ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

OPTIMIZACIÓN DISPERSA DE ECUACIONES DIFERENCIALES PARCIALES ELÍPTICAS EN UN ESPACIO DE CONTROL FINITO CON RESTRICCIONES DE CONTROL Y ESTADO

TRABAJO PREVIO A LA OBTENCIÓN DEL TÍTULO DE MATEMÁTICO

PROYECTO DE INVESTIGACIÓN

HERNÁN ALEXANDER NENJER MORILLO
hernan.alex_nenjer@hotmail.com

Director: PEDRO MARTIN MERINO ROSERO, PH.D.
pedro.merino@epn.edu.ec

QUITO, OCTUBRE 2017

DECLARACIÓN

Yo HERNÁN ALEXANDER NENJER MORILLO, declaro bajo juramento que el trabajo aquí escrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y que he consultado las referencias bibliográficas que se incluyen en este documento.

A través de la presente declaración cedo mis derechos de propiedad intelectual, correspondientes a este trabajo, a la Escuela Politécnica Nacional, según lo establecido por la Ley de Propiedad Intelectual, por su reglamento y por la normatividad institucional vigente.

Hernán Alexander Nenjer Morillo

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por HERNÁN ALEXANDER NENJER MORILLO, bajo mi supervisión.

Pedro Martin Merino Rosero, Ph.D.
Director del Proyecto

AGRADECIMIENTOS

A Pedro Merino por haberme asignado este interesante tema de investigación y apoyarme con su paciencia, motivación y conocimiento.

Al grupo de MODEMAT por la ayuda brindada en el desarrollo de este proyecto.

A mis amigos por su privilegiada compañía . En particular, a Jonathan Ortiz por la ayuda recibida a lo largo de la carrera.

Finalmente quiero agradecer a mis Padres, Hernán y Germania, a mis hermanos Melani y Eddy, a mis abuelitos, a mis tíos, a mis primos, y a todos mis seres queridos, por haber estado siempre a mi lado.

DEDICATORIA

A mis padres y hermanos.

Índice general

Resumen

El objetivo del presente trabajo es el de obtener una estimación del error generado al aproximar numéricamente, mediante el método de elementos finitos, un problema de control óptimo gobernado por una ecuación diferencial parcial lineal elíptica con condiciones de frontera Dirichlet homogénea, con las siguientes particularidades:

- El espacio de los controles es de dimensión finita.
- Restricciones puntuales sobre el estado son impuestas en el dominio de la ecuación.
- El funcional de costo posee una penalización en norma-ℓ_1, la cual no es diferenciable.

Se demuestra que el orden de la estimación para el error es $h^2|\log h|$, donde h representa el tamaño de la malla. Finalmente, para ilustrar esta estimación se realizan experimentos numéricos.

Palabras clave: problema de control óptimo elíptico, restricciones de estado, estimación de errores, método de elementos finitos, optimización dispersa.

Abstract

The purpose of this study is deriving an estimation of the generated error due to the numerical approximation of a optimal control problem governed by linear elliptic partial differential equations with homogeneous Dirichlet boundary conditions (by the finite element method) with the following particularities:

- The control space is finite dimentional.
- Pointwise state constraints are imposed in the domain of equation.
- The functional cost includes a ℓ_1-norm penalization, which is non-differentiable.

It is proved that the order of convergence for the error is $h^2|\log h|$, where h represents the size of the mesh. Finally, to illustrate this estimation, numerical experiments are conducted.

Keywords: elliptic optimal control problem, state constraints, error estimates; finite element method, sparse optimization.

Notaciones

$\mathbb{R}^n$	Espacio de dimensión n de números reales
$\mathbb{R}^{n-}$	$\{z \in \mathbb{R}^n : z_j < 0,\ \forall j = 1, \ldots, l\}$.
$\mathbb{R}^n_+$	$\{z \in \mathbb{R}^n : z_j \geq 0,\ \forall j = 1, \ldots, l\}$.
$\|\|x\|\|_1 = \sum_{i=1}^{M} \|x_i\|$	La norma ℓ_1 en $\mathbb{R}^M$.
$\|\|x\|\|_2 = \|\|x\|\| = \left(\sum_{i=1}^{M} x_i^2\right)^{1/2}$	La norma euclidiana en $\mathbb{R}^M$.
$\langle x, y\rangle = \sum_{i=1}^{M} x_i y_i$	El producto escalar sobre $\mathbb{R}^M$.
$\nabla f(x)$	El gradiente de una función f diferenciable en x.
$\nabla^2 f(x)$	La Hessiana de una función f evaluado en x.
I_C	La función indicatriz del conjunto C.
$\bar{C}$	Clausura de C.
$\operatorname{int} C$	Interior del conjunto C.
Ω	Un dominio Lipschitz acotado en $\mathbb{R}^2$.
Γ	Frontera de Ω.
$\|\| \cdot \|\|$	La norma correspondiente al espacio funcional $L^2(\Omega)$.
$(\cdot, \cdot)$	El producto interno de $L^2(\Omega)$.
X^*	Espacio dual de X.
$B(u, \rho)$	Bola abierta de radio ρ con centro en u.
$\langle \cdot, \cdot \rangle_{X^*, X}$	Evaluación de un elemento de X^* en un punto de X.
$\|\| \cdot \|\|_{X^*}$	Norma del espacio dual de X^*.
$[v_i]_{i=j+1}^{n} = \begin{pmatrix} v_{j+1} \\ v_{j+2} \\ \vdots \\ v_n \end{pmatrix}$	Un vector en $\mathbb{R}^{n-j}$.

