2, duality theory is used to study this problem. We show that the Separation Theorem does not hold since the associated 'control' problem is dependent on the solution of the 'filtering' problem. Rather, a result which we have called a Quasi-separation Theorem holds instead. However, this interdependence of the 'control' and 'filtering' problems presents no difficulties when calculating the optimal control, which can be determined by solving a finite dimensional optimization problem. As in the linearly constrained case, this finite dimensional problem can be solved using standard numerical software packages, so long as the gradient of the cost functional can be calculated. We conclude this chapter with a derivation of this gradient formula. In particular, we show that solving an unconstrained LQG problem is an essential step
is calculating the gradient.

In Chapter 4, we consider an extension of the results in Chapter 3. Recently, new results on the optimal control of stochastic LQG problems with an indefinite control weight have been obtained by Chen, Li and Zhou [58]. They prove that a common assumption that is made in LQG control, namely that the control weighting matrix is positive definite, is not necessary when the diffusion term in the linear system is dependent on the control. In fact, they derive a necessary and sufficient condition under which the problem is well posed (that is, has finite optimal cost that can be achieved by a unique optimal control) and show that the control weighting matrix may have negative eigenvalues. In this chapter, we combine the results obtained by Chen, Li and Zhou [58] with those from Chapter 3 by studying LQG control problems with finitely many integral quadratic constraints, where the control weighting matrix in any of the cost or constraint functionals may have negative eigenvalues. In particular, we show that the necessary and sufficient condition for well posedness of the unconstrained problem, as derived in [58], is actually a necessary and sufficient condition for strict convexity of the corresponding (indefinite) cost functional. Duality theory is once again used to derive the expression for the optimal control.

In Chapter 5, we generalize a path following IPM so that it can be used to solve quadratic optimization problems with infinitely many linear constraints, where the variable may belong to an infinite dimensional Hilbert space. As mentioned before, the purpose of this work is to present alternative techniques for solving linearly constrained LQ problems, and is in keeping with the idea that the results on linearly constrained LQG derived in Chapter 2 have the potential for wide applicability if efficient methods for calculating the optimal control are available. We prove global convergence of the generalized algorithm. Moreover, we show how this algorithm can be used to solve linearly constrained LQ problems. Actually, the key step in the implementation of any IPM is one known as the ‘Newton step’, and efficient methods for doing this are required if the IPM is to perform efficiently. We show that calculating the Newton step is equivalent to solving an unconstrained LQ problem together with an integral equation. This work raises several open questions. In particular, it has not yet been determined whether it is possible to derive complexity bound for this algorithm. When the number of constraints is finite, this algorithm is proven to have a worse case polynomial complexity. The efficiency of this algorithm in numerical simulations also needs to be explored.
CHAPTER 1. INTRODUCTION

In Chapter 6, we generalize the potential reduction IPM of Ye [56] to infinite dimensions and show how it can be used to solve continuous linear programming, and linear optimal control problems with continuous state constraints. Convergence properties of this algorithm are presented in some detail, and under certain conditions, global convergence is proven. As in the case of the path following IPM studied in Chapter 5, the key step in applying this IPM is calculating the associated Newton step. We show that in the case of continuous linear programming and state constrained linear optimal control, the problem of calculating the Newton step is equivalent to solving a pair of integral equations. A formulation of complexity bounds remains an open question, as does generalization to infinite quadratic programming. The efficiency of this algorithm in numerical implementations also needs to be explored.

We conclude our work in Chapter 7 with a summary of our results, and a discussion of future work.
Chapter 2

Linearly constrained LQG control

In this chapter, we consider the linearly constrained LQG control problem. In particular, we focus on generalizing results that are closely associated with unconstrained LQG theory such as the existence of a closed form expression for the optimal control and the Separation Theorem. In deriving these results, duality theory plays a fundamental role. In the case of finitely many integral linear constraints, we prove that the optimal control is obtained by solving a finite dimensional optimization problem. One major advantage of this result is that the optimal control problem (an infinite dimensional, closed loop optimization problem) is transformed to an equivalent finite dimensional problem that can be solved using standard nonlinear optimization algorithms, so long as the gradient of the cost functional can be calculated. We show that the key step in calculating the gradient is solving an unconstrained LQG control problem. In the case of linearly constrained LQG however, this finite dimensional optimization problem can be simplified even further. In fact, it is equivalent to a linearly constrained quadratic optimization problem for which a large number of specialized algorithms can be used to obtain the optimal solution. We shall be studying LQG problems with linear integral constraints. In fact, this class of constraints is quite general. For example, LQG control with linear terminal state inequality constraints as well as finitely many intermediate state inequality constraints are special cases of the problem with finitely many integral linear constraints. In the case when there are infinitely many integral linear constraints, the optimal control is obtained by solving a nonlinear infinite programming problem. As in the finite constraint case,
this problem includes continuous linear state constraints such as \( c_1(t) \leq x(t) \leq c_2(t) \) as a special case.

The second issue we address is that of the Separation Theorem. The Separation Theorem is a fundamental result associated with unconstrained LQG control. It is a major reason why LQG control for partially observable linear systems is so popular. Indeed, a consequence of the Separation Theorem is that the optimal control is elegant and easy to calculate. It is obtained by solving the minimum error variance (Kalman) filtering problem and the unconstrained LQ optimal control problem and putting the solutions of the two together, with the state estimates replacing the states. Importantly, the problems of ‘filtering’ and ‘control’ can be solved independently. This result was proven by Wohnam in [55]. In this chapter, we prove that the Separation Theorem holds for LQG control problems subject to finitely many and infinitely many integral linear constraints. In our knowledge, this is the first such result for constrained LQG control.

An outline of this chapter is as follows. In Section 2.1, we introduce definitions and results from convex analysis and optimization theory that are important in this chapter. In Section 2.2, we consider the LQG problem with finitely many linear integral constraints. We derive the optimal control for the deterministic, stochastic full observation and stochastic partial observation problems, and show (using duality theory) how these problems can be transformed into finite dimensional open loop optimization problems. Furthermore, we prove the Separation Theorem. We study the LQG problem subject to infinitely many linear integral constraints in Section 2.3. As for the case of finitely many constraints, we derive a closed form expression for the optimal control and prove the Separation Theorem.

These results can be found in the papers by Lim, Moore and Faybusovich [25] and Lim and Moore [31]. A discrete time version can be found in [24].

2.1 Mathematical preliminary

The constrained LQG problems are solved using techniques from convex optimization. A summary of relevant results is as follows.

Let \( X \) be a vector space and \( \Omega \) a convex subset of \( X \). Let \( f : X \to \mathbb{R} \) be a real
valued convex functional defined on $X$. Let $Z$ be a normed space (possibly infinite dimensional) and $P$ be a given cone in $Z$ with vertex at the origin which is defined as follows (see [34]):

**Definition 2.1** A subset $P$ of a linear vector space is said to be a cone with vertex at the origin if $x \in P$ implies that $\alpha x \in P$ for all $\alpha \geq 0$.

Given this cone $P \subset Z$, we can define the relation $\leq$ between elements $y, z \in Z$ as follows: if $y, z \in Z$, then $y \leq z$ if and only if $z - y \in P$. Sometimes, we will represent such a relation between $y$ and $z$ by $z \geq y$. The cone $P \subset Z$ through which this relation is defined is called the positive convex cone in $Z$. Let $G : X \to Z$ be a convex mapping; that is, given $x, y \in X$ and $\lambda \in [0, T]$, $G(\lambda x + (1 - \lambda) y) \leq \lambda G(x) + (1 - \lambda) G(y)$ where the relation $\leq$ is defined by the convex cone $P$ in $Z$. We consider the optimization problem:

$$\begin{align*}
\begin{cases}
\text{minimize: } & f(x) \\
\text{subject to: } & x \in \Omega, G(x) \leq \theta
\end{cases}
\end{align*}$$

(2.1)

where $\theta$ denotes the null element of $Z$. Unless otherwise mentioned, we shall assume the following:

**Assumption 2.1** The convex cone $P \subset Z$ contains an interior point.

We shall write $y > \theta$ if $y$ is an interior point of $P$.

We denote the dual space (that is, the space of bounded linear functionals) of $Z$ by $Z^*$. For every $z^* \in Z^*$ and $z \in Z$, the functional $z^*$ evaluated at $z$ is $\langle z, z^* \rangle$. If $Z$ has a positive convex cone $P$, we can define a corresponding convex positive cone $P^\oplus$ in $Z^*$ by

$$P^\oplus = \{ z^* \in Z^* : \langle z, z^* \rangle \geq 0 \text{ for all } z \in P \}$$

As before, if $y^*, z^* \in Z^*$ we write $z^* \geq y^*$ if $z^* - y^* \in P^\oplus$. The dual functional $\varphi$ defined on the positive cone $P^\oplus$ in $Z^*$ is

$$\varphi(z^*) = \inf_{x \in \Omega} \left[ f(x) + \langle G(x), z^* \rangle \right], \quad z^* \geq \theta$$

(2.2)

Note moreover that the dual functional (2.2) has the following important property [34]:
Theorem 2.1 If \( f: X \to \mathbb{R}, G: X \to Z \) and \( \Omega \) are convex, then the dual functional \( \varphi: \mathcal{P}^\theta \to \mathbb{R} \) is a concave functional.

Often it is convenient to consider the expression \( f(x) + \langle G(x), z^* \rangle \) separately. This expression is called the Lagrangian and written as

\[
L(x, z^*) = f(x) + \langle G(x), z^* \rangle \tag{2.3}
\]

The central result of this section is the Lagrange Duality Theorem [3, 34] for which the following assumption is required.

Assumption 2.2 For every \( z \in Z^* \) such that \( z \geq \theta \) and \( z \neq \theta \), there exists \( x \in \Omega \) such that \( \langle G(x), z \rangle < 0 \).

Remark 2.1 This assumption is automatically satisfied if there exists \( x_1 \in \Omega \) such that \( G(x_1) < \theta \). Such an \( x_1 \) is said to be strictly feasible for the problem (2.1).

Note however that in the case of finitely many linear inequality constraints, Assumption 2.2 can be relaxed to the existence of \( x_1 \in \Omega \) such that \( G(x_1) \leq \theta \). We shall return to this issue in the relevant section.

The Lagrange Duality Theorem can be stated as follows:

Theorem 2.2 (Lagrange Duality) Let \( f: X \to \mathbb{R} \) be a real valued convex functional defined on a convex subset \( \Omega \) of a vector space \( X \), and let \( G: X \to \mathbb{R} \) be a convex mapping of \( X \) into a normed space \( Z \) with convex cone \( \mathcal{P} \). Suppose that there exists \( x_1 \in \Omega \) such that \( G(x_1) < \theta \) and that \( \mu_0 = \inf \{ f(x) : G(x) \leq \theta, x \in \Omega \} \) is finite. Then

\[
\inf_{G(x) \leq \theta} f(x) = \max_{x \in \Omega} \varphi(z^*) \tag{2.4}
\]

and the maximum on the right is achieved by some \( z^* \geq \theta \). If the infimum on the left is achieved by some \( x^* \in \Omega \), then

\[
\langle G(x^*), z^* \rangle = 0
\]

and \( x^* \) minimizes \( f(x) + \langle G(x), z^* \rangle, x \in \Omega \).
Note that Theorem 2.2 does not guarantee the existence of an optimal solution $x^*$ for the problem (2.1). In the following sections of this chapter, we shall use the results in Theorem 2.2 to prove generalizations of the Separation Theorem for various constrained LQ/LQG control problems and derive computational methods to solve them. In each case, we will prove the existence of an optimal solution $x^*$.

2.2 LQG control with finitely many linear constraints

In this section, we turn our attention to the LQG control problem with finitely many integral linear constraints. We consider the deterministic, the full observation and the partial observation versions of this problem. In the partially observed case, we prove that the Separation Theorem holds. For the deterministic problem, this framework includes problems with finitely many hard constraints, for example $|x_1(t)| \leq c$ for $t = 1, \cdots, T$ and in the stochastic case, it includes constraints of the form $E|x_1(t)| \leq c$. In practice, stochastic control problems with constraints of this form arise less frequently than problems with hard constraints (and in particular, problems with pointwise hard constraints). Problems with hard constraints are considerably more difficult than problems with constraints on the mean of the components of the state. However, problems of this form can be approached by placing constraints on the variance of the state and/or control in addition to constraints on the mean of the state. We consider this problem in Chapter 3.

2.2.1 Mathematical Preliminary

We begin by refining some of the results stated in Section 2.1 to the case of finitely many linear constraints. In particular, we consider the issue of solving the dual problem, as given by the right hand side of (2.4), which forms the basis of our algorithm for solving the LQG problem with finitely many real valued linear constraints. In such a case, the dual problem is a finite dimensional optimization problem, and we derive the gradient of the dual cost functional so that it can be solved using gradient-type optimization algorithms.

In this section, suppose that $X$ is a Hilbert space, $\Omega$ a closed affine subspace of $X$ and $f : X \rightarrow \mathbb{R}$ a strictly convex linear-quadratic functional. It follows that $\Omega = z + \mathcal{Y}$.
where $z \in X$ and $Y$ is a closed subspace of $X$. Let $G : X \to \mathbb{R}^N$ be given by

$$G(x) = \begin{bmatrix} f_1(x) - c_1 \\ \vdots \\ f_N(x) - c_N \end{bmatrix}$$

(2.5)

where $f_i \in X^*$ and $c = [c_1, \ldots, c_N] \in \mathbb{R}^N$ are given a priori. In this case the convex cone in $\mathbb{R}^N$ is

$$P = \{x \in \mathbb{R}^N : x_i \geq 0, \ i = 1, \ldots, N\}$$

and Assumption 2.1 is clearly satisfied. Moreover, it is proven in [10] that Assumption 2.2 can be relaxed to the following:

**Assumption 2.3** There exists $x \in \Omega$ such that $G(x) \leq \theta$.

Under the assumptions above, we have the following existence result for the problem (2.1):

**Theorem 2.3** There exists an optimal solution $x^*$ of (2.1).

**Proof:** Let $x$ be feasible for (2.1) and $\sigma = f(x)$. Define

$$H_\sigma = \{x \in \Omega : f(x) \leq \sigma\}$$

and consider the problem

$$\begin{cases} 
    f(x) \to \max \\
    f(x) \leq \sigma \\
    G(x) \leq \theta \\
    x \in \Omega 
\end{cases}$$

(2.6)

Since $f(x)$ is strictly convex on $X$, the constraints of (2.6) define a bounded, closed, convex subset of the Hilbert space $X$. Since every continuous, convex functional defined on a Hilbert space achieves its minimum on every bounded, closed, convex set [3, Theorem 2.6.1], it follows that there exists $x^*$ which is optimal for (2.6). Furthermore, the uniqueness of $x^*$ follows from the strict convexity of $f(x)$. Now we show that $x^*$ is optimal for (2.1). Suppose that this is not true. Then there exists a $\bar{x}$ which is feasible for (2.1) such that $f(\bar{x}) < f(x^*) \leq \sigma$. However, this implies that
2.2. LQG CONTROL WITH FINITELY MANY LINEAR CONSTRAINTS

\( \bar{x} \) is feasible for (2.6) and hence, \( f(x^*) \leq f(\bar{x}) \) - a contradiction. Therefore, by the definition of \( x^* \) we have \( f(x^*) \leq f_0(\bar{x}) \) for all feasible solutions \( \bar{x} \) of (2.1). The result follows.

The dual space of \( \mathbb{R}^N \) is \( (\mathbb{R}^N)^* = \mathbb{R}^N \) and the positive cone \( P^\rho \) on \( X^* = X \) induced by \( P \) is \( P^\rho = P \). The Lagrangian defined by (2.3) becomes

\[
L(x, \lambda) = f(x) + \lambda'G(x) = f(x) + \sum_{i=1}^{N} \lambda_i (f_i(x) - c_i) \tag{2.7}
\]

Given any \( \lambda \geq 0 \), the dual function defined by (2.2) becomes

\[
\varphi(\lambda) = \min_{x \in \Omega} L(x, \lambda) \tag{2.8}
\]

Noting the result in Theorem 2.3, it follows that Theorem 2.2 can be strengthened to the following:

**Theorem 2.4** Under the conditions stated above, there exists a unique optimal solution \( x^* \) for the optimization problem (2.1). For every \( \lambda \geq 0 \), let \( \varphi(\lambda) \) be given by (2.8) and \( x(\lambda) \) be defined by

\[
x(\lambda) = \arg \min_{x \in \Omega} L(x, \lambda) \tag{2.9}
\]

Then \( x(\lambda) \) exists and the Lagrange multiplier \( \lambda^* \) defined by

\[
\lambda^* = \arg \max_{\lambda \geq 0} \min_{x \in \Omega} L(x, \lambda) \tag{2.10}
\]

\[
= \arg \max_{\lambda \geq 0} \varphi(\lambda) \tag{2.11}
\]

eexists. Also, \( x^* = x(\lambda^*) \).

**Proof:** We need only show that \( x(\lambda) \) exists for every \( \lambda \geq 0 \). Every other result follows from Theorem 2.2 and Theorem 2.3. Since \( f_i(x), i = 1, \ldots, N \) are bounded linear functionals and \( f(x) \) is a strictly convex linear-quadratic functional, it follows that for every given \( \lambda \geq 0 \), \( L(x, \lambda) \) is a strictly convex linear-quadratic functional over \( X \). Let \( \bar{x} \in \Omega \) and let \( \sigma = L(\bar{x}, \lambda) \). Since \( L(x, \lambda) \) is strictly convex over \( X \), \( H_\sigma = \{x \in \Omega : L(x, \lambda) \leq \sigma\} \) is a bounded, closed convex subset. By [3, Theorem 2.6.1], every continuous convex functional defined on a Hilbert space achieves its minimum on every bounded, closed, convex subset of the Hilbert space. Hence, \( L(x, \lambda) \) achieves its minimum \( \bar{x}(\lambda) \) on \( H_\sigma \). Moreover, strict continuity of \( L(x, \lambda) \) with
respect to $x \in X$ implies that $\bar{x}(\lambda)$ is the unique optimal solution over $H_\sigma$. Now, $\bar{x}(\lambda)$ is optimal over $\Omega$ since the existence of $y \in \Omega$ such that $L(y, \lambda) < L(\bar{x}(\lambda), \lambda)$ implies that $y \in H_\sigma$ and hence, $L(\bar{x}(\lambda), \lambda) \leq L(y, \lambda)$ by the optimality of $\bar{x}(\lambda)$ over $H_\sigma$ - contradiction. Therefore, $L(\bar{x}(\lambda), \lambda) \leq L(y, \lambda)$ for every $y \in \Omega$ so $x(\lambda) = \bar{x}(\lambda)$ by the definition of $x(\lambda)$. 

The solution $x(\lambda)$ of (2.9) is uniquely characterized by the condition

$$\nabla_x L(x, \lambda) \in \mathcal{Y}^\perp$$

(2.12)

where for any functional $h : X \to \mathbb{R}$ and any $x \in X$, $\nabla_x h(x) \in X^* = X$ is the gradient of $h$ at $x$. Since $f(x)$ is a convex linear-quadratic functional, it follows that

$$f(x) = \frac{1}{2} \langle Q \cdot x, x \rangle + \langle a, x \rangle + b$$

where $Q : X \to X$ is a symmetric positive operator and $a \in X$ and $b \in \mathbb{R}$. Without loss of generality, we assume that $b = 0$. Also, since $f_i \in X^*$, there exists $a_i \in X$, $i = 1, \cdots, N$ such that

$$f_i(x) = \langle a_i, x \rangle$$

Hence

$$L(x, \lambda) = \frac{1}{2} \langle Q \cdot x, x \rangle + \langle a(\lambda), x \rangle - \lambda^T c$$

where $a(\lambda) = a + \sum_{i=1}^{N} \lambda_i a_i$, and the characterization (2.12) of $x(\lambda) \in \Omega$ becomes

$$Q \cdot x(\lambda) + a(\lambda) \in \mathcal{Y}^\perp$$

(2.13)

We re-iterate the fact that for any given $\lambda \geq 0$, the problem of finding $x(\lambda) \in \Omega$ such that (2.13) holds is equivalent to solving the unconstrained convex quadratic optimization problem (2.9).

When $x = x(\lambda)$, the dual optimization problem (2.11) over $\lambda$ becomes

$$\begin{cases}
\varphi(\lambda) = \frac{1}{2} \langle Q \cdot x(\lambda), x(\lambda) \rangle + \langle a(\lambda), x(\lambda) \rangle - \lambda^T c \to \text{max} \\
Q \cdot x(\lambda) + a(\lambda) \in \mathcal{Y}^\perp \\
x(\lambda) \in \Omega \\
\lambda \geq 0
\end{cases}$$

(2.14)

This is a finite dimensional optimization problem over $\mathbb{R}^N$. It can be solved using algorithms for solving finite dimensional optimization problems if for any given $\lambda \geq 0$,}
2.2. LQG CONTROL WITH FINITELY MANY LINEAR CONSTRAINTS

\( \varphi(\lambda) \in \mathbb{R} \) and the gradient vector \( \frac{d\varphi(\lambda)}{d\lambda} \in \mathbb{R}^N \) can be calculated. The problem of calculating \( \varphi(\lambda) \) is easy: Given \( \lambda \geq 0 \), one solves the unconstrained quadratic optimization problem (2.9) for \( x(\lambda) \) and \( \varphi(\lambda) \) is given by \( \varphi(\lambda) = L(x(\lambda), \lambda) - \lambda c \).

The gradient \( \frac{d\varphi(\lambda)}{d\lambda} \) is stated in the following theorem.

**Theorem 2.5** Let \( \lambda \in \mathbb{R}^N, \lambda \geq 0 \) be given. The gradient of the dual cost functional \( \varphi(\lambda) \) with respect to \( \lambda \) is

\[
\frac{d\varphi(\lambda)}{d\lambda} = \left[ \frac{\partial \varphi(\lambda)}{\partial \lambda_1}, \ldots, \frac{\partial \varphi(\lambda)}{\partial \lambda_n} \right]
\]

where

\[
\frac{\partial \varphi(\lambda)}{\partial \lambda_i} = \langle a_i, x(\lambda) \rangle - c_i = f_i(x(\lambda)) - c_i \tag{2.15}
\]

and \( x(\lambda) \) is the optimal solution of (2.9) which is uniquely characterized by

\[
Q \cdot x(\lambda) + a(\lambda) \in \mathcal{Y}^\perp
\]

**Proof:** By the chain rule

\[
\frac{\partial \varphi(\lambda)}{\partial \lambda_i} = \frac{\partial \varphi(\lambda)}{\partial \lambda_i} + \frac{\partial \varphi(\lambda)}{\partial x(\lambda)} \cdot \frac{\partial x(\lambda)}{\partial \lambda_i} = \langle a_i, x \rangle - c_i + (Q \cdot x(\lambda) + a(\lambda)) \cdot \frac{\partial x(\lambda)}{\partial \lambda_i}
\]

Since \( x(\lambda) \in \Omega = z + \mathcal{Y} \) for every \( \lambda \geq 0 \), we have

\[
\frac{\partial x(\lambda)}{\partial \lambda_i} \in \mathcal{Y}
\]

However, \( Q \cdot x(\lambda) + a(\lambda) \in \mathcal{Y}^\perp \) so it follows that

\[
(Q \cdot x(\lambda) + a(\lambda)) \cdot \frac{\partial x(\lambda)}{\partial \lambda_i} = 0
\]

from which we obtain our result.

2.2.2 Deterministic case

We consider now the deterministic LQ optimal control problem subject to finitely many linear integral constraints. This problem can be cast into the form (2.1) where the cost is a linear-quadratic functional, and the constraints are linear functionals.
The optimal control is determined by solving the related dual problem which in this case, is a finite dimensional optimization problem. We derive the gradient of the dual cost functional using Theorem 2.5 and show that it is calculated by solving an unconstrained LQ problem and evaluating the constraint functionals at the optimal control. We also show that in the case of LQG control with linear constraints, the dual problem is equivalent to a finite dimensional linearly constrained quadratic optimization problem.

Assume that $T < \infty$. Let $X$ be a Hilbert space with norm $\| \cdot \|_X$ induced by the inner product $\langle \cdot, \cdot \rangle_X$ on $X$. We define the Hilbert space

$$L^2(0, T; X) = \{ \phi : [0, T] \to X | \phi(\cdot) \text{ is an } X\text{-valued Borel measurable function on } [0, T] \text{ with } \int_0^T \| \phi(t) \|_X^2 \, dt < \infty \} \quad (2.16)$$

where the inner product $\langle \cdot, \cdot \rangle : L^2(0, T; X) \times L^2(0, T; X) \to \mathbb{R}$ is given by

$$\langle x, y \rangle = \int_0^T \langle x(t), y(t) \rangle_x \, dt \quad (2.17)$$

In particular, $L^2(0, T; \mathbb{R}^n)$ refers to the space of $\mathbb{R}^n$-valued measurable, square integrable functions on $[0, T]$ with inner product

$$\langle x, y \rangle = \int_0^T x(t)' \cdot y(t) \, dt$$

For each $t \in [0, T]$ let $A(t) \in \mathbb{R}^{n \times n}$ and $B(t) \in \mathbb{R}^{n \times m}$. Assume that $A(t)$ and $B(t)$ are continuous, Borel measurable functions of $t \in [0, T]$. Consider the deterministic linear system

$$x(t) = A(t)x(t) + B(t)u(t), \quad x_0 = \xi \quad (2.18)$$

Let $U = L^2(0, T; \mathbb{R}^m)$ be the class of feasible controls. If $x \in L^2(0, T; \mathbb{R}^n)$ is the solution of (2.18) corresponding to $u \in U$, we shall refer to $(x, u)$ as an admissible state-control pair. Let $\mathcal{Y}$ be the closed subspace of the Hilbert space $X = L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m)$ defined by

$$\mathcal{Y} = \{(x, u) \in X : \dot{z}(t) = A(t)x(t) + B(t)u(t), \quad z(0) = 0\}$$

and $z = (\bar{z}, 0) \in L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m)$ where $\bar{z}$ is the solution of the differential equation

$$\dot{\bar{z}}(t) = A(t)\bar{z}(t); \quad \bar{z}(0) = \xi$$
2.2. LQG CONTROL WITH FINITELY MANY LINEAR CONSTRAINTS

Clearly, the set $\Omega = z + \mathcal{Y}$ is a closed affine subspace of $X$. Let the cost functional $f : X \to \mathbb{R}$ be given by

$$f(x, u) = \frac{1}{2} \int_0^T [x'(t)Q(t)x(t) + u(t)R(t)u(t)] \, dt + \frac{1}{2} x(T)'Hx(T)$$

(2.19)

where for each $t \in [0, T]$, $Q(t) \in \mathbb{R}^{n \times n}$ and $H \in \mathbb{R}^{n \times n}$ are symmetric positive semi-definite matrices and $R(t) \in \mathbb{R}^{m \times m}$ is symmetric positive definite matrix. Furthermore, we shall assume that $Q(t)$ and $R(t)$ are continuous, Borel measurable functions of $t \in [0, T]$. Let the constraint functionals $f_i : X \to \mathbb{R}$ ($i = 1, \cdots, N$) be given by

$$f_i(x, u) = \int_0^T [a_i(t)x(t) + b_i(t)u(t)] \, dt + h_i x(T)$$

(2.20)

where $a_i(t) \in \mathbb{R}^n$ and $b_i(t) \in \mathbb{R}^m$ for each $t \in [0, T]$, and $h_i \in \mathbb{R}^n$. We assume that $a_i(t)$ and $b_i(t)$ are piecewise continuous, Borel measurable functions of $t \in [0, T]$. It should be noted by the reader that for deterministic systems, open-loop and closed-loop control is equivalent. The LQ optimal control problem with integral linear constraints can be described as follows:

**Problem:** Find the control function $u \in \mathcal{U}$ which minimizes the cost functional $f(x)$ given by (2.19) and satisfies the linear system (2.18) and the constraints $f_i(x) \leq c_i$ for $i = 1, \cdots, N$ where $c_i \in \mathbb{R}$ are assigned a priori.

It is clear from the discussion above that the LQ optimal control problem with finitely many linear integral constraints is a linearly constrained (convex) quadratic optimization problem on the affine subspace $\Omega = z + \mathcal{Y}$ of $X$, and can be written in the following way:

$$\left\{ \begin{array}{l}
\text{minimize: } f(x, u) \\
\text{subject to: } f_i(x, u) \leq c_i, \ i = 1, \cdots, N \\
(x, u) \in \Omega
\end{array} \right. \tag{2.21}$$

We shall assume that Assumption 2.3 holds for (2.21); that is:

**Assumption 2.4** There exists a feasible solution $(x, u)$ for (2.21).

Before proceeding any further, let us consider more carefully the constraint functionals $f_i(x, u)$ defined by (2.20). In fact, many useful constraints can be considered as special cases of (2.20). For example, it follows from (2.18) that

$$x(t) - \xi = \int_0^t (A(s) x(s) + B(s) u(s)) \, ds$$
Therefore, terminal state constraints of the form \( x_T \leq k \) can be written in the form (2.20) by choosing \( a_i(t) \) as the \( i^{th} \) row of \( A(t) \), \( b_i(t) \) as the \( i^{th} \) row of \( B(t) \), \( h_i = 0 \) and \( c = k - x(0) \). Using such a technique, it is clear that problems with finitely many intermediate state constraints (that is, \( x(t_i) \leq k_i, i = 1, \cdots, N \) where \( t_i \in [0, T] \)) can also be considered as a special case of (2.20). For convenience and generality however, we shall treat constraints in the form (2.20).

Returning to the problem (2.21), we can (under Assumption 2.4), use Theorem 2.4 to determine the optimal solution. As in (2.7), for every \( \lambda \geq 0 \) and \( (x, u) \in \Omega \) we define the functional

\[
L((x, u), \lambda) = f(x, u) + \sum_{i=1}^{N} \lambda_i(f_i(x, u) - c_i) .
\]

For \( \lambda \geq 0 \) define

\[
a(\lambda, t) = \sum_{i=1}^{N} \lambda_i a_i(t) \quad b(\lambda, t) = \sum_{i=1}^{N} \lambda_i b_i(t) \quad h(\lambda) = \sum_{i=1}^{N} \lambda_i h_i
\]

(2.22)

with \( c(\lambda) \) defined similarly. It follows that

\[
L((x, u), \lambda) = \frac{1}{2} \int_0^T [x'(t)Q(t)x(t) + u(t)R(t)u(t)] \, dt + \frac{1}{2} x(T)'Hx(T)
\]

\[
+ \int_0^T [a'(\lambda, t)x(t) + b'(\lambda, t)u(t)] \, dt + h'(\lambda)x(T) - \lambda'c
\]

(2.23)

Given any \( \lambda \geq 0 \), the dual functional \( \varphi(\lambda) \) is defined by (2.2). In the following proposition, we derive a closed form expression for \( \varphi(\lambda) \) by solving the optimization problem, as defined in (2.2), for (2.21). In particular, \( \varphi(\lambda) \) is the optimal cost associated with an unconstrained LQ problem with cost functional \( L((x, u), \lambda) \). The following is easily derived using dynamic programming.

**Proposition 2.1** Let \( \lambda \geq 0 \) be given and consider the problem of finding the optimal \( (x(\lambda), u(\lambda)) \in \Omega \) for the problem

\[
\begin{align*}
\min_{(x, u)} & \ L((x, u), \lambda) \\
\text{subject to:} & \ (x, u) \in \Omega
\end{align*}
\]

(2.24)

The optimal control \( u(t, \lambda) \) is given by

\[
u(t, \lambda) = -R^{-1}(t) \left[ B'(t)P(t)x(t) + B'(t)d(\lambda, t) + b(\lambda, t) \right]
\]

(2.25)
2.2. LQG CONTROL WITH FINITELY MANY LINEAR CONSTRAINTS

where $P(t)$, $d(\lambda,t)$ and $p(\lambda,t)$ are the solutions of

\begin{align}
\dot{P} &= -PA - A'P + PBR^{-1}B'P - Q, \quad P(T) = H \tag{2.26} \\
\dot{d} &= -[A - BR^{-1}B']d - a(\lambda) + PBR^{-1}b(\lambda), \quad d(T,\lambda) = h(\lambda) \tag{2.27} \\
\dot{p} &= [B'd + b(\lambda)]'R^{-1}[B'd + b(\lambda)], \quad p(\lambda,T) = 0 \tag{2.28}
\end{align}

and $x(\lambda)$ is the solution of the differential equation (2.18) when $u = u(\lambda)$. The value of the dual functional $\varphi(\lambda)$ is the optimal cost associated with (2.24), and is given by

$$
\varphi((x,u),\lambda) = \frac{1}{2}\xi'P(0)\xi + d'(\lambda,0)\xi + \frac{1}{2}p(\lambda,0) - \lambda'c \tag{2.29}
$$

By Theorem 2.4, the optimal admissible pair $(x^*,u^*)$ for (2.21) can be obtained by solving the associated dual (maximization) problem. The result is summarized in the following theorem.

**Theorem 2.6** Let $\lambda^*$ be the optimal solution of

$$
\max_{\lambda \geq 0}\left\{ d'(\lambda,0)\xi + \frac{1}{2}p(\lambda,0) - \lambda'c \right\} \tag{2.30}
$$

subject to (2.26)-(2.28). Then the optimal control for the LQ control problem with integral linear constraints (2.21) is

$$
u^*(t) = -R^{-1}(t) [B'(t)P(t)x(t) + B'(t)d(\lambda^*,t) + b(\lambda^*,t)] \tag{2.31}$$

**Proof:** From Theorem 2.4 the optimal solution of (2.21) is $(x^*,u^*) = (x(\lambda^*),u(\lambda^*))$ where $(x(\lambda),u(\lambda))$ is the solution of (2.24) and $\lambda^*$ is the solution of

$$
\lambda^* = \arg \max_{\lambda \geq 0} \varphi(\lambda)
$$

where $\varphi(\lambda)$ is given by (2.29). (2.30) follows by noting that the term $\frac{1}{2}\xi'P(0)\xi$ is independent of $\lambda$. The existence of $x^*$ and $\lambda^*$ is guaranteed by Theorem 2.4.

We make the following observations. First, the optimal control $u^*$ is not a feedback control - at every time $t > 0$, $u^*_t$ depends on the initial state $\xi$ since the optimal solution $\lambda^*$ of (2.30) depends on $\xi$. This dependence on $\xi$ cannot be avoided because the value of the constraint

$$
\int_0^t (a'_i(s)x(s) + b'_i(s)u(s)) \, ds
$$
at any time \( t \), depends on past behavior and in particular, on the value of \( \xi \).

Second, the problem (2.30) is a finite dimensional, open-loop optimization problem over \( \mathbb{R}^N \). In the literature, problems of the form (2.30) are known as optimal parameter selection problems, and can be solved using algorithms which solve finite dimensional optimization problems. We now give a brief overview of this idea but for a detailed account, the reader is encouraged to see [46].

Many gradient type optimization algorithms are iterative algorithms which improve current estimates \( \lambda \) of the optimal solution \( \lambda^* \) using gradient information. Therefore, to use these techniques to solve (2.30), we need to be able to calculate the value of the cost functional

\[
\varphi(\lambda) = d'(\lambda,0) \xi + \frac{1}{2} p(\lambda,0) - \lambda' c
\]

and the gradient

\[
\frac{d\varphi(\lambda)}{d\lambda} = \left[ \frac{\partial \varphi(\lambda)}{\partial \lambda_1}, \ldots, \frac{\partial \varphi(\lambda)}{\partial \lambda_N} \right]
\]

The value of the cost functional \( \varphi(\lambda) \) as given by (2.32) is easy to calculate. For any given \( \lambda \geq 0 \), one solves the differential equations (2.27)-(2.28) for \( d(\lambda,0) \), \( p(\lambda,0) \) and the cost functional can then be evaluated. The gradient functional can be determined by applying Theorem 2.5:

**Theorem 2.7** Let \( \lambda \geq 0 \) be given. The gradient of \( \varphi(\lambda) \) with respect to \( \lambda \) is

\[
\frac{d\varphi(\lambda)}{d\lambda} = \left[ \frac{\partial \varphi(\lambda)}{\partial \lambda_1}, \ldots, \frac{\partial \varphi(\lambda)}{\partial \lambda_N} \right]
\]

\[
\frac{\partial \varphi(\lambda)}{\partial \lambda_i} = \int_0^T \left( a'_i(t) x(t) + b'_i(t) u(t) \right) dt + h'_i x(T) - c_i
\]

where \((x(\lambda), u(\lambda))\) is the solution of

\[
\begin{align*}
\dot{x}(t) &= A(t) x(t) + B(t) u(t), \quad x(0) = \xi \\
u(t) &= -R^{-1}(t) \left[ B'(t) P(t) + B'(t) d(\lambda,t) + b(\lambda,t) \right]
\end{align*}
\]

**Proof:** By Theorem 2.5, for any fixed \( \lambda \geq 0 \), the components of the gradient vector are given by (2.34) where \((x(\lambda), u(\lambda))\) is the solution of the unconstrained LQ problem (2.24) from which the result follows immediately. \( \blacksquare \)
2.2. LQG CONTROL WITH FINITELY MANY LINEAR CONSTRAINTS

An immediate consequence of Theorem 2.7 is that a key step in calculating the gradient is solving an unconstrained LQ control problem. Optimization algorithms for solving (2.30) calculate \( \varphi(\lambda) \) and \( \frac{d\varphi(\lambda)}{d\lambda} \) at each \( \lambda \geq 0 \), and a step direction and step length for updating \( \lambda \) is determined using this information. Hence, the problem of calculating \( \lambda^* \) (and hence \( u^* \)) is equivalent to solving a sequence of unconstrained LQ problems since calculating the gradient is equivalent to solving an unconstrained LQ problem.

In fact, the dual problem (2.30) can be simplified further. We now show that it is equivalent to a finite dimensional, convex, linearly constrained quadratic optimization problem. To make this transformation, note the following. First, let \( \tilde{d}(t) = [d_1(t) \cdots d_N(t)] \in \mathbb{R}^{n \times N} \) where \( d_i(t) \) is the solution of the \( \lambda \)-independent differential equation

\[
\dot{d}_i = - [A - BR^{-1}B'] d_i - a_i + PB^{-1}b_i, \quad d_i(T) = h_i \tag{2.35}
\]

Therefore, it follows that for every \( \lambda \geq 0 \)

\[
d(\lambda, t) = d_0 + \tilde{d}(t) \lambda
\]

Similarly, by putting \( \tilde{b}(t) = [b_1(t) \cdots |b_N(t)] \in \mathbb{R}^{m \times N} \) we can write

\[
p(\lambda, t) = a(t) + 2 \beta(t) \lambda + \lambda' \Gamma(t) \lambda
\]

where

\[
\dot{\alpha} = (B' d_0 + b_0)' R^{-1} (B' d_0 + b_0), \quad \alpha(T) = 0
\]

\[
\dot{\beta} = (B' \tilde{d} + \tilde{b})' R^{-1} (B' \tilde{d} + \tilde{b}), \quad \beta(T) = 0
\]

\[
\dot{\Gamma} = (B' \tilde{d} + \tilde{b})' R^{-1} (B' \tilde{d} + \tilde{b}), \quad \Gamma(T) = 0
\]

Therefore, (2.30) is equivalent to

\[
\max_{\lambda \geq 0} \left\{ \frac{1}{2} \lambda' \cdot \Lambda(0) \cdot \lambda + \gamma' \cdot \lambda \right\} \tag{2.36}
\]

where \( \Lambda = \Gamma(0) \) and \( \gamma = \tilde{d}(0)' \xi - 2 \beta(0) - c \). In particular, note that (2.36) is a convex quadratic optimization problem. This is easily seen by writing \( \Gamma(0) \) as

\[
\Gamma(0) = - \int_0^T (B' \tilde{d} + \tilde{b})' R^{-1} (B' \tilde{d} + \tilde{b}) \, dt
\]

However, it may not be strictly convex!
2.2.3 Full observation case

We consider now the full observation LQG control problem subject to finitely many linear integral constraints. In particular, we prove that the optimal control for this problem can be obtained by solving the deterministic LQ problems as considered in Section 2.2.2.

Let \((\Omega, \mathcal{F}, P)\) be a probability space. Let \(\{\mathcal{F}_t\}\) be an increasing family of \(\sigma\)-algebras such that \(\mathcal{F}_t \subset \mathcal{F}\). Let \(\{W(t) : t \in [0, T]\}\) be a standard Brownian motion such that \(W(t)\) is an \(\mathbb{R}^2\)-valued random variable. Assume that \(\{W(t)\}\) is adapted to \(\{\mathcal{F}_t\}\). For a given Hilbert space \(X\) with norm \(\| \cdot \|_X\) (induced by the inner product \(\langle \cdot, \cdot \rangle_X\) defined on \(X\)), a constant \(1 \leq p \leq \infty\) and \(a, b \in \mathbb{R}\) such that \(a \leq b\), define the Banach space of stochastic processes on \((\Omega, \mathcal{F}, P)\)

\[
L^p_{\mathcal{F}}(a, b; X) = \left\{ \phi(\cdot) = \{\phi(t, \omega) : a \leq t \leq b\} : \phi(\cdot) \text{ is an } X\text{-valued } \mathcal{F}_t\text{-measurable process on } [0, T] \text{ with } E \int_a^b \|\phi(t, \omega)\|_X^p \, dt < \infty \right\} \tag{2.37}
\]

with norm

\[
\|\phi(\cdot)\|_{p, \mathcal{F}} = \left( E \int_a^b \|\phi(t, \omega)\|_X^p \, dt \right)^{1/p} \tag{2.38}
\]

Note in particular that \(L^p_{\mathcal{F}}(a, b; X)\) is a Hilbert space. Define the sets

\[
\mathcal{X} = L^2_{\mathcal{F}}(a, b; \mathbb{R}^n) \tag{2.39}
\]

\[
\mathcal{U} = L^2_{\mathcal{F}}(a, b; \mathbb{R}^n) \tag{2.40}
\]

For every \(u \in \mathcal{U}\), let \(x \in \mathcal{X}\) be the solution of the stochastic differential equation

\[
dx(t) = (A(t)x(t) + B(t)u(t)) \, dt + C(t) dW(t), \quad x(0) = \xi \tag{2.41}
\]

with \(A(t), B(t)\) as in (2.18) and \(C(t)\) an \(\mathbb{R}^{n \times j}\)-valued, Borel measurable, continuous function. We assume that \(\xi \in \mathbb{R}^n\). Let

\[
\mathcal{Y} = \{(x, u) \in \mathcal{X} \times \mathcal{U} : dx(t) = (A(t)x(t) + B(t)u(t)) \, dt, x(0) = 0\} \tag{2.42}
\]

and \(z = (\bar{z}, 0) \in \mathcal{X} \times \mathcal{U}\) where \(\bar{z}\) is the solution of the stochastic differential equation

\[
d\bar{z}(t) = A(t)\bar{z}(t) \, dt + C(t) dW(t), \quad \bar{z}(0) = \xi \tag{2.43}
\]
Then \( z + \mathcal{Y} \) is a closed affine subspace of \( \mathcal{X} \times \mathcal{U} \), and is the set of solutions of (2.41). Define the cost functional \( f : \mathcal{X} \times \mathcal{U} \to \mathbb{R} \) by

\[
f(x, u) = E \left[ \frac{1}{2} \int_0^T \left[ x'(t)Q(t)x(t) + u'(t)R(t)u(t) \right] dt + \frac{1}{2}z'(T)Hz(T) \right]
\] (2.44)

where \( Q(t) \), \( R(t) \) and \( H \) satisfy the same assumptions as in the deterministic case (2.19). Under these assumptions, \( f(x, u) \) is a strictly convex quadratic functional on \( \mathcal{X} \times \mathcal{U} \). Similarly, let the \( N \) linear constraint functionals \( f_i : \mathcal{X} \times \mathcal{U} \to \mathbb{R} \) be given by

\[
f_i(x, u) = E \left[ \int_0^T \left[ a'_i(t)x(t) + b'_i(t)u(t) \right] dt + h'_i z(T) \right]
\] (2.45)

The stochastic full observation LQG optimal control problem subject to integral linear constraints can be described as follows:

**Problem:** Find the optimal control policy \( u^* \in \mathcal{U} \) which minimizes \( f(x, u) \) such that \( (x, u) \in \mathcal{X} \times \mathcal{U} \) satisfies the linear system (2.41) and the constraints \( f_i(x, u) \leq c_i \) where \( c_i \in \mathbb{R} \) \((i = 1, \cdots, N)\) are given a priori.

Again, we note that the constraint functionals (2.45) include certain classes of constraints as a special case. For instance, recalling that the stochastic differential equation (2.41) is shorthand for

\[
x(t) - x(0) = \int_0^t (A(s) x(s) + B(s) u(s)) ds + \int_0^t C(s) dW(s)
\]

where \( \int_0^t C(s) dW(s) \) is an integral in the Ito sense \([37]\), it follows that

\[
E x(t) = \xi + E \int_0^t (A(s) x(s) + B(s) u(s)) ds
\]

Therefore, inequality constraints on the expectation of the terminal state can be expressed in the form (2.45). Generalizing this technique to incorporate finitely many inequality constraints on the expectation of intermediate values of the state, in the framework of (2.45) is straightforward. Unfortunately, inequality constraints on the expectation of the state are not common in practice. Indeed, hard constraints of the form \( x(T) \leq c \) a.s. (as opposed to \( E x(T) \leq c \)) are much more common in applications. However, constraints of this form are difficult. We do not consider such issues here.

Once again the full observation LQG control problem with linear integral constraints is a linearly constrained convex quadratic optimization problem on an affine subspace of a Hilbert space and may be stated in the following way:

\[
\begin{aligned}
\text{minimize: } & f(x, u) \\
\text{subject to: } & f_i(x, u) \leq c_i, \quad i = 1, \cdots, N \\
&(x, u) \in z + \mathcal{Y}
\end{aligned}
\] (2.46)
We assume that Assumption 2.3 holds for (2.46); that is:

**Assumption 2.5** There exists an admissible pair \((x, u)\) which is feasible for (2.46).

Under Assumption 2.5, it follows from Theorem 2.3 that there exists a unique optimal admissible pair \((x^*, u^*)\) for (2.46).

The optimal solution of (2.46) can be determined using Theorem 2.4. Once again, for \(\lambda \geq 0\) and \((x, u) \in X \times U\) we define the Lagrangian

\[
L((x, u), \lambda) = f(x, u) + \sum_{i=1}^{N} \lambda_i (f_i(x, u) - c_i)
\]

With \(f(x, u)\) and \(f_i(x, u)\) given by (2.44)-(2.45), it follows that

\[
L((x, u), \lambda) = E \left[ \frac{1}{2} \int_0^T \left[ x'(t)Q(t)x(t) + u'(t)R(t)u(t) \right] dt + \frac{1}{2} x'(T)Hx(T) 
+ \int_0^T \left[ a'(\lambda, t)x(t) + b'(\lambda, t)u(t) \right] dt + h'(\lambda)z(T) \right] - \lambda' c
\]

where \(a(\lambda, t), b(\lambda, t)\) and \(h(\lambda)\) as defined by (2.22). Noting that \(\lambda' c\) is independent of \((x, u)\) for every fixed \(\lambda \geq 0\), it follows that \(L((x, u), \lambda)\) is a standard LQG cost functional.

Given any \(\lambda \geq 0\), it follows from (2.8) that the dual cost functional \(\varphi(\lambda)\) is the optimal cost associated with an unconstrained full observation LQG problem with cost functional (2.47). This can be stated as follows.

**Proposition 2.2** Let \(\lambda \geq 0\) be given and consider the problem of finding the optimal admissible pair \((x, u)\) for the problem

\[
\left\{ \begin{array}{c}
\min L((x, u), \lambda) \\
\text{subject to: } (x, u) \in X + Y
\end{array} \right. \quad (2.48)
\]

The optimal control \(u(\lambda, t)\) is given by

\[
u(\lambda, t) = -R^{-1}(t) \left[ B'(t)P(t)x(t) + B'(t)d(\lambda, t) + b(\lambda, t) \right]
\]

where \(P(t), d(\lambda, t)\) and \(p(\lambda, t)\) are the solutions of (2.26)-(2.28). The value of the dual functional \(\varphi(\lambda)\) is the optimal cost associated with (2.48), and is given by

\[
\varphi(\lambda) = \frac{1}{2} \xi'P(0)\xi + d'(\lambda, 0)\xi + \frac{1}{2} b(\lambda, 0) + \frac{1}{2} \int_0^T \text{tr} \{C'(t)P(t)C(t)\} dt - \lambda' c
\]

(2.50)
2.2. LQG CONTROL WITH FINITELY MANY LINEAR CONSTRAINTS

Proof: When $\lambda \geq 0$, and the class of feasible controls is the set of feedback controls $\mathcal{V}$ where

$$\mathcal{V} = \{u \in U : u(t) \text{ is measurable with respect to } X_t\}$$

it is a classical result of stochastic control that the optimal control of the unconstrained full information LQG problem

$$\begin{cases} 
\min L((x, u), \lambda) \\
\text{subject to: } u \in \mathcal{V} \text{ and } (x, u) \text{ satisfies (2.41)}
\end{cases}$$

is given by (2.49). It is proven in [15, Corollary 4.1, pp 163] that (2.49) is also optimal over the class $x + \mathcal{V}$ from which we obtain the optimal control (2.49) and the resulting optimal cost (2.50). By the definition (2.8), it follows that $\varphi(\lambda)$ is given by (2.50).

The expression of the optimal control for the full information LQG problem follows immediately from Theorem 2.4.

