CENTRO DE INVESTIGACIÓN Y DE ESTUDIOS
AVANZADOS DEL INSTITUTO POLITÉCNICO NACIONAL

UNIDAD ZACATENCO

DEPARTAMENTO DE CONTROL AUTOMÁTICO

Control Óptimo tipo LQ para una Clase de Sistemas Lineales con Entradas Constantes a Trozos

T E S I S

Que presenta
Félix Alfredo Miranda Villatoro

Para obtener el grado de
Maestro en Ciencias

En la especialidad de
Control Automático

Directores de Tesis
Dr. Vadim Azhmyakov
Dr. Fernando Castaños Luna

México, D.F. Noviembre 2012

Agradecimientos

Al CONACyT por el apoyo económico brindado a lo largo de la duración del programa de Maestría. A mis asesores el Dr. Vadim Azhmyakov y el Dr. Fernando Castaños Luna por sus consejos y el apoyo brindado a lo largo del desarrollo de este trabajo de Tesis. También quiero agradecer a todos los profesores del Departamento de Control Automático.

Por otra parte, quiero agradecer a Auroris por todo el apoyo brindado durante mucho años y a toda su familia por su comprensión y aceptación. Y finalmente, agradezco a mis padres y hermanos.

Índice general

Agradecimientos **III**

Resumen **VII**

Abstract **IX**

1. Introducción **11**

2. Marco Teórico **13**
2.1. Elementos de Análisis Convexo 13
2.2. El Problema Clásico de Control Óptimo 14
2.3. El Principio del Máximo de Pontryagin 16
2.4. Programación Dinámica y la Ecuación de Hamilton-Jacobi-Bellman 17
2.5. El Regulador Cuadrático Lineal 20
2.5.1. El principio del máximo aplicado al problema LQ 21
2.5.2. La ecuación de Riccati asociada al problema 23
2.5.3. La ecuación HJB aplicada al problema LQ 24

3. Planteamiento del Problema **27**
3.1. El Problema Original 27
3.2. El Problema Relajado 31

4. Resultado Principal **37**
4.1. Solución Analítica del Problema Relajado 37
4.2. Solución Numérica del Problema Relajado 44
4.2.1. El algoritmo del gradiente con proyección 44
4.3. Ejemplos 48
4.3.1. Satélite 49
4.4. Solución del Problema Original 65

5. Conclusiones y Trabajos Futuros **75**

Bibliografía **77**

Resumen

Este trabajo de Tesis es motivado a un problema de control óptimo tipo LQ asociado con una familia específica de sistemas lineales en la presencia de restricciones adicionales en la señal de control. Se consideran sistemas dinámicos gobernados por ecuaciones diferenciales lineales con un conocimiento *a priori* de la estrategia de conmutación de la señal de control y el funcional de costo cuadrático. La estructura constante a trozos de los controles admisibles en consideración, es motivada por una variedad de aplicaciones de ingeniería concretas y además, puede ser interpretado como el resultado de un proceso de cuantización asociado con la dinámica original. Se propone un algoritmo numérico implementable que hace posible calcular de manera efectiva una solución aproximada a el problema restringido tipo LQ. En este trabajo se discuten algunos aspectos teóricos del esquema computacional obtenido y también se muestran algunos ejemplos numéricos.

Abstract

This thesis is devoted to an LQ-type optimal control problem (OCP) associated with a specific family of linear systems in the presence of additional control constraints. In this work is consider control processes governed by linear differential equations with a priori known control switching strategies and quadratic costs functional. The piecewise constant structure of the admissible controls under consideration is motivated by variety of concrete engineering applications and moreover, can be interpreted as a result of a quantization procedure associated with the original dynamics. This work proposes an implementable numeric algorithm that makes it possible to calculate effectively an approximating solution to the constrained LQ-type OCPs. This contribution discusses some theoretic aspects of the obtained computational scheme and also contains an illustrative numerical example.

CAPÍTULO 1

Introducción

Las técnicas de diseño de controladores basadas en técnicas de optimización convencionales y avanzadas, son hoy en día una metodología madura y relativamente simple para la síntesis práctica de varios tipos de controladores para sistemas dinámicos, tanto convencionales como híbridos (ver por ejemplo: [1], [3], [8], [24]).

Recientemente, el problema de métodos numéricos efectivos para el caso donde se busca la minimización de un funcional de costo cuadrático asociado a un sistema lineal (problema LQ) ha atraído mucha atención, por lo que fueron desarrollados resultados teóricos y prácticos. (ver por ejemplo: [2], [9], [12]).

Nótese que manejar restricciones en el diseño de un sistema práctico es una cuestión importante a tomar en cuenta, en todas las aplicaciones del mundo real. Se aprecia fácilmente que los sistemas de control implementables tienen un conjunto de restricciones intrínseco. Por ejemplo, el conjunto de entradas admisibles siempre tiene un nivel mínimo y un máximo, más aún, los estados siempre está restringidos a permanecer en ciertos conjuntos. Por supuesto, el diseñador del sistema de control podría ignorar estas condiciones y esperar que no haya consecuencias serias, como resultado de tomar tal enfoque.

Por otra parte, es generalmente cierto que los niveles óptimos de rendimiento están asociados con la operación sobre o cerca de la frontera del conjunto de restricciones. Por lo tanto, un ingeniero de control no puede simplemente ignorar las restricciones impuestas sin incurrir en una degradación en el rendimiento.

El objetivo de este trabajo de Tesis es elaborar un algoritmo computacional consistente para un problema de control óptimo tipo LQ en la presencia de señales de entrada constantes a trozos con tiempos de conmutación fijos y niveles restringidos a ciertos conjuntos. La

estructura dada al conjunto de controles admisibles, es motivada por varias aplicaciones de control prácticas (ver [11], [25]) así como por procedimientos de cuantización aplicados a la dinámica original (ver [6], [17]).

En el presente trabajo se propone un método numérico basado en una combinación de un esquema de relajación clásico y un enfoque de proyección. Además, debe ser notado que el algoritmo propuesto puede ser efectivamente usado para llevar a cabo la síntesis de controladores asociados con sistemas lineales convencionales y algunos sistemas lineales conmutados.

Recordando que un sistema conmutado general constituye una clase de modelos donde están presente dos tipos de dinámica: continua y discreta (ver por ejemplo [16]). Con el fin de entender como este tipo de sistemas pueden ser operados de manera eficiente, ambos tipos de dinámicas deben de ser consideradas durante el proceso de diseño del control óptimo.

El resto del trabajo esta organizado de la siguiente manera: en el capítulo 2 se presenta el marco teórico necesario para la formulación del problema y la versión relajada, mostrando algunos resultados clásicos del análisis convexo. En el capítulo 3 se hace una descripción de los conjunto de restricciones, para posteriormente realizar el planteamiento formal del problema, a continuación se estudian las propiedades de convexidad del funcional de costo asociado y se plantea el problema relajado. En el capítulo 4 se presenta el algoritmo para encontrar la solución del problema relajado de forma numérica, además se plantea (cuando es posible) la solución analítica del problema relajado, posteriormente se plantea el problema existente al buscar la solución al problema original, a lo largo de este capítulo, se muestra un ejemplo de aplicación para este tipo de sistemas. Finalmente, en el capítulo 5 se presentan las conclusiones obtenidas como producto de este trabajo.

CAPÍTULO 2

Marco Teórico

2.1. Elementos de Análisis Convexo

A continuación se presentan algunas definiciones y propiedades básicas de los elementos del análisis convexo que serán de utilidad en lo que sigue, ver [10], [19], [22].

Definición 2.1 *Un conjunto $\mathcal{A}$ es convexo si para cada par de elementos de $\mathcal{A}$, a_1 y a_2, tenemos que:*

$$(1-\lambda)a_1 + \lambda a_2 \in \mathcal{A}, \qquad 0 \leq \lambda \leq 1.$$

El conjunto $\{(1-\lambda)x + \lambda y : 0 \leq \lambda \leq 1\}$ es llamado el segmento de línea cerrado que une los puntos x, y. Entonces, un conjunto es convexo si y sólo si para cualquiera dos puntos del conjunto, el segmento de línea que une dichos puntos está completamente contenido en el conjunto.

Proposición 2.1 *La intersección de una colección arbitraria de conjuntos convexos es un conjunto convexo.*

Prueba. Sea $\mathcal{B} = \cap_{i\in I} B_i$ para algún conjunto de índices arbitrario I. Sean b_1, b_2 elementos de $\mathcal{B}$, entonces, b_1, $b_2 \in B_i \quad \forall i \in I$. Puesto que cada conjunto B_i es convexo, se tiene:

$$(1-\lambda)b_1 + \lambda b_2 \in B_i, \quad 0 \leq \lambda \leq 1, \quad \forall i \in I$$

Por lo tanto, $(1-\lambda)b_1 + \lambda b_2 \in \mathcal{B}$.

Definición 2.2 *Sea $\mathcal{A} \subseteq \mathbb{R}^n$, la intersección de todos los conjuntos convexos que contienen al conjunto $\mathcal{A}$ es llamada el 'convex hull' de $\mathcal{A}$ y se denota por conv$(\mathcal{A})$.*

Nota. El $conv(\mathcal{A})$ es un conjunto convexo por la proposición 2.1.

Definición 2.3 *Una combinación de elementos $a_1, a_2, \ldots, a_n \in \mathcal{A}$ de la forma:*

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n, \quad \sum_{i=1}^{n} \lambda_i = 1, \quad \lambda_i \geq 0, \quad i = 1, \ldots, n.$$

Se denomina combinación convexa de elementos de $\mathcal{A}$.

Lema 2.1 *Para cualquier conjunto $\mathcal{A} \subseteq \mathbb{R}^n$, el conjunto $conv(\mathcal{A})$ es igual al conjunto formado por todas las combinaciones convexas de los elementos de $\mathcal{A}$.*

Definición 2.4 *Una función $f : C \subseteq \mathbb{R}^n \to \mathbb{R}$, (donde C es el dominio efectivo de f), se denomina convexa si para todo $x, y \in C$ tenemos:*

$$f\left((1-\lambda)x + \lambda y\right) \leq (1-\lambda)f(x) + \lambda f(y), \qquad 0 \leq \lambda \leq 1 \tag{2.1}$$

Nota. El dominio efectivo $C \subseteq \mathbb{R}^n$ de una función f está dado por:

$$C = \{x \in \mathbb{R}^n : f(x) < \infty\}$$

A la función g que cumple con la desigualdad opuesta ($\geq$) de la ecuación (2.1) se le denomina cóncava.

Definición 2.5 *Una función $h : C \subseteq \mathbb{R}^n \to \mathbb{R}$ se le denomina afín si es cóncava y convexa a la vez. Es decir:*

$$h\left((1-\lambda)x + \lambda y\right) = (1-\lambda)h(x) + \lambda h(y), \qquad 0 \leq \lambda \leq 1 \tag{2.2}$$

2.2. El Problema Clásico de Control Óptimo

Se considera un sistema dinámico descrito por la siguiente ecuación diferencial ordinaria:

$$\begin{cases} \dot{x}(t) = f(t, x(t), u(t)), & t \in [t_0, t_f] \\ x(t_0) = x_0 \end{cases} \tag{2.3}$$

Donde $f : [t_0, t_f] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}^n$. El mapeo $u(\cdot) : [t_0, t_f] \to \mathbb{R}^m$ es llamado el *control*, $x_0 \in \mathbb{R}^n$ es el *estado inicial* y la solución de (2.3) es una función absolutamente continua $x(\cdot) : [t_0, t_f] \to \mathbb{R}^n$ llamada la *trayectoria de estado* correspondiente al control $u(\cdot)$.

De aquí en adelante se asume que para cualquier x_0 y cualquier control $u(\cdot)$ existe una única solución de la ecuación (2.3) dada por $x(\cdot) \equiv x(\cdot, u(\cdot))$.

Además, también está dado un funcional de costo que mide el rendimiento del control aplicado al sistema:

$$J(u(\cdot)) = \int_{t_0}^{t_f} g(t, x(t), u(t))dt + h(x(t_f)). \tag{2.4}$$

Donde $g : [t_0, t_f] \times \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ y $h : \mathbb{R}^n \to \mathbb{R}$. El primer término del lado derecho de la ecuación (2.4) se le conoce como *'running cost'* y al segundo término se le conoce como *'terminal cost'*.

En los problemas de control óptimo es común encontrar restricciones sobre los estados y/o sobre el control, estas restricciones pueden caracterizarse como:

$$x(t) \in S, \quad u(t) \in U \quad \forall t \in [t_0, t_f].$$

donde $S \subseteq \mathbb{R}^n$ y $U \subseteq \mathbb{R}^m$. Un control $u(\cdot)$ es llamado control admisible y el par $(x(\cdot), u(\cdot))$ par admisible si:

- $u(\cdot) \in U$.
- $x(\cdot)$ es la única solución absolutamente continua de (2.3).
- La restricción de estado es satisfecha.
- $t \mapsto f(t, x(t), u(t)) \in \mathbb{L}_1[t_0, t_f]$.

El conjunto de todos los controles admisibles es denotado por $U_{adm}[t_0, t_f]$.

El problema de control óptimo se enuncia a continuación:

Problema A Minimizar (2.4) sobre $U_{adm}[t_0, t_f]$.

El problema A se dice que es *finito* si (2.4) tiene una cota inferior finita, y se dice que es (*únicamente*) *soluble* si existe un (único) control $u^*(\cdot) \in U_{adm}[t_0, t_f]$ que satisface:

$$J(u^*(\cdot)) = \inf_{u(\cdot) \in U_{adm}[t_0, t_f]} J(u(\cdot)). \tag{2.5}$$

Cualquier control $u^*(\cdot) \in U_{adm}[t_0, t_f]$ que satisface (2.5) es llamado *control óptimo*, la correspondiente trayectoria de estado $x^*(\cdot) \equiv x(\cdot, u^*(\cdot))$ es llamada *trayectoria óptima* y $(x^*(\cdot), u^*(\cdot))$ es el *par óptimo*.

Definición 2.6 *Un control $u^*(\cdot) \in U_{adm}[t_0, t_f]$ es llamado un mínimo local de (2.4) si:*

$$J(u^*(\cdot)) \leq J(u(\cdot)) \qquad \forall u(\cdot) \in W_\varepsilon(u^*(\cdot)).$$

Donde:

$$W_\varepsilon = \{u(\cdot) \in U_{adm}[t_0, t_f] \mid \|u(\cdot) - u^*(\cdot)\| < \varepsilon\}.$$

En el caso donde $W_\varepsilon = U_{adm}[t_0, t_f]$ entonces el mínimo es global.

En las siguientes dos subsecciones se enuncian dos teoremas fundamentales en la teoría de control óptimo: El Principio del Máximo de Pontryagin y el Principio de Optimalidad de Bellman, los cuales son de gran utilidad en el desarrollo de las secciones subsecuentes, en especial para la solución del problema del Regulador Cuadrático Lineal y la extensión propuesta a una clase de sistemas lineales híbridos con funcional de costo cuadrático.

2.3. El Principio del Máximo de Pontryagin

Uno de los principales enfoques en optimización es obtener un conjunto de condiciones necesarias que debe satisfacer una solución óptima. Estas condiciones necesarias también se vuelven suficientes bajo ciertas condiciones de convexidad en el funcional de costo. Los problemas de control óptimo pueden ser tratados como problemas de optimización en espacios de dimensión infinita que son bastante difíciles de resolver.

El principio del máximo es una base importante en la teoría de control óptimo, enuncia que cualquier control óptimo junto con la trayectoria de estado óptima debe de resolver el llamado sistema Hamiltoniano, el cual es un problema de 2 puntos con valores en la frontera, además de cumplir con una condición de máximo de una función llamada Hamiltoniano. La importancia matemática del principio del máximo recae en el hecho de que es más fácil maximizar el Hamiltoniano del sistema que el problema original de dimensión infinita.

Retomamos el problema clásico de control óptimo descrito en la sección anterior. Es decir, se considera el sistema dado por (2.3) y el funcional (2.4) donde además se asume que los mapeos $f(\cdot,x,\cdot), g(\cdot,x,\cdot), h(x) \in C^1[t_0,t_f]$ lo que implica que existe una solución única $x(\cdot)$ de (2.3).

El siguiente teorema enuncia el bien conocido *Principio del Máximo de Pontryagin* el cual nos proporciona un conjunto de condiciones necesarias para los pares óptimos.

Teorema 2.1 ***(Principio del Máximo)****. Sea $(x^*(\cdot),u^*(\cdot))$ un par óptimo del problema clásico de control óptimo. Entonces, existe una función absolutamente continua $p(\cdot):[t_0,t_f]\to\mathbb{R}^n$ que satisface las siguientes condiciones:*

$$\begin{cases} \dot{p}(t) = -f_x(t,x^*(t),u^*(t))^T p(t) + g_x(t,x^*(t),u^*(t)), \quad t\in[t_0,t_f]. \\ p(t_f) = -h_x(x^*(t_f)). \end{cases} \tag{2.6}$$

Donde el subíndice x indica derivada parcial, es decir:

$$f_x = \frac{\partial f}{\partial x}, \quad g_x = \frac{\partial g}{\partial x} \quad h_x = \frac{\partial h}{\partial x}$$

y

$$H(t,x^*(t),u^*(t),p(t)) = \max_{u\in U} H(t,x^*(t),u,p(t)), \quad t\in[t_0,t_f]. \tag{2.7}$$

donde U representa el conjunto de controles admisibles y además:

$$H(t,x,u,p) = \langle p, f(t,x,u)\rangle - g(t,x,u), \quad (t,x,u,p) \in [t_0,t_f]\times\mathbb{R}^n\times U\times\mathbb{R}^n. \tag{2.8}$$

con $\langle\cdot,\cdot\rangle$ representa el producto interno estándar en $\mathbb{R}^n$.

La prueba del teorema se puede consultar en [28].

A la función $p(\cdot)$ se le denomina *variable/función adjunta* y a (2.6) la ecuación adjunta (correspondientes al par óptimo $(x^*(\cdot),u^*(\cdot))$).

La función H definida por (2.8) es llamada el Hamiltoniano del sistema. La ecuación de estado (2.3), la correspondiente ecuación adjunta (2.6) y la condición de máximo (2.7), puede ser escrita como:

$$\begin{cases} \dot{x}(t) = H_p(t,x(t),u(t),p(t)) \\ \dot{p}(t) = -H_x(t,x(t),u(t),p(t)) \\ x(0) = x_0, \qquad p(T) = -h_x(x(T)) \\ H(t,x^*(t),u^*(t),p(t)) = \max_{u\in U} H(t,x^*(t),u,p(t), \quad t\in[t_0,t_f]. \end{cases} \tag{2.9}$$

El sistema (2.9) es conocido como sistema Hamiltoniano. Este sistema caracteriza parcialmente la optimalidad del problema. En casos donde ciertas condiciones de convexidad están presentes, el sistema Hamiltoniano completamente caracteriza un control óptimo.

2.4. Programación Dinámica y la Ecuación de Hamilton-Jacobi-Bellman

Al igual que en la sección anterior se considera un sistema dinámico como el descrito por (2.3):

$$\begin{cases} \dot{x}(t) = f(t,x(t),u(t)), \qquad t\in[t_0,t_f] \\ x(t_0) = x_0 \end{cases} \tag{2.10}$$

y se busca minimizar el funcional de costo dado por:

$$J(u(\cdot)) = \int_{t_0}^{t_f} g(t,x(t),u(t))dt + h(x(t_f)) \tag{2.11}$$

Cabe mencionar que el tiempo inicial t_0 y el estado inicial $x(t_0)=x_0$ permanecen fijos en la formulación del problema. La idea básica del método de programación dinámica es encontrar las relaciones existentes en toda un familia de problemas de control óptimo tomando como parámetros que describen la familia el tiempo inicial y el estado inicial.

Para precisar lo mencionado en el párrafo anterior se considera el par $(s, y) \in [t_0, t_f) \times \mathbb{R}^n$ y el siguiente sistema de control:

$$\begin{cases} \dot{x}(t) = f(t, x(t), u(t)), & t \in [s, t_f] \\ x(s) = y \end{cases} \tag{2.12}$$

junto con el funcional de costo:

$$J(s, y; u(\cdot)) = \int_s^{t_f} g(t, x(t), u(t))dt + h(x(t_f)) \tag{2.13}$$

donde:

$$u(\cdot) \in \Gamma[s, t_f] = \{u(\cdot) : [s, t_f] \to U \subset \mathbb{R}^m \mid u(\cdot) \text{ es medible}\}$$

Problema B. Minimizar (2.13) sujeto a (2.12) sobre $\Gamma[s, t_f]$.

El problema anterior en realidad es una familia de problemas de control óptimo parametrizadas por $(s, y) \in [t_0, t_f) \times \mathbb{R}^n$ en donde el problema original se encuentra embebido[1].

Además, se asume que las funciones $f : [t_0, t_f] \times \mathbb{R}^n \times U \to \mathbb{R}^n$, $g : [t_0, t_f] \times \mathbb{R}^n \times U \to \mathbb{R}^n$ y $h : \mathbb{R}^n \to \mathbb{R}$ son uniformemente continuas y satisfacen la condición de Lipschitz en la variable x y están acotadas uniformemente en t y u.

Las condiciones anteriores garantizan que para cualquier $(s, y) \in [t_0, t_f) \times \mathbb{R}^n$ y $u \in \Gamma[s, t_f]$ la ecuación (2.12) tiene una única solución $x(\cdot) \equiv x(\cdot, s, y; u(\cdot))$ y (2.13) está bien definido.

Se define la siguiente función:

$$\begin{cases} V(s, y) = \text{ínf}_{u(\cdot) \in \Gamma[s, t_f]} J(s, y, ; u(\cdot)), & \forall (s, y) \in [t_0, t_f) \times \mathbb{R}^n. \\ V(t_f, y) = h(y) \qquad \forall y \in \mathbb{R}^n. \end{cases} \tag{2.14}$$

la cual es conocida como *'value function'* del problema original.

A continuación se presenta un teorema importante conocido como el *principio de optimalidad de Bellman.*

Teorema 2.2 *Bajo las consideraciones mencionadas arriba, para cualquier* $(s, y) \in [t_0, t_f) \times \mathbb{R}^n$,

$$\begin{aligned} V(s, y) \quad = \quad & \inf_{u(\cdot) \in \Gamma[s, t_f]} \left\{ \int_s^{\hat{s}} g(t, x(t; s, y, u(\cdot)), u(t))dt + V(\hat{s}, x(\hat{s}; s, y, u(\cdot))) \right\} \\ & \forall 0 \le t_0 \le s \le \hat{s} \le t_f. \end{aligned} \tag{2.15}$$

[1] El problema original se obtiene al hacer $s = 0$ e $y = x_0$.

La prueba del teorema anterior se puede consultar en [28].

Supongamos que el par $(x^*(\cdot), u^*(\cdot))$ es un par óptimo del problema B y $\hat{s} \in (s, t_f)$. Entonces:

$$\begin{aligned} V(s,y) = J(s,y;u^*(\cdot)) &= \int_s^{\hat{s}} g(t, x^*(t), u^*(t))dt + J(\hat{s}, x^*(\hat{s}), u^*(\cdot)) \\ &\geq \int_s^{\hat{s}} g(t, x^*(t), u^*(t))dt + V(\hat{s}, x^*(\hat{s})) \geq V(s,y). \end{aligned}$$

Donde la ultima desigualdad se obtiene a partir de (2.15), por lo que tenemos la siguiente igualdad:

$$V(\hat{s}, x^*(\hat{s})) = J(\hat{s}, x^*(\hat{s}), u^*(\cdot)) = \int_{\hat{s}}^{t_f} g(t, x^*(t), u^*(t))dt + h(x^*(t_f)).$$

Esta ultima relación muestra la esencia del principio de optimalidad de Bellman, es decir:

$$\begin{aligned} &u^*(\cdot) \text{ es óptimo en } [s, t_f](\text{ con condición inicial } (s,y)) \\ \Rightarrow\ &u^*|_{[\hat{s},t_f]}(\cdot) \text{ es óptimo en } [\hat{s}, t_f](\text{ con condición inicial } (\hat{s}, x^*(\hat{s}))). \end{aligned}$$

es decir:

$$\text{Optimalidad global} \Rightarrow \text{Optimalidad local.}$$

La ecuación (2.15)caracteriza la solución al problema B vía la 'value function', aunque esta ecuación es bastante complicada de resolver. Es posible obtener una ecuación más sencilla que caracterice el control óptimo a través del siguiente teorema.

Teorema 2.3 *Supongamos que $V \in C^1([t_0, t_f] \times \mathbb{R}^n)$, entonces V es una solución al siguiente problema de valor final de la siguiente ecuación diferencial parcial de primer orden:*

$$\begin{cases} -V_t + \sup_{u \in U} H(t, x, u, -V_x) = 0, \quad (t,x) \in [t_0, t_f] \times \mathbb{R}^n, \\ V|_{t=t_f} = h(x), \quad x \in \mathbb{R}^n. \end{cases} \tag{2.16}$$

donde:

$$H(t,x,u,p) \triangleq \langle p, f(t,x,u) \rangle - g(t,x,u) \quad \forall (t,x,u,p) \in [t_0, t_f] \times \mathbb{R}^n \times U \times \mathbb{R}^n$$

Una prueba completa del teorema puede ser consultada en [28].

La ecuación (2.16) es llamada la ecuación de Hamilton-Jacobi-Bellman (HJB) asociada al problema A.Si a partir de la ecuación HJB se encuentra la *'value function'* V entonces es

posible construir un par óptimo para cada problema B en especial para el problema original A haciendo uso de la condición de supremo en el Hamiltoniano de la ecuación (2.16), es decir:

$$H(t, x^*(t), u^*(t), -V_x(t, x^*(t))) = \sup_{u \in U} H(t, x^*(t), u, -V_x(t, x^*(t))) \tag{2.17}$$

El procedimiento descrito es conocido como *técnica de verificación*. La técnica de verificación involucra los siguientes pasos:

- Resolver la ecuación HJB (2.16) para encontrar la *value function* $V(t, x)$.
- Encontrar $u^*(t)$ a través de (2.17).
- Resolver (2.12) con $(s = t_0, y = x_0)$ para obtener el par óptimo $(x^*(\cdot), u^*(\cdot))$.