Capítulo 1

Introducción

Existen numerosos fenómenos físicos que pueden ser descritos mediante las ecuaciones diferenciales parciales por ejemplo: en la propagación del sonido, transferencia de calor, la electrostática, la electrodinámica, la dinámica de fluidos, la elasticidad, la mecánica cuántica, etc. cf. [14]. Una vez modelados dichos fenómenos, es de interés controlarlos de manera óptima. De aquí la importancia de los problemas de control óptimo para este tipo de ecuaciones. La elección de un espacio de control de dimensión finita está motivada por requerimientos tecnológicos ya que, en la práctica, varias cantidades finitas son seleccionadas para controlar ciertos procesos en aplicaciones reales cf. [13], [18], [28]. De hecho, hay aplicaciones en las que resulta muy costoso o poco realista trabajar con controles de funciones, por lo cual es necesario que las variables de decisión sean finitas (controles finitos). Adicionalmente, en el caso de controles finitos, cuando estas variables son muy costosas o muy numerosas, una estrategia que puede significar en la práctica un ahorro importante de recursos es determinar las variables más relevantes de control. Esto nos motiva a considerar la norma-ℓ_1 como término de penalización, debido a que es conocido que ésta induce soluciones que tienen gran cantidad de componentes iguales a cero cf. [28]. Esta propiedad es conocida como ”dispersión” o en inglés ”sparsity”, la cual nos permite disminuir los costos de la aplicación y nos da una idea de cuales son las componentes más importantes de control. Por otro lado, las restricciones puntuales sobre el estado también son importantes en aplicaciones en las cuales se específica mantener el estado restringido cf. [3], [20]. Aunque en la literatura existen varios trabajos que consideran la aproximación y análisis del error (por elementos finitos) para problemas con dispersión cf. [5], [6], [31], son pocos los que consideran dispersión y restricciones de estado cf. [7].

Este tipo de problemas han sido estudiados en cf. [19] con un término de penalización en la norma de L^2. Posteriormente en cf. [31] se considera términos sparse, sin embargo estos trabajos están dedicados a la resolución numérica más a no a la pregunta sobre la estimación de errores de aproximación por el método de elementos finitos.

La novedad de este trabajo consiste en que abordamos la interrogante sobre el orden del error de aproximación de la resolución del problema de control óptimo con penalización en norma-ℓ_1 por el método de elementos finitos para problemas de controles finitos y restricciones de estado. A continuación se describe el contenido de este proyecto: en el Capítulo 2 se introduce algunas definiciones y resultados fundamentales, así como también se resumen trabajos preliminares los cuales nos ayudarán con el entendimiento de toda la investigación. En el Capítulo 3 analizamos la formulación del problema, además de estudiar la existencia, unicidad y regularidad de la ecuación de estado a partir de la cual demostrar la existencia de una solución óptima y finalmente realizar un estudio sobre las condiciones de optimalidad que satisface dicha solución. El Capítulo 4 presenta todo lo correspondiente a la aproximación numérica del problema de control, obtenida a partir del método de elementos finitos. Principalmente, la aproximación se hace sobre la ecuación de estado en el cual se estima el orden de los errores. El Capítulo 5 presenta el estudio del problema sin regularización de Tikhonov, es decir analizamos el caso $\alpha = 0$. En el Capítulo 6 se ilustran experimentos numéricos para verificar los resultados teóricos obtenidos. Finalmente, en el Capítulo 7 se exponen algunas conclusiones y comentarios del trabajo realizado.

Capítulo 2

Definiciones y resultados preliminares

2.1. Espacios funcionales

Los siguientes conceptos y resultados fueron tomados de [2], [4], [14], [16] y [19]. Para lo cual, consideramos Ω un subconjunto abierto y acotado de $\mathbb{R}^m$.

DEFINICIÓN 2.1. *Diremos que la frontera $\Gamma = \bar{\Omega}\backslash\Omega$ es continua (respectivamente Lipschitz, continuamente diferenciable, de clase $C^{k,1}$, n veces diferenciable) si para cada $x \in \Gamma$ existe una vecindad V de x en $\mathbb{R}^m$ y nuevas coordenadas ortogonales $\{y_1, y_2, \ldots, y_m\}$ tales que*

1. *V es un hipercubo en las nuevas coordenadas, es decir*

$$V = \{(v_1, \ldots, v_m) : -a_i < v_i < a_i, \quad i = 1, \ldots, m\};$$

2. *Existe una función ϕ continua (respectivamente Lipschitz, continuamente diferenciable, de clase $C^{k,1}$, n veces continuamente diferenciable), definida en*

$$V' = \{(v_1, \ldots, v_{m-1}) : -a_i < v_i < a_i, \quad i = 1, \ldots, m-1\}$$

y tal que

$$|\phi(y')| \leq \frac{a_m}{2}, \text{ para cada } y' = (y_1, \ldots, y_{m-1}) \in V',$$
$$\Omega \cap V = \{y = (y', y_m) \in V : y_m < \phi(y')\},$$
$$\Gamma \cap V = \{y = (y', y_m) \in V : y_m = \phi(y')\}.$$

En particular, un subconjunto abierto acotado de $\mathbb{R}^2$ cuya frontera es un polígono, tiene una frontera Lipschitz pero esta no es continuamente diferenciable.

DEFINICIÓN 2.2. *Definimos el espacio de funciones continuas como el conjunto*

$$C(\bar{\Omega}) = \{f : \bar{\Omega} \to \mathbb{R} : f \text{ es continua}\}$$

normado por

$$||f||_{C(\bar{\Omega})} = \sup_{x\in\Omega} |f(x)|.$$

Así mismo consideramos $C^k(\bar{\Omega})$ el espacio de funciones reales definidas en $\bar{\Omega}$, k veces continuamente diferenciables, que pueden ser normado por

$$||f||_{C^k(\bar{\Omega})} = \sum_{|\alpha|\leq k} ||D^\alpha f||_{C(\bar{\Omega})}.$$

En esta definición, para $\alpha = (\alpha_1, \ldots, \alpha_n) \in \mathbb{N}^n$ hemos utilizado las siguiente notación

$$|\alpha| = \sum_{i=1}^{n} \alpha_i \quad \text{y} \quad D^\alpha f = \frac{\partial^{|\alpha|} f}{\partial_{x_1}^{\alpha_1} \ldots \partial_{x_n}^{\alpha_n}}.$$

Para $k = 0, 1, 2, \ldots$, el espacio $(C^k(\bar{\Omega}), ||\cdot||_{C^k(\bar{\Omega})})$ es un espacio de Banach.