Theorem 2.8 The unique optimal control for the full observation LQG optimal control problem with finitely many integral linear constraints is

$$u(t)^* = -R^{-1}(t) \left[ B'(t)P(t)x(t) + B'(t)d(\lambda^*, t) + b(\lambda^*, t) \right]$$

where $\lambda^*$ is the optimal solution of (2.26)-(2.28), (2.30).

Proof: Under Assumption 2.5, it follows from Theorem 2.3 that there exists a unique admissible pair $(x^*, u^*)$ that is optimal for (2.46). By Theorem 2.4, $(x^*, u^*) = (x(\lambda^*), u(\lambda^*))$ where $u(\lambda)$ is given by (2.49) and $\lambda^*$ is the optimal solution of

$$\lambda^* = \arg \max_{\lambda \geq 0} \varphi(\lambda)$$

In Proposition 2.2, it is shown that $\varphi(\lambda)$ is given by (2.50). The result follows by noting that

$$\frac{1}{2}\xi'P(0)\xi + \frac{1}{2} \int_0^T \text{tr} \{C'(t)P(t)C(t)\} \, dt$$

is a constant that is independent of $\lambda$. The existence of $\lambda^*$ follows from Theorem 2.4.
Note in particular that as in the case of the unconstrained LQG problem, certainty equivalence holds. That is, the optimal control for the full observation problem (2.46) can be obtained from that of the deterministic problem (2.21), simply by replacing the deterministic state \( x(t) \) in (2.31) by the solution \( x(t) \) of the stochastic differential equation (2.41). This is not true for more general constrained LQG problems. In particular, we show in Chapter 3 that certainty equivalence does not hold when there are integral quadratic constraints. Therefore, by solving the optimal parameter selection problem (2.26)-(2.28), (2.30) associated with the deterministic problem (2.21), the optimal control can be determined.

2.2.4 Partial observation case

We now consider the partially observed LQG problem with finitely many linear constraints. In particular, we prove that the Separation Theorem can be generalized to this case.

The following assumptions are made in addition to the ones for the full observation case in Section 2.2.3. Let \( F(t) \) and \( G(t) \) be continuous, Borel measurable, matrix valued functions of \( t \in [0,T] \) such that \( F(t) \in \mathbb{R}^{p \times n} \) and \( G(t) \in \mathbb{R}^{p \times k} \). Let \( \{V(t) : t \in [0,T]\} \) be a standard Brownian motion such that \( V(t) \) is an \( \mathbb{R}^k \)-valued random variable for every \( t \in [0,T] \). We assume that \( x(0), \{V(t)\} \) and \( \{W(t)\} \) are mutually independent. Consider the partially observed linear system

\[
\begin{align*}
\dot{x}(t) &= A(t)x(t)dt + B(t)u(t)dt + C(t)dW(t), \quad x(0) \sim N(\xi, \Sigma_0) \quad (2.51) \\
\dot{y}(t) &= F(t)x(t)dt + G(t)dV(t), \quad x(0)y(0) = 0. \quad (2.52)
\end{align*}
\]

The class of feasible controls for the system (2.51) is defined as in [15, 55] for the unconstrained partially observed LQG problem: Let \((C(0,T; \mathbb{R}^p), \| \cdot \|)\) be the Banach space of continuous \( \mathbb{R}^p \)-valued functions on \([0,T]\) with the sup norm \( \| \cdot \| \) defined by

\[
\|g\| = \sup_{t \in [0,T]} |g(t)|, \quad g \in C(0,T; \mathbb{R}^p)
\]

where \( | \cdot | \) is the Euclidean norm on \( \mathbb{R}^p \). For every \( t \in [0,T] \), define the operator \( \pi_t : C(0,T; \mathbb{R}^p) \to C(0,T; \mathbb{R}^p) \) by

\[
(\pi_t g)(s) = \begin{cases} 
  g(s), & s \in [0,t] \\
  g(t), & s \in [t,T]
\end{cases}
\]

Let \( \Psi \) be the set of functions \( \psi : [0,T] \times C(0,T; \mathbb{R}^p) \to \mathbb{R}^m \) which satisfy the following properties:
2.2. LQG CONTROL WITH FINITELY MANY LINEAR CONSTRAINTS

1. For every \( \psi \in \Psi \), there exists \( K_\psi \in \mathbb{R} \) such that \(|\psi(t, g) - \psi(t, h)| \leq K_\psi \|g - h\|\) for every \( g, h \in C(0, T; \mathbb{R}^p) \) and \( t \in [0, T] \). (Uniform Lipschitz condition).

2. \( \psi(\cdot, \cdot) \) is Borel measurable.

3. \( \psi(t, 0) \) is bounded.

The class of feasible controls is the set

\[
\mathcal{U} = \{ u : [0, T] \times C(0, T; \mathbb{R}^p) \to \mathbb{R}^m \mid \text{there exists } \psi \in \Psi \text{ with } u(t) = \psi(t, \pi_y) \text{ for every } t \in [0, T], y \text{ given by (2.52)} \}
\] (2.53)

Given \( u \in \mathcal{U} \), let \( x \) denote the solution of the state equation (2.51). We refer to the pair \((x, u)\) as an admissible pair. The cost functional associated with each admissible pair is defined by

\[
f(x, u) = E \left[ \frac{1}{2} \int_0^T \left[ x'(t)Q(t)x(t) + u(t)'R(t)u(t) \right] dt + \frac{1}{2} x'(T)Hx(T) \right]
\] (2.54)

where \( Q(t), R(t) \) and \( H \) satisfy the same assumptions as in the deterministic case (2.19). Similarly, let the \( N \) linear constraint functionals are given by

\[
f_i(x, u) = E \left[ \int_0^T \left[ a_i'(t)x(t) + b_i'(t)u(t) \right] dt + h'_ix(T) \right]
\] (2.55)

The partially observed LQG optimal control problem subject to integral linear constraints can be stated as follows:

**Problem:** Find the optimal control policy \( u^* \in \mathcal{U} \) which minimizes \( f(x, u) \) such that \((x, u)\) satisfies the linear system (2.51) and the constraints \( f_i(x, u) \leq c_i \) (\( i = 1, \ldots, N \)) where \( c_i \in \mathbb{R} \) are given a priori.

Equivalently, this problem can be written in the form:

\[
\begin{align*}
\min & \ f(x, u) \\
& f_i(x, u) \leq c_i, \ i = 1, \ldots, N \\
& (x, u) \text{ satisfies (2.51), } u \in \mathcal{U}
\end{align*}
\] (2.56)

As in the full observation case, inequality constraints on the expectation of the terminal state or finitely many intermediate states are special cases of this problem. We make the following assumption about the partially observed problem (2.56).

**Assumption 2.6** There exists a feasible solution for the problem (2.56).
Using standard techniques, we can transform the partial observation problem (2.56) into a full observation problem of the form (2.46). To summarize this, we need the following basic results from Kalman filtering theory. For every \( u \in \mathcal{U} \) let \( \mathcal{G}_t^u = \sigma \{ y(s) : s \in [0,t] \} \) be the \( \sigma \)-field generated by the output process \( y(\cdot) \) of (2.51)-(2.52) when the input of (2.51) is \( u \). Let \( \mathcal{G}_t^0 \) correspond to the case when \( u = 0 \). The following result is proven in [15].

**Lemma 2.1** If \( u(t) \) is \( \mathcal{G}_t^0 \)-measurable and \( \mathcal{G}_t^0 = \mathcal{G}_t^u \), then the conditional distribution of \( x(t) \) given \( \mathcal{G}_t^u \) is Gaussian with mean \( \hat{x}(t) = E[x(t)|\mathcal{G}_t^u] \) and covariance \( \Sigma(t) \) where

\[
\dot{\hat{x}}(t) = A(t)\hat{x}(t)dt + B(t)u(t)dt - \Sigma(t)F(t)\left( G(t)G'(t) \right)^{-1}d\nu(t), \quad \hat{x}_0 = \xi \quad (2.57)
\]

\[
\dot{\Sigma} = \Sigma A + A'\Sigma - \Sigma F'(GG')^{-1}FS + CC', \quad \Sigma(0) = \Sigma_0 \quad (2.58)
\]

and the innovations process \( \{ \nu(t) \} \) is given by \( d\nu(t) = dy(t) - F(t)\hat{x}(t)dt = G(t)d\hat{w}(t) \) where \( \hat{w}(\cdot) \) is a Brownian motion adapted to \( \{ \mathcal{G}_t^0 \} \). Furthermore

\[
E[\nu(t)] = 0, \quad E[\nu(t)\nu(t)'] = \int_0^t G(s)G'(s)ds, \quad E[\nu(t)\hat{x}(t)'] = 0,
\]

and the optimal state estimate and the optimal state estimate error are orthogonal; that is

\[
E[(x(t) - \hat{x}(t))\hat{x}'(t)] = 0
\]

In particular, when \( u \in \bar{\mathcal{U}} \), the conditions of Lemma 2.1 are satisfied. Using the results in Lemma 2.1, it is easy to show that

\[
E[x'(t)Q(t)x(t)] = E[\hat{x}'(t)Q(t)\hat{x}(t)] + \text{tr} \{Q(t)\Sigma(t)\}
\]

and hence, (2.44) becomes

\[
f(x, u) = \frac{1}{2} E \left[ \int_0^T [\hat{x}'(t)Q(t)\hat{x}(t) + u'(t)R(t)u(t)] dt + \hat{x}'(T)H\hat{x}(T) \right] \\
+ \frac{1}{2} \int_0^T \text{tr} \{Q(t)\Sigma(t)\} dt + \frac{1}{2} \text{tr} \{H \Sigma(T)\}
\]

\[
f(\hat{x}, u) + \frac{1}{2} \int_0^T \text{tr} \{Q(t)\Sigma(t)\} dt + \frac{1}{2} \text{tr} \{H \Sigma(T)\}
\]

Note that the terms \( \Sigma(t) \) is deterministic and independent of the input \( u \).
To deal with the constraints, note that when (2.57) is subtracted from (2.51), we have
\[ d(x(t) - \hat{x}(t)) = A(t)(x(t) - \hat{x}(t))dt + C(t)dW(t) + \Sigma(t)H(t) \left( G(t)G'(t) \right)^{-1} d\nu(t) \]
with \( E[x(0)] = \xi \) and \( \hat{x}(0) = \xi \). Since \( \{W(t)\} \) and \( \{\nu(t)\} \) are Brownian motions, they are both zero mean processes and it follows that
\[ E[x(t) - \hat{x}(t)] = \int_0^T A(t)E[x(t) - \hat{x}(t)] dt, \quad E[x_0 - \hat{x}_0] = 0 \]
and hence \( E[x(t)] = E[\hat{x}(t)] \). Therefore, the constraint functionals (2.45) becomes
\[
\begin{align*}
  f_i(x, u) &= \int_0^T [a'_i(t)E[x(t)] + b'_i(t)E[u(t)]] dt + h'_iE[x(T)] \\
  &= \int_0^T [a'_i(t)E[\hat{x}(t)] + b'_i(t)E[u(t)]] dt + h'_iE[\hat{x}(T)] \\
  &= E \left[ \int_0^T [a'_i(t)\hat{x}(t) + b'_i(t)u(t)] dt + h'_i\hat{x}(T) \right]
\end{align*}
\]
Therefore, the partially observed LQG control problem with linear integral constraints is equivalent to the full observation problem
\[
\begin{align*}
  \text{minimize: } f(\hat{x}, u) \\
  \text{subject to: } f_i(\hat{x}, u) \leq c_i, \quad i = 1, \ldots, N \\
  (\hat{x}, u) \text{ satisfies (2.57), } u \in \tilde{U}
\end{align*}
\]
We are now in the position to prove the following generalization of the Separation Theorem.

**Theorem 2.9 (Separation Theorem)** There exists a \( \lambda^* \) which is optimal for the problem (2.26)-(2.28),(2.30). Furthermore, the optimal control for the partially observed LQG problem with integral linear constraints (2.56) is
\[
u^*(t) = -R^{-1}(t) \left[ B'(t)P(t)\hat{x}(t) + B'(t)d(\lambda^*, t) + b(\lambda^*, t) \right] \tag{2.60}
\]
where \( \hat{x} \) is the output of the Kalman filter (2.57).

**Proof:** It is shown in [15, Lemma 11.3, pp 191] that if \( u \in \tilde{U} \), then the conditions of Lemma 2.1 are satisfied; that is \( u \) is non-anticipative with respect to \( \{G^0_t\} \) and \( G^0_t = G^u_t \). Consider the full observation problem
\[
\begin{align*}
  f_0(\hat{x}, u) \to \min \\
  f_i(\hat{x}, u) \leq \hat{c}_i, \quad i = 1, \ldots, N \\
  u(t) \text{ is } G^0_t\text{-measurable and } (\hat{x}, u) \text{ satisfies (2.57).}
\end{align*}
\]
Then (2.61) is exactly of the form (2.46). Moreover, the class of feasible controls for (2.61) contains the set $\tilde{U}$. By Assumption 2.6, the problem (2.61) satisfies the conditions of Theorem 2.4, so there exists a unique optimal admissible pair $(x^*, u^*)$ for (2.61). Moreover, the optimal control $u^*$ for the problem (2.61) is given by (2.60). Since $u^* \in \tilde{U}$ [15], it follows that $u^*$ is optimal for (2.59) and hence, optimal for (2.56).

Note that the Separation Theorem is true for the partially observed LQG problem with finitely many linear integral constraints: The optimal control (2.60) is determined by solving a ‘control problem’, namely (2.26)-(2.28), (2.29), and a (Kalman) filtering problem and these two problems can be solved independently. Moreover, the ‘control problem’ (2.26)-(2.28), (2.29) is a generalization of the ‘control problem’ associated with the Separation Theorem for the unconstrained partially observed LQG problem. In the unconstrained case, the ‘control problem’ involves solving the Riccati equation (2.26) only. Furthermore, certainty equivalence holds; that is, the optimal control for the partially observed problem is obtained by solving the deterministic problem (2.21) (as outlined in Theorem 2.6), and replacing the deterministic state $x(t)$ in the deterministic optimal control $u^*(t)$ as given by (2.31), with the output $\hat{x}(t)$ of the Kalman filter. As mentioned before, certainty equivalence does not generally hold for constrained LQG problems. Indeed, we will show in Chapter 3 that it does not hold for the LQG problem with integral quadratic constraints.

### 2.3 LQG control with infinitely many linear constraints

In this section we generalize the results of Section 2.2 by turning our attention to the LQG problem with infinitely many linear constraints. In particular, this includes the important class of LQG problems with state constraints at every time instant, for example $x(t) \leq c(t)$ in the deterministic case. As in the case of finitely many constraints, we can view this problem in the general framework of Section 2.1. The main difference between the case of infinitely many linear constraints and the case of finitely many linear constraints is that the normed space $Z$ is now infinite dimensional. However, like the case of finitely many constraints, we show that the Separation Theorem holds. We derive a computational algorithm to solve this problem and show that by solving the dual problem, the optimal control can be obtained. Although the dual problem is an infinite dimensional optimization problem (because $Z$ is infinite
2.3. LQG CONTROL WITH INFINITELY MANY LINEAR CONSTRAINTS

dimensional) it is significantly easier than the original control problem because it is
an open-loop optimization problem.

2.3.1 Mathematical preliminary

Let $X$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$, $\Omega$ a closed affine subspace of $X$
and $f : X \to \mathbb{R}$ a strictly convex, linear quadratic functional. Therefore, it follows
that

$$f(x) = \frac{1}{2} \langle Qx, x \rangle + \langle a, x \rangle + b$$

where $Q : X \to X$ is a symmetric positive operator, $a \in X$ and $b \in \mathbb{R}$. Without loss
of generality, we can assume that $b = 0$. Let $C(0,T;\mathbb{R}^N)$ denote the Banach space of
continuous $\mathbb{R}^N$-valued functions with norm

$$||f|| = \sup_{\eta \in [0,T]} |f(\eta)|$$

where $| \cdot |$ denotes the usual Euclidean norm on $\mathbb{R}^N$. Let $G : X \to C(0,T;\mathbb{R}^N)$ be
given by

$$G(x) = \begin{bmatrix} f_1(x) - c_1 \\
\vdots \\
f_N(x) - c_N \end{bmatrix} \quad (2.62)$$

where $c_i \in C(0,T;\mathbb{R})$ and $f_i : X \to C(0,T;\mathbb{R})$ are given by

$$f_i(x) = \langle a_i(\eta), x \rangle, \quad \eta \in [0,T]$$

with $a_i(\eta) \in X$ for every $\eta \in [0,T]$. We assume that $a_i(\eta)$ is a continuous function of
$\eta$. The convex cone $P$ in $C(0,T;\mathbb{R}^N)$ is

$$P = \{ x \in C(0,T;\mathbb{R}^N) : x_i(t) \geq 0 \text{ for all } t \in [0,T] \}$$

It is important to note that $P$ satisfies Assumption 2.1. Recall that $P$ defines the
relation $\leq$ between $y, z \in C(0,T)$ in the following way: $y \leq z$ if and only if $z - y \in P$;
that is, $y \leq z$ if and only if $z(t) - y(t) \geq 0$ for every $t \in [0,T]$. The following
assumption is necessary for the existence and uniqueness of an optimal solution of
(2.1):

**Assumption 2.7** There exists $x_1 \in \Omega$ such that $G(x_1) \leq \theta$. 
Theorem 2.10 Suppose that \( f, G \) and \( \Omega \) satisfy the conditions stated above, and that Assumption 2.7 holds. Then there exists a unique optimal solution \( x^* \) of (2.1).

Proof: Let \( \bar{x} \) be feasible for (2.1) with \( f \) and \( G \) given as above. Let \( \sigma = f(\bar{x}) \).

Consider the problem

\[
\begin{aligned}
&\text{minimize } f(x) \\
&\text{subject to } f(x) \leq \sigma \\
&\text{ } G(x) \leq \theta \\
&\text{ } x \in \Omega
\end{aligned}
\]  

(2.63)

Since \( f(x) \) is a strictly convex linear-quadratic functional, the set of feasible solutions of (2.63) is a non-empty, bounded, closed convex subset of \( X \). By Theorem 2.6.1 in [3], every non-empty, convex, continuous functional defined on a Hilbert space achieves its minimum on every closed, convex, bounded set. Hence, there exists a solution \( x^* \) of (2.63). The uniqueness of \( x^* \) follows from the strict convexity of \( f(x) \). We now prove that \( x^* \) is optimal for (2.1). Suppose that this is not true. Then there exists \( y \in \Omega \) such that \( G(y) \leq \theta \) and \( f(y) < f(x^*) \leq \sigma \). However, this implies that \( y \) is feasible for (2.63) and by the optimality of \( x^* \), it follows that \( f(x^*) \leq f(y) \) – a contradiction. Therefore, \( f(x^*) \leq f(y) \) for every \( y \) which is feasible for (2.1). The result follows.

The dual space of \( C(0, T; \mathbb{R}^N) \) is the set of functions of bounded variation which we denote by \( BV^N(0, T) \). Given \( f \in C(0, T; \mathbb{R}^N) \) and \( v \in BV^N(0, T) \), the value of the functional \( v \) evaluated at \( f \) is

\[
\langle f, v \rangle = \sum_{i=1}^{N} \int_{0}^{T} f_i(\eta) \, dv_i(\eta)
\]

where the integral is the Riemann-Stieltjes integral of \( f_i \in C(0, T) \) with respect to \( v_i \in BV(0, T) \), \( i = 1, \cdots, N \). The closed convex cone \( P^\ominus \) on \( BV^N(0, T) \) induced by \( P \) is

\[
P^\ominus = \{ g \in BV^N(0, T) : g_i \text{ is a non-decreasing function of } t \in [0, T], \ i = 1, \cdots, N \}
\]

Note in particular that given \( v \in BV(0, T) \), \( v \geq \theta \) means that \( v \in P^\ominus \).

The Lagrangian \( L(x, v) \) defined by (2.3) becomes

\[
L(x, v) = f(x) + \langle G(x), v \rangle
\]
2.3. LQG CONTROL WITH INFINITELY MANY LINEAR CONSTRAINTS

\[ \frac{1}{2} \langle Q x, x \rangle + \langle a, x \rangle + \sum_{i=1}^{N} \int_{0}^{T} (\langle a_i(\eta), x \rangle - c_i(\eta)) \, dv_i(\eta) \]

\[ = \frac{1}{2} \langle Q x, x \rangle + \left( a + \sum_{i=1}^{N} \int_{0}^{T} a_i(\eta) \, dv_i(\eta), x \right) - \sum_{i=1}^{N} \int_{0}^{T} c_i(\eta) \, dv_i(\eta) \] (2.64)

where the integral \( \int_{0}^{T} a_i(\eta) \, dv_i(\eta) \) is to be interpreted in the following way.

**Lemma 2.2** Let \((X, \cdot, \cdot)\) be a Hilbert space and \(\alpha(\cdot)\) be an \(X\)-valued function that is continuous with respect to \(\eta \in [0, T]\). Let \(v(\eta) \in BV(0, T)\) and define \(T : X \to \mathbb{R}\) by

\[ Ty = \int_{0}^{T} \langle \alpha(\eta), y \rangle \, dv(\eta) \] (2.65)

where the integral in (2.65) is the Riemann-Stieltjes integral of \(\langle \alpha(\eta), y \rangle \in C(0, T)\) with respect to \(v(\eta) \in BV(0, T)\). Then \(T\) is a bounded linear functional and

\[ Ty = \left \langle \int_{0}^{T} \alpha(\eta) \, dv(\eta), y \right \rangle \] (2.66)

where

\[ \int_{0}^{T} \alpha(\eta) \, dv(\eta) = \lim_{N \to \infty} \sum_{j=1}^{2N} \alpha(\eta_i) \cdot [v(\eta_i) - v(\eta_{i-1})], \quad \eta_i = \frac{j}{2N} \] (2.67)

**Proof:** \(T\) is clearly linear. Since \(\alpha(\eta)\) is continuous on \([0, T]\), it is uniformly bounded and hence, \(T\) is a bounded linear functional. By the Riesz Representation Theorem for functionals on a Hilbert space, there exists \(\Phi \in X\) such that \(Ty = \langle \Phi, y \rangle\).

On the other hand, by the definition of the Riemann-Stieltjes integral

\[ Ty = \lim_{N \to \infty} \sum_{j=1}^{2N} \langle \alpha(\eta_j), y \cdot [v(\eta_j) - v(\eta_{j-1})] \rangle \]

\[ = \lim_{N \to \infty} \left \langle \sum_{j=1}^{2N} \alpha(\eta_j) \cdot [v(\eta_j) - v(\eta_{j-1})], y \right \rangle \]

Putting

\[ z_N = \sum_{j=1}^{2N} \alpha(\eta_j) \cdot [v(\eta_j) - v(\eta_{j-1})] \]

it follows that for every \(y \in X\), \((z_N, y) \to \langle \Phi, y \rangle\) as \(N \to \infty\); that is, \(z_N\) converges weakly to \(\Phi\) (which we denote by \(z_N \rightharpoonup \Phi\)). Let \(N, k \in \mathbb{Z}^+\). Noting that

\[ z_{N+k} = \sum_{j=1}^{2N} \sum_{m=1}^{2k} \alpha(\eta_{j-1} + \frac{m}{2^{N+k}}) \cdot \left [ v \left ( \eta_{j-1} + \frac{m}{2^{N+k}} \right ) - v \left ( \eta_{j-1} + \frac{m-1}{2^{N+k}} \right ) \right ] \]
2N 2k
ZN = \sum_{j=1}^{2N} \alpha(\eta_j) \cdot \sum_{m=1}^{2^k} \left[ v(\eta_{j-1} + \frac{m}{2^{N+k}}) - v(\eta_{j-1} + \frac{m-1}{2^{N+k}}) \right]

it is easily shown that

\|z_{N+k} - z_N\| \leq C_N \cdot TV(v)

where TV(v) is the total variation of v(\eta) and

C_N = \max \left( \frac{\max_{j=1,\ldots,2^N, \eta \in [0,T]} \| \alpha(\eta_j) - \alpha(\eta_{j-1} + \frac{\eta_j}{2^N}) \|} \right)

Since \alpha(\eta) is continuous on the compact subset [0, T], it is uniformly continuous on [0, T]; that is, for all \epsilon > 0, there is \bar{N} \in \mathbb{Z}^+ such that C_N < \frac{\epsilon}{TV(v)} for all \eta \geq \bar{N} and k \in \mathbb{Z}^+. Hence, \|z_{N+k} - z_N\| < \epsilon for all \eta \geq \bar{N} and k \in \mathbb{Z}^+. It follows that \{z_N\}^[\infty] is a Cauchy sequence and thus converges; that is,

z_N \rightarrow \int_0^T a(\eta) \, dv(\eta) \in H \text{ as } N \rightarrow \infty

Since strong convergence implies weak convergence, it follows from the uniqueness of weak limits that

\Phi = \int_0^T a(\eta) \, dv(\eta)

For every non-decreasing v \in BV(0, T), the dual functional \varphi(v) is defined as follows:

\varphi(v) = \inf_{x \in \Omega} L(x, v) \quad (2.68)

In fact, for every non-decreasing v \in BV(0, T) the infimum in (2.68) is achieved by x(v) \in \Omega; (that is, the infimum in (2.68) can actually be replaced by a minimum).

In the next Theorem, we strengthen the results of Theorem 2.2. In order to do so however, we need to strengthen our assumptions. So far, we have only assumed the existence of a feasible point x_1 \in \Omega \ (Assumption 2.7). This is the necessary and sufficient condition to guarantee existence of a unique optimal solution, as proven in Theorem 2.10. However, for Theorem 2.2 to hold, we need to make the following assumption:

**Assumption 2.8** For every v \in BV^N(0, T) such that v \geq \theta and v \neq \theta, there exists x \in \Omega such that

\sum_{i=1}^{N} \int_0^T (\langle a_i(\eta), x \rangle - c_i(\eta)) \, dv_i(\eta) < 0
2.3. LQG CONTROL WITH INFINITELY MANY LINEAR CONSTRAINTS

Remark 2.2 This assumption is automatically satisfied if there exists \( x_1 \in \Omega \) such that \( (a_i(\eta), x_1) < c_i(\eta) \) for \( i = 1, \ldots, N \) and \( \eta \in [0, T] \). Such an \( x_1 \) is said to be strictly feasible for the problem (2.1) with constraints given by (2.62).

Theorem 2.11 Suppose that Assumption 2.2 and the conditions stated in Theorem 2.10 are true. Then there exists a unique optimal solution \( x^* \) of (2.1). For every non-decreasing \( v \in BV^N(0, T) \), let \( \varphi(v) \) be given by (2.68) and \( x(v) \) be defined by

\[
x(v) = \arg \min_{x \in \Omega} L(x, v)
\]

Then \( x(v) \) exists. Moreover, the Lagrange multiplier \( v^* \in BV^N(0, T) \) defined by

\[
v^* = \arg \max_{v \geq \Omega} \min_{x \in \Omega} L(x, v) = \arg \max_{v \geq \Omega} \varphi(\lambda)
\]

exists and the optimal solution is given by \( x^* = x(v^*) \).

Proof: Let \( v \in BV^N(0, T) \), \( v \geq \Omega \) be given. From Lemma 2.2, \( \int_0^T a_i(\eta) \, dv_i(\eta) \in X \). Therefore, it follows from (2.64) that

\[
L(x, v) = \frac{1}{2} (Q x, x) + (a(v), x) - \sum_{i=1}^N \int_0^T c_i(\eta) \, dv_i(\eta)
\]

where \( a(v) = a + \sum_{i=1}^N \int_0^T a_i(\eta) \, dv_i(\eta) \in X \). Since \( \int_0^T c_i(\eta) \, dv_i(\eta) \in \mathbb{R} \) is a constant, it follows from (2.71) that the optimization problem (2.68) is exactly the same as that in (2.9). Therefore, it follows from Theorem 2.4 that \( x(v) \) exists for every non-decreasing \( v \in BV(0, T) \). The existence of \( x^* \) is proven in Theorem 2.10, and the remaining results follow from Theorem 2.2.

2.3.2 Deterministic case

Consider once again the deterministic linear system:

\[
x(t) = A(t)x(t) + B(t)u(t), \quad x(0) = \xi
\]

We assume that the assumptions stated for (2.18) hold. As in Section 2.2.2, we take the class of feasible controls to be \( \mathcal{U} = L^2(0, T; \mathbb{R}^m) \). Given any \( u \in \mathcal{U} \), suppose that \( x \in L^2(0, T; \mathbb{R}^n) \) is the corresponding solution of (2.72). We shall refer to the pair...
(x, u) as an admissible pair. The set of admissible pairs for (2.72) is given by the closed affine subspace \( z + \mathcal{Y} \) of \( L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m) \), where

\[
\mathcal{Y} = \{ (x, u) \in L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m) : \dot{x}(t) = A(t)x(t) + B(t)u(t), \ x(0) = 0 \}
\]

and \( z = (\tilde{z}, 0) \in L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m) \) where \( \tilde{z} \) is the solution of

\[
\tilde{z}(t) = A(t)\tilde{z}(t); \quad \tilde{z}(0) = \xi
\]

The cost functional \( f : L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m) \to \mathbb{R} \) is given by

\[
f(x, u) = \frac{1}{2} \int_0^T [x'(t)Q(t)x(t) + u'(t)R(t)u(t)] dt + \frac{1}{2} x'(T)Hx(T) \tag{2.73}
\]

where \( Q(t), H \in \mathbb{R}^{n \times n}, R(t) \in \mathbb{R}^{m \times m}, Q(t) \geq 0, R(t) > 0 \) for every \( t \in [0, T] \) and \( Q(t), R(t) \) are Borel measurable and continuous with respect to \( t \in [0, T] \).

We define the constraint functionals in the following way: For every fixed \( \eta \in [0, T], a_i(\cdot, \eta) \) is an \( \mathbb{R}^n \)-valued square integrable measurable functional; that is, for every \( \eta \) we have \( a_i(\cdot, \eta) \in L^2(0, T; \mathbb{R}^n) \). Furthermore, we assume that \( a_i(\cdot, \eta) \) is measurable and continuous with respect to \( \eta \in [0, T] \); that is

\[
\int_0^T (a_i(t, \eta_1) - a_i(t, \eta_0))' \cdot (a_i(t, \eta_1) - a_i(t, \eta_0)) dt \to 0
\]

as \( \eta_1 \to \eta_0 \). Similarly, for every \( \eta \in [0, T] \), we assume that \( b_i(\cdot, \eta) \in L^2(0, T; \mathbb{R}^m) \) and is a measurable function that is continuous with respect to \( \eta \in [0, T] \). Also, we assume that \( h_i(\cdot) \) is an \( \mathbb{R}^n \)-valued, measurable, continuous function of \( \eta \in [0, T] \); that is, \( h_i(\eta) \in \mathbb{R}^n \) for every \( \eta \in [0, T] \) and is measurable, and continuous with respect to \( \eta \). The \( N \) constraint functionals \( f_i : L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m) \to C(0, T; \mathbb{R}) \) are given by

\[
f_i(x, u) = \int_0^T [a_i'(t, \eta) x(t) + b_i'(t, \eta) u(t)] dt + h_i'(\eta) x(T) \tag{2.74}
\]

Note in particular that for any given \( (x, u) \in L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m), f_i(x, u) \) is a continuous functional of \( \eta \in [0, T] \); that is, \( f_i(x, u) \in C(0, T; \mathbb{R}) \). Let \( c_i(\eta) \in C(0, T; \mathbb{R}), i = 1, \ldots, N \) be given a priori. The inequality constraint \( f_i(x, u) \leq c_i \) is equivalent to \( f_i(x, u)(\eta) \leq c_i(\eta) \), for every \( \eta \in [0, T] \). With this understanding, the deterministic LQ problem subject to infinitely many linear integral constraints can be stated in the following way.
2.3. LQG CONTROL WITH INFINITELY MANY LINEAR CONSTRAINTS

**Problem:** Find the optimal control policy \( u^* \in U \) such that \((x^*, u^*)\) is an admissible pair which minimizes \( f(x, u) \) and satisfies \( f_i(x, u) \leq c_i \) \((i = 1, \cdots, N)\) where \( c_i \in C(0, T) \) are given a priori.

Equivalently, this can be stated in the following way.

\[
\begin{align*}
\min & \ f(x, u) \\
\text{s.t.} & \ f_i(x, u) \leq c_i, \quad i = 1, \cdots, N \\
& \ (x, u) \in x + Y
\end{align*}
\]

We shall make the following assumptions for (2.75):

**Assumption 2.9** For every \( v \in BV^N(0, T) \) such that \( v \geq \theta \) and \( v \neq \theta \), there exists an admissible pair \((x, u)\) for (2.72) such that

\[
\sum_{i=1}^{N} \int_0^T (f_i(x, u)(\eta) - c_i(\eta)) \, dv_i(\eta) < 0
\]

Note that this will be satisfied if there exists an admissible pair \((x, u)\) such that \( f_i(x, u) < c_i \) for \( i = 1, \cdots, N \). Such an admissible pair is referred to as a strictly feasible admissible pair. Under Assumption 2.9, it follows that Theorem 2.11 applies. This guarantees the existence of a unique optimal admissible pair for (2.75) that can be found by solving the associated dual problem.

As in the case of finitely many linear constraints, the constraints defined by the functionals (2.74) include important classes of constraints as special cases. For example, we can express pointwise state constraints of the form \( x(t) \leq k(t) \) in the general framework arising from (2.74) in the following way: Note that the differential equation (2.72) is equivalent to

\[
x(\eta) - x(0) = \int_0^\eta (A(t) x(t) + B(t) u(t)) \, dt
\]

for every \( \eta \in [0, T] \). For every given \( \eta \in [0, T] \), let \( A_i(t) \) denote the \( i^{th} \) row of \( A(t) \), and \( B_i(t) \) the \( i^{th} \) row of \( B(t) \) and consider the following functions \( a_i(\cdot, \eta) \) and \( b_i(\cdot, \eta) \) of \( t \):

\[
a'_i(t, \eta) = \begin{cases} 
A_i(t), & 0 \leq t \leq \eta \\
0, & t > \eta
\end{cases}
\]

\[
b'_i(t, \eta) = \begin{cases} 
B_i(t), & 0 \leq t \leq \eta \\
0, & t > \eta
\end{cases}
\]
It follows that
\[ x_i(\eta) = \xi_i + \int_0^\eta (A_i(t) x(t) + B_i(t) u(t)) dt \]
and the constraint \( x_i(\eta) \leq k_i(\eta), \eta \in [0,T] \) can be written as
\[ \int_0^T (a_i(t, \eta)' x(t) + b_i(t, \eta)' u(t)) dt \leq c_i(\eta) \]
where \( c_i(\eta) = k_i - (\eta) \xi \). That is, linear state inequality constraints are a special case of the constraints in (2.75). If we choose \( h_i(\eta) = 0 \) and \( c(\eta) = k(\eta) - \xi \), then the linear inequality constraints in (2.75) represent \( x(\eta) \leq c(\eta) \). Note also that for every \( \eta \in [0,T] \), \( a_i(\cdot, \eta) \in L^2(0,T;\mathbb{R}^n) \), \( b_i(\cdot, \eta) \in L^2(0,T;\mathbb{R}^m) \) and both \( a_i(\cdot, \eta) \) and \( b_i(\cdot, \eta) \) are continuous \( L^2(0,T;\mathbb{R}^n) \) and \( L^2(0,T;\mathbb{R}^m) \) valued functions of \( \eta \in [0,T] \), respectively. This is easily seen by observing that for any \( \eta_0, \eta_1 \in [0,T] \), we have
\[ \|a_i(\cdot, \eta_1) - a_i(\cdot, \eta_0)\|^2 = \int_0^T (a_i(t, \eta_1) - a_i(t, \eta_0))' (a_i(t, \eta_1) - a_i(t, \eta_0)) dt \]
\[ = \int_{\eta_0}^{\eta_1} a_i(t, \eta_1)' a_i(t, \eta_1) dt \to 0 \]
as \( \eta_1 \to \eta_0 \) since (as assumed) the elements of the matrix \( A(t) \) are continuous functions of \( t \). It follows that \( a_i(\cdot, \eta) \) is a continuous \( L^2(0,T;\mathbb{R}^n) \)-valued function of \( \eta \). Similarly, it can be shown that \( b_i(\cdot, \eta) \) is a continuous \( L^2(0,T;\mathbb{R}^m) \)-valued function of \( \eta \in [0,T] \).

Let \( v \in BV^N(0,T) \), such that \( v \geq \theta \) and \( \zeta = (x, u) \in z + \mathcal{V} \) be an admissible pair for (2.72). Defining the Lagrangian \( L(\zeta, v) \) as in (2.64), we obtain
\[ L(\zeta, v) = \frac{1}{2} \int_0^T [x'(t)Q(t)x(t) + u'(t)R(t)u(t)] dt + \frac{1}{2}x'(T)Hx(T) + \int_0^T (a'(t; v) x(t) + b'(t; v) u(t)) dt + h'(v) x(T) \]
\[ - \sum_{i=1}^N \int_0^T c_i(\eta) dv_i(\eta) \]
where
\[ a(t; v) = \sum_{i=1}^N \int_0^T a_i(t, \eta) dv_i(\eta) \]
\[ b(t; v) = \sum_{i=1}^N \int_0^T b_i(t, \eta) dv_i(\eta) \]
2.3. LQG CONTROL WITH INFINITELY MANY LINEAR CONSTRAINTS

\[ h(v) = \sum_{i=1}^{N} \int_{0}^{T} h_i(\eta) \, dv_i(\eta) \]

Given any non-decreasing \( v \in BV(0,T) \), the following result gives us the value of the dual cost functional \( \varphi(\lambda) \) evaluated at \( v \), and the solution \((x(v), u(v))\) of the optimization problem associated with calculating \( \varphi(v) \), as defined by (2.68).

**Proposition 2.3** Let \( v \in BV(0,T) \) be a non-decreasing function of \( \eta \in [0,T] \). Let \( \zeta(v) = x((v), u(v)) \) denote the optimal state-control pair for the unconstrained deterministic LQ problem

\[
\begin{bmatrix}
\min_{\zeta} L(\zeta, v) \\
\zeta \in x + Y
\end{bmatrix}
\] (2.79)

Then the optimal control \( u(t, v) \) is

\[
u(t, v) = -R^{-1}(t) \left[ B'(t)P(t)x(t) + B'(t) d(t; v) + b(t; v) \right]
\] (2.80)

where

\[
\dot{P} = -PA - A'P + PB R^{-1}B'P - Q, \quad P(T) = H
\] (2.81)

\[
d = -[A - BR^{-1}B'] d - a(v) + PB R^{-1}b(v), \quad d(T, v) = h(v)
\] (2.82)

\[
\dot{p} = [B' d + b(v)]' R^{-1} [B' d + b(v)], \quad p(T, v) = 0
\] (2.83)

and \( x(v) \) is the solution of (2.72) with \( u = u(\lambda) \). The value of the dual functional is the optimal cost associated with (2.79), and is given by

\[
\varphi(v) = \frac{1}{2} \xi' P(0) \xi + d'(0; v) \xi + \frac{1}{2} p(0; v) - \sum_{i=1}^{N} \int_{0}^{T} c_i(\eta) \, dv_i(\eta)
\] (2.84)

**Proof:** Note that for any given non-decreasing \( v \in BV(0,T) \), \( L(\zeta, v) \) as given by (2.78), is a standard LQ cost functional, and (2.79) is an unconstrained LQ problem. It follows that the optimal control for (2.79) is given by (2.80) and the associated optimal cost by (2.84).

Note that \( P(t) \) is independent of \( v \), and hence, can be ignored.

Under Assumptions 2.9, it follows that Theorem 2.11 applies to (2.75). Therefore, we have the following result.
CHAPTER 2. LINEARLY CONSTRAINED LQG CONTROL

**Theorem 2.12** There exists \( v^* \geq 0 \) which is optimal for

\[
\max_{v \geq 0} \left\{ d'(0; v)\xi + \frac{1}{2} p(0; v) - \sum_{i=1}^{N} \int_{0}^{T} c_i(\eta) \, dv_i(\eta) \right\} \tag{2.85}
\]

subject to (2.81)-(2.83). Furthermore, the optimal control \( u^* \) of (2.75) exists, and is given by

\[
u^*(t) = -R^{-1}(t) \left[ B'(t) P(t) x(t) + B'(t) d(t; v^*) + b(t; v^*) \right] \tag{2.86}
\]

**Proof:** By Theorem 2.10, there exists a unique optimal control \((x^*, u^*)\) for (2.75). By Theorem 2.11, it follows that \((x^*, u^*) = (x(v^*), u(v^*))\) where \(v^*\) is the optimal solution of (2.85) and \(u(v)\) is given by (2.80). Existence of \(v^*\) is guaranteed by Theorem 2.11, and existence of \(u(v)\) by Proposition 2.3. 

From Theorem 2.12, we see that the optimal control \(u^*\) is obtained by solving the optimization problem (2.81)-(2.83), (2.85) over \(v \in BV^N(0, T)\). Several things should be noted. First, (2.81)-(2.83), (2.85) is an infinite dimensional optimization problem because the space of variables is the infinite dimensional set \(BV^N(0, T)\). However, one crucial difference between this (infinite dimensional) dual problem and the original (infinite dimensional) control problem is that the dual problem is an open-loop optimization problem since the dual variable \(v(t)\) needs only be a function of time as opposed to the control problem where the optimal control \(u^*(t)\) needs also to be a function of the current state. To solve problems of the form (2.81)-(2.83), (2.85), one approach is to make a finite dimensional approximation - one that can be solved using optimization algorithms for solving finite dimensional optimization problems - and a sequence of sub-optimal solutions \(u_j^*\) is obtained by solving ‘better’ approximations of the original infinite dimensional problem. In this way, an approximate solution to the dual problem (2.81)-(2.83), (2.85) can be obtained. A finite dimensional approximation the infinite dimensional problem (2.81)-(2.83), (2.85) is obtained by approximating the integrals

\[
\int_{0}^{T} a_i(t, \eta) \, dv_i(\eta) \quad \int_{0}^{T} b_i(t, \eta) \, dv_i(\eta) \quad \int_{0}^{T} h_i(\eta) \, dv_i(\eta)
\]

which appear in (2.81)-(2.83), (2.85) in the following way. Recall from Lemma 2.2 that

\[
\int_{0}^{T} a_i(t, \eta) \, dv_i(\eta) = \lim_{N \to \infty} \sum_{j=1}^{K-1} a_i(t, \eta_j) \left( v_i(\eta_{j+1}) - v_i(\eta_j) \right)
\]
2.3. LQG CONTROL WITH INFINITELY MANY LINEAR CONSTRAINTS

By replacing \( \int_0^T a_i(t, \eta) \, dv_i(\eta) \) with \( \sum_{j=1}^{K} a_i(t, \eta_j) \lambda_{ij} \) where \( \lambda_{ij} = v_i(\eta_j) - v_i(\eta_j-1) \), \( j = 1, \cdots, N \) the problem (2.81)-(2.83), (2.85) over \( v(t) \) becomes a finite dimensional problem over \( \lambda_{ij} \in \mathbb{R} \) for \( i = 1, \cdots, N \) and \( j = 1, \cdots, K \). Making this approximation for \( \int_0^T a_i(t, \eta) \, dv_i(\eta), \int_0^T b_i(t, \eta) \, dv_i(\eta) \) and \( \int_0^T h_i(\eta) \, dv_i(\eta) \), we see that (2.81)-(2.83), (2.85) is approximated by a problem of the form (2.26)-(2.28), (2.30) and can be solved as outlined at the end of Section 2.2.2. Note that solving this approximation of (2.81)-(2.83), (2.85) is equivalent to solving the dual problem which corresponds to approximating the constraints (2.75) by the finite subset

\[
E \left[ \int_0^T \left( a'_i(t, \eta_j) x(t) + b'_i(t, \eta_j) u(t) \right) \, dt + h'_i(t, \eta_j) x(T) \right] \leq c_i(\eta_j)
\]

for \( i = 1, \cdots, N, \ j = 1, \cdots, K \).

2.3.3 Full observation case

We consider the linear system

\[
\frac{dx(t)}{dt} = (A(t)x(t) + B(t)u(t)) \, dt + C(t)dW(t), \quad x(0) = \xi
\]

(2.87)

under the same assumptions as in Section 2.2.3. In particular, let \( \mathcal{X} \) be defined as in (2.39) and the class of admissible controls \( \mathcal{U} \) by (2.40). As shown in Section 2.2.3, the set of all admissible pairs \( (x, u) \) for (2.87) is a closed affine subspace \( z + \mathcal{Y} \) of \( \mathcal{X} \times \mathcal{U} \), where

\[
\mathcal{Y} = \{(x, u) \in \mathcal{X} \times \mathcal{U} : \frac{dx(t)}{dt} = (A(t)x(t) + B(t)u(t)) \, dt, \ x(0) = 0 \}
\]

and \( z = (\bar{z}, 0) \in \mathcal{X} \times \mathcal{U} \) with \( \bar{z} \) being the solution of

\[
\frac{d\bar{z}(t)}{dt} = A(t)\bar{z}(t) \, dt + C(t)dW(t), \quad \bar{z}(0) = \xi
\]

The cost functional \( f : \mathcal{X} \times \mathcal{U} \to \mathbb{R} \) is given by

\[
f(x, u) = \frac{1}{2} E \left[ \int_0^T \left[ x'(t)Q(t)x(t) + u'(t)R(t)u(t) \right] \, dt + x'(T)Hx(T) \right]
\]

(2.88)

where \( Q(t), R(t) \) and \( H \) satisfy the same conditions as the cost functional (2.73). The constraints \( f_i : \mathcal{X} \times \mathcal{U} \to \mathbb{R}, \ i = 1, \cdots, N \) are given by

\[
f_i(x, u) = E \left[ \int_0^T \left[ a'_i(t, \eta) \, x(t) + b'_i(t, \eta) \, u(t) \right] \, dt + h'_i(\eta) \, x(T) \right]
\]

(2.89)
where \( a_i(t, \eta), b_i(t, \eta), \) and \( h_i(\eta) \) satisfy the same conditions as (2.74). The full observation LQG problem with infinitely many linear constraints can be stated as follows

**Problem:** Find the optimal control policy \( u^* \in \mathcal{U} \) which minimizes \( f(x, u) \) such that \( (x, u) \) satisfies the linear system (2.87) and the constraints \( f_i(x, u) \leq c_i \) (\( i = 1, \cdots, N \)), where \( c_i \in C(0, T; \mathbb{R}) \) are given a priori.

As before, \( f_i(x, u) \in C(0, T; \mathbb{R}) \) and the inequality constraints \( f_i(x, u) \leq c_i \) means that \( f_i(x, u)(\eta) \leq c_i(\eta) \) for every \( \eta \in [0, T] \).

Equivalently, this problem can be expressed in the form (2.1); that is

\[
\begin{align*}
\min f(x, u) \\
\text{subject to: } f_i(x, u) \leq c_i, \quad i = 1, \cdots, N \\
(x, u) \in z + \mathcal{Y}
\end{align*}
\]

Parallel to the case of finitely many linear constraints, the constraints defined by the functionals (2.89) include pointwise linear inequality constraints of the form \( Ex(t) \leq k(t), t \in [0, T] \) as a special case. This follows by noting that the stochastic differential equation (2.87) can be written as

\[
x(t) - \xi = \int_0^t (A(s)x(s) + B(s)u(s)) \, ds + \int_0^t C(s) \, dW(s)
\]

for every \( t \in [0, T] \). Taking expectations on both sides gives

\[
Ex(t) = \xi + E \int_0^t (A(s)x(s) + B(s)u(s)) \, ds
\]

and picking \( a_i(t, \eta), b_i(t, \eta) \) as in (2.76)-(2.77), \( h_i = 0 \) and \( c(t) = k(t) - \xi \), we see that pointwise constraints on the expected value of the state can be represented by the functional (2.89). For each \( v \in BV^N(0, T) \) such that \( v \geq \theta \), we define the Lagrangian \( L((x, u), v) \) by

\[
L((x, u), v) = f(x, u) + \sum_{i=1}^N \int_0^T (f_i(x, u)(\eta) - c_i(\eta)) \, dv_i(\eta)
\]

\[
= \frac{1}{2} E \left[ \int_0^T \left[ x'(t)Q(t)x(t) + u'(t)R(t)u(t) \right] dt + x'(T)Hx(T) \\
+ \int_0^T \left[ a'(t; v)x(t) + b'(t; v)u(t) \right] dt + h'(v)x(T) \right] \\
- \sum_{i=1}^N \int_0^T c_i(\eta) \, dv_i(\eta)
\]
As in the deterministic case, \( a(t; v), b(t; v) \) and \( h(v) \) are given by

\[
\begin{align*}
    a(t; v) &= \sum_{i=1}^{N} \int_{0}^{T} a_i(t, \eta) \, dv_i(\eta) \\
    b(t; v) &= \sum_{i=1}^{N} \int_{0}^{T} b_i(t, \eta) \, dv_i(\eta) \\
    h(v) &= \sum_{i=1}^{N} \int_{0}^{T} h_i(\eta) \, dv_i(\eta)
\end{align*}
\]

respectively. The following result is a direct consequence of the optimal solution for the unconstrained LQG control problem. It gives the value of the dual functional \( \varphi \) associated with (2.90), for any \( v \in BV^N(0, T) \).