Hasta el momento se ha asumido que la *value function* $V \in C^1$ y además que las ecuación HJB tiene una solución única los cual no es cierto en general, por lo que se vuelve necesario redefinir el concepto de solución para la ecuación HJB tomando en cuenta las soluciones de viscosidad las cuales se encuentran fuera del alcance de este trabajo.

2.5. El Regulador Cuadrático Lineal

Se considera el siguiente sistema lineal:

$$\dot{x}(t) = A(t)x(t) + B(t)u(t) + b(t), \qquad x(t_0) = x_0 \tag{2.18}$$

Donde:

$$A(\cdot) \in \mathbb{L}_\infty[t_0, t_f; \mathbb{R}^{n\times n}], \qquad B(\cdot) \in \mathbb{L}_\infty[t_0, t_f; \mathbb{R}^{n\times m}], \qquad b(\cdot) \in \mathbb{L}_2[t_0, t_f; \mathbb{R}^n]$$

En donde los espacios $\mathbb{L}_\infty[t_0, t_f; \mathbb{R}^{n\times n}]$ y $\mathbb{L}_\infty[t_0, t_f; \mathbb{R}^{n\times m}]$ corresponden a los clásicos espacios de Lebesgue compuesto por todas las funciones matriciales esencialmente acotadas definidas sobre el intervalo $[t_0, t_f]$. Es decir:

$$\mathbb{L}_\infty[t_0, t_f; \mathbb{R}^{n\times n}] = \left\{ F : [t_0, t_f] \to \mathbb{R}^{n\times n} \middle| \, \|F\|_\infty = ess \sup_{t\in[t_0,t_f]} |F(t)| < \infty \right\}$$

De igual manera el espacio $\mathbb{L}_2[t_0, t_f; \mathbb{R}^n]$ define el clásico espacio de Lebesgue compuesto por todas la funciones vectoriales cuadrático integrables en el intervalo de tiempo $[t_0, t_f]$.

$$\mathbb{L}_2[t_0, t_f; \mathbb{R}^n] = \left\{ u(\cdot) : [t_0, t_f] \to \mathbb{R}^n | \int_{t=t_0}^{t_f} u(t)^T u(t) dt < \infty \right\}$$

Siguiendo con la descripción del regulador cuadrático lineal, tenemos que el funcional de costo toma la forma:

$$\begin{aligned} J(u(\cdot)) \;\; = \;\; & \frac{1}{2} \int_{t_0}^{t_f} \langle Q(t)x(t), x(t)\rangle + 2\langle S(t)x(t), u(t)\rangle + \\ & \langle R(t)u(t), u(t)\rangle] dt + \frac{1}{2} \langle Gx(t_f), x(t_f)\rangle \end{aligned} \tag{2.19}$$

Donde:

$$G \in \mathbb{R}^{n\times n}, \quad Q(\cdot) \in \mathbb{L}_\infty[t_0, t_f; \mathbb{R}^{n\times n}]$$
$$S(\cdot) \in \mathbb{L}_\infty[t_0, t_f; \mathbb{R}^{n\times m}], \quad R(\cdot) \in \mathbb{L}_\infty[t_0, t_f; \mathbb{R}^{m\times m}]$$

Además para casi todo $t \in [t_0, t_f]$:

$$G \geq 0, \quad Q(t) \geq 0, \quad R(t) \geq \delta I \quad \text{con: } \delta > 0$$

Se asume que el control admisible es cuadrático integrable en $[t_0, t_f]$, y no existen restricciones terminales, es decir:

$$\mathcal{U}_{adm}[t_0, t_f] = \{u(\cdot) | u(\cdot) \in \mathbb{L}_2[t_0, t_f; \mathbb{R}^m]\}$$

El problema del regulador lineal cuadrático (LQR por sus siglas en inglés), consiste en encontrar la señal de control $u^{opt}(t) \in \mathcal{U}_{adm}[t_0, t_f]$ que hace el funcional de costo dado por (2.19) tan pequeño como sea posible, es decir:

$$J(u^{opt}(\cdot)) = \inf_{u\in\mathcal{U}_{adm}[t_0,t_f]} \{J(u(\cdot))\} \tag{2.20}$$

2.5.1. El principio del máximo aplicado al problema LQ

Aplicando el teorema 2.1 al problema del regulador cuadrático, tenemos el siguiente corolario:

Corolario 2.1 *Si el par $(x^{opt}(\cdot), u^{opt}(\cdot))$ es óptimo, entonces:*

1. *Existe una solución $p(\cdot)$ a la ecuación:*

$$\begin{cases} \dot{p}(t) = -A^T(t)p(t) + Q(t)x^{opt}(t) + S^T(t)u^{opt}(t) \\ p(t_f) = -Gx^{opt}(t_f) \end{cases} \tag{2.21}$$

2. *Además el control $u^{opt}(t)$ cumple con:*

$$R(t)u^{opt}(t) - B^T(t)p(t) + S(t)x^{opt}(t) = 0, \qquad R(t) \geq 0 \tag{2.22}$$

Prueba: Puesto que el problema es regular (ver [21], [28]), es aplicable el principio del máximo, de tal manera que el hamiltoniano del sistema está dado por:

$$\begin{aligned} H(t, x, u, p) &= \langle p, A(t)x + B(t)u + b(t)\rangle - \frac{1}{2}(\langle Q(t)x, x\rangle \\ &\quad +2\langle S(t)x, u\rangle + \langle R(t)u, u\rangle) \end{aligned}$$

Entonces:

$$\begin{aligned}
\dot{p}(t) &= -\frac{\partial H(t,x,u,p)}{\partial x} \\
\dot{p}(t) &= \frac{\partial}{\partial x}\left(-\langle p(t), A(t)x(t)\rangle + \frac{1}{2}\langle Q(t)x(t), x(t)\rangle + \langle S(t)x(t), u(t)\rangle\right) \\
\dot{p}(t) &= -A^T(t)p(t) + Q(t)x^{opt}(t) + S^T(t)u^{opt}(t) \\
p(t_f) &= -\frac{\partial}{\partial x}\left(\frac{1}{2}\langle Gx(t_f), x(t_f)\rangle\right) = -Gx^{opt}(t_f)
\end{aligned}$$

La cual es la misma condición que la ecuación (2.21). Aplicando el principio del máximo, es decir:

$$u^{opt}(t) = \arg\max_{u\in\mathbb{R}^m} H(x^{opt}(t), u, p(t))$$

Y puesto que $H(\cdot,\cdot,u,\cdot)$ es una función cóncava, tenemos:

$$\begin{aligned}
0 &= \frac{\partial H(t,x,u,p)}{\partial u} \\
0 &= \frac{\partial}{\partial u}\langle p, B(t)u\rangle - \frac{1}{2}(2\langle S(t)x, u\rangle + \langle R(t)u, u\rangle) \\
0 &= B^T(t)p(t) - S(t)x^{opt}(t) - R(t)u^{opt}(t)
\end{aligned}$$

con $R(t) \geq 0$

■

A partir de (2.22) puede verse fácilmente que el control $u^{opt}(t)$ es único, si $R(t) \geq \delta I$, $\delta > 0$. lo cual nos lleva al siguiente lema.

Lema 2.2 *Si el control $u^{opt}(\cdot)$ está dado como en (2.22), y además:*

$$Q(t) - S(t)R^{-1}S^T(t) \geq 0$$

entonces el control óptimo es único.

Prueba: El Hessiano de $H(t,x,u,p)$ está dado por:

$$\begin{bmatrix} \frac{\partial^2}{\partial x^2}H(t,x,u,p) & \frac{\partial^2}{\partial x \partial u}H(t,x,u,p) \\ \frac{\partial^2}{\partial u \partial x}H(t,x,u,p) & \frac{\partial^2}{\partial u^2}H(t,x,u,p) \end{bmatrix} = -\begin{bmatrix} Q(t) & S(t) \\ S^T(t) & R(t) \end{bmatrix}$$

Verificamos que $\begin{bmatrix} Q(t) & S(t) \\ S^T(t) & R(t) \end{bmatrix} \geq 0$, esto pasa si y solo si:

$$Q(t) \geq 0 \qquad y \qquad Q(t) - S(t)R^{-1}S^T(t) \geq 0 \tag{2.23}$$

Lo anterior implica que el Hessiano es una función cóncava en $u(\cdot)$ lo cual asegura que $u^{opt}(\cdot)$ es como en (2.22). La unicidad se sigue de la condición sobre $R(t)$.

■

2.5.2. La ecuación de Riccati asociada al problema

Una forma de encontrar el control óptimo $u^{opt}(\cdot)$ es encontrando la solución a la ecuación diferencial de Riccati asociada como se enuncia en el siguiente teorema.

Teorema 2.4 *Bajo la consideracion de que $P(t) = P^T(t)$ es solución de la ecuación diferencial de Riccati dada por:*

$$\begin{cases} \dot{P}(t) + A^T(t)P(t) + P(t)A(t) + Q(t) - \\ \left[B^T(t)P(t) + S(t)\right]^T R^{-1}(t)\left[B^T(t)P(t) + S(t)\right] = 0 \\ P(t_f) = G \end{cases} \tag{2.24}$$

Entonces, el control óptimo $u^{opt}(\cdot) \in \mathcal{U}_{adm}[t_0, t_f]$ que satisface (2.20) está dado por:

$$u^{opt}(t) = -R^{-1}(t)\left(\left[B^T(t)P(t) + S(t)\right]x^{opt}(t) + B^T(t)\varphi(t)\right) \tag{2.25}$$

Prueba: Se propone que la solución de la ecuación (2.21) tiene la siguiente forma:

$$p(t) = -P(t)x^{opt}(t) - \varphi(t) \tag{2.26}$$

Sustituyendo (2.26) en (2.22), obtenemos:

$$u^{opt}(t) = -R^{-1}(t)\left[\left(B^T(t)P(t) + S(t)\right)x^{opt}(t) + B^T(t)\varphi(t)\right] \tag{2.27}$$

Si sustituimos (2.26) y (2.27) en el lado derecho de (2.21), obtenemos:

$$\begin{aligned} \dot{p}(t) &= -A^T(t)\left[-P(t)x^{opt}(t) - \varphi(t)\right] + Q(t)x^{opt}(t) + \\ &\quad S^T(t)\left[-R^{-1}(t)\left(\left(B^T(t)P(t) + S(t)\right)x^{opt}(t) + B^T(t)\varphi(t)\right)\right] \\ \dot{p}(t) &= \left[A^T(t)P(t) + Q(t) - S^T(t)R^{-1}(t)\left(B^T(t)P(t) + S(t)\right)\right]x^{opt}(t) + \\ &\quad \left[A^T(t) - S^T(t)R^{-1}(t)B^T(t)\right]\varphi(t) \end{aligned} \tag{2.28}$$

Ahora bien, derivando la ecuación (2.26) resulta:

$$\begin{aligned} \dot{p}(t) &= -\dot{P}(t)x^{opt}(t) - P(t)\dot{x}^{opt}(t) - \dot{\varphi}(t) \\ &= -\dot{P}(t)x^{opt}(t) - P(t)\left[A(t)x^{opt}(t) + B(t)u^{opt}(t) + b(t)\right] - \dot{\varphi}(t) \\ &= \left[-\dot{P}(t) - P(t)A(t)\right]x^{opt}(t) - \dot{\varphi}(t) - P(t)b(t) + \\ &\quad P(t)B(t)R^{-1}(t)\left[\left(B^T(t)P(t) + S(t)\right)x^{opt}(t) + B^T(t)\varphi(t)\right] \\ \dot{p}(t) &= \left[-\dot{P}(t) - P(t)A(t) + P(t)B(t)R^{-1}(t)\left(B^T(t)P(t) + S(t)\right)\right]x^{opt}(t) - \\ &\quad \dot{\varphi}(t) + P(t)B(t)R^{-1}(t)B^T(t)\varphi(t) - P(t)b(t) \end{aligned} \tag{2.29}$$

Igualando (2.28) y (2.29):

$$\begin{aligned} 0 &= [\dot{P}(t) + P(t)A(t) + A^T(t)P(t) + Q(t) - \\ &\quad \left(B^T(t)P(t) + S(t)\right)^T R^{-1}(t)\left(B^T(t)P(t) + S(t)\right)]x^{opt}(t) + \dot{\varphi}(t) + \\ &\quad \left[\left(A^T(t) - \left(S^T(t) + P(t)B(t)\right)R^{-1}(t)B^T(t)\right)\varphi(t)\right] + P(t)b(t) \end{aligned}$$

Por lo tanto, si seleccionamos $P(t)$ y $\varphi(t)$ tales que se satisfacen las siguientes ecuaciones:

$$\begin{cases} \dot{P}(t) + A^T(t)P(t) + P(t)A(t) + Q(t) - \\ \left[B^T(t)P(t) + S(t)\right]^T R^{-1}(t)\left[B^T(t)P(t) + S(t)\right] = 0 \\ P(t_f) = G \end{cases}$$

y,

$$\begin{cases} \dot{\varphi}(t) + (A^T(t) - S^T(t)R^{-1}(t)B^T(t) - P(t)B(t)R^{-1}(t)B^T(t))\varphi(t) \\ +P(t)b(t) = 0 \\ \varphi(t_f) = 0 \end{cases}$$

Entonces se cumple la ecuación (2.26) y la ley de control dada por (2.25) es válida.

■

A partir del Teorema 2.4 obtenemos el siguiente Corolario:

Corolario 2.2 *Si en la ecuación (2.18) el termino $b(t) = 0$. Entonces, el control óptimo $u^{opt}(t)$ debe cumplir con:*

$$u^{opt}(t) = -R^{-1}(t)\left[B^T(t)P(t) + S(t)\right]x^{opt}(t) \tag{2.30}$$

Hasta este punto se concluye con las condiciones necesarias para el control óptimo proporcionadas por el principio del máximo de Pontryagin, en la siguiente sub-sección se abordan las condiciones suficientes proporcionadas por el principio de optimalidad de Bellman y la ecuación HJB.

2.5.3. La ecuación HJB aplicada al problema LQ

El control dado por (2.30) cumple con las condiciones necesarias impuestas por el principio del máximo, en esta sub-sección trataremos acerca de las condiciones suficientes del control óptimo y veremos como en el caso del regulador lineal con funcional de costo cuadrático las condiciones necesarias son también suficientes.

De la sección 2.4 tenemos que para el caso del problema del LQR la ecuación HJB (2.16) toma la forma:

$$V_t = \sup_{u\in U}\left\{\langle -V_x, A(t)x + B(t)u\rangle - \frac{1}{2}\left(\langle Q(t)x, x\rangle + 2\langle S(t)x, u\rangle + \langle R(t)u, u\rangle\right)\right\} \tag{2.31}$$

con la condición de frontera:

$$V(t_f, x) = \frac{1}{2} x^T G x \tag{2.32}$$

En vista de que $R(t) \geq \delta$, $\delta > 0$ $\forall t \in [t_0, t_f]$ es posible alcanzar el máximo en el Hamiltoniano aplicando el enfoque clásico, es decir se calcula el gradiente de H con respecto a u e igualando a cero, resulta que el control que maximiza el Hamiltoniano está dado por:

$$u^*(t, x) = -R^{-1}(t) \left[B^T(t) V_x(t, x) + S(t) x \right] \tag{2.33}$$

Sustituyendo (2.33) en la ecuación HJB (2.31) se obtiene:

$$\begin{aligned} -V_t(t, x) = &- \frac{1}{2} \left[B^T(t) V_x(t, x) + S(t) x \right]^T R^{-1}(t) \left[B^T(t) V_x(t, x) + S(t) x \right] + \\ &\frac{1}{2} x^T Q(t) x + V_x^T(t, x) A(t) x \end{aligned} \tag{2.34}$$

Con el fin de aplicar las condiciones suficientes vistas en la sección 2.4, es necesario encontrar la solución $V(\cdot, \cdot)$ de (2.34). Entonces, para que la ley de control dada por (2.33) sea óptima, ésta debe de satisfacer la condición necesaria (2.30). Igualando (2.30) y (2.33) se obtiene que:

$$V_x(t, x) = P(t) x$$

Donde $P(\cdot)$ satisface la ecuación diferencial de Ricatti dada por (2.24). Además, se tiene que $V(\cdot, \cdot)$ debe de satisfacer la condición de frontera (2.32), tomando en cuanta estos hechos se hace evidente que la selección de la *value function*:

$$V(t, x) = \frac{1}{2} x^T P(t) x, \tag{2.35}$$

satisface ambas condiciones.

Resta mostrar que la *value function* dada por (2.35) satisface la ecuación HJB (2.34). Calculando las derivadas parciales de $V(t, x)$, obtenemos:

$$V_t(t, x) = x^T \dot{P}(t) x, \qquad V_x(t, x) = P(t) x$$

sustituyendo en (2.34):

$$\begin{aligned} -x^T \dot{P}(t) x = &- \frac{1}{2} \left[B^T(t) P(t) x + S(t) x \right]^T R^{-1}(t) \left[B^T(t) P(t) x + S(t) x \right] + \\ &\frac{1}{2} x^T Q(t) x + (P(t) x)^T A(t) x \\ = &- \frac{1}{2} x^T \left[B^T(t) P(t) + S(t) \right]^T R^{-1}(t) \left[B^T(t) P(t) + S(t) \right] x + \\ &\frac{1}{2} x^T Q(t) x + x^T P(t) A(t) x \end{aligned}$$

Y puesto que $P(\cdot)$ satisface (2.24) se tiene la igualdad deseada.

Por lo tanto la función (2.35) es la *value function* y la ley control retroalimentada dada por (2.30) es el control óptimo.

Resumiendo, se tiene que para un sistema lineal de la forma:

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \qquad x(0) = x_0$$

el control que minimiza el funcional de costo (2.19), está dado por (2.30), y tiene la forma de un control por retroalimentación de estado. Es decir, el control óptimo puede expresarse de la forma:

$$u^{opt}(t) = -K(t)x^{opt}(t)$$

Hasta este punto termina el marco teórico necesario para el planteamiento del problema para una clase especial de sistemas lineales híbridos en la siguiente sección.

Cabe mencionar que en la solución al problema LQR no se tiene en cuenta ningún tipo de restricción en el control ni en la trayectoria de estado, pues el conjunto de controles admisibles corresponde a todo el espacio $\mathbb{L}_2[t_0, t_f; \mathbb{R}^m]$. La falta de inclusión de restricciones en el algoritmo del LQR representa una desventaja significante al tratar de extender esta teoría al campo de los sistemas híbridos donde es común encontrar restricciones sobre el control.

CAPÍTULO 3

Planteamiento del Problema

3.1. El Problema Original

Se considera el siguiente sistema lineal:

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \qquad x(t_0) = x_0 \tag{3.1}$$

con:

$$A(\cdot) \in \mathbb{L}_\infty[t_0, t_f; \mathbb{R}^{n\times n}], \qquad B(\cdot) \in \mathbb{L}_\infty[t_0, t_f; \mathbb{R}^{n\times m}]$$

Es bien sabido que si el par $(A(t), B(t))$ es controlable para todo $t \in [t_0, t_f]$, es posible asignar los polos del sistema al lugar deseado a través de una retroalimentación de estado lineal de la forma:

$$u(t, x(\cdot)) = -K(t)x(t)$$

Si además se desea que la señal de control cumpla con ciertas restricciones de consumo de energía y comportamiento de los estados, estas restricciones pueden ser caracterizadas por el siguiente funcional de costo cuadrático:

$$J\left(u(\cdot)\right) \;=\; \frac{1}{2}\int_{t_0}^{t_f} [\langle Q(t)x(t), x(t)\rangle + \langle R(t)u(t), u(t)\rangle]dt + \frac{1}{2}\langle Gx(t_f), x(t_f)\rangle$$

donde:

$$G \geq 0, \qquad Q(t) \geq 0, \qquad R(t) \geq \delta I, \quad \delta > 0, \qquad \forall t \in [t_0, t_f].$$

Entonces, el problema se reduce al problema del regulador lineal cuadrático estándar (LQR por sus siglas en inglés), tal que el control que asegura el rendimiento está dado por:

$$u^{opt}(t) = -R^{-1}(t)\left[B^T(t)P(t)\right]x^{opt}(t) \tag{3.2}$$

con:

$$\begin{cases} \dot{P}(t) + A^T(t)P(t) + P(t)A(t) + Q(t) \\ - \left[B^T(t)P(t)\right]^T R^{-1}(t) \left[B^T(t)P(t)\right] = 0 \\ P(t_f) = G \end{cases}$$

Además:

$$J(u^{opt}(\cdot)) = \min_{u(\cdot)\in U} \{J(u(\cdot))\} \tag{3.3}$$

Donde U es el conjunto de todos los controles admisibles que para el caso del LQR es todo el espacio $\mathbb{L}_2[t_0, t_f; \mathbb{R}^m]$ pues en la teoría del LQR no se toma en cuenta ningún tipo de restricción en los estados ni el control.

Para la clase de sistemas híbridos que se estudian en este trabajo, es posible caracterizarla como una familia de sistemas lineales variantes en el tiempo con restricciones en la señal de control u de tal forma que u es una señal de *tipo escalera*.

Un ejemplo de este tipo de control es dado en la figura 3.1, en donde la señal de control unidimensional $u(t)$ solo puede tomar un valor (nivel) dentro del conjunto $\mathcal{Q} = \{-4, -3, -2, -1, 0, 1, 2, 3, 4\}$ durante el intervalo de tiempo $[t_{i-1}, t_i), i = 1, \ldots, 12$. Además, a la señal de control solo se le permite cambiar su valor en los tiempos $t_0, t_1, \ldots, t_{12} = t_f$ manteniéndose fija entre estos tiempos, en otras palabras u es constante a trozos.

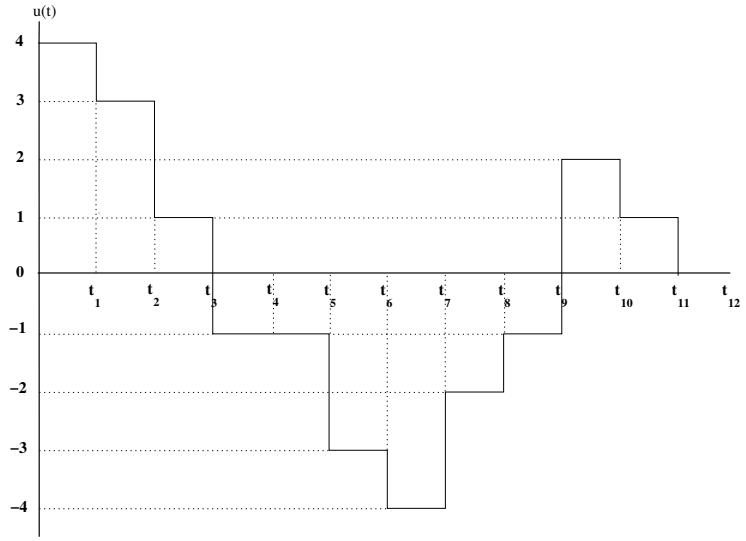

Figura 3.1: Control $u(t) \in \mathcal{S}$

Es fácil ver que la función $u(\cdot) \in \mathbb{L}_2[t_0, t_f]$, pero el enfoque clásico del LQR, en general

no proporciona un resultado de este tipo.

En el caso multidimensional se trata con controles $u(t) \in \mathbb{R}^m$, donde cada componente $u_k(t)$, $(k = 1, \ldots, m)$, está restringido a un conjunto finito de niveles admisibles denotado por $\mathcal{Q}_k$, $(k = 1, \ldots, m)$.

De manera general, se tiene para cada componente de $u(\cdot) = [u_1(\cdot), \ldots, u_m(\cdot)]^T$ el siguiente conjunto finito de niveles admisibles:

$$\mathcal{Q}_k := \{q_k^{(j)} \in \mathbb{R} : q_k^{(j)} < q_k^{(j+1)}; j = 1, \ldots, M_k\} \tag{3.4}$$

donde no necesariamente todos los conjuntos $\mathcal{Q}_k$ son iguales, ni tampoco tienen el mismo número de elementos.

También, se define la sucesión finita de tiempos de conmutación para cada componente $u_k(\cdot)$, $k = 1, \ldots, m$ como:

$$\mathcal{T}_k := \{t_i^{(k)} \in \mathbb{R}^+, i = 0, \ldots, N_k\} \tag{3.5}$$

donde $\mathbb{R}^+$ denota el conjunto de los reales no negativos. Es decir, $\mathbb{R}^+ = \{x \in \mathbb{R} | x \geq 0\}$. Cada sucesión $\mathcal{T}_k$ es estrictamente creciente, es decir:

$$t_0^{(k)} < t_1^{(k)} < \ldots < t_{N_k}^{(k)}$$

Nota. Aunque las sucesiones $\mathcal{T}_k$ en general son diferentes, se tienen las siguientes restricciones:

$$t_0 = t_0^{(1)} = \ldots = t_0^{(m)} \tag{3.6}$$

Es decir, se pide que los tiempos iniciales y finales para cada $\mathcal{T}_k$ sean iguales.