DEFINICIÓN 2.3. *Para $0 < \gamma \leq 1$, el espacio de Hölder $C^{k,\gamma}(\bar{\Omega})$ consiste de las funciones $f \in C^k(\bar{\Omega})$ para el cual la norma*

$$||f||_{C^{k,\gamma}(\bar{\Omega})} = \sum_{|\alpha|\leq k} ||D^\alpha f||_{C^k(\bar{\Omega})} + \sum_{|\alpha|=k} |D^\alpha f|_{C^{0,\gamma}(\bar{\Omega})}$$

es finita, donde $|\cdot|_{C^{0,\gamma}(\bar{\Omega})}$ denota la semi-norma γ-Hölder, definida por

$$|f|_{C^{0,\gamma}(\bar{\Omega})} = \sup_{\substack{x_1,x_2\in\bar{\Omega} \\ x_1\neq x_2}} \left\{ \frac{|f(x_1) - f(x_2)|}{|x_1 - x_2|^\gamma} \right\}.$$

Con esta norma $C^{k,\gamma}(\bar{\Omega})$ es un espacio de Banach.

DEFINICIÓN 2.4. *El conjunto $C_c(\Omega)$ representa el espacio de funciones continuas con soporte compacto en Ω. $C_c^k(\Omega)$ es el espacio de funciones continuas k veces diferenciables con soporte compacto.*

DEFINICIÓN 2.5. *Definimos los espacios de Lebesgue $L^p(\Omega)$, con $1 \leq p < \infty$, como el conjunto*

$$L^p(\Omega) = \left\{ f : \Omega \to \mathbb{R} \text{ medible} : \int_\Omega |f(x)|^p dx < \infty \right\}.$$

Este espacio es de Banach con la siguiente norma

$$||f||_{L^p(\Omega)} = \left[\int_\Omega |f(x)|^p dx\right]^{1/p}.$$

DEFINICIÓN 2.6. *Definimos el espacio de Lebesgue cuando $p = \infty$, como*

$$L^\infty(\Omega) = \{f : \Omega \to \mathbb{R} \text{ medible} : \ \exists C \geq 0 \text{ tal que } |f(x)| < C \quad \text{c.t.p. en } \Omega\}.$$

Este espacio es de Banach con la siguiente norma

$$||f||_{L^\infty(\Omega)} = \operatorname*{ess\,sup}_{x\in\Omega} |f(x)| = \text{ínf}\{C \in \mathbb{R}_+ : |f(x)| < C \quad \text{c.t.p. en } \Omega\}.$$

OBSERVACIÓN. $|f(x)| \leq ||f||_{L^\infty(\Omega)}$ c.t.p en Ω.

PROPOSICIÓN 2.1. *Sea $f : \bar{\Omega} \to \mathbb{R}$ tal que $f \in L^\infty(\Omega) \cap C(\bar{\Omega})$ entonces*

$$||f||_{L^\infty(\Omega)} = ||f||_{C(\bar{\Omega})}.$$

Demostración. Puesto que $|f(x)| \leq ||f||_{C(\bar{\Omega})}$ para todo $x \in \bar{\Omega}$, de la definición de $||\cdot||_{L^\infty(\Omega)}$ se sigue que

$$||f||_{L^\infty(\Omega)} \leq ||f||_{C(\bar{\Omega})}.$$

Para la otra desigualdad, vamos a suponer lo contrario. Es decir, suponemos que $||f||_{L^\infty(\Omega)} < ||f||_{C(\bar{\Omega})}$, por lo tanto existe $x_0 \in \Omega$ tal que $||f||_{L^\infty(\Omega)} < |f(x_0)|$. Luego, puesto que f es continua existe una bola abierta $B(x_0, r)$ donde se cumple que

$$||f||_{L^\infty(\Omega)} < |f(x)|, \quad \forall x \in B(x_0, r),$$

lo cual no puede ser, pues la medida de $B(x_0, r)$ es no nula. □

DEFINICIÓN 2.7. *Sean $1 \leq p \leq \infty$ y $k \in \mathbb{N}$, definimos el espacio de Sobolev $W^{k,p}(\Omega)$ como el conjunto de todas las funciones $f \in L^p(\Omega)$, tales que para todo multi-índice $\alpha \in \mathbb{N}^n$ con $|\alpha| \leq k$, se tiene que $D^\alpha f$ existe en el sentido débil y pertenece a $L^p(\Omega)$, es decir*

$$W^{k,p}(\Omega) = \{f \in L^p(\Omega) : D^\alpha f \in L^p(\Omega), \ \forall |\alpha| \leq k\}.$$

En este espacio se define la siguiente norma

$$||f||_{W^{k,p}(\Omega)} = ||f||_{k,p,\Omega} = \begin{cases} \left(\sum_{|\alpha|\leq k} \int_\Omega |D^\alpha f(x)|^p dx \right)^{1/p}, & si \quad 1 \leq p < \infty, \\ \sum_{|\alpha|\leq k} \operatorname{ess\,sup}_{x\in\Omega} |D^\alpha f(x)|, & si \quad p = \infty. \end{cases}$$

Notemos que si $k = 0$ entonces $W^{0,p}(\Omega) = L^p(\Omega)$.

Si $p = 2$ denotamos $H^k(\Omega) = W^{k,2}(\Omega)$ para $k \in \mathbb{N}$. Además $H^k(\Omega)$ es un espacio de Hilbert con el producto escalar

$$(f, g)_{H^k(\Omega)} = \sum_{|\alpha|\leq k} \int_\Omega D^\alpha f(x) D^\alpha g(x) dx.$$

DEFINICIÓN 2.8. *Para $1 \leq p < +\infty$, se define $W_0^{1,p}(\Omega)$ como la clausura de $C_c^1(\Omega)$ en $W^{1,p}(\Omega)$.*

Denotamos $H_0^1(\Omega) = W_0^{1,2}(\Omega)$ el cual es un espacio de Hilbert.