**Proposition 2.4** Let \( v \in BV^N(0, T) \), \( v \geq \theta \) be given. Let \( \zeta(v) = x((v), u(v)) \) denote the optimal state-control pair for the problem

\[
\begin{align*}
    \min_{\zeta} L(\zeta, v) \\
    \zeta \in z + Y
\end{align*}
\]

(2.92)

Then the optimal control \( u(t, v) \) is

\[
u(t, v) = -R^{-1}(t) \left[ B'(t) P(t) x(t) + B'(t) d(t; v) + b(t; v) \right]
\]

(2.93)

where \( x(v) \) is the solution of (2.87) with \( u = u(\lambda) \). The value of the dual function \( \varphi(\lambda) \) associated with (2.92), is given by optimal cost is

\[
\varphi(v) = \frac{1}{2} \zeta' P(0) \zeta + d'(0; v) \zeta + \frac{1}{2} p(0; v) + \frac{1}{2} \int_{0}^{T} \text{tr} \left\{ C'(t) P(t) C(t) \right\} dt - \sum_{i=1}^{N} \int_{0}^{T} c_i(\eta) \, dv_i(\eta)
\]

(2.94)

where \( P(t), d(t; v) \) and \( p(t; v) \) are given by (2.81)-(2.83) respectively.

Under the following assumption, we can apply Theorem 2.11 to obtain the optimal control

**Assumption 2.10** For every \( v \in BV^N(0, T) \) such that \( v \geq \theta \) and \( v \neq \theta \), there exists an admissible pair \((x, u)\) for (2.87) such that

\[
\sum_{i=1}^{N} \int_{0}^{T} (f_i(x, u)(\eta) - c_i(\eta)) \, dv_i(\eta) < 0
\]
Theorem 2.13 There exists a unique optimal control $u^* \in \mathcal{U}$ for the full observation LQG optimal control with infinitely many constraints (2.90). Moreover,

$$u^*(t) = -R^{-1}(t) \left[ B'(t)P(t)x(t) + B'(t)d(t; v^*) + b(t; v^*) \right]$$

(2.95)

where $v^* \in BV^N(0,T)$ is the optimal solution of the optimization problem defined by (2.81)-(2.83), (2.85).

Note in particular that certainty equivalence holds; that is, by solving the deterministic problem, the optimal control for the full observation LQG problem is obtained by replacing the deterministic state with the output of the stochastic differential equation (2.87).

2.3.4 Partial observation case

Consider the stochastic linear system with output:

\[
\begin{align*}
\dot{x}(t) &= (A(t)x(t) + B(t)u(t))dt + C(t)dW(t), \quad x(0) \sim N(\xi, \Sigma_0), \\
\dot{y}(t) &= F(t)x(t)dt + G(t)dV(t), \quad y(0) = 0.
\end{align*}
\]

(2.96) (2.97)

We assume that the conditions, as stated for (2.51)-(2.52) hold for (2.96)-(2.97). In particular, we shall take the class of feasible controls to be the set $\mathcal{U}$ as defined in (2.53). Given $u \in \mathcal{U}$, let $x$ be the solution of the stochastic differential equation (2.96). The cost functional is given by

\[
f(x, u) = \frac{1}{2} E \left[ \int_0^T \left[ x'(t)Q(t)x(t) + u'(t)R(t)u(t) \right] dt + x'(T)Hx(T) \right]
\]

(2.98)

where $Q(t)$, $R(t)$ and $H$ satisfy the same conditions as the cost functional (2.73) for the deterministic (and hence, the full observation) problem. The $N$ constraint functionals are given by

\[
f_i(x, u) = E \left[ \int_0^T a'_i(t, \eta) x(t) + b'_i(t, \eta) u(t) \right] dt + h'_i(\eta) x(T)
\]

(2.99)

where $a_i(t, \eta)$, $b_i(t, \eta)$ and $h_i(\eta)$ satisfy the same conditions as in the deterministic (and hence, full observation) case (2.74). In particular, given and admissible pair $(x, u)$, it follows that $f_i(x, u) \in C(0,T; \mathbb{R})$. The partially observed LQG problem with infinitely many linear constraints can be stated as follows:
2.3. LQG CONTROL WITH INFINITELY MANY LINEAR CONSTRAINTS

**Problem:** Find the optimal control policy \( u^* \in \bar{U} \) which minimizes \( f(x, u) \) such that \((x, u)\) satisfies the linear system (2.96) and the constraints \( f_i(x, u) \leq c_i \) \((i = 1, \ldots, N)\) where \( c_i \in C(0, T) \) are given a priori.

Equivalently, this can be written:

\[
\begin{align*}
\min f(x, u) \\
f_i(x, u) \leq c_i, & \quad i = 1, \ldots, N \\
u \in \bar{U} \text{ and } (x, u) \text{ satisfies (2.96).}
\end{align*}
\]  

(2.100)

As before, we solve this problem by transforming it into a full observation problem. It turns out that this full observation problem can be solved using the results obtained in Section 2.3.3, and the optimal control is determined in this way.

In order to use the results in Section 2.3.3 to solve (2.100), we need to make the following assumption. It is required for Theorem 2.13 to be applicable to the equivalent full observation problem.

**Assumption 2.11** For every \( v \in BV^N(0, T) \) such that \( v \geq \theta \) and \( v \neq \theta \), there exists an admissible control \( u \in \bar{U} \) such that the admissible pair \((x, u)\) for (2.96)-(2.97) such that

\[
\sum_{i=1}^{N} \int_0^T (f_i(x, u)(\eta) - c_i(\eta)) \, dv_i(\eta) < 0
\]

We solve (2.100) by transforming it into an equivalent full observation problem in the standard way. First, it can be shown that the cost functional \( f(x, u) \) is equivalent to

\[
f(x, u) = \frac{1}{2} E \left[ \int_0^T \left[ \dot{z}'(t)Q(t)\dot{z}(t) + u'(t)R(t)u(t) \right] dt + \dot{z}'(T) H \dot{z}(T) \right]
\]

\[
+ \frac{1}{2} \int_0^T \text{tr} \{Q(t)\Sigma(t)\} dt + \frac{1}{2} \text{tr} \{H \Sigma(T)\}
\]

\[
= f(\hat{x}, u) + \frac{1}{2} \int_0^T \text{tr} \{Q(t)\Sigma(t)\} dt + \frac{1}{2} \text{tr} \{H \Sigma(T)\}
\]

where \( \hat{x} \) is given by the Kalman filter associated with (2.96)-(2.97):

\[
d\hat{x}(t) = (A(t)\hat{x}(t) + B(t)u(t))dt - \Sigma(t)F(t) (G(t)G'(t))^{-1} dv(t), \quad \hat{x}(0) = \xi \]  

(2.101)

\[
\dot{\Sigma} = \Sigma A + A'\Sigma - \Sigma F' (GG')^{-1} F \Sigma + CC', \quad \Sigma(0) = \Sigma_0
\]  

(2.102)
Relevant properties of (2.101)-(2.102) are stated in Lemma 2.1. Note in particular that

\[ \frac{1}{2} \int_0^T \text{tr} \{ Q(t)\Sigma(t) \} \, dt + \frac{1}{2} \text{tr} \{ H\Sigma(T) \} \]

is a constant that is independent of \( u \), and hence, this term in the cost functional can be ignored. Similarly, it can be shown that

\[ f_i(x, u) = E \left[ \int_0^T \left[ a'_i(t, \eta) \dot{z}(t) + b'_i(t, \eta) u(t) \right] \, dt + h'_i(\eta) \dot{z}(T) \right] = f_i(\hat{x}, u) \]

In particular, \( f_i(\hat{x}, u) \) are of the form (2.89) and satisfy the conditions stated for these constraint functionals. Therefore, the partially observed problem (2.100) is equivalent to the full observation problem

\[
\begin{align*}
\min \ f(\hat{x}, u) \\
\text{subject to: } f_i(\hat{x}, u) \leq c_i, \ i = 1, \ldots, N \\
(\hat{x}, u) \text{ satisfies (2.101) and } u \in \bar{U}
\end{align*}
\]

(2.103)

Using similar arguments to the ones presented in the proof of Theorem 2.9, it can be shown by virtue of Theorem 2.13 for the full observation LQG problem with infinitely many constraints, we have the following result:

**Theorem 2.14 (Separation Theorem)**  The optimal control \( u^* \in \bar{U} \) for (2.100) is given by

\[
u^*(t) = -R^{-1}(t) \left[ B'(t)P(t)\hat{x}(t) + B'(t)\dot{d}(v^*, t) + b(v^*, t) \right]
\]

for the partially observed LQG problem with infinitely many constraints, where \( \hat{x} \) is the output of the Kalman filter (2.57) and \( v^* \) the solution of the optimization problem (2.81)-(2.83), (2.85).

It is important to note that the optimal control (2.104) is determined by solving a ‘control problem’ represented by (2.81)-(2.83), (2.85) and the Kalman Filtering problem (2.57). Moreover, these two problems can be solved independently and hence, the Separation Theorem holds. Furthermore, certainty equivalence also holds since the ‘control problem’ (2.81)-(2.83), (2.85) is the same for both the partial observation and deterministic problems.
2.4 Conclusion

In this chapter, we have studied the LQG control problem subject to finitely many, and infinitely many integral linear constraints. In particular, we have derived the optimal control for these problems and proven the Separation Theorem holds. In the case of finitely many integral linear constraints, we showed that the optimal control for the deterministic problem is determined by solving a finite dimensional (open-loop) optimization problem. In the case of deterministic LQ subject to infinitely many linear integral constraints, the optimal control is determined by solving an (open-loop) infinite programming problem. These equivalent finite and infinite dimensional optimization problems are derived using duality theory. In the partially observed case, we proved that the Separation Theorem holds for both problems. Moreover, the optimal control is determined by replacing the state in the deterministic optimal control by the optimal state estimate given by the Kalman filter; that is, certainty equivalence holds.

We feel that an interesting and exciting area for future research is the application of these results to constrained nonlinear optimal control problems in the context of sequential quadratic programming (SQP) [17, 39, 44, 52]. Such a generalization would involve extending SQP to an infinite dimensional setting. SQP is one of the most efficient methods for solving finite dimensional constrained nonlinear optimization problems. The basic idea behind this scheme is solving a sequence of quadratic optimization problems where the cost is a quadratic approximation of the nonlinear cost, and constraints are linear approximations of the nonlinear constraints. Under certain conditions, quadratic convergence of this algorithm near the optimal solution can be proven. Clearly, the SQP algorithm strongly relies on being able to solve the linearly constrained quadratic optimization problems efficiently. In the context of nonlinear constrained optimal control, quadratic approximations result in linearly constrained LQG control problems of the form studied in this Chapter. In making such a generalization of SQP, one interesting issue is determining the conditions under which quadratic convergence near the optimal solution can be guaranteed. Note once again that the optimal control for the linearly constrained LQG problem is in the form of linear state feedback (with an unavoidable nonlinear dependence on the initial condition). It follows immediately that the (sub)optimal control obtained by the SQP algorithm will be in this elegant form as well. This has the advantage that such controllers are easy to implement.
CHAPTER 2. LINEARLY CONSTRAINED LQG CONTROL

\[ \text{Equation} \]

\[ \text{Example} \]

\[ \text{Theorem} \]

\[ \text{Proof} \]

\[ \text{Conclusion} \]
Chapter 3

LQG control with IQ constraints

In this chapter, we consider the deterministic, the full observation and the partial observation LQG optimal control problems subject to integral quadratic constraints. In the unconstrained case, Wohnam's Separation Theorem [55] is a fundamental result. It states that optimal control for the partially observed LQG problem is obtained by solving both a 'filtering' and a 'control' problem, and that these two problems can be solved separately. In Chapter 2, we have shown that a Separation Theorem holds in the LQG problem with finitely many and infinitely many linear integral constraints. In this chapter, we consider the case of integral quadratic constraints.

We show that unlike the unconstrained or the linearly constrained cases, the Separation Theorem does not hold for the LQG problem subject to integral quadratic constraints. Although the optimal control is calculated by solving a control problem and a filtering problem, the control and filtering problems can not be solved separately - the solution of the control problem is dependent on the solution of the filtering Riccati equation. However, this dependence adds no complication to the control or the filtering problems. It is in this context that the label Quasi-Separation Theorem should be understood.

In the literature, it has been said that the use of multiple-objective LQG (and in particular, LQG control subject to integral quadratic constraints) in practise has been limited because it is computationally demanding. We conclude this chapter by examining computational issues. Like the case of finitely many linear constraints, we show that the optimal control can be calculated by solving an optimal parameter selection problem [46]. Optimal parameter selection problems can be solved using standard
gradient-type optimization algorithms so long as the gradient of the cost functional can be calculated. We show how this gradient can be determined. Indeed, it is an easy problem. In the deterministic case, it is equivalent to solving an unconstrained LQ optimal control problem. In the full observation and partial observation cases, it is equivalent to solving a full observation, and a partial observation LQG problem respectively. Thus, the gradient can be easily calculated and efficient optimization algorithms can be used to solve the optimal parameter selection problem which in turn, gives the optimal control. It is appropriate to mention here that the software package MISER 3.1 [35] is designed to solve optimization problems of this type.

Thus a brief outline of this chapter is as follows. In Section 3.1, we examine the deterministic LQ problem with IQ constraints and derive an expression for the optimal control. In Section 3.2, we consider the full observation case. In particular, we show that (unlike the unconstrained and linearly constrained cases) certainty equivalence does not hold. In Section 3.3, we solve the partially observed problem. In particular, we prove that the Separation Theorem does not hold, but a result that we call a Quasi-separation Theorem does. We end by examining computational issues in Section 3.4. In particular, we show that solving the LQG problem with IQ constraints can be viewed as solving a sequence of unconstrained LQG problems.

Before presenting our results, we end with a reminder that throughout this chapter, we are making standard assumptions for the LQG cost and constraint functionals. In particular, we are assuming that the control weighting matrix is strictly positive definite in the cost functional, and positive semidefinite in the constraints. In Chapter 4, we relax these assumptions and consider problems with indefinite control weighting matrices in the cost and constraints.

These results can be found in the papers by Lim and Moore [30] and Lim, Liu, Teo and Moore [27].

### 3.1 Deterministic Case

In this section, we consider the deterministic LQG control problem with integral quadratic constraints. Assume that \( T < \infty \). Let \( L^2(0, T; \mathbb{R}^n) \) denote the Hilbert space of \( \mathbb{R}^n \)-valued, measurable, square integrable functions on \([0, T]\), as defined by (2.16) in Section 2.2.2. As in Section 2.2.2, we take \( \mathcal{U} = L^2(0, T; \mathbb{R}^m) \) to be the class
3.1. DETERMINISTIC CASE

of feasible controls. For every $u \in \mathcal{U},$ define $x \in L^2(0, T; \mathbb{R}^n)$ as the solution of the linear system

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \quad x(0) = \xi \quad (3.1)$$

where $A(t), B(t)$ and $\xi \in \mathbb{R}^n$ satisfy the conditions as stated in Section 2.2.2. In such a case, $(x, u)$ is referred to as an admissible pair.

Let $z = (\bar{x}, 0) \in L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m)$ where $\bar{x}$ is the solution of the differential equation

$$\dot{\bar{x}}(t) = A(t)\bar{x}(t); \quad \bar{x}(0) = \xi \quad (3.2)$$

Define the set

$$\mathcal{Y} = \{(x, u) \in L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m) : \dot{x}(t) = A(t)x(t) + B(t)u(t), x(0) = 0\} \quad (3.3)$$

It follows that $z + \mathcal{Y}$ is a closed subspace, and consists of all the solutions of the linear system (3.1).

Define the cost ($i = 0$) and constraint ($i = 1, \cdots, N$) functionals $f_i : L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m) \to \mathbb{R}$ by

$$f_i(x, u) = \frac{1}{2} \int_0^T \left( x'(t)Q_i(t)x(t) + u'(t)R_i(t)u(t) \right) dt + \frac{1}{2} x'(T)H_ix(T)$$

$$+ \int_0^T \left( a'_i(t)x(t) + b'_i(t)u(t) \right) dt + h'_ix(T) \quad (3.4)$$

where $H_i, Q_i(t) \in \mathbb{R}^{n \times n}, R_i(t) \in \mathbb{R}^{m \times m}, a_i(t) \in \mathbb{R}^n, b_i(t) \in \mathbb{R}^m$ are measurable, continuous functions of $t \in [0, T]$ and $R_i(t) \geq 0 \ (i = 1, \cdots, N), Q_i(t) \geq 0, H_i \geq 0 \ (i = 0, \cdots, N)$ and $R_0(t) > 0$ for each $t \in [0, T].$ Note that this allows for the case of linear integral constraints. Let $c_i \in \mathbb{R} \ i = 1, \cdots, N$ be given constants. The deterministic LQ optimal control problem subject to integral quadratic constraints can be stated as follows:

$$\begin{align*}
& f_0(x, u) \to \text{max} \\
& f_i(x, u) \leq c_i, \quad i = 1, \cdots, N \\
& (x, u) \in z + \mathcal{Y}
\end{align*} \quad (3.5)$$

We make the following assumption:
Assumption 3.1  There exists an admissible pair \((x, u)\) which is feasible for (3.5).

We have the following result on the existence of an optimal solution \((x^*, u^*)\) for (3.5).

Theorem 3.1  Suppose that Assumption 3.1 is true. Then there exists a unique optimal solution \((x^*, u^*)\) of (3.5).

Proof:  Let \((x, u)\) be feasible for (3.5) and \(\sigma = f_0(x, u)\). Define

\[
H_\sigma = \{(x, u) \in L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m) : f_0(x, u) \leq \sigma\}
\]

and consider the problem

\[
\begin{cases}
  f_0(x, u) \to \text{max} \\
  f_0(x, u) \leq \sigma \\
  f_i(x, u) \leq c_i, \quad i = 1, \ldots, N \\
  (x, u) \in z + \mathcal{Y}
\end{cases}
\]  \hspace{1cm} (3.6)

Since \(f_0(x, u)\) is strictly convex on \(L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m)\), the constraints of (3.6) define a bounded, closed, convex subset of the Hilbert space \(L^2(0, T; \mathbb{R}^n) \times L^2(0, T; \mathbb{R}^m)\). Since every continuous, convex functional defined on a Hilbert space achieves its minimum on every bounded, closed, convex set [3, Theorem 2.6.1], it follows that there exists \((x^*, u^*)\) which is optimal for (3.6). Furthermore, the uniqueness of \((x^*, u^*)\) follows from the strict convexity of \(f_0(x, u)\). Now we show that \((x^*, u^*)\) is optimal for (3.5). Suppose that this is not true. Then there exists \((\bar{x}, \bar{u})\) which is feasible for (3.5) and \(f_0(\bar{x}, \bar{u}) < f_0(x^*, u^*) \leq \sigma\). However, this implies that \((\bar{x}, \bar{u})\) is feasible for (3.6) and hence, \(f_0(x^*, u^*) \leq f_0(\bar{x}, \bar{u})\) - a contradiction. Therefore, by the definition of \((x^*, u^*)\) we have \(f_0(x^*, u^*) \leq f_0(\bar{x}, \bar{u})\) for all feasible solutions \((\bar{x}, \bar{u})\) of (3.5). The result follows.

Since (3.5) is a convex optimization problem, it can be studied using the Lagrange Duality Theorem (Theorem 2.2). In this way, the infinite dimensional problem (3.5) over the space \(z + \mathcal{Y}\) can be converted into a related finite dimensional optimization problem over \(\mathbb{R}^N\), where \(N\) is the number of constraints. In order to do this, we need to introduce the Lagrangian for (3.5). For every \(\lambda = (\lambda_1, \cdots, \lambda_N) \geq 0\) and \(\zeta = (x, u) \in z + \mathcal{Y}\) let the Lagrangian be given by

\[
L(\zeta, \lambda) = f_0(\zeta) + \sum_{i=1}^{N} \lambda_i(f_i(\zeta) - c_i)
\]  \hspace{1cm} (3.7)
3.1. DETERMINISTIC CASE

Note that

\[ L((x, u), \lambda) = \frac{1}{2} \int_0^T (x'(t)Q(t, \lambda)x(t) + u'(t)R(t, \lambda)u(t)) \, dt + \frac{1}{2}x'(T)H(\lambda)x(T) \]

\[ + \int_0^T (a'(t, \lambda)x(t) + b'(t, \lambda)u(t)) \, dt + h'(\lambda)x(T) - \lambda'c \]

where

\[ Q(t, \lambda) = \sum_{i=1}^N \lambda_i Q_i(t) \quad H(\lambda) = \sum_{i=1}^N \lambda_i H_i \quad a(t, \lambda) = \sum_{i=1}^N \lambda_i a_i(t) \]

with a similar interpretation for \( R(t, \lambda), b(t, \lambda) \) and \( h(\lambda) \).

For any \( \lambda \geq 0 \), the dual functional \( \varphi(\lambda) \) for problem (3.5) is defined by (2.2) and stated here again for convenience:

\[ \varphi(\lambda) = \inf_{\zeta \in z + Y} L(\zeta, \lambda) \quad \text{(3.8)} \]

Recalling the definition of \( z + Y \) and \( L(\zeta, \lambda) \), it follows that for any given \( \lambda \geq 0 \), \( \varphi(\lambda) \) as given by (3.8) is calculated by solving an unconstrained LQ problem. An explicit formula for \( \varphi(\lambda) \) is given in the following proposition.

**Proposition 3.1** Let \( \lambda \geq 0 \) be given and \( \zeta = (x(\lambda), u(\lambda)) \) denote the optimal state-control pair for the problem

\[ \left\{ \begin{array}{l} \min_{\zeta} L(\zeta, \lambda) \\
\zeta \in z + Y \end{array} \right. \quad \text{(3.9)} \]

Then the optimal control \( u(\lambda) \) is

\[ u(t, \lambda) = -R^{-1}(t, \lambda) \left[ B'(t)P(t, \lambda)x(t) + B'(t)d(t, \lambda) + b(t, \lambda) \right] \quad \text{(3.10)} \]

and \( x(\lambda) \) is the solution of (3.1) with \( u(t) = u(\lambda, t) \). The optimal cost is

\[ \varphi(\lambda) = \frac{1}{2}\xi'P(0, \lambda)\xi + d'(0, \lambda)\xi + \frac{1}{2}p(0, \lambda) - \lambda'c \quad \text{(3.11)} \]

**Proof:** Note that for any given \( \lambda^* \geq 0 \), \( \lambda^*c \) is a constant. Therefore, (3.9) is a standard unconstrained LQ control problem from which the result follows.

The following assumption is required for the Lagrange Duality Theorem (Theorem 2.2) to be applied. It is commonly referred to as a Slater condition.
CHAPTER 3. LQG CONTROL WITH IQ CONSTRAINTS

Assumption 3.2 For every $\lambda_i \geq 0, i = 1, \cdots, N$ (not all equal to zero), there exists $(x, u) \in z + \mathcal{Y}$ such that

$$
\sum_{i=1}^{N} \lambda_i (f_i(x, u) - c_i) < 0
$$

Remark 3.1 A sufficient condition for Assumption 3.2 to hold is the existence of $(\bar{x}, \bar{u}) \in z + \mathcal{Y}$ such that $f_i(\bar{x}, \bar{u}) < c_i$ for $i = 1, \cdots, N$.

We are now in the position to apply Theorem 2.2 to the problem (3.5). Theorem 2.2 states the equivalence between two optimization problems, the ‘primal’ problem (3.5) and its associated ‘dual’ problem, a maximization problem with cost functional (3.11) which is stated in Theorem 3.2. The primal problem is a complicated infinite dimensional optimization problem for which an optimal solution has to be found. In the case of optimal control problems such as (3.5), it is also desirable that the optimal solution be in ‘feedback’ form; that is, the optimal solution $u^*(t)$ at time $t \in [0, T]$ is a function of the state $x(t)$ at time $t$. This is certainly the case with the optimal control for the unconstrained LQ problem. On the other hand, the dual problem associated with (3.5) is an $N$-dimensional optimization problem in which feedback is not an issue. In the next theorem, we state the dual problem associated with (3.5), and show how the optimal control of (3.5) can be calculated from this.

Theorem 3.2 Let $\varphi(\lambda)$ be given by (3.11). Then there exists a $\lambda^* \geq 0$ which is optimal for

$$
\max_{\lambda} \left\{ \frac{1}{2} \xi' P(0, \lambda) \xi + d'(0, \lambda) \xi + \frac{1}{2} p(0, \lambda) - \lambda' c \right\} \tag{3.12}
$$

$$
\dot{P} = -PA - A'P + PBR^{-1}(\lambda)B'P - Q(\lambda), \quad P(T) = H(\lambda) \tag{3.13}
$$

$$
\dot{d} = -[A - BR^{-1}(\lambda)B'] d - a(\lambda) + PBR^{-1}(\lambda) b(\lambda), \quad d(T) = h(\lambda) \tag{3.14}
$$

$$
\dot{p} = [B'd + b(\lambda)]' R^{-1}(\lambda) [B'd + b(\lambda)], \quad p(T) = 0 \tag{3.15}
$$

$$
\lambda \geq 0 \tag{3.16}
$$

Furthermore, the optimal control of (3.5) is

$$
u^*(t) = -R^{-1}(t, \lambda^*) \left[ B'(t) P(t, \lambda^*) x(t) + B'(t) d(t, \lambda^*) + b(t, \lambda^*) \right] \tag{3.17}$$
3.2. FULL OBSERVATION STOCHASTIC CASE

**Remark 3.2** The notation \( P(t, \lambda^*) \) is to be interpreted as the solution of the Riccati equation (3.13) when \( \lambda = \lambda^* \). A similar interpretation holds for \( d(t, \lambda^*), p(t, \lambda^*) \).

From Theorem 3.2, it follows that the optimal control \( u^* \) for the LQ problem subject to integral quadratic constraints is calculated by solving the finite dimensional optimization problem (3.12)-(3.16). In the literature, (3.12)-(3.16) is known as an optimal parameter selection problem, and can be solved using gradient-type optimization algorithms if the gradient of the cost functional (3.12) with respect to \( \lambda \) can be determined. In Section 3.4, we show how the gradient of (3.12) can be calculated.

Note also the relationship between the optimal control (3.17) for (3.5) and the optimal control for the traditional unconstrained full observation LQ problem. In particular, the optimal control for the unconstrained LQ problem is a special case of (3.17), obtained by setting \( \lambda^* = 0 \) (that is, we get the unconstrained solution when none of the constraints are active, which is expected). However, it is important to realize that despite this relationship, (3.17) is not in feedback form! To see this, simply observe that \( u^*(t) \) depends not only on the current state \( x(t) \), but also on the initial state \( x(0) = \xi \). Indeed, the optimal Lagrange multiplier \( \lambda^* \), determined by (3.12)-(3.16) is a function of \( x(0) = \xi \) via the dual cost functional (3.12). This dependence of the optimal control on the initial state is due to the nature of the problem for a consequence of the integral constraints is that past behavior of the system is not forgotten. For this reason, the initial state is 'remembered' in the constraints and has influence on future values of the control.

### 3.2 Full observation stochastic case

Consider the stochastic differential equation

\[
\begin{align*}
    dx(t) &= (A(t) x(t) + B(t) u(t)) \, dt + C(t) \, dW(t), \\
    x(0) &= \xi
\end{align*}
\]  

(3.18)

We assume that (3.18) satisfies the same properties as the stochastic differential equation (2.41), as discussed in Section 2.2.3. The class of feasible controls \( U \) is defined by (2.40). For any \( u \in U \), there is a solution \( x \in X \) of the stochastic differential equation (3.18), where \( X \) is defined by (2.39). The pair \( (x, u) \) is called an admissible pair. The set of admissible pairs is \( X + Y \) where

\[
Y = \{(x, u) \in X \times U : dx(t) = (A(t)x(t) + B(t)u(t)) \, dt, \, x(0) = 0\}
\]  

(3.19)
and \( z = (\bar{z}, 0) \in \mathcal{X} \times \mathcal{U} \) such that

\[
\begin{align*}
    d\bar{z}(t) &= A(t)\bar{z}(t)dt + C(t)dW(t), \\
    \bar{z}(0) &= \xi
\end{align*}
\]  

(3.20)

Given any feasible control \( u \in \mathcal{U} \), the cost \((i = 0)\) and constraint \((i = 1, \ldots, N)\) functionals associated with the admissible pair \((x, u)\) are given by

\[
\begin{align*}
    f_i(x, u) &= E \left[ \frac{1}{2} \int_{0}^{T} \left[ x'(t)Q_i(t)x(t) + u'(t)R_i(t)u(t) \right] dt + \frac{1}{2} x'(T)H_i x(T) \\
    &+ \int_{0}^{T} \left( a'_i(t)x(t) + b'_i(t)u(t) \right) dt + h'_i x(T) \right] 
\end{align*}
\]

(3.21)

where \( Q_i(t), R_i(t), H_i, a_i(t), b_i(t), \) and \( h_i \) are the same as in (3.4). The full information LQG problem subject to integral quadratic constraints is

\[
\begin{align*}
    \begin{cases}
        f_0(x, u) \to \min \\
        f_i(x, u) \leq c_i, \quad i = 1, \ldots, N \\
        (x, u) \in z + \mathcal{Y}
    \end{cases}
\end{align*}
\]

(3.22)

Analogous to Assumption 3.1, we make the following assumption:

**Assumption 3.3** There exists an admissible pair \((x, u)\) which is feasible for (3.22).

The following result relating to the existence of an optimal solution \((x^*, u^*)\) for (3.22) follows from Assumption (3.3). It can be proved in exactly the same manner as Theorem 3.1.

**Theorem 3.3** Suppose that Assumption 3.3 holds. Then there exists a unique optimal solution \((x^*, u^*)\) of (3.22).

Let \( \zeta = (x, u) \). We define, as in the deterministic case, the Lagrangian \( L(\zeta, \lambda) \) by

\[
L(\zeta, \lambda) = f_0(\zeta) + \sum_{i=1}^{N} \lambda_i (f_i(\zeta) - c_i)
\]

(3.23)

By definition (see (2.2)), the value of the dual functional \( \varphi(\lambda) \) associated with (3.22) is given by the optimal cost of the following minimization problem over \( \zeta = z + \mathcal{Y} \) with cost functional \( L(\zeta, \lambda) \); that is,

\[
\varphi(\lambda) = \min_{\zeta \in z + \mathcal{Y}} L(\zeta, \lambda)
\]

(3.24)
3.2. FULL OBSERVATION STOCHASTIC CASE

In this case, the minimization problem in (3.24) is a standard unconstrained full observation LQG problem. A closed form expression for the dual functional $\varphi(\lambda)$ is given in the following proposition.

**Proposition 3.2** Let $\lambda \geq 0$ be given. Let $(x(\lambda), u(\lambda))$ denote the optimal state-control pair for the problem

$$
\begin{cases}
\min_\zeta L(\zeta, \lambda) \\
\zeta \in z + \mathcal{Y}
\end{cases}
$$

Then the optimal control is

$$u(t)(t, \lambda) = -R^{-1}(t, \lambda) \left[ B'(t) P(t, \lambda) x(t) + B'(t) d(t, \lambda) + b(t, \lambda) \right] \quad (3.25)
$$

and $x(\lambda)$ is the solution of (2.41) with $u = u(\lambda)$. The optimal cost is

$$
\varphi(\lambda) = \frac{1}{2} \xi' P(0, \lambda) \xi + d'(0, \lambda) \xi + \frac{1}{2} p(0, \lambda) + \frac{1}{2} \int_0^T \text{tr} \left\{ C'(t) P(t, \lambda) C(t) \right\} dt - \lambda' c \quad (3.26)
$$

**Proof:** Suppose that $\lambda \geq 0$ is fixed. Let $X_t = \sigma \{ x(s) : s \in [0, t] \}$ and

$$
\mathcal{V} = \{ u \in U : u(t) \text{ is measurable with respect to } X_t \}
$$

Then the optimal control for the problem

$$
\begin{cases}
\min_{(x, u)} L((x, u), \lambda) \\
(x, u) \text{ satisfies (2.41)} \\
u \in \mathcal{V}
\end{cases}
$$

is given by (3.25). In [15, Corollary 4.1, pp 163] it is also shown that (3.25) is also optimal over the class $u \in \mathcal{U}$ and the result follows immediately.

Theorem 2.2 shows that $\varphi(\lambda)$ is the cost functional of a maximization problem, the solution of which gives the optimal control of (3.22). To apply Theorem 2.2, we need to make the following assumption.

**Assumption 3.4** For every $\lambda_i \geq 0$, $i = 1, \cdots, N$ (not all equal to zero), there exists $(x, u) \in z + \mathcal{Y}$ such that

$$
\sum_{i=1}^N \lambda_i (f_i(x, u) - c_i) < 0
$$
Under Assumption 3.4, we have the following result:

**Theorem 3.4** There exists a $\lambda^* \geq 0$ which is optimal for the problem

$$\max_{\lambda} \varphi(\lambda)$$
subject to: (3.13) - (3.15) \hspace{1cm} (3.27)

$\lambda \geq 0$

Furthermore, the unique optimal control for (3.27) is

$$u^*(t) = -R^{-1}(t, \lambda^*) [B'(t)P(t, \lambda^*)x(t) + B'(t)d(t, \lambda^*) + b(t, \lambda^*)]$$ \hspace{1cm} (3.28)

**Proof:** This is an immediate consequence of the Lagrange Duality Theorem (Theorem 2.2) which is true under Assumption 3.4. $\blacksquare$

Unlike the unconstrained [55] and the linearly constrained (see Sections 2.2.3 and 2.3.3) LQG problems, certainty equivalence does not hold in the full-information LQG problem subject to integral quadratic constraints because the optimization problems (3.12)-(3.16) and (3.27) do not generally have the same optimal solution. On the issue of calculating $\lambda^*$, the problem (3.27) (as in the deterministic case (3.12)-(3.16)) is an optimal parameter selection problem. This is a finite dimensional optimization problems over $\lambda \in \mathbb{R}^N$ and can be solved using gradient-type optimization algorithms so long as the gradient of the cost functional $\varphi(\lambda)$ as given in (3.27), with respect to the parameter $\lambda$ can be calculated.

### 3.3 Partial observation stochastic case

Consider the partially observed linear system

$$\begin{align*}
\frac{dx}{dt} &= (A(t)x(t) + B(t)u(t))dt + C(t)dW(t) \quad x(0) \sim N(\xi, \Sigma_0) \quad (3.29) \\
\frac{dy}{dt} &= F(t)x(t)dt + G(t)dV(t), \quad y(0) = 0. \quad (3.30)
\end{align*}$$

We assume in this section that the conditions stated in Section 2.2.4 for the system (2.51)-(2.52) hold for (3.29)-(3.30). We take the class of feasible controls to be $\bar{U}$, as defined in (2.53).
The cost \( i = 0 \) and constraint \( i = 1, \ldots, N \) functionals associated with a feasible \( u \in \bar{U} \) and the resulting solution \( x \) of (3.29) is

\[
 f_i(x, u) = \mathbb{E} \left[ \frac{1}{2} \int_0^T \left[ x'(t)Q_i(t)x(t) + u'(t)R_i(t)u(t) \right] dt + \frac{1}{2} x'(T)H_i x(T) \\
 + \int_0^T \left( a'_i(t)x(t) + b'_i(t)u(t) \right) dt + h'_i x(T) \right] \tag{3.31}
\]

The partially observed LQG optimal control problem subject to integral quadratic constraints can be defined as follows:

\[
\begin{aligned}
\min f_0(x, u) \\
f_i(x, u) \leq c_i, \quad i = 1, \ldots, N \\
(x, u) \text{ satisfies (2.51), } u \in \bar{U}
\end{aligned} \tag{3.32}
\]

We solve this problem by converting it into an equivalent full observation problem that can be solved using the techniques in Section 3.2. As with the full observation case, we need the following assumptions:

**Assumption 3.5** There exists \( (x, u) \) such that \( u \in \bar{U}, x \) satisfies (2.51) and \( f_i(x, u) \leq c_i \) for \( i = 1, \ldots, N \).

**Assumption 3.6** For every \( \lambda_i \geq 0, i = 1, \ldots, N \) (not all zero), there exists \( (x, u) \) satisfying \( u \in \bar{U} \) and (2.51) such that

\[
\sum_{i=1}^N \lambda_i (f_i(x, u) - c_i) < 0
\]

Using standard techniques, we can show that

\[
f_i(x, u) = \mathbb{E} \left[ \frac{1}{2} \int_0^T \left[ \hat{x}'(t)Q_i(t)\hat{x}(t) + \hat{u}'(t)R_i(t)\hat{u}(t) \right] dt + \frac{1}{2} \hat{x}'(T)H_i \hat{x}(T) \\
+ \int_0^T \left( a'_i(t)\hat{x}(t) + b'_i(t)\hat{u}(t) \right) dt + h'_i \hat{x}(T) \right] + \frac{1}{2} \int_0^T \text{tr}\{Q_i(t)\Sigma(t)\} dt \tag{3.33}
\]

where \( \hat{x}(t) \) is the output of the Kalman Filter

\[
d\hat{x}(t) = A(t)\hat{x}(t)dt + B(t)\nu(t)dt - \Sigma(t)F(t) \left( G(t)G'(t) \right)^{-1} d\nu(t), \quad \hat{x}(0) = \xi \tag{3.34}
\]

associated with (3.29)-(3.30), and \( \Sigma(t) \) is the output of the filtering Riccati equation

\[
\dot{\Sigma} = \Sigma A + A'\Sigma - \Sigma F'(GG')^{-1} F\Sigma + CC', \quad \Sigma(0) = \Sigma_0 \tag{3.35}
\]
Relevant properties of (3.34)-(3.35) are stated in Lemma 2.1. Denoting
\[
\dot{c}_i = c_i - \frac{1}{2} \int_0^T \text{tr}\{Q_i(t) \Sigma(t)\} dt - \frac{1}{2} H_i \Sigma(T)
\]
(3.36) and
\[
f_i(\hat{x}, u) = E \left[ \frac{1}{2} \int_0^T \left[ \dot{x}'(t) Q_i(t) \dot{x}(t) + u'(t) R_i(t) u(t) \right] dt + \frac{1}{2} \dot{x}'(T) H_i \dot{x}(T) 
+ \int_0^T \left( a_i'(t) \dot{x}(t) + b_i'(t) u(t) \right) dt + h_i' \dot{x}(T) \right]
\]
(3.37)

it follows that (3.32) can be written as the full observation problem
\[
\begin{cases}
    f_0(\hat{x}, u) \to \min \\
    f_i(\hat{x}, u) \leq \hat{c}_i, \ i = 1, \cdots, N \\
    (\hat{x}, u) \text{ satisfies (3.34) and } u \in \tilde{U}
\end{cases}
\]
(3.38)
Moreover, if Assumption 3.5 holds for the partially observed problem (3.32), then Assumption 3.3 is satisfied for the equivalent full observation problem (3.38). Therefore, the following existence result is an immediate consequence of Theorem 3.3.

**Theorem 3.5** Suppose that Assumption 3.5 holds. Then there exists a unique optimal solution \((x^*, u^*)\) of (3.32).

For every \(\lambda \geq 0\), we can define the Lagrangian \(L(\hat{x}, u, \lambda)\) as
\[
L(\hat{x}, u, \lambda) = f_0(\hat{x}, u) + \sum_{i=1}^N \lambda_i (f_i(\hat{x}, u) - c_i)
\]
(3.39)
where \(f_i(\hat{x}, u)\) are given by (3.37). Under Assumption 3.6 together with the arguments used in the proof of Theorem 2.9, Theorem 3.4 can be used to solve (3.38). In particular, the optimal control for (3.38) (and therefore for (3.32)) can be obtained by solving the optimal parameter selection problem (3.27), but with the cost functional \(\varphi(\lambda)\) replaced by:
\[
\varphi(\lambda) = \frac{1}{2} \xi' P(0, \lambda) \xi + d'(0, \lambda) \xi + \frac{1}{2} P(0, \lambda)
+ \frac{1}{2} \int_0^T \text{tr} \{\Gamma'(t) P(t, \lambda), \Gamma(t)\} dt - \lambda' c
\]
(3.40)
where
\[
\Gamma(t) = \Sigma(t) H(t)(G(t)G'(t))^{-1} F(t)
\]
This is summarized as follows.
3.4. OPTIMAL PARAMETER SELECTION PROBLEMS

Theorem 3.6 (Quasi-Separation Theorem) Let $\varphi(\lambda)$ be given by (3.40). Then there exists a $\lambda^* \geq 0$ which is optimal for the problem:

$$\begin{aligned}
\max_{\lambda} \varphi(\lambda) \\
\text{Subject to: (3.13) - (3.15)} \\
\lambda \geq 0
\end{aligned}$$

(3.41)

Furthermore, the unique optimal control for (3.32) is

$$u_t^* = -R^{-1}(t, \lambda^*) \left[ B'(t)P(t, \lambda^*)\hat{x}_t + B'(t)d(t, \lambda^*) + b(t, \lambda^*) \right]$$

(3.42)

where $\hat{x}_t$ is the solution of (3.34).

The reader should note the following. First, certainty equivalence does not hold. Second and more importantly, the Separation Theorem does not hold in the sense of [55]. To see this, the reader should observe (3.36), (3.40) and (3.41). From these equations, it is clear that the solution $\lambda^*$ of (3.41) is dependent on the error covariance $\Sigma(t)$ associated with the filtering problem (3.34)-(3.35). Thus the problems of filtering and control are not independent. However, when solving (3.41) the dependence of $\lambda^*$ (and hence the solution of the control problem) on $\Sigma(t)$ adds no complication. Since $\Sigma(t)$ is independent of $\lambda$, it needs to be calculated only once, and the optimization problem (3.41) may be solved with no further re-calculation of $\Sigma(t)$. It is in this sense that the control and filtering problems are independent, and hence our naming Theorem 3.6 a Quasi-Separation Theorem.

3.4 Optimal parameter selection problems

In view of Theorems 3.2, 3.4 and 3.6, the optimal control for the deterministic, the full observation and the partial observation LQG problems with integral quadratic constraints is obtained by solving the finite dimensional optimization problems (3.12)-(3.16), (3.27) and (3.41) respectively. In the literature, these finite dimensional optimization problems are known as optimal parameter selection problems, and the interested reader may refer to [46] for more details. Optimal parameter selection problems can be solved as mathematical programming problems (using efficient gradient-type optimization algorithms) so long as the value of the cost functional and gradient of the cost functional can be calculated for any given $\lambda$. In this section, we derive the gradient of the cost functional for the problems (3.12)-(3.16), (3.27) and (3.41).
In the optimal parameter selection problem for the deterministic LQ problem (3.12)-(3.16), we show that solving an unconstrained LQ problem is the key step in calculating the gradient of the cost functional. In the optimal parameter selection problems for the full observation and partial observation problems, similar results hold; that is, the key step in calculating the gradient of the cost functional is solving an unconstrained full observation and an unconstrained partial observation LQG problem respectively. Interestingly, by virtue of the Separation theorem for unconstrained LQG, it follows that there is a Separation theorem associated with calculating the gradient.

We shall calculate the gradient of the optimal parameter selection problems (3.12)-(3.16), (3.27) and (3.41) by working with a general optimization problem with quadratic cost and quadratic constraints. The LQG problems as stated in (2.6), (3.22) are special cases of problems of this form and the partially observed LQG problem is solved by transforming it into a problem of this type (namely, a full observation problem). The general problem which we shall work with can be stated as follows: Let $X$ be a Hilbert space with inner product $\langle \cdot, \cdot \rangle$, $z \subset X$ and $\mathcal{Y}$ a closed subspace of $X$. Let $Q_i : X \to X$, $i = 0, \cdots, N$ be symmetric operators such that $Q_0 > 0$ and $Q_i \geq 0$ for $i = 1, \cdots, N$. Let $a_i \in X$ for $i = 1, \cdots, N$. The general quadratic optimization problem is

\[
\begin{cases}
  f_0(x) \to \min \\
  f_i(x) \leq c_i, \quad i = 1, \cdots, N \\
  x \in z + \mathcal{Y}
\end{cases}
\]  

(3.43)

where $f_i : X \to \mathbb{R}$ are the linear-quadratic cost functionals

\[
f_i(x) = \frac{1}{2} \langle Q_i x, x \rangle + \langle a_i, x \rangle
\]  

(3.44)

The optimal parameter selection problems (3.12)-(3.16), (3.27) and (3.41) correspond to the dual problem of (3.43) with $f_i(x)$ given by (3.44). We shall derive the gradient of the cost functional of the optimal parameter selection problems by examining the dual problem associated with (3.43). The dual problem associated with (3.43) is an optimization problem over $\lambda \in \mathbb{R}^N$ and may be stated as follows:

\[
\begin{cases}
  J(\lambda) = \frac{1}{2} \langle Q(\lambda) x, x \rangle + \langle a(\lambda), x \rangle \to \max \\
  Q(\lambda) x + a(\lambda) \in \mathcal{Y}^\perp \\
  x \in z + \mathcal{Y}, \quad \lambda \geq 0
\end{cases}
\]  

(3.45)

It should be noted that for every $\hat{\lambda} \in \mathbb{R}^N$, $\lambda \geq 0$, there is a unique $x(\lambda) \in z + \mathcal{Y}$
3.4. OPTIMAL PARAMETER SELECTION PROBLEMS

satisfying the constraint

$$Q(\lambda) \cdot x + a(\lambda) \in \mathcal{Y}^\perp$$

and that $x(\lambda)$ is the optimal solution of the following optimization problem over $x \in z + \mathcal{Y}$

$$\{ \frac{1}{2} \langle Q(\lambda) \cdot x, x \rangle + \langle a(\lambda), x \rangle \to \min \quad x \in z + \mathcal{Y} \}$$

(3.46)

Moreover, $x(\lambda)$ is a smooth function of $\lambda \geq 0$. The optimal parameter selection problems (3.12)-(3.16), (3.27) and (3.41) correspond to the dual problem (3.45). The following theorem gives the gradient $\frac{dJ(\lambda)}{d\lambda}$ where $J(\lambda)$ is as stated in (3.45).

**Theorem 3.7** For every $\lambda \in \mathbb{R}^N$, $\lambda \geq 0$ the gradient of $J(\lambda)$ with respect to $\lambda$ is

$$\frac{dJ(\lambda)}{d\lambda} = \left[ \frac{\partial J(\lambda)}{\partial \lambda_1}, \cdots, \frac{\partial J(\lambda)}{\partial \lambda_N} \right]'$$

$$\frac{\partial J(\lambda)}{\partial \lambda_i} = f_i(x(\lambda)) - c_i$$

(3.47)

where $x(\lambda)$ is the unique $x \in z + \mathcal{Y}$ which satisfies the constraint $Q(\lambda) \cdot x + a(\lambda) \in \mathcal{Y}^\perp$ and $f_i(\lambda)$ is given by (3.44).

**Proof:** By the chain rule

$$\frac{dJ(\lambda)}{d\lambda_i} = \frac{\partial J(\lambda)}{\partial \lambda_i} + \frac{\partial J(\lambda)}{\partial x(\lambda)} \cdot \frac{\partial x(\lambda)}{\partial \lambda_i}$$

From (3.45), we obtain

$$\frac{\partial J(\lambda)}{\partial x(\lambda)} = Q(\lambda) \cdot x(\lambda) + a(\lambda) \in \mathcal{Y}^\perp$$

On the other hand, $x(\lambda) \in z + \mathcal{Y}$ for every $\lambda \geq 0$ and hence

$$\frac{\partial x(\lambda)}{\partial \lambda_i} \in \mathcal{Y}$$

It follows that for every $\lambda \geq 0$

$$\frac{\partial J(\lambda)}{\partial x(\lambda)} \cdot \frac{\partial x(\lambda)}{\partial \lambda_i} = 0$$

The result follows from the fact that

$$\frac{\partial J(\lambda)}{\partial \lambda_i} = \frac{1}{2} \langle Q_i \cdot x(\lambda), x(\lambda) \rangle + \langle a_i, x(\lambda) \rangle - c_i = f_i(x(\lambda)) - c_i$$
Thus, for every \( \lambda \geq 0 \), the gradient of \( J(\lambda) \) is obtained in the following way. First, we solve the unconstrained optimization problem (3.46) over \( x \in \Omega \) and obtain the optimal solution \( x(\lambda) \). Once \( x(\lambda) \) has been obtained, the components of \( \frac{dJ(\lambda)}{d\lambda} \) are obtained by evaluating the constraint functionals \( f_i(x) - c_i \) at \( x(\lambda) \). Since the cost functionals of the optimal parameter selection problems (3.12)-(3.16), (3.27) and (3.41) correspond to the cost functional \( J(\lambda) \) of a dual problem of the form (3.45), we obtain the following result from Theorem 3.7.

**Theorem 3.8** Let \( \lambda \in \mathbb{R}^N \), \( \lambda \geq 0 \) and \( J(\lambda) = g(\zeta(\lambda), \lambda) - \lambda c \) where \( g(\zeta(\lambda), \lambda) \) is given by (3.11). The gradient of the cost functional \( J(\lambda) \) evaluated at \( \lambda \) is

\[
\frac{dJ(\lambda)}{d\lambda} = \left[ \frac{\partial J(\lambda)}{\partial \lambda_1}, \ldots, \frac{\partial J(\lambda)}{\partial \lambda_N} \right]
\]

\[
\frac{\partial J(\lambda)}{\partial \lambda_j} = \frac{1}{2} \int_0^T \left[ \beta' Q_j \beta + \eta' R_j \eta \right] dt + \frac{1}{2} \beta'(T) H_j \beta(T)
+ \int_0^T \left[ a'_j \beta + b'_j \eta \right] dt + h'_j(T) \beta(T) - c_j
\]

where \( \beta(t) \) is the solution of the equation

\[
\dot{\beta}(t) = A(t) \beta(t) + B(t) \eta(t), \quad \beta(0) = \xi
\]

with \( \eta(t) \) given by

\[
\eta(t) = -R^{-1}(\lambda, t) \left[ B'(t) P(\lambda, t) \beta(t) + B'(t) d(\lambda, t) + b(\lambda, t) \right]
\]

From Theorem 3.8, it is clear that calculating the gradient of \( J(\lambda) \) with respect to \( \lambda \) is equivalent to solving an unconstrained LQ optimal control problem, and evaluating the value of the constraints with this optimal control.