Teniendo en cuenta las restricciones sobre los niveles de u y los tiempos de cambio, el conjunto de controles admisibles denotado por $\mathcal{S}$ se define como:

$$\begin{aligned} \mathcal{S} =& \mathcal{S}_1 \times \ldots \times \mathcal{S}_m \\ =& \{u(t) = [u_1(t), \ldots, u_m(t)]^T \in \mathbb{R}^m : u_k(\cdot) \in \mathcal{S}_k, k = 1, \ldots, m\} \end{aligned} \tag{3.7}$$

Con cada conjunto $\mathcal{S}_k$ definido como:

$$\begin{aligned} \mathcal{S}_k :=& \{v : [t_0, t_f] \to \mathbb{R} | v(t) = \sum_{i=1}^{N_k} I_{[t_{i-1}^{(k)}, t_i^{(k)})}(t) q_k^{(j_i)}; \quad q_k^{(j_i)} \in \mathcal{Q}_k; \\ & j_i \in \mathbb{Z} \cap [1, M_k]; t_i^{(k)} \in \mathcal{T}_k \quad \forall i = 0, \ldots, N_k\}. \end{aligned}$$

Donde la función $I_{[t_{i-1}^k, t_i^k)}(t)$ es la función indicadora del intervalo $[t_{i-1}^k, t_i^k)$ dada por:

$$I_{[t_{i-1}^k, t_i^k)}(t) = \begin{cases} 1, & \text{si: } t \in [t_{i-1}^k, t_i^k) \\ 0, & \text{si: } t \notin [t_{i-1}^k, t_i^k) \end{cases}$$

Entonces, el conjunto $\mathcal{S}$ puede ser visto como el conjunto de todas la funciones posibles $u : [t_0, t_f] \to \mathbb{R}^m$, tal que, para cada intervalo $[t_{i-1}^k, t_i^k)$ cada componente de $u(\cdot)$ permanece en un nivel constante $q_{j_i}^{(k)} \in \mathcal{Q}^k$, pudiendo cambiar de nivel solamente en los tiempos $t_i^k \in \mathcal{T}_k$ especificados , $i = 0, \ldots, N_k$.

Ejemplo 3.1 *Supongamos que $u(t) \in \mathbb{R}^2$. Es decir, la señal de control $u(\cdot)$ tiene dos componentes denotados por: $u_1(\cdot)$, $u_2(\cdot)$, y además se tienen restricciones sobre estos componentes de la siguiente forma:*

$$\mathcal{Q}_1 = \{0, 1, 2\}, \qquad \mathcal{Q}_2 = \{0, -1\}$$

Es decir, el componente u_1 solo puede tomar valores dentro del conjunto descrito por $\mathcal{Q}_1$ y u_2 en $\mathcal{Q}_2$. Además, la secuencia de tiempos de cambio para cada componente está dada por:

$$\mathcal{T}_1 = \{0, 0.5, 1\}, \qquad \mathcal{T}_2 = \{0, 0.33, 0.66, 1\}$$

Por lo que el conjunto de controles admisibles $\mathcal{S}$ resulta en:

$$\mathcal{S} = \mathcal{S}_1 \times \mathcal{S}_2 = \{[u_1, u_2]^T \in \mathbb{R}^2 : u_1 \in \mathcal{S}_1, u_2 \in \mathcal{S}_2\}$$

donde:

$$\begin{aligned}
\mathcal{S}_1 =& \{v : [0,1] \to \mathbb{R} : v(t) = I_{[0,0.5)}(t) q_1^{(j_1)} + I_{[0.5,1)}(t) q_1^{(j_2)}, \\
& q_1^{(j_i)}, \in \mathcal{Q}^1, j_i \in \{1,2,3\}, \forall i = 1,2.\} \\
\mathcal{S}_2 =& \{w : [0,1] \to \mathbb{R} : w(t) = I_{[0,0.33)}(t) q_2^{(j_1)} + I_{[0.33,0.66)}(t) q_2^{(j_2)} + I_{[0.66,1)}(t) q_2^{(j_3)}, \\
& q_2^{(j_i)} \in \mathcal{Q}^2, j_i \in \{1,2\}, i = 1,2,3\}
\end{aligned}$$

En este caso tenemos: $M_1 = 3$, $M_2 = 2$, $N_1 = 2$ y $N_2 = 3$. La cardinalidad del conjunto $\mathcal{S}$ en el ejemplo está dada por:

$$|\mathcal{S}| = 3^2 \cdot 2^3 = 72$$

Es decir, se tienen 72 controles admisibles, de entre los cuales tenemos que encontrar el que minimiza el criterio de desempeño.

De manera general, la cardinalidad del conjunto de controles admisibles $\mathcal{S}$ como el descrito por (3.7), para el caso donde el control $u(t) \in \mathbb{R}^m$, está dado por:

$$|\mathcal{S}| = \prod_{l=1}^{m} M_l^{N_l} \tag{3.8}$$

Es fácil ver que todos los controles en $\mathcal{S}$ son cuadrático integrables, es decir $\mathcal{S} \subset \mathbb{L}_2[t_0, t_f; \mathbb{R}^m]$. En comparación con el problema clásico del LQR sin restricciones, el control óptimo $u^{opt}(\cdot)$

también es un elemento de $\mathbb{L}_2[t_0, t_f; \mathbb{R}^m]$ aunque en general este último no pertenece a $\mathcal{S}$, ya que la estructura de *escalera* de la señal $u(\cdot) \in \mathcal{S}$, caracterizada por las restricciones sobre los niveles constantes y los tiempos de cambio específicos, no son contemplados en el diseño del LQR clásico.

Resumiendo, la teoría clásica del LQR no puede ser aplicada directamente en la clase de sistemas híbridos estudiada aquí, por lo que se presenta un nuevo problema, el cual, puede ser formulado como:

Problema 3.1 *Encontrar el control $u^*(\cdot) \in \mathcal{S}$ tal que:*

$$J\left(u^*(\cdot)\right) = \min_{u(\cdot)\in\mathcal{S}} \{J\left(u(\cdot)\right)\} \tag{3.9}$$

El principal problema para encontrar el control $u^*(\cdot) \in \mathcal{S}$ que cumpla con (4.29), se encuentra al no tener una herramienta matemática que ataque el problema directamente. Para resolver este inconveniente se propone el uso de elementos de análisis convexo presentando una versión relajada del problema original y así obtener un conjunto de soluciones[1] del problema relajado, a partir del cual se puede obtener la solución al problema original.

3.2. El Problema Relajado

Para el problema (4.29) descrito en la sección anterior no existe un herramental matemático bien definido para resolver el problema directamente, es por eso que se utilizan herramientas de análisis convexo para hacer una generalización del problema. Esta generalización consiste en llevar el problema particular a una versión relajada que será convexa, por lo que surgirá un nuevo problema de programación convexa[2]. La ventaja de usar esta metodología es que sí existe el herramental matemático para resolver el problema relajado de manera eficiente.

Para llevar el problema original a un problema de programación convexa, primero es necesario estudiar las propiedades del funcional de costo asociado al problema, así como las propiedades del conjunto de controles admisibles $\mathcal{S}$.

A continuación se muestran algunas propiedades del funcional de costo $J(u(\cdot))$ que serán de utilidad para el planteamiento del problema relajado.

[1] Si el funcional de costo es estrictamente convexo y el conjunto de controles admisibles es convexo, entonces, la solución es única

[2] El problema de programación convexa consiste en encontrar el mínimo de una función convexa en un conjunto de restricciones que también es convexo.

Proposición 3.1 *Sea el sistema dado por (3.1)*

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \qquad x(t_0 = 0) = x_0 \tag{3.10}$$

Entonces, la solución $x(t)$ de (3.10) es una función afín en $u(\cdot)$.

Prueba. Aplicando la fómula de Cauchy a la ecuación (3.10), tenemos:

$$x(t)^{u(\cdot)} = \Phi(t,0)x(0) + \int_{\tau=0}^{t} \Phi(t,\tau)B(\tau)u(\tau)d\tau \tag{3.11}$$

donde $\Phi(t,t_0)$ es la matriz de transición de estados del sistema (3.10) (ver [7]), definida por: $\Phi(t,t_0) = X(t)X^{-1}(t_0)$ y $X(t)$ es una solución a la ecuación homogénea.

$$\dot{X}(t) = A(t)X(t) \tag{3.12}$$

con $X(t_0)$ invertible. De la definición 2.5, tenemos:

$$\begin{aligned}
x(t)^{(1-\lambda)u_1(\cdot)+\lambda u_2(\cdot)} &= \Phi(t,0)x(0) + \int_{\tau=0}^{t} \Phi(t,\tau)B(\tau)\left[(1-\lambda)u_1(\tau) + \lambda u_2(\tau)\right]d\tau \\
&= \Phi(t,0)x(0) + (1-\lambda)\int_{\tau=0}^{t} \Phi(t,\tau)B(\tau)u_1(\tau)d\tau \\
&\quad +\lambda \int_{\tau=0}^{t} \Phi(t,\tau)B(\tau)u_2(\tau)d\tau \\
&= (1-\lambda)\left(\Phi(t,0)x(0) + \int_{\tau=0}^{t} \Phi(t,\tau)B(\tau)u_1(\tau)d\tau\right) + \\
&\quad \lambda\left(\Phi(t,0)x(0) + \int_{\tau=0}^{t} \Phi(t,\tau)B(\tau)u_2(\tau)d\tau\right) \\
&= (1-\lambda)x(t)^{u_1(\cdot)} + \lambda x(t)^{u_2(\cdot)}. \quad \blacksquare
\end{aligned}$$

Proposición 3.2 *Sea $f : A \subset \mathbb{R}^m \to \mathbb{R}^p$ una función convexa y $g : B \subset \mathbb{R}^n \to \mathbb{R}^m$ una función afín, tal que, $g(B) \subset A$. Entonces, la composición de f y g denotada por $h = f \circ g$ es una función convexa.*

Prueba. Sean x_1 y x_2 elementos de B, se tiene:

$$\begin{aligned}
h\left((1-\lambda)x_1 + \lambda x_2\right) &= f\left[g\left((1-\lambda)x_1 + \lambda x_2\right)\right] \\
&= f\left[(1-\lambda)g(x_1) + \lambda g(x_2)\right] \\
&\leq (1-\lambda)f(g(x_1)) + \lambda f(g(x_2)) = (1-\lambda)h(x_1) + \lambda h(x_2) \quad \blacksquare
\end{aligned}$$

En base a las proposiciones anteriores tenemos la siguiente propiedad fundamental sobre el funcional de costo dado por (3.2).

Proposición 3.3 *El funcional de costo dado por*

$$J\left(u(\cdot)\right) = \frac{1}{2}\int_0^T [\langle Q(t)x(t), x(t)\rangle + \langle R(t)u(t), u(t)\rangle]dt + \frac{1}{2}\langle Gx(T), x(T)\rangle$$

es una función convexa en u.

Prueba El funcional $J(u(\cdot))$ puede ser descompuesto en la suma de dos funciones convexas de la manera siguiente.

$$\begin{aligned} J(u(\cdot)) &= J_1(u(\cdot)) + J(u(\cdot)) \\ &\text{donde:} \\ J_1(u(\cdot)) &= \frac{1}{2}\int_0^T [\langle R(t)u(t), u(t)\rangle]dt \\ J_2(u(\cdot)) &= \frac{1}{2}\int_0^T [\langle Q(t)x(t), x(t)\rangle]dt + \frac{1}{2}\langle Gx(T), x(T)\rangle \end{aligned}$$

Ahora bien, $J_1(u(\cdot))$ es una función convexa en u, pues su matriz Hessiana es positiva definida. Por otra parte, $J_2(u(\cdot))$ es la composición de una función convexa en x y x a su vez es una función afín en u y a partir de la proposición 3.2, se tiene que $J_2(u(\cdot))$ es una función convexa en u y puesto que la suma de dos funciones convexas es convexa, tenemos el resultado deseado. ■

Ahora cambiamos al estudio del conjunto de controles admisibles $\mathcal{S}$. De la definición 2.1 es fácil ver que el conjunto $\mathcal{S}$ no es convexo. En este punto es donde se aplica la relajación del problema al tomar un nuevo conjunto de restricciones que resultará en un conjunto convexo, la forma de obtenerlo es a partir del lema 2.1, por lo que el nuevo conjunto de restricciones está dado por $conv(\mathcal{S})$ de la siguiente manera:

$$\begin{aligned} conv(S) &= \{v \in \mathbb{R}^m : v(\cdot) = \sum_{i=1}^{|\mathcal{S}|} \mu_i u_i(\cdot), \quad \sum_{i=1}^{|\mathcal{S}|} \mu_i = 1, \\ &\qquad u_i \in \mathcal{S}, \mu_i \geq 0, \quad \forall i = 1, \ldots, |\mathcal{S}|\} \end{aligned} \tag{3.13}$$

En este punto es útil mencionar un par de lemas que serán de utilidad en la sección siguiente.

Lema 3.1 *Sean $\mathcal{A} \subset \mathbb{R}^p$ y $\mathcal{B} \subset \mathbb{R}^q$ conjuntos finitos con $|\mathcal{A}| = M_1$ y $|\mathcal{B}| = M_2$. Entonces:*

$$conv(\mathcal{A} \times \mathcal{B}) = conv(\mathcal{A}) \times conv(\mathcal{B}).$$

Prueba: Primero es fácil ver que el conjunto $conv(\mathcal{A}) \times conv(\mathcal{B})$ es un conjunto convexo que pertenece al espacio de números reales $p+q$ - dimensional. Además, $\mathcal{A} \times \mathcal{B} \subset conv(\mathcal{A}) \times conv(\mathcal{B})$ y por la definición de *convex hull*, tenemos:

$$conv(\mathcal{A} \times \mathcal{B}) \subseteq conv(\mathcal{A}) \times conv(\mathcal{B})$$

Por otra parte, a partir de la caracterización del *convex hull* dada por el lema 2.1, los elementos del conjunto $conv(\mathcal{A}) \times conv(\mathcal{B})$ tienen la siguiente forma:

$$\left[\sum_{i=1}^{M_1} \alpha_i a_i^T, \sum_{j=1}^{M_2} \beta_j b_j^T\right]^T \tag{3.14}$$

con:

$$a_i \in \mathcal{A},\ \alpha_i \geq 0,\ \forall i = 1, \ldots, M_1, \quad \sum_{i=1}^{M_1} \alpha_i = 1,$$
$$b_j \in \mathcal{B},\ \beta_j \geq 0,\ \forall j = 1, \ldots, M_2, \quad \sum_{j=1}^{M_2} \beta_j = 1.$$

Entonces, a partir de la caracterización de los elementos de $conv(\mathcal{A}) \times conv(\mathcal{B})$ se tiene:

$$\begin{aligned}\sum_{i=1}^{M_1} \alpha_i a_i^T &= \left(\sum_{j=1}^{M_2} \beta_i\right)\left(\sum_{i=1}^{M_1} \alpha_i a_i^T\right) = \sum_{i=1}^{M_1}\left(\alpha_i a_i^T \sum_{j=i}^{M_2} \beta_j\right) \\ &= \sum_{i=1}^{M_1}\sum_{j=1}^{M_2} \alpha_i \beta_j a_i^T\end{aligned}$$

y

$$\begin{aligned}\sum_{j=1}^{M_2} \beta_j b_j^T &= \left(\sum_{i=1}^{M_1} \alpha_i\right)\left(\sum_{j=1}^{M_2} \beta_j b_j^T\right) = \sum_{j=1}^{M_2}\left(\beta_j b_j^T \sum_{i=1}^{M_1} \alpha_i\right) \\ &= \sum_{j=1}^{M_2}\sum_{i=1}^{M_1} \alpha_i \beta_j b_j^T = \sum_{i=1}^{M_1}\sum_{j=1}^{M_2} \alpha_i \beta_j b_j^T\end{aligned}$$

Lo cual implica que:

$$\begin{aligned}\left[\sum_{i=1}^{M_1} \alpha_i a_i^T, \sum_{j=1}^{M_2} \beta_j b_j^T\right]^T &= \left[\sum_{i=1}^{M_1}\sum_{j=1}^{M_2} \alpha_i \beta_j a_i^T, \sum_{i=1}^{M_1}\sum_{j=1}^{M_2} \alpha_i \beta_j b_j^T\right]^T \\ &= \sum_{i=1}^{M_1}\sum_{j=1}^{M_2} \left[\alpha_i \beta_j a_i^T, \alpha_i \beta_j b_j^T\right]^T \\ &= \sum_{i=1}^{M_1}\sum_{j=1}^{M_2} \alpha_i \beta_j \left[a_i^T, b_j^T\right]^T\end{aligned}$$

Además:

$$\sum_{i=1}^{M_1}\sum_{j=1}^{M_2}\alpha_i\beta_j = \sum_{i=1}^{M_1}\alpha_i = 1,\ \alpha_i\beta_j \geq 0\ \forall i = 1,\ldots,M_1,\ \forall j = 1,\ldots,M_2.$$

Por lo tanto:

$$conv(\mathcal{A}) \times conv(\mathcal{B}) \subseteq conv(\mathcal{A}\times\mathcal{B})\blacksquare$$

A partir del lema 3.1 se obtiene el siguiente corolario:

Corolario 3.1 *Sea una familia finita de conjuntos finitos* $\{\mathcal{A}_i : |\mathcal{A}_i| < \infty, \forall i = 1,\ldots,N\}$. *Entonces:*

$$conv(\mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_N) = conv(\mathcal{A}_1) \times conv(\mathcal{A}_2) \times \ldots \times conv(\mathcal{A}_N)$$

Prueba: Aplicando inducción matemática y el lema 3.1 se obtiene el resultado deseado. ■

Haciendo uso del corolario 3.1, se tiene que el conjunto de restricciones está dado por:

$$conv(\mathcal{S}) = conv(\mathcal{S}_1) \times \ldots \times conv(\mathcal{S}_m) \tag{3.15}$$

donde cada conjunto $conv(\mathcal{S}_k)$ está dado por:

$$\begin{aligned} conv(\mathcal{S}_k) = \{w(\cdot) : [t_0,t_f] \to \mathbb{R} | w(t) &= \sum_{j=1}^{|\mathcal{S}_k|}\lambda_j v_j(t); \\ v_j(\cdot) \in \mathcal{S}^k, \lambda_j \geq 0,\ \forall j &= 1,\ldots,|\mathcal{S}_k|\} \\ = \{w(\cdot) : [t_0,t_f] \to \mathbb{R} | w(t) &= \sum_{j=1}^{N_k} I_{[t_{i-1}^{(k)}, t_i^{(k)})}(t)\bar{q}_i, \\ \bar{q}_i \in conv(\mathcal{Q}^k)\} & \end{aligned}$$

Esta última caracterización del conjunto $conv(\mathcal{S})$ será de gran utilidad en la siguiente sección. Es fácil ver que los conjuntos $\mathcal{S}$ y $conv(\mathcal{S})$ son cerrados y acotados. Una propiedad importante a considerar, es que mientras el conjunto $\mathcal{S}$ es finito (ver ecuación (3.8)), el conjunto $conv(\mathcal{S})$ no lo es, pero a cambio se obtiene un conjunto de restricciones convexo y se pueden aplicar métodos númericos eficientemente.

Teniendo en cuenta lo anterior se puede formular el siguiente problema:

Problema 3.2 *Encontrar el control* $\hat{u}(\cdot) \in conv(\mathcal{S})$ *tal que:*

$$J(\hat{u}(\cdot)) = \min_{u(\cdot)\in conv(\mathcal{S})}\{J(u(\cdot))\} \tag{3.16}$$

Hasta este punto termina la formulación del problema. Resumiendo tenemos un sistema en tiempo continuo, lineal y variante en el tiempo con un conjunto de controles admisibles del *tipo escalera*, es decir señales de naturaleza discreta, con un secuencia de conmutación fija, por lo que surge un sistema híbrido.

Los sistemas híbridos vistos en este capítulo pueden ser caracterizados como sistemas lineales con controles constantes a trozos, la principal dificultad está en no contar con un método eficiente para el cálculo del control óptimo. A partir de este inconveniente, se propone relajar las restricciones, haciendo el conjunto de controles admisibles convexo, con el fin de aplicar métodos numéricos de manera eficiente.

En el siguiente capítulo se abordarán las cuestiones relacionadas al uso del algoritmo del gradiente con proyección para la búsqueda del control óptimo del problema relajado, así como las cuestiones relacionadas con la obtención del control óptimo para el problema original y se presentan algunos ejemplos que muestran la efectividad del enfoque propuesto.

CAPÍTULO 4

Resultado Principal

El problema 3.2 de la página 65 consiste en un problema de programación convexa con restricciones, el cual es sencillo de resolver analíticamente en el caso de sistemas lineales invariantes en el tiempo y para sistemas variantes en el tiempo (con matriz de transición de estados conocida), donde el conjunto $conv(\mathcal{Q}_1 \times \ldots \times \mathcal{Q}_m) = \mathbb{R}^m$ y donde se tiene cierto desacoplamiento entre los componentes del control u.

En el caso general, no es posible resolver analíticamente el problema, por lo que se justifica el uso de métodos numéricos. Técnicas basadas en gradiente son usadas para el problema restringido ($conv(\mathcal{Q}_1 \times \ldots \times \mathcal{Q}_m) \subset \mathbb{R}^m$) obteniendo resultados satisfactorios.

4.1. Solución Analítica del Problema Relajado

Se considera el siguiente sistema lineal variante en el tiempo:

$$\dot{x}(t) = A(t)x(t) + B(t)u(t), \quad x(t_0) = x_0. \tag{4.1}$$

y con el funcional de costo asociado:

$$J(u(\cdot)) = \frac{1}{2}\int_{t_0}^{t_f} \left(\langle Q(t)x(t), x(t)\rangle + \langle R(t)u(t), u(t)\rangle\right) dt + \frac{1}{2}\langle Gx(t_f), x(t_f)\rangle. \tag{4.2}$$

Para la solución analítica del problema 3.2, primero se asume que se conoce la matriz de transición de estados del sistema y posteriormente se plantea un problema equivalente, el cual resulta más sencillo de analizar.

El funcional de costo dado por (4.2) se puede reescribir de la siguiente manera:

$$J(u(\cdot)) = \sum_{i=1}^{N} \tilde{J}_i(u(\cdot)) \tag{4.3}$$

Donde:

$$\begin{aligned}
\tilde{J}_i(u(\cdot)) =& \frac{1}{2}\int_{t_{i-1}}^{t_i} \left(\langle Q(t)x(t), x(t)\rangle + \langle R(t)u(t), u(t)\rangle\right) dt. \quad \forall i = 1, \ldots, N-1. \\
\tilde{J}_N(u(\cdot)) =& \frac{1}{2}\int_{t_{N-1}}^{t_N} \left(\langle Q(t)x(t), x(t)\rangle + \langle R(t)u(t), u(t)\rangle\right) dt + \frac{1}{2}\langle Gx(t_f), x(t_f)\rangle
\end{aligned}$$

Ahora como primer paso, se considera que todas las secuencias de tiempos de conmutación $\mathcal{T}^k$ son iguales y solo se considera la secuencia de conmutación $\mathcal{T}$ dada por:

$$\mathcal{T} = \{t_i \in \mathbb{R}^+ : t_0 < t_1 < \ldots < t_N = t_f\} \tag{4.4}$$

Entonces, la descomposición del funcional original dada por (4.3) se pude realizar de tal forma que en cada intervalo de integración $[t_{i-1}, t_i], t_i \in \mathcal{T},\ \forall i = 1, \ldots, N$ el control $u(\cdot)$ sea una constante, es decir:

$$\begin{aligned}
\tilde{J}_i(u(\cdot)) =& \frac{1}{2}\int_{t_{i-1}}^{t_i} \left(\langle Q(t)x(t, \bar{u}^i), x(t, \bar{u}^i)\rangle + \langle R(t)\bar{u}^i, \bar{u}^i\rangle\right) dt. \\
& \forall i = 1, \ldots, N-1. \\
\tilde{J}_N(u(\cdot)) =& \frac{1}{2}\int_{t_{N-1}}^{t_N} \left(\langle Q(t)x(t, \bar{u}^N), x(t, \bar{u}^N)\rangle + \langle R(t)\bar{u}^N, \bar{u}^N\rangle\right) dt + \\
& \frac{1}{2}\langle Gx(t_f, \bar{u}^N), x(t_f, \bar{u}^N)\rangle
\end{aligned} \tag{4.5}$$

De esta forma cada control $u(\cdot)$ admisible está dado por:

$$u(t) = \sum_{i=1}^{N} I_{[t_{i-1}, t_i)}(t)\bar{u}^i. \quad \bar{u}^i \in \Omega = conv(\mathcal{Q}_1) \times \ldots \times conv(\mathcal{Q}_m)$$

En esta nueva descomposición del funcional original, es posible ver una propiedad fundamental, pues el hecho de que el control $\bar{u}^i$ es constante en cada $J_i(u(\cdot))$, el problema original realmente trata de N problemas de dimensión finita. Es decir, cada problema de dimensión finita consiste en encontrar el control $\bar{u}^{i*}$ tal que:

$$\tilde{J}_i(\bar{u}^{i*}) = \min_{\substack{\bar{u}_k^i \in conv(\mathcal{Q}_k) \\ k=1,\ldots,m}} (\tilde{J}_i(\bar{u}^i)) \tag{4.6}$$

Para resolver este último problema, en primer lugar se tiene que la solución del sistema dado por (4.1) en el intervalo de tiempo $[t_{i-1}, t_i]$ se puede expresar como:

$$x(t, \bar{u}^i) = \Delta_i(t) + \Gamma_i(t)\bar{u}^i \; \forall t \in [t_{i-1}, t_i] \tag{4.7}$$

donde:

$$\Delta_i(t) = \Phi(t, t_{i-1})x(t_{i-1})$$
$$\Gamma_i(t) = \int_{t_{i-1}}^{t} \Phi(t, \sigma)B(\sigma)d\sigma$$

Sustituyendo (4.7) en (4.5), el problema descrito por (4.6) es equivalente a minimizar:

$$\begin{aligned} \bar{J}_i(\bar{u}^i) =& \frac{1}{2}\int_{t_{i-1}}^{t_i} \left[2\bar{\Delta}_i^T(t)\bar{u}^i + \left(\bar{u}^i\right)^T \bar{\Gamma}_i(t)\bar{u}^i\right] dt, \quad i = 1, \ldots, N-1. \\ \bar{J}_N(\bar{u}^N) =& \frac{1}{2}\int_{t_{N-1}}^{t_N} \left[2\bar{\Delta}_N^T(t)\bar{u}^N + \left(\bar{u}^N\right)^T \bar{\Gamma}_N(t)\bar{u}^N\right] dt + \frac{1}{2}\left[2\tilde{\Delta}_N^T\bar{u}^N + \left(\bar{u}^N\right)^T \tilde{\Gamma}_N\bar{u}^N\right] \end{aligned} \tag{4.8}$$

donde:

$$\begin{aligned} \bar{\Delta}_i(t) &= \Gamma_i(t)^T Q(t)\Delta_i(t) = \left(\int_{t_{i-1}}^{t} \Phi(t, \sigma)B(\sigma)d\sigma\right)^T Q(t)\Phi(t, t_{i-1})x(t_{i-1}) \\ \bar{\Gamma}_i(t) &= \Gamma_i^T(t)Q(t)\Gamma_i(t) + R(t) = \left(\int_{t_{i-1}}^{t} \Phi(t, \sigma)B(\sigma)d\sigma\right)^T Q(t)\left(\int_{t_{i-1}}^{t} \Phi(t, \sigma)B(\sigma)d\sigma\right) + R(t) \\ \tilde{\Delta}_N &= \Gamma_N(t_N)^T G\Delta_N(t_N) \\ \tilde{\Gamma}_N &= \Gamma_N^T(t_N)G\Gamma_N(t_N) \end{aligned} \tag{4.9}$$

Ahora bien, en el caso donde $\bar{u}^{i*}$ está en el interior de $\Omega = conv(\mathcal{Q}_1) \times conv(\mathcal{Q}_m)$, $\forall i = 1, \ldots, N$, se tiene el siguiente teorema:

Teorema 4.1 *El control $\hat{u}(t)$ que es solución al Problema 2 (sin restricciones en los conjuntos $\mathcal{Q}_k$, es decir, cada conjunto $\mathcal{Q}_k = \mathbb{R}$), está dado por:*

$$\hat{u}(t) = \sum_{k=1}^{N} I_{[t_{i-1}, t_i)}(t)\bar{u}^{i*}$$

donde:

$$\begin{aligned} \bar{u}^{i*} &= -\left[\int_{t_{i-1}}^{t_i} \bar{\Gamma}_i(t)dt\right]^{-1}\left[\int_{t_{i-1}}^{t_i} \bar{\Delta}_i(t)dt\right], \quad \forall i = 1, \ldots, N-1. \\ \bar{u}^{N*} &= -\left[\int_{t_{N-1}}^{t_N} \bar{\Gamma}_N(t)dt + \tilde{\Gamma}_N\right]^{-1}\left[\int_{t_{N-1}}^{t_N} \bar{\Delta}_N(t)dt + \tilde{\Delta}_N\right] \end{aligned} \tag{4.10}$$

y las matrices $\bar{\Delta}_i(t)$, $\bar{\Gamma}_i(t)$, $\tilde{\Delta}_N$ y $\tilde{\Gamma}_N$ están dadas por (4.9).