2.2. Cálculo Subdiferencial

En esta sección revisamos brevemente algunos resultados del análisis convexo, el cual engloba una gran parte del cálculo subdiferencial para funciones convexas no diferenciables. Los resultados presentados a continuación fueron tomados de [9] y [10].

A lo largo de este capítulo denotamos por X, W dos espacios de Banach, mientras que Y denota un subconjunto de X.

DEFINICIÓN 2.9. *Una función $f : Y \to \mathbb{R}$ es Lipschitz sobre Y, si existe $K \geq 0$ tal que*

$$|f(y) - f(x)| \leq K||y - x||, \quad \forall x, y \in Y.$$

Diremos que f es localmente Lipschitz en x, si existe una vecindad V_x de x donde se cumple que

$$|f(y) - f(z)| \leq K||y - z||, \quad \forall y, z \in V_x.$$

DEFINICIÓN 2.10. *Sea $x \in Y$ y $f : Y \to \mathbb{R}$. La derivada direccional generalizada de*

f en x en una dirección $v \in X$, denotada por $f^\circ(x;v)$, está definida como

$$f^\circ(x;v) = \limsup_{\substack{y \to x \\ t \downarrow 0}} \frac{f(y+tv) - f(y)}{t} = \lim_{h \downarrow 0} \sup_{y \in B(x,h), t \in (0,h)} \frac{f(y+tv) - f(y)}{t}.$$

El gradiente generalizado de f en x, se define por

$$\partial_c f(x) := \{\zeta \in X^* : f^\circ(x;v) \geq \langle \zeta, v \rangle, \quad \forall v \in X\}.$$

Presentamos algunas propiedades básicas del gradiente generalizado.

PROPOSICIÓN 2.2. *Si f es localmente Lipschitz en x, entonces:*

1. *$\partial_c f(x)$ es un subconjunto no vacío de X^*.*
2. *$||\zeta||_* \leq K$ para todo $\zeta \in \partial_c f(x)$.*
3. *Para cada $v \in X$, se tiene $f^\circ(x;v) = \max\{\langle \zeta, v \rangle : \zeta \in \partial_c f(x)\}$.*

Para la demostración ver [10], páginas 27-28, Proposición 2.1.2.

DEFINICIÓN 2.11. *Sea $f : X \to W$ y $x \in X$, se define la derivada direccional de f en x en una dirección $v \in X$ (en el sentido usual), como*

$$f'(x;v) := \lim_{t \downarrow 0} \frac{f(x+tv) - f(x)}{t}$$

cuando este límite existe.

DEFINICIÓN 2.12. *Decimos que f es Gâteaux diferenciable en x, si la función*

$$\begin{aligned} D_G f(x) : \quad & X \to W \\ & v \mapsto D_G f(x)v = f'(x;v) \end{aligned}$$

es lineal y continua.

Decimos que f es Gâteaux diferenciable en X, si f es Gâteaux diferenciable en todo $x \in X$.

DEFINICIÓN 2.13. *f es continuamente Gâteaux diferenciable en x, si existe una vecindad V_x de x, tal que*

$$\begin{aligned} D_G f : \quad & V_x \to \mathcal{L}(X, W) \\ & x \mapsto D_G f(x) \end{aligned}$$

es una función continua en x.

f es continuamente Gâteaux diferenciable en X, si f es continuamente Gâteaux diferenciable en todo $x \in X$.

DEFINICIÓN 2.14. *Un conjunto $U \subset X$ es convexo, si para todo $x, y \in U$ y $\alpha \in [0,1]$ se tiene que*

$$\alpha x + (1-\alpha) \in U.$$

DEFINICIÓN 2.15. *Sea $f : U \to \mathbb{R}$ una función definida sobre un conjunto convexo U, decimos que f es convexa si para todo $x, y \in U$ y $\alpha \in [0,1]$ se tiene que*

$$f(\alpha x + (1-\alpha)y) \leq \alpha f(x) + (1-\alpha) f(y).$$

Si la desigualdad es estricta para $\alpha \in (0,1)$, entonces f se dice estrictamente convexa.

DEFINICIÓN 2.16. *Decimos que $\zeta \in X^*$ es un subgradiente de f en x (en el sentido del análisis convexo) si cumple la siguiente desigualdad:*

$$f(y) - f(x) \geq \langle \zeta, y - x \rangle, \quad \forall y \in X.$$

DEFINICIÓN 2.17. *El subdiferencial de f en x, denotado por $\partial f(x)$, es el conjunto de todos los subgradientes de f en x.*

OBSERVACIÓN. El subdiferencial de f en x es un subconjunto no vacío de X^*.

El siguiente resultado se encuentra en [9], página 60, Ejercicio 4.2.

PROPOSICIÓN 2.3. *Consideremos $f(x) = ||x||_1$,*

1. *$\partial f(0)$ es la bola cerrada unitaria en X^*.*
2. *si $\zeta \in \partial f(x)$ con $x \neq 0$, entonces $\langle \zeta, x \rangle = ||x||_1$ y $||\zeta||_{X^*} = 1$.*

Una aplicación de este resultado corresponde al cálculo del subdiferencial de $f(x) = ||x||_1$ con $X = \mathbb{R}^M$. Para ello, sea $\zeta \in \partial f(x)$ de la Proposición 2.3 se tiene que $||\zeta||_\infty \leq 1$ y $\langle \zeta, x \rangle = ||x||_1$, de lo cual se sigue que $\zeta_1 x_1 + \cdots + \zeta_M x_M = |x_1| + \cdots + |x_M|$ y $|\zeta_i| \leq 1$, para cada $i = 1, \ldots, M$. Luego, analizando por casos, se concluye que

$$\zeta_i \in \begin{cases} \{1\}, & \text{si } x_i > 0, \\ \{-1\}, & \text{si } x_i < 0, \\ [-1,1], & \text{si } x_i = 0, \end{cases}$$

para todo $i = 1, \ldots, M$.