The optimal parameter selection problem (3.27) is the dual problem of the full observation problem (3.22). By Theorem 3.7, the gradient of the cost functionals of (3.27) and (3.41) is given as follows.

**Theorem 3.9** Let \( \lambda \in \mathbb{R}^N \), \( \lambda \geq 0 \) be given. Then the cost functional of the problem (3.27) is of the form

\[
J(\lambda) = \frac{1}{2} \xi' P(\lambda, 0) \xi + d'(\lambda, 0) \xi + \frac{1}{2} b(\lambda, 0)
+ \frac{1}{2} \int_0^T \text{tr} \left\{ C'(t) P(t, \lambda), C(t) \right\} dt - \lambda c
\]

(3.51)
The gradient of $J(\lambda)$ evaluated at $\lambda$ is

$$\frac{dJ(\lambda)}{d\lambda} = \left[ \frac{\partial J(\lambda)}{\partial \lambda_1}, \cdots, \frac{\partial J(\lambda)}{\partial \lambda_N} \right]$$

$$\frac{\partial J(\lambda)}{\partial \lambda_j} = E \left[ \frac{1}{2} \int_0^T \left[ \beta' Q \beta + \eta' R \eta \right] dt + \frac{1}{2} \beta'(T) H_j \beta(T) \right.$$

$$\left. + \int_0^T \left[ a'_j \beta + b'_j \eta \right] dt + h'_j(T) \beta(T) \right] - c_j$$

(3.52)

where

$$d\beta(t) = (A(t)\beta(t) + B(t)\eta(t)) dt + C(t)dW(t) \quad \beta(0) \sim N(\xi, \Sigma_0)$$

(3.53)

with $\eta(t)$ given by

$$\eta_t = -R^{-1}(\lambda, t) \left[ B'(t) P(\lambda, t) \beta_t + B'(t) d(\lambda, t) + b(\lambda, t) \right]$$

(3.54)

As in the deterministic case, the problem of calculating the gradient is equivalent to solving an unconstrained full observation LQG control problem, and evaluating the constraints with this optimal control.

The partially observed problem (3.32) is solved by transforming it into the full observation problem (3.38). This full observation problem is solved by finding the optimal solution of the optimal parameter selection problem (3.41). By Theorem 3.7, the gradient of the cost functional of the optimal parameter selection problem (3.41) is obtained by solving the unconstrained version of this full observation problem. However, this unconstrained full observation problem is obtained from the unconstrained version of the partially observed problem. For this reason, there is a Separation Theorem for calculating the gradient.

**Theorem 3.10 (Gradient Separation Theorem)** Let $\lambda \in \mathbb{R}^N$, $\lambda \geq 0$ be given. Then the cost functional of the problem (3.41) is

$$J(\lambda) = \frac{1}{2} \xi' P(0, \lambda) \xi + d'(0, \lambda) \xi + \frac{1}{2} p(0, \lambda)$$

$$+ \frac{1}{2} \int_0^T \text{tr} \left\{ \Gamma'(t) P(t, \lambda), \Gamma(t) \right\} dt - \lambda' c$$

(3.55)

where $\Gamma(t) = \Sigma(t) F(t)(G(t) G'(t))^{-1} G(t)$. The gradient of $J(\lambda)$ evaluated at $\lambda$ is

$$\frac{dJ(\lambda)}{d\lambda} = \left[ \frac{\partial J(\lambda)}{\partial \lambda_1}, \cdots, \frac{\partial J(\lambda)}{\partial \lambda_N} \right]$$
\[
\frac{\partial J(\lambda)}{\partial \lambda_j} = E \left[ \frac{1}{2} \int_0^T \left[ \dot{\beta}' Q_j \dot{\beta} + \eta' R_j \eta \right] dt + \frac{1}{2} \dot{\beta}'(T) H_j \dot{\beta}(T) \right. \\
+ \left. \int_0^T \left[ a_j' \dot{\beta} + b_j' \eta \right] dt + h_j'(T) \beta(T) \right] - c_j
\] (3.56)

where

\[
d\dot{\beta}(t) = A(t) \dot{\beta}(t) dt + B(t) \eta(t) dt - \Sigma(t) F(t) (G(t) G'(t))^{-1} d\nu(t), \quad (3.57)
\]
\[
\beta(0) = \xi
\]
\[
\eta(t) = -R^{-1}(\lambda, t) \left[ B'(t) P(\lambda, t) \dot{\beta}(t) + B'(t) d(\lambda, t) + b(\lambda, t) \right] \quad (3.58)
\]
\[\nu(\cdot)\] is the innovations process given by \(d\nu(t) = d\eta(t) - F(t) \dot{\beta}(t) dt\) and \(y(\cdot)\) is the solution of the stochastic differential equations

\[
d\dot{\beta}(t) = (A(t) \beta(t) + B(t) \eta(t)) dt + C(t) dW(t), \quad \beta(0) \sim N(\xi, \Sigma_0) \quad (3.59)
\]
\[
d\dot{y}(t) = F(t) \beta(t) dt + G(t) dV(t), \quad y(0) = 0 \quad (3.60)
\]

where \(W(\cdot)\) and \(V(\cdot)\) are standard Brownian motions which satisfy the conditions stated in Section 3.3.

For computational purposes, the following expression for the gradient is the most useful. As observed in Theorem 3.6, when calculating the optimal control for the partially observed case, the problems of ‘control’ and ‘filtering’ are not independent but rather, the ‘control’ problem (namely, the optimal parameter selection problem (3.41)) depends on the solution \(\Sigma(t)\) of the filtering Riccati equation. This dependence on \(\Sigma(t)\) can be seen in the expression for the gradient of the cost functional of (3.41) which is stated in Theorem 3.10. Once again however, once \(\Sigma(t)\) has been determined, the problem of calculating the gradient can be carried out independently of the ‘filtering’ problem and hence, the control problem can be solved independently of the filtering problem.

**Theorem 3.11** Let \(K(t) = 0\) for the optimal parameter selection problem (3.12)-(3.16), \(K(t) = C(t)\) for (3.27) and \(K(t) = \Sigma(t) H(t)(G(t) G'(t))^{-1} F(t)\) for (3.41). Let \(\lambda > 0\) be given. Then the gradient of the cost functional \(J(\lambda)\) evaluated at \(\lambda\) for (3.12)-(3.16), (3.27) and (3.41) is

\[
\frac{dJ(\lambda)}{d\lambda} = \left[ \frac{\partial J(\lambda)}{\partial \lambda_1}, \ldots, \frac{\partial J(\lambda)}{\partial \lambda_N} \right]
\]
3.5. CONCLUSION

\[
\frac{\partial J(\lambda)}{\partial \lambda_j} = \frac{1}{2} \int_0^T [\beta' Q_j \beta + \eta' R_j \eta] \, dt + \frac{1}{2} \beta'(T) H_j \beta(T) \\
+ \int_0^T [a'_j \beta + b'_j \eta] \, dt + h'_p(T) \beta(T) \\
+ \frac{1}{2} \int_0^T \text{tr} \left\{ (PBR^{-1}(\lambda) R_j R^{-1}(\lambda) B' P + Q_j) \Lambda_K \right\} \, dt \\
+ \frac{1}{2} \text{tr} \{ \Lambda_K(T) \cdot H_j \} - c_j
\]

(3.61)

where \( \beta(t) \) is the solution of the equation

\[
\dot{\beta}(t) = A(t) \beta(t) + B(t) \eta(t), \quad \beta(0) = \xi
\]

(3.62)

with \( \eta(t) \) given by

\[
\eta(t) = -R^{-1}(\lambda, t) \left[ B'(t) P(\lambda, t) \beta(t) + B'(t) d(\lambda, t) + b(\lambda, t) \right]
\]

(3.63)

\( \Lambda_K(t) \) is the solution of

\[
\dot{\Lambda}_K = (A - BR^{-1}(\lambda) B' P) \Lambda_K + \Lambda_K (A - BR^{-1}(\lambda) B' P)' + KK',
\]

(3.64)

\( \Lambda_K(0) = 0 \)

and \( P(\lambda, t), d(\lambda, t) \) are the solutions of the differential equations (2.26)-(2.27).

3.5 Conclusion

We have studied the LQ and LQG optimal control problems subject to finitely many integral quadratic constraints. In Chapter 2, it was shown that the classic Separation Theorem result of Wohnam holds in the case on finitely many and infinitely many integral linear constraints. In this Chapter, we have shown that it does not hold in the case of integral quadratic constraints. However, a generalization of the Separation Theorem which we call a Quasi-Separation Theorem holds instead. Furthermore, we have derived computational methods for calculating the optimal control by examining the dual problem. One advantage of this approach is that the infinite dimensional, closed-loop constrained LQG problem is converted into a finite dimensional open loop problem that can be solved using standard numerical optimization packages. In the case of finitely many quadratic constraints, the dual problem is a finite dimensional optimization problem, and for this reason, we have derived the gradient of the dual cost functional so that efficient algorithms for finite dimensional optimization problems can be used to calculate the optimal solution. The key step in calculating the
gradient is solving an unconstrained LQG problem, with the components of the gradient vector being given by the value of the constraint functionals evaluated with this unconstrained optimal control. Although the Separation Theorem does not hold for the LQG problem with integral quadratic constraints, it is interesting to see that there is a Separation Theorem associated with the problem of calculating the gradient.
Chapter 4

Indefinite constrained LQG control

In traditional LQG theory, it is a standard assumption that the control weighting matrix be strictly positive definite; for example, see Anderson and Moore [1]. In the deterministic case, this is necessary for there to exist a finite optimal cost that is achievable by a unique optimal control. Such a problem is said to be well posed. In the traditional stochastic full observation problem, this positive definite condition is generally assumed (see Bensoussan [5] for example). In such a case, there appears little difference between the deterministic, and the full observation stochastic LQG problems. Indeed, the optimal control in both these problems is given by a linear state feedback; the feedback gain being identical in both cases and determined by the solution of a deterministic, backward Riccati equation.

However, recent results by Chen, Li and Zhou [58] show that this assumption is not necessary in the full observation case when the diffusion term in the state equation is dependent on the control. Such a situation corresponds to the case when the control input influences the variance of the disturbances to the system. Necessary and sufficient conditions for solvability are derived which show the allowable values of the control weight for which the stochastic LQG problem is well posed. In particular, the control weighting matrix may have negative eigenvalues.

In this paper, we consider the full observation LQG problem with integral quadratic constraints. Unlike the results in Chapter 3, we do not assume that the control weighting matrix in any of the cost or constraint functionals is positive definite. Building
on the results of Chen, Li and Zhou [58], sufficient conditions under which the optimal control can be determined explicitly are presented. In fact, it is shown that the conditions obtained in [58] for well posedness of the unconstrained LQG problem are actually the necessary and sufficient conditions for strict convexity of the cost functional (with a possibly indefinite control weighting matrix).

The paper is organized as follows. In Section 4.1, we present the (indefinite) full observation LQG problem with integral quadratic constraints. In Section 4.2, we present a brief overview of the results obtained in [58]. In particular, we present conditions for solvability of the unconstrained indefinite stochastic LQG problem. In Section 4.3, we derive the conditions for strict convexity of the cost and constraint functionals which in turn, gives a sufficient condition for which the optimal control can be obtained explicitly. We end with a brief conclusion in Section 4.4.

The results of this chapter can be found in the paper by Lim and Zhou [32].

4.1 Problem statement

Let \((\Omega, \mathcal{F}, P)\) be a given probability triple and consider the following Ito stochastic differential equation:

\[
\begin{align*}
\frac{d}{dt} x(t) &= (A(t) x(t) + B(t) u(t) + g(t)) \ dt + (h(t) + D(t) u(t)) \ dW(t), \\
x(s) &= y
\end{align*}
\] (4.1)

where \((s, y) \in [0, T) \times \mathbb{R}^n\) are the initial time and state respectively, associated with (4.1). For convenience, we shall assume that the process \(W(\cdot)\) is a one-dimensional Brownian motion on \([0, T]\). Generalization of these results to the multi-dimensional case is straightforward. Associated with \(W(\cdot)\) is the filtration

\[
\mathcal{F}_t = \sigma\{W(s) : 0 \leq s \leq t\}.
\]

For a given Hilbert space \(X\) with norm \(\| \cdot \|_X\), \(1 \leq p \leq \infty\) and \(a, b \in \mathbb{R}\) such that \(a \leq b\), we define the following Banach space:

\[
L^p_x(a, b; X) = \{ \phi(\cdot) = \{\phi(t, \omega) : a \leq t \leq b\} : \phi(\cdot) \text{ is an } X\text{-valued} \\
\mathcal{F}_t\text{-measurable process on } [0, T] \text{ with } E \int_a^b \|\phi(t, \omega)\|_X^p \ dt < \infty\}
\] (4.3)
4.1. PROBLEM STATEMENT

with norm

$$\|\phi(\cdot)\|_{L^p} = \left( E \int_a^b \|\phi(t, \omega)\|_X^p \, dt \right)^{\frac{1}{p}} \quad (4.4)$$

In the case when the functions we are interested in are deterministic, we shall sometimes write the space (4.3) as $L^p(a, b; X)$ and its associated norm (4.4) as $\| \cdot \|_p$. We also denote by $C(a, b; X)$ the Banach space of $X$-valued continuous functions on $[a, b]$ with the max-norm induced by $\| \cdot \|_X$; that is, for any $K \in C(a, b; X)$ we have

$$\|K\| = \max_{a \leq t \leq b} \|K(t)\|_X$$

In certain cases, it will be convenient for us to somewhat abuse the notation (4.3) for $L^p_x(a, b; X)$ in the following way. Suppose $Y$ is a subset of $X$. Then we will use $L^p_x(a, b; Y)$ to denote the following subset of $L^p_x(a, b; X)$:

$$L^p_x(a, b; Y) = \{ \phi(\cdot) \in L^p_x(a, b; X) : \phi(\cdot) \text{ is } Y\text{-valued} \}$$

Note in particular that $L^p_x(a, b; Y)$ is not even a vector space (let alone a Banach space) unless $Y$ is a closed subspace of $X$. A similar comment applies to our use of the notation $C(a, b; Y)$.

The class of admissible controls associated with (4.1) is the set $U_{ad} = L^2_x(0, T; \mathbb{R}^m)$. Given $u(\cdot) \in U_{ad}$, the pair $(x(\cdot), u(\cdot))$ shall be referred to as an admissible pair if $x(\cdot) \in L^2_x(0, T; \mathbb{R}^n)$ is the solution of the stochastic differential equation (4.1) associated with $u(\cdot) \in U_{ad}$.

Given any $u(\cdot) \in U_{ad}$, the cost $(i = 0)$ and constraint $(i = 1, \cdots, m)$ functionals are given by

$$J_i(s, y; u(\cdot)) = E \left\{ \frac{1}{2} \int_s^T \left( x'(t) Q_i(t) x(t) + u'(t) R_i(t) u(t) \right) \, dt + \frac{1}{2} x'(T) M_i x(T) \right\} \quad (4.5)$$

where $E(\cdot) = E(\cdot|\mathcal{F}_s)$. Throughout this paper, we shall assume that

$$A \in L^\infty(0, T; \mathbb{R}^{n \times n}) \cap C(0, T; \mathbb{R}^{n \times n})$$
$$B, D \in L^\infty(0, T; \mathbb{R}^{n \times m}) \cap C(0, T; \mathbb{R}^{n \times m})$$
$$g, h \in L^\infty(0, T; \mathbb{R}^n) \cap C(0, T; \mathbb{R}^n)$$
$$Q_i \in L^2(0, T; S^n_+) \cap C(0, T; S^n_+)$$
$$R_i \in L^2(0, T; S^n) \cap C(0, T; S^n)$$
$$M_i \in S^n_+ \quad (4.6)$$
where \( S^n \) and \( S^n_+ \) refer to the set of all symmetric, and non-negative symmetric \( \mathbb{R}^{n \times n} \)-valued matrices respectively. Note that \( S^n \) with the inner product

\[
\langle X, Y \rangle = \text{tr} (XY)
\]

is a Hilbert space. We note here once again that unlike the standard case, the control weighting matrices \( R_i \) may be indefinite. Given \( c_1, \cdots, c_m \in \mathbb{R} \), the optimal control problem associated with the system (4.1) and cost/constraint functionals (4.5) is as follows:

\[
\begin{align*}
\min J_0(s, y; u(\cdot))
\end{align*}
\]

\[
\min J_1(s, y; u(\cdot)) \leq c_1
\]

\[
\vdots
\]

\[
\min J_m(s, y; u(\cdot)) \leq c_m
\]

\[
J(\cdot) \in U_{ad}, \quad x(\cdot) \text{satisfies (4.1)}.
\]

4.2 Preliminary results

In standard unconstrained LQG control theory, a basic assumption is that the control weighting matrix \( R(t) \) is strictly positive definite. In the paper [58] by Chen, Li and Zhou, there is a detailed study of the unconstrained LQG problem for the case when the control weighting matrix may be indefinite and the diffusion term in the stochastic differential equation is linearly dependent on the control. In particular, issues such as existence and uniqueness of solutions for the indefinite LQG problem, its relation to the solvability of a certain Riccati equation, and the conditions under which this Riccati equation admits a unique positive definite solution are examined. In this section, we summarize key results pertaining to this work.

The stochastic LQG problem considered in [58] is stated as follows:

\[
\min J(s, y; u(\cdot)) = E \left\{ \frac{1}{2} \int_s^T \left( x'(t) Q(t) x(t) + u'(t) R(t) u(t) \right) dt + \frac{1}{2} x'(T) M x(T) \right\}
\]

\[
dx(t) = (A(t) x(t) + B(t) u(t) + g(t)) dt + (h(t) + D(t) u(t)) dW(t),
\]

\[
x(s) = y
\]

where it is assumed that the cost functional (4.8) satisfies the same conditions as the functionals (4.5) and the stochastic differential equation (4.9) satisfies the same
conditions as (4.1). As in the constrained case, the class of feasible controls is $U_{ad}$. In particular, we emphasize again that the control weighting matrix $R(t)$ in (4.8) need not be positive definite.

As in the case of the standard unconstrained LQG problem (that is, when $R(t)$ is strictly positive definite and $D(t) = 0$), solvability of (4.8)-(4.9) is closely related to that of a certain Riccati differential equation. In [58], it is shown that the indefinite LQG problem (4.8)-(4.9) is related closely related to a Riccati equation of the following type:

$$
\begin{aligned}
\dot{P} + PA + A'P - PB(R + D'PD)^{-1}B'P + Q &= 0 \\
P(T) &= M \\
R + D'PD > 0
\end{aligned}
$$

(4.11)

More precisely, this relationship can be stated as follows.

**Theorem 4.1** If the Riccati equation (4.11) admits a solution, then the stochastic LQG problem (4.8)-(4.9) is well posed. Moreover, the optimal control is given by

$$
u^*(t) = -(R(t) + D'(t)P(t)D(t))^{-1}B'(t)P(t)x(t)$$

(4.12)

and the corresponding optimal cost is

$$J(s, y; \nu^*(\cdot)) = y'P(s)y$$

(4.13)

Therefore, the questions of existence and uniqueness of solutions of the Riccati equation (4.11) are fundamental when addressing the issue of solvability of (4.8)-(4.9). The next result relates to the issue of uniqueness.

**Theorem 4.2** If $P$ is a solution to the Riccati equation (4.11), then $P \in C(0, T; S^t_\mathbb{R})$ and it is the only solution.

The final result in this section deals with the necessary and sufficient conditions for solvability of the Riccati equation (4.11). In order to state the result, we need to introduce some notation. First, let $\hat{S}_\mathbb{R}$ denote the set of all symmetric, positive definite $\mathbb{R}^{m \times m}$-valued matrices and consider the set

$$\mathcal{K} = \{K \in L^\infty(0, T; \hat{S}_\mathbb{R}^t) : K^{-1} \in L^\infty(0, T; \hat{S}_\mathbb{R}^t)\}$$

(4.14)
Note in particular that the set of continuous, symmetric, positive definite $\mathbb{R}^{m \times m}$-valued matrices $C(0,T;\mathbb{S}_+^m) \subset \mathcal{K}$. Given any $K \in \mathcal{K}$, it is well known that the corresponding (traditional) LQG problem

$$\min E \left\{ \frac{1}{2} \int_0^T (x'(t) Q(t) x(t) + u'(t) K(t) u(t)) dt + \frac{1}{2} x'(T) M x(T) \right\}$$

(4.15)

$$dx(t) = (A(t) x(t) + B(t) u(t) + g(t)) dt + h(t) dW(t), \quad x(s) = y$$

(4.16)

is well posed, and the Riccati equation

$$\dot{P} + PA + A'P - PBK^{-1}B'P + Q = 0$$

$$P(T) = M$$

(4.17)

has a unique positive definite solution; that is, to each $K \in \mathcal{K}$, there corresponds a $P \in C(0,T;\mathbb{S}_+^m)$ that is determined via (4.17). Therefore, we can define the mapping $\Psi : \mathcal{K} \rightarrow C(0,T;\mathbb{S}_+^m)$ where $P = \Psi(K)$ is the solution of the Riccati equation (4.17). The necessary and sufficient conditions for solvability of the Riccati equation (4.11) can be expressed in terms of this mapping $\Psi$ as follows:

**Theorem 4.3** The Riccati equation (4.11) admits a solution if and only if there exists $K \in C(0,T;\mathbb{S}_+^m)$ such that

$$R + D' \Psi(K) D \geq K$$

(4.18)

In the following sections, we shall examine the LQG problem with integral quadratic constraints, where the control weighting matrix in neither the cost nor the constraint functionals is assumed to be positive definite. The necessary and sufficient condition (4.18) for solvability of the Riccati equation (4.11) is shown to be a condition for strict convexity of the cost functional (4.8). In this way, sufficient conditions for the solvability of (4.7) are derived, and the optimal control determined.

### 4.3 Main results

In this section, we derive a set of sufficient conditions under which the constrained optimal control problem (4.7) can be solved and the optimal control determined explicitly using results from duality theory. We begin by examining some of the convexity properties of the cost and constraint functionals defined by (4.5) following which, this set of conditions for solvability are derived.
4.3. MAIN RESULTS

4.3.1 Conditions for convexity

Let \( X \subset L^2_T(0, T; \mathbb{R}^n) \times L^2_T(0, T; \mathbb{R}^m) \) be the set of admissible pairs for the stochastic differential equation (4.1); that is

\[
X = \left\{ (x(\cdot), u(\cdot)) \in L^2_T(0, T; \mathbb{R}^n) \times L^2_T(0, T; \mathbb{R}^m) : u(\cdot) \in U_{ad} \right\} (4.19)
\]

and \((x(\cdot), u(\cdot))\) satisfy (4.1) \( (4.20) \)

Let \( X_0 \subset L^2_T(0, T; \mathbb{R}^n) \times L^2_T(0, T; \mathbb{R}^m) \) be defined by

\[
X_0 = \left\{ (x(\cdot), u(\cdot)) \in L^2_T(0, T; \mathbb{R}^n) \times L^2_T(0, T; \mathbb{R}^m) : u(\cdot) \in U_{ad} \right\} (4.21)
\]

and the process \( \tilde{x}(\cdot) \in L^2_T(0, T; \mathbb{R}^n) \) be the solution of (4.1) corresponding to \( u(t) = 0 \) a.e. on \([0, T]\); that is

\[
d\tilde{x}(t) = (A(t)x(t) + B(t)u(t)) dt + D(t)u(t) dW(t), \quad \tilde{x}(0) = 0
\]

(4.22)

It follows then that \( X_0 \) is a linear subspace, \( x(\cdot) = (\tilde{x}(\cdot), 0) \) an element of \( L^2_T(0, T; \mathbb{R}^n) \times L^2_T(0, T; \mathbb{R}^m) \) and \( X = x(\cdot) + X_0 \). That is, \( X \) is an affine subspace of \( L^2_T(0, T; \mathbb{R}^n) \times L^2_T(0, T; \mathbb{R}^m) \).

We are now in the position to determine the conditions under which the functionals \( J_i(0, y; u(\cdot)) \) are convex. In the preceeding discussion, we have shown that the class of admissible pairs \( X = z + X_0 \) is an affine subspace. Therefore, the domain \( X \) of \( J_i(0, y; u(\cdot)) \) is convex, as required.

Let \( \lambda \in [0, 1] \) and suppose that \((x_1(\cdot), u_1(\cdot)), (x_2(\cdot), u_2(\cdot)) \in X \) are given admissible pairs. We now derive a set of conditions under which the inequality

\[
\lambda J_i(0, y; u_1(\cdot)) + (1 - \lambda) J_i(0, y; u_2(\cdot)) \geq J_i(0, y; \lambda u_1(\cdot) + (1 - \lambda) u_2(\cdot)) (4.23)
\]

holds. In fact, it can be shown (after some simple algebra) that

\[
\lambda J_i(0, y; u_1(\cdot)) + (1 - \lambda) J_i(0, y; u_2(\cdot)) - J_i(s, y; \lambda u_1(\cdot) + (1 - \lambda) u_2(\cdot)) = \lambda (1 - \lambda) \tilde{J}_i(0, 0; \tilde{u}(\cdot)) (4.24)
\]

where \( \tilde{u}(\cdot) \in U_{ad} \) is given by \( \tilde{u}(t) = u_1(t) - u_2(t) \) and \( \tilde{x}(t) = x_1(t) - x_2(t) \). Note that \((\tilde{x}(\cdot), \tilde{u}(\cdot)) \in X_0 \). The functional \( \tilde{J}_i(0, 0; \tilde{u}(\cdot)) \) is given by

\[
\tilde{J}_i(0, 0; \tilde{u}(\cdot)) = E \left\{ \frac{1}{2} \int_0^T \tilde{x}'(t) Q_i(t) \tilde{x}(t) + \tilde{u}'(t) R_i(t) \tilde{u}(t) dt + \frac{1}{2} \tilde{x}'(T) M_i \tilde{x}(T) \right\} (4.25)
\]
with \((\bar{x}(\cdot), \bar{u}(\cdot)) \in X_0\). It follows from (4.24) and the definition (4.23) of convexity that \(J_i(0, y; u(\cdot))\) is a convex functional defined on \(X\) if and only if \(\bar{J}_i(0, 0; \bar{u}(\cdot)) \geq 0\) for every \((\bar{x}(\cdot), \bar{u}(\cdot)) \in X_0\). At this stage, we note several things. First, if there exists \((x(\cdot), u(\cdot)) \in X_0\) such that \(J_i(0, 0; u(\cdot)) < 0\), then \(J_i(0, 0; u(\cdot))\) is unbounded on \(X_0\) since for every \(k \in \mathbb{R}\), \((k \bar{x}(\cdot), k \bar{u}(\cdot)) \in X_0\) and \(\bar{J}_i(0, 0; k \bar{u}(\cdot)) = k^2 \bar{J}_i(0, 0; \bar{u}(\cdot))\) which can be made arbitrarily negative by choosing \(|k|\) sufficiently large. Therefore, \(J_i(0, y; u(\cdot))\) is convex over \(X\) if and only if the optimal control problem

\[
\min \bar{J}_i(0, 0; \bar{u}(\cdot))
\]

\[
dx(t) = (A(t)x(t) + B(t)u(t)) \, dt + D(t)u(t) \, dW(t), \quad x(0) = 0
\]  

(4.26)
is well posed (with a necessarily non-negative infimal cost). Note that (4.26) is an optimal LQG control problem with a possibly indefinite control weight, as considered in [58] and for which the main results have been summarized in Section 4.2.

Second, the infimal cost associated with (4.26) is \(J_i(0, 0; u(\cdot)) = 0\) and is achieved by the admissible pair \((\bar{x}(\cdot), \bar{u}(\cdot)) \in X_0\) where \(\bar{x}(t) = 0\) and \(\bar{u}(t) = 0\) a.e. on \([0, T]\). Note however that since \((\bar{x}(t), \bar{u}(t)) = (x_1(t) - x_2(t), u_1(t) - u_2(t)) \in X_0\) for every \((x_1(\cdot), u_1(\cdot)), (x_2(\cdot), u_2(\cdot)) \in X\), it follows that \(J_i(0, y; u(\cdot))\) is strictly convex over \(X\) if and only if the admissible pair \((\bar{x}(\cdot), \bar{u}(\cdot)) \in X_0\) is the unique solution of (4.26).

To explore these issues further, consider the Riccati equations:

\[
\begin{cases}
\dot{P} + PA + A'P - PB(R_i + D'PD)^{-1}B'P + Q_i = 0 \\
P(T) = M_i \\
R_i + D'PD > 0
\end{cases}
\]  

(4.27)

and

\[
\begin{cases}
\dot{P} + PA + A'P - PBK^{-1}B'P + Q_i = 0 \\
P(T) = M_i \\
R_i + D'PD > 0
\end{cases}
\]  

(4.28)

where \(K \in \mathcal{K}\) with \(\mathcal{K}\) defined as in (4.14). As discussed in Section 4.2, (4.28) admits a unique solution \(P \in C(0, T; S^n_+\)) for every \(K \in \mathcal{K}\) and we can define a mapping \(\Psi_i : \mathcal{K} \rightarrow C(0, T; S^n)\) which maps every \(K \in \mathcal{K}\) to the solution \(P = \Psi_i(K)\) of the Riccati equation (4.28) associated with \(K\). This gives rise to a sufficient condition for the strict convexity of the functional \(J_i(0, y; u(\cdot))\) over \(X\).

**Proposition 4.1** If there exists \(K_i \in \mathcal{K}\) such that \(R_i + D'\Psi_i(K_i)D \geq K_i\), then \(J_i(0, y; u(\cdot))\) is a strictly convex functional over the class of admissible controls \(U_{ad}\).

**Proof:** If there exists \(K_i \in \mathcal{K}\) such that \(R_i + D'\Psi_i(K_i)D \geq K_i\), then by Theorem 4.3, the Riccati equation admits a solution \(P_i \in C(0, T; S^n_+)\). By Theorem 4.2, \(P_i\) is
4.3. MAIN RESULTS

the unique solution of (4.27). From Theorem 4.1, it then follows that the optimal control problem (4.26) is well posed and has a minimal cost of $J_i(0, 0; \bar{u}^*(\cdot)) = 0$. Therefore, $J_i(0, y; u(\cdot))$ is convex over $U_{ad}$.

To prove strict convexity, we show that the unique optimal solution of (4.26) is $\bar{u}^*(t) = 0$ (a.e.). This follows by noting that if $P_i$ is a solution of (4.27), then $\bar{K}_i = R_i + D'P_iD > 0$. Moreover, by the completion of squares technique, we can write the cost functional $\bar{J}_i(0, 0; u(\cdot))$ as

$$\bar{J}_i(0, 0; u(\cdot)) = E \left\{ \frac{1}{2} \int_0^T \left[ u(t) + \bar{K}_i^{-1}B'(t)P_i(t)z(t) \right] \bar{K}_i \left[ u(t) + \bar{K}_i^{-1}B(t)P_i(t)z(t) \right] dt \right\} \quad (4.29)$$

Since $\bar{K}_i > 0$, there is a unique optimal control for (4.29) and hence for (4.26), namely $\bar{u}^*(t) = -\bar{K}_i^{-1}(t)B'(t)P_i(t)z(t)$ a.e. on $[0, T]$ which achieves the optimal cost $\bar{J}_i(0, 0; \bar{u}^*(\cdot)) = 0$. Moreover, the corresponding state process satisfies $\bar{x}^*(t) = 0$ a.e. on $[0, T]$. Therefore, $\bar{u}^*(t) = 0$ a.e. on $[0, T]$ is the unique optimal control, from which we conclude that $J_i(0, y; u(\cdot))$ is strictly convex over $U_{ad}$.

At this point, it is convenient to make several observations which are relevant to the condition in Proposition 4.1 for strict convexity of $J_i(0, y; u(\cdot))$. First, if $R_i > 0$, then this condition is automatically satisfied. Indeed, by choosing $K_i = R_i$, we obtain

$$R_i + D'\Psi_i(K_i)D = R_i + D'\Psi_i(R_i)D \geq R_i = K_i$$

since $\Psi_i(R_i)$ is always positive definite. Second, given any $K_i \in \mathcal{K}$, we can always choose $D$ such that the condition $R_i + D'\Psi(K_i)D \geq K_i$ is satisfied. In fact, by choosing $D$ ‘sufficiently large’, this condition will be satisfied and the cost functional $J_i(0, y; u(\cdot))$ will be convex over $U_{ad}$. To clarify this point, consider the following example:

$$J = E \left\{ \frac{1}{2} \int_0^1 (x^2(t) + r(t)u(t)) \, dt + \frac{1}{2}x^2(1) \right\}$$

$$dx(t) = D(t)u(t) \, dW(t), \quad x(0) = 0$$

For this problem, the Riccati equation (4.28) is

$$\dot{P} + 1 = 0$$

$$P(1) = 1$$

so for any $K \in \mathcal{K}$, the corresponding solution $P = \Psi_i(K)$ is $P(t) = 2 - t$. Note that $P(t)$ is positive definite for each $t \in [0, 1]$. By Proposition 4.1, $J$ is convex over $U_{ad}$ if
there exists \( K \in \mathcal{K} \) such that
\[
r(t) + D^2(t) (2 - t) \geq K(t)
\]
For any given \( K \), this condition will be satisfied if \( D(t) \) is 'sufficiently large'. For this problem, 'sufficiently large' means
\[
D^2(t) \geq \frac{K(t) - r(t)}{2 - t}
\]
Finally, the result in Proposition 4.1 shows that the condition (4.18) for solvability of the unconstrained LQG problem with an indefinite control weight (as studied in [58]) is actually a condition for strict convexity of the cost functional associated with the LQG problem.

### 4.3.2 Optimal control

By Proposition 4.1, we have a sufficient condition which guarantees that the cost and constraint functionals \( J_i(0, y; u(\cdot)) \) for \( i = 0, \cdots, m \) are convex; namely, for every \( i = 0, \cdots, m \) there exists \( K_i \in \mathcal{K} \) such that \( R_i + D^i \Psi_i(K_i) D \geq K_i \). In fact, under this condition, the problem (4.7) can be solved explicitly, and the optimal control determined.

To begin, we give sufficient conditions under which the problem (4.7) has a unique optimal control. The reader may recall from Theorem 2.2 in Section 2.1 that the Lagrange Duality Theorem does not guarantee the existence and uniqueness of an optimal control. However, under the following assumption, we can prove this result.

**Assumption 4.1** There exists \( (x(\cdot), u(\cdot)) \in \mathcal{X} \) which is feasible for (4.7).

Under this assumption, we have the following existence and uniqueness result for (4.7).

**Theorem 4.4** Suppose that Assumption 4.4 holds and that for every \( i \in \{0, \cdots, m\} \), there exists \( K_i \in \mathcal{K} \) such that \( R_i + D^i \Psi_i(K_i) D \geq K_i \). Then (4.7) has a unique optimal control \( u^*(\cdot) \in U_{ad} \).

**Proof:** Let \( u(\cdot) \in U_{ad} \) and \( (x(\cdot), u(\cdot)) \in \mathcal{X} \) be the corresponding admissible pair. Since \( \mathcal{X} = z(\cdot) + X_0 \) where \( X_0 \) and \( z(\cdot) = (\bar{z}(\cdot), 0) \) are given by (4.21) and (4.22)
4.3. MAIN RESULTS

respectively, it follows that we can decompose $x(t) = \bar{x}(t) + \bar{z}(t)$ where $\bar{x}(\cdot) \in X_0$. Making this decomposition, (4.7) can easily be shown to be equivalent to the following optimization problem over $X_0$:

$$
\begin{array}{l}
\min f_0(x(\cdot), u(\cdot)) \\
f_1(x(\cdot), u(\cdot)) \leq c_1 \\
\vdots \\
f_m(x(\cdot), u(\cdot)) \leq c_m \\
(x(\cdot), u(\cdot)) \in X_0
\end{array}
$$

where

$$f_i(x(\cdot), u(\cdot)) = E \left\{ \frac{1}{2} \int_0^T [x'(t) Q_i(t) x(t) + 2 q_i(t) x(t) + u'(t) R_i u(t)] dt \\
+ \frac{1}{2} [x'(T) M_i x(T) + 2 m_i x(T)] \right\} + E \left\{ \frac{1}{2} \int_0^T \bar{z}'(t) Q_i(t) \bar{z}(t) dt + \frac{1}{2} \bar{z}'(T) M_i \bar{z}(T) \right\}
$$

and $q_i(t) = Q_i(t) \bar{z}(t)$, $m_i(t) = M_i \bar{z}(T)$. By the assumptions of the theorem, it follows from Proposition 4.1 that $J_i(0, y; u(\cdot))$ are strictly convex functionals on $X$ and hence, $f_i(x(\cdot), u(\cdot))$ are strictly convex on $X_0$ for each $i$. Thus, (4.30) is a strictly convex optimization problem on the Hilbert space $X_0$. By Assumption 4.1, there exists $(\hat{x}(\cdot), \hat{u}(\cdot)) \in X_0$ which is feasible for (4.30). Since $f_i(x(\cdot), u(\cdot))$ are strictly convex on $X_0$ for all $i$, it follows from Assumption 4.1 that the constraints of (4.30) define a non-empty, bounded, closed, convex subset of the Hilbert space $X_0$. Since every continuous, convex functional defined on a Hilbert space achieves its minimum on every non-empty, bounded, closed, convex set [3, Theorem 2.6.1], it follows that there exists $(x^*(\cdot), u^*(\cdot)) \in X_0$ which is optimal for (4.30). Furthermore, the uniqueness of $(x^*(\cdot), u^*(\cdot))$ follows from the strict convexity of $f_0(x(\cdot), u(\cdot))$. Clearly, $z(\cdot) + (x^*(\cdot), u^*(\cdot)) \in X$ is the unique optimal admissible pair for (4.7).

Under the assumption that the so-called Slater condition is satisfied, duality theory can be used to determine the optimal control when the cost and each of the constraint functionals $J_i(0, y; u(\cdot))$ satisfy the condition for strict convexity, as stated in Proposition 4.1. The Slater condition, which we shall assume, is stated as follows:

**Assumption 4.2 (Slater condition)** For every $\lambda_i \geq 0$, $i = 1, \ldots, m$ (not all equal
to zero), there exists an admissible pair \((x(\cdot), u(\cdot))\) \(\in X\) such that
\[
\sum_{i=1}^{m} \lambda_i (J_i(0, y; u(\cdot)) - c_i) < 0
\]

**Remark 4.1** A sufficient condition for Assumption 4.2 to hold is existence of an admissible pair \((\bar{x}(\cdot), \bar{u}(\cdot))\) \(\in X\) such that \(J_i(0, y; \bar{u}(\cdot)) < c_i\), for every \(i = 1, \ldots, m\).

To begin, for every \(\lambda \geq 0\), we define the Lagrangian as follows:
\[
L(u(\cdot), \lambda) = J_0(0, y; u(\cdot)) + \sum_{i=1}^{m} \lambda_i (J_i(0, y; u(i)) - c_i)
\]
(4.31)
where \((x(\cdot), u(\cdot))\) \(\in X\). It follows that
\[
L(u(\cdot), \lambda) = E \left\{ \frac{1}{2} \int_0^T (x'(t) Q(\lambda)(t) x(t) + u'(t) R(\lambda)(t) u(t)) \, dt + \frac{1}{2} x'(T) M(\lambda) x(T) \right\} - \lambda c
\]
where \(Q(\lambda) = Q_0 + \sum_{i=1}^{m} \lambda_i Q_i\), with \(R(\lambda)\) and \(M(\lambda)\) defined similarly and \(x(\cdot)\) is the solution of the stochastic differential equation (4.1) corresponding to \(u \in U_{ad}\). Note also that for any given \(\lambda \geq 0\), \(Q(\lambda) \geq 0\) and \(M(\lambda) \geq 0\) while \(R(\lambda)\) may be positive, negative or indefinite. Given any \(\lambda \geq 0\), we have the associated dual functional:
\[
\varphi(\lambda) = \inf_{u(\cdot) \in U_{ad}} L(u(\cdot), \lambda)
\]
(4.32)
This is an LQG problem with an indefinite control weighting matrix in the cost, as discussed in Section 4.2 and studied in [58]. Hence, associated with the minimization in (4.32) is the following (\(\lambda\)-dependent) Riccati equation:
\[
\begin{align*}
\dot{P} + PA + A'P - PB (R(\lambda) + D'PD)^{-1} B'P + Q(\lambda) &= 0 \\
P(T) &= M(\lambda) \\
R(\lambda) + D'PD &> 0
\end{align*}
\]
(4.33)
An important issue associated with the dual functional (4.32) is determining the values of \(\lambda \geq 0\) for which (4.32) exists. Indeed, the dual problem associated with (4.7) is an optimization problem over \(\lambda \geq 0\), where (4.32) is the cost functional. Since for any \(\lambda \geq 0\), calculating the value of \(\varphi(\lambda)\) is equivalent to solving an LQG problem with indefinite control weight, it follows from the discussion in Section 4.2 and the paper [58] that to address this issue, we need to consider the values of \(\lambda \geq 0\) for which there
exists solutions of the Riccati equation (4.33). In fact, we shall prove in Proposition 4.2 that if the cost and each constraint functional is convex (that is, if the conditions in Proposition 4.1 are satisfied for the cost and each constraint functional), then the Riccati equation (4.33) has a unique solution (and hence, (4.32) will be defined) for every $\lambda \geq 0$. However, in order to prove this result, we need the results stated in the following two lemmas which state some properties of Riccati equations.

To begin, let $(M, K, Q) \in S_{+}^{T} \times \mathcal{K} \times (L^{2}(0, T; S_{+}^{n}) \cap C(0, T; S_{+}^{n}))$. The associated Riccati equation is given by

$$
\begin{align*}
\dot{P} + PA + A'P - PBK^{-1}B'P + Q &= 0 \\
P(T) &= M
\end{align*}
$$

As in the earlier discussions of this Riccati equation, we can define a map $\Psi : S_{+}^{T} \times \mathcal{K} \times (L^{2}(0, T; S_{+}^{n}) \cap C(0, T; S_{+}^{n})) \rightarrow C(0, T; S^{n})$ which takes $(M, K, Q)$ to the solution $P = \Psi(M, K, Q)$ of (4.34) resulting from $(M, K, Q)$. The following two lemmas state some properties of this map $\Psi$, and are required to investigate the existence of solutions of (4.7) and the well posedness of (4.32).

**Lemma 4.1** $\Psi$ is a concave mapping from $S_{+}^{T} \times \mathcal{K} \times (L^{2}(0, T; S_{+}^{n}) \cap C(0, T; S_{+}^{n}))$ to $C(0, T; S^{n})$.

**Proof:** Note first that $S_{+}^{T} \times \mathcal{K} \times (L^{2}(0, T; S_{+}^{n}) \cap C(0, T; S_{+}^{n}))$ is a convex set.

Given (4.34), we have the following standard deterministic LQG control problem.

$$
\begin{align*}
\min \int_{s}^{T} \left( x'(t)Q(t)x(t) + u'(t)K(t)u(t) \right) dt + \frac{1}{2}x'(T)Mx(T) \\
x(t) &= A(t)x(t) + B(t)u(t), \quad x(s) = y
\end{align*}
$$

Let $\lambda \in [0, 1]$ and $(M_1, K_1, Q_1), (M_2, K_2, Q_2) \in S_{+}^{T} \times \mathcal{K} \times (L^{2}(0, T; S_{+}^{n}) \cap C(0, T; S_{+}^{n}))$ be given. Suppose that

$$
\begin{align*}
P &= \Psi[\lambda(M_1, K_1, Q_1) + (1 - \lambda)(M_2, K_2, Q_2)] \\
P_1 &= \Psi(M_1, K_1, Q_1) \\
P_2 &= \Psi(M_2, K_2, Q_2)
\end{align*}
$$

Then

$$
\begin{align*}
y'P(s)y \\
&= \min_{u(.) \in U_{ad}} \left\{ \frac{1}{2} \int_{s}^{T} \left[ x'(t)(\lambda Q_1 + (1 - \lambda)Q_2)x(t) + u'(t)(\lambda K_1 + (1 - \lambda)K_2)u(t) \right] dt \right\}
\end{align*}
$$
CHAPTER 4. INDEFINITE CONSTRAINED LQG CONTROL

\[ + \frac{1}{2} z'(T)[\lambda M_1 + (1 - \lambda) M_2] z(T) \]

\[ = \min_{u(\cdot) \in \mathcal{U}_{ad}} \left\{ \lambda \left( \frac{1}{2} \int_s^T [z'(t) Q_1(t) z(t) + u'(t) K_1(t) u(t)] \, dt + \frac{1}{2} z'(T) M_1 z(T) \right) + \frac{1}{2} z'(T) M_2 z(T) \right\} \]

\[ + (1 - \lambda) \left( \frac{1}{2} \int_s^T [z'(t) Q_2(t) z(t) + u'(t) K_2(t) u(t)] \, dt + \frac{1}{2} z'(T) M_2 z(T) \right) \]

\[ \geq \lambda \min_{u(\cdot) \in \mathcal{U}_{ad}} \left\{ \frac{1}{2} \int_s^T [z'(t) Q_1(t) z(t) + u'(t) K_1(t) u(t)] \, dt + \frac{1}{2} z'(T) M_1 z(T) \right\} \]

\[ + (1 - \lambda) \min_{u(\cdot) \in \mathcal{U}_{ad}} \left\{ \frac{1}{2} \int_s^T [z'(t) Q_2(t) z(t) + u'(t) K_2(t) u(t)] \, dt \right\} \]

\[ + \frac{1}{2} z'(T) M_2 z(T) \]

\[ \geq \lambda y' P_1(s) y + (1 - \lambda) y' P_2(s) y \]

Since this is true for every \( y \in \mathbb{R}^n \) and \( s \in [0, T] \), it follows that

\[ P(s) \geq \lambda P_1(s) + (1 - \lambda) P_2(s) \]

The result follows from the definition of \( \Psi(M, K, Q) \).

Lemma 4.2 The operator \( \Psi \) is continuous.

\[ y' \left( \frac{1}{2} \int_s^T (u + K_i^{-1} B' P_i x)' K_i (u + K_i^{-1} B' P_i x) \, dt \right) \]

Now

\[ y' P_1(s) y \]

\[ = 2 \min_{u(\cdot)} J_1(s, y; u(\cdot)) \]

\[ = 2 \min_{u(\cdot)} \left\{ J_0(s, y; u(\cdot)) + \frac{1}{2} \int_s^T \left[ z'((Q_1 - Q_0)x + u'(K_1 - K_0)u) \right] \, dt \right\} \]

\[ + \frac{1}{2} z'(T)(M_1 - M_0) z(T) \]

\[ = \min_{u(\cdot)} \left\{ \int_s^T \left[ z'((Q_1 - Q_0)x + u'(K_1 - K_0)u) \right] \right\} \]
4.3. MAIN RESULTS

\[ + (u + K_0^{-1} B' P_0 x)' K_0 (u + K_0^{-1} B' P_0 x) \right] \ dt \\
+ x'(T) (M_1 - M_0) x(T) \right] + y' P_0(s) y \]

\[ \leq y' P_0(s) y + \int_s^T \dot{x}' \Gamma \dot{x} \ dt + x'(T) (M_1 - M_0) x(T) \]

where

\[ \Gamma = (Q_1 - Q_0) + P_0 B K_0^{-1} (K_1 - K_0) K_0^{-1} B' P_0 \]

and \( \dot{x}(t) \) is the state trajectory resulting from (4.35) when the control law is \( \dot{u}(t) = -K_0(t)^{-1} B'(t) P_0(t) x(t) \). That is, \( \dot{x}(\cdot) \) is the solution of

\[ \dot{x} = (A - B K_0^{-1} B' P_0) x, \quad x(s) = y \]

(4.37)

and is given by \( \dot{x}(t) = \Phi(t, s) y \) where \( \Phi(t, s) \) is the transition matrix associated with (4.37).