Prueba. Derivando (4.8) con respecto a $\bar{u}^i$, se obtiene:

$$\nabla \bar{J}_i = \int_{t_{i-1}}^{t_i} \bar{\Delta}_i(t)dt + \left(\int_{t_{i-1}}^{t_i} \bar{\Gamma}_i(t)dt \right) \bar{u}^i, \quad \forall i = 1, \ldots, N-1.$$
$$\nabla \bar{J}_N = \int_{t_{i-N}}^{t_N} \bar{\Delta}_N(t)dt + \left(\int_{t_{N-1}}^{t_N} \bar{\Gamma}_N(t)dt \right) \bar{u}^N + \tilde{\Delta}_N + \tilde{\Gamma}_N \bar{u}^N$$

Y aplicando las condiciones necesarias de primer orden para la optimalidad de $\bar{u}^i$, se tiene:

$$0 = \int_{t_{i-1}}^{t_i} \bar{\Delta}_i(t)dt + \left(\int_{t_{i-1}}^{t_i} \bar{\Gamma}_i(t)dt \right) \bar{u}^{i*}$$
$$0 = \int_{t_{N-1}}^{t_N} \bar{\Delta}_N(t)dt + \left(\int_{t_{N-1}}^{t_N} \bar{\Gamma}_N(t)dt \right) \bar{u}^{N*} + \tilde{\Delta}_N + \tilde{\Gamma}_N \bar{u}^{N*}$$

Debido a que las matrices $Q(t) \geq 0$ y $R(t) > \delta$, $\delta > 0$, $\forall t$ es fácil ver que la matriz $\bar{\Gamma}_i(t)$ es simétrica, y además es estrictamente positiva definida, pues:

$$\begin{aligned} \eta^T \bar{\Gamma}_i(t)\eta =& \eta^T \left[\Gamma_i^T(t) Q(t) \Gamma_i(t) + R(t) \right] \eta \\ =& \eta^T \Gamma_i^T(t) Q(t) \Gamma_i(t) \eta + \eta^T R(t) \eta > \delta \end{aligned}$$

La integral de $\bar{\Gamma}_i(t)$ en el intervalo $[t_{i-1}, t_i]$, es estrictamente positiva definida, por lo que la inversa siempre existe y entonces, despejando $\bar{u}^{(i)*}$ de la última ecuación se obtiene el resultado deseado. Además, puesto que cada $\bar{J}_i(\cdot)$, $i = 1, \ldots, N$ es una función convexa, se tiene que la condición de primer orden también es suficiente. ■

Ahora bien, para el caso con restricciones en los niveles $\mathcal{Q}_k$, el problema se vuelve más complicado. En primer lugar se hace una caracterización funcional de las restricciones, es decir, en especial se buscan restricciones de la forma $r(\bar{u}^i) \leq 0$. Obtener este tipo de caracterización es muy fácil para el conjunto de restricciones dadas por Ω, pues para cada componente $\bar{u}_k^i$ de $\bar{u}^i$ se tiene:

$$q_0^{(k)} \leq \bar{u}_k^i \leq q_M^{(k)}, \ \forall k = 1, \ldots, m, \ \forall i = 1, \ldots, N, \ q_j^{(k)} \in \mathcal{Q}_k, \ j = 1, \ldots, M.$$

por lo que la restricción, $\bar{u}^i \in \Omega$, se puede interpretar como $2m$ restricciones de la forma:

$$\begin{array}{cc} h_1(\bar{u}^i) = q_0^{(1)} - \bar{u}_1^i \leq 0, & g_1(\bar{u}^i) = \bar{u}_1^i - q_M^{(1)} \leq 0 \\ h_2(\bar{u}^i) = q_0^{(2)} - \bar{u}_2^i \leq 0, & g_2(\bar{u}^i) = \bar{u}_2^i - q_M^{(2)} \leq 0 \\ \vdots & \vdots \\ h_m(\bar{u}^i) = q_0^{(m)} - \bar{u}_m^i \leq 0, & g_m(\bar{u}^i) = \bar{u}_m^i - q_M^{(m)} \leq 0 \end{array} \tag{4.11}$$

cada punto $\bar{u}^i$ que cumple con las restricciones anteriores se denomina **realizable**.

Se puede consultar en [5], [18] y [20], las condiciones necesarias y suficientes para un extremo con restricciones del tipo desigualdad. En especial las condiciones de Karush-Kuhn-Tucker para las cuales son necesarias las siguientes definiciones tomadas de [18].

Definición 4.1 *Una restricciíon de desigualdad $r_j(u) \leq 0$ se dice que está **activa** en el punto realizable u si $r_j(u) = 0$ e **inactiva** en u si $r_j(u) < 0$.*

Entonces, es claro que las restricciones activas en un punto realizable u restringen el dominio de realizabilidad en vecindades de u, mientras que, las restricciones inactivas no tienen influencia en las vecindades de u.

Definición 4.2 *Sea u^* un punto que satisface las restricciones:*

$$r_j(u^*) \leq 0,\ j = 1, \ldots, 2m. \tag{4.12}$$

*donde cada $r_j \in \mathcal{C}^1$ y sea $\mathcal{A}(u^*)$ un conjunto de índices j para los cuales $r_j(u^*) = 0$. Entonces, u^* se dice que es un **punto regular** de las restricciones (4.12) si los vectores gradiente $\nabla r_j(u^*)$, $j \in \mathcal{A}(u^*)$ son linealmente independientes.*

Es decir, un punto u^* es regular si los gradientes de las restricciones activas son linealmente independientes.

Para el caso de las restricciones descritas por (4.11) es claro que solamente se tienen a lo mucho m restricciones activas. Además, debido a la simplicidad de las restricciones, los gradientes de éstas siempre son linealmente independientes, por lo que todos los puntos en Ω son regulares.

Teorema 4.2 *(Condiciones **Karush-Kuhn-Tucker**). Sea u^* un mínimo relativo para el problema:*

$$\begin{aligned} &\textit{minimizar } J(u) \\ &\textit{sujeto a } h(u) \leq 0, g(u) \leq 0 \end{aligned}$$

y supóngase que u^* es un punto regular para las restricciones. Entonces, existe un vector $\lambda \in \mathbb{R}^m$ y un vector $\mu \in \mathbb{R}^m$ tal que:

$$\begin{aligned} \nabla J(u^*) + \sum_{l=1}^{m} [\lambda_l \nabla h_l(u^*) + \mu_l \nabla g_l(u^*)] = 0 \\ \lambda_l \geq 0,\ \lambda_l h_l(u^*) = 0,\ l = 1, \ldots, m. \\ \mu_l \geq 0,\ \mu_l g_l(u^*) = 0,\ l = 1, \ldots, m. \end{aligned}$$

Además, en presencia de convexidad las condiciones de Karush-Kuhn-Tucker son también suficientes. Entonces, aplicando el teorema 4.2 al problema 2 del capítulo anterior, se obtiene la solución del problema relajado. En general, la determinación de la forma analítica del control óptimo $\hat{u}$ no es simple, pues se trata más de una búsqueda usando el método de prueba y error (ver [20]). En el caso en donde cada matriz $\bar{\Gamma}_i(t)$ es diagonal $\forall i = 1, \ldots, N$ es posible calcular el control $\hat{u}$ de forma analítica como se enuncia en el siguiente teorema.

Teorema 4.3 *Para el caso donde las matrices* $\bar{\Gamma}_i(t)$, $\forall i = 1, \ldots, N-1$ *y* $\tilde{\Gamma}_N$ *dadas por (4.9) son diagonales. El control* $\hat{u}$ *que minimiza el criterio de desempeño dado por (4.2) para el sistema descrito por (4.1) está dado por:*

$$\hat{u}(t) = \sum_{k=1}^{N} I_{[t_{i-1},t_i)}(t) \left[\bar{u}^{i*}\right]_{\Omega}^{+}$$

donde, $\bar{u}^{i*}$ *está dado por (4.10) y* $[\bar{u}^{i*}]_{\Omega}^{+}$ *representa la proyección del elemento* $\bar{u}^{i*}$ *sobre el conjunto* $\Omega = conv(\mathcal{Q}_1 \times \ldots \times \mathcal{Q}_m)$ *definida como:*

$$[\bar{u}^{i*}]_{\Omega}^{+} = \min_{\bar{v}\in\Omega} \|\bar{u}^{i*} - \bar{v}\| \tag{4.13}$$

Prueba. Como se mostró anteriormente, la minimización de (4.2) es equivalente a la minimización de las N funciones dadas por (4.8), donde el conjunto de restricciones está descrito por (4.11). Entonces, aplicando el teorema 4.2 a cada $\bar{J}_i$, se tiene:

$$\begin{aligned} \nabla \bar{J}_i(\tilde{u}^{i*}) + \sum_{l=1}^{m} \left[\lambda_l \nabla h_l(\tilde{u}^{i*}) + \mu_l \nabla g_l(\tilde{u}^{i*})\right] = 0 \\ \lambda_l \geq 0,\ \lambda_l h_l(\tilde{u}^{i*}) = 0,\ l = 1, \ldots, m. \\ \mu_l \geq 0,\ \mu_l g_l(\tilde{u}^{i*}) = 0,\ l = 1, \ldots, m. \end{aligned}$$

la cual es equivalente a:

$$\int_{t_{i-1}}^{t_i} \bar{\Delta}_i(t)dt + \left(\int_{t_{i-1}}^{t_i} \bar{\Gamma}_i(t)dt\right) \tilde{u}^{i*} + \sum_{l=1}^{m} (\mu_l - \lambda_l)\, \bar{e}_l = 0$$

donde cada vector $\bar{e}_l$ corresponde al l-ésimo elemento de la base canónica de $\mathbb{R}^m$. Tomando en cuenta que la matriz $\bar{\Gamma}_i(t)$ es diagonal, se tiene:

$$\begin{bmatrix} \bar{\delta}_1 \\ \bar{\delta}_2 \\ \vdots \\ \bar{\delta}_m \end{bmatrix} + \begin{bmatrix} \bar{\gamma}_{11}\tilde{u}_1^{i*} \\ \bar{\gamma}_{22}\tilde{u}_2^{i*} \\ \vdots \\ \bar{\gamma}_{mm}\tilde{u}_m^{i*} \end{bmatrix} + \begin{bmatrix} \mu_1 - \lambda_1 \\ \mu_2 - \lambda_2 \\ \vdots \\ \mu_m - \lambda_m \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{4.14}$$

donde se consideró que:

$$\int_{t_{i-1}}^{t_i} \bar{\Delta}_i(t)dt = \begin{bmatrix} \bar{\delta}_1 \\ \bar{\delta}_2 \\ \vdots \\ \bar{\delta}_m \end{bmatrix},\ \int_{t_{i-1}}^{t_i} \bar{\Gamma}_i(t)dt = \begin{bmatrix} \bar{\gamma}_{11} & 0 & \ldots & 0 \\ 0 & \bar{\gamma}_{22} & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & \bar{\gamma}_{mm} \end{bmatrix}$$

Cada una de las m ecuaciones que aparecen en (4.14) está desacoplada completamente del resto, es decir, cada término de la k-ésima ecuación aparece solamente en dicha ecuación y no vuelve a presentarse en otra. Con esto en mente, el análisis se desarrolla solamente para la k-ésima ecuación sindo el análisis igual para todas. Entonces, tenemos que:

$$\bar{\delta}_k + \bar{\gamma}_{kk}\tilde{u}_k^{i*} + \mu_k - \lambda_k = 0 \tag{4.15}$$

existen 3 casos a tomar en cuenta, los cuales se describen a continuación:

1. $\bar{u}_k^{i*} \in conv(\mathcal{Q}_k)$. En este caso las k-ésimas restricciones ($h_k(\tilde{u}^{i*})$ y $g_k(\tilde{u}^{i*})$) están desactivadas, y a partir de la condiciones de Karush-Kuhn-Tucker se tiene que, $\lambda_k = \mu_k = 0$, lo cual implica:
$$\tilde{u}_k^{i*} = \bar{u}_k^{i*}, \text{ con: } \bar{u}^{i*_k} \in conv(\mathcal{Q})_k$$
con $\bar{u}_k^{i*}$ dado por (4.10).

2. $\bar{u}_k^{i*} < q_0^{(k)}$. En este caso la restricción $h_k(\tilde{u}^{i*})$ está activada (es decir, $\tilde{u}_k^{i*} = q_0^{(k)}$) y la restricción $g_k(\tilde{u}^{i*})$ está desactivada, lo cual implica que $\lambda_k \neq 0$ y $\mu_k = 0$. Despejando λ_k de (4.15) se tiene:
$$\lambda_k = \bar{\delta}_k + \bar{\gamma}_k q_0^{(k)} > \bar{\delta}_k + \bar{\gamma}_k \bar{u}_k^{i*} = 0$$
a partir del teorema 4.2 se tiene que:
$$\tilde{u}_k^{i*} = q_0^{(k)}, \text{ con: } \bar{u}_k^{i*} < q_0^{(k)}$$

3. $\bar{u}_k^{i*} > q_M^{(k)}$. En este caso la restricción $h_k(\tilde{u}^{i*})$ está desactivada y la restricción $g_k(\tilde{u}^{i*})$ está activada (es decir, $\tilde{u}_k^{i*} = q_M^{(k)}$), lo cual implica que $\lambda_k = 0$ y $\mu_k \neq 0$. Despejando μ_k de (4.15) se tiene:
$$\mu_k = -\bar{\delta}_k - \bar{\gamma}_k q_M^{(k)} > -\bar{\delta}_k - \bar{\gamma}_k \bar{u}_k^{i*} = 0$$
a partir del teorema 4.2 se tiene que:
$$\tilde{u}_k^{i*} = q_M^{(k)}, \text{ con: } q_M^{(k)} < \bar{u}_k^{i*}$$

Resumiendo, el k-ésimo componente del control óptimo $\tilde{u}_k^{i*}$ está dado por:

$$\tilde{u}_k^{i*} = \begin{cases} q_0^{(k)}, & \bar{u}_k^{i*} < q_0^{(k)} \\ -\dfrac{\bar{\delta}_k}{\bar{\gamma}_{kk}}, & q_0^{(k)} \leq \bar{u}_k^{i*} \leq q_M^{(k)} \\ q_M^{(k)}, & q_M^{(k)} \leq \bar{u}_k^{i*} \end{cases}$$

o equivalentemente:

$$\tilde{u}_k^{i*} = [\bar{u}^{i*}]^+_{conv(\mathcal{Q}_k)}$$

y realizando el análisis anterior para cada coordenada, se tiene que:

$$\tilde{u}^{i*} = [\bar{u}^{i*}]^+_{\Omega}$$

de donde se obtiene el resultado deseado. ■

4.2. Solución Numérica del Problema Relajado

El teorema 4.3 resuelve el problema con restricciones cuando las matrices $\bar{\Gamma}_i(t)$ son diagonales. El problema general (problema 3.16) no es posible resolverlo de forma analítica, pues se necesitan saber cuales son las restricciones activas, es por eso que se hace uso de métodos numéricos, en concreto se hace uso de métodos de conjuntos activos. Además, el caso general donde se tienen secuencias $\mathcal{T}_k$ fijas y no necesariamente iguales entre ellas, no es posible resolverlo analíticamente.

La idea detrás de los métodos de conjuntos activos consiste en dividir las restricciones de desigualdad en dos grupos: activas e inactivas. Es claro que si en el problema general se conocen *a priori* cuales son las restricciones activas, el problema original puede ser reemplazado por un problema donde solamente aparecen restricciones de igualdad [18].

Es en esta categoría de métodos numéricos donde pertenece el algoritmo de gradiente con proyección (ver [18] para detalles acerca de los métodos de conjuntos activos y como el algoritmo del gradiente pertenece a esta familia).

4.2.1. El algoritmo del gradiente con proyección

En la sección anterior, los teoremas desarrollados recaen en el hecho de que el problema original (de dimensión infinita) se puede descomponer en N problemas de dimensión finita. Aunque este enfoque se puede seguir tomando, el caso de dimensión infinita puede ser resuelto directamente con este enfoque por lo que es el enfoque se que discutirá a lo largo de la sección.

Recordando el problema relajado original, se tiene que encontrar el control $\hat{u}(\cdot) \in conv(\mathcal{S})$ tal que:

$$J\left(\hat{u}(\cdot)\right) = \min_{u(\cdot)\in conv(\mathcal{S})} \left\{J\left(u(\cdot)\right)\right\} \tag{4.16}$$

Para resolver este problema, se propone el uso del método del gradiente con proyección, el cual es enuncia en el siguiente teorema tomado de [15].

Teorema 4.4 *Sea $\mathcal{A}$ un conjunto convexo cerrado y acotado en un espacio de Hilbert $\mathcal{H}$, $J(u)$ una función diferenciable en $\mathcal{A}$, donde $\nabla J(u)$ satisface la condición de Lipschitz con constante M. Se considera la secuencia:*

$$u^{\nu+1}(\cdot) = [u^{\nu}(\cdot) - \alpha_\nu \nabla J(u^{\nu}(\cdot))]^{+}_{\mathcal{A}} \tag{4.17}$$

donde $[\cdot]^{+}_{\mathcal{A}}$ es el operador proyección sobre $\mathcal{A}$, dado por:

$$[x]^{+}_{\mathcal{A}} = arg\min_{z\in\mathcal{A}}(\parallel z - x \parallel)$$

Y α_ν cumple con:

$$0 < \epsilon_1 \leq \alpha_\nu \leq \frac{2}{(M + 2\epsilon_2)}, \quad \epsilon_2 > 0$$

Si además el funcional J es convexo. Entonces, la secuencia (4.17) es convergente a el mínimo $u^*(\cdot)$.

La demostración se puede consultar en [15].

A partir de capítulo anterior se tiene que el conjunto $conv(\mathcal{S})$ es cerrado, acotado y convexo. Además también se mostró que el funcional de costo dado por (4.2) es continuo, de clase $\mathcal{C}^2$ y convexo, y puesto que el espacio $\mathbb{L}_2[t_0, t_f; \mathbb{R}^m]$ es Hilbert, el método de gradiente con proyección es adecuado para resolver el problema relajado descrito por (4.16).

Para la implementación del algoritmo del gradiente con proyección es necesario calcular la derivada del funcional de costo (4.2) con respecto a u, el cálculo de esta derivada puede ser en el sentido de Gateaux como en [13]. Una forma sencilla de obtener tal derivada es mostrada en [4] y es calculada como:

$$\begin{aligned} \nabla J(u) &= -H_u(t, x(t,u), u, p(t,u)) \\ \dot{p}(t,u) &= -H_x(t, x(t,u), u(t), p(t,u)) \\ p(t_f, u) &= -Gx(t_f, u) \\ \dot{x}(t,u) &= H_p(t, x(t,u), u(t), p(t,u)) \\ x(t_0) &= x_0 \end{aligned} \tag{4.18}$$

donde H es el Hamiltoniano del sistema dado por:

$$\begin{aligned} H(t, x(t,u), u(t), p(t,u)) =& \langle p(t), A(t)x(t,u) + B(t)u(t)\rangle - \\ & \frac{1}{2}\left[\langle Q(t)x(t,u), x(t,u)\rangle + \langle R(t)u(t), u(t)\rangle\right] \end{aligned} \tag{4.19}$$

La selección del parámetro α_ν en el teorema 4.4 se puede llevar a cabo de distintas maneras, una forma común de establecer este parámetro es usando la conocida regla de Armijo con proyección (ver [26]), la cual consiste en seleccionar el máximo $\alpha_\nu \in \{1, 1/2, 1/4, \ldots\}$ tal que:

$$J\left([u^\nu(\cdot) - \alpha_\nu \nabla J(u^\nu(\cdot))]^+_{\mathcal{A}}\right) - J(u^\nu(\cdot)) \leq -\frac{\gamma}{\alpha_\nu}\left\| [u^\nu(\cdot) - \alpha_\nu \nabla J(u^\nu(\cdot))]^+_{\mathcal{A}} - u^\nu(\cdot)\right\|^2 \tag{4.20}$$

donde $\gamma \in (0,1)$ constante. Es demostrado en [26] que la selección de α_k usando la regla de Armijo resulta en convergencia al mínimo global en el caso convexo.

Cabe notar que, en algoritmo del gradiente con proyección es necesario calcular la proyección sobre el conjunto de restricciones en cada iteración. En el caso donde el conjunto $\mathcal{A}$ es convexo, tal proyección siempre existe y es única. Además, calcular la proyección consiste en resolver un problema de minimización restringida, por lo que, para que el algoritmo del gradiente con proyección sea implementable, es necesario que la proyección sobre el conjunto de restricciones sea sencilla de calcular.