PROPOSICIÓN 2.4. *Si f es una función convexa, entonces*

$$\partial f(x) = \left\{\zeta \in X^* : f'(x;v) \geq \langle \zeta, v\rangle_{X^*,X}, \quad \forall v \in X\right\}.$$

Para la demostración ver [9], página 61, Proposición 4.3.

PROPOSICIÓN 2.5. *Sea $f : X \to \mathbb{R}$ una función convexa y $x \in X$. Si f es Gâteaux diferenciable, entonces*

$$\partial f(x) = \{D_G f(x)\}.$$

Demostración. De la Proposición 2.4, se tiene que $D_G f(x) \in \partial f(x)$. Sea $\zeta \in \partial f(x)$ y $v \in X$ entonces

$$f(x+tv) - f(x) \geq t\langle \zeta, v\rangle, \quad \forall t \geq 0,$$

se sigue que

$$\lim_{t\downarrow 0} \frac{f(x+tv) - f(x)}{t} \geq \langle \zeta, v\rangle,$$

entonces

$$\langle D_G f(x), v\rangle \geq \langle \zeta, v\rangle.$$

De igual manera, si consideramos $-v$ de obtiene que

$$-\langle D_G f(x), v\rangle \geq -\langle \zeta, v\rangle,$$

por lo tanto

$$\langle D_G f(x), v\rangle = \langle \zeta, v\rangle$$

para todo $v \in X$, lo que implica que $D_G f(x) = \zeta$. □

El siguiente resultado nos permite calcular el subdiferencial de la función indicatriz de un conjunto convexo, .

PROPOSICIÓN 2.6. *Sea $C \subset \mathbb{R}^M$ un conjunto convexo cerrado y $x \in C$, entonces*

$$\partial I_C(x) = N_C(x) = \{y \in \mathbb{R}^M : \langle y, c - x\rangle \leq 0, \quad \forall c \in C\}.$$

La demostración es una implicación directa del Ejercicio 4.5 [página 61] y la Proposición 2.9 [página 30] de [9].

PROPOSICIÓN 2.7 (Subdiferencial del producto por un escalar)**.** *Sea $f : X \to \mathbb{R}$ una función Lipschitz cerca de $x \in X$ y $s \in \mathbb{R}$, entonces*

$$\partial(sf)(x) = s\partial f(x).$$

Para la demostración ver [10], página 38, Proposición 2.3.1.

PROPOSICIÓN 2.8 (Subdiferencial de la suma)**.** *Sean $f, g : X \to \mathbb{R}$ funciones convexas las cuales admiten un punto en X en el cual f es continua, entonces*

$$\partial(f+g)(x) = \partial f(x) + \partial g(x), \quad \forall x \in X.$$

Para la demostración ver [9], Página 63, Teorema 4.10.

PROPOSICIÓN 2.9. *Sea f una función convexa y Lipschitz cerca de x, entonces $\partial_c f(x)$ coincide con el subdiferencial en x (en el sentido del análisis convexo) y $f^\circ(x;v)$ coincide con la derivada direccional $f'(x;v)$, para todo $v \in X$.*

Para la demostración ver [10], Páginas 36, Proposición 2.2.7.

PROPOSICIÓN 2.10. *Si X es un espacio de dimensión finita, entonces cualquier función convexa $f : X \to \mathbb{R}$ es Lipschitz cerca de todo punto en X.*

Para la demostración ver [9], Páginas 39-40, Corolario 2.35.

DEFINICIÓN 2.18. *Sea F una función de X en W. Decimos que F es estrictamente diferenciable en $x \in X$, si la aplicación definida por*

$$\langle D_s F(x), v\rangle := \lim_{\substack{y \to x \\ t \downarrow 0}} \frac{F(y+tv) - F(y)}{t}, \qquad \forall v \in X$$

es lineal continuo y además el límite es uniforme con respecto a v en conjuntos compactos.

PROPOSICIÓN 2.11. *Si F es continuamente Gâteaux diferenciable en x, entonces F es estrictamente diferenciable en x, y*

$$D_s F(x) = D_G F(x).$$

Para la demostración ver [10], Página 32, Corolario.

2.3. Ecuaciones generalizadas

En esta sección se proporcionará una herramienta útil para la obtención de la estimación del error para la aproximación del problema de control óptimo con restricciones puntuales en el estado. Las definiciones y resultados presentados a conti-

nuación fueron tomados de [19], [25], [26] y [27].

Si A es un conjunto en un espacio normado $(X, ||\cdot||)$, la distancia de un punto dado y al conjunto A, está dado por

$$\text{dist}[y, A] = \begin{cases} \text{ínf}\{||y-a|| : a \in A\}, & \text{si } A \neq \emptyset, \\ +\infty, & \text{si } A = \emptyset. \end{cases}$$

Sean $(X, ||\cdot||_X)$, $(Y, ||\cdot||_Y)$ y $(Z, ||\cdot||_Z)$ espacios de Banach, H un subconjunto abierto de X, $G : H \to Y$ una función Fréchet diferenciable, y K un cono convexo contenido en Y. Definimos la relación $y \leq_K 0$ si $-y \in K$, con lo cual el sistema de nuestro interés es el siguiente

$$G(u) \leq_K 0, \qquad \forall u \in C, \tag{2.1}$$

donde C es un conjunto convexo cerrado no vacío en X tal que $C \cap H \neq \emptyset$.

Para nuestro estudio, consideramos $\bar{u} \in C \cap H$ que satisface (2.1).