To begin, note first that for any \( \Sigma \in S^n \) and \( y \in \mathbb{R}^n \),

\[ |y' \Sigma y| = |\text{tr} \Sigma Y|, \quad Y = y y \]

\[ \leq ||\Sigma||_{S^n} ||Y||_{S^n} \quad \text{by the Cauchy-Schwarz inequality} \]

Since \( ||Y||_{S^n} = (\text{tr} Y Y')^{\frac{1}{2}} = |y|^2 \), it follows that

\[ |y' \Sigma y| \leq ||\Sigma||_{S^n} |y|^2 \]

(4.38)

Therefore, noting that there exists \( 0 \leq k < \infty \) such that

\[ 0 \leq \text{ess sup}_{0 \leq t \leq T} ||\Phi(t, s)||_{S^n} \leq k \]

and (4.38), we have

\[ |y'(P_1(s) - P_0(s)) y| \]

\[ = | \int_s^T y' \Phi(t, s) \Gamma(t) \Phi(t, s) y \ dt + y' \Phi(T, s)' (M_1 - M_0) \Phi(T, s) y | \]

\[ \leq \int_s^T ||\Phi(t, s)||_{S^n} ||\Gamma(t)||_{S^n} |y|^2 \ dt + ||\Phi(T, s)||_{\mathcal{L}^\infty(0, T; S^n)} ||M_1 - M_0||_{S^n} |y|^2 \]

\[ \leq (T k \text{ ess sup}_{0 \leq t \leq T} ||\Gamma(t)||_{S^n} + k ||M_1 - M_0||_{S^n}) |y|^2 \]

\[ = (T k ||\Gamma(t)||_{\mathcal{L}^\infty(0, T; S^n)} + k ||M_1 - M_0||_{S^n}) |y|^2 \]

\[ \leq (C_1||Q_1 - Q_0||_{C^\infty(0, T; S^n)} + C_2||K_1 - K_0||_{\mathcal{L}^\infty(0, T; S^n)} + C_3||M_1 - M_0||_{S^n}) |y|^2 \]

\[ \leq C (||Q_1 - Q_0||_{C^\infty(0, T; S^n)} + ||K_1 - K_0||_{\mathcal{L}^\infty(0, T; S^n)} + ||M_1 - M_0||_{S^n}) |y|^2 \]

\[ = C ||(M_1, K_1, Q_1) - (M_0, K_0, Q_0)|| |y|^2 \]
where \( C_1, C_2, C_3 \) are constants dependent on \((M_0, K_0, Q_0)\), \( C = \max \{ C_1, C_2, C_3 \} \) and \( \| \cdot \| \) is the metric on \( S^n \times \mathcal{K} \times (L^2(0,T;S^2_+)) \cap C(0,T;S^2_+) \) induced by those on \( S^n \), \( \mathcal{K} \) and \( L^2(0,T;S^2_+) \cap C(0,T;S^2_+) \); that is

\[
\|(M, K, Q)\| = \|M\|_{S^n} + \|K\|_{L^\infty(0,T;S^m)} + \|Q\|_{C^\infty(0,T;S^n)}
\]

Therefore

\[
\|P_1(s) - P_0(s)\| = \sup_{|y| \neq 0} \frac{|y'(P_1(s) - P_0(s))y|}{|y|^2} \leq C\|(M_1, K_1, Q_1) - (M_0, K_0, Q_0)\|
\]

Noting that \( C\|(M_1, K_1, Q_1) - (M_0, K_0, Q_0)\| \) is independent of \( s \in [0,T] \), it follows that

\[
\|\Psi(M_1, K_1, Q_1) - \Psi(M_0, K_0, Q_0)\| = \sup_{0 \leq s \leq T} \|P_1(s) - P_0(s)\| \leq C\|(M_1, K_1, Q_1) - (M_0, K_0, Q_0)\|
\]

from which the result follows.

This next result deals with the solvability of the Riccati equation (4.33) and hence, the existence of the dual functional \( \varphi(\lambda) \) for any given \( \lambda \geq 0 \). It shows that convexity of the cost and each of the constraint functionals is equivalent to convexity (and hence solvability) of the indefinite LQG problem associated with the dual functional (4.32) for every \( \lambda \geq 0 \). This result is important for what follows since the optimal control for (4.7) is determined by solving a finite dimensional optimization problem over \( \lambda \geq 0 \) for which (4.32) is the cost functional. The next result shows that convexity of the cost and constraint functionals is necessary and sufficient for the dual functional (4.32) to exist for every \( \lambda \geq 0 \).

**Proposition 4.2** Let \( K_i \in \mathcal{K} \) for every \( i \in \{0, \cdots, m\} \). Then

\[
R_i + D'\Psi_i(K_i) D \geq K_i
\]

if and only if for every \( \lambda = [\lambda_1, \cdots, \lambda_m] \geq 0 \)

\[
R(\lambda) + D'\Psi(M(\lambda), K(\lambda), Q(\lambda)) D \geq K(\lambda).
\]

Moreover, \( K(\lambda) \in \mathcal{K} \).

**Proof:** Let \( K_i \in \mathcal{K} \) be given. Let \( \lambda = [\lambda_1, \cdots, \lambda_m] \geq 0 \) and define

\[
\gamma = 1 + \sum_{i=1}^{m} \lambda_i > 0
\]
4.3. MAIN RESULTS

Let
\[ K(\lambda) = K_0 + \sum_{i=1}^{m} \lambda_i K_i, \quad Q(\lambda) = Q_0 + \sum_{i=1}^{m} \lambda_i Q_i, \quad M(\lambda) = M_0 + \sum_{i=1}^{m} \lambda_i M_i \]

Clearly \( K(\lambda) \in \mathcal{K} \). Suppose (4.39) holds, and consider the Riccati equation
\[
\begin{aligned}
\hat{P} + PA + A'P - PBK(\lambda)^{-1}B'P + Q(\lambda) &= 0 \\
P(T) &= M(\lambda)
\end{aligned}
\]  
(4.41)

which has the solution \( \Psi(M(\lambda), K(\lambda), Q(\lambda)) \). Dividing both sides of (4.41) by \( \gamma > 0 \) gives rise to the Riccati equation:
\[
\begin{aligned}
\hat{P} + PA + A'P - PB\hat{K}(\alpha)^{-1}B'P + \hat{Q}(\alpha) &= 0 \\
P(T) &= \hat{M}(\alpha)
\end{aligned}
\]  
(4.42)

where
\[
\hat{Q}(\alpha) = \sum_{i=0}^{m} \alpha_i Q_i \\
\alpha_i = \begin{cases} 
\frac{1}{\gamma}, & i = 0 \\
\frac{\alpha_i}{\gamma}, & i = 1, \ldots, m 
\end{cases}
\]

and similarly for \( \hat{K}(\alpha) \) and \( \hat{M}(\alpha) \). For any \( \lambda \geq 0 \) (and hence \( \gamma > 0 \)), the solution of (4.42) is given by \( \Psi(M(\alpha), K(\alpha), Q(\alpha)) \). By the concavity of \( \Psi(M, K, Q) \) with respect to \( (M, K, Q) \), it follows that
\[
\Psi(M(\alpha), K(\alpha), Q(\alpha)) \geq \sum_{i=0}^{m} \alpha_i \Psi(M_i, K_i, Q_i) = \sum_{i=0}^{m} \alpha_i \Psi_i(K_i)
\]  
(4.44)

Multiplying both sides of (4.44) by \( \gamma \), and noting that
\[
\gamma \Psi(M(\alpha), K(\alpha), Q(\alpha)) = \Psi(M(\lambda), K(\lambda), Q(\lambda))
\]
we get
\[
\Psi(M(\lambda), K(\lambda), Q(\lambda)) \geq \Psi_0(K_0) + \sum_{i=1}^{m} \lambda_i \Psi_i(K_i)
\]

Therefore
\[
R(\lambda) + D' \Psi(M(\lambda), K(\lambda), Q(\lambda)) D \\
\geq R_0 + D' \Psi_0(K_0) D + \sum_{i=1}^{m} \lambda_i (R_i + D' \Psi_i(K_i) D) \\
\geq K_0 + \sum_{i=1}^{m} \lambda_i K_i \\
= K(\lambda)
\]
To prove the converse result, suppose that (4.40) holds for every \( \lambda \geq 0 \). It follows immediately that (4.39) holds for \( i = 0 \). (Just choose \( \lambda = 0 \)). Suppose that there exists \( j \in \{1, \ldots, m\} \) such that (4.39) does not hold; that is

\[
R_j + D' \Psi_j(K_j) D < K_j
\]

(4.45)

Consider \( \bar{\lambda} = [\bar{\lambda}_1, \ldots, \bar{\lambda}_m] \) where \( \bar{\lambda}_i = 0 \) for \( i \neq j \) and \( \bar{\lambda}_j > 0 \). As before, we denote the solution of (4.34) with \( M(\bar{\lambda}) = M_0 + \bar{\lambda}_j M_j, K(\bar{\lambda}) = K_0 + \bar{\lambda}_j K_j \) and \( Q(\bar{\lambda}) = Q_0 + \bar{\lambda}_j Q_j \), by \( \Psi(M(\bar{\lambda}), K(\bar{\lambda}), Q(\bar{\lambda})) \). Dividing both sides of this Riccati equation by \( \bar{\lambda}_j \) gives rise to another Riccati equation of the form (4.34) with solution \( \Psi(\bar{M}, \bar{K}, \bar{Q}) \) where

\[
\begin{align*}
\bar{M} &= \frac{1}{\bar{\lambda}_j} M_0 + M_j, \\
\bar{K} &= \frac{1}{\bar{\lambda}_j} K_0 + K_j, \\
\bar{Q} &= \frac{1}{\bar{\lambda}_j} Q_0 + Q_j
\end{align*}
\]

Note that

\[
\Psi(M(\bar{\lambda}), K(\bar{\lambda}), Q(\bar{\lambda})) = \bar{\lambda}_j \Psi(\bar{M}, \bar{K}, \bar{Q})
\]

Moreover, since \( \bar{M} \rightarrow M_j, \bar{K} \rightarrow K_j \) and \( \bar{Q} \rightarrow Q_j \) as \( \bar{\lambda}_j \rightarrow \infty \), it follows from the continuity of \( \Psi \) with respect to \( (M, K, Q) \) (see Lemma 4.2) that \( \Psi(\bar{M}, \bar{K}, \bar{Q}) \rightarrow \Psi(M_j, K_j, Q_j) = \Psi_j(K_j) \) as \( \lambda_j \rightarrow \infty \). Therefore

\[
\bar{R} + D' \Psi(\bar{M}, \bar{K}, \bar{Q}) D \rightarrow R_j + D' \Psi_j(K_j) D < K_j
\]

where \( \bar{R} = \frac{1}{\bar{\lambda}_j} R_0 + R_j \). Hence, there exists a constant \( c \geq 0 \) such that

\[
\bar{R} + D' \Psi(\bar{M}, \bar{K}, \bar{Q}) D < K_j
\]

(4.46)

for all \( \bar{\lambda}_j \geq c \). Multiplying both sides of (4.46) by this \( \bar{\lambda}_j \), and noting that \( K_0 > 0 \), we have

\[
R(\bar{\lambda}) + D' \Psi(M(\bar{\lambda}), K(\bar{\lambda}), Q(\bar{\lambda})) D < \bar{\lambda}_j K_j < K(\bar{\lambda})
\]

which is a contradiction to (4.40). This proves the converse result.

An immediate consequence of Proposition 4.2 is the following Theorem. It gives a set of conditions that guarantee the existence and uniqueness of solutions of the Riccati equation (4.33) for every \( \lambda \geq 0 \). In turn, this gives us the value of the dual functional (4.32). This is summarized as follows.

**Theorem 4.5** Suppose that for every \( i \in \{0, \ldots, m\} \), there exists \( K_i \in K \) such that

\[
R_i + D' \Psi_i(K_i) D \geq K_i.
\]

Then, for every \( \lambda \geq 0 \), the Riccati equation (4.39) has a unique solution \( P \in C(0, T; S^n) \) and the value of the dual functional (4.32) is

\[
\varphi(\lambda) = \frac{1}{2} y' P(0) y - \lambda' c
\]

(4.47)
4.3. MAIN RESULTS

Proof: By Proposition 4.2, it follows that for every $\lambda \geq 0$, we have $K(\lambda) \in K$ and $R(\lambda) + D'\Psi(M(\lambda), K(\lambda), Q(\lambda))D \geq K(\lambda)$. Hence, by Theorems 4.2 and 4.3, there exists a unique solution $P \in C(0, T; S^2_0)$ of (4.33). Therefore, by Theorem 4.1, the optimal control problem associated with the dual cost functional (4.32) is well posed, with optimal cost given by (4.47).

By the Lagrange Duality Theorem (Theorem 2.2), the optimal control for (4.7) can be determined by solving the finite dimensional dual problem associated with (4.7). This can be stated as follows.

Theorem 4.6 Consider the full observation LQG problem with integral quadratic constraints (4.7). Suppose that for every $i \in \{0, \ldots, m\}$, there exists $K_i \in K$ such that $R_i + D'\Psi_i(K_i)D \geq K_i$. Then there exists $\lambda^* = [\lambda_1^*, \ldots, \lambda_m^*] \geq 0$ which is optimal for the problem

$$\begin{align*}
\max_{\lambda} \quad & \frac{1}{2} y' P(0) y - \lambda^t c \\
\text{s.t.} \quad & \dot{P} + PA + A'P - PB(R(\lambda) + D'PD)^{-1}B'P + Q(\lambda) = 0 \\
& P(T) = M(\lambda) \\
& R(\lambda) + D'PD > 0 \\
& \lambda \geq 0
\end{align*}$$

(4.48)

Furthermore, the unique optimal control for (4.7) is

$$u^*(t) = -(R(\lambda^*) + D'P^*D)^{-1}B'P^*x(t)$$

where $P^*(t)$ is the solution of the Riccati equation in (4.48) corresponding to $\lambda^*$.

Proof: Under the conditions of the theorem, it is a consequence of Proposition 4.1 that all the cost and constraint functionals $J_i(0, y; u(\cdot))$ are strictly convex. Therefore, (4.7) is a convex optimization problem over $U_{ad}$. The existence of a unique optimal control follows from Theorem 4.4 and the remainder of the result follows from the Lagrange Duality Theorem (Theorem 2.2) which is necessary and sufficient for optimality when the cost and constraint functionals in (4.7) are convex with respect to $u(\cdot)$, and Assumption 4.2 is satisfied.

Note that the dual problem (4.48) is concave in $\lambda$ (since the dual problem associated with a convex primal problem is always concave; see Theorem 2.1 in Section 2.1). Moreover, the dual problem (4.48) is a finite-dimensional optimization problem that can be solved using gradient-type optimization algorithms, so long as the gradient of
the dual cost functional (4.47) with respect to $\lambda$ can be calculated. These gradients are derived in Section 3.7 for the general convex quadratic optimization problem with convex quadratic constraints. In the case of (4.48), the gradient of the cost functional can be calculated as follows.

**Theorem 4.7** Suppose that for every $i \in \{0, \cdots, m\}$, there exists $K_i \in K$ such that $R_i + D' \Psi_i(K_i) D \geq K_i$. Let $\lambda = [\lambda_1, \cdots, \lambda_m] \geq 0$ be given. Then

$$\frac{d\varphi(\lambda)}{d\lambda} = \begin{bmatrix} \frac{\partial \varphi(\lambda)}{\partial \lambda_1} & \cdots & \frac{\partial \varphi(\lambda)}{\partial \lambda_m} \end{bmatrix}$$

where

$$\frac{\partial \varphi(\lambda)}{\partial \lambda_i} = \text{tr} \left\{ \frac{1}{2} \int_0^T \dot{Q}_i(t) \Lambda(t) dt + \frac{1}{2} M_i \Lambda(T) \right\} - c_i$$

$$\dot{\Lambda}(t) = \bar{A}(t) \Lambda(t) + \Lambda(t) \bar{A}(t) + \bar{C}(t) \Lambda(t) \bar{C}(t)^T$$  \hspace{1cm} (4.49)

$$\Lambda(0) = y' y$$  \hspace{1cm} (4.50)

$$\bar{A} = A - B \left( R(\lambda) + D' P D \right)^{-1} B' P$$

$$\bar{C} = D \left( R(\lambda) + D' P D \right)^{-1} B' P$$

$$\dot{Q}_i = Q_i + P B \left( R(\lambda) + D' P D \right)^{-1} R_i \left( R(\lambda) + D' P D \right)^{-1} B' P$$

with $P$ denoting the solution of the Riccati equation (4.33).

**Proof:** As shown in Theorem 3.7, for convex quadratic optimization problems with convex quadratic constraints

$$\frac{\partial \varphi(\lambda)}{\partial \lambda_i} = J_i(0, y; \bar{u}^*(\cdot)) - c_i$$  \hspace{1cm} (4.51)

where $\bar{u}^*(\cdot)$ is the optimal control for the dual cost functional (4.32) corresponding to the given $\lambda \geq 0$. By the conditions of the theorem, it follows from Theorem 4.5 that for every $\lambda \geq 0$, (4.33) has a unique solution $P(t)$. Moreover, as stated in Theorem 4.1, the optimal control for (4.32) corresponding to this $\lambda \geq 0$ is given by

$$\bar{u}^*(t) = -(R(\lambda) + D' P D)^{-1} B' P x(t)$$  \hspace{1cm} (4.52)

Substituting (4.52) into (4.51) gives

$$\frac{\partial \varphi(\lambda)}{\partial \lambda_i} = E \left\{ \frac{1}{2} \int_0^T x'(t) \dot{Q}_i(t) x(t) dt + \frac{1}{2} x'(T) M_i x(t) \right\}$$

$$= \text{tr} \left\{ \frac{1}{2} \int_0^T \dot{Q}_i(t) \Lambda(t) dt + \frac{1}{2} M_i \Lambda(T) \right\} - c_i$$

where $\Lambda(t) = E[x(t) x'(t)]$. By the Ito differential rule, it is easily shown that $\Lambda(t)$ is given by the equation (4.50). \qed
4.4 Conclusion

In this chapter, we have studied the LQG control problem with integral quadratic constraints. In particular, we have shown that the assumptions made in Chapter 3, namely that the control weight in the cost functional and each of the constraint functionals is strictly positive definite, is not necessary for us to obtain the optimal control explicitly. Indeed, when the diffusion term in the state equation is dependent on the control term, the control weighting matrices may have negative eigenvalues. By showing that the solvability condition for the unconstrained LQG problem is actually the condition for the strict convexity of the cost functional, we have derived a set of sufficient conditions for which the optimal control for the constrained problem can be determined explicitly.
Chapter 5

Infinite quadratic programming

In recent years, it has become clear that interior point methods are an efficient tool for solving many classes of constrained optimization problems. The distinguishing feature of IPM's is an attractive theoretical computational complexity which has led (in many cases) to efficient numerical computer packages. In particular, a large class of IPM's have been shown to have a worse case polynomial complexity, meaning that the number of iterations required to obtain the optimal solution is bounded above by a polynomial in the problem size. In practice moreover, it is becoming accepted that many IPM's have superior performance when compared with other widely used algorithms (such as the Simplex method for LP), especially when the number of variables and constraints becomes large. Furthermore, interior point algorithms possess an additional property that is useful in many applications, this being that suboptimal iterates of the algorithm also satisfy the constraints of the problem being solved. For this reason, IPM's are well suited for engineering systems that require 'online' optimization. They can be stopped whenever a feasible solution is required and the current suboptimal iterate used with no fear of violating the constraints.

Recent papers by Faybusovich and Moore [11, 12] has seen further generalizations of IPM's which has resulted in new algorithms for solving constrained optimal control problems. By generalizing a certain path following algorithm to the case where the variable belongs to an infinite dimensional Hilbert space, interior point algorithms for solving LQ control problems with finitely many linear constraints [11] and finitely many integral quadratic constraints [12] are derived and studied. Interestingly, despite the set of variables being an infinite dimensional Hilbert space, the algorithms run in polynomial time!
In this chapter, we consider extensions of the algorithm derived in [11] for linearly constrained quadratic optimization. In particular, we consider extensions of the algorithm so that it can be used to solve quadratic optimization problems with infinitely many linear constraints where the variable may belong to a (generally infinite-dimensional) Hilbert Space. As shown in Section 2.3, this includes the class of LQ control problems with continuous state inequality constraints. Note however that we have not derived a complexity bound for this algorithm. This is an important open problem. Nevertheless, despite the absence of a complexity bound, IPM's are an appropriate method with which to approach this problem. Their efficiency in solving problems with many constraints suggests that it is not unreasonable to expect them to perform well.

This chapter is divided into two sections. In Section 5.1, we introduce the problem and the concept of a central path: a trajectory in the space of feasible solutions that plays a fundamental role in many IPM's. In particular, we show that the points lying on the central trajectory are smoothly parametrized by a real variable we call \( \beta \), and that points on this path converge to the optimal solution of the problem as \( \beta \to \infty \). In practice, points on the central path can not be determined exactly (since they can not be calculated in finite time). Rather, they are approximately determined for increasing values of \( \beta \). An approximation to the optimal solution is determined in this way. We show that every point on the central path corresponds to the optimal solution of an unconstrained convex optimization problem that can be solved using Newton’s Method. In fact, calculating the updates for the iterates in Newton’s Method (called ‘Newton Steps’) is the most complicated step in any IPM. In Section 5.2, we show how the algorithm can be applied to the LQ optimal control problem with infinitely many linear constraints. As mentioned above, this includes as a special case the situation when there are constraints of the form \( c_1(t) \leq x(t) \leq c_2(t) \) at every time instant. We focus in particular on the issue of calculating the Newton Step. We show that calculating the Newton Step is equivalent to solving an unconstrained LQ control problem together with an integral equation.

### 5.1 Central trajectory analysis

Let \((H, \langle \cdot, \cdot \rangle)\) be a Hilbert space and \(X\) a closed subspace of \(H\). Given \(a, a_i(\eta) \in H\), \(b_i(\eta) \in \mathbb{R}\) for all \(\eta \in [0, 1]\), \(i = 1, \ldots, N\) and a symmetric bounded strictly positive
5.1. CENTRAL TRAJECTORY ANALYSIS

operator $Q : H \rightarrow H$, consider the following optimization problem:

$$f(x) = (a, x) + \frac{1}{2} \langle Qx, x \rangle \rightarrow \min$$

$$\langle a_i(\eta), x \rangle \leq b_i(\eta), \quad \eta \in [0, 1], i = 1, \cdots N$$

$$x \in X$$

We assume throughout that $a_i(\eta)$ and $b_i(\eta)$ are continuous, measurable $H$-valued functions of $\eta$.

We denote the set of feasible points $x \in X$ determined by the constraints (5.2)-(5.3) by $P$. We shall assume throughout this paper that

$$\text{int}(P) = \{ x \in P : \langle a_i(\eta), x \rangle < b_i(\eta), \eta \in [0, 1], i = 1, \cdots, N \}$$

is non-empty.

Consider the following dual problem of (5.1)-(5.3):

$$h(x, v) = -\frac{1}{2} \langle Qx, x \rangle - \sum_{i=1}^{N} \int_{0}^{1} b_i(\eta)dv_i(\eta) \rightarrow \max$$

$$a + Qx + \sum_{i=1}^{N} \int_{0}^{1} a_i(\eta)dv_i(\eta) \in X^\perp$$

$v(t) \in BV[0, 1]$ a non-decreasing functional

where $BV[0, 1]$ is the set of all functionals of bounded variation on the domain $[0, 1]$. For functionals $g(t) \in C[0, 1]$ and $v(t) \in BV[0, 1]$, the integral

$$\int_{0}^{1} g(t)dv(t)$$

is the Riemann-Stieltjes integral of $g(t)$ with respect to $v(t)$. The integrals

$$\int_{0}^{1} b_i(\eta)dv_i(\eta)$$

in (5.5) should be understood in this way. On the other hand, recalling that $a_i(\eta)$ is a $H$-valued function on $[0, 1]$, the integrals

$$\int_{0}^{1} a_i(\eta)dv_i(\eta)$$
CHAPTER 5. INFINITE QUADRATIC PROGRAMMING

should be understood as generalized Riemann-Stieltjes integrals, as explained in Lemma 2.2. Also, we define the orthogonal complement $X^\perp$ of the subspace $X \in H$ as follows:

$$X^\perp = \{ x \in H : \langle x, y \rangle = 0 \text{ for every } y \in X \}$$

Every feasible solution for the dual problem (5.5)-(5.7) gives a lower bound for the optimal cost for the primal problem (5.1)-(5.3), and every feasible solution of the primal problem gives rise to an upper bound for the optimal cost to the dual problem. A precise statement of this is given as follows:

**Lemma 5.1** Suppose $x \in P$ and $(\bar{x}, v)$ satisfies (5.6)-(5.7). Then

$$f(x) - h(\bar{x}, v) = \frac{1}{2} \langle Q(x - \bar{x}), x - \bar{x} \rangle + \sum_{i=1}^{N} \int_{0}^{1} [b_i(\eta) - \langle a_i(\eta), x \rangle] dv(\eta) \geq 0$$

**Proof:** Since $x \in X$, it follows from (5.6)-(5.7) that

$$\left< a + Qx + \sum_{i=1}^{N} \int_{0}^{1} a_i(\eta) dv(\eta), x \right> = 0$$

By Lemma 2.2

$$\left< \int_{0}^{1} a_i(\eta) dv_i(\eta), x \right> = \int_{0}^{1} \langle a_i(\eta), x \rangle dv_i(\eta)$$

so it follows that

$$f(x) - h(\bar{x}, v) = \frac{1}{2} \langle Q(x - \bar{x}), x - \bar{x} \rangle + \sum_{i=1}^{N} \int_{0}^{1} b_i(\eta) dv_i(\eta)$$

$$- \left< \sum_{i=1}^{N} \int_{0}^{1} a_i(\eta) dv_i(\eta), x \right>$$

$$= \frac{1}{2} \langle Q(x - \bar{x}), x - \bar{x} \rangle + \sum_{i=1}^{N} \int_{0}^{1} [b_i(\eta) - \langle a_i(\eta), x \rangle] dv(\eta)$$

$$\geq 0$$

since $Q$ is strictly positive and $b_i(\eta) - \langle a_i(\eta), x \rangle \geq 0$ by the assumptions of the Lemma.

Consider the problem:

$$f_\beta(x) = \beta((a, x) + \frac{1}{2} \langle Qx, x \rangle) - \sum_{i=1}^{N} \int_{0}^{1} ln(b_i(\eta) - \langle a_i(\eta), x \rangle) d\eta \rightarrow \min (5.8)$$
5.1. CENTRAL TRAJECTORY ANALYSIS

Given any $\beta \geq 0$, the functional $f_\beta : \text{int}(P) \to \mathbb{R}$ is referred to as the internal penalized function, and is the combination of the cost functional (5.1) and the barrier function

$$\int_0^1 \ln(b_i(\eta) - \langle a_i(\eta), x \rangle) \, d\eta$$

associated with the constraints (5.3). It expresses the tradeoff between two objectives: minimizing the cost functional (5.1) and staying as far as possible from the boundary of the feasible set $P$ defined by (5.2)-(5.3). Given any $\beta \geq 0$, the optimal solution $x(\beta)$ of (5.8)-(5.9) defines a point in the set of points we refer to as the ‘central trajectory’.

In fact, the central trajectory comprises of the set of all optimal solutions of (5.8)-(5.9) for $\beta \geq 0$. Intuitively, one may suspect that points $x(\beta)$ on the central trajectory converge to the optimal solution of the primal problem (5.1)-(5.3) as $\beta \to \infty$. In this section we shall confirm this result. The next few results deal with some of the properties of the problem (5.8)-(5.9). We begin by clarifying the issue of existence and uniqueness of optimal solutions $x(\beta)$.

**Proposition 5.1** For any $\beta > 0$ the problem (5.8)-(5.9) has a unique solution $x(\beta) \in \text{int}(P)$

**Proof:** Let $\bar{x} \in \text{int}(P)$ and suppose that $f_\beta(\bar{x}) = \sigma$. Define

$$P_\sigma = \{x \in \text{int}(P) : f_\beta(x) \leq \sigma\}$$

$P_\sigma$ is a convex subset of $X$ so strong closure and weak closure of $P_\sigma$ are equivalent (see Theorem 8.1 in Section 8.1 in the Appendix). We first show that $P_\sigma$ is a strongly (and hence weakly) closed in $\text{int}(P)$. Indeed let $x_i, i = 1, 2, \cdots$ be a sequence of points in $P_\sigma$ and suppose that $x_i \to \bar{x} \in P$ as $i \to \infty$. If $\bar{x} \in \partial P = P \setminus \text{int}(P)$, then $f_\beta(x_i) \to \infty$ as $i \to \infty$. However, $f_\beta(x_i) \leq \sigma$ for all $i$ - a contradiction. Therefore, $\bar{x} \in \text{int}(P)$. We now show that $P_\sigma$ is bounded. The first Frechet derivative $Df_\beta(x)$ of the functional $f_\beta(x)$ at $x \in \text{int}(P)$ in the direction $\xi \in H$ is given by

$$Df_\beta(x) \cdot \xi = \beta \langle a + Qx, \xi \rangle + \sum_{i=1}^N \int_0^1 \frac{\langle a_i(\eta), \xi \rangle}{s_i(x, \eta)} \, d\eta$$

$$s(x, \eta) = b_i(\eta) - \langle a_i(\eta), x \rangle, \quad \eta \in [0, 1]$$

(5.10) (5.11)

The second Frechet derivative $D^2f_\beta(x)$ of $f_\beta$ in the directions $\xi, \gamma \in H$ is given by

$$D^2f_\beta(x) \cdot [\xi, \gamma] = \beta \langle Q\xi, \gamma \rangle + \sum_{i=1}^N \int_0^1 \frac{\langle a_i(\eta), \xi \rangle \langle a_i(\eta), \gamma \rangle}{s_i^2(x, \eta)} \, d\eta$$

(5.12)
The condition that $Q$ is strictly positive means

$$
\langle Q\xi, \xi \rangle \geq c \langle \xi, \xi \rangle, \quad \xi \in H
$$

(5.13)

for some $c > 0$. It follows from (5.12)-(5.13) that

$$
D^2 f_\beta(x) \cdot [\xi, \xi] \geq \langle Q\xi, \xi \rangle \geq c \langle \xi, \xi \rangle, \quad \xi \in H, \; x \in \text{int}(P)
$$

(5.14)

Hence (see [38])

$$
f_\beta(x) - f_\beta(\bar{x}) - Df_\beta(\bar{x}) \cdot (x - \bar{x}) \geq \frac{c}{2} \|x - \bar{x}\|^2
$$

(5.15)

for all $x \in \text{int}(P)$. If $x_i, \; i = 1, 2, \cdots$ is a sequence of points in $P_\sigma$ such that $\|x_i\| \to \infty$ as $i \to \infty$, it follows from (5.15) that $f_\beta(x_i) \to \infty$ - a contradiction. Therefore, $P_\sigma$ is bounded and hence, weakly compact. Since $f_\beta(x)$ is weakly semi-continuous, it attains its minimum on $P_\sigma$. Furthermore, $f_\beta(x)$ is strictly convex so this minimum is unique.

One consequence of this result is that a point $x(\beta)$ on the central path is defined for every $\beta \geq 0$.

Our next objective is to show that the points $x(\beta)$ on the central path are smoothly parametrized by $\beta \geq 0$. To do this however, we need the following alternative characterization of the central path. Let $\pi : H \to X$ be the orthogonal projection of $H$ onto $X$. Then $x(\beta)$, the optimal solution of (5.8)-(5.9) is uniquely determined by the condition

$$
\nabla f_\beta(x(\beta)) \in X^\perp
$$

(5.16)

where $\nabla f_\beta(x)$, the gradient of $f_\beta(x)$ is defined by

$$
Df_\beta(x) \cdot \xi = \langle \nabla f_\beta(x), \xi \rangle, \quad \xi \in H
$$

Clearly, (5.16) is equivalent to

$$
\pi \left[ \beta a + \beta Qx(\beta) + \sum_{j=1}^N \int_0^1 \frac{a_j(\eta)}{s_j(x(\beta), \eta)} \; d\eta \right] = 0
$$

(5.17)

For $\beta > 0$ consider the Legendre transform map $\psi_\beta : \text{int}(P) \to X$ defined by

$$
\psi_\beta(x) = \pi \left[ \beta Qx + \sum_{i=1}^N \int_0^1 \frac{a_i(\eta)}{s_i(x, \eta)} \; d\eta \right]
$$

(5.18)
5.1. CENTRAL TRAJECTORY ANALYSIS

It follows from (5.18) that (5.17) can be written as

$$\psi_\beta(x(\beta)) = -\beta \pi(a)$$  \hspace{1cm} (5.19)

Note that the Legendre transform (5.18) maps points $x(\beta)$ of the central trajectory onto a ray starting from the origin. The following result is needed to show the smoothness properties of the central path.

**Proposition 5.2** For each $\beta > 0$, the map $\psi_\beta(x)$ is a smooth isomorphism of $\text{int}(P)$ onto $X$ such that $\psi_\beta^{-1}(x)$ depends smoothly on $\beta$.

**Proof:** By Proposition 5.1, for any $c \in X$, the problem

$$-\langle c, x \rangle + \frac{\beta}{2} \langle Qx, x \rangle - \sum_{i=1}^{N} \int_{0}^{1} \ln(b_i(\eta) - \langle a_i(\eta), x \rangle) d\eta \rightarrow \min, \quad x \in \text{int}(P)$$  \hspace{1cm} (5.20)

has a unique solution $x_c \in \text{int}(P)$ and by (5.19), $\psi_\beta(x_c) = c$. Hence, $\psi_\beta$ is surjective. On the other hand, if $\psi_\beta(x_1) = \psi_\beta(x_2) = c$, and $x_1, x_2 \in \text{int}(P)$, then we have again by (5.19) that both $x_1$ and $x_2$ are solutions of (5.20). Hence by Proposition 5.1, $x_1 = x_2$. In other words, $\psi_\beta$ is injective. We now show that $\psi_\beta^{-1}$ is smooth. Observe that for every $\xi \in X, x \in \text{int}(P)$,

$$D\psi_\beta(x) \cdot \xi = \pi \left( \beta Q\xi + \sum_{i=1}^{N} \int_{0}^{1} \frac{\langle a_i(\eta), \xi \rangle}{s_i^2(x, \eta)} a_i(\eta) d\eta \right)$$  \hspace{1cm} (5.21)

If $D\psi_\beta(x) \cdot \xi = 0$, then

$$\beta Q\xi + \sum_{i=1}^{N} \int_{0}^{1} \frac{\langle a_i(\eta), \xi \rangle}{s_i^2(x, \eta)} a_i(\eta) d\eta \in X^\perp$$

or

$$\beta \langle Q\xi, \xi \rangle + \sum_{i=1}^{N} \int_{0}^{1} \frac{(a_i(\eta), \xi)^2}{s_i^2(x, \eta)} d\eta = 0$$

which implies that $\xi = 0$. Thus $D\psi_\beta(x)$ induces an injective map from $X$ into $X$.

It can be seen that

$$D\psi_\beta(x)|_X = \beta \pi \circ Q \circ i + \pi \left[ \sum_{i=1}^{N} \int_{0}^{1} \frac{a_i \otimes a_i}{s_i^2(x, \eta)} \right]$$  \hspace{1cm} (5.22)

where $i : X \to H$ is the canonical immersion. Since the operator $Q$ is strictly positive, (5.22) easily implies that $D\psi_\beta(x)|_X$ is surjective. We can now apply the
implicit function theorem to conclude that $\psi_{-1}^\beta$ is smooth and depends smoothly on $\beta$.

The next result shows that points on the central path are smoothly parametrized by $\beta \geq 0$.

**Corollary 5.1** The solution $x(\beta)$ of the problem (5.8)-(5.9) depends smoothly on $\beta$.

**Proof:** It follows from (5.19) that $x(\beta) = \psi_{-1}^\beta(-\beta \pi(a_0))$. The result is a consequence of Proposition 5.2.

Thus far, we have introduced the concept of the central path. For every $\beta \geq 0$, the optimization problem (5.8)-(5.9) yields a unique optimal solution $x(\beta)$, and the set of all optimal solutions defines the central path. In fact, we have also shown that the central path depends smoothly on $\beta \geq 0$.

We now address the problem of 'calculating' $x(\beta)$. Suppose that $\beta \geq 0$ is given and fixed. The optimization problem (5.8)-(5.9) is solved using Newton's method; that is, given any $x \in \text{int}(P)$ for (5.8)-(5.9), an update $y$ is chosen by minimizing the quadratic approximation of $f_\beta$ at $x$. More precisely, the quadratic approximation $q$ of $f_\beta$ is given by

$$q(y) = f_\beta(x) + \langle \nabla f_\beta(x), y - x \rangle + \frac{1}{2} \langle \gamma_\beta(x) \cdot (y - x), y - x \rangle$$

where $\gamma_\beta(x) : H \to H$ denotes the Hessian of $f_\beta(x)$, and is given by

$$\gamma_\beta(x) = \beta Q + \sum_{j=1}^{N} \int_{0}^{1} a_j(\eta) \otimes a_j(\eta) \frac{d}{dy} s_j^2(x, \eta) \, d\eta, \quad \beta > 0 \quad (5.23)$$

By minimizing $q(y)$ with respect to $y$, it follows that the update $y \in X$ is characterized by the condition

$$\nabla q(y) = \nabla f_\beta(x) + \gamma_\beta(x) \cdot (y - x) \in X^\perp$$

Assuming the existence of $\xi(x) \in X^\perp$ and $y \in X$ such that $\nabla q(y) = \xi(x) \in X^\perp$, it follows that

$$y = x - \gamma_\beta(x)^{-1} \cdot (\nabla f_\beta(x) - \xi(x))$$
The term \( \gamma \beta(x)^{-1} \cdot (\nabla f_\beta(x) - \xi(x)) \) is called the Newton step, and must satisfy the conditions

\[
\gamma \beta(x)^{-1} \cdot (\nabla f_\beta(x) - \xi(x)) \in X
\]
\[
\xi(x) \in X^\perp
\]
(5.24)

The problem of calculating the Newton step (5.24) is the most computationally demanding process in any interior point method. Our immediate aim is to clarify how (5.24) can be calculated. In fact, we will show that the problem of calculating (5.24) is equivalent to solving a minimum norm problem. In order to be able to do this however, we need to introduce the idea of a Riemannian metric.

Given a Hilbert space \((H, \langle \cdot, \cdot \rangle)\), let \( \gamma : H \to H \) be any strictly positive symmetric linear operator. Associated with \( \gamma \) is the function \( \langle \cdot, \cdot \rangle_\gamma : H \times H \to \mathbb{R} \) where \( \langle x, y \rangle_\gamma = \langle \gamma \cdot x, y \rangle \). It is easy to show that \( \langle \cdot, \cdot \rangle_\gamma \) satisfies the properties of an inner product and hence, \((H, \langle \cdot, \cdot \rangle_\gamma)\) is a Hilbert space. We say that \( \gamma \) determines a Riemannian metric on \( H \). Let \( X \) be a closed subspace of \( H \). Given a smooth function \( \varphi : X \to \mathbb{R} \), the gradient of \( \varphi \) evaluated at \( x \in X \) (with respect to the original inner product \( \langle \cdot, \cdot \rangle \)) is denoted by \( \nabla \varphi(x) \). \( \nabla \varphi(x) \) is the unique element in \( X \) for which the property

\[
D \varphi(x) \cdot \eta = \langle \nabla \varphi(x), \eta \rangle
\]
(5.25)

holds for every \( \eta \in X \), where \( D \varphi(x) : X \to \mathbb{R} \) is the Fréchet differential of \( \varphi \) at \( x \in X \). Note in particular that \( \nabla \varphi(x) \) depends on the inner product \( \langle \cdot, \cdot \rangle \). If the inner product \( \langle \cdot, \cdot \rangle_\gamma \) is used instead of \( \langle \cdot, \cdot \rangle \) in (5.25), then we have

\[
D \varphi(x) \cdot \eta = \langle \nabla \gamma \varphi(x), \eta \rangle_\gamma = \langle \gamma \cdot \nabla \gamma \varphi(x), \eta \rangle
\]
(5.26)

where \( \nabla \gamma \varphi(x) \) is referred to as the gradient of \( \varphi \) with respect to the Riemannian metric induced by \( \gamma \). Note that \( \nabla \varphi(x) \) and \( \nabla \gamma \varphi(x) \) are related in the following way:

\[
\nabla \varphi(x) - \gamma \cdot \nabla \gamma \varphi(x) \in X^\perp
\]

In fact, the Newton step (5.24) is the steepest descent direction of the cost functional \( f_\beta(x) \) with respect to a certain Riemannian metric, which we now show.

To begin, note that \( f_\beta(x) \) is strictly convex on the set \( \{ x \in H : \langle a_i(\eta), x \rangle < b_i(\eta), \ i = 1, \cdots, N \ \text{and} \ \eta \in [0,1] \} \). Therefore, for any \( \beta > 0 \) and \( x \in H \) satisfying \( \langle a_i(\eta), x \rangle < b_i(\eta) \), the Hessian \( \gamma \beta(x) : H \to H \) of \( f_\beta(x) \), as given by (5.23) is a strictly positive symmetric operator. Therefore, \( \gamma \beta(x) \) defines an inner product \( g_\beta(x; \cdot, \cdot) :
CHAPTER 5. INFINITE QUADRATIC PROGRAMMING

\( H \times H \to \mathbb{R} \) where

\[
g_\beta(x; \xi, \eta) = D^2 f_\beta(x) \cdot (\xi, \eta) = \langle \gamma_\beta(x) \cdot \xi, \eta \rangle
\]  

(5.27)

for any \( \xi, \eta \in H \). The conditions \( \langle a_i(\eta), x \rangle < b_i(\eta), \eta \in [0, 1] \) and \( i = 1, \ldots, N \) define an open subset of \( H \). We can think of \( \text{int}(P) \) as a submanifold of this open subset. Given any smooth functional \( \varphi \) defined on an open neighbourhood of \( \text{int}(P) \) in \( H \), let \( \nabla_\beta \varphi(x) \in X \) denote the gradient of \( \varphi \) relative to the Riemannian metric \( g_\beta \). That is, for every \( \eta \in X \) and \( x \in \text{int}(P) \)

\[
D\varphi(x) \cdot \eta = \langle \nabla \varphi(x), \eta \rangle = \langle \gamma_\beta(x) \cdot \nabla_\beta \varphi(x), \eta \rangle = g_\beta(x; \nabla_\beta \varphi(x), \eta), \nabla_\beta \varphi(x) \in X
\] (5.28)

The next result shows the relationship between \( \nabla \varphi(x) \) and \( \nabla_\beta \varphi(x) \).

**Proposition 5.3**

\[
\nabla_\beta \varphi(x) = \gamma_\beta(x)^{-1}(\nabla \varphi(x) - \xi(x))
\] (5.29)

where \( \xi(x) \) is the unique vector in \( X^\perp \) such that \( \nabla_\beta \varphi(x) \in X \).

**Proof:** This proof is exactly the same as in [11]. For completeness, we will include it here. Suppose that \( \xi(x) \in X^\perp \) is such that \( \nabla \varphi(x) \in X \). We have

\[
g_\beta(x; \xi, \nabla \varphi(x)) = \langle \gamma_\beta(x) \xi, \gamma_\beta(x)^{-1}(\nabla \varphi(x) - \xi(x)) \rangle \\
= \langle \xi, \nabla \varphi(x) - \xi(x) \rangle
\]

for \( \xi \in X \). To prove that there exists a unique \( \xi(x) \) satisfying this property, consider the following minimum norm optimization problem:

\[
\langle \gamma_\beta(x)^{-1}(\nabla \varphi(x) - \xi), (\nabla \varphi(x) - \xi) \rangle \to \min
\]

(5.30)

(5.31)

This is a minimum norm problem, and the unique solution satisfies the condition \( \nabla \varphi(x) - \bar{\xi} \perp X \) with respect to the Riemannian metric defined by \( g_\beta \); that is

\[
\langle \gamma_\beta(x)^{-1}(\nabla \varphi(x) - \bar{\xi}), z \rangle = 0 \text{ for every } z \in X^\perp
\]

That is

\[
\gamma_\beta(x)^{-1}(\nabla \varphi(x) - \bar{\xi}) \in X
\]
5.1. **CENTRAL TRAJECTORY ANALYSIS**

Hence $\xi = \xi(x)$. 

Note in particular that $\xi(x)$ in (5.29) is determined by solving the minimum norm problem (5.30)-(5.31). Moreover, if $\varphi(x) = f_\beta(x)$ in (5.29), it follows from Proposition 5.3 that there exists a unique $\xi(x) \in X$ (depending on $x$) such that (5.24) holds, and that the Newton step (5.24) coincides with the gradient of $f_\beta(x)$ with respect to the Riemannian metric induced by its Hessian $\gamma_\beta(x)$. Furthermore, the Newton step (5.24) is calculated by solving the minimum norm problem (5.30)-(5.31). As stated earlier, the problem of calculating the Newton step is the most computationally demanding process in an IPM. In Section 5.2, we shall look at this issue more closely.

The following result is required later when we derive a differential equation which is satisfied by the central path.

**Corollary 5.2** Under the assumptions of Proposition 5.2, we have

$$D\psi_\beta(x) \cdot \nabla_\beta \psi(x) = \pi(\nabla \varphi(x))$$  (5.32)

**Proof:** This is proven in [11] and we include it here for the sake of completeness.

By (5.21) we have

$$\nabla = D\psi_\beta(x) \cdot \nabla_\beta \psi(x) = \pi(\gamma_\beta(x) \nabla \varphi(x))$$

Using (5.29) we obtain

$$\nabla = \pi(\nabla \varphi(x) - \xi(x)) = \pi(\nabla \varphi(x))$$

because $\xi(x) \in X^\perp$ so $\pi(\xi(x)) = 0$. 

The following few results relate to the properties of points on the central path. In particular, they are used to show that the cost functional $f(x)$ given by (5.1) is an increasing function when evaluated on the central path. This is ultimately used to prove the convergence properties of points on the central path.

As shown earlier, the central path is a curve that is smoothly parametrized by $\beta \geq 0$. In fact, we can derive a differential equation which has as its solution the points lying on the central path. This is given as follows.
CHAPTER 5. INFINITE QUADRATIC PROGRAMMING

Proposition 5.4 Let \( \psi(x) = -\langle a, x \rangle - \frac{1}{2} \langle Qx, x \rangle \). then

\[
\frac{dx(\beta)}{d\beta} = \nabla_{\beta} \varphi(x(\beta)), \quad \beta > 0 \tag{5.33}
\]

Proof: Let \( \alpha(\beta) = \psi_{\beta}(x(\beta)) \). We have

\[
\frac{d\alpha}{d\beta}(\beta) = D\psi_{\beta}(x(\beta)) \cdot \frac{dx(\beta)}{d\beta} + \pi(Qx(\beta))
\]

On the other hand, by (5.19)

\[
\frac{d\alpha}{d\beta}(\beta) = -\pi(a)
\]

Hence

\[
D\psi_{\beta}(x(\beta)) \cdot \frac{dx(\beta)}{d\beta} = -\pi(a + Qx(\beta))
\]

Combining this with (5.32) and using the fact that \( D\psi_{\beta}(x) \) restricted to \( X \) is bijective (from Proposition 5.2, we arrive at our result.

This next result shows that the cost functional \( f(x) \) is a monotonically non-increasing function when evaluated on the central path for increasing values of \( \beta \geq 0 \). This suggests that by solving (5.8)-(5.9) for increasing values of \( \beta \geq 0 \), improved estimates of the optimal solution of (5.1)-(5.3) may be obtained. This idea forms the basis of this (and many other) path following algorithms.

Corollary 5.3 Let

\[
f(x) = \langle a, x \rangle + \frac{1}{2} \langle Qx, x \rangle \tag{5.34}
\]

Then the function \( f(x(\beta)) \) is a monotonically non-increasing function of \( \beta \) for \( \beta > 0 \).

Proof: This result is proven in [11]. We include it here for the sake of completeness.

We have

\[
\frac{d}{d\beta} f(x(\beta)) = \left\langle \nabla f(x(\beta)), \frac{dx(\beta)}{d\beta} \right\rangle
\]

\[
= g_{\beta}(x(\beta); \nabla f(x(\beta)), \frac{dx(\beta)}{d\beta})
\]

\[
= -g_{\beta}(x(\beta); \nabla f(x(\beta)), \nabla f(x(\beta)))
\]

\[
\leq 0
\]
5.1. CENTRAL TRAJECTORY ANALYSIS

We turn our attention now to the dual problem (5.5)-(5.7) and in particular, properties of the dual cost functional \(h(x, v)\) on the central path. Recall that the cost functional \(f(x)\) of the primal problem (5.1)-(5.3) is a non-increasing function function on the central path. We now show that for every point \(x(\beta)\) on the central path, we can find a feasible dual variable \(v(\beta)\) for (5.5)-(5.7) such that \(h(x(\beta), v(\beta))\) is a non-decreasing functional of \(\beta \geq 0\).

To begin, we shall examine the dual problem (5.5)-(5.7). In particular, we wish to draw the reader’s attention to the constraints defined by (5.6)-(5.7). The dual space is the class of functions of bounded variation, \(BV^N[0, 1]\). In the analysis which follows, it is in fact more convenient to restrict the class of dual functionals to the subset of functionals in \(BV^N[0, 1]\) which are continuous and once differentiable. More precisely, we consider the following (approximate) dual problem in place of the original dual problem (5.5)-(5.7).

\[
h(x, v) = -\frac{1}{2} \langle Q x, x \rangle - \sum_{j=1}^{N} \int_{0}^{1} b_j(\eta) \, dv_j(\eta) \rightarrow \max
\]

\[
Q x + a + \sum_{j=1}^{N} \int_{0}^{1} a_j(\eta) \, dv_j(\eta) \in X^\perp
\]

\(v \in BV[0, 1]\) continuous, once differentiable and \(\dot{v}_j(t) \geq 0\) for all \(t \in [0, 1]\) (5.37)

Note that any feasible solution of (5.35)-(5.37) is feasible and sub-optimal for (5.5)-(5.7).