En el caso del problema relajado descrito por (4.16), la proyección de un elemento $u(\cdot)$ en el conjunto $conv(\mathcal{S})$ está dada por:

$$[u(\cdot)]^+_{conv(\mathcal{S})} = \min_{v(\cdot)\in conv(\mathcal{S})} \|u(\cdot) - v(\cdot)\|_{\mathbb{L}_2[t_0,t_f;\mathbb{R}^m]}$$

donde la norma $\|\cdot\|_{\mathbb{L}_2[t_0,t_f;\mathbb{R}^m]}$, es la norma del espacio $\mathbb{L}_2[t_0,t_f;\mathbb{R}^m]$ dada por:

$$\|u(\cdot)\|_{\mathbb{L}_2[t_0,t_f;\mathbb{R}^m]} = \left(\int_{t_0}^{t_f} u(t)^T u(t) dt\right)^{1/2}$$

Entonces, se tiene que:

$$[u(\cdot)]^+_{conv(\mathcal{S})} = \arg\min_{v(\cdot)\in conv(\mathcal{S})} \left(\int_{t_0}^{t_f} (u(t) - v(t))^T (u(t) - v(t))\, dt\right)^{1/2}$$

De manera equivalente:

$$\begin{aligned}[u(\cdot)]^+_{conv(\mathcal{S})} &= \arg\min_{v(\cdot)\in conv(\mathcal{S})} \left(\int_{t_0}^{t_f} \left(\sum_{i=1}^{m} (u_k(t) - v_k(t))^2\right) dt\right)^{1/2} \\ &= \arg\min_{v(\cdot)\in conv(\mathcal{S})} \left(\sum_{i=1}^{m} \left(\int_{t_0}^{t_f} (u_k(t) - v_k(t))^2\, dt\right)\right)^{1/2}\end{aligned}$$

aplicando el corolario 3.1 del capítulo 3, se tiene que si $v(\cdot) \in conv(\mathcal{S})$, entonces, $v_k(\cdot) \in conv(\mathcal{S}_k)\ \forall k = 1,\ldots,m$. Además, de la última ecuación se puede apreciar que cada componente de la raíz cuadrada es no negativo, por lo tanto, la proyección se puede llevar a cabo sobre cada elemento del vector $u(\cdot)$, es decir:

$$[u(\cdot)]^+_{conv(\mathcal{S})} = \left[[u_1(\cdot)]^+_{conv(\mathcal{S}_1)} \quad ,\ldots, \quad [u_m(\cdot)]^+_{conv(\mathcal{S}_m)}\right]^T \tag{4.21}$$

donde:

$$[u_k]^+_{conv(\mathcal{S}_k)} = \arg\min_{v_k(\cdot)\in conv(\mathcal{S}_k)} \left(\int_{t_0^{(k)}}^{t_f^{(k)}} (u_k(t) - v_k(t))^2\, dt\right),\ t_i^{(k)} \in \mathcal{T}_k$$

Ahora bien, considerando que los elementos de $conv(\mathcal{S}_k)$ son funciones constantes a trozos, se tiene que la proyección de cada u_k es de la forma:

$$[u_k]^+_{conv(\mathcal{S}_k)} = \sum_{i=1}^{N_k} I_{[t_{i-1}^{(k)}, t_i^{(k)})}(t)\zeta_k^{(j_i)},\ \zeta_k^{(j_i)} \in conv(\mathcal{Q}_k).$$

Entonces, para calcular cada $\zeta_k^{(j_i)}$ procedemos como sigue:

$$\begin{aligned}[u_k(\cdot)]^+_{conv(\mathcal{S}_k)} &= \arg \min_{v_k(\cdot)\in conv(\mathcal{S}_k)} \left(\int_{t_0^{(k)}}^{t_f^{(k)}} \left(u_k(t) - v_k(t)\right)^2 dt \right) \\ &= \arg \min_{v_k(\cdot)\in conv(\mathcal{S}_k)} \sum_{i=1}^{N_k} \left(\int_{t_{i-1}^{(k)}}^{t_i^{(k)}} \left(u_k(t) - v_k^{(j_i)}\right)^2 dt \right)\end{aligned}$$

y se tiene que:

$$\begin{aligned}\zeta_k^{(j_i)} &= \arg \min_{v_k^{(j_i)}\in conv(\mathcal{Q}_k)} \left(\int_{t_{i-1}^{(k)}}^{t_i^{(k)}} \left(u_k(t) - v_k^{(j_i)}\right)^2 dt \right) \\ &= \arg \min_{v_k^{(j_i)}\in conv(\mathcal{Q}_k)} \left(\int_{t_{i-1}^{(k)}}^{t_i^{(k)}} \left(-2u_k(t)v_k^{(j_i)} + \left(v_k^{(j_i)}\right)^2\right) dt \right)\end{aligned}$$

Aplicando la condición necesaria de primer orden de optimalidad a la última expresión, se obtiene lo siguiente:

$$-2\zeta_k^{(j_i)} \int_{t_{i-1}^{(k)}}^{t_i^{(k)}} u_k(t)dt + 2\zeta_k^{(j_i)} \int_{t_{i-1}^{(k)}}^{t_i^{(k)}} dt = 0$$

lo cual implica que:

$$\zeta_k^{(j_i)} = \frac{1}{\Delta_{i,k}} \int_{t_{i-1}^{(k)}}^{t_i^{(k)}} u_k(t)dt,\ \Delta_{i,k} = t_i^{(k)} - t_{i-1}^{(k)} > 0$$

siempre y cuando

$$\frac{1}{\Delta_{i,k}} \int_{t_{i-1}^{(k)}}^{t_i^{(k)}} u_k(t)dt \in conv(\mathcal{Q}_k)$$

Por lo tanto, la proyección del k-ésimo componente de $u(\cdot)$ sobre el conjunto $conv(\mathcal{S}_k)$ se obtiene como:

$$\begin{aligned}[u_k(\cdot)]^+_{conv(\mathcal{S}_k)} =& \sum_{i=1}^{N} I_{[t_{i-1},t_i)}(t)\zeta_k^{(j_i)} \\ \zeta_k^{(j_i)} =& \begin{cases} q_0^{(k)},\ \rho_{i,k} < q_0^{(k)} \\ \rho_{i,k},\ q_0^{(k)} \le \rho_{i,k} \le q_{M_k}^{(k)} \\ q_{M_k}^{(k)},\ q_{M_k}^{(k)} < \rho_{i,k} \end{cases} \\ \rho_{i,k} =& \frac{1}{\Delta_{i,k}} \int_{t_{i-1}^{(k)}}^{t_i^{(k)}} u_k(t)dt,\ \Delta_{i,k} = t_i^{(k)} - t_{i-1}^{(k)} > 0.\end{aligned} \tag{4.22}$$

En este caso puede verse como las restricciones sobre los tiempos de conmutación $\mathcal{T}_k$ son consideradas de manera natural al poder descomponer la proyección del vector $u(\cdot)$ como la proyección sobre cada uno de sus componentes.

Todo lo descrito hasta el momento nos lleva a formular el siguiente algoritmo para resolver el problema relajado dado por (4.16).

1. Establecer el valor del índice ν a cero.
2. Proponer la condición inicial $u^0(\cdot)$ como:

$$u^0(\cdot) = [u^{opt}(\cdot)]^+_{conv(\mathcal{S})}$$

   donde $u^{opt}(\cdot)$ representa el control óptimo para el sistema (4.1) sin restricciones (es decir, $u^{opt}(\cdot)$ resuelve el problema LQ clásico). En caso de no contar con el valor de $u^{opt}(\cdot)$ proponer cualquier condición inicial que cumpla con $u^0(\cdot) \in conv(\mathcal{S})$.
3. Calcular la correspondiente trayectoria del sistema $x^\nu(\cdot)$ asociada al control $u^\nu(\cdot)$.
4. Calcular el gradiente del funcional de costo $\nabla J(u^\nu(\cdot))$ a partir de (4.18).
5. Calcular el valor del tamaño de paso α_ν como lo indica la regla de Armijo (4.20).
6. Calcular el elemento:

$$u^{\nu+1}(\cdot) = [u^\nu(\cdot) + \alpha_\nu \nabla J(u^\nu(\cdot))]^+_{conv(\mathcal{S})}$$

   a partir de (4.21) y (4.22).
7. Si $|J(u^\nu(\cdot)) - J(u^{\nu+1}(\cdot))| < \varepsilon$, para algún $\varepsilon > 0$ pre-especificado, entonces, detener el algoritmo. De lo contrario, incrementar el índice ν en uno y regresar al paso 3.

Al finalizar el algoritmo se obtiene una aproximación al control óptimo $\hat{u}(\cdot)$ para el problema relajado.

4.3. Ejemplos

En esta sección se presentan un par de ejemplos numéricos sobre la utilización del algoritmo del gradiente con proyección para resolver el problema relajado. El ejemplo en concreto trata sobre el control de la trayectoria de un satélite, el cual se obtuvo de [23]. Debido a que la dinámica asociada a tal sistema es no lineal, se hace la linealización alrededor de una trayectoria nominal para posteriormente aplicar el algoritmo propuesto.

Como un primer caso, se considera una trayectoria nominal circular, resultando en un sistema lineal invariante en el tiempo, y en el segundo caso la trayectoria nominal corresponde a una elipse, resultando en un sistema lineal variante en el tiempo. Se considera que los actuadores del satélite se encuentran alineados en la dirección normal y tangencial de movimiento, proporcionando una cantidad constante de aceleración en la dirección correspondiente ante una entrada $u(\cdot)$.

4.3.1. Satélite

Un satélite de masa unitaria puede ser modelado como una masa puntual moviéndose en un plano mientras es atraído al centro del plano por la ley de fuerza cuadrática inversa. Para este caso es conveniente el uso de coordenadas polares, donde: $r(t)$ representa la distancia medida desde el origen a la masa y $\theta(t)$ representa el ángulo medido a partir de un eje de referencia dado. Se asume que al satélite se la aplica una fuerza $u_1(t)$, (la cual actúa en la dirección radial del movimiento) y una fuerza $u_2(t)$ (actuando en la dirección tangencial) como se muestra en la figura 4.1.

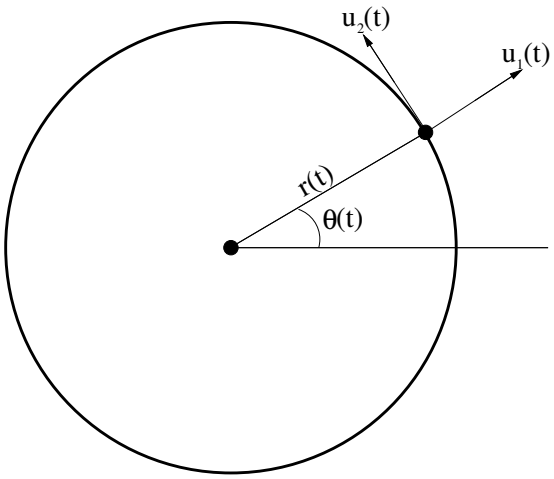

Figura 4.1: Un satélite en órbita circular

Las ecuaciones de movimiento tienen la forma:

$$\begin{aligned} \ddot{r}(t) =& r(t)\dot{\theta}^2(t) - \frac{\beta}{r^2(t)} + u_1(t) \\ \ddot{\theta}(t) =& \frac{-2\dot{r}(t)\dot{\theta}(t)}{r(t)} + \frac{u_2(t)}{r(t)} \end{aligned} \tag{4.23}$$

donde β es una constante. Para construir una representación en espacio estado de las ecuaciones de movimiento, sea:

$$x_1(t) = r(t),\ x_2(t) = \dot{r}(t),\ x_3(t) = \theta(t),\ x_4(t) = \dot{\theta}(t)$$

entonces, las ecuaciones toman la forma:

$$\dot{x}(t) = \begin{bmatrix} \dot{x}_1(t) \\ \dot{x}_2(t) \\ \dot{x}_3(t) \\ \dot{x}_4(t) \end{bmatrix} = F(x(t), u(t)) = \begin{bmatrix} x_2(t) \\ x_1(t)x_4^2(t) - \dfrac{\beta}{x_1^2(t)} + u_1(t) \\ x_4(t) \\ \dfrac{-2x_2(t)x_4(t)}{x_1(t)} + \dfrac{u_2(t)}{x_1(t)} \end{bmatrix} \tag{4.24}$$

Para el primer caso se considera la órbita más simple, la cual consiste en una circunferencia donde $r(t)$ y $\dot{\theta}(t)$ permanecen constantes. Por lo que la trayectoria nominal está dada por:

$$\tilde{r}(t) = r_0,\ \tilde{\theta}(t) = \omega_0 t + \theta_0,\ \tilde{u}_1(t) = \tilde{u}_2(t) = 0$$

o equivalentemente, en representación matricial:

$$\tilde{x}(t) = \begin{bmatrix} \tilde{x}_1(t) \\ \tilde{x}_2(t) \\ \tilde{x}_3(t) \\ \tilde{x}_4(t) \end{bmatrix} = \begin{bmatrix} r_0 \\ 0 \\ \omega_0 t + \theta_0 \\ \omega_0 \end{bmatrix},\ \tilde{u}(t) = \begin{bmatrix} \tilde{u}_1(t) \\ \tilde{u}_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{4.25}$$

Linealizando el sistema (4.24) alrededor de la trayectoria nominal (4.25) se obtiene:

$$\begin{aligned} A &= \left.\frac{\partial F}{\partial x}\right|_{\substack{x=\tilde{x} \\ u=\tilde{u}}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ x_4^2 + 2\dfrac{\beta}{x_1^3} & 0 & 0 & 2x_1x_4 \\ 0 & 0 & 0 & 1 \\ \dfrac{2x_2x_4}{x_1^2} - \dfrac{u_2}{x_1^2} & -\dfrac{2x_4}{x_1} & 0 & \dfrac{-2x_2}{x_1} \end{bmatrix}_{\substack{x=\tilde{x} \\ u=\tilde{u}}} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 3\omega_0^2 & 0 & 0 & 2r_0\omega_0 \\ 0 & 0 & 0 & 1 \\ 0 & \dfrac{-2\omega_0}{r_0} & 0 & 0 \end{bmatrix} \\ B &= \left.\frac{\partial F}{\partial u}\right|_{\substack{x=\tilde{x} \\ u=\tilde{u}}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & \dfrac{1}{x_1} \end{bmatrix}_{\substack{x=\tilde{x} \\ u=\tilde{u}}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & \dfrac{1}{r_0} \end{bmatrix} \end{aligned} \tag{4.26}$$

Por lo tanto, el sistema:

$$\dot{z}(t) = Az(t) + Bu(t)$$

donde A y B están dadas por (4.26), representa la linealización del sistema (4.24) alrededor de la trayectoria nominal dada por (4.25).

Ahora bien, para este caso se consideran los siguientes valores para los parámetros ω_0 y r_0:

$$\omega_0 = 0.1\ [rad/s],\ r_0 = 200\ [m]$$

Por otra parte, se asume que los conjuntos $\mathcal{Q}_k$ de restricciones sobre los niveles de $u_k(t)$ están dados por:

$$\begin{aligned}\mathcal{Q}_1 =&\{q_1^j \in \mathbb{R} | q_1^j = j/10, j \in [-50, 50] \cap \mathbb{Z}\}\\ \mathcal{Q}_2 =&\{q_2^j \in \mathbb{R} | q_2^j = j/100, j \in [-100, 100] \cap \mathbb{Z}\}\end{aligned}$$

Con la siguiente secuencia de tiempos finita:

$$\begin{aligned}\mathcal{T}_1 = \mathcal{T}_2 = \mathcal{T} :=\{t \in \mathbb{R} | 0 = t_0 < t_1 < \ldots < t_N = 50,\\ t_i - t_{i-1} = 1\ \forall i = 1, \ldots, N\}\end{aligned}$$

Es decir, el primer componente de $u(t) \in \mathcal{S}$ solo puede tomar niveles constantes desde -5 hasta 5 con incrementos de 0.1 unidades, mientras que el segundo componente solo puede tomar niveles constantes desde -1 hasta 1 con incrementos de 0.05 unidades. Además, cada componente permanece constante en un intervalo de tiempo de longitud 1, desde $t_0 = 0s$. hasta $t_N = 50s$.

Se busca minimizar el siguiente funcional de costo:

$$J(u) = \frac{1}{2}\int_{t_0}^{t_N} \left(x^T(t)Qx(t) + u^T(t)Ru(t)\right) dt$$

donde:

$$Q = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix},\ R = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

Para el caso sin restricciones, es decir, se trata de un problema LQ clásico, el control óptimo denotado por $u^{opt}(t)$ está dado por:

$$u^{opt}(t) = -R^{-1}B^T P x^{opt}(t)$$

donde P es la solución a la ecuación de Ricatti asociada al sistema. En esta caso (sin restricciones) el costo óptimo genera un valor de 149.3759 unidades.

Al aplicar el algoritmo del gradiente con proyección al problema relajado, después de 4 iteraciones se obtienen las trayectorias $\hat{x}(t)$ del sistema, mostradas en las figuras 4.2 - 4.5 y la

señal de control $\hat{u}(t)$, la cual se muestra en las figuras 4.6 - 4.7 y el costo asociado al control $\hat{u}(\cdot)$ es de aproximadamente 149.7898 unidades.

El costo asociado a $\hat{u}(t)$ es un 0.2771 % mayor comparado con el costo obtenido con el control $u^{opt}(t)$.

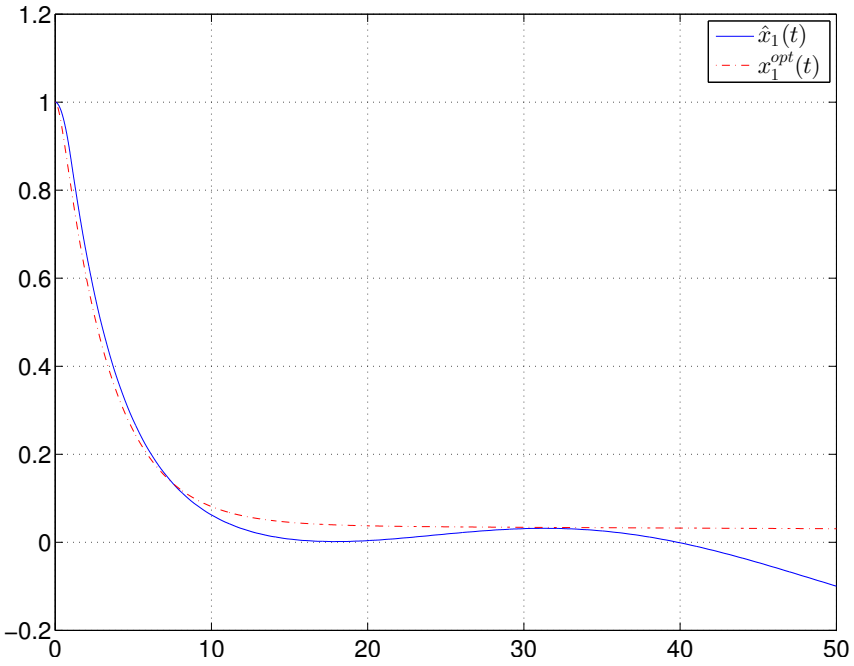

Figura 4.2: $\hat{x}_1(t)$ vs. $x_1^{opt}(t)$

Para el segundo caso, se tiene que la trayectoria nominal representa una elipse en $\mathbb{R}^2$ descrita por:

$$E = \{x \in \mathbb{R}^2 : x^T P x = 1\}$$

donde P es una matriz simétrica positiva definida. En coordenadas polares se tiene la siguiente representación:

$$\tilde{r}(\theta) = \frac{1}{\sqrt{p_{11}\cos^2\theta + 2p_{12}\cos\theta\sin\theta + p_{22}\sin^2\theta}}$$

como en el caso anterior se busca velocidad angular constante, es decir: $\dot{\tilde{\theta}} = \omega_0$ constante, lo cual implica $\tilde{\theta} = \omega_0 t + \theta_0$. Entonces la representación matricial de la trayectoria nominal

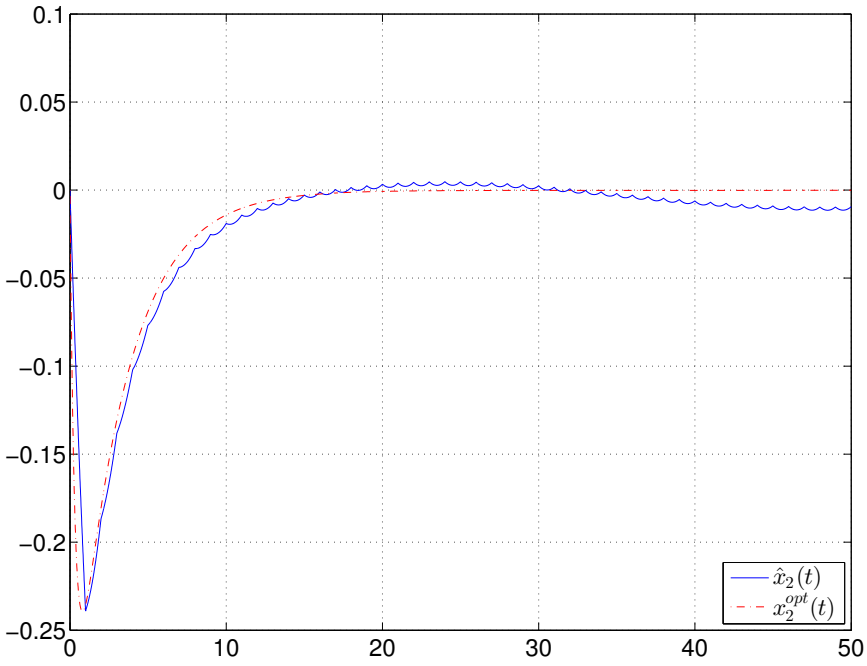

Figura 4.3: $\hat{x}_2(t)$ vs. $x_2^{opt}(t)$

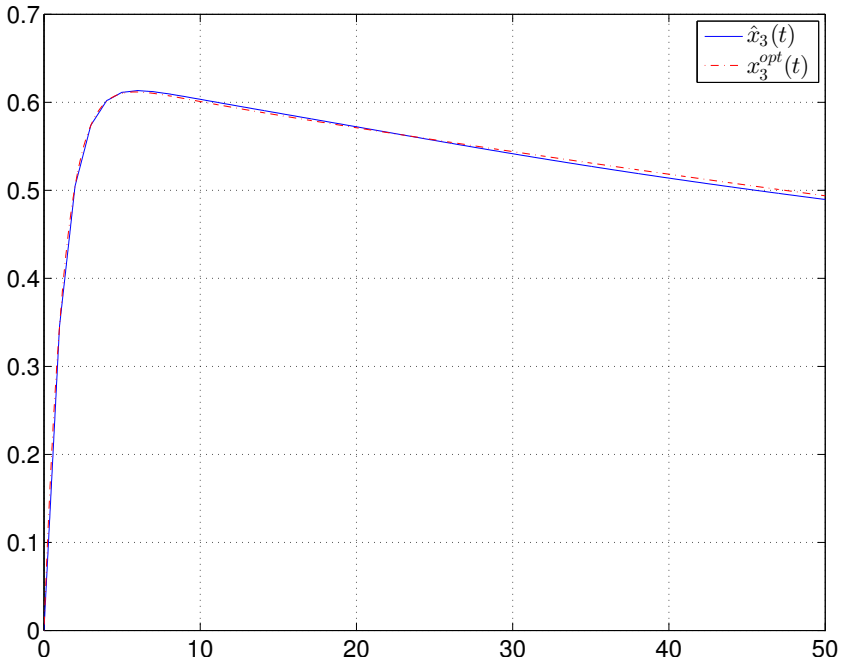

Figura 4.4: $\hat{x}_3(t)$ vs. $x_3^{opt}(t)$

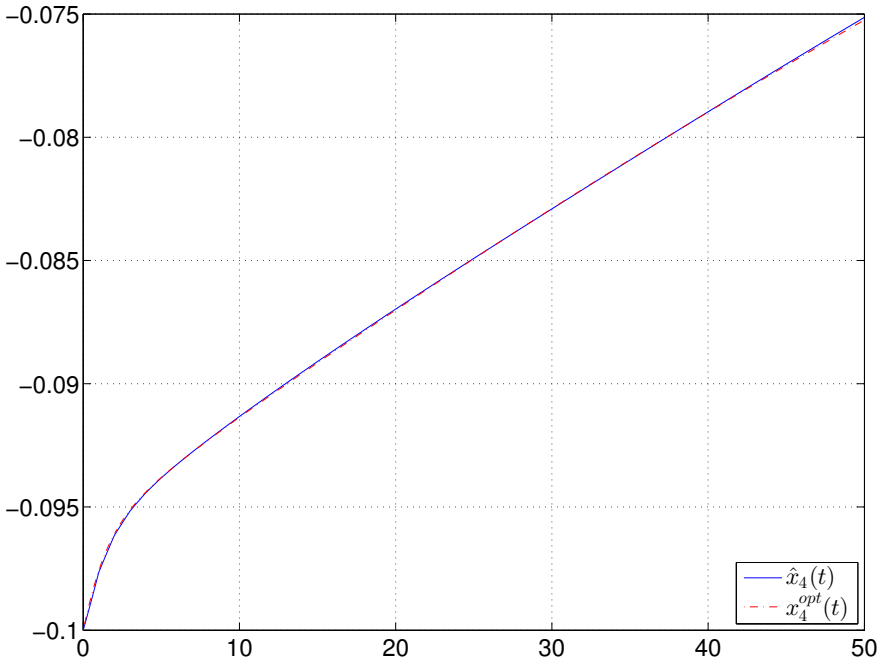

Figura 4.5: $\hat{x}_4(t)$ vs. $x_4^{opt}(t)$

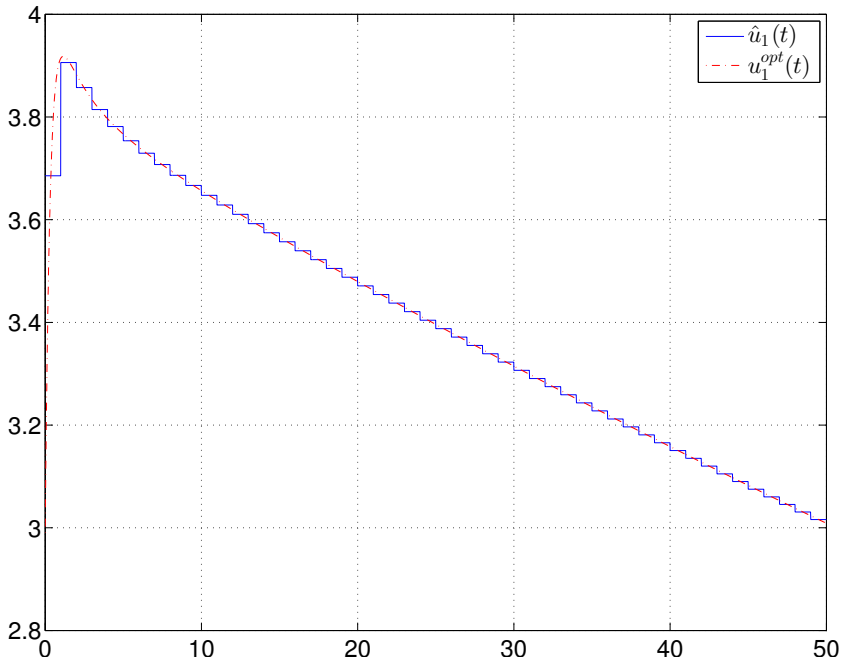

Figura 4.6: $\hat{u}_1(t)$ vs. $u_1^{opt}(t)$

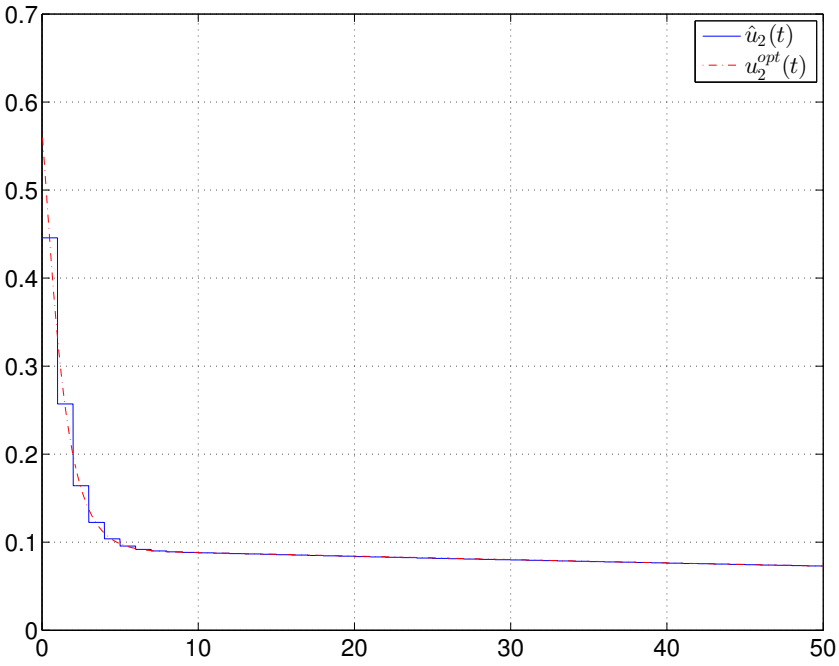

Figura 4.7: $\hat{u}_2(t)$ vs. $u_2^{opt}(t)$

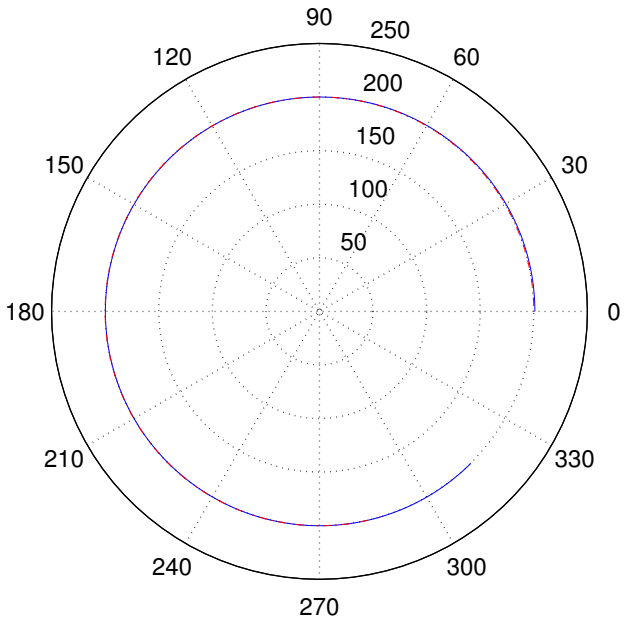

Figura 4.8: Trayectoria del satélite ($\hat{x}_1$ vs $\hat{x}_3$) al aplicar el control $\hat{u}(t)$

está dada por:

$$\tilde{x}(t) = \begin{bmatrix} \tilde{x}_1(t) \\ \tilde{x}_2(t) \\ \tilde{x}_3(t) \\ \tilde{x}_4(t) \end{bmatrix} = \begin{bmatrix} \dfrac{1}{\sqrt{\xi(t)}} \\ \omega_0 \dfrac{\eta(t)}{\xi(t)^{3/2}} \\ \omega_0 t + \theta_0 \\ \omega_0 \end{bmatrix}, \ \tilde{u}(t) = \begin{bmatrix} \tilde{u}_1(t) \\ \tilde{u}_2(t) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{4.27}$$

donde:

$$\xi(t) = p_{11} cos^2(\omega_0 t + \theta_0) + 2p_{12} \cos(\omega_0 t + \theta_0) \sin(\omega_0 t + \theta_0) + p_{22} \sin^2(\omega_0 t + \theta_0)$$
$$\eta(t) = p_{12} \left(\sin^2(\omega_0 t + \theta_0) - \cos^2(\omega_0 t + \theta_0) \right) + (p_{11} - p_{22}) \cos(\omega_0 t + \theta_0) \sin(\omega_0 t + \theta_0)$$