DEFINICIÓN 2.19. *Una perturbación admisible del problema (2.1) en $\bar{u}$ es una tripleta $\{P, \bar{p}, G(\cdot,\cdot)\}$, donde P es un espacio topológico, $\bar{p} \in P$, y $G(\cdot,\cdot)$ es una función de $P \times H$ en Y, tal que*

1. $G(\bar{p}, u) = G(u), \quad \forall u \in H.$
2. *$G(\cdot,\cdot)$ es parcialmente Fréchet diferenciable con respecto a la segunda variable para todo $(p, u) \in P \times H$.*
3. *$G(\cdot,\cdot)$ y $\dfrac{\partial G}{\partial u}(\cdot,\cdot)$ son continuas en $(\bar{p}, \bar{u})$, donde $\dfrac{\partial G}{\partial u}(\cdot,\cdot)$ denota la derivada parcial mencionada en 2.*

En general, el sistema perturbado (2.1), lo expresamos como

$$G(p, u) \leq_K 0, \quad \forall u \in C. \tag{2.2}$$

Presentamos las condiciones para las cuales el sistema (2.2) tiene solución en una vecindad de $\bar{p}$ y si, además, el conjunto de soluciones del sistema perturbado tiene un buen comportamiento cerca de $\bar{u}$, es decir, si al realizar una perturbación el conjunto de las soluciones permanece cerca a $\bar{u}$, esto será medido por la dist$[0, \mathcal{G}(p, u)]$, donde $\mathcal{G}$ está definido por

$$\mathcal{G}(p, u) = \begin{cases} G(p, u) + K, & \text{si } u \in C \cap H, \\ \emptyset, & \text{si } u \notin C \cap H. \end{cases}$$

El conjunto de soluciones del sistema perturbado (2.2), para toda perturbación admisible p, está dado por

$$\Sigma(p) = \{u \in C \cap H : G(p,u) \leq_K 0\}.$$

DEFINICIÓN 2.20. *(Condición de regularidad de Robinson) Sea G una función continuamente Fréchet diferenciable en el punto $\bar{u}$. Decimos que (2.1) es regular en $\bar{u}$ si se cumple que*

$$0 \in \text{int}\{G(\bar{u}) + G'(\bar{u})(C - \bar{u}) + K\}. \tag{2.3}$$

Tenemos los siguiente resultados sobre la estabilidad de sistemas no lineales.

PROPOSICIÓN 2.12. *Supóngase que (2.1) es regular en $\bar{u} \in C \cap H$ para el cual se satisface que $G(\bar{u}) \leq_K 0$ y en el cual, G es continuamente Fréchet diferenciable. Entonces, existe una constante $c > 0$ tal que para cualquier perturbación admisible $\{P, \bar{p}, G(\cdot,\cdot)\}$ de (2.1) en $\bar{u}$, y cualquier $\varepsilon > 0$, existen vecindades $\mathcal{N}$ de $\bar{p}$, y $\mathcal{O}$ de $\bar{u}$ tal que para cada $p \in \mathcal{N}$ el sistema perturbado (2.2) tiene solución, y la estimación*

$$\text{dist}[u, \Sigma(p)] \leq c\, \text{dist}[0, \mathcal{G}(p,u)],$$

se cumple para todo $u \in \mathcal{O}$.

Para la demostración ver [25], páginas 501-503, Teorema 1.

DEFINICIÓN 2.21. *Una ecuación generalizada consiste en inclusiones de la forma:*

$$0 \in F(x) + \partial\psi_C(x), \tag{2.4}$$

donde F es una función de un espacio normado X a su dual X^*, C es un conjunto convexo cerrado no vacío en X y $\partial\psi_C(x)$ denota el operador cono normal definido por

$$\partial\psi_C(x) = \begin{cases} z \in X^* \text{ con } \langle z, c - x\rangle_{X^*,X} \leq 0,\ \forall c \in C, & \text{si } x \in C, \\ \emptyset, & \text{si } x \notin C. \end{cases}$$

PROPOSICIÓN 2.13. *Si C es un conjunto convexo cerrado no vacío de $\mathbb{R}^m$ y K un cono convexo de $\mathbb{R}^n$, entonces $\partial\psi_{C\times K}(x,\nu) = \partial\psi_C(x) \times \partial\psi_K(\nu)$.*

Demostración. Primero vamos a probar que $\partial\psi_{C\times K}(x,\nu) \subset \partial\psi_C(x) \times \partial\psi_K(\nu)$. Para ello, suponemos que $(x,\nu) \in C \times K$, entonces para todo $(z_1,z_2)^T \in \partial\psi_{C\times K}(x,\nu)$ se tiene que

$$z_1^T(c - x) + z_2^T(k - \nu) = (z_1,z_2)^T\Big((c,k) - (x,\nu)\Big) \leq 0,$$

para todo $(c,k) \in C \times K$. En particular, si tomamos $(c,k) = (c,\nu)$ y $(c,k) = (x,\nu)$ se obtiene que

$$z_1^T(c-x) \leq 0, \quad \forall c \in C,$$
$$z_2^T(k-\nu) \leq 0, \quad \forall k \in K,$$

esto implica que $(z_1,z_2)^T \in \partial\psi_C(x) \times \partial\psi_K(\nu)$. Para el caso en que $(x,\nu) \notin C \times K$, el resultado es inmediato.

Para la otra contenencia, sea $(z_1,z_2)^T \in \partial\psi_C(x) \times \partial\psi_K(\nu)$ y analizamos el caso en que $x \in C$ y $\nu \in K$, del cual se tiene que

$$z_1^T(c-x) \leq 0, \quad \forall c \in C,$$
$$z_2^T(k-\nu) \leq 0, \quad \forall k \in K.$$

y sumando estás desigualdades se concluye que

$$(z_1,z_2)^T((c,k)-(x,\nu)) \leq 0, \quad \forall (c,k) \in C \times K,$$

es decir $(z_1,z_2)^T \in \partial\psi_{C\times K}(x,\nu)$. De igual manera, la conclusión es inmediata para el caso en que $x \notin C$ o $\nu \notin K$. □

Para el análisis de estabilidad de la ecuación generalizada (2.4), presentamos las siguientes definiciones y resultados.