Related to the dual problem (5.35)-(5.37) is the following \(\beta\)-dependent family of optimization problems:

\[
\chi_{\beta}(x, v) = -\beta \left[ \frac{1}{2} \langle Q x, x \rangle + \sum_{j=1}^{N} \int_{0}^{1} b_j(\eta) \, dv_j(\eta) \right] + \sum_{j=1}^{N} \int_{0}^{1} \ln(\dot{v}_j(\eta)) \, d\eta \rightarrow \max
\]

\[
Q x + a + \sum_{j=1}^{N} \int_{0}^{1} a_j(\eta) \, dv_j(\eta) \in X^\perp
\]

(5.39)
\[ v \in BV[0,1] \text{ continuous, once differentiable and } \psi_j(t) > 0 \text{ for all } t \in [0, 1] \quad (5.40) \]

Note in particular that our restriction of the dual variables to the continuous, once- differentiable subset of \( BV^N[0,1] \) allows us to add the barrier terms

\[ \int_0^1 \ln(\hat{\psi}_j(\eta)) \, d\eta \]

to the cost functional \( (5.38) \). Such terms would not make sense if the restriction of continuous once-differentiability were not imposed. For every \( \beta > 0 \), we consider the optimal solution of the problem \( (5.38)-(5.40) \). This can be expressed as follows:

**Proposition 5.5** For any \( \beta > 0 \), let \( x(\beta) \) denote the optimal solution of the problem \( (5.1)-(5.2) \). Then \( (x(\beta), v(\beta)) \) with

\[ \frac{\partial \psi_j(\beta, \eta)}{\partial \eta} = \frac{1}{\beta(b_j(\eta) - \langle a_j(\eta), x(\beta) \rangle)}, \quad \eta \in [0, 1] \]

is an optimal solution for the problem \( (5.38)-(5.40) \).

**Proof:** The condition \( (5.40) \) can be written as

\[ \pi \left[ Q x + \sum_{j=1}^N \int_0^1 a_j(\eta) \, d\psi_j(\eta) \right] = -\pi(a) \quad (5.41) \]

\( (5.38)-(5.40) \) is a convex optimization problem and the constraints define an open convex region. Therefore, \( (x^*, v^*) \) is the optimal solution if and only if \( D\chi_\beta(x^*, v^*) \cdot (\xi, \mu) = 0 \) for all \( (\xi, \mu) \in H \times BV^N[0, 1] \) such that \( \mu \) is continuous, once differentiable and

\[ \pi \left[ Q \xi + \sum_{j=1}^N \int_0^1 a_j(\eta) \, d\mu_j(\eta) \right] = 0 \quad (5.42) \]

For \( (x, v) \) satisfying \( (5.39)-(5.40) \) and \( (\xi, \mu) \in H \times BV^N[0, 1] \) satisfying \( (5.42) \), let

\[
D\chi_\beta(x, v) \cdot (\xi, \mu) = \lim_{\alpha \to 0} \frac{\chi_\beta(x + \alpha \xi, v + \alpha \mu) - \chi_\beta(x, v)}{\alpha}
= -\beta \langle Q x, \xi \rangle - \beta \sum_{j=1}^N \int_0^1 b_j(\eta) \, d\mu_j(\eta) + \lim_{\alpha \to 0} \sum_{j=1}^N \frac{1}{\alpha} \int_0^1 \ln \left[ 1 + \alpha \left( \frac{\hat{\psi}_j(\eta)}{\psi_j(\eta)} \right) \right] \, d\eta
\]
5.1. CENTRAL TRAJECTORY ANALYSIS

Since \( \mu(t) \) is once differentiable on \([0, 1]\), \( |\dot{\mu}_j(t)| < \infty \) for \( j = 1, \ldots, N \). Furthermore, \( \dot{v}_j(t) > 0 \) for all \( t \in [0, 1] \), \( j = 1, \ldots, N \). Therefore, \( \left| \frac{\dot{\mu}_j(t)}{\dot{v}_j(t)} \right| < \infty \) for all \( t \in [0, 1] \), \( j = 1, \ldots, N \). It follows that for all \( \alpha > 0 \) sufficiently small

\[
\ln \left[ 1 + \alpha \left( \frac{\dot{\mu}_j(\eta)}{\dot{v}_j(\eta)} \right) \right] = \alpha \left( \frac{\dot{\mu}_j(\eta)}{\dot{v}_j(\eta)} \right) - \frac{\alpha^2}{2} \left( \frac{\dot{\mu}_j(\eta)}{\dot{v}_j(\eta)} \right)^2 + o \left[ \alpha^2 \left( \frac{\dot{\mu}_j(\eta)}{\dot{v}_j(\eta)} \right)^2 \right]
\]

Since the Taylor expansion is uniformly convergent for sufficiently small \( \alpha > 0 \), it follows that

\[
\frac{1}{\alpha} \int_0^1 \ln \left[ 1 + \alpha \left( \frac{\dot{\mu}_j(\eta)}{\dot{v}_j(\eta)} \right) \right] d\eta = \int_0^1 \left( \frac{\dot{\mu}_j(\eta)}{\dot{v}_j(\eta)} \right) d\eta - \frac{\alpha^2}{2} \int_0^1 \left( \frac{\dot{\mu}_j(\eta)}{\dot{v}_j(\eta)} \right)^2 d\eta + \int_0^1 o \left[ \alpha^2 \left( \frac{\dot{\mu}_j(\eta)}{\dot{v}_j(\eta)} \right)^2 \right] d\eta
\]

from which we conclude that

\[
\lim_{\alpha \to 0} \frac{1}{\alpha} \int_0^1 \ln \left[ 1 + \alpha \left( \frac{\dot{\mu}_j(\eta)}{\dot{v}_j(\eta)} \right) \right] d\eta = \int_0^1 \left( \frac{\dot{\mu}_j(\eta)}{\dot{v}_j(\eta)} \right) d\eta = \int_0^1 \frac{1}{\dot{v}_j(\eta)} d\mu_j(\eta)
\]

Therefore

\[
D \chi_{\beta}(x, v) \cdot (\xi, \mu) = -\beta \langle Q x, \xi \rangle - \beta \sum_{j=1}^N \int_0^1 b_j(\eta) \, d\mu_j(\eta)
\]

\[
+ \sum_{j=1}^N \int_0^1 \frac{1}{\dot{v}_j(\eta)} \, d\mu_j(\eta)
\]

(5.43)

Let \( \beta > 0 \) be fixed. Suppose \( x = x(\beta) \) and let

\[
\dot{v}_j(\beta, \eta) = \frac{1}{\beta(b_j(\eta) - \langle a_j(\eta), x(\beta) \rangle)}, \quad \eta \in [0, 1] \text{ and } j = 1, \ldots, N
\]

(5.44)

Since \( x(\beta) \in \text{int}(P) \), it follows that \( b_j(\eta) - \langle a_j(\eta), x(\beta) \rangle > 0 \) for all \( \eta \in [0, 1] \), \( j = 1, \ldots, N \). Therefore, \( 0 < \dot{v}_j(\beta, \eta) < \infty \) for all \( \eta \in [0, 1] \). Furthermore for fixed \( \beta > 0 \) we can construct a continuous \( v_j(\beta, \eta) \in BV[0, 1] \) of \( \eta \) satisfying (5.44); that is \( v(\beta, \eta) \) satisfies (5.40). Furthermore, for any \( (\xi, \mu) \) satisfying (5.42)

\[
D \chi_{\beta}(x(\beta), v(\beta)) \cdot (\xi, \mu)
\]

\[
= \beta \left( -\langle Q x(\beta), \xi \rangle - \sum_{j=1}^N \int_0^1 b_j(\eta) \, d\mu_j(\eta) + \sum_{j=1}^N \int_0^1 (b_j(\eta) - \langle a_j(\eta), x(\beta) \rangle) \, d\mu_j(\eta) \right)
\]

\[
= -\beta \left( x(\beta), Q \xi + \sum_{j=1}^N \int_0^1 a_j(\eta) \, d\mu_j(\eta) \right)
\]

\[
= 0
\]
Therefore by the optimality of \(x(j_3)\) and the condition that \(\nabla f(x(j_3))\) is satisfied (5.39)-(5.40). Furthermore, for each \((\xi, \mu)\) satisfying (5.42), the optimality condition

\[
D_{x_\beta}(x(\beta), v(\beta)) \cdot (\xi, \mu) = 0
\]

is satisfied.

Parallel to Corollary 5.3, we show that the dual functional \(h(x(\beta), v(\beta))\) is monotonically non-decreasing. That is, the dual cost is non-decreasing on the central path.

**Proposition 5.6** The function \(a(\beta) = h(x(\beta), v(\beta))\) is a monotonically non-decreasing function of \(\beta\) for \(\beta > 0\).

**Proof:** Let \(\Delta(\beta) = \chi_\beta(x(\beta), v(\beta)), \beta > 0\) and let \(\beta_2 > \beta_1 > 0\). Since \((x(\beta), v(\beta))\) is the optimal solution of (5.38)-(5.40), it follows that:

\[
\Delta(\beta_1) \geq \chi_{\beta_1}(x(\beta_2), v(\beta_2)) \\
\Delta(\beta_2) \geq \chi_{\beta_2}(x(\beta_1), v(\beta_1))
\]

Since \(\chi_\beta = \beta h + \sum_{j=1}^N \int_0^1 \ln(\hat{v}_j(\eta)) d\eta\), adding these inequalities gives

\[
(\beta_2 - \beta_1) [h(x(\beta_2), v(\beta_2)) - h(x(\beta_1), v(\beta_1))] \geq 0
\]
Since $\beta_2 > \beta_1$, it follows that $h(x(\beta_2), v(\beta_2)) \geq h(x(\beta_1), v(\beta_1))$.

We arrive now to the main results of this section; the convergence properties of the central path as $\beta \to \infty$. The next result shows that the duality gap $f(x) - h(x, v)$ decreases on the central path as $\beta \to \infty$ and that points on the central path converge to the global optimal solution of (5.1)-(5.3).

**Corollary 5.4**

$$f(x(\beta)) - h(x(\beta), v(\beta)) = \frac{N}{\beta}$$

(5.45)

Furthermore, there exists a unique optimal solution $x^*$ to the problem (5.1)-(5.3) and $x(\beta) \to x^*$ as $\beta \to \infty$.

**Proof:** By Lemma 5.1

$$f(x(\beta)) - h(x(\beta), v(\beta)) = \sum_{j=1}^{N} \int_{0}^{1} (b_j(\eta) - \langle a_j(\eta), x(\beta) \rangle) \, dv_j(\beta; \eta)$$

Since $(x(\beta), v(\beta))$ satisfies (5.39)-(5.40), it follows from Proposition 5.5 and the change of variable formula that

$$f(x(\beta)) - h(x(\beta), v(\beta))$$

$$= \sum_{j=1}^{N} \int_{0}^{1} \frac{1}{\beta (b_j(\eta) - \langle a_j(\eta), x(\beta) \rangle)} \, d\eta$$

$$= \frac{N}{\beta}$$

Since $f(x(\beta))$ is a monotonically non-increasing, and $h(x(\beta), v(\beta))$ a monotonically non-decreasing function of $\beta > 0$, there exists $f^*$ such that

$$h(x(\beta), v(\beta)) \leq f^* \leq f(x(\beta))$$

for all $\beta > 0$. Therefore, $f(x(\beta)) \to f^*$ as $\beta \to \infty$. Let $\bar{\beta} > 0$ be fixed and $\sigma = f(x(\bar{\beta}))$. Define $H_\sigma = \{x \in H : f(x) \leq \sigma\}$. Since $Q$ is a strictly positive operator, $H_\sigma$ is bounded. Furthermore, $H_\sigma$ is convex and strongly closed and hence, is also weakly closed (see Section 8.1 of Appendix for a definition of weak closure). Therefore, $H_\sigma$ is weakly compact (see Theorem 8.2 Section 8.1 of Appendix). Note also that $x(\beta) \in H_\sigma$ for $\beta \geq \bar{\beta}$ since $f(x(\beta))$ is a monotonically non-increasing function of $\beta > 0$. In particular, $f^* \leq \sigma$. The existence of $x^*$ follows from the following observation. Let the
problem \((P)\) be given by (5.1)-(5.3) together with the additional constraint \(x \in H_\sigma\). Note that \(x^*\) is optimal for (5.1)-(5.3) if and only if \(x^*\) is optimal for \((P)\). The constraints (5.2)-(5.3), \(x \in H_\sigma\) define a closed, convex, bounded subset of \(H\) and \(f\) is a strictly convex, continuous functional. Therefore, by [3, pp 50], \(f\) achieves its minimum value at a unique \(x^*\). Therefore, (5.1)-(5.3) achieves its minimum \(x^*\) which is unique. We now prove that \(x(\beta) \to x^*\). First we show that \(x(\beta) \to x^*\) weakly. If this is not true, it follows from weak compactness of \(H_\sigma\) that there exists a subsequence \(\beta_1 < \beta_2 < \cdots \) with \(\beta_i \to \infty\) such that \(x(\beta_i) \to y\) weakly, where \(y \in H_\sigma\) and \(y \neq x^*\). Since \(x^*\) is the unique optimal solution, we have \(f(y) > f(x^*)\). Let

\[
\epsilon = \frac{f(y) - f(x^*)}{2}
\]

and define \(U = \{x \in H : f(x) > f(y) - \epsilon\}\). Now \(U\) is a weakly open neighbourhood of \(y\), so \(x(\beta_i) \to y\) weakly implies that \(x(\beta_i) \in U\) for \(i\) sufficiently large. Since \(f(x(\beta_i)) \to f(x^*)\), it follows that

\[
f(x^*) \geq f(y) - \epsilon = \frac{f(y) + f(x^*)}{2}
\]

or \(f(x^*) \geq f(y)\); a contradiction to \(f(y) > f(x^*)\). Therefore, \(x(\beta) \to x^*\) weakly. Since \(Q\) is strictly positive, it follows that \(f(x(\beta)) - f(x^*) \geq \langle \nabla f(x^*), x(\beta) - x^* \rangle + c\|x(\beta) - x^*\|^2\) for some \(c > 0\). Since \(f(x(\beta)) \to f(x^*)\) and \(x(\beta) \to x^*\) weakly as \(\beta \to \infty\), it follows that \(\|x(\beta) - x^*\| \to 0\) as \(\beta \to \infty\). 

In a similar way, we also show that \((x(\beta), v(\beta)) \to (x^*, v^*)\), the optimal solution of the dual problem (5.5)-(5.7) as \(\beta \to \infty\). From this and Corollary 5.4, we conclude that the optimal primal cost \(f(x^*)\) and the optimal dual cost \(h(x^*, v^*)\) are equal.

**Corollary 5.5**

\[
f(x^*) = h(x^*, v^*)
\]
which seem to perform better in practice. The following result shows how the decrease in \( f(x(\beta)) \) on the central path is related to the change in \( \beta \).

**Corollary 5.6** If \( \beta_2 > \beta_1 > 0 \), then

\[
 f(x(\beta_1)) - f(x(\beta_2)) \leq \frac{N}{\beta_1} - \frac{N}{\beta_2} \tag{5.47}
\]

**Proof:** By Proposition 5.6, \( h(x(\beta_2), v(\beta_2)) \geq h(x(\beta_1), v(\beta_1)) \). Therefore

\[
 f(x(\beta_1)) - f(x(\beta_2)) \\
 \leq \left[ f(x(\beta_1)) - h(x(\beta_1), v(\beta_1)) \right] + f(x(\beta_1)) \\
 \leq \left[ f(x(\beta_2)) - h(x(\beta_2), v(\beta_2)) \right]
\]

from which the result follows.

### 5.2 Example

In this section, we show how the path following IPM studied in this chapter can be used to solve LQ control problems with continuous linear state inequality constraints. For simplicity, we shall restrict ourself to scalar time invariant systems. However, generalization to the time varying vector case can be carried out in much the same manner.

Consider the deterministic linear system

\[
 \dot{x}(t) = a \cdot x(t) + b \cdot u(t), \quad x(0) = \xi \tag{5.48}
\]

where \( a, b, \xi \in \mathbb{R} \) are fixed constants. We assume that the class of feasible controls is \( U = L^2(0, 1; \mathbb{R}) \). Let \( q, r \in \mathbb{R} \) be given constants such that \( q \geq 0 \) and \( r > 0 \). The cost functional is given by

\[
 f(x, u) = \frac{1}{2} \int_0^1 (q \cdot x^2(t) + r \cdot u^2(t)) \, dt \tag{5.49}
\]

Let \( c \in \mathbb{R} \) such that \( c > \xi \) be given. Suppose that the constraint functional associated with (5.48)-(5.49) is \( x(t) \leq c \). The problem is to find \( u \in U \) which minimizes (5.49) and satisfies the constraint \( x(t) \leq c \) for every \( t \in [0, 1] \). From (5.49) it follows that

\[
 x(\eta) = \xi + \int_0^\eta (a \cdot x(s) + b \cdot u(s)) \, ds \tag{5.50}
\]
Therefore
\[ x(\eta) = \xi + \int_0^1 (a(s, \eta) \cdot x(s) + b(s, \eta) \cdot u(s)) ds \]  
(5.51)

where
\[
\begin{align*}
a(s, \eta) &= \begin{cases} 
    a, & s \leq \eta \\
    0, & s > \eta
\end{cases} \\
b(s, \eta) &= \begin{cases} 
    b, & s \leq \eta \\
    0, & s > \eta
\end{cases}
\end{align*}
\]
(5.52)
(5.53)

Defining the constraint functional
\[ f_1((x, u), \eta) = \int_0^1 (a(s, \eta) \cdot x(s) + b(s, \eta) \cdot u(s)) ds \]  
(5.54)

it follows that the problem (5.48)-(5.49) with constraint \( x(t) \leq c \) is equivalent to the following:
\[
\begin{cases} 
    f(\zeta) \rightarrow \min \\
    f_1(\zeta, \eta) \leq c, \eta \in [0, 1] \\
y \in z + X
\end{cases}
\]  
(5.55)

where
\[ X = \{(x, u) \in L^2(0, 1; \mathbb{R}) \times L^2(0, 1; \mathbb{R}) : \dot{z} = a \cdot z + b \cdot u, \ x(0) = 0\} \]  
(5.56)

and \( z = (\bar{z}, 0) \in L^2(0, 1; \mathbb{R}) \times L^2(0, 1; \mathbb{R}) \) such that
\[ \dot{\bar{z}}(t) = a \cdot z(t), \ \bar{z}(0) = \xi \]  
(5.57)

Then, it follows that (5.55) is of the form (5.1)-(5.3).

To solve (5.55), we need to minimize the functional \( f_\beta \) with respect to \( (x, u) \in z + X \) for an increasing sequence of \( \beta_i > 0 \) with \( \beta_i \rightarrow \infty \). As in the finite dimensional case, we do this using Newton’s method. A crucial part of Newton’s method is calculating the Newton step which is defined by (5.29) in Proposition 5.3. We now show how the Newton step for the problem (5.55) can be calculated.

For convenience we put \( y = (x, u) \). From (5.8) we obtain
\[
f_\beta(y) = \beta f(y) - \int_0^1 \ln(c - f_1(y, \eta)) d\eta
\]  
(5.58)
5.2. EXAMPLE

Denoting $s(y, \eta) = c(\eta) - f_1(y, \eta)$, $\eta \in [0, 1]$ we have

$$\nabla f_\beta(y) = \begin{bmatrix} \beta \cdot q(t) \cdot x(t) + \int_0^1 \frac{a(t, \eta)}{s(y, \eta)} d\eta \\ \beta \cdot r(t) \cdot u(t) + \int_0^1 \frac{b(t, \eta)}{s(y, \eta)} d\eta \end{bmatrix} = \begin{bmatrix} \beta \cdot q(t) \cdot x(t) + \int_t^1 \frac{a}{s(y, \eta)} d\eta \\ \beta \cdot r(t) \cdot u(t) + \int_t^1 \frac{b}{s(y, \eta)} d\eta \end{bmatrix}$$

(5.59)

For any $\begin{bmatrix} \xi \\ \sigma \end{bmatrix} \in X$, we have by the definition of $\gamma_\beta(y)$ (see (5.23)) that

$$\gamma_\beta(y) \cdot \begin{bmatrix} \xi \\ \sigma \end{bmatrix} = \begin{bmatrix} \beta q\xi(t) + \int_0^1 f_1((\xi, \sigma), \eta) \frac{a(t, \eta)}{s(y, \eta)} d\eta \\ \beta r\sigma(t) + \int_0^1 f_1((\xi, \sigma), \eta) \frac{b(t, \eta)}{s(y, \eta)} d\eta \end{bmatrix} = \begin{bmatrix} \beta q\xi(t) + \int_t^1 f_1((\xi, \sigma), \eta) \frac{a}{s(y, \eta)} d\eta \\ \beta r\sigma(t) + \int_t^1 f_1((\xi, \sigma), \eta) \frac{b}{s(y, \eta)} d\eta \end{bmatrix}$$

(5.60)

When $X$ is given by (5.56), the orthogonal complement of $X$ is

$$X^\perp = \left\{ \begin{bmatrix} \dot{p} + ap \\ bp \end{bmatrix} : p \text{ absolutely continuous on}[0, 1], p(1) = 0, \dot{p} \in L^2[0, 1; \mathbb{R}] \right\}$$

(5.61)

The Newton step $p_\beta(y)$ associated with $f_\beta(y)$ is as defined in Proposition 5.3. In Proposition 5.3, it is proven that

$$p_\beta(y) = -\gamma_\beta(y)^{-1}(\nabla f_\beta(y) - \zeta(y))$$

(5.62)

where $\zeta(y) \in X^\perp$ is the unique element in $X^\perp$ such that $p_\beta(y) \in X$. Therefore, the problem of calculating the Newton step $p_\beta(y)$ is equivalent to finding this unique $\zeta(y) \in X^\perp$. In Proposition 5.3, it is shown that $\zeta(y)$ is found by solving the minimum norm problem (5.30). To calculate $p_\beta(y)$, we proceed in the following way. From (5.62), it follows that

$$\gamma_\beta(y) \cdot p_\beta(y) + \nabla f_\beta(y) = \zeta(y) \in X^\perp$$

(5.63)

For convenience, put

$$\nabla f_\beta(y) = \begin{bmatrix} \ddot{x} \\ \ddot{u} \end{bmatrix}$$

(5.64)
and
\[
d(\eta) = f_t((\xi, \sigma), \eta) \quad (5.65)
\]

It follows from (5.59) and (5.64) that
\[
\ddot{x}(t) = \beta q(t) x(t) + \int_t^1 \frac{a}{s(y, \eta)} d\eta
\]
\[
\ddot{u}(t) = \beta r(t) u(t) + \int_t^1 \frac{b}{s(y, \eta)} d\eta
\]

Note in particular that \( \ddot{x} \) and \( \ddot{u} \) are both known functions of \( t \). We wish to calculate \( p_\beta(y) \in X \) such that (5.63) and hence (5.62) is satisfied. We do this in the following way. First, put
\[
p_\beta(y) = \begin{bmatrix} \xi \\ \sigma \end{bmatrix} \in X \quad (5.66)
\]

Substituting (5.60) and (5.64) into (5.63), it follows the definition of \( X \) and \( X^\perp \) that
\[
\dot{\xi} = a \cdot \xi + b \cdot \sigma, \quad \xi(0) = 0 \quad (5.67)
\]
\[
\beta q(t) \xi(t) + \int_t^1 \frac{a \cdot d(\eta)}{s^2(y, \eta)} d\eta + \ddot{x}(t) = \ddot{p}(t) + a \ddot{p}(t) \quad (5.68)
\]
\[
\beta r(t) \cdot \sigma(t) + \int_t^1 \frac{b \cdot d(\eta)}{s^2(y, \eta)} d\eta + \ddot{u}(t) = b \ddot{p}(t) \quad (5.69)
\]
\[
p(1) = 0 \quad (5.70)
\]

We look for a solution \( p, (\xi, \sigma) \) of the system of equations (5.67)-(5.70) of the form
\[
p(t) = -\beta k(t) \xi(t) - \rho(t) \quad (5.71)
\]
where \( k(t) \) and \( \rho(t) \) are functions that need to be determined. From (5.69), since \( r > 0 \), we immediately obtain
\[
\sigma(t) = \frac{1}{\beta r} \left[ b \rho(t) - \int_t^1 \frac{b \cdot d(\eta)}{s^2(y, \eta)} d\eta - \ddot{u}(t) \right] \quad (5.72)
\]

Substituting (5.71)-(5.72) into (5.67)-(5.68) gives
\[
\dot{k}(t) + 2a \cdot k(t) + \frac{k^2(y) \cdot b^2}{r} + q = 0, \quad k(1) = 0 \quad (5.73)
\]
5.2. EXAMPLE

\[ \dot{\rho}(t) = -(a - \frac{b^2 \cdot k(t)}{r})\rho + \alpha(t), \quad \rho(1) = 0 \]  \hspace{1cm} (5.74)

where

\[ \alpha(t) = \frac{b \cdot k(t)}{r} \left[ \int_t^1 \frac{b \cdot d(\eta)}{s^2(y, \eta)} \, d\eta + \bar{u}(t) \right] - \left[ \int_t^1 \frac{a \, d(\eta)}{s^2(y, \eta)} \, d\eta + \bar{x}(t) \right] \]

Note that (5.73) is just the standard (scalar) Riccati equation associated with unconstrained LQG control. Substituting (5.72) and (5.71) into (5.67) we obtain

\[ \dot{\xi}(t) = (a - \frac{b^2 \cdot k(t)}{r})\xi(t) - \frac{b^2 \rho(t)}{\beta \cdot r} - \gamma(t), \quad \xi(0) = 0 \]  \hspace{1cm} (5.75)

where \( \gamma(t) \) is given by

\[ \gamma(t) = \frac{b}{\beta \cdot r} \left[ \int_t^1 \frac{b \cdot d(\eta)}{s^2(y, \eta)} \, d\eta + \bar{u}(t) \right] \]

Let \( \phi(t, \tau) \) be the fundamental solution of

\[ \dot{\rho} = -(a - \frac{b^2 k}{r})\rho \]

and \( \psi(t, \tau) = [\phi(t, \tau)]^{-1} \) be the fundamental solution of

\[ \dot{\xi} = (a - \frac{b^2 k}{r})\xi \]  \hspace{1cm} (5.76)

It follows that

\[ \rho(t) = \int_T^t \phi(t, \tau) \alpha(\tau) \, d\tau \]  \hspace{1cm} (5.77)

\[ \xi(t) = \int_0^t \psi(t, \tau) \left[ -\frac{b^2 \cdot \rho(\tau)}{\beta \cdot r} - \gamma(\tau) \right] \, d\tau \]  \hspace{1cm} (5.78)

Observe that the only unknown in (5.78) is \( d(\eta) \). Moreover, since \( d(\eta) \) appears linearly in \( \alpha(t) \) and \( \gamma(t) \), it follows that \( \xi(t) \) can be expressed in the form

\[ \xi(t) = \theta(t) + \int_0^1 d(\eta) \, \bar{\theta}(t, \eta) \, d\eta \]  \hspace{1cm} (5.79)

where \( \theta(t) \) and \( \bar{\theta}(t, \eta) \) are known functions and \( d(\eta) \) needs to be calculated. Substituting (5.79) into (5.72) gives

\[ \sigma(t) = \mu(t) + \int_0^1 d(\eta) \, \bar{\mu}(t, \eta) \, d\eta \]  \hspace{1cm} (5.80)

where \( \mu(t) \) and \( \bar{\mu}(t, \eta) \) are also known. Substituting (5.79)-(5.80) into (5.65), we obtain (after some algebra) the following equation for \( d(\eta) \):

\[ d(\eta) = \int_0^1 d(v) \, f(\eta, v) \, dv + g(\eta) \]  \hspace{1cm} (5.81)
CHAPTER 5. INFINITE QUADRATIC PROGRAMMING

where

\[ f(\eta, v) = \int_0^1 [a(t, \eta) \bar{\theta}(t, v) + b(t, \eta) \bar{\mu}(t, v)] \, dt \]
\[ g(\eta) = \int_0^1 [a(t, \eta) \theta(t) + b(t, \eta) \mu(t)] \, dt \]

Therefore, the Newton step \( p_\beta(y) \) (5.66) can be calculated using the following steps:

1. Solve the Riccati equation (5.73).
2. Solve for the fundamental matrix \( \psi(t) \) of (5.76).
3. Solve the integral equation (5.81)

Performing steps (1)-(2) is equivalent to solving an unconstrained LQ optimal control problem. Importantly, the Newton step can be calculated.

5.3 Conclusion

We have generalized the path following interior point method studied in [11] so that it can be used to solve infinite dimensional quadratic optimization problems with infinitely many linear inequality constraints. We have proven that this generalization of the algorithm converges to the global optimal solution. However, we have not considered the issue of complexity. This is an important open problem that needs to be answered. It is important to note here that in finite dimensional IPM's, the size of a problem is measured in terms of the number of variables and the number of constraints. In this problem however, such a measurement of problem size is inadequate since the variable space and the constraint space are infinite dimensional. Rather, a generalized definition of problem size, as given in the paper by Renegar [42], is required. As an example, we considered the deterministic LQ control problem with infinitely many constraints, as considered in Section 2.3.2. As already shown, this includes the class of LQG control problems with continuous linear state inequality constraints. As with all interior point methods, the key step of the algorithm is calculating the Newton step. In this case of LQ control with infinitely many integral linear constraints, calculating the Newton step is equivalent to solving an unconstrained LQ problem together with an integral equation.
Chapter 6

Infinite linear programming

Many optimization problems are naturally cast as infinite programming problems; for instance, continuous time, optimal control problems subject to continuous linear state inequality constraints can be written in such a form. In this chapter, we consider a infinite linear programming problem; that is, a linear programming problem with infinitely many linear inequality constraints where the variables belong to an infinite-dimensional Hilbert space. We generalize a potential reduction method introduced by Ye [56] so that it can be used to solve this problem.

Although much of the research on interior point methods (IPM's) has focused on finite-dimensional problems, interest in infinite-dimensional problems has begun to attract more attention, especially of late. In recent papers by Faybusovich and Moore [11, 12] the logarithmic barrier method is extended so as to be amendable to infinite-dimensional quadratic optimization problems subject to finitely many linear or quadratic constraints. On the other hand, Ferris and Philpott [13, 14], Powell [40] and Todd [53] study a class of semi-infinite linear programming problems. Semi-infinite programming is a term commonly used to describe linear programming problems where the variable belongs to a finite dimensional space, but for which there are infinitely many constraints. Todd shows that a concept termed invariance is fundamental to whether a given IPM can be generalized to solve a semi-infinite problem. The theory determines which IPM's converge to a sensible limiting algorithm as the number of constraints goes to infinity. It does not deal with the issue of convergence of the iterates produced by the limiting algorithm. In particular, Todd shows that the potential reduction method introduced by Ye [56] has a sensible generalization to the semi-infinite programming case. In the papers by Ferris and Philpott [13, 14] and
Powell [40] the affine scaling algorithm, and Karmarkar's algorithm respectively are
generalized to the semi-infinite setting. In particular, Powell shows that convergence
to a non-optimal point can occur if the semi-infinite generalization of Karmarkar's
algorithm is used.

In this chapter we combine features of [11, 12] and [13, 14, 40, 53] to study infinite
linear programming (that is, where the variable can belong to an arbitrary infinite-
dimensional Hilbert space and there are infinitely many constraints). In Section
6.1, we introduce the infinite linear programming problem and the potential function
associated with this problem. In particular, we show that the 'closeness to optimality'
of a given 'strictly feasible solution' can be measured in some way by this potential
function. In Section 6.2, we derive an algorithm based on minimizing this potential
function. It is a generalization of the method studied in [56]. In Section 6.3, we
examine convergence issues. Under certain conditions, we prove that convergence
to the optimal solution occurs. In Section 6.4, we introduce the continuous linear
programming problem of Bellman [4], and show how this algorithm can be used to
solve this problem, as well as linear optimal control problems with continuous linear
state inequality constraints.

We wish to emphasize that in this chapter, we focus on generalizing the algo-
rithm by Ye so that it can be used to solve infinite linear programming problems.
It is far from complete. For instance, we have not tackled the issue of complexity
which remains an important open problem. Nor have we conducted any numerical
implementation, which is the real gauge of the success of an algorithm. Also, many
of the assumptions we have evoked in our convergence analysis are quite stringent.
It remains an open question which of these assumptions can be relaxed to something
less restrictive. In this way, this work should be viewed as presenting new methods
that can be explored rather than the final definitive statement. It is the topic of
continuing research.

These results can be found in the paper by Lim and Moore [29].

6.1 Infinite LP and the potential function

In this section, we introduce the infinite linear programming problem. In particular,
we show that under certain conditions, the 'closeness to optimality' can be measured
6.1. INFINITE LP AND THE POTENTIAL FUNCTION

by the value of a certain ‘potential function’. The algorithm obtained in Section 6.2 is based on the idea of minimizing this potential function.

To begin, let \((H, \langle \cdot, \cdot \rangle)\) be a Hilbert Space and \(X \subset H\) a closed subspace of \(H\). In particular, \(H\) and \(X\) may be infinite dimensional. For \(i = 1, \ldots, m\) and \(\eta \in [0, 1]\) let \(a_i(\eta) \in H, b_i(\eta) \in \mathbb{R}\) and \(c \in X\) be given. Assume that \(a_i(\eta)\) and \(b_i(\eta)\) are measurable functions that depend continuously on \(\eta \in [0, 1]\). For the moment, we shall assume that \(X\) satisfies the following condition:

\[ X = \text{span} \{a_i(\eta) : i = 1, \ldots, m\ \text{and} \ \eta \in [0, 1]\} \]  

(6.1)

The infinite linear programming problem we wish to consider has the primal form

\[
\begin{aligned}
(P) \quad & \langle c, x \rangle \rightarrow \min \\
& \langle a_i(\eta), x \rangle \leq b_i(\eta), \ i = 1, \ldots, m, \ \eta \in [0, 1] \\
& x \in X 
\end{aligned}
\]

This class of problems includes linear optimal control problems with pointwise state constraints. Moreover, an important class of optimization problems known as continuous linear programming (introduced by Bellman [4] in 1953) can be ‘approximated’ by problems of this form. We examine this issue further in Section 6.6.2.

The problem dual to \((P)\) is

\[
\begin{aligned}
(D) \quad & - \sum_{i=1}^{m} \int_{0}^{1} b_i(\eta) \ dv_i(\eta) \rightarrow \max \\
& c + \sum_{i=1}^{m} \int_{0}^{1} a_i(\eta) \ dv_i(\eta) = 0 \\
& v_i(\eta) \in BV[0, 1], \text{ non-decreasing}
\end{aligned}
\]

where \(BV[0,1]\) denotes the space of functions of bounded variation. The integral \(\int_{0}^{1} b_i(\eta) \ dv_i(\eta)\) is the Riemann-Stieltjes integral of \(b_i(\eta)\) with respect to \(v_i(\eta)\). The integral \(\int_{0}^{1} a_i(\eta) \ dv_i(\eta)\) in the constraint of \((D)\) is to be interpreted as in Lemma 2.2.

Throughout this chapter, we shall assume the following.

**Assumption 6.1** There exists a unique optimal solution \(x^* \in X\) for \((P)\) such that \(z^* = \langle c, x^* \rangle > -\infty\).

**Assumption 6.2** There exists \(x \in X\) such that \(\langle a_i(\eta), x \rangle < b_i(\eta)\) for \(i = 1, \ldots, m\) and \(\eta \in [0, 1]\).
**Assumption 6.3** There is $0 < M < \infty$ such that $\|x\| \leq M$ for every primal feasible $x \in X$.

Any $x \in X$ satisfying the conditions in Assumption 6.2 is referred to as being strictly feasible for $(P)$.

Note that the primal problem $(P)$ is a convex optimization problem. Moreover, by Assumption 6.1, the optimal solution is finite and by Assumption 6.2, there exists a strictly feasible solution. Therefore, the Lagrange Duality Theorem (Theorem 2.2) applies. Most importantly, it follows that if $x \in X$ is feasible for $(P)$ and $v_i(\cdot) \in BV[0,1]$ is feasible for $(D)$, then

$$- \sum_{i=1}^{m} \int_{0}^{1} b_i(\eta) \, dv_i(\eta) \leq z^* \leq \langle c, x \rangle \tag{6.2}$$

where $z^*$ is the optimal cost for $(P)$ which is equal to the optimal cost of $(D)$. Clearly, it follows from (6.2) that for any primal feasible $x$ and dual feasible $v_i$, the duality gap

$$\langle c, x \rangle - \left( - \sum_{i=1}^{m} \int_{0}^{1} b_i(\eta) \, dv_i(\eta) \right) \geq 0 \tag{6.3}$$

is an upper bound for $\langle c, x \rangle - z^* \geq 0$. By making (6.3) ‘small’, the difference $\langle c, x \rangle - z^* \geq 0$ can also be made ‘small’. To achieve this aim, we introduce the following potential function.

The potential function is defined as follows. Let $z \leq z^*$ and $x \in X$ be strictly primal feasible. Note that $z \leq z^*$ represents the dual cost corresponding to a feasible dual variable for $(D)$. The primal potential function for $(P)$ is

$$\phi(x, z) = \rho \ln |\langle c, x \rangle - z| - \sum_{i=1}^{m} \int_{0}^{1} s_i(x, \eta) \, d\eta \tag{6.4}$$

where $s_i(x, \eta) = b_i(\eta) - \langle a_i(\eta), x \rangle$.

We show now that by making $\phi(x, z)$ ‘sufficiently negative’, the duality gap $\langle c, x \rangle - z \geq \langle c, x \rangle - z^* \geq 0$ can be made arbitrarily small. In fact, for every strictly primal feasible $x^0 \in X$, it follows from Assumption 6.3 that there exists $0 < K(x^0) < \infty$ such that

$$K(x^0) > \exp \left[ \int_{0}^{1} \ln s_i(x, \eta) \, d\eta - \int_{0}^{1} \ln s_i(x^0, \eta) \, d\eta \right]$$

for every strictly primal feasible $x \in X$. This implies the following relationship between the duality gap, and the value of the potential function.
Proposition 6.1 Let $\epsilon > 0$ be given. If $x^0, x \in X$ are strictly primal feasible and $z, z^0$ are lower bounds of $z^*$, then
\[
\phi(x, z) < \phi(x^0, z^0) + \rho \ln \frac{\epsilon}{K(x^0) \cdot (\langle c, x^0 \rangle - z^0)}
\]
implies that $\langle c, x \rangle - z < \epsilon$.

Proof: This is similar to the proof for the finite dimensional case found in [57]. From (6.4) we obtain
\[
\frac{\langle c, x \rangle - z}{\langle c, x^0 \rangle - z^0} = \gamma(x) \exp \left\{ \frac{1}{\rho} \left[ \phi(x, z) - \phi(x^0, z^0) \right] \right\}
\]
\[
\gamma(x) = \exp \left\{ \frac{1}{\rho} \sum_{i=1}^{m} \int_{0}^{1} \left[ \ln s_i(x, \eta) - \ln s_i(x^0, \eta) \right] d\eta \right\}
\]
It follows from Assumption 6.3 that there exists $K < \infty$ such that $|\gamma(x)| \leq K$ for all primal feasible $x$ and the result follows.

Therefore, by making $\phi(x, z)$ 'sufficiently negative', the duality gap can be made sufficiently small.

6.2 Algorithm

We have shown in Section 6.1 that the duality gap can be reduced by making the potential function $\phi(x, z)$ more negative. In this section, we derive an interior point algorithm which can be used to reduce $\phi(x, z)$. The algorithm seeks to decrease $\phi(x, z)$ in two stages: first by changing $z$, and then by changing $x$. The process for changing $z$ is derived from the constraints of the dual problem while the method for changing $x$ is obtained by minimizing an upper bound on the difference $\phi(x, \bar{z}) - \phi(\bar{x}, \bar{z})$.

First, we introduce some notation. Given the Hilbert space $(H, \langle \cdot , \cdot \rangle)$, let $L^2(0, 1; H)$ be defined as in (2.16). Let $X(\cdot) \in L^2(0, 1; \mathbb{R}^m)$ be a continuous function. For such an $X(\cdot)$, define $\| \cdot \|_2$ and $\| \cdot \|_{\infty}$ as
\[
\|X(\eta)\|_2 = \left[ \sum_{i=1}^{m} \int_{0}^{1} |X_i(\eta)|^2 d\eta \right]^{\frac{1}{2}}
\]
\[
\|X(\eta)\|_{\infty} = \max_{1 \leq i \leq m} \sup_{\eta \in [0,1]} |X_i(\eta)|
\]
where $|\cdot|$ is the standard absolute value operation in $\mathbb{R}$. Given a strictly primal feasible solution $x^0 \in X$, it follows from the assumptions on $a(\cdot)$ and $b(\cdot)$ that $\frac{a(\eta)}{s(x^0, \eta)}$ is a continuous $X$-valued function of $\eta \in [0, 1]$. If $x^1 \in X$, then $x^1 - x^0 \in X$ and $\left\langle \frac{a(\eta)}{s(x^0, \eta)}, x^1 - x^0 \right\rangle$ is an $\mathbb{R}$-valued continuous function of $\eta \in [0, 1]$; that is, $\left\langle \frac{a(\eta)}{s(x^0, \eta)}, x^1 - x^0 \right\rangle \in C[0, 1]$. Somewhat abusing notation, we shall use $\left\langle \frac{a(\eta)}{s(x^0, \eta)}, x^1 - x^0 \right\rangle$ to denote the following $\mathbb{R}^n$-valued continuous function:

$$
\left\langle \frac{a(\eta)}{s(x^0, \eta)}, x^1 - x^0 \right\rangle = \begin{bmatrix}
\left\langle \frac{a_1(\eta)}{s_1(x^0, \eta)}, x^1 - x^0 \right\rangle \\
\vdots \\
\left\langle \frac{a_m(\eta)}{s_m(x^0, \eta)}, x^1 - x^0 \right\rangle
\end{bmatrix}
$$

The following result is an inequality which we shall require to derive bounds for the decrease in the potential function $\phi(x, z)$. It is a generalization of a result from [56].

**Lemma 6.1** Let $X(\eta) \in L^2(0, 1; \mathbb{R}^m)$ be continuous and satisfy $X_i(\eta) > 0$ on $\eta \in [0, 1]$. If $\epsilon = [1, \cdots, 1]' \in L^2(0, 1; \mathbb{R}^m)$ and $\|X(\eta) - \epsilon\|_\infty < 1$, then

$$
\sum_{i=1}^m \int_0^1 \ln X_i(\eta) \, d\eta \geq \sum_{i=1}^m \int_0^1 X_i(\eta) \, d\eta - m - \frac{\|X(\eta) - \epsilon\|_2^2}{2 (1 - \|X(\eta) - \epsilon\|_\infty)}
$$

**Proof:** Let $N \in \mathbb{Z}^+$. It follows from [56] that

$$
\sum_{i=1}^m \sum_{j=1}^N \ln X_i \left( \frac{j}{N} \right) \geq \sum_{i=1}^m \sum_{j=1}^N X_i \left( \frac{j}{N} \right) - m N - \frac{\sum_{i=1}^m \sum_{j=1}^N X_i \left( \frac{j}{N} \right) - 1}{2 \left( 1 - \sup_{i, j} X_i \left( \frac{j}{N} \right) - 1 \right)}
$$

The result is obtained by dividing both sides by $N$ and letting $N \to \infty$. $\blacksquare$

Observe that the potential function $\phi(x, z)$ may be reduced by changing $x$ and by changing $z$. The next result is the theoretical basis for the method of changing $x$ in the algorithm. It is an infinite dimensional generalization of a result from [56].

**Proposition 6.2** Let $x^0, x^1 \in X$ be strictly primal feasible and $z^0 \leq z^*$ where $z^*$ satisfies (6.2). If

$$
\left\| \left\langle \frac{a(\eta)}{s(x^0, \eta)}, x^1 - x^0 \right\rangle \right\|_\infty = \max_{1 \leq i \leq m} \sup_{\eta \in [0, 1]} \left\| \left\langle \frac{a_i(\eta)}{s_i(x^0, \eta)}, x^1 - x^0 \right\rangle \right\| < 1 \quad (6.5)
$$

then

$$
\phi(x^1, z^0) - \phi(x^0, z^0) \leq \langle \nabla \phi(x^0, z^0), x^1 - x^0 \rangle + \frac{\left\| \left\langle \frac{a(\eta)}{s(x^0, \eta)}, x^1 - x^0 \right\rangle \right\|_2^2}{2 \left( 1 - \left\| \left\langle \frac{a(\eta)}{s(x^0, \eta)}, x^1 - x^0 \right\rangle \right\|_\infty \right)} \quad (6.6)
$$
6.2. ALGORITHM

where

$$\nabla \phi(x,z) = \frac{\rho}{\langle c, x \rangle - z} c + \sum_{i=1}^{m} \int_{0}^{1} \frac{a_i(\eta)}{s_i(x, \eta)} d\eta$$

**Proof:** From the concavity of \(\ln(\cdot)\) and the result stated in Lemma 6.1, it follows that if

$$\left\| \left\langle \frac{a(\eta)}{s(\eta, x^0)}, x^1 - x^0 \right\rangle \right\|_\infty < 1$$

then

$$\phi(x^1, z^1) - \phi(x^0, z^1) \leq \left( \frac{\rho}{\langle c, x^0 \rangle - z^1 c} + \sum_{i=1}^{m} \int_{0}^{1} \left\langle \frac{a(\eta)}{s(x^0, \eta)}, x^1 - x^0 \right\rangle d\eta \right)$$

$$+ \frac{\left\| \left\langle \frac{a(\eta)}{s(x^0, \eta)}, x^1 - x^0 \right\rangle \right\|_2^2}{2 \left\{ 1 - \left\| \left\langle \frac{a(\eta)}{s(x^0, \eta)}, x^1 - x^0 \right\rangle \right\|_\infty \right\}}.$$  

Since

$$\frac{a_i(\eta)}{s_i(x^0, \eta)} \in X, \ \eta \in [0, 1]$$

is continuous, we obtain

$$\int_{0}^{1} \left\langle \frac{a_i(\eta)}{s_i(x^0, \eta)}, x^1 - x^0 \right\rangle d\eta = \left\langle \int_{0}^{1} \frac{a_i(\eta)}{s_i(x^0, \eta)} d\eta, x^1 - x^0 \right\rangle$$

where

$$\int_{0}^{1} \frac{a_i(\eta)}{s_i(x^0, \eta)} d\eta$$

is defined as in Lemma 2.2. The result follows. \(\blacksquare\)

Given a strictly primal feasible solution \(x^0 \in X\) and feasible lower bound \(z^1 \leq z^*\), we can use (6.6) to determine another strictly primal feasible solution \(x^1 \in X\) by minimizing the right hand side of the inequality (6.6). In this way, we are in some sense minimizing an upper bound on the difference

$$\phi(x^1, z^1) - \phi(x^0, z^1)$$

(6.7)

Noting that

$$\left\| \left\langle \frac{a(\eta)}{s(x^0, \eta)}, x^1 - x^0 \right\rangle \right\|_2^2 = \langle Q(x^0) \cdot (x^1 - x^0), x^1 - x^0 \rangle$$
where the operator \( Q(x^0) : X \to X \) is defined by
\[
Q(x^0) \cdot y = \sum_{i=1}^{m} \int_{0}^{1} \left( \frac{a_i(\eta)}{s_i(x^0, \eta)} \right) \frac{a_i(\eta)}{s_i(x^0, \eta)} \, d\eta
\]
we obtain from (6.6) the following optimization problem over \( x \):
\[
\langle \nabla \phi(x^0, z^1), x - x^0 \rangle \to \min
\]
\[
\langle Q(x^0) \cdot (x - x^0), x - x^0 \rangle^2 \leq \beta^*
\]
\[x - x^0 \in X\]  

Note that the inequality (6.6) holds only if (6.5) is satisfied. On the other hand, unless \( \beta = 0 \) the inequality in (6.9) being true will not guarantee that (6.5) is true. That is, a solution \( x^1 \) of (6.9), with arbitrary \( \beta^* > 0 \) will generally not give rise to a minimal upper bound on the difference (6.7) as suggested by the inequality (6.6).

For the moment, we assume that there exists \( \beta^* > 0 \) such that the optimal solution of (6.9) gives rise to a solution \( x^1 \) which satisfies the inequality (6.5) and hence (6.6). We defer the issue of calculating \( \beta^* \) until later and for the moment, assume that \( \beta^* \) has been found.

We solve (6.9) by transforming it into an equivalent minimization problem over a sphere, with a linear cost functional. The optimal solution for such a problem is given by the scaled gradient of the linear cost functional. To make such a transformation, we must observe the following: Since \( Q(x^0) : X \to X \) is a strictly positive operator on \( X \) for each \( x^0 \in X \), it follows that \( \langle \cdot, \cdot \rangle_{x^0} : X \times X \to \mathbb{R}, \langle \cdot, \cdot \rangle_{x^0} = \langle Q(x^0) \cdot, \cdot \rangle \) defines an inner product on \( X \) and hence, a Riemannian metric on \( X \). The gradient of \( \phi(x, z) \) with respect to this Riemannian metric is given as follows.