Linealizando el sistema (4.24) alrededor de la trayectoria nominal descrita por (4.27) se obtienen la siguientes matrices $A(t)$ y $B(t)$ como sigue:

$$\begin{aligned} A(t) = \left. \frac{\partial F}{\partial x} \right|_{\substack{x=\tilde{x}(t) \\ u=\tilde{u}(t)}} &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ x_4^2 + 2\dfrac{\beta}{x_1^3} & 0 & 0 & 2x_1 x_4 \\ 0 & 0 & 0 & 1 \\ \dfrac{2x_2 x_4}{x_1^2} - \dfrac{u_2}{x_1^2} & -\dfrac{2x_4}{x_1} & 0 & \dfrac{-2x_2}{x_1} \end{bmatrix}_{\substack{x=\tilde{x} \\ u=\tilde{u}}} \\ &= \begin{bmatrix} 0 & 1 & 0 & 0 \\ \omega_0^2 + 2\beta\xi(t)^{3/2} & 0 & 0 & \dfrac{2\omega_0}{\sqrt{\xi(t)}} \\ 0 & 0 & 0 & 1 \\ \dfrac{2\omega_0^2 \eta(t)}{\xi(t)^{3/2}} & -2\omega_0\sqrt{\xi(t)} & 0 & \dfrac{-2\omega_0 \eta(t)}{\xi(t)} \end{bmatrix} \\ B(t) = \left. \frac{\partial F}{\partial u} \right|_{\substack{x=\tilde{x}(t) \\ u=\tilde{u}(t)}} &= \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & \dfrac{1}{x_1} \end{bmatrix}_{\substack{x=\tilde{x}(t) \\ u=\tilde{u}(t)}} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & \sqrt{\xi(t)} \end{bmatrix} \end{aligned} \tag{4.28}$$

En este caso, se consideran los siguientes conjuntos de restricciones:

$$\mathcal{Q}_1 = \{q_1^j \in \mathbb{R} | q_1^j = j/10, j \in [-10, 10] \cap \mathbb{Z}\}$$
$$\mathcal{Q}_2 = \{q_2^j \in \mathbb{R} | q_2^j = j/200, j \in [-200, 200] \cap \mathbb{Z}\}$$

Con la siguiente secuencia de tiempos finita:

$$\mathcal{T}_1 = \mathcal{T}_2 = \mathcal{T} := \{t \in \mathbb{R} | 0 = t_0 < t_1 < \ldots < t_N = 50, \\ t_i - t_{i-1} = 1 \; \forall i = 1, \ldots, N\}$$

De manera similar, como en el ejemplo anterior se busca minimizar un criterio de costo cuadrático dado por:

$$J(u(\cdot)) = \frac{1}{2} \int_{t_0}^{t_f} \left(x(t)^T Q x(t) + u(t)^T R u(t) \right) dt$$

donde:

$$Q = \begin{bmatrix} 0.5 & 0 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 0.3 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \; R = \begin{bmatrix} 0.5 & 0 \\ 0 & 0.5 \end{bmatrix}$$

La solución al problema LQ clásico (sin restricciones) se obtiene como:

$$u^{opt}(t) = -R^{-1} B(t)^T P(t) x^{opt}(t)$$

donde $P(t)$ es la solución a la ecuación diferencial de Ricatti asociada al sistema. En esta caso el control $u^{opt}(t)$ genera un costo de 1.7795 unidades. Después de aplicar el algoritmo del gradiente con proyección se obtienen después de 3 iteraciones el control $\hat{u}(t)$ y las correspondientes las trayectorias del sistema $\hat{x}(t)$. En las figuras 4.9 - 4.12 se muestran la trayectorias $\hat{x}(t)$ del sistema y en las figuras 4.13 - 4.14 se muestran los componentes del control $\hat{u}(t)$.

El costo generado por aplicar el control $\hat{u}(\cdot)$ al sistema es de aproximadamente 3.1202 unidades, lo cual representa un incremento del 75.3414 % con respecto al costo generado por $u^{opt}(\cdot)$.

Esta diferencial porcentual es mayor comparada con la diferencia porcentual del caso anterior, esto se debe principalmente a que la dinámica del sistema variante en el tiempo es más rápida comparada con la velocidad de respuesta del control $\hat{u}(\cdot)$. Por ejemplo, si se aumenta la taza de conmutación de la señal de control de 1 s a 0.5 s. Es decir, el nuevo conjunto de tiempos de conmutación está dado por:

$$\mathcal{T}_1 = \mathcal{T}_2 = \mathcal{T} := \{t \in \mathbb{R} | 0 = t_0 < t_1 < \ldots < t_N = 50, \\ t_i - t_{i-1} = 0.5 \; \forall i = 1, \ldots, N\}$$

Entonces, se obtiene una mejora en el rendimiento general del sistema, como se muestra en las figuras 4.16 - 4.22. En este caso, el control calculado por el algoritmo del gradiente con proyección (denotado por $\hat{u}^{0.5s}(\cdot)$) genera un costo de 2.0783 unidades el cual representa un incremento del 16.7912 % con respecto al costo generado por $u^{opt}(\cdot)$.

Hasta este punto se tiene solución numérica para el problema relajado en general, haciendo uso del método del gradiente con proyección expuesto en la sub-sección anterior. El problema relajado fue relativamente sencillo de resolver debido a una propiedad clave del conjunto de restricciones a saber la convexidad.

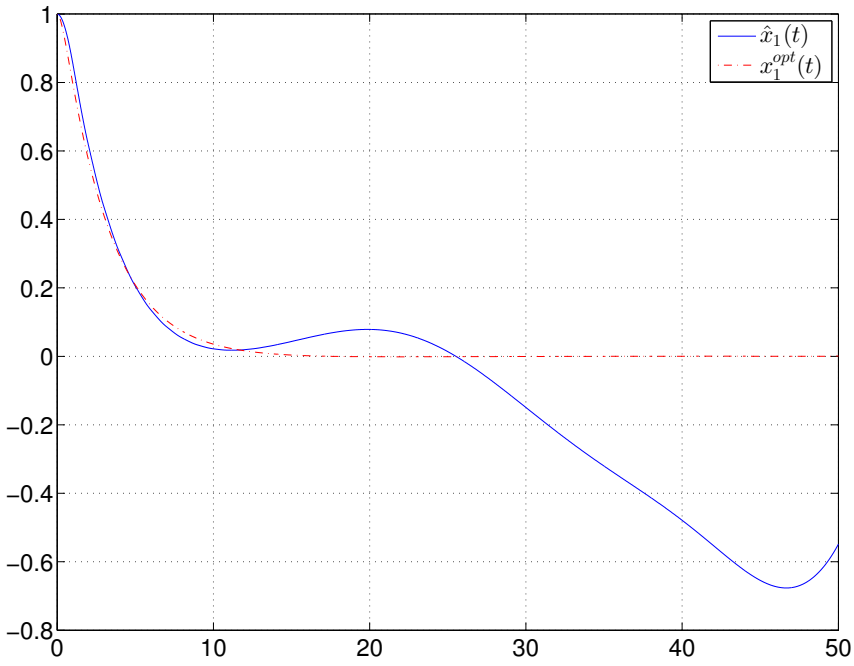

Figura 4.9: $\hat{x}_1(t)$ vs. $x_1^{opt}(t)$

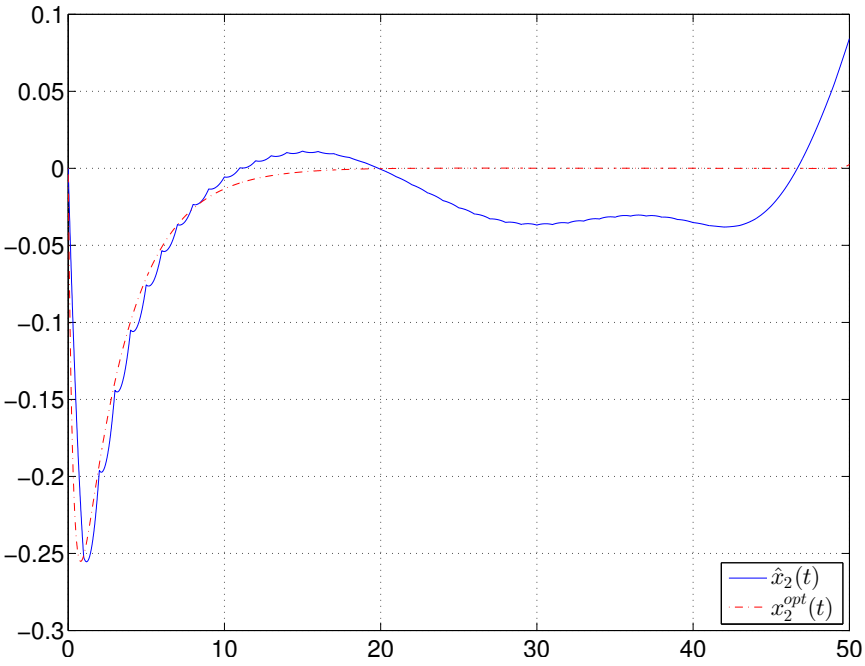

Figura 4.10: $\hat{x}_2(t)$ vs. $x_2^{opt}(t)$

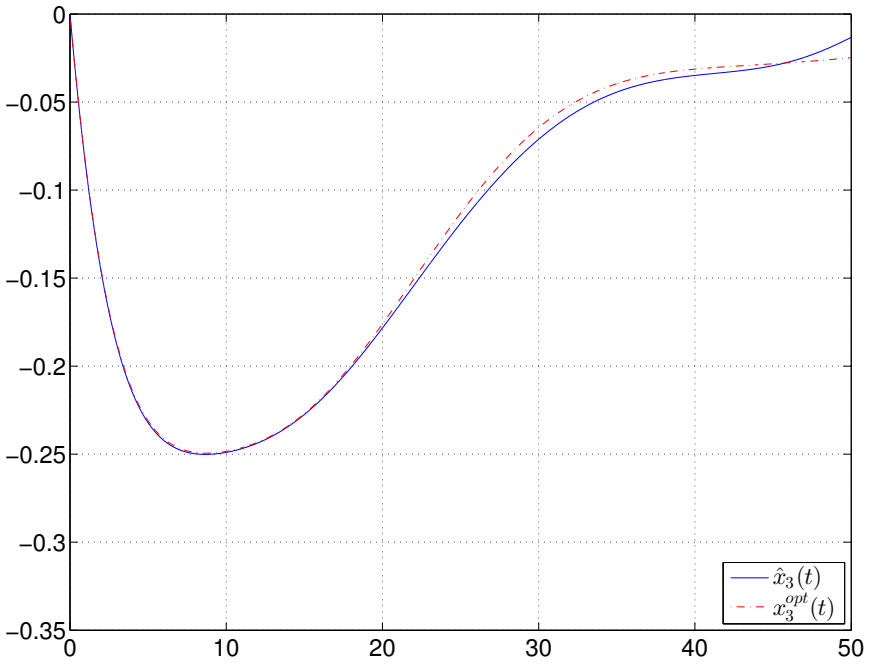

Figura 4.11: $\hat{x}_3(t)$ vs. $x_3^{opt}(t)$

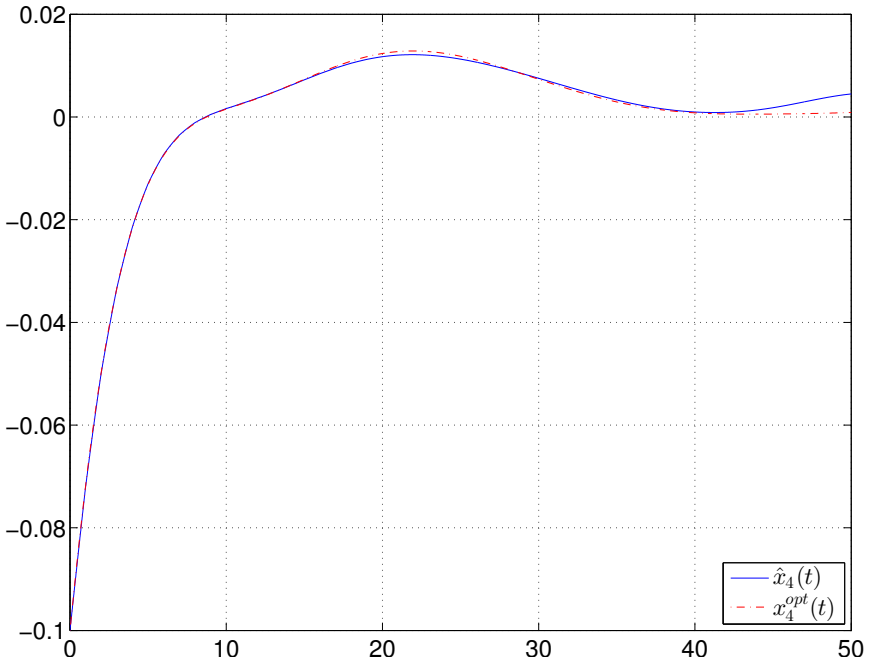

Figura 4.12: $\hat{x}_4(t)$ vs. $x_4^{opt}(t)$

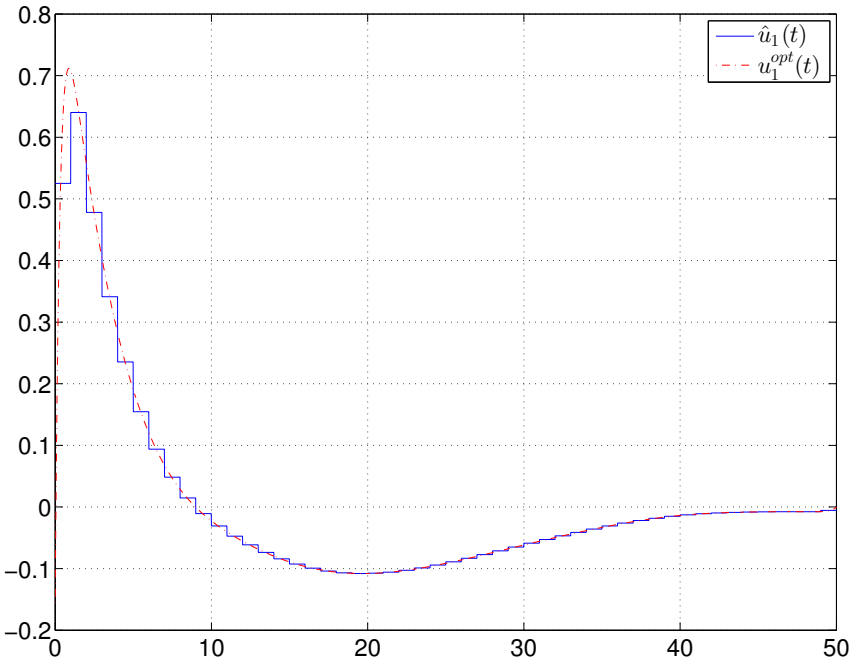

Figura 4.13: $\hat{u}_1(t)$ vs. $u_1^{opt}(t)$

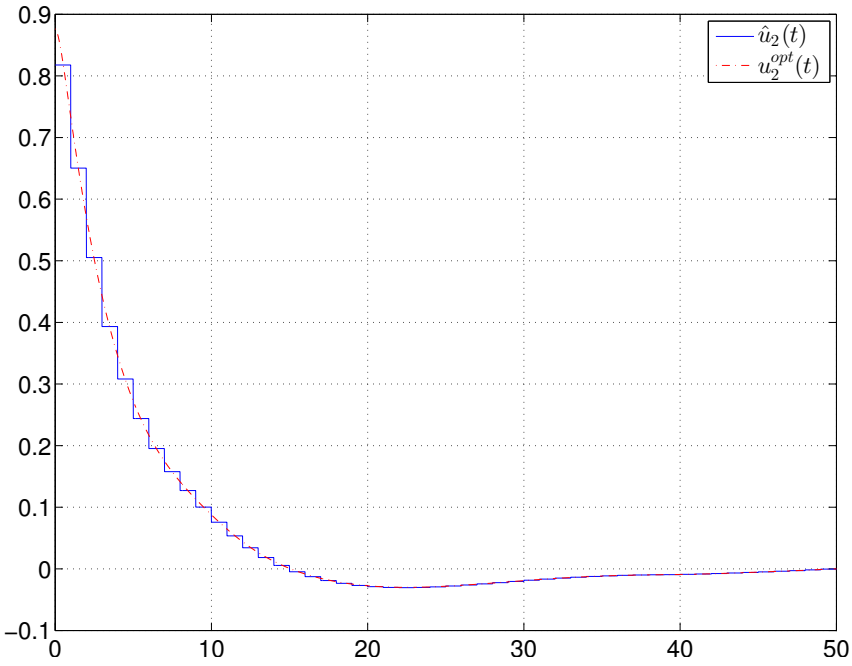

Figura 4.14: $\hat{u}_2(t)$ vs. $u_2^{opt}(t)$

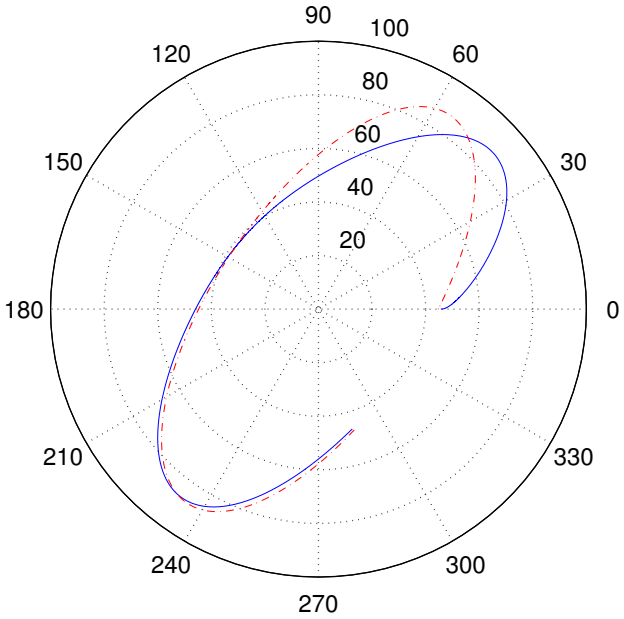

Figura 4.15: Trayectoria deseada vs. Trayectoria real del satélite al aplicar el control $\hat{u}(t)$