DEFINICIÓN 2.22 (Condición de regularidad fuerte)**.** *Para $\bar{x} \in X$ definimos la función $T : X \to X^* + 2^{X^*}$, tal que*

$$Tx = F(\bar{x}) + F'(\bar{x})(x-\bar{x}) + \partial\psi_C(x).$$

Sea $\bar{x}$ una solución de (2.4), decimos que (2.4) es fuertemente regular en $\bar{x}$ con una constante Lipschitz asociada C_L, si existen vecindades $\mathcal{U}$ del origen de X^* y $\mathcal{V}$ de $\bar{x}$ tal que la intersección de $T^{-1} \cap \mathcal{V}$ es una función Lipschitz con constante C_L, de $\mathcal{U}$ en $\mathcal{V}$.

PROPOSICIÓN 2.14 (Teorema de la función implícita de Robinson)**.** *Sea $\mathcal{O}$ un subconjunto abierto de un espacio normado X, P un espacio topológico, $F : \mathcal{O} \times P \to X^*$ una función, y C un subconjunto convexo cerrado de X. Suponemos que la derivada parcial de Fréchet de F con respecto a la primera componente, denotada por $F'(\cdot,\cdot)$, existe sobre $\mathcal{O} \times P$, tal que $F(\cdot,\cdot)$ y $F'(\cdot,\cdot)$ son continuas en $(x_0,p_0) \in \mathcal{O} \times P$ y que x_0 resuelve*

$$0 \in F(x,p_0) + \partial\psi_C(x). \tag{2.5}$$

Si (2.5) es fuertemente regular en x_0 con una constante Lipschitz asociada C_L, entonces para cualquier $\varepsilon > 0$, existen vecindades N_ε de p_0 y W_ε de x_0, y una función $x : N_\varepsilon \to W_\varepsilon$, tal que, para cualquier $p \in N_\varepsilon$, $x(p)$ es la única solución en W_ε de la inclusión

$$0 \in F(x, p) + \partial\psi_C(x).$$

Más aún, para cada p y q en N_ε, se tiene que

$$||x(p) - x(q)|| \leq (C_L + \varepsilon)||F(x(q), p) - F(x(q), q)||.$$

Notemos que $x(p_0) = x_0$, es decir la única solución de (2.5) es x_0 .

Para la demostración ver [26], página 45, Teorema 2.1.

2.4. Optimización con restricciones en dimensión finita

Basados en [19], presentamos en esta sección algunos resultados de la teoría de la optimización con restricciones en $\mathbb{R}^m$.

DEFINICIÓN 2.23. *Un problema de programación no lineal estándar en dimensiones finitas, consiste en minimizar una función objetivo $f : \mathbb{R}^m \to \mathbb{R}$ sobre un conjunto admisible, que se define como un conjunto convexo cerrado, y funciones de restricción $G_i : \mathbb{R}^m \to \mathbb{R}$, para $i = 1, ..., l$. La formulación para este problema está dada por*

$$(NP) \begin{cases} \min\limits_{u \in \mathcal{U}_{ad}} f(u) \\ \text{sujeto a:} \\ G_i(u) = 0, \quad i = 1, \ldots, k, \\ G_i(u), \leq 0 \quad i = k+1, \ldots, l. \end{cases}$$

En particular, nos interesa el caso en que u esté restringido a una caja, es decir, para algunos $u_a, u_b \in \mathbb{R}^m$, $\mathcal{U}_{ad} = \{u \in \mathbb{R}^m : u_{a,i} \leq u_i \leq u_{b,i}, \quad \forall i = 1, \ldots, m\}$. Para este caso podemos definir las siguientes funciones de restricción

$$G_{i+l}(u) = u_{a,i} - u_i, \quad \forall i = 1, \ldots, m,$$
$$G_{i+l+m}(u) = u_i - u_{b,i}, \quad \forall i = 1, \ldots, m.$$

Así, reescribimos el problema (NP) en su formulación equivalente

$$(NP') \begin{cases} \text{mín } f(u) \\ \text{sujeto a:} \\ G_i(u) = 0, \quad i = 1, \ldots, k, \\ G_i(u) \leq 0, \quad i = k+1, \ldots, l, l+1, \ldots, l+2m. \end{cases}$$

Consideremos ahora K el cono definido por

$$K = \{z \in \mathbb{R}^l : z_i = 0,\ i = 1, ..., k, \quad z_i \geq 0,\ i = k+1, ..., l\}$$

Entonces, las restricciones $G_i(u) = 0,\ i = 1, \ldots, k, \quad G_i(u) \leq 0,\ i = k+l, \ldots, l$ pueden ser expresadas como $G(u) \leq_K 0$, donde $z \leq_K 0$ significa $-z \in K$, y G es el vector definido por $G = [G_1, \ldots, G_l]$. Así, el problema (NP) puede ser escrita de la forma equivalente:

$$(P) \begin{cases} \text{mín } f(u) \\ G(u) \leq_K 0, \\ u \in U_{ad}. \end{cases}$$

DEFINICIÓN 2.24. *Definimos el conjunto factible para el problema (NP), como*

$$\mathcal{U}_{feas} = \{u \in \mathbb{R}^m : u \in \mathcal{U}_{ad} \text{ y } G(u) \leq_K 0\}.$$

Si $u \in \mathcal{U}_{feas}$, decimos que u es factible para el problema (NP).