**Lemma 6.2** Let \( Q(x^0) : X \to X \) be given by (6.8). The gradient of \( \phi(x^0, z^1) \) with respect to the Riemannian metric defined by the inner product \( \langle \cdot, \cdot \rangle_{x^0} \) is
\[
\nabla_{x^0} \phi(x^0, z^1) = Q(x^0)^{-1} \cdot \nabla \phi(x^0, z^1)
\]

**Proof:** The functional
\[
\langle \nabla \phi(x^0, z^1), \cdot \rangle : X \to \mathbb{R}
\]
is a bounded linear functional on the Hilbert (sub)space \( X \), so by the Riesz Representation Theorem for functionals on a Hilbert space, there exists a unique \( \nabla_{x^0} \phi(x^0, z^0) \in \)
6.2. ALGORITHM

$X$ such that

$$\langle \nabla \phi(x^0, z^1), y \rangle = \langle Q(x^0) \cdot \nabla \phi(x^0, z^1), y \rangle$$

for all $y \in X$. Therefore

$$\nabla \phi(x^0, z^1) - Q(x^0) \cdot \nabla \phi(x^0, z^1) \in X^\perp$$

The result follows from the fact that

$$\nabla \phi(x^0, z^1) - Q(x^0) \cdot \nabla \phi(x^0, z^1) \in X$$

It follows from Lemma 6.2 that the problem (6.9) can be transformed into the following equivalent one:

$$\langle \nabla \phi(x^0, z^1), y \rangle_{z^0} \rightarrow \min$$

$$\langle y, y \rangle_{z^0}^{\frac{1}{2}} \leq \beta^*$$

$$y \in X$$

where we have replaced $x - x^0 \in X$ in (6.9) with $y$. The problem (6.11) is a minimization problem over the sphere $\{y \in X : \langle y, y \rangle_{z^0}^{\frac{1}{2}} \leq \beta^*\}$ with a linear cost functional $\langle \nabla \phi(x^0, z^1), y \rangle_{z^0}$. The optimal solution of (6.11) is given by the scaled gradient of $\langle \nabla \phi(x^0, z^1), y \rangle_{z^0}$, namely

$$y^* = -\beta \nabla \phi(x^0, z^1)$$

$$\beta = \frac{\beta^*}{\langle \nabla \phi(x^0, z^1), \nabla \phi(x^0, z^1) \rangle_{z^0}^{\frac{1}{2}}}$$

Therefore, it follows that the optimal solution $x^1$ of (6.9) is

$$x^1 = x^0 - \beta Q(x^0)^{-1} \cdot \nabla \phi(x^0, z^1)$$

$$\beta = \frac{\beta^*}{\langle Q(x^0)^{-1} \cdot \nabla \phi(x^0, z^1), \nabla \phi(x^0, z^1) \rangle_{z^0}^{\frac{1}{2}}}$$

We shall refer to the update (6.12) as performing a Newton step.

We return now to the problem of choosing $\beta^*$ such that $x^1 - x^0$ satisfies the inequality (6.5). Recall that in general, $\beta^* > 0$ will not guarantee that (6.5) holds and hence, will not guarantee the reduction in $\phi(x, z)$ which (6.6) suggests. From (6.12)-(6.13), an equivalent problem to finding $\beta^* > 0$ is choosing $\beta > 0$ such that the
inequality (6.5) holds. We show in Proposition 6.3 that it is possible to chose $\beta > 0$ such that a decrease in the potential function is guaranteed. This result is an infinite dimensional generalization of a result from [56].

**Proposition 6.3** If $\pi(x^0, z^1) \in X$ is given by

$$\pi(x^0, z^1) = Q(x^0)^{-1} \cdot \nabla \phi(x^0, z^1)$$

(6.14)

and $0 < \beta \leq \frac{1}{2}$ satisfies the inequality

$$\beta \left( \frac{a_i(\eta)}{s_i(x^0, \eta)} \pi(x^0, z^1) \right) < \frac{1}{2} \quad i = 1, \ldots, m; \quad \eta \in [0, 1]$$

then

$$\phi(x^0 - \beta \pi, z^1) - \phi(x^0, z^1) < -\frac{\beta}{2} \left( Q(x^0) \cdot \pi(x^0, z^1), \pi(x^0, z^1) \right)$$

(6.15)

**Proof:** For simplicity, we shall put $\pi = \pi(x^0, z^1)$ in this proof.

$$\phi(x^0 - \beta \pi, z^1) - \phi(x^0, z^1)$$

$$= \rho \ln \left( \langle c, x^0 \rangle - z^1 - \beta \langle c, \pi \rangle \right) - \sum_{i=1}^{m} \int_{0}^{1} \ln \left( s_i(x^0, \eta) + \beta \langle a_i(\eta), \pi \rangle \right) d\eta$$

$$- \rho \ln \left( \langle c, x^0 \rangle - z^1 \right) + \sum_{i=1}^{m} \int_{0}^{1} \ln s_i(x^0, \eta) d\eta$$

$$= \rho \ln \left[ 1 - \beta \frac{\langle c, x^0 \rangle - z^1 \langle c, \pi \rangle}{\langle c, x^0 \rangle - z^1} \right] - \sum_{i=1}^{m} \int_{0}^{1} \ln \left[ 1 + \frac{\beta \langle a_i(\eta), \pi \rangle}{s_i(x^0, \eta)} \right] d\eta$$

$$\leq -\beta \left[ \frac{\rho}{\langle c, x^0 \rangle - z^1} \langle c, \pi \rangle - \sum_{i=1}^{m} \int_{0}^{1} \left[ \beta \frac{\langle a_i(\eta), \pi \rangle}{s_i(x^0, \eta)} - \frac{\left( \frac{\beta \langle a_i(\eta), \pi \rangle}{s_i(x^0, \eta)} \right)^2}{2 \left( 1 - \beta \frac{\langle a_i(\eta), \pi \rangle}{s_i(x^0, \eta)} \right)} \right] d\eta \right]$$

(because $\ln(1 + \theta) \geq \theta - \frac{\theta^2}{2(1 - |\theta|)}$ for $|\theta| < 1$ and $\ln(1 + \theta) < \theta$ for $\theta > 0$)

$$= -\beta \left[ \frac{\rho}{\langle c, x^0 \rangle - z^1} c + \sum_{i=1}^{m} \int_{0}^{1} \frac{a_i(\eta)}{s_i(x^0, \eta)} d\eta, \pi \right]$$

$$+ \beta^2 \sum_{i=1}^{m} \int_{0}^{1} \frac{\left( \frac{a_i(\eta), \pi}{s_i(x^0, \eta)} \right)^2}{2 \left( 1 - \beta \frac{\langle a_i(\eta), \pi \rangle}{s_i(x^0, \eta)} \right)} d\eta$$

$$< -\beta \langle \nabla \phi(x^0, z^1), \pi \rangle + \beta^2 \sum_{i=1}^{m} \int_{0}^{1} \left( \frac{a_i(\eta), \pi}{s_i(x^0, \eta)} \right)^2 d\eta$$

(because $\frac{1}{2 \left( 1 - \beta \frac{\langle a_i(\eta), \pi \rangle}{s_i(x^0, \eta)} \right)} < 1$)
Unlike the finite dimensional case, a reduction of $\phi(x^0, z)$ by at least a fixed constant after each iteration is not guaranteed because $\langle Q(x^0) \cdot \pi(x^0, z^1), \pi(x^0, z^1) \rangle$ in (6.15) depends on $x^0$. Thus, (6.15) on its own does not guarantee that $\phi(x, z)$ can be made arbitrarily negative. Also, for this reason, the theoretically attractive upper bound polynomial complexity on the number of iterates required to obtain the optimal solution cannot be proven by making a direct generalization of the techniques used in [56].

Thus far, we have derived a technique for updating a given strictly feasible solution $x \in X \text{ of } (P)$ so as to guarantee a decrease in the value of the potential function. Next, we consider the problem of updating the lower bound $z^0$ of the optimal cost $z^*$. To do this, we examine the constraints of the dual problem $(D)$.

Let $x^0 \in X$ be a strictly primal feasible solution. If $v_i(\eta) \in BV[0, 1]$ is given by

$$v_i(\eta) = \int_0^\eta f_i(x^0, z, t) \, dt$$  \hfill (6.16)

where

$$f_i(x^0, z, \eta) = \left[ \frac{\langle c, x^0 \rangle - z}{\rho} \right] \left[ s_i(x^0, \eta)^{-1} - \left\langle \frac{a_i(\eta)}{s_i(x^0, \eta)^2} \cdot Q(x^0)^{-1} \cdot \nabla \phi(x^0, z) \right\rangle \right]$$  \hfill (6.17)

then $v_i(\eta)$ satisfies the constraint

$$c + \sum_{i=1}^m \int_0^1 a_i(\eta) \, dv_i(\eta) = c + \sum_{i=1}^m \int_0^1 a_i(\eta) \, \dot{v}_i(\eta) \, d\eta = 0$$  \hfill (6.18)

for any $z \in \mathbb{R}$. This can be verified by substitution. Therefore, for any $z \in \mathbb{R}$, $v_i(\eta)$ given by (6.16) is a feasible dual variable if and only if it is a non-decreasing function of $\eta \in [0, 1]$. Furthermore, the lower bound $z^0$ can be increased to $z^1 > z^0$ so long as
the resulting \( v_i(\eta) \) corresponding to \( z^1 \) is a non-decreasing function of \( \eta \); that is

\[
    f_i(x^0, z^1, \eta) \geq 0, \quad i = 1, \cdots, m; \quad \eta \in [0, 1]
\]

Hence, by solving:

\[
    (ZP) \quad \begin{cases}
        z^1 = \arg \max z \\
        f_i(x^0, z, \eta) \geq 0; \quad i = 1, \cdots, m; \quad \eta \in [0, 1] \\
        z \geq z^0
    \end{cases}
\]

we obtain \( z^1 = \max\{z^0, z^1\} \). If \( z^1 \) can be made arbitrarily large, then \((P)\) has no solution. Note also that for \( z^0 < z^1 \leq z^* \)

\[
    \phi(x^0, z^1) < \phi(x^0, z^0)
\]

Summarizing, we have shown in Proposition 6.1 that the duality gap \( \langle c, x \rangle - z \geq 0 \) can be made arbitrarily small by making the penalty function \( \phi(x, z) \) 'sufficiently negative'. The algorithm for solving \((P)\) is based on this fact. At each iteration, \( \phi(x, z) \) is decreased in two stages: by changing \( x \) and changing \( z \). Given a strictly primal feasible solution \( x^i \), and a lower bound \( z^i \) of the optimal cost \( z^* \), an update \( z^{i+1} \geq z^i \) is obtained by solving the optimization problem \((ZP)\) which is derived from the constraints of the dual problem \((D)\). An update of the strictly feasible solution \( x^i \) of the primal problem is obtained by minimizing the upper bound of the difference \( \phi(x, z^{i+1}) - \phi(x^i, z^{i+1}) \), as suggested by the inequality (6.6). This leads to the following algorithm.

**Algorithm \((P)\):**

\[
    x^0 \in X, \langle a_i(\eta), x^0 \rangle < b_i(\eta) \quad \eta \in [0, 1], \quad i = 1, \cdots, m; \\
    z^0 \leq z^*; \\
    \epsilon > 0; \\
    k = 0; \\
\textbf{while} \quad \langle c, x^k \rangle - z^k > \epsilon \textbf{ do} \\
\textbf{begin} \\
\quad \text{solve} \quad (ZP) \quad \text{and obtain} \quad z^{k+1}; \\
\quad \text{compute} \quad \pi(x^k, z^{k+1}) \quad \text{according to} \quad (6.14); \\
\quad x^{k+1} = x^k - \beta^* \pi(x^k, z^{k+1}) \quad \text{where} \quad \beta^* = \arg \min_{\beta \geq 0} \phi(x^k - \beta \pi(x^k, z^{k+1})); \\
\quad k = k + 1 \\
\textbf{end} \\
\textbf{end}
\]
6.3. CONVERGENCE RESULTS

6.3 Convergence results

We now examine the convergence properties of the iterates \((x^k, z^k)\) produced by Algorithm \((P)\). By considering a family of finite dimensional approximations of \((P)\), we prove (under certain assumptions) convergence of this sequence \((x^k, z^k)\).

Consider the following discretization of \((P)\): let \(I_N \subset [0, 1]\) be a finite subset of the form

\[
    I_N = \left\{ \frac{i}{2^N} : i = 1, \ldots, 2^N \right\}
\]

which partitions \([0, 1]\) into \(2^N\) subintervals. Clearly \(I_N \subset I_{N+1}\). Let \(c_N \in X_N\) be the orthogonal projection of \(c \in X\) onto the subspace

\[
    X_N = \text{span} \{ a_i(\eta) : i = 1, \ldots, m; \eta \in I_N \}
\]

A finite-dimensional discretization of \((P)\) with respect to the partition \(I_N\) is:

\[
    (P_N) \quad \begin{cases} 
        \langle c_N, x \rangle \rightarrow \min \\
        \langle a_i(\eta), x \rangle \leq b_i(\eta); \quad \eta \in I_N, i = 1, \ldots, m \\
        x \in X_N 
    \end{cases}
\]

Dual to \((P_N)\) is

\[
    (D_N) \quad \begin{cases} 
        - \sum_{i=1}^m \int_0^1 \left[ \sum_{j=1}^{2^N} b_i(\eta_j) I_A_j(\eta) \right] dv_i(\eta) \rightarrow \max \\
        c_N + \sum_{i=1}^m \int_0^1 \left[ \sum_{j=1}^{2^N} a_i(\eta_j) I_A_j(\eta) \right] dv_i(\eta) = 0 \\
        v_i(\eta) \in BV[0, 1], \text{ increasing}
    \end{cases}
\]

where

\[
    I_A(\eta) = \begin{cases} 
        1, & \eta \in A \\
        0, & \eta \notin A 
    \end{cases} \quad A_j = \left[ \frac{j-1}{2^N}, \frac{j}{2^N} \right]
\]

and \(\eta_i = \frac{i}{2^N}\). Since \(X_N\) is a finite dimensional subset of \(X\), it follows that \((P_N)\) is equivalent to a finite-dimensional LP problem over \(\mathbb{R}^g(N)\) where \(g(N) = m \cdot 2^N\). Let \(x_N^*\) denote the optimal solution for \((P_N)\). To begin, we examine the sequence of optimal costs \(\{\langle c_N, x_N^* \rangle\}_{N=1}^\infty\) associated with the family of discretizations \(\{(P_N)\}_{N=1}^\infty\), and its relation to the optimal cost \(\langle c, x^* \rangle\) of \((P)\).
Lemma 6.3 Let $x_N^*$ denote the optimal solution for the problem $(P_N)$ and $x^*$ the optimal solution for $(P)$. Then $(c_N, x_N^*) \to (c, x^*)$

Proof: See Section 6.6.1 in the Appendix.

Our results concerning the convergence of the sequence $\{x^k\}_{k=1}^{\infty}$ resulting from Algorithm $(P)$ are obtained by examining the convergence behavior of the family of finite dimensional approximations $\{(P_N)\}_{N=1}^{\infty}$. We have already shown in Lemma 6.3 that the sequence of optimal costs associated with $\{(P_N)\}_{N=1}^{\infty}$ converges to that of $(P)$ (that is, $(c_N, x_N^*) \to (c, x^*)$). Now we turn our attention to the sequence of optimal solutions $\{x_N^*\}_{N=1}^{\infty}$ and show that $x_N^*$ converges to $x^*$ weakly. The definition of weak convergence on a Hilbert space can be found in Section 8.1 of the Appendix at the end of the thesis.

Proposition 6.4 If $x^*$ is the optimal solution for $(P)$ then $x_N^* \to x^*$ as $N \to \infty$.

Proof: Let $y \in X$ satisfy the constraint $\langle a_i(\eta), y \rangle \leq b_i(\eta)$ for $i = 1, \cdots, m$ and suppose that $\langle c, y \rangle > \langle c, x^* \rangle$. Put $\sigma = \langle c, y \rangle$ and let $\delta > 0$ be given. Define the set

$$H_\sigma = \{x \in X : \langle c, x \rangle \leq \sigma, \langle a_i(\eta), x \rangle \leq b_i(\eta) + \delta, i = 1, \cdots, m\}$$

By Lemma 6.3 $\langle c, x_N^* \rangle = \langle c_N, x_N^* \rangle \to \langle c, x^* \rangle$ as $N \to \infty$. Therefore, $\langle c, x_N^* \rangle \leq \sigma$ for $N$ sufficiently large. Furthermore, $\langle a_i(\eta), x_N^* \rangle \leq b_i(\eta)$ for $\eta = \frac{1}{2N}, j = 1, \cdots, 2N$. From this, it can be shown that for $N$ sufficiently large, $\langle a_i(\eta), x_N^* \rangle \leq b_i(\eta) + \delta$ for every $\eta \in [0, 1]$. Hence, it follows that $x_N^* \in H_\sigma$ for $N \in \mathbb{Z}^+$ sufficiently large. We now prove that $x_N^* \to x^*$. Suppose that this is not true. Then it follows from the weak compactness of $H_\sigma$ (it is closed, bounded and convex; see Section 8.1 in the Appendix at the end of the thesis) that there exists an increasing sequence $N_1, N_2, N_3, \cdots$ such that $x_{N_j}^* \to y$ where $y \neq x^*$ (and hence, $\langle c, y \rangle > \langle c, x^* \rangle$). Let

$$U = \{x \in X : \langle c, x \rangle > \langle c, y \rangle - \epsilon\}, \quad \epsilon = \frac{\langle c, y \rangle - \langle c, x^* \rangle}{2}$$

Then $U$ is a weakly open neighbourhood of $y$ (since $U^c$ is weakly closed; see Section 8.1 in the Appendix at the end of the thesis). Hence $x_{N_j}^* \in U$ for all $j$ sufficiently large. Since $\langle c, x_{N_j} \rangle \to \langle c, x^* \rangle$, it follows that $\langle c, x_N^* \rangle \geq \langle c, y \rangle - \epsilon$ or equivalently, $\langle c, x^* \rangle \geq \langle c, y \rangle$. This contradicts $\langle c, y \rangle > \langle c, x^* \rangle$. Therefore, $x_N^* \to x^*$.

However, stronger results can be obtained if we make the following assumption:
6.3. CONVERGENCE RESULTS

Assumption 6.4 The sequence \( \{x_N^*\}_{N=1}^{\infty} \) converges as \( N \to \infty \).

Under Assumption 6.4, we have the following strong convergence result for the sequence \( x_N^* \).

Proposition 6.5 Let \( x^* \) be the optimal solution for \((P)\) and \( x_N^* \in X_N \) be the optimal solution for \((P_N)\). Under Assumption 6.4, \( x_N^* \to x^* \) as \( N \to \infty \).

Proof: By Assumption 6.4, there exists \( x^*_\infty \in X \) such that \( x_N^* \to x^*_\infty \) as \( N \to \infty \). Furthermore, it is easily shown that such an \( x^*_\infty \) must be feasible for \((P)\). Since \( x^* \) is optimal for \((P)\), it follows that \( \langle c, x^*_\infty \rangle \leq \langle c, x^*_N \rangle \). On the other hand, it is easily shown that \( \pi_N x^*_\infty \) is feasible for \((P_N)\), where \( \pi_N : X \to X_N \) is the orthogonal projection of \( X \) onto \( X_N \). Since \( x_N^* \) is optimal for \((P_N)\), it follows that \( \langle c_N, x_N^* \rangle \leq \langle c_N, \pi_N x^* \rangle \) for all \( N \). Letting \( N \to \infty \), we have by continuity of the inner product that \( \langle c, x^*_\infty \rangle \leq \langle c, x^* \rangle \) and hence \( \langle c, x^*_\infty \rangle = \langle c, x^* \rangle \) from which it follows that \( x^*_\infty \) is optimal for \((P)\).

We consider now a finite dimensional discretization of the potential function \( \phi(x, z) \), as defined by (6.4). In Algorithm \((P)\), updates of \( x \) are derived by minimizing an upper bound on the change in \( \phi(x, z) \) resulting from changes in \( x \). This idea is used to construct the Newton step (6.12) for Algorithm \((P)\). Our purpose for introducing this discretization of \( \phi(x, z) \) is to study the convergence behavior of the Newton step (6.12) by comparing it with its finite dimensional counterpart, as derived in [56]. Ultimately, this leads to our results on the convergence behavior of the iterates \( \{x_k\}_{k=1}^{\infty} \) produced by Algorithm \((P)\). Also linked to this is the behavior of the algorithm \((ZP)\) for updating the lower bound \( z \leq z^* \).

To begin, let \( z_N^* \) be the optimal cost for \((P_N)\) and \((D_N)\). For primal feasible \( x \) and dual feasible \( \eta \) we have

\[
- \sum_{i=1}^{m} \int_{0}^{1} \left[ \sum_{j=1}^{2N} b_i(\eta_j) I_{A_j}(\eta) \right] d\nu_i(\eta) \leq z_N^* \leq \langle c_N, x \rangle \tag{6.20}
\]

The potential function for \((P_N)\) with \( z \leq z_N^* \) and \( x \) strictly primal feasible is

\[
\phi_N(x, z) = \rho \ln (\langle c_N, x \rangle - z) - \frac{1}{2N} \sum_{i=1}^{m} \sum_{j=1}^{2N} \ln s_i(x, \eta_j) \tag{6.21}
\]
and the gradient of $\phi_N(x, z)$ with respect to $x$ is

$$
\nabla \phi_N(x, z) = \frac{\rho}{\langle c_N, z \rangle - z} c_N + \frac{1}{2N} \sum_{i=1}^m \sum_{j=1}^{2N} \frac{a_i(\eta_j)}{s_i(x_N, \eta_j)}
$$

Analogous to (6.14), the Newton step $\pi_N \in X_N$ for $(P_N)$ is

$$
\pi_N(x, z) = Q_N(x)^{-1} \cdot \nabla \phi_N(x, z)
$$

(6.22)

where for each $y \in X$

$$
Q_N(x_N^0) \cdot y = \frac{1}{2N} \sum_{i=1}^m \sum_{j=1}^{2N} \left( \frac{a_i(\eta_j)}{s_i(x_N^0, \eta_j)} \right) y.
$$

(6.23)

A finite dimensional version of $(ZP)$ can be derived as follows. Let $z \in \mathbb{R}$ and $v_i(\eta) \in BV[0, 1]$ be given by

$$
v_i(\eta) = \left[ \sum_{j=1}^{2N} f_i^{(N)}(x_N^0, z, \eta_j) I_{A_j}(\eta) \right] \eta.
$$

(6.24)

where $\eta_i = \frac{1}{2N}$, $I_{A_i}(\eta)$ is defined by (6.19) and

$$
f_i^{(N)}(x_N^0, z, \eta) = \left[ \frac{\langle c_N, x_N^0 \rangle - z}{\rho} \right]$$

$$
\times \left[ s_i(x_N^0, \eta)^{-1} - \frac{a_i(\eta)}{s_i(x_N^0, \eta)^2} \right]
$$

(6.25)

It can be shown by direct substitution that for any $z \in \mathbb{R}$, $v_i(\eta)$ as defined by (6.24) satisfies the equality constraint in $(DN)$. Therefore, $v_i(\eta)$ is dual feasible if and only if it is an increasing function of $\eta$; that is, if and only if $f_i^{(N)}(x^0, z, \eta_j) \geq 0$ for all $i = 1, \cdots, m$ and $j = 1 \cdots, 2N$. Therefore, a lower bound $z_N^0 \leq z_N^*$ can be updated to $z_N^1$ by solving the following problem, analogous to $(ZP)$:

$$
\begin{align*}
\underline{z}_N^1 = \arg \max z \\
(P_N) \quad \begin{cases} 
 f_i^{(N)}(x^0, z, \eta_j) \geq 0; \quad i = 1, \cdots, m; \quad \eta_j \in I_N \\
 z \geq z_N^0 
\end{cases}
\end{align*}
$$

with $z_N^1 = \max\{z_N^0, \underline{z}_N^1\}$. The algorithm for solving $(P_N)$ is identical in philosophy to that for solving $(P)$, and the properties and convergence behavior of this algorithm is studied in [56]. For $(P_N)$, we have the following algorithm:
6.3. CONVERGENCE RESULTS

Algorithm $(P_N)$:

\[ x_N^0 \in X_N, \langle a_i(\eta), x_N^0 \rangle < b_i(\eta) \eta \in I_N, i = 1, \cdots, m; \]
\[ z_N^0 \leq z_N^*; \]
\[ \epsilon > 0; \]
\[ k = 0; \]

while $\langle c_N, x_N^k \rangle - z_N^k > \epsilon$ do

begin

solve $(ZP_N)$ and obtain $z_N^{k+1}$;
compute $\pi_N(x_N^k, z_N^{k+1})$ according to (6.22);

\[ x_N^{k+1} = x_N - \beta_N \pi(x_N^k, z_N^{k+1}) \text{ where } \beta_N = \arg \min_{\beta \geq 0} \phi(x_N^k - \beta(x_N^k, z_N^{k+1})); \]
\[ k = k + 1 \]

end

end

By simple modifications of the proofs in [56], it can be shown that the iterates $(x_N^k, z_N^k)$ produced by Algorithm $(P_N)$ have the following convergence results.

**Proposition 6.6** If $x_N^k, k \in \mathbb{Z}^+$ is the sequence of feasible solutions produced by Algorithm $(P_N)$ for the problem $(P_N)$, then $x_N^k \rightarrow x_N^*$ as $k \rightarrow \infty$.

In fact, we need a stronger statement about convergence of the iterates $x_N^k$ than is stated in Proposition 6.6, namely that $x_N^k \rightarrow x_N^*$ uniformly. We shall assume that this is true. A precise statement of this is as follows:

**Assumption 6.5** For all $\epsilon > 0$ there exists $N, \bar{q}$ such that $\|x_q^k - x_q^*\| < \epsilon$ for all $q > \bar{q}$ and $k > N$.

Assumption 6.5 is needed in the proofs of two technical lemmas which state some of the properties of the operators $Q_N(x_N^k) : X \rightarrow X$ and $Q(x^k) : X \rightarrow X$. They are required to prove the convergence properties associated with the algorithm $(ZP)$. Interested readers should consult Lemmas 6.6 and 6.7 in Section 6.6.1 of the Appendix for these results. It is an open question whether Assumption 6.5 can be relaxed. This is the topic of ongoing research.

The next result deals with the convergence of the sequence of functionals $f_i^N(x_N^k, z, \eta)$
and the functional $f_i(x^k, z, \eta)$ which are associated with the dual problems $(DN)$ and $(D)$ respectively.

**Lemma 6.4** Let $x^k$ be strictly feasible for $(P)$, $x^k_N$ be strictly feasible for $(PN)$ and $z \leq z^*$. If $x^k_N \rightarrow x^k$ as $N \rightarrow \infty$, then $f_i^N(x^k_N, z, \eta) \rightarrow f_i(x^k, z, \eta)$ in $C[0, 1]$.

**Proof:** See Section 6.6.1 in the Appendix.

The following regularity assumption is required in the proof of the next lemma.

**Assumption 6.6** Let $z^1$ and $z^2$ be feasible for $(ZP)$ associated with $x \in X$. For every $z^1 > z^2$ such that $f_i(x, z^1, \eta) \geq 0$ and $f_i(x, z^2, \eta) \geq 0$, there exists $z$ such that $z^1 > z > z^2$ and $f_i(x, z, \eta) > 0$. Let $z^*$ be optimal for $(ZP)$ associated with $x \in X$. For all $\epsilon > 0$, there exists $z \in \mathbb{R}$ such that $z^* > z > z^* - \epsilon$ and $f_i(x, z, \eta) > 0$ for all $\eta \in [0, 1]$.

**Lemma 6.5** Let $\{N_j\}_{j=1}^{\infty}$ be an increasing sequence in $\mathbb{Z}^+$. Let $x^k_{N_j}$ be feasible for $(PN_j)$ and $x^k$ be feasible for $(P)$. Suppose that $x^k_{N_j} \rightarrow x^k$ as $j \rightarrow \infty$. If $z^k_{N_j+1}$ denotes the optimal solution of $(ZPN_{N_j})$ associated with $x^k_{N_j}$ and $z^k_{N_j+1} \rightarrow z^*$ as $j \rightarrow \infty$, then $z^*$ is the global optimal solution of $(ZP)$ associated with $x^k$.

**Proof:** See Section 6.6.1 in the Appendix.

This next result shows that the relationship between the optimal solution of the problem $(ZPN)$, and that of $(ZP)$. Under certain conditions, it shows that the optimal solution of $(ZP)$ exists and can be approximated by that of $(ZPN)$, for sufficiently large $N$. It is later used to prove convergence properties of the Newton step.

**Proposition 6.7** Let $x^k$ be feasible for $(P)$, $x^k_N$ be feasible for $(PN)$ and $z^k_{N+1}$ be the optimal solution of $(ZPN)$ associated with $x_N$. Suppose that $x^k_N \rightarrow x^k$ as $N \rightarrow \infty$. If there exists $N \in \mathbb{Z}^+$ and $\bar{z} > -\infty$ such that $\bar{z} < z^k_{N+1} < \infty$ for each $N > N$, then the problem $(ZP)$ associated with $x^k$ has an optimal solution $z^{k+1}$ and $z^k_{N+1} \rightarrow z^{k+1}$ as $N \rightarrow \infty$.

**Proof:** See Section 6.6.1 of the Appendix.
6.3. CONVERGENCE RESULTS

We are now able to prove the following result relating to the Newton step (6.14). We show conditions under which (6.14) can be approximated by the finite dimensional Newton step (6.22) for \((PN)\). As a result, the updates \(x^{k+1}\) of \(x^k\), as produced by Algorithm \((P)\), can be approximated by updates resulting from Algorithm \((PN)\). This is later used to show that the iterates produced by Algorithm \((P)\) can be approximated by those produced by Algorithm \((PN)\).

**Proposition 6.8** Let \(k \in \mathbb{Z}^+\) be given. For every \(N \in \mathbb{Z}^+\) let \(x_N^k \in X_N\) be strictly feasible for \((PN)\), \(x^k \in X\) be strictly feasible for \((P)\) with \(x_N^k \rightarrow x^k\) as \(N \rightarrow \infty\). Suppose there exists \(\tilde{N} \in \mathbb{Z}^+\) and \(\bar{z} > -\infty\) such that \(\bar{z} < z_{N+1}^{k+1} < \infty\) for each \(N > \tilde{N}\) where \(z_{N+1}^{k+1}\) is the optimal solution of \((ZP_N)\) associated with \(x_N^k\). Then \(z_{N+1}^{k+1} \rightarrow z^{k+1}\) as \(N \rightarrow \infty\).

**Proof:** This proof is rather technical. The interested reader may consult Section 6.6.1 in the Appendix.

We now show that the iterates produced by Algorithm \((P)\) can in fact be approximated by those produced by Algorithm \((PN)\). More precisely, we have the following result:

**Proposition 6.9** Let \(x^0\) be the initial point for Algorithm \((P)\), and \(x_N^0 = \pi_N x^0\) the initial point of Algorithm \((PN)\), where \(\pi_N : X \rightarrow X_N\) is the orthogonal projection of \(X\) onto \(X_N\). Let \(\{x_N^k\}_{k=0}^\infty\) and \(\{x^k\}_{k=0}^\infty\) be iterates produced by \((PN)\) and \((P)\) respectively. For every \(k \in \mathbb{Z}^+\), \(x_N^k \rightarrow x^k\) as \(N \rightarrow \infty\).

**Proof:** We prove this using induction.

This is true for \(k = 0\), for let \(x^0 \in X\) be strictly feasible for \((P)\). Denote by \(x_N^0\) the projection of \(x^0\) onto \(X_N\). Then \(x_N^0\) is strictly feasible for \((PN)\) and \(x_N^0 \rightarrow x^0\) as \(N \rightarrow \infty\). Let \(z^1 \leq z^*\) solve \((ZP)\) and \(z_N^1 \leq z_N^*\) solve \((ZP_N)\). Then by Proposition 6.7 \(z_N^1 \rightarrow z^1\) as \(N \rightarrow \infty\). Thus \(\{(x_N^0, z_N^1)\}_{N=1}^\infty\) satisfies the conditions of Proposition 6.8 and \(x_N^1 \rightarrow x^1\) as \(N \rightarrow \infty\).

Suppose there is \(k \in \mathbb{Z}^+\) such that \(x_N^k \rightarrow x^k\) and \(z_N^{k+1} \rightarrow z^{k+1}\) as \(N \rightarrow \infty\). Then the conditions of Proposition 6.7 are satisfied and hence, \(x_N^{k+1} \rightarrow x^{k+1}\) as \(N \rightarrow \infty\).
The result follows from induction.

Under Assumptions 6.1, 6.2, 6.3, 6.5 and 6.6, the sequence of iterates \( \{x^k\}_{k=1}^\infty \) produced by Algorithm \((P)\) has the following weak convergence properties:

**Theorem 6.1** If \( \{x^k\}_{k=1}^\infty \) be the sequence of iterates produced by Algorithm \((P)\) and \( x^* \) is the optimal solution of \((P)\) then \( x^k \to x^* \) as \( k \to \infty \).

**Proof:** Let \( \{x^k_N\}_{k=1}^\infty \) be the sequence of iterates which result from applying Algorithm \((P_N)\) to \((P)\) with initial point \( x^0_N = \pi_N x^0 \), the orthogonal projection of \( x^0 \) on \( X_N \). Let \( \epsilon > 0 \) and \( z \in X \) be given. By Proposition 6.9, \( x^k_p \to x^k \) as \( p \to \infty \). Since strong convergence implies weak convergence, there exists \( N_1 \in \mathbb{Z}^+ \) such that \( \langle z, x^k - x^k_p \rangle < \frac{\epsilon}{3} \) for all \( p > N_1 \). By Proposition 6.6 and Assumption 6.5 \( x^k_q \to x^*_q \) uniformly so there exists \( K \in \mathbb{Z}^+ \) and \( N_2 \in \mathbb{Z}^+ \) such that \( \langle z, x^k_q - x^*_q \rangle < \frac{\epsilon}{3} \) for all \( q > N_2 \). By Proposition 6.4, there exists \( N_3 \in \mathbb{Z}^+ \) such that \( \langle z, x^*_q - x^* \rangle < \frac{\epsilon}{3} \) for all \( q > N_3 \). Thus, with \( p, q > \max\{N_1, N_2, N_3\} \) and \( k > K \), we have \( |\langle z, x^k - x^* \rangle| < \frac{\epsilon}{3} \). An immediate consequence of the definition of weak convergence is the following:

**Corollary 6.2** Let \( \{x^k\}_{k=0}^\infty \) and \( x^* \) be as in Theorem 6.1. Then \( \langle c, x^k \rangle \to \langle c, x^* \rangle \) as \( k \to \infty \).

If in addition we assume that Assumption 6.4 holds, we have the following strong convergence result:

**Theorem 6.3** Let \( \{x^k\}_{k=0}^\infty \) and \( x^* \) be as in Theorem 6.1. Then \( x^k \to x^* \) as \( k \to \infty \).

**Proof:** This proof is similar to that of Theorem 6.1. Let \( \epsilon > 0 \) be given and \( \{x^k_N\}_{k=1}^\infty \) be the sequence of iterates obtained by applying Algorithm \((P_N)\) to \((P_N)\) with initial point \( x^0_N = \pi_N x^0 \). By Proposition 6.9, there exists \( N_1 \in \mathbb{Z}^+ \) such that \( \|x^k_N - x^k_p\| < \frac{\epsilon}{3} \) for all \( q > N_1 \). By Proposition 6.6 and Assumption 6.5, there exists \( N_2, K \in \mathbb{Z}^+ \) such that \( \|x^k_q - x^*_q\| < \frac{\epsilon}{3} \) for all \( k > K \) and \( q > N_2 \). Under Assumption 6.4, it follows from Proposition 6.5 that there exists \( N_3 \in \mathbb{Z}^+ \) such that...
6.4 Applications and examples

In 1953, Bellman introduced a class of infinite linear programming problems that he called continuous linear programming (CLP):

\[
\begin{aligned}
\int_0^1 c'(s) x(s) ds & \rightarrow \min \\
\text{subject to: } B(t) \cdot x(t) + \int_0^t K(s, t) \cdot x(s) ds & \leq b(t) \\
x(t) & \geq 0
\end{aligned}
\]

(6.26)

These problems were introduced to model economic systems which he called 'Bottleneck Processes'. As its name suggests, CLP is a generalization of LP. Bellman hoped that it could be solved by making appropriate extensions of the Simplex method. However, this was not to be. In fact, Bellman even wrote that 'the analysis is decidedly difficult and it cannot be said that these problems have in any sense been tamed' [4].

Many situations are well modelled by the CLP problem (6.26); for instance, dynamic network flow problems with storage permitted at the nodes. For a more comprehensive list, refer to the work of Pullan [41] and the references cited there in. However, despite this fact, the problem (6.26) is not well known. In fact, there exists no algorithm that can solve CLP in its most general form (6.26). Several algorithms have been developed for solving a subclass of CLP called separated continuous linear programming (SCLP). Mostly, these involve some sort of discretization of the time domain of the SCLP. For more details, see [41].

In this section, we show that the potential reduction method for infinite linear programming, as stated in Algorithm (P), provides a new way to tackle (6.26). This can be done as follows. First we show that (6.26) can be approximated by a linear optimal control problem. Consider the linear system:

\[
\begin{aligned}
\dot{x}(t) &= u(t) \\
x(0) &= \xi
\end{aligned}
\]
where \( \xi \in \mathbb{R}^n \) and \( u \in L^2(0,1;\mathbb{R}^n) \) are the variables. Then
\[
x(t) = \xi + \int_0^t u(s) \, ds
\]
and hence
\[
B(t) x(t) + \int_0^t K(s,t) x(s) \, ds = B(t) \cdot (\xi + \int_0^t u(s) \, ds) + \int_0^t K(s,t) x(s) \, ds = B(t) \xi + \int_0^t (K(t,s) x(s) + B(s) u(s)) \, ds
\]
The constraint \( x(t) \geq 0 \) becomes
\[
\xi + \int_0^t u(s) \, ds \geq 0
\]
Hence, (6.26) can be written as an optimal control problem of the form
\[
\begin{aligned}
\int_0^1 c'(s) x(s) \, ds &\rightarrow \min \\
B(t) \xi + \int_0^1 (K(t,s) x(s) + B(t) u(s)) \, ds &\leq b(\eta) \\
\xi + \int_0^t u(s) \, ds &\geq 0 \\
\dot{x}(t) &= u(t), \quad x(0) = \xi
\end{aligned}
\] (6.27)
where the variable is \((\xi, (x, u)) \in \mathbb{R}^n \times L^2(0,1;\mathbb{R}^n) \times L^2(0,1;\mathbb{R}^n)\). Importantly, this is a problem of the form \((P)\) to which we can apply Algorithm \((P)\).

More generally, (6.27) is a problem of the form
\[
\begin{aligned}
\int_0^1 (c_1 \cdot x(t) + c_2 \cdot u(t)) \, dt \\
d_i(\eta) \xi + \int_0^1 (a_i(t, \eta)' \cdot x(t) + b_i(t, \eta)' \cdot u(t)) \, dt &\leq \gamma_i(\eta), \quad i = 1, \ldots, m \\
\dot{x}(t) &= A(t) x(t) + B(t) u(t), \quad x(0) = \xi
\end{aligned}
\] (6.28)
where the variable is \((\xi, (x, u)) \in \mathbb{R}^n \times L^2(0,1;\mathbb{R}^n) \times L^2(0,1;\mathbb{R}^m)\). When \( \xi \in \mathbb{R}^n \) is fixed, then (6.28) is a linear optimal control problem with infinitely many linear constraints, and is a problem of the form \((P)\).

In the remainder of this section, we show how Algorithm \((P)\) can be used to tackle the problem (6.28). However, we begin with a word of caution. We remind the reader that in the convergence analysis of Algorithm \((P)\) in Section 6.3, many assumptions were made. Indeed, determining which of these assumptions can be relaxed and which are necessary is the topic of continuing research. Some of these assumptions may not
be satisfied by the problem (6.28). However, the example in this section serves to show how Algorithm \((P)\) can be applied to (6.34). How this algorithm may behave is another matter.

In Algorithm \((P)\), the most difficult step is the Newton step, as defined by (6.10). In order to implement Algorithm \((P)\), it is a necessary requirement that the Newton step is implementable. In the remainder of this section, we show how the Newton step can be calculated for (6.28). For simplicity, we shall restrict ourselves to the scalar time invariant and assume that \(\xi\) is not a variable but rather, that \(\xi = 0\). This method of calculating the Newton step can be applied to the more general problem (6.28). More specifically, let \(c_1, c_2 \in \mathbb{R}\) be constants and define the cost functional by

\[
f(x, u) = \int_0^1 (c_1 \cdot x(t) + c_2 \cdot u(t)) \, dt	ag{6.29}
\]

Let \(a, b \in \mathbb{R}\) be fixed constants, and define

\[
a(t, \eta) = \begin{cases} a, & t \leq \eta \\ 0, & t > \eta \end{cases}
\]

\[
b(t, \eta) = \begin{cases} b, & t \leq \eta \\ 0, & t > \eta \end{cases}
\]

The constraint functional is given by

\[
f_1((x, u), \eta) = \int_0^1 (a(t, \eta) \cdot x(t) + b(t, \eta) \cdot u(t)) \, dt
\]

Let \(\alpha, \beta \in \mathbb{R}\) be fixed and

\[
X = \{(x, u) \in L^2(0, 1; \mathbb{R}) \times L^2(0, 1; \mathbb{R}) : \dot{x}(t) = \alpha \cdot x(t) + \beta \cdot u(t), \, x(0) = 0\}
\]

Denoting \(y = (x, u)\), we consider the problem

\[
\begin{aligned}
f(y) & \rightarrow \min \\
f_1(y, \eta) & \leq \gamma, \quad \eta \in [0, 1] \\
y & \in X
\end{aligned}
\]

where \(\gamma \in \mathbb{R}\) is also fixed. Let \(z^*\) denote the optimal cost and \(z \leq z^*\) be any lower bound. Suppose that \(y \in X\) is strictly feasible for (6.34). Then the potential function for (6.34) as defined by (6.4) is given by

\[
\phi(y, z) = \rho \ln (f(y) - z) - \int_0^1 s(y, \eta) \, d\eta
\]
where \( s(y, \eta) = \gamma - f_1(y, \eta) \). The gradient of \( \phi(y, z) \) with respect to \( y \) is given by

\[
\nabla \phi(y, z) = \begin{bmatrix}
\frac{\rho c_1}{f(y) - z} + \int_0^1 \frac{a(t, \eta)}{s(y, \eta)} d\eta \\
\frac{\rho c_2}{f(y) - z} + \int_0^1 \frac{b(t, \eta)}{s(y, \eta)} d\eta \\
\frac{\rho c_1}{f(y) - z} + \int_t^1 \frac{a}{s(y, \eta)} d\eta \\
\frac{\rho c_2}{f(y) - z} + \int_t^1 \frac{b}{s(y, \eta)} d\eta
\end{bmatrix}
\]  

(6.36)

The Newton step \( \nabla_y \phi(y, z) \in X \) is defined by (6.10) and we denote this by

\[
\nabla_y \phi(y, z) = \begin{pmatrix} \xi \\ r \end{pmatrix}
\]

Let \( Q(y) \) be defined as in (6.8). It follows that

\[
Q(y) \cdot \nabla_y \phi(y, z) = \begin{bmatrix}
\int_0^1 d(\eta) \frac{a(t, \eta)}{s(y, \eta)} d\eta \\
\int_0^1 d(\eta) \frac{b(t, \eta)}{s(y, \eta)} d\eta \\
\int_t^1 d(\eta) \frac{a}{s(y, \eta)} d\eta \\
\int_t^1 d(\eta) \frac{b}{s(y, \eta)} d\eta
\end{bmatrix}
\]  

(6.37)

where

\[
d(\eta) = \int_0^1 [a(t, \eta) \cdot \xi(t) + b(t, \eta) \cdot r(t)] d\eta = \int_0^\eta (a \cdot \xi(t) + b \cdot r(t)) dt
\]

(6.38)

By (6.10), we also have that

\[
Q(y) \cdot \nabla_y \phi(y, z) = \nabla \phi(y, z)
\]

(6.39)

Therefore, it follows from (6.36), (6.37) and (6.39) that

\[
\int_t^1 d(\eta) \frac{a}{s(y, \eta)} d\eta = \frac{\rho c_1}{f(y) - z} + \int_t^1 \frac{a}{s(y, \eta)} d\eta
\]

(6.40)

\[
\int_t^1 d(\eta) \frac{b}{s(y, \eta)} d\eta = \frac{\rho c_2}{f(y) - z} + \int_t^1 \frac{b}{s(y, \eta)} d\eta
\]

(6.41)

To calculate \( \nabla_y \phi(y, z) = (\xi(t), r(t)) \in X \), let \( \psi(t, s) \) be the fundamental solution of \( \dot{x}(t) = \alpha x \). It follows from \( \nabla_y \phi(y, z) = (\xi(t), r(t)) \in X \) and the definition of \( X \) that

\[
\xi(t) = \int_0^t \psi(t, s) \beta r(s) ds
\]

(6.42)
6.5. **CONCLUSIONS**

Substituting (6.42) into (6.38) gives the integral equation

\[ d(\eta) = \int_0^\eta a \int_0^\eta \psi(t, s) \cdot \beta \cdot r(s) \, ds \, dt + \int_0^\eta b \cdot r(t) \, dt \]  \hspace{1cm} (6.43)

Hence, to calculate the Newton step \( \nabla_y \phi(y, z) = (\xi, r) \) we need to solve (6.40)-(6.41) for \( d(\eta) \), and the integral equation (6.42)-(6.43) for \( \nabla_y \phi(y, z) \).

### 6.5 Conclusions

We have derived an interior point algorithm for solving the infinite linear programming problem on a Hilbert space by generalizing the potential reduction interior point method introduced by Ye in [56]. Under certain (rather stringent) assumptions, we have proven that the iterates produced by this algorithm converge to the optimal solution. As an example, we showed how this algorithm could be used to solve continuous linear programming problems, and linear optimal control problems with continuous state constraints. In applying this algorithm to these problems, the key step in implementations (as with all interior point methods) is performing the so called Newton step. In this case, the Newton step is determined by solving a pair of integral equations.

Several open questions remain. In particular, we have not derived a complexity bound for this infinite dimensional generalization of the algorithm. Many interior point methods for finite dimensional problems are known to have a polynomial worse case complexity bound. In deriving complexity, a measure of problem size is required. In the finite dimensional case, this is given in terms of the number of variables and the number of constraints. In our case however, such a definition does not make sense since both the variable and constraint spaces are infinite dimensional. In order to study complexity, a generalized definition of problem size, as introduced in the paper by Renegar [42], is needed.

In the convergence proof of this algorithm, several assumptions are invoked. It is currently unclear whether these assumptions are necessary, or whether they can be relaxed. Indeed, infinite dimensional generalizations of finite dimensional problems can converge to non-optimal points, as shown by Powell in [40]. An important open question is determining which of these assumptions can be relaxed, and which are necessary for the algorithm to converge to an optimal solution.
Finally, it remains to be seen how well this algorithm performs numerically. There is scope for further work in the numerical implementation of this algorithm.

6.6 Appendix

6.6.1 Relegated proofs from Section 6.3

Proof of Lemma 6.3:

Let $\epsilon > 0$ be given and consider the sets

$$
\Gamma = \{ x \in X : \langle a_i(\eta), x \rangle \leq b_i(\eta), i = 1 \cdots, m \} 
$$

$$
\Gamma^\epsilon = \{ x \in X : \langle a_i(\eta), x \rangle \leq b_i(\eta) + \epsilon, i = 1 \cdots, m \} 
$$

$$
\Gamma_N = \{ x \in X_N : \langle a_i(\eta_j), x \rangle \leq b_i(\eta_j), \eta_j = \frac{j}{2^N}, j = 1 \cdots, 2^N \text{ and } i = 1 \cdots, m \} 
$$

Since $\Gamma$ is a strict subset of $\Gamma^\epsilon$, it follows that for $N$ sufficiently large, $\Gamma_N \subset \Gamma^\epsilon$. Let $(P^\epsilon)$ be the problem $(P)$ with constraint set $\Gamma$ replaced by $\Gamma^\epsilon$ and $(D^\epsilon)$ the problem dual to $(P^\epsilon)$. Then

$$
(D^\epsilon) \begin{cases}
- \sum_{i=1}^{m} \int_{0}^{1} b_i(\eta) \, dv_i(\eta) - \epsilon \sum_{i=1}^{m} (v_i(1) - v_i(0)) \to \max \\
\quad c + \sum_{i=1}^{m} \int_{0}^{1} a_i(\eta) \, dv_i(\eta) = 0 \\
\quad v_i(\eta) \in BV[0,1], \text{ increasing}
\end{cases}
$$

Let $x^*, \epsilon$ denote the optimal for $(P^\epsilon)$ and $v^*(\eta)$ the optimal solutions for $(D)$. Then $v^*(\eta)$ is feasible for $(D^\epsilon)$. It follows that

$$
\langle c, x^* \rangle \geq \langle c, x^{*, \epsilon} \rangle 
$$

$$
\geq - \sum_{i=1}^{m} \int_{0}^{1} b_i(\eta) \, dv_i^*(\eta) - \epsilon \sum_{i=1}^{m} (v_i^*(1) - v_i^*(0)) 
$$

$$
= z^* - \epsilon \sum_{i=1}^{m} (v_i^*(1) - v_i^*(0))
$$

The first inequality follows from the fact that $\Gamma \subset \Gamma^\epsilon$; the second from the feasibility of $v(\eta)$ for $(D^\eta)$ and the fact that the dual cost corresponding to a dual feasible solution is a lower bound for the optimal (primal) cost. The equality follows from the optimality of $v^*(\eta)$ for $(D)$. Since $z^* - \epsilon \sum_{i=1}^{m} (v_i^*(1) - v_i^*(0)) \to \langle c, x^* \rangle = z^*$ as $\epsilon \to 0$, it follows that $\langle c, x^{*, \epsilon} \rangle \to \langle c, x^* \rangle$ as $\epsilon \to 0$. On the other hand, $\Gamma$ is a
strict subset of $\Gamma^*$ so it follows that for $N$ sufficiently large, $\Gamma_N \subseteq \Gamma^*$. Therefore 
\[ \langle c_N, x_N^* \rangle = \langle c, x_N^* \rangle \geq \langle c, x^* \rangle \] for all such $N$. Let $\pi_N : X \rightarrow X_N$ be the orthogonal projection of $X$ onto $X_N$. It is easily shown that $\pi_N x^*$ is feasible for $(P_N)$. Therefore, for all $\epsilon > 0$, $\langle c_N, \pi_N x^* \rangle \geq \langle c_N, x_N^* \rangle \geq \langle c, x^* \epsilon \rangle$ for $N$ sufficiently large. However, $\langle c_N, \pi_N x^* \rangle \rightarrow \langle c, x^* \rangle$ as $N \rightarrow \infty$ and from above, $\langle c, x^* \epsilon \rangle \rightarrow \langle c, x^* \rangle$ as $\epsilon \rightarrow 0$ from which the result follows.