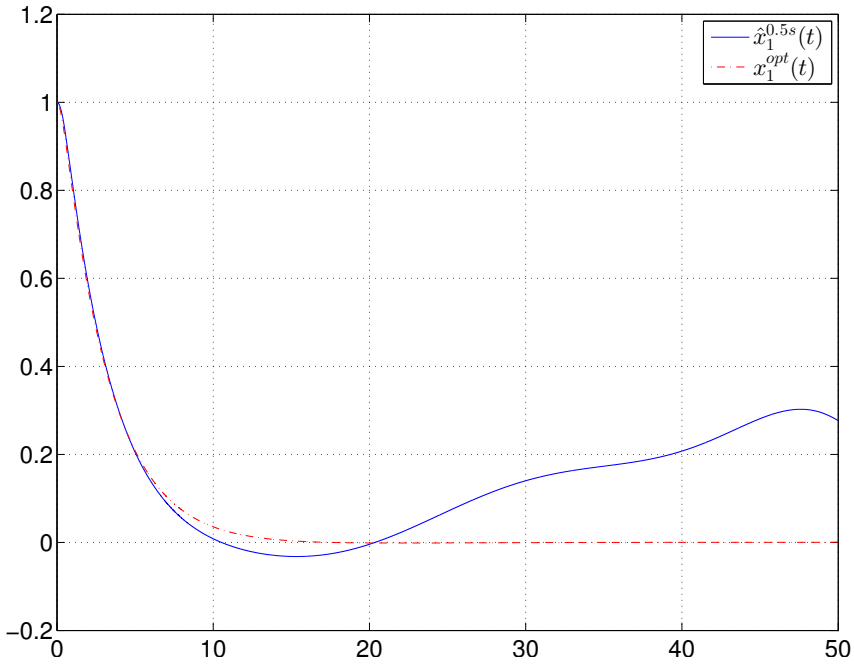

Figura 4.16: $\hat{x}_1^{0.5s}(t)$ vs. $x_1^{opt}(t)$

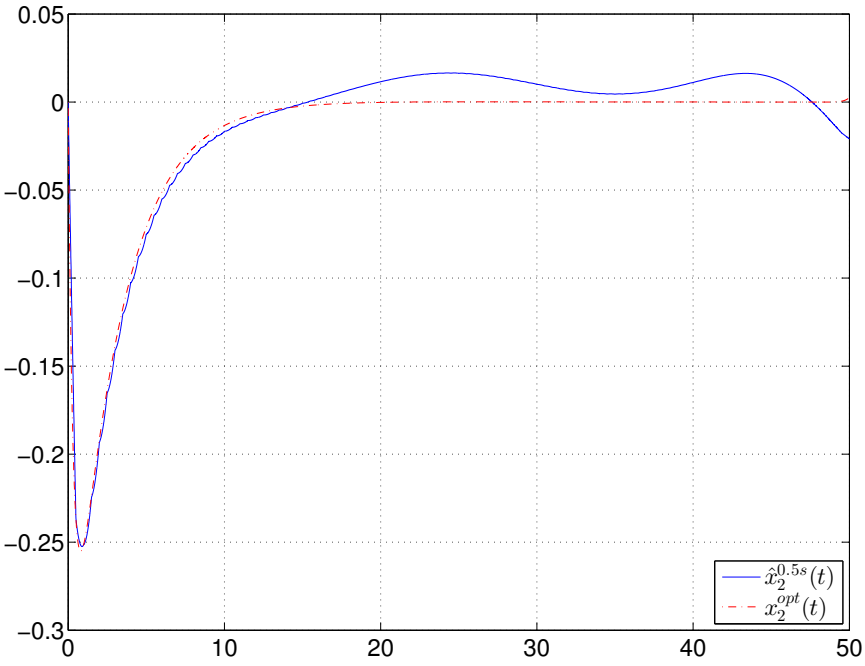

Figura 4.17: $\hat{x}_2^{0.5s}(t)$ vs. $x_2^{opt}(t)$

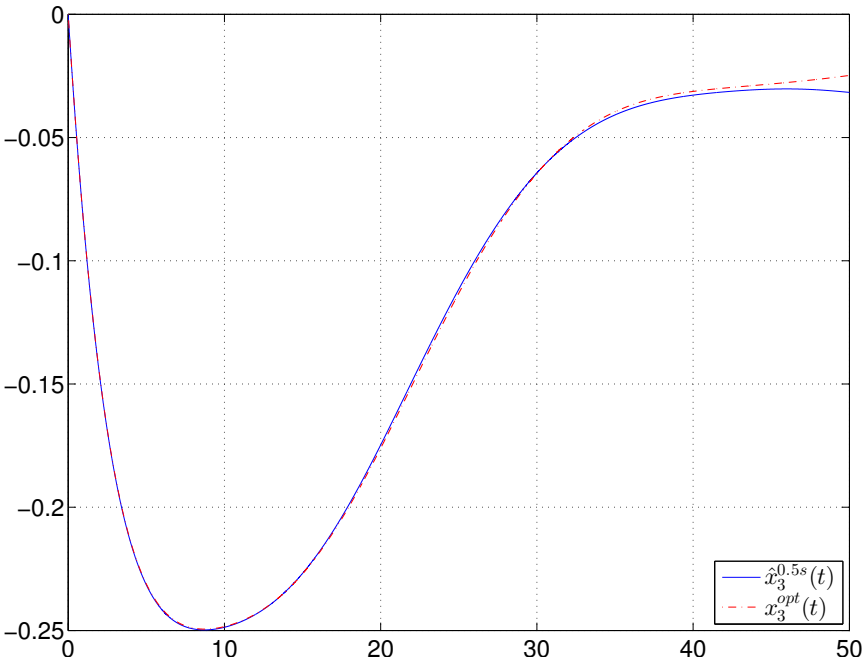

Figura 4.18: $\hat{x}_3^{0.5s}(t)$ vs. $x_3^{opt}(t)$

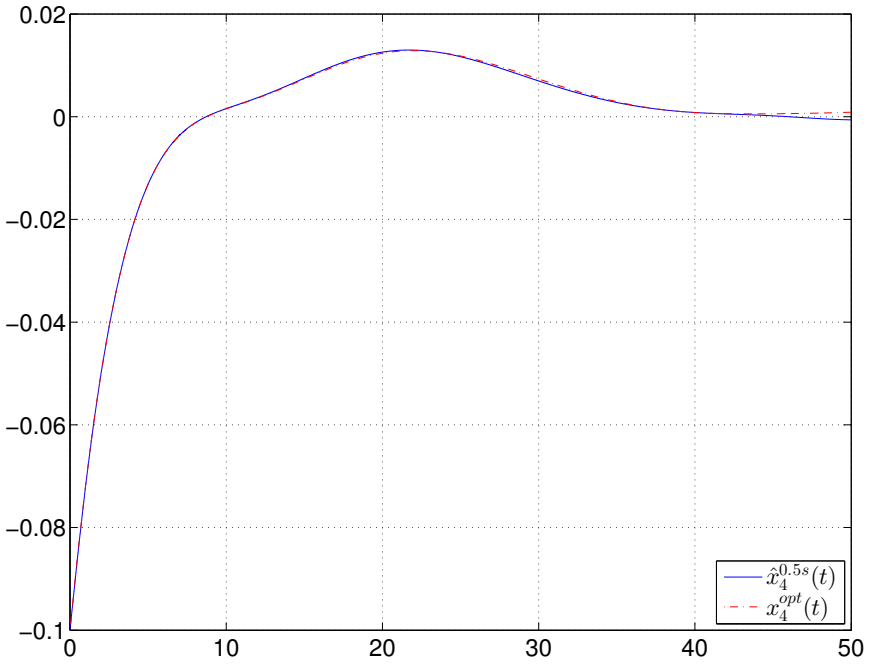

Figura 4.19: $\hat{x}_4^{0.5s}(t)$ vs. $x_4^{opt}(t)$

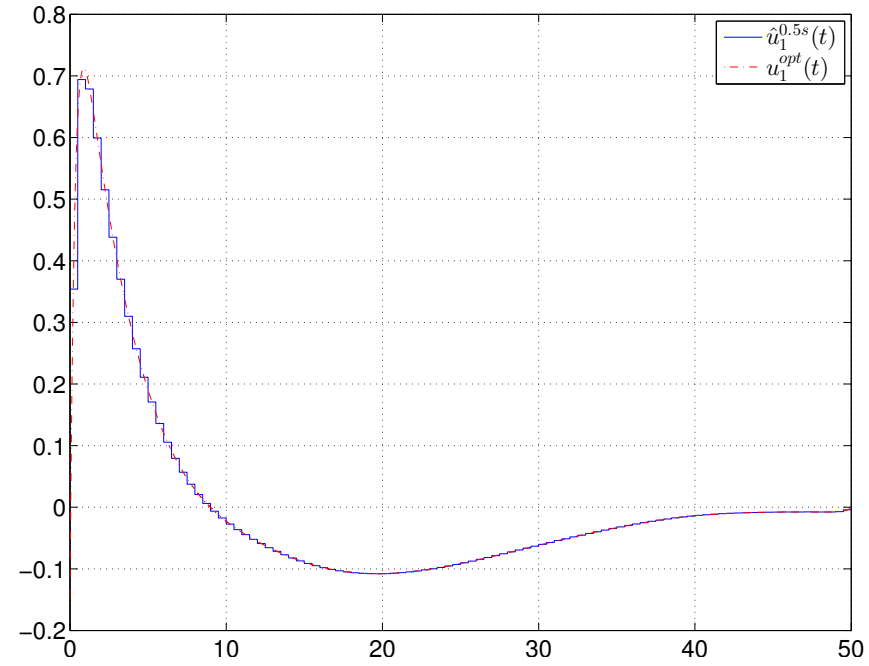

Figura 4.20: $\hat{u}_1^{0.5s}(t)$ vs. $u_1^{opt}(t)$

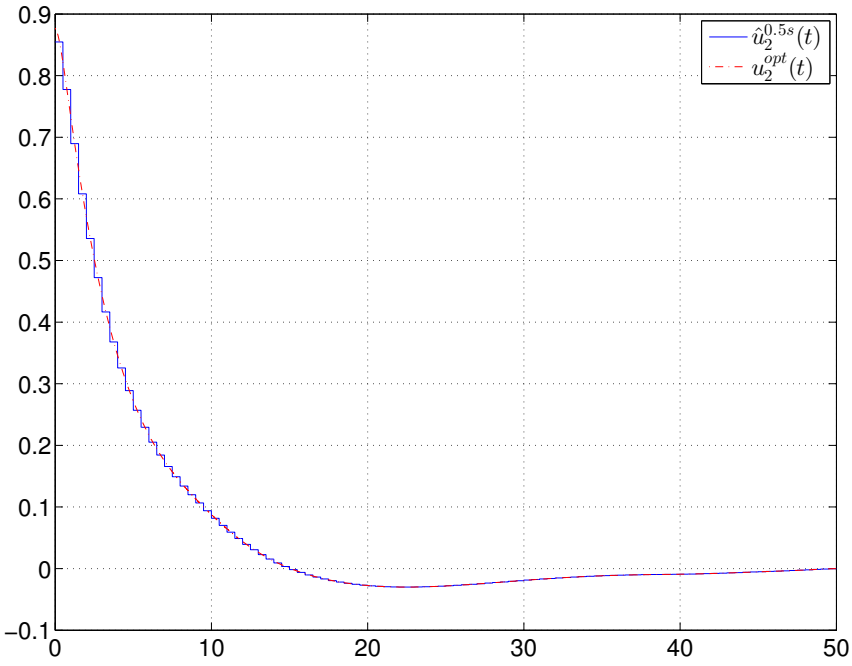

Figura 4.21: $\hat{u}_2^{0.5s}(t)$ vs. $u_2^{opt}(t)$

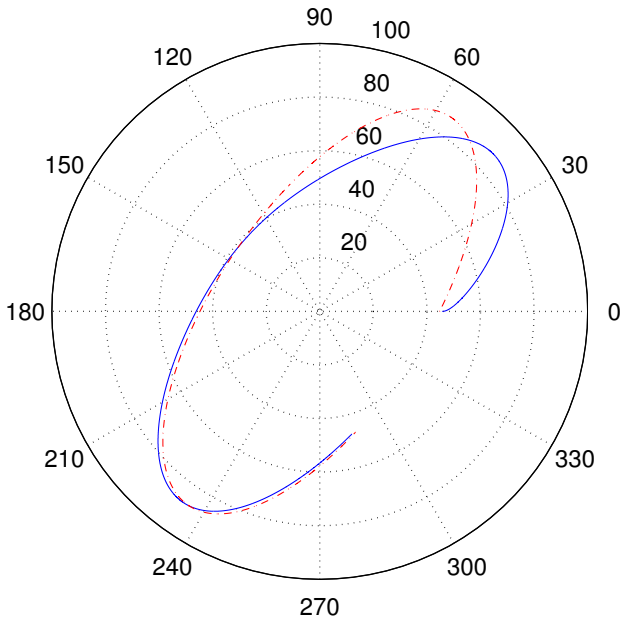

Figura 4.22: Trayectoria deseada vs. Trayectoria real del satélite al aplicar el control $\hat{u}^{0.5s}(t)$

4.4. Solución del Problema Original

Recordando el problema original planteado en el capítulo anterior, se tiene:

Problema 4.1 *Encontrar el control $u^*(\cdot) \in \mathcal{S}$ tal que:*

$$J\left(u^*(\cdot)\right) = \min_{u(\cdot)\in\mathcal{S}} \{J\left(u(\cdot)\right)\} \tag{4.29}$$

donde el conjunto de restricciones $\mathcal{S}$ está dado por:

$$\begin{aligned}\mathcal{S} =& \mathcal{S}_1 \times \ldots \times \mathcal{S}_m \\ =& \{u(t) = [u_1(t), \ldots, u_m(t)]^T \in \mathbb{R}^m : u_k(\cdot) \in \mathcal{S}_k, k = 1, \ldots, m\}\end{aligned} \tag{4.30}$$

Con cada conjunto $\mathcal{S}_k$ definido como:

$$\begin{aligned}\mathcal{S}_k :=& \{v : [t_0, t_f] \to \mathbb{R} | v(t) = \sum_{i=1}^{N_k} I_{[t_{i-1}^{(k)}, t_i^{(k)})}(t) q_k^{(j_i)}; \quad q_k^{(j_i)} \in \mathcal{Q}_k; \\ & j_i \in \mathbb{Z} \cap [1, M_k]; t_i^{(k)} \in \mathcal{T}_k \quad \forall i = 0, \ldots, N_k\}.\end{aligned}$$

Puesto que el conjunto $\mathcal{S}$ no es convexo, entonces no se puede aplicar directamente el enfoque anterior. Más aún, el conjunto $\mathcal{S}$ está completamente contenido en su complementación convexa $conv(\mathcal{S})$, por lo cual surge la idea intuitiva de poder obtener la solución del problema original a partir de la solución del problema relajado.

A partir de la continuidad del funcional, se sigue que, la ley de control perteneciente a $\mathcal{S}$ lo *suficientemente cercana* a el control $\hat{u}(\cdot)$ producirá un costo pequeño comparada con cualquier otro control en $\mathcal{S}$. Por supuesto, esta es solo una idea intuitiva, la cual, como veremos más adelante no necesariamente se cumple.

Basados en esta idea intuitiva, se puede calcular el elemento $\bar{u}^*(\cdot) \in \mathcal{S}$ más *cercano* al control $\hat{u}(\cdot)$ como:

$$\bar{u}^*(\cdot) = \arg \min_{v(\cdot)\in\mathcal{S}} \|\hat{u}(\cdot) - v(\cdot)\| \tag{4.31}$$

Cabe mencionar que debido a que el conjunto $\mathcal{S}$ no es convexo, el elemento $\bar{u}^*(\cdot)$ no necesariamente es único, su existencia queda garantizada debido a que $\mathcal{S}$ es un conjunto finito. Además, debido a la naturaleza constante a trozos de las señales involucradas, el problema de optimización restringida motivado para encontrar $\bar{u}^*(\cdot)$ tiene una solución bastante sencilla. Primero, recordando que el control $\hat{u}(\cdot)$ tiene la siguiente forma:

$$\begin{aligned}\hat{u}(\cdot) &= \begin{bmatrix}\hat{u}_1(\cdot) & \ldots & \hat{u}_m(\cdot)\end{bmatrix} \\ \hat{u}_k(\cdot) &= \sum_{i=1}^{N_k} I_{[t_{i-1}^k, t_i^k)}(t) \zeta_k^{(j_i)}, \ \zeta_k^{(j_i)} \in conv(\mathcal{Q}_k), \ t_i \in \mathcal{T}_k.\end{aligned}$$

Entonces, el control $\bar{u}^*(\cdot)$ más cercano a $\hat{u}(\cdot)$ está dado por:

$$\begin{aligned}
\bar{u}^*(\cdot) &= \begin{bmatrix} \bar{u}_1^*(\cdot) & \dots & \bar{u}_m^*(\cdot) \end{bmatrix} \\
\bar{u}_k^*(t) &= \sum_{i=1}^{N_k} I_{[t_{i-1}^k, t_i^k)}(t) \bar{\zeta}_k^{(j_i)},\ j_i \in [1, M_k] \cap \mathbb{Z},\ \bar{\zeta}_k^{(j_i)} \in \mathcal{Q}_k,\ t_i \in \mathcal{T}_k. \\
\bar{\zeta}_k^{(j_i)} &= \arg \min_{q_k \in \mathcal{Q}_k} |\bar{\zeta}_k^{(j_i)} - q_k|
\end{aligned}$$

Así, de esta forma el control $\bar{u}^*(\cdot)$ representa una aproximación al control $u^*(\cdot) \in \mathcal{S}$ que resuelve el problema original. Además, otra consecuencia de la no convexidad del conjunto $\mathcal{S}$ es que existe una *brecha* entre el control $\hat{u}(\cdot) \in \mathcal{S}$ y el control $\bar{u}^*(\cdot)$, la cual afecta de manera directa el rendimiento del sistema. Esta *brecha* está principalmente afectada por los conjuntos de restricciones en los niveles constantes $\mathcal{Q}_k$, donde como consecuencia se tiene que:

$$\min_{u(\cdot) \in conv(\mathcal{S})} J(u(\cdot)) \leq \min_{u(\cdot) \in \mathcal{S}} J(u(\cdot))$$

Con la caracterización del control más cercano a $\hat{u}(\cdot)$ se tiene una aproximación a la solución del problema original, siempre y cuando el funcional de costo sea *bien comportado*. Por ejemplo, en la figura 4.23 se muestra un ejemplo de un funcional de costo *mal comportado* con un control de dos entradas.

El término *mal comportado* es asignado en el sentido de que el control $\bar{u}^*(\cdot)$ genera un costo mayor que algún otro elemento dentro de $\mathcal{S}$. Por ejemplo, en la figura 4.23 se muestra como el control $u^{(1)}(\cdot)$ genera un costo mayor que el control $u^{(4)}(\cdot)$ a pesar de ser $u^1(\cdot)$ el elemento más cercano a $\hat{u}(\cdot)$ en el sentido de la definición (4.31).

La existencia de estos ejemplos *mal comportados* representa todo un reto a resolver, pues al no tener la convexidad del conjunto $\mathcal{S}$ no existe herramental matemático para resolver el problema de manera directa y eficiente. Cabe notar que debido a la convexidad del funcional de costo J, éste presenta una conducta monótona creciente en cierta dirección y decreciente en otra a partir del mínimo $J(\hat{u}(\cdot))$. Entonces, el control que resuelve el problema original se debe de buscar en una vecindad del control $\hat{u}(\cdot)$ volviéndose un problema combinatorio.

Para los ejemplos de la subsección anterior, se muestra el control *más cercano* $\bar{u}^*(\cdot)$ al control $\hat{u}(\cdot)$ en el sentido de (4.31) en ambos casos (sistema LTI y sistema LTV). El costo generado por $\bar{u}_{LTI}^*(\cdot)$ es de 152.3109 unidades para el sistema LTI, mientras que el costo generado por $\bar{u}_{LTV}^*(\cdot)$ es de 12.602 unidades para el sistema LTV. En las figuras 4.24 - 4.30 se muestran las trayectorias en el caso LTI y en las figuras 4.31 - 4.37 se muestran para el caso LTV.

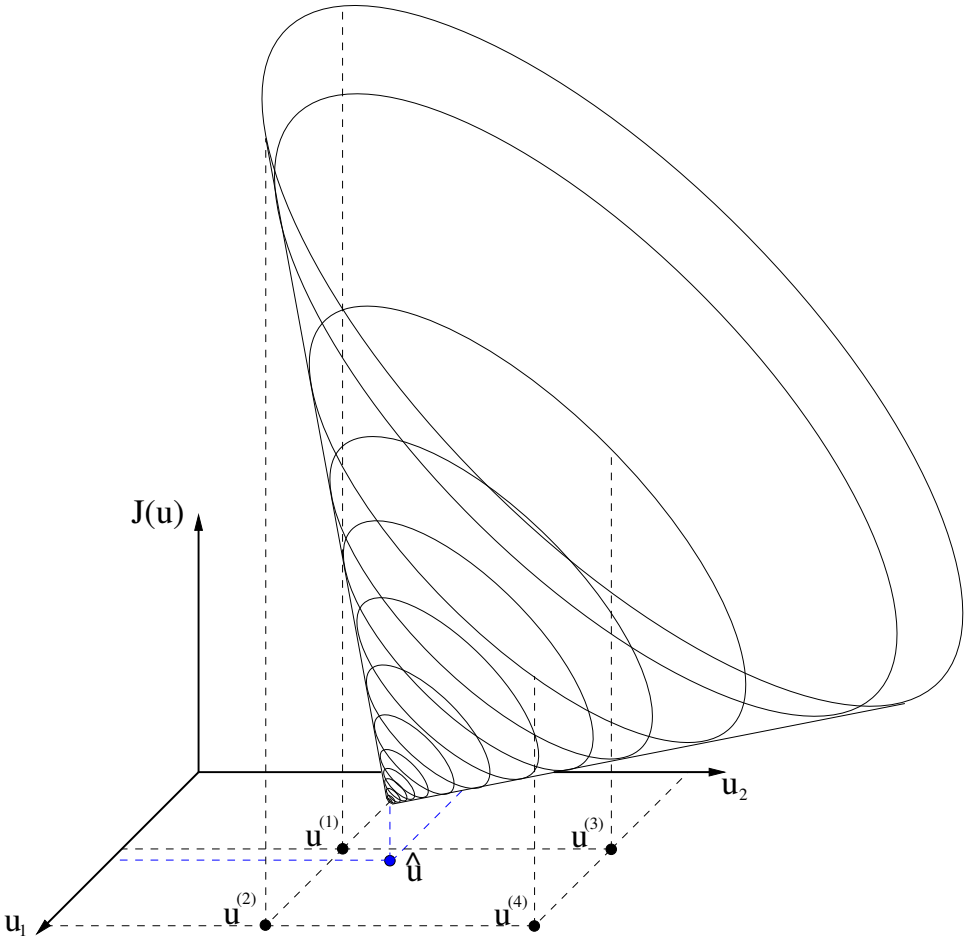

Figura 4.23: Funcional de costo *mal comportado*

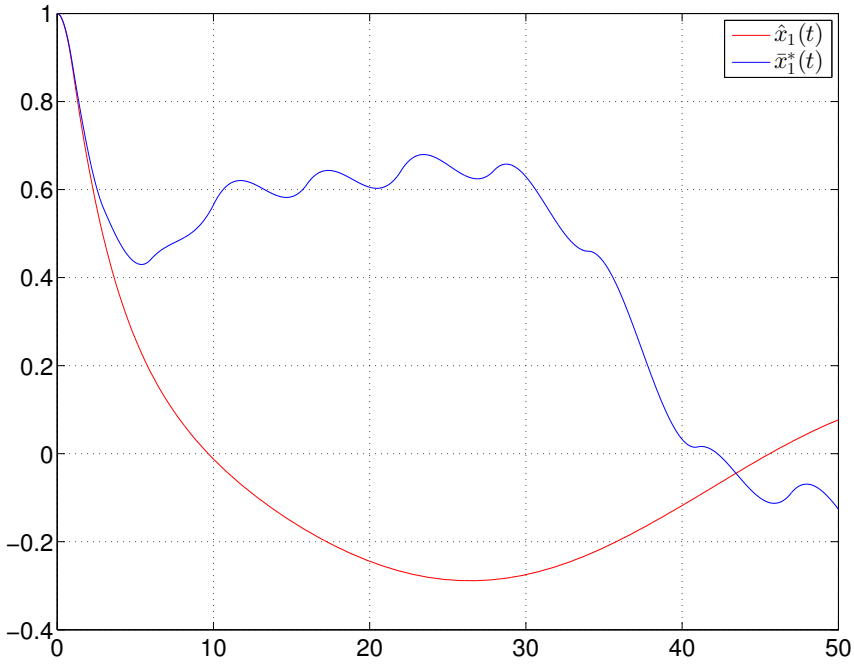

Figura 4.24: $\bar{x}_1^*(t)$ vs. $\hat{x}_1(t)$

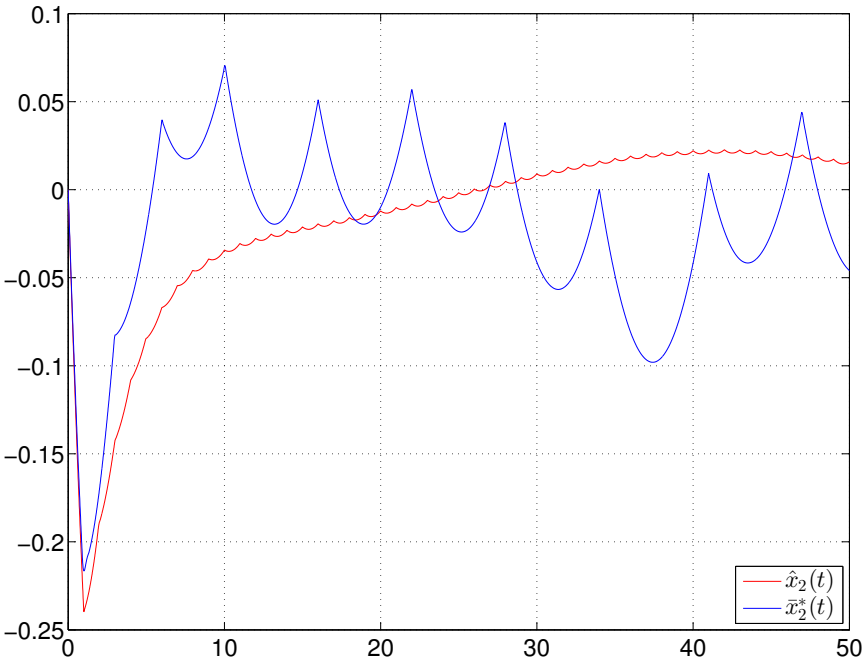

Figura 4.25: $\bar{x}_2^*(t)$ vs. $\hat{x}_2(t)$

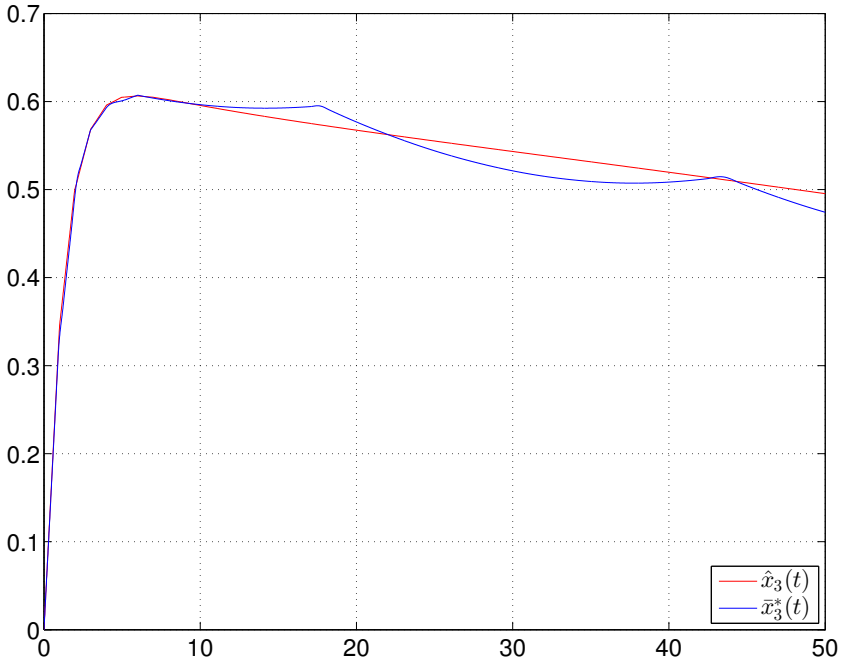

Figura 4.26: $\bar{x}_3^*(t)$ vs. $\hat{x}_3(t)$

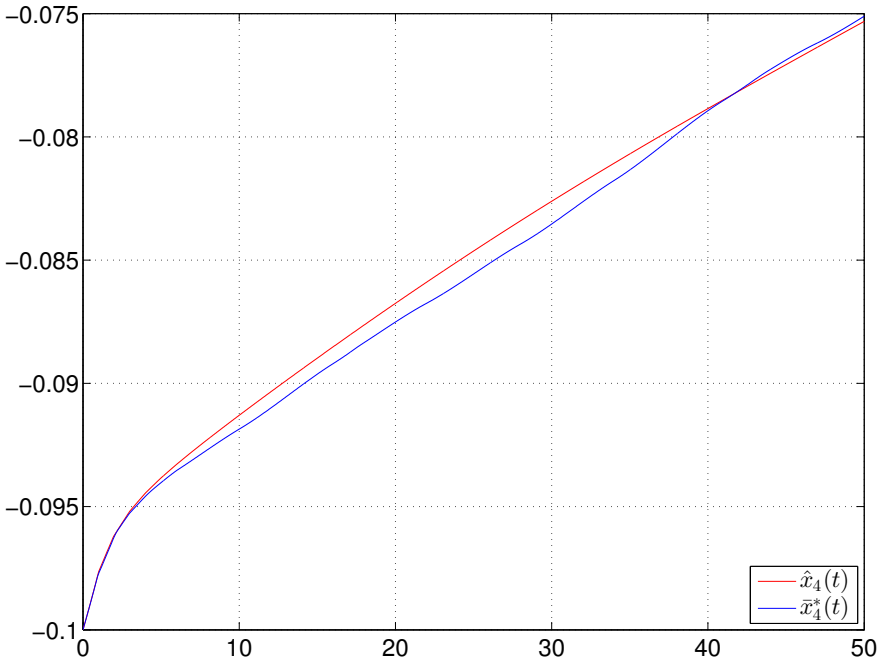

Figura 4.27: $\bar{x}_4^*(t)$ vs. $\hat{x}_4(t)$

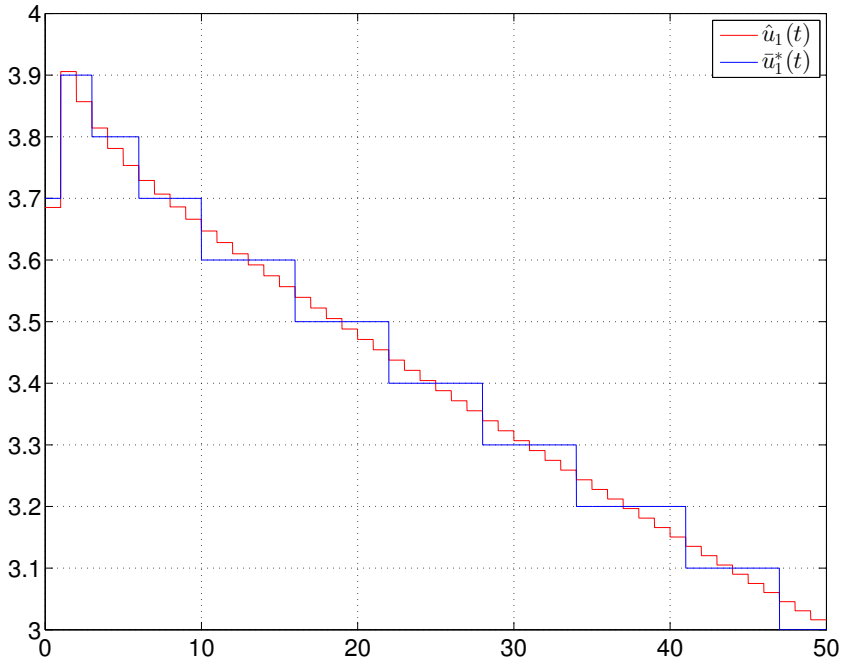

Figura 4.28: $\bar{u}_1^*(t)$ vs. $\hat{u}_1(t)$

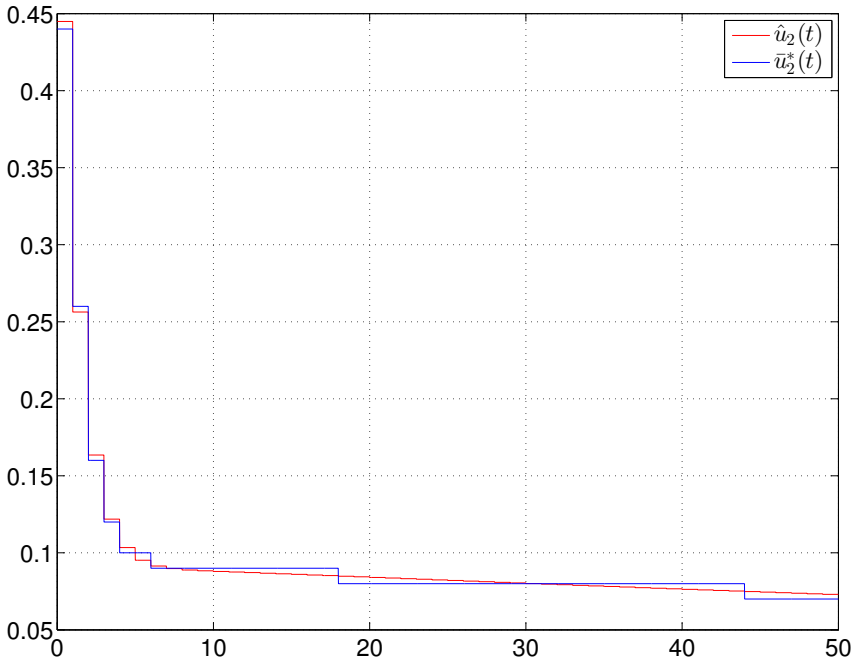

Figura 4.29: $\bar{u}_2^*(t)$ vs. $\hat{u}_2(t)$

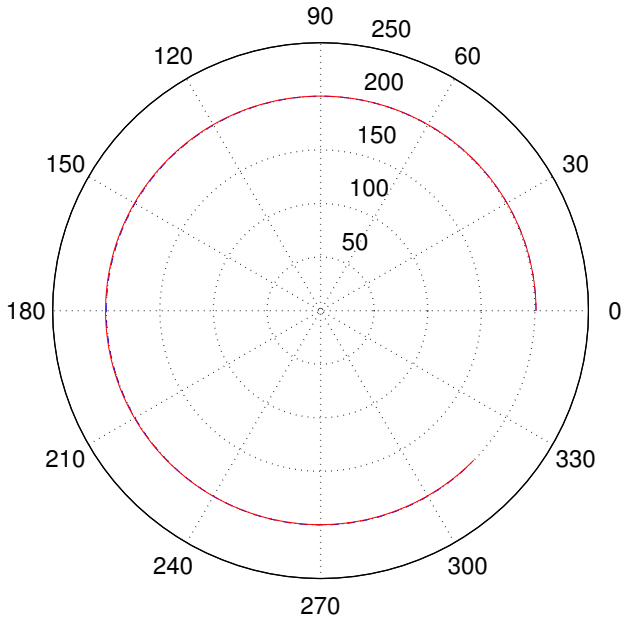

Figura 4.30: Trayectoria deseada vs. Trayectoria real del satélite al aplicar el control $\bar{u}^*(t)$

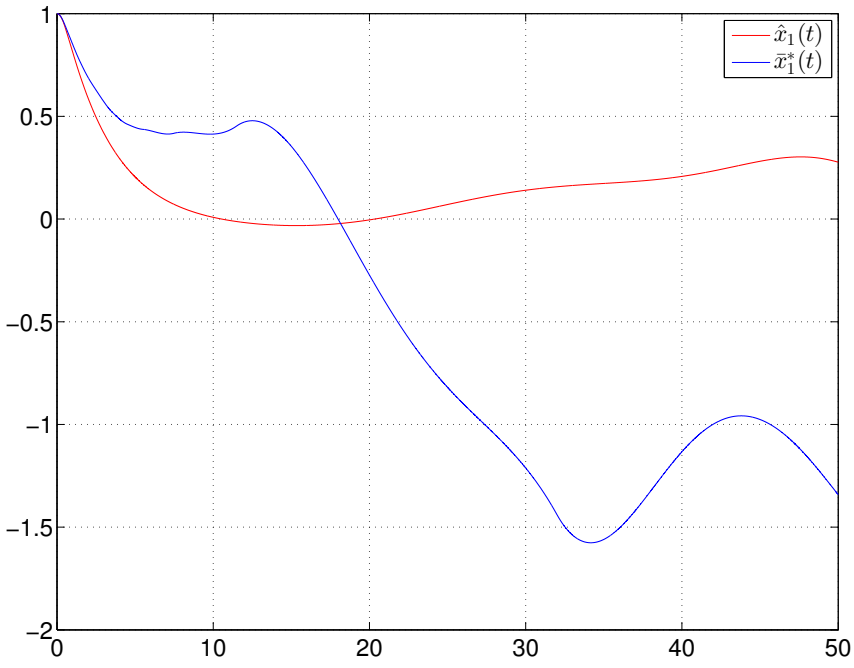

Figura 4.31: $\bar{x}_1^*(t)$ vs. $\hat{x}_1(t)$

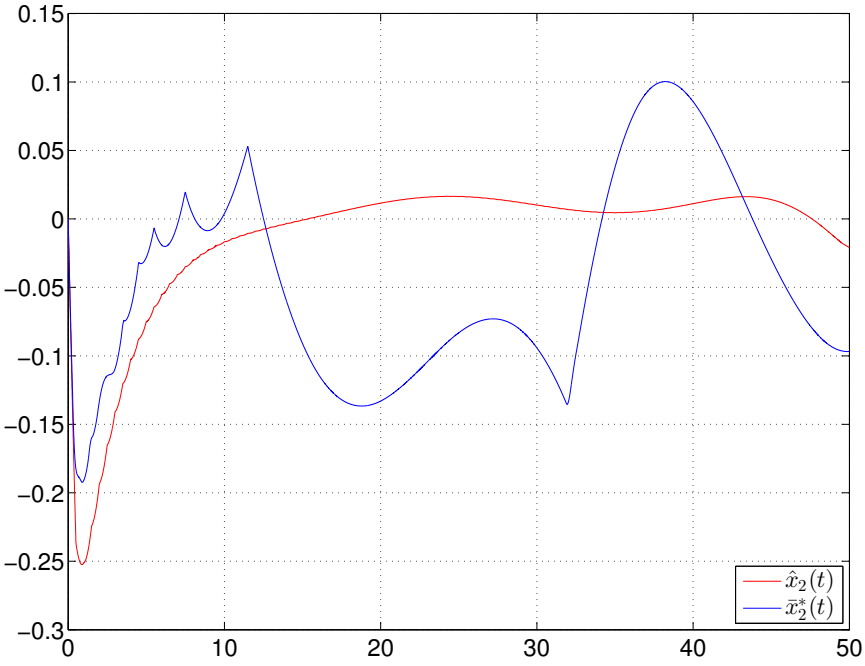

Figura 4.32: $\bar{x}_2^*(t)$ vs. $\hat{x}_2(t)$

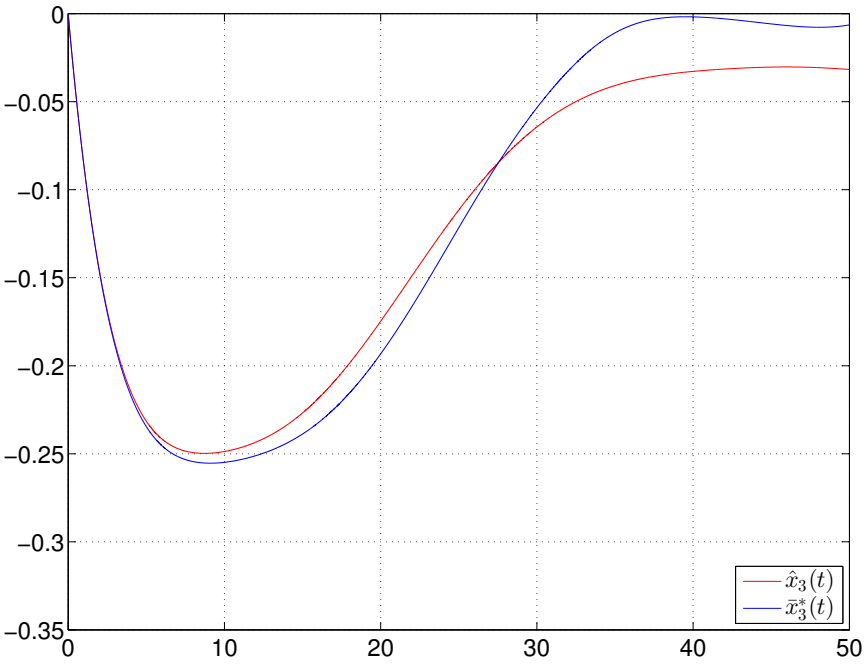

Figura 4.33: $\bar{x}_3^*(t)$ vs. $\hat{x}_3(t)$

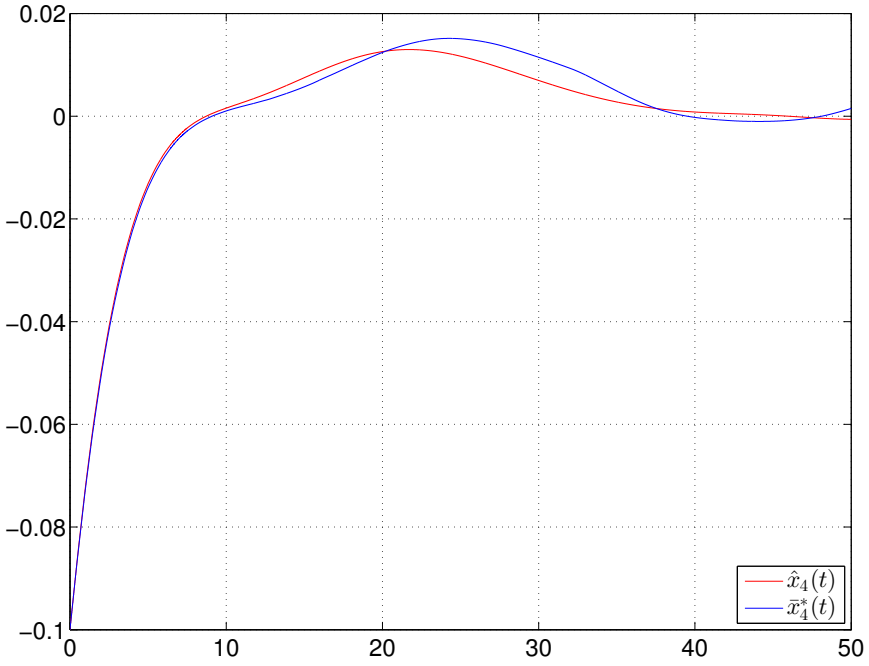

Figura 4.34: $\bar{x}_4^*(t)$ vs. $\hat{x}_4(t)$

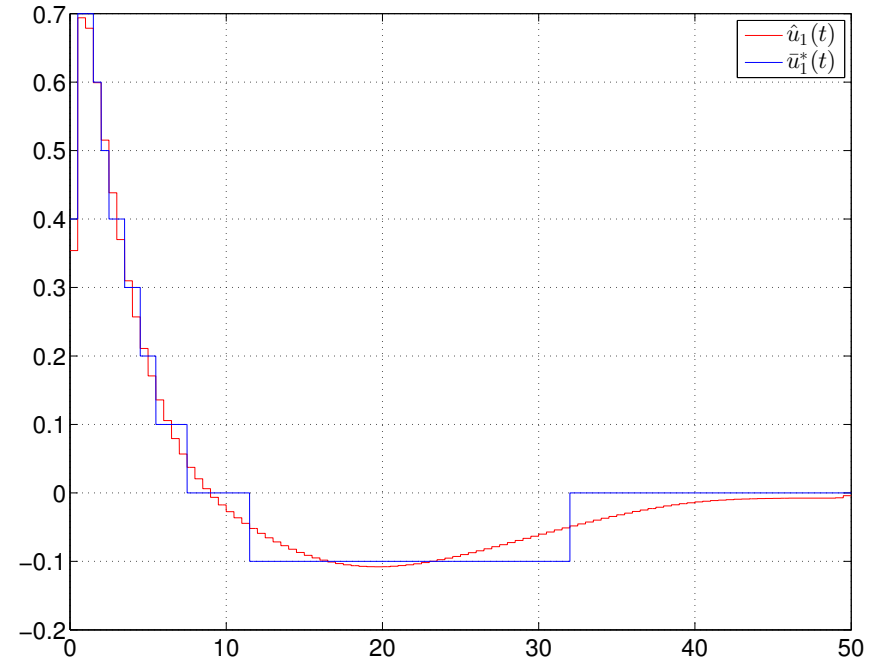

Figura 4.35: $\bar{u}_1^*(t)$ vs. $\hat{u}_1(t)$

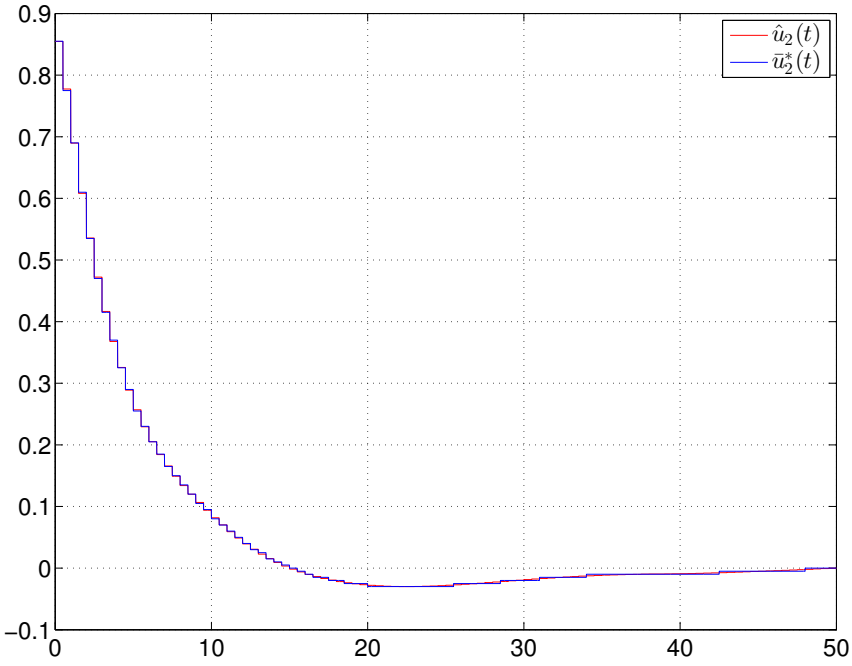

Figura 4.36: $\bar{u}_2^*(t)$ vs. $\hat{u}_2(t)$

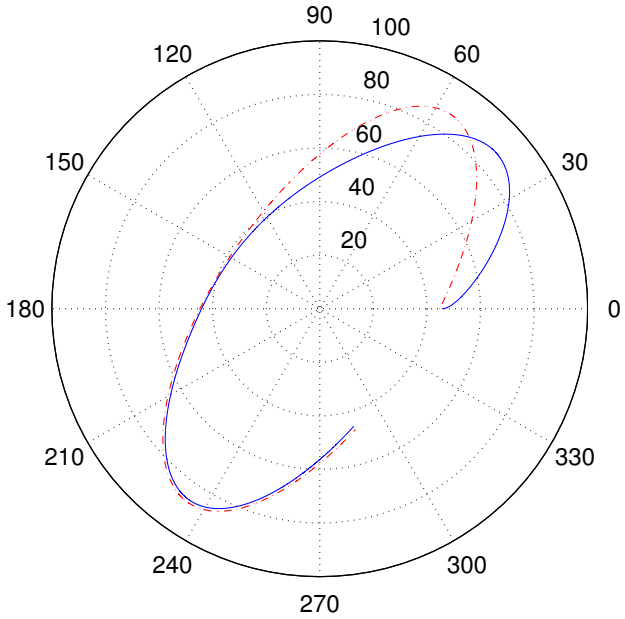

Figura 4.37: Trayectoria deseada vs. Trayectoria real del satélite al aplicar el control $\bar{u}^*(t)$

CAPÍTULO 5

Conclusiones y Trabajos Futuros

En el presente trabajo se propuso un nuevo enfoque numérico para resolver un problema tipo LQ para un sistema lineal en tiempo continuo con controles constantes a trozos y secuencias de conmutación fijas.Además, se considera que el conjunto de controles admisibles queda restringido por un conjunto de niveles aceptables, los cuales dependen de cada aplicación y afectan directamente el rendimiento del sistema.

El esquema computacional propuesto se basa en obtener una versión relajada del problema, primero tomando el conjunto convexo más pequeño que contiene a todas las restricciones, para posteriormente aplicar técnicas numéricas clásicas tipo gradiente. Al reformular el problema original en su versión relajada, se hace un estudio de las propiedades de convexidad de sus elementos, haciendo posible el uso de técnicas clásicas de programación convexa.

También se hizo un pequeño estudio para buscar la solución analítica del problema, aunque para este caso dicha representación está lejos de ser alcanzable, pues nuevamente las restricciones salen a flote para mostrar la complejidad del problema, ya que en el estudio analítico no fue posible agregar los tiempos de conmutación distintos entre componentes del control de una manera natural y sencilla de analizar y solamente se encontró dicha solución para casos particulares donde hay cierto desacoplamiento entre los componentes del control ó cuando el conjunto de restricciones en los niveles es muy grande.

A lo largo de este trabajo, se ha buscado obtener una solución al problema original 4.1 (p. 65), obteniendo una versión relajada de éste, para posteriormente hacer una aproximación al control que resuelve el problema original, explotando el hecho de que el funcional $J(\cdot)$ es una función continua en $u(\cdot)$. Esta última aproximación resulta no muy satisfactoria para algunos problemas, (como en los ejemplos de seguimiento de trayectoria del satélite en el caso de trayectoria elíptica), pero debe considerarse que la ley de control admisible está restringida

a un conjunto muy pobre (comparado con $\mathbb{L}_2$), por lo que la degradación en el rendimiento es debida a las restricciones y no al enfoque propuesto.

Ahora bien, el enfoque propuesto resultó satisfactorio para sistemas lineales invariantes en el tiempo donde la dinámica del sistema no es muy rápida en comparación con los tiempos de conmutación de los controles admisibles, y donde las restricciones en los niveles del control no representan una restricción activa para el sistema.

El método propuesto en este trabajo puede ser visto como una parte de un algoritmo más complejo, por ejemplo, para el caso donde la secuencia de tiempos no es fija, entonces se puede hacer una búsqueda recursiva sobre la secuencias de tiempos admisibles y en cada iteración del algoritmo aplicar el método propuesto para así obtener la siguiente iteración, como es el caso en la Etapa 1 del algoritmo presentado en [27].

BIBLIOGRAFÍA

[1] V. Azhmyakov, V.G. Boltyanski, and A. Poznyak. Optimal control of impulsive hybrid systems. *Nonlinear Analysis: Hybrid Systems*, 2(4):1089 – 1097, 2008.

[2] V. Azhmyakov, R. Galvan-Guerra, and M. Egerstedt. Hybrid lq-optimization using dynamic programming. In *Proceedings of the 2009 conference on American Control Conference*, ACC'09, pages 3617–3623, Piscataway, NJ, USA, 2009. IEEE Press.

[3] Vadim Azhmyakov, SidAhmed Attia, and Jörg Raisch. On the maximum principle for impulsive hybrid systems. In *Hybrid Systems: Computation and Control*, volume 4981 of *Lecture Notes in Computer Science*, pages 30–42. Springer Berlin Heidelberg, 2008.

[4] Vadim Azhmyakov, Michael Basin, and Jörg Raisch. A Proximal Point Based Approach to Optimal Control of Affine Switched Systems. *Discrete Event Dynamic Systems*, pages 1–21, June 2011.

[5] D.P. Bertsekas. *Constrained optimization and Lagrange multiplier methods*. Optimization and neural computation series. Athena Scientific, 1996.

[6] R.W. Brockett and D. Liberzon. Quantized feedback stabilization of linear systems. *Automatic Control, IEEE Transactions on*, 45(7):1279 –1289, jul 2000.

[7] Chi-Tsong Chen. *Linear System Theory and Design*. Oxford University Press, 1999.

[8] X.-C. Ding, Y. Wardi, and M. Egerstedt. On-line optimization of switched-mode dynamical systems. *Automatic Control, IEEE Transactions on*, 54(9):2266 –2271, sept. 2009.

[9] M. Egerstedt, Y. Wardi, and H. Axelsson. Transition-time optimization for switched-mode dynamical systems. *Automatic Control, IEEE Transactions on*, 51(1):110 – 115, jan. 2006.

[10] Ivar Ekeland and Roger Témam. *Convex Analysis and Variational Problems*. SIAM, second edition, 1999.

[11] G. Goodwin, M.M. Seron, and J.A. de Doná. *Constrained Control and Estimation: An Optimisation Approach*. Communications and Control Engineering. Springer, 2004.

[12] Akira Kojima and Manfred Morari. Lq control for constrained continuous-time systems. *Automatica*, 40(7):1143 – 1155, 2004.

[13] A.J. Kurdila and M. Zabarankin. *Convex Functional Analysis*. Systems & Control. Birkhäuser Basel, 2005.

[14] Leslie Lamport. *Latex: A Document Preparation System*. Addison-Wesley, second edition, 1994.

[15] E. S. Levitin and B. T. Polyak. Constrained minimization methods. *URSS Comput. Math. Math. Phys.*, 6:787–823, 1965.

[16] D. Liberzon. *Switching in Systems and Control*. Systems & Control. Birkhäuser, 2003.

[17] Daniel Liberzon. Hybrid feedback stabilization of systems with quantized signals. *Automatica*, 39(9):1543 – 1554, 2003.

[18] D.G. Luenberger and Y. Ye. *Linear and Nonlinear Programming*. International Series in Operations Research & Management Science. Springer, 2008.

[19] G. G. Magaril-Il'yaev and V. M. Tikhomirov. *Convex Analysis: Theory and Applications*. American Mathematical Society, 2003.

[20] B.T. Polyak. *Introduction to optimization*. Translations series in mathematics and engineering. Optimization Software, Publications Division, 1987.

[21] Alexander S. Poznyak. *Advanced Mathematical Tools for Engineers: Deterministic Techniques*, volume 1. Elsevier, first edition, 2008.

[22] R. Tyrrell Rockafellar. *Convex Analysis*. Princenton University Press, second edition, 1972.

[23] Wilson Rugh. *Linear System Theory*. Prentice Hall, second edition, 1996.

[24] M.S. Shaikh and P.E. Caines. On the hybrid optimal control problem: Theory and algorithms. *Automatic Control, IEEE Transactions on*, 52(9):1587 –1603, sept. 2007.

[25] Stuart A. Stanton and Belinda G. Marchand. Finite set control transcription for optimal control applications. *Jorunal of Spacecraft and Rockets*, 47(3):457 – 471, 2010.

[26] Michael Ulbrich. Optimization methods in banach spaces. In *Optimization with PDE Constraints*, volume 23 of *Mathematical Modelling: Theory and Applications*, pages 97–156. Springer Netherlands, 2009.

[27] Xuping Xu and P.J. Antsaklis. An approach for solving general switched linear quadratic optimal control problems. In *Decision and Control, 2001. Proceedings of the 40th IEEE Conference on*, volume 3, pages 2478 –2483 vol.3, 2001.

[28] Jiongmin Yong and Xun Yu Zhou. *Stochastic Controls: Hamiltonian Systems and HJB Equations*. Springer, 1999.