The following two lemmas are required in the proof of Lemma 6.4.

**Lemma 6.6** Let $x_N^k \in X_N$ be strictly feasible for $(P_N)$ and $x^k \in X$ be strictly feasible for $(P)$ such that $x_N^k \rightarrow x^k$ as $N \rightarrow \infty$. Then $Q_N(x_N^k) \rightarrow Q(x^k)$ uniformly as $N \rightarrow \infty$.

**Proof:** Let $y \in X$. Noting that

\[
\begin{align*}
Q_N(x_N^k) \cdot y &= \sum_{i=1}^{m} \sum_{j=1}^{2^N} \left\langle \frac{a_i(\eta_j)}{s_i(x_N^k, \eta_j)}, y \right\rangle \frac{a_i(\eta_j)}{s_i(x_N^k, \eta_j)} d\eta \\
&= \sum_{i=1}^{m} \int_{0}^{1} \left[ \sum_{j=1}^{2^N} \left\langle \frac{a_i(\eta_j)}{s_i(x_N^k, \eta_j)}, y \right\rangle \frac{a_i(\eta_j)}{s_i(x_N^k, \eta_j)} I_{A_i}(\eta) \right] d\eta 
\end{align*}
\]

it follows that if $\|y\| = 1$, we have

\[
\begin{align*}
\left\| Q(x^k) \cdot y - Q_N(x_N^k) \cdot y \right\| &\leq \sum_{i=1}^{m} \left\| \int_{0}^{1} \left\langle \frac{a_i(\eta)}{s_i(x^k, \eta)}, y \right\rangle \frac{a_i(\eta)}{s_i(x^k, \eta)} d\eta \\
&- \int_{0}^{1} \left[ \sum_{j=1}^{2^N} \left\langle \frac{a_i(\eta)}{s_i(x_N^k, \eta_j)}, y \right\rangle \frac{a_i(\eta)}{s_i(x_N^k, \eta_j)} I_{A_i}(\eta) \right] d\eta \right\|
\end{align*}
\]

Now, let

\[
S_N^i(y) = \left\| \int_{0}^{1} \left\langle \frac{a_i(\eta)}{s_i(x^k, \eta)}, y \right\rangle \frac{a_i(\eta)}{s_i(x^k, \eta)} d\eta \\
- \int_{0}^{1} \left[ \sum_{j=1}^{2^N} \left\langle \frac{a_i(\eta)}{s_i(x_N^k, \eta_j)}, y \right\rangle \frac{a_i(\eta)}{s_i(x_N^k, \eta_j)} I_{A_i}(\eta) \right] d\eta \right\|
\]

\[
\leq \left\| \int_{0}^{1} \left\langle \frac{a_i(\eta)}{s_i(x^k, \eta)} - \sum_{j=1}^{2^N} \frac{a_i(\eta)}{s_i(x_N^k, \eta_j)} I_{A_i}(\eta), y \right\rangle \frac{a_i(\eta)}{s_i(x^k, \eta)} d\eta \right\|
\]

\[
+ \left\| \int_{0}^{1} \sum_{j=1}^{2^N} \left( \frac{1}{s_i(x^k, \eta_j)} - \frac{1}{s_i(x_N^k, \eta_j)} \right) \left\langle \frac{a_i(\eta)}{s_i(x^k, \eta)}, y \right\rangle \frac{a_i(\eta)}{s_i(x_N^k, \eta_j)} I_{A_i}(\eta) d\eta \right\|
\]
By continuity of \( a_i(\eta) \) and the strict feasibility of \( x^k \) (that is \( s_i(x^k, \eta) > 0 \) for all \( \eta \in [0,1] \)), \( \frac{a_i(\eta)}{s_i(x^k, \eta)} \) and \( |a_i(\eta)| \) are bounded on \([0,1]\). Since \( s_i(x^k, \eta) \to s_i(x^k, \eta) \) uniformly on \([0,1]\) as \( N \to \infty \), it follows from \( s_i(x^k, \eta) > 0 \) for all \( \eta \in [0,1] \) that there exists \( N_1 \in \mathbb{Z}^+ \) and \( \delta > 0 \) such that \( s_i(x^k, \eta) \geq \delta \) on \([0,1]\) for all \( N > N_1 \). Hence, there exists \( \gamma < \infty \) such that for all \( N > N_1 \) and \( \eta \in [0,1] \)

Therefore there exists \( K < \infty \) such that for all \( N > N_1 \)

\[
S^N_i(\eta) \leq \frac{K}{4} \sup_{\eta \in [0,1]} \left\| \frac{a_i(\eta)}{s_i(x^k, \eta)} - \sum_{j=1}^{2N} \frac{a_i(\eta_j)}{s_i(x^k, \eta_j)} I_{A_j}(\eta) \right\| + \frac{K}{4} \left\| \frac{1}{s_i(x^k, \eta)} - \frac{1}{s_i(x^k, \eta)} \right\|_{C[0,1]} 
\]

\[
+ \frac{K}{4} \sup_{\eta \in [0,1]} \sum_{j=1}^{2N} \left| \frac{1}{s_i(x^k, \eta_j)} - \frac{1}{s_i(x^k, \eta_j)} \right| I_{A_j}(\eta) 
\]

\[
+ \frac{K}{4} \sup_{\eta \in [0,1]} \sum_{j=1}^{2N} |a_i(\eta) - a_i(\eta_j)| I_{A_j}(\eta) 
\]

Since \( a_i(\eta) \) is continuous on the compact set \([0,1]\), it is uniformly continuous; that is for all \( \varepsilon > 0 \) there exists \( \delta > 0 \) such that \( |\eta_1 - \eta_0| < \delta \) implies \( |a_i(\eta_1) - a_i(\eta_0)| < \varepsilon \) for all \( \eta_0, \eta_1 \in [0,1] \). Hence, there exists \( N_1 \in \mathbb{Z}^+ \) such that for all \( N > N_1 \)

\[
\sup_{\eta \in [0,1]} |a_i(\eta) - a_i(\eta_j)| I_{A_j}(\eta) < \frac{\varepsilon}{K} 
\]
Similarly, by uniform continuity of \( \frac{a_i(\eta)}{s_i(x^k, \eta)} \), there exists \( \bar{N}_2 \in \mathbb{Z}^+ \) such that for all \( N > \bar{N}_2 \)

\[
\sup_{\eta \in [0,1]} \left\| \frac{a_i(\eta)}{s_i(x^k, \eta)} - \sum_{j=1}^{2N} \frac{a_i(\eta_j)}{s_i(x^k, \eta_j)} I_{A_j}(\eta) \right\| < \frac{\epsilon}{K}
\]

Since \( s_i(x^k_N, \eta) \rightarrow s_i(x^k, \eta) \) uniformly on \( C[0,1] \) and \( s_i(x^k, \eta) > 0 \) for all \( \eta \in [0,1] \), there exists \( \bar{N}_3 \in \mathbb{Z}^+ \) such that for all \( N > \bar{N}_3 \)

\[
\left\| \frac{1}{s_i(x^k, \eta)} - \frac{1}{s_i(x^k_N, \eta)} \right\|_{C[0,1]} < \frac{\epsilon}{K}
\]

Finally, observing that

\[
\sum_{j=1}^{2N} \left| \frac{1}{s_i(x^k, \eta)} - \frac{1}{s_i(x^k_N, \eta_j)} \right| I_{A_j}(\eta)
\]

\[
< \sum_{j=1}^{2N} |b_i(\eta) - b_i(\eta_j)| I_{A_j}(\eta) + \left\| x^k - x^k_N \right\| \left( \sum_{j=1}^{2N} ||a_i(\eta_j)|| \right)
\]

\[
+ \left\| x^k \right\| \left( \sum_{j=1}^{2N} |a_i(\eta) - a_i(\eta_j)| I_{A_j}(\eta) \right)
\]

It follows from the continuity (and hence uniform continuity) of \( a_i(\eta), b_i(\eta) \) and \( x^k_N \rightarrow x^k \) as \( N \rightarrow \infty \) that there exists \( \bar{N}_4 \in \mathbb{Z}^+ \) such that for all \( N > \bar{N}_4 \)

\[
\sup_{\eta \in [0,1]} \sum_{j=1}^{2N} \left| \frac{1}{s_i(x^k, \eta)} - \frac{1}{s_i(x^k_N, \eta_j)} \right| I_{A_j}(\eta) < \frac{\epsilon}{K}
\]

for all \( N > \bar{N}_4 \). Therefore, for every \( N > \max\{\bar{N}_1, \ldots, \bar{N}_4\} \) and \( y \in X, ||y|| = 1 \) we have \( S^N_i(y) < \epsilon \). Hence, there exists \( \bar{N} \in \mathbb{Z}^+ \) such that for all \( N > \bar{N} \) and \( y \in X \) with ||y|| = 1, we have

\[
\left\| Q(x^k) \cdot y - Q_N(x^k_N) \cdot y \right\| \leq \sum_{i=1}^{m} S^N_i(y) < \epsilon
\]

and it follows that \( Q_N(x^k_N) \rightarrow Q(x^k) \) uniformly.

**Lemma 6.7** Let \( x^k_N \) and \( x^k \) be as in Lemma 6.6. Then there exists \( \lambda > 0 \) and \( \bar{N} \in \mathbb{Z}^+ \) such that for all \( N > \bar{N}, \) \( \langle Q_N(x^k_N) \cdot x, x \rangle \geq \lambda \|x\|^2 \) for all \( x \in X \).
Proof: Since $Q(z^k) : X \to X$ is a strictly positive, bounded self-adjoint linear operator, there exists $\lambda > 0$ such that $\langle Q(z^k) \cdot z, z \rangle \geq 2\lambda$ for all $x \in X$, $\|x\| = 1$. By Lemma 6.6, there exists $\tilde{N} \in \mathbb{Z}^+$ such that 

$$\left| \langle Q(z^k) \cdot z, x \rangle - \langle Q_N(z_N^k) \cdot z, x \rangle \right| \leq \|Q(z^k) - Q_N(z_N^k)\| < \lambda$$

for all $N > \tilde{N}$ and $x \in X$ with $\|x\| = 1$. Since $\langle Q(z^k) \cdot z, x \rangle \geq 2\lambda$ for all $x \in X$, $\|x\| = 1$, it follows that $\langle Q_N(z_N^k) \cdot z, x \rangle \geq \lambda$ for all $N > \tilde{N}$ from which the result follows.

Proof of Lemma 6.4:

First we show that

$$\sum_{j=1}^{2N} \frac{a_i(\eta_j)}{s_i(x_N^k, \eta_j)} \frac{1}{2^N} \to \int_0^1 \frac{a_i(\eta)}{s_i(x^k, \eta)} \, d\eta \text{ as } N \to \infty$$

where $\eta_j = \frac{1}{2^j}$. Indeed it can be shown that

$$\left\| \int_0^1 \frac{a_i(\eta)}{s_i(x^k, \eta)} \, d\eta - \sum_{j=1}^{2N} \frac{a_i(\eta_j)}{s_i(x_N^k, \eta_j)} \frac{1}{2^N} \right\| \leq \left\| \int_0^1 \frac{a_i(\eta)}{s_i(x^k, \eta)} \, d\eta - \sum_{j=1}^{2N} \frac{a_i(\eta_j)}{s_i(x_N^k, \eta_j)} \frac{1}{2^N} \right\| + \sum_{j=1}^{2N} \frac{\|a_i(\eta_j)\|^2}{s_i(x_N^k, \eta_j) \cdot s_i(x_N^k, \eta_j)} \frac{1}{2^N} \cdot \|x_N^k - x^k\|$$

Since $s_i(x^k, \eta) = b_i(\eta) - \langle a_i(\eta), x^k \rangle > 0$ for all $\eta \in [0, 1]$ and $x_N^k \to x^k$ as $N \to \infty$, it follows that $s_i(x_N^k, \eta) \to s_i(x^k, \eta)$ uniformly. Therefore, $s_i(x_N^k, \eta) > 0$ for all $\eta \in [0, 1]$ for $N$ sufficiently large. Furthermore, by the boundedness of $\|a_i(\eta)\|$ that there is a finite $K > 0$ and $N_1 \in \mathbb{Z}^+$ such that

$$\sum_{j=1}^{2N} \frac{\|a_i(\eta_j)\|^2}{s_i(x_N^k, \eta_j) \cdot s_i(x_N^k, \eta_j)} \frac{1}{2^N} < K$$

for all $N > N_1$. Since $\|x_N^k - x^k\| \to 0$ as $N \to \infty$, there exists $N_2$ such that

$$\sum_{j=1}^{2N} \frac{\|a_i(\eta_j)\|^2}{s_i(x_N^k, \eta_j) \cdot s_i(x_N^k, \eta_j)} \frac{1}{2^N} \cdot \|x_N^k - x^k\| < K \cdot \frac{\epsilon}{2K} = \frac{\epsilon}{2}$$
for every $N > \max\{N_1, N_2\}$. By (2.67) in Lemma 2.2, there exists $N_3 \in \mathbb{Z}^+$ such that
\[
\left\| \int_0^1 \frac{a_i(\eta)}{s_i(x^k, \eta)} \, d\eta - \sum_{j=1}^{2N} \frac{a_i(\eta_j)}{s_i(x^k, \eta_j)} \frac{1}{2^N} \right\| < \frac{\epsilon}{2}
\]
for $N > N_3$. Therefore
\[
\left\| \int_0^1 \frac{a_i(\eta)}{s_i(x^k, \eta)} \, d\eta - \sum_{j=1}^{2N} \frac{a_i(\eta_j)}{s_i(x^k, \eta_j)} \frac{1}{2^N} \right\| < \epsilon
\]
for $N > \max\{N_1, N_2, N_3\}$. Next we show that
\[
Q_N(x^k_N)^{-1} \cdot \nabla \phi_N(x^k_N, z) \rightarrow Q(x^k)^{-1} \cdot \nabla \phi(x^k, z)^{-1} \text{ as } N \rightarrow \infty.
\]
For convenience we shall denote
\[
\xi^k_N = Q_N(x^k_N)^{-1} \cdot \nabla \phi_N(x^k_N, z), \quad \xi^k = Q(x^k)^{-1} \cdot \nabla \phi(x^k, z).
\]
Since $\nabla \phi_N(x^k_N, z) \rightarrow \nabla \phi(x^k, z)$ as $N \rightarrow \infty$, it follows that $Q_N(x^k_N)^{-1} \cdot \xi^k_N \rightarrow Q(x^k)^{-1} \cdot \xi^k$ and hence, $Q(x^k)^{-1} \cdot Q_N(x^k_N)^{-1} \cdot \xi^k_N \rightarrow \xi^k$ as $N \rightarrow \infty$. Note also that the sequence $\{\xi^k_N\}_{N=1}^\infty$ is bounded. Indeed, by Lemma 6.7 there exists an $\bar{N} \in \mathbb{Z}^+$ and $\lambda > 0$ such that
\[
\left\| \xi^k_N \right\| < \frac{1}{\lambda} \left\| Q_N(x^k_N)^{-1} \cdot \xi^k_N \right\|
\]
for all $N > \bar{N}$. The boundedness of the sequence $\{\left\| Q_N(x^k_N) \cdot \xi^k_N \right\|\}_{N=1}^\infty$ follows from the convergence of the sequence $Q_N(x^k_N)^{-1} \cdot \xi^k_N \rightarrow Q(x^k)^{-1} \cdot \xi^k$ and fact that every convergent sequence is bounded. Also, by virtue of the Bounded Inverse Theorem [21], the inverse operator $Q(x^k)^{-1} : X \rightarrow X$ is a bounded linear operator. Since
\[
\left\| \xi^k - \xi^k_N \right\| \leq \left\| \xi^k - Q(x^k)^{-1} \cdot Q_N(x^k_N)^{-1} \cdot \xi^k_N \right\|
\]
\[
+ \left\| Q(x^k)^{-1} \right\| \cdot \left\| Q_N(x^k_N) - Q(x^k) \right\| \left\| \xi^k_N \right\|
\]
it follows that $\xi^k_N \rightarrow \xi^k$ as $N \rightarrow \infty$.

We now prove that $f^N_i(x^k_N, z, \eta) \rightarrow f_i(x^k, z, \eta)$ as $N \rightarrow \infty$. Indeed, it can be shown (after some manipulation) that there exists $\alpha, \beta < \infty$ such that
\[
\max_{\eta \in [0, 1]} \left| f_i(x^k, z, \eta) - f^N_i(x^k_N, z, \eta) \right| \leq \alpha \left\| x^k - x^k_N \right\| + \beta \left\| \xi^k - \xi^k_N \right\|
\]
from which the convergence in $C[0, 1]$ follows.
Proof of Lemma 6.5:

First we show that \( z^* \) is feasible for \((ZP)\). For if not, there exists an open subinterval \( I \subset [0, 1] \) such that \( f_i(\eta, x^k, z^*) < 0 \) for all \( \eta \in I \). On the other hand, for every \( j \in \mathbb{Z}^+ \) and \( \eta \in I_{Nj}, f_i^N(\eta, x^k_{Nj}, z^k_{Nj}) \geq 0 \). From Lemma 6.4, \( f_i^N(\eta, x^k_{Nj}, z^k_{Nj}) \rightarrow f_i(\eta, x^k, z^*) \) in \( C[0, 1] \) as \( N \rightarrow \infty \). Therefore, \( f_i(\eta, x^k, z^*) \geq 0 \) for all \( \eta \in [0, 1] \) which is a contradiction.

Next we show that \( z^* \) is the unique global maximum. Suppose that this is not true. Then there exists \( \bar{z} > z^* \) such that \( \bar{z} \) is feasible. By Assumption 6.6, there exists \( z \in \mathbb{R} \) such that \( \bar{z} > z > z^* \) with \( f_i(x^k, z, \eta) > 0 \) for all \( \eta \in [0, 1] \). Now \( f_i^N(x^k_{Nj}, z, \eta) \rightarrow f_i(x^k, \bar{z}, \eta) > 0 \) in \( C[0, 1] \) as \( N \rightarrow \infty \) so there exists \( \bar{N} \in \mathbb{Z}^+ \) such that \( f_i^N(x^k_{Nj}, \bar{z}, \eta) > 0 \) for all \( N > \bar{N} \) and \( \eta \in [0, 1] \). Therefore, \( \bar{z} > z^* \) is feasible which contradicts the optimality of \( z^* \) for \((ZP)\) since \( z^* \rightarrow z^* \). Hence, \( z^* \) must be the global maximum for \((ZP)\).

Proof of Proposition 6.7:

Consider first the convergence of the sequence \( \{z^{k+1}_{Nj}\}_{j=1}^\infty \). From Lemma 6.3 and the fact that \( \langle c, x^*_N \rangle = \langle c, x^*_N \rangle \), it follows that the sequence of optimal costs \( \langle c, x^*_N \rangle \rightarrow \langle c, x^* \rangle \) as \( N \rightarrow \infty \). Therefore, since every convergent sequence is bounded there exists \( \gamma \in \mathbb{R} \) such that \( \langle c, x^*_N \rangle < \gamma \) for all \( N \in \mathbb{Z}^+ \). Therefore, \( \bar{z} < z^k_{Nj+1} < \gamma \) for every \( N > \bar{N} \) so \( \{z^k_{Nj+1}\}_{j=1}^\infty \) is a bounded sequence in \( \mathbb{R} \). By compactness of the interval \([\bar{z}, \gamma]\), there exists a convergent subsequence \( \{z^k_{Nj}\}_{j=1}^\infty \) such that \( z^k_{Nj} \rightarrow z^k+1 \) as \( j \rightarrow \infty \) for some \( z^k+1 \in [\bar{z}, \gamma] \). It follows from Lemma 6.5 that \( z^k+1 \) is the unique global optimal solution for \((ZP)\) associated with \( x^k \). Suppose that \( z^k+1 \neq z^k+1 \) as \( N \rightarrow \infty \). Then there exists a subsequence \( \{z^k_{Nj}\}_{j=1}^\infty \) of \( \{z^k_{Nj}\}_{j=1}^\infty \) with \( z^k_{Nj} \rightarrow z^k+1 \) as \( j \rightarrow \infty \), \( z^k+1 \neq z^k+1 \). By Lemma 6.5, \( z^k+1 \) is the unique global maximum of \((ZP)\) associated with \( x^k \) and hence, \( z^k+1 = z^k+1 \) which is a contradiction. Therefore, the sequence \( \{z^k_{Nj}\}_{j=1}^\infty \) converges to the unique global optimum.

Proof of Proposition 6.8:

By Proposition 6.7, it follows that \( z^k_{Nj} \rightarrow z^k+1 \) as \( N \rightarrow \infty \) where \( z^k+1 \) is optimal of \((ZP)\) associated with \( x^k \). Therefore, \( \nabla \phi_N(x^k_{Nj}, z^k_{Nj}) \rightarrow \nabla \phi(x^k, z^k+1) \) as \( N \rightarrow \infty \).
6.6. APPENDIX

∞. With this result, the proof of Lemma 6.4 is easily generalized to show that $\pi_N(x_N^k, z_N^{k+1}) \to \pi(x^k, z^{k+1})$ as $N \to \infty$. Let

$$
\beta_N^* = \arg \min_{\beta \geq 0} \phi_N(x_N^k - \beta \pi_N(x_N^k, z_N^{k+1}), z_N^{k+1})
$$

$$
\beta^* = \arg \min_{\beta \geq 0} \phi(x^k - \beta \pi(x^k, z^{k+1}), z^{k+1})
$$

We now show that $\beta_N^* \to \beta^*$ as $N \to \infty$. Suppose that $\beta_N^* \not\to \beta^*$. Since $\{\beta_N^*\}_{N=1}^{\infty}$ is a bounded sequence (since the feasible set is bounded) with $\beta_N^* \leq \gamma$ say, it follows from the compactness of $[0, \gamma]$ that there exists a subsequence $\{\beta_{N_j}^*\}_{j=1}^{\infty}$ such that $\beta_{N_j} \to \xi^*$, $\xi^* \neq \beta^*$. For convenience, we denote

$$
y_{N_j} = x_N^k - \beta_{N_j}^* \pi_N(x_N^k, z_N^{k+1})
$$

$$
y^{k*} = x^k - \xi^* \pi(x^k, z^{k+1})
$$

Then $y_{N_j} \to y^{k*}$ as $j \to \infty$. For such $y_{N_j}$ we define the following functional on $[0, 1]$:

$$
g_{N_j}(y_{N_j}, \eta) = \sum_{k=1}^{2N} \ln \left( b_i(\eta_k) - \langle a_i(\eta_k), y_{N_j} \rangle \right) \cdot I_{A_k}(\eta)
$$

Then $g_{N_j}(y_{N_j}, \eta)$ is a piecewise constant functional of $\eta \in [0, 1]$. Furthermore, if for each $y \in X$ satisfying $b_i(\eta) - \langle a_i(\eta), y \rangle > 0$ we define

$$
g(y, \eta) = \ln \left( b_i(\eta) - \langle a_i(\eta), y \rangle \right)
$$

then $g(y^{k*}, \eta)$ is defined and $g_{N_j}(y_{N_j}, \eta) \to g(y^{k*}, \eta)$ for all $\eta \in [0, 1]$. Furthermore, since $b_i(\eta) - \langle a_i(\eta), y^{k*} \rangle > 0$ for all $\eta \in [0, 1]$ and $b_i(\eta) - \langle a_i(\eta), y_{N_j} \rangle \to b_i(\eta) - \langle a_i(\eta), y^{k*} \rangle$ uniformly on $C[0,1]$ as $j \to \infty$, it follows that there exists a $K > 0$ such that $|g(y_{N_j}, \eta)| \leq K$ for all $\eta \in [0,1]$, and $N_j$ sufficiently large. By the Lebesgue dominated convergence theorem

$$
\int_0^1 g_{N_j}(y_{N_j}, \eta) \, d\eta \to \int_0^1 g(y^{k*}, \eta) \, d\eta
$$

Noting that

$$
\int_0^1 g_{N_j}(y_{N_j}, \eta) \, d\eta = \frac{1}{2N_j} \sum_{j=1}^{2N_j} \ln \left( b_i(\eta_j) - \langle a_i(\eta_j), y_{N_j} \rangle \right)
$$

and $\ln(\langle c_{N_j}, y_{N_j} \rangle - z_{N_j}^{k+1}) \to \ln(\langle c, y^{k*} \rangle - z^{k+1})$ as $j \to \infty$, it follows that $\phi_{N_j}(y_{N_j}, z_{N_j}^{k+1}) \to \phi(y^{k*}, z^{k+1})$ as $j \to \infty$. Let

$$
y^* = x^k - \beta^* \pi(x^k, z^{k+1})
$$
By the definition of $\beta^*$, we have $\phi(y^*, z^{k+1}) < \phi(y^{*,*}, z^{k+1})$. On the other hand, since $\|a_i(\eta)\|$ is uniformly bounded on $[0, 1]$ and $b_i(\eta) - \langle a_i(\eta), y^* \rangle > 0$ for all $i = 1, \ldots, m$, $\eta \in [0, 1]$ and

$$y^*_{N_j} = x^{k}_{N_j} - \beta^*_{N_j}(x^{k}_{N_j}, z^{k+1}_{N_j}) \rightarrow y^*$$

it follows from

$$\left| b_i(\eta) - \langle a_i(\eta), y^* \rangle - \left( b_i(\eta) - \langle a_i(\eta), y^*_{N_j} \rangle \right) \right| \leq \|a_i(\eta)\| \cdot \|y^*_{N_j} - y^*\|$$

that there exists $\bar{N} \in \mathbb{Z}^+$ such that $y^*_{N_j}$ is strictly feasible for $(P)$ when $N_j > \bar{N}$; that is, $b_i(\eta) - \langle a_i(\eta), y^*_{N_j} \rangle > 0$ when $N_j > \bar{N}$. Using the same argument as above, it can be shown that

$$\phi_{N_j}(y^*_{N_j}, z^{k+1}_{N_j}) \rightarrow \phi(y^*, z^{k+1}) \text{ as } j \rightarrow \infty$$

Since $\phi(y^*, z^{k+1}) < \phi(y^{*,*}, z^{k+1})$, it follows that for $j$ sufficiently large

$$\phi_{N_j}(y^*_{N_j}, z^{k+1}_{N_j}) < \phi_{N_j}(y_{N_j}, z^{k+1}_{N_j})$$

which contradicts the definition of $\beta^*_{N_j}$. Therefore, $\beta^*_N \rightarrow \beta^*$. Since

$$x^{k+1}_N = x^k - \beta^*_N \pi_N(x^k, z^{k+1}_N)$$
$$x^{k+1} = x^k - \beta^* \pi(x^k, z^{k+1})$$

we have $x^{k+1}_N \rightarrow x^{k+1}$ as $N \rightarrow \infty$. ■

### 6.6.2 Generalizations

We now generalize Algorithm $(P)$ to the case where $X$ satisfies less restrictive constraints that (6.1). Let $\pi : H \rightarrow X$ be the orthogonal projection of $H$ onto $X$. We now consider the case when

$$X = \text{span}\{\pi a_i(\eta), \ldots, \pi a_m(\eta) : \eta \in [0, 1]\} \quad (6.44)$$

Let $(P^*)$ refer to the problem $(P)$ with (6.1) replaced by (6.44). The problem dual to $(P^*)$ is

$$(D^*) \begin{cases} -\sum_{i=1}^m \int_0^1 b_i(\eta) \, dv_i(\eta) \rightarrow \max \\ c + \sum_{i=1}^m \int_0^1 a_i(\eta) \, dv_i(\eta) \in X^\perp \\ v_i(\eta) \in BV[0, 1], \text{ increasing} \end{cases}$$
From \((P^*)\) we can define the related problem

\[
(P_\pi) \begin{cases} 
\langle \pi c, x \rangle \to \min \\
\langle \pi a_i(\eta), x \rangle \leq b_i(\eta), \quad i = 1, \ldots, m, \quad \eta \in [0, 1] \\
x \in X
\end{cases}
\]

\((P^*)\) and \((P_\pi)\) are related in the sense that \(x\) is feasible for \((P^*)\) if and only if it is feasible for \((P_\pi)\), and \(x^*\) is optimal for \((P^*)\) if and only if it is optimal for \((P_\pi)\).

Moreover, \((P_\pi)\) is of the form \((P)\) in Section 6.1. In this section, we generalize Algorithm \((P)\) to the problem \((P^*)\), and will refer to this new algorithm as Algorithm \((P^*)\). We prove convergence of Algorithm \((P^*)\) by comparing the iterates produced by this algorithm to the iterates obtained when Algorithm \((P)\) is applied to \((P_\pi)\). (ie Algorithm \((P_\pi)\)). Convergence of Algorithm \((P^*)\) is deduced from the convergence properties of Algorithm \((P_\pi)\) as stated in Theorem 6.1, Corollary 6.2 and Theorem 6.3. Let \(x^0 \in X\) be strictly feasible for \((P^*)\) and define the operator \(Q(x^0) : X \to Y\) by

\[
Q(x^0) \cdot y = \sum_{i=1}^{m} \int_{0}^{1} \left\langle \frac{a_i(\eta)}{s_i(x^0, \eta)}, y \right\rangle \frac{a_i(\eta)}{s_i(x^0, \eta)} \, d\eta
\]

Then \(\langle \cdot, \cdot \rangle_{x^0} = \langle Q(x^0) \cdot, \cdot \rangle : X \times X \to \mathbb{R}\) defines an inner product on \(X\) and hence, a Riemannian metric on \(X\). The gradient of \(\phi(x^0, z^0)\) with respect to this Riemannian metric \(\nabla_{x^0} \phi(x^0, z^0)\) is given as follows:

**Lemma 6.8** Let \(Q(x^0) : X \to Y\) be given by (6.45). The gradient of \(\phi(x^0, z^0)\) with respect to the Riemannian metric defined by the inner product \(\langle \cdot, \cdot \rangle_{x^0}\) is the unique element \(\nabla_{x^0} \phi(x^0, z^0) \in X\) such that

\[
\nabla \phi(x^0, z^0) - Q(x^0) \cdot \nabla_{x^0} \phi(x^0, z^0) \in X^\perp
\]

**Proof:** Since \(\langle \cdot, \cdot \rangle_{x^0}\) defines an inner product on \(X\), it follows from the Riesz Representation Theorem for linear functionals on a Hilbert space that there exists a unique \(\nabla_{x^0} \phi \in X\) such that

\[
\langle \nabla \phi(x^0, z^0), y \rangle = \langle Q(x^0) \cdot \nabla_{x^0} \phi(x^0, y^0), y \rangle
\]

for every \(y \in X\). (6.46) follows immediately.

Let \(x^0\) be strictly feasible for \((P^*)\). Then \(x^0\) is feasible for \((P_\pi)\) also. Let \(\nabla_{x^0} \phi(x^0, z^0)\) be defined by (6.46) and associated with \((P_\pi)\), let \(\nabla_{x^0} \phi(x^0, z^0)\) be defined by (6.10).
Let $y \in X$. By (6.45)
\[
Q(x^0) \cdot y = \sum_{i=1}^{m} \int_0^1 \left< \frac{a_i(\eta)}{s_i(x^0, \eta)}, y \right> \frac{\pi a_i(\eta)}{s_i(x^0, \eta)} d\eta \\
+ \sum_{i=1}^{m} \int_0^1 \left< \frac{a_i(\eta)}{s_i(x^0, \eta)}, y \right> (I - \pi) a_i(\eta) \frac{d\eta}{s_i(x^0, \eta)}
\in X \oplus X^1
\]

It follows that
\[
(\pi Q(x^0)) \cdot y = \sum_{i=1}^{m} \int_0^1 \left< \frac{\pi a_i(\eta)}{s_i(x^0, \eta)}, y \right> \frac{\pi a_i(\eta)}{s_i(x^0, \eta)} d\eta
\]

Denote by $\hat{Q}(x^0)$ the operator $\pi Q(x^0): X \to X$. Then $\hat{Q}(x^0)$ is equal to the operator defined by (6.8) for $(P^*)$. From (6.46) we obtain
\[
\pi \nabla \phi(x^0, z^0) - \hat{Q}(x^0) \cdot \nabla x^0 \phi(x^0, z^0) = 0
\]

Since $\hat{Q}(x^0)$ is invertible, it follows that
\[
\nabla x^0 \phi(x^0, z^0) = \hat{Q}(x^0)^{-1} \cdot \pi \nabla \phi(x^0, z^0)
\]

from which we conclude the following:

**Proposition 6.10** Let $x^0$ be strictly feasible for $(P^*)$ and $z^0 \leq z^*$. Then $\nabla x^0 \phi(x^0, z^0) \in X$ given by (6.46) is
\[
\nabla x^0 \phi(x^0, z^0) = \hat{Q}(x^0)^{-1} \cdot \pi \nabla \phi(x^0, z^0)
\]

We now consider the problem of updating the lower bound $z^0 \leq z^*$ of the optimal cost functional value. Let $x^0$ be feasible for $(P^*)$. Suppose that $v_i(\eta)$ is given by (6.16) where $f_i(x^0, z^0, \eta)$ is defined by
\[
f_i(x^0, z^0, \eta) = \frac{\langle c, x^0 \rangle - z^1}{\rho} \left[ s_i(x^0, \eta)^{-1} - \left< \frac{a_i(\eta)}{s_i(x^0, \eta)^2}, \nabla x^0 \phi(x^0, z^0) \right> \right]
\]

(6.47)

where $s_i(x^0, \eta) = b_i(\eta) - \langle a_i(\eta), x^0 \rangle$. Noting that
\[
f_i(x^0, z^1, \eta) = \frac{\langle \pi c, x^0 \rangle - z^1}{\rho} \left[ s_i(x^0, \eta)^{-1} - \left< \frac{a_i(\eta)}{s_i(x^0, \eta)^2}, \hat{Q}(x^0)^{-1} \cdot \pi \nabla \phi(x^0, z^0) \right> \right]
\]

(6.48)
where \( s_i(x^0, \eta) = b_i(\eta) - \langle \pi a_i(\eta), x^0 \rangle \), it follows that

\[
c + \sum_{i=1}^{m} \int_0^1 a_i(\eta) \, dv_i(\eta)
\]
\[
= \left[ c + \sum_{i=1}^{m} \int_0^1 \pi a_i(\eta) \, dv_i(\eta) \right] + \left[ (I - \pi) c + \sum_{i=1}^{m} \int_0^1 (I - \pi) a_i(\eta) \, dv_i(\eta) \right]
\]
\[
= 0 + \left[ (I - \pi) c + \sum_{i=1}^{m} \int_0^1 (I - \pi) a_i(\eta) \, dv_i(\eta) \right]
\]
\[
\in X^\perp
\]

Therefore, \( v_i(\eta) \) is dual feasible so long as \( f_i(x^0, z^1, \eta) \geq 0 \). We shall refer to \((ZP)\) with \( f_i(x^0, z, \eta) \) given by (6.47) as \((ZP^*)\), and to \((ZP)\) with \( f_i(x^0, z, \eta) \) given by (6.48) as \((ZP_\pi)\). Let \( z^1 \) be optimal for \((ZP^*)\) associated with \( x^0 \), and \( z^1_\pi \) the optimal solution of \((ZP_\pi)\) associated with \( x^0 \). A consequence of the equality in (6.47)-(6.48) and Proposition 6.10 is the following:

**Proposition 6.11** Let \( x^0 \) be strictly feasible for \((P^*)\). If \( z^1 \) is the optimal solution of \((ZP^*)\) associated with \( x^0 \), and \( z^1_\pi \) the optimal solution of \((ZP_\pi)\) associated with \( x^0 \), then \( z^1 = z^1_\pi \).

It should be noted that \( \nabla_x \phi(x^0, z) \) defined by (6.46) depends on \( z \). Since

\[
\nabla \phi(x^0, z^1) = \nabla \phi(x^0, x^0) + \frac{\rho(z^1 - z^0) \, c}{((c, x^0) - z^1) \, ((c, x^0) - z^0)}
\]

if we put

\[
\nabla_x \phi(x^0, z^1) = \nabla_x \phi(x^0, z^0) + \frac{\rho(z^1 - z^0) \, \Delta}{((c, x^0) - z^1) \, ((c, x^0) - z^0)} \tag{6.49}
\]

for some \( \Delta \in X \), then (6.46) becomes

\[
\nabla \phi(x^0, z^0) - Q(x^0) \cdot \nabla_x \phi(x^0, z^0) + \frac{\rho(z^1 - z^0)}{((c, x^0) - z^1) \, ((c, x^0) - z^0)} \, (c - Q(x^0) \cdot \Delta) \in X^\perp
\]

To solve \((ZP^*)\), we need to solve

\[
\nabla \phi(x^0, z^0) - Q(x^0) \cdot \nabla_x \phi(x^0, z^0) \in X^\perp \tag{6.50}
\]

\[
c - Q(x^0) \cdot \Delta \in X^\perp \tag{6.51}
\]
Note that \( \nabla_{x_0} \phi(x_0, z^0), \Delta \in X \) are independent of \( z^1 \) and \( \nabla_{x_0} \phi(x_0, z^1) \) is given by (6.49). In this way, (6.46) need not be solved for each new value of \( z \).

To show convergence of Algorithm (\( P^* \)) we observe the following: given any strictly feasible \( x^0 \in X \) and lower bound \( z^0 \) of the optimal cost, the Algorithm (\( P^* \)) produces iterates which are equal to the iterates produced by Algorithm (\( P_\pi \)). It has already been shown in Proposition 6.11 that (\( ZP^* \)) and (\( ZP_\pi \)) will give equal values of the lower bound update \( z^1 \). Hence it only remains to be shown that the updates \( x^1 \) produced by Algorithm (\( P^* \)) and Algorithm (\( P \)) are equal. The update of \( x^0 \) produced by Algorithm (\( P^* \)) is

\[
x^1_{\pi} = x^0 - \beta^* \hat{Q}(x^0)^{-1} \cdot \pi \phi(x^0, z^1)
\]

\[
\beta^* = \max_{\beta \geq 0} \phi(x^0 - \beta \hat{Q}(x^0)^{-1} \cdot \pi \phi(x^0, z^1), z^1)
\]

It can be shown that Proposition 6.2 is true for (\( P^* \)), and the update of \( x^0 \) given by Algorithm (\( P^* \)) is

\[
x^1 = x^0 - \beta^* \nabla_{x_0} \phi(x^0, z^1)
\]

\[
\beta^* = \max_{\beta \geq 0} \phi(x^0 - \beta \nabla_{x_0} \phi(x^0, z^1), z^1)
\]

Using Proposition 6.10, it follows that \( \nabla_{x_0} \phi(x^0, z^1) = \hat{Q}(x^0)^{-1} \cdot \pi \nabla \phi(x^0, z^1) \). Therefore, \( \beta^* = \beta^\pi \) and it follows that \( x^1 = x_{\pi}^1 \). Under the Assumptions of Theorem 6.1 we conclude the following

**Theorem 6.4** Let Assumptions 6.1, 6.2, 6.3, 6.5 and 6.6 hold. If \( \{x_k\}_{k=1}^\infty \) is the sequence of iterates produced by Algorithm (\( P^* \)) and \( x^* \) is an optimal solution of (\( P^* \)) then \( x^k \to x^* \) as \( k \to \infty \).

**Corollary 6.5** Let \( \{x_k\}_{k=0}^\infty \) and \( x^* \) be as in Theorem 6.1. Then \( \langle c, x^k \rangle \to \langle c, x^* \rangle \) as \( k \to \infty \).

If we assume in addition that that Assumption 6.4 holds for (\( P_\pi \)), then Theorem 6.3 can be generalized in the following way:

**Theorem 6.6** Let \( \{x_k\}_{k=0}^\infty \) and \( x^* \) be as in Theorem 6.4. Then \( x^k \to x^* \) as \( k \to \infty \).
Chapter 7

Concluding remarks and future directions

In this thesis, we have explored issues relating to the optimal control of systems with constraints. We have focused on constrained LQG control problems with particular emphasis on generalizing well known results for the unconstrained LQG control to the linearly constrained and quadratically constrained case. Also, we have derived new algorithms for solving these problems by generalizing certain interior point methods (IPM's) to infinite dimensions. The new algorithms appear powerful and convenient for implementation. However, it remains to be seen how well these algorithms perform in real world applications, although there is some current progress in one application area.

Although our focus has been on LQG control with linear and quadratic constraints, we feel that these results open up exciting possibilities. By generalizing sequential quadratic programming to infinite dimensions, our results on constrained LQG control can be used as part of an algorithm for constructing optimal feedback controls for general nonlinear optimal control problems with constraints.

In this final chapter, we briefly summarize the results of our work, and indicate some directions for future research.
CHAPTER 7. CONCLUDING REMARKS AND FUTURE DIRECTIONS

Linearly constrained LQG control

In Chapter 2, we studied the LQG control problem with finitely many and infinitely many integral linear constraints. We derived closed form expressions for the optimal control, and in the partially observed case, proved that the Separation Theorem holds. In the case of finitely many constraints, the deterministic optimal control is obtained by solving a finite dimensional optimization problem. In the case of infinitely many constraints, it is obtained by solving an infinite programming problem. Moreover, we proved in both cases that the optimal control for the partially observed problem is obtained by replacing the state in the deterministic optimal control by the output of the Kalman filter.

An interesting and exciting area for future research is the application of these results in the context of sequential quadratic programming (SQP) for constrained nonlinear optimal control problems. Such a generalization would involve extending SQP to an infinite dimensional setting. SQP is one of the most efficient methods for solving finite dimensional constrained nonlinear optimization problems. The basic idea behind this scheme is solving a sequence of quadratic optimization problems where the cost is a quadratic approximation of the nonlinear cost, and constraints are linear approximations of the nonlinear constraints. Under certain conditions, quadratic convergence of this algorithm near the optimal solution can be proven. Clearly, the SQP algorithm strongly relies on being able to solve the linearly constrained quadratic optimization problems efficiently. In the context of nonlinear constrained optimal control, quadratic approximations result in linearly constrained LQG control problems of the form studied in Chapter 2. In making such a generalization of SQP, one interesting issue is determining the conditions under which quadratic convergence near the optimal solution can be guaranteed. Note once again that the optimal control for the linearly constrained LQG problem is a linear state feedback. It follows immediately that the (sub)optimal control obtained by using the SQP algorithm to solve a constrained nonlinear optimal control problems will be a linear state feedback. This has the advantage that such controllers are easy to implement.

LQG control with IQ constraints

In Chapter 3, we considered the convex LQG control problem with integral quadratic constraints. Using duality theory, we derived the optimal control, and showed that
it can be obtained by solving the finite dimensional dual optimization problem. This can be solved using finite dimensional optimization algorithms, so long as the gradient of the cost functional can be calculated. We showed that a key step in calculating the gradient is solving an unconstrained LQG control problem. In the partially observed case, we proved that the Separation Theorem does not hold but rather, a result that we called a Quasi-separation Theorem does instead.

Stochastic LQG control with IQ constraints and indefinite control weights

Using results obtained by Chen, Li and Zhou [58], we have shown in Chapter 4 that in the case of full observation LQG control with IQ constraints, the control weighting matrices in the cost and constraint functionals need not be positive definite if the diffusion term depends linearly on the control. In fact, we proved in this chapter that the condition derived in [58] for the existence and uniqueness of solutions of a certain Riccati equation is actually the necessary and sufficient condition for strict convexity of the associated indefinite LQG cost functional. From this, we derived sufficient conditions under which duality theory can be used to solve the indefinite constrained LQG problem. A closed form expression for the optimal control is derived. As in the case of Chapter 3, this can be obtained by solving a finite dimensional optimization problem.

Infinite quadratic programming - a path following method

In Chapter 5, we generalized a path following method known as the logarithmic barrier method so that it can be used to solve infinite quadratic programming problems. Moreover, we proved global convergence of this algorithm. As an application, we showed how this algorithm could be used to solve LQ control problems with infinitely many linear integral constraints. With all IPM's, the most complicated step is the so called 'Newton step'. In the case of linearly constrained LQ control, we showed that calculating the Newton step is equivalent to solving an unconstrained LQ problem together with an integral equation.

Much work remains to be done on this problem. Theoretically, the most interesting question pertains to deriving a complexity bound for this algorithm. Many IPM's
have a worse case polynomial complexity. When there are finitely many constraints, this algorithm in known to have a polynomial complexity bound [11]. Traditionally, polynomial complexity is expressed in terms of problem size. In finite dimensional problems, problem size is given in terms of the number of variables and the number of constraints in the problem. In this case though, both the variable space and the constraint space are infinite dimensional and hence, a generalized definition of problem size, as given in [42], is needed if complexity is to be studied.

**Infinite linear programming - a potential reduction method**

In Chapter 6, we generalized the potential reduction IPM of Ye [56] to infinite dimensions and showed how it could be used to solve problems such as continuous linear programming [4, 41] and linear optimal control with continuous linear state constraints. Furthermore, we proved the convergence of this algorithm under certain conditions. As with the path following method in Chapter 5, the so called Newton step is the most computationally demanding operation when performing this algorithm. When applied to CLP and linear optimal control with infinitely many state constraints, a pair of integral equations needs to be solved to carry out the Newton step.

This work is far from complete. The issue of complexity is the most interesting and important open problem. As already mentioned, problem size is usually defined in terms of the number of variables and the number of constraints. In this problem, the variable and constraint spaces are infinite dimensional and hence, a generalized definition of problem size [42] is required to address the issue of complexity.

In our study of this algorithm, we focused on the issue of convergence. In our analysis, many assumptions were made. It remains to be seen which of these assumptions can be relaxed, and which are necessary. Indeed, it is shown in [40] that the generalization of Karmarkar's algorithm to the semi-infinite case can converge to a non-optimal point.

Finally, there is scope for further work in the numerical implementation of this algorithm.
Chapter 8

Appendix

8.1 Definitions and theorems from analysis

Definition 8.1 (Weak convergence) Let \( \{x^k\}_{k=1}^\infty \) be a sequence in the Hilbert space \( (H, \langle \cdot, \cdot \rangle) \). Then \( x^k \) is said to converge to \( x^* \in H \) weakly (denoted \( x^k \rightharpoonup x^* \)) if \( \langle x^k, z \rangle \to \langle x^*, z \rangle \) as \( k \to \infty \) for every \( z \in H \).

Definition 8.2 (Weak closure) A subset \( M \) of a Hilbert space \( (H, \langle \cdot, \cdot \rangle) \) is said to be weakly closed if it contains all of its weak limits. That is, if \( \{x^k\}_{k=1}^\infty \in M \) and \( x^k \rightharpoonup \bar{x} \) as \( k \to \infty \), then \( \bar{x} \in M \).

Theorem 8.1 A convex subset of a Hilbert space \( (H, \langle \cdot, \cdot \rangle) \) is weakly closed if and only if it is strongly closed.

Theorem 8.2 A subset \( M \) of a Banach space \( X \) is weakly compact if and only if it is norm bounded and weakly closed.

8.2 On the baking of almond biscuits

Almond biscuits are enjoyed world wide in cafes, coffee shops and homes. However, many have complained about the price of these biscuits. In this section, we discuss briefly some issues associated with the synthesis of almond biscuits. In fact, we show that the problem of consistently producing a biscuit that meets the high quality...
standards of most reputable cafes is a difficult task, which partly accounts for the high cost of this product. We hope that this discussion will eventually lead to automated procedures for producing these biscuits and ultimately, a revolution in the biscuit making industry. The recipe below is taken from the unpublished preprint [33].

8.2.1 Production procedure

Ingredients:
3 egg whites
$\frac{1}{2}$ cup castor sugar
1 cup plain flour
12 oz. almond

Method:
Beat egg white and sugar together until stiff. Add plain flour, mixing in on low speed. Add almond and mix with a wooden spoon. Pour into lightly greased loaf tin.

Bake at moderate temperature until slightly golden. Test with finger until it feels firm. Remove and cut into very, very thin slices. (This may be difficult.) Lay on a tray and replace in oven on a low temperature - 100 degrees C for one or two hours or until crispy.

8.2.2 Concluding remarks

As already mentioned, the major difficulty in this process is slicing the baked biscuits into thin slices. This is a difficult problem for several reasons. First, it seems that no machine has yet been developed for automatically performing this task. Of course, there are many slicing machines available on the market, but whether these are capable of slicing brittle almond biscuits into thin slices is another matter. Until now, slicing has had to be done by hand and for this reason, the production of almond biscuits has been limited to small quantities at a time. (From personal experience, the cutting of large quantities of almond biscuits is likely to cause much pain). It is appropriate to mention here that knives with serrated edges seem to be the most effective tool for this cutting task. However, the problems in almond biscuit production are not limited to the slicing process. Indeed, a major variable in the cutting task is the quality of the baked almond biscuit. If the baked biscuit is too hard and
brittle, cutting thin biscuit slices is difficult, time consuming and energy sapping. Moreover, biscuit rupture is likely to occur while cutting. On the other hand, if the biscuit is insufficiently firm, the desirable 'delicate and thinly sliced' appearance of the final product will be compromised. It seems that developing a method for baking a firm yet moist biscuit is an important step towards a fully automated almond biscuit production procedure.
Bibliography