DEFINICIÓN 2.25. *Decimos que $\bar{u} \in \mathbb{R}^m$ es una solución global del problema (NP) si $\bar{u}$ es factible para (NP) y $f(\bar{u}) \leq f(u)$ para todo u factible. Si se cumple la desigualdad estricta $f(\bar{u}) < f(u)$ para todo $u \neq \bar{u}$ factible para (NP), entonces $\bar{u}$ se dice solución global estricta para (NP).*

DEFINICIÓN 2.26. *$\bar{u} \in \mathbb{R}^m$ es una solución local del problema (NP) si $\bar{u}$ es factible para (NP) y existe una bola abierta $B(\bar{u}, \rho)$ tal que $f(\bar{u}) \leq f(u)$ para todo $u \in B(\bar{u}, \rho)$ factible. Si se cumple la desigualdad estricta $f(\bar{u}) < f(u)$ para todo $u \neq \bar{u}$ factible en $B(\bar{u}, \rho)$, entonces $\bar{u}$ se dice solución local estricta para (NP).*

Análogamente estas definiciones se cumplen para (NP'), considerando el problema equivalente

$$(P') \begin{cases} \text{mín } f(u) \\ G(u) \leq_K 0, \end{cases}$$

con

$$K = \{z \in \mathbb{R}^{l+2m} : z_i = 0,\ i = 1, ..., k, \quad z_i \geq 0,\ i = k+1, ..., l+2m\}. \tag{2.6}$$

PROPOSICIÓN 2.15. *Si $\bar{u} \in \mathbb{R}^m$ es una solución local del problema (NP) y f es convexa entonces $\bar{u}$ es una solución global para (NP). Además si f es estrictamente convexa entonces $\bar{u}$ es una solución global estricta.*

Para obtener las condiciones de optimalidad asumiremos en lo que sigue que f y G son lo suficientemente suaves.

Decimos que una restricción es activa en algún punto factible u si $G_i(u) = 0$, e inactiva en caso contrario.

DEFINICIÓN 2.27. *La función de Lagrange asociada a (NP') está definida por*

$$\mathcal{L}(u, \lambda) = f(u) + \sum_{i=1}^{l+2m} \lambda_i G_i(u).$$

DEFINICIÓN 2.28. *Sea $\mathcal{I} = \{1, \ldots, l+2m\}$. El conjunto de índices $\mathcal{A}(\bar{u})$ de restricciones activas en $\bar{u}$ está definido por*

$$\mathcal{A}(\bar{u}) = \{i \in \mathcal{I} : G_i(\bar{u}) = 0\}.$$

DEFINICIÓN 2.29. *Dado el vector $\bar{u}$ y el conjunto de los índices activos $\mathcal{A}(\bar{u})$, diremos que $\bar{u}$ satisface la condición de calificación de restricciones de independencia lineal (LICQ) si el conjunto de los gradientes activos*

$$\{\nabla G_i(\bar{u}) : i \in \mathcal{A}(\bar{u})\}$$

es linealmente independiente.

DEFINICIÓN 2.30. *Dado el vector $\bar{u}$, diremos que $\bar{u}$ satisface la condición de calificación de restricciones de Mangasarian-Fromovitz (MFCQ) si el conjunto de los gradientes*

$$\{\nabla G_i(\bar{u}) : i = 1, \ldots, k\}$$

es linealmente independiente y existe algún $u \in \mathcal{U}_{ad}$ tal que

$$\begin{aligned} &\nabla G_i(\bar{u})u = 0, \quad \text{para } i = 1, \ldots, k, \\ &\nabla G_i(\bar{u})u < 0, \quad \text{para } i = k+1, \ldots, l+2m, \text{ con } G_i(\bar{u}) = 0. \end{aligned}$$

PROPOSICIÓN 2.16. *En dimensión finita, la condición de regularidad de Robinson*

(2.3) es equivalente a (MFCQ).

Para la demostración ver [27], página 206.

PROPOSICIÓN 2.17. *La condición (LICQ) implica (MFCQ).*

Para la demostración ver [17], página 188.

La demostración del siguiente resultado es una consecuencia directa de la Proposición 2.16 y 2.17.

PROPOSICIÓN 2.18. *La condición (LICQ) implica la condición de regularidad de Robinson.*

Las siguientes condiciones conocidas como Karush-Kuhn-Tucker (o KKT system) expresan las condiciones necesarias de primer orden para el problema (NP').

PROPOSICIÓN 2.19. *Suponemos que $\bar{u}$ es un óptimo local para (NP'), y que (LICQ) se satisface en $\bar{u}$. Entonces existe un vector $\bar{\lambda} \in \mathbb{R}^{l+2m}$, llamado multiplicador de Lagrange, tal que se cumple*

$$\frac{\partial \mathcal{L}}{\partial u}(\bar{u}, \bar{\lambda}) = 0 \tag{2.7}$$

y

$$\begin{aligned} G_i(\bar{u}) &= 0, \quad \text{para } i = 1, \ldots, k, \\ G_i(\bar{u}) &\leq 0, \quad \text{para } i = k+1, \ldots, l+2m, \\ \bar{\lambda}_i &\geq 0, \quad \text{para } i = k+1, \ldots, l+2m, \\ \bar{\lambda}_i G_i(\bar{u}) &= 0, \quad \text{para } i = 1, \ldots, l+2m. \end{aligned} \tag{2.8}$$

En particular, de (2.7) podemos omitir los índices $i \notin \mathcal{A}(\bar{u})$, pues en estos índices $\bar{\lambda}_i = 0$, obteniendo así

$$\nabla_u \mathcal{L}(\bar{u}, \bar{\lambda}) = \nabla f(\bar{u}) + \sum_{i \in \mathcal{A}(\bar{u})} \bar{\lambda}_i \nabla G_i(\bar{u}) = 0.$$

La demostración puede ser encontrada en [22], páginas 321-329 ,Teorema 12.1.

DEFINICIÓN 2.31. *La condición de complementariedad estricta se cumple en $\bar{u}$ si $\bar{u}$ es una solución local que satisface (2.7)-(2.8), y si además satisface la implicación*

$$G_i(\bar{u}) = 0 \Rightarrow \bar{\lambda}_i > 0,$$

para todo $i \in \{k+1, \ldots, l+2m\} \cap \mathcal{A}(\bar{u})$.

En la Sección 4.4 veremos que las condiciones de optimalidad de primer orden (2.7)-(2.8) de (NP') pueden ser expresadas como una ecuación generaliza de la forma

$$0 \in \begin{bmatrix} \nabla_u \mathcal{L}(\bar{u}, \bar{\lambda}) \\ -G(\bar{u}) \end{bmatrix} + \begin{bmatrix} \{0\} \\ \partial \psi_{K_+}(\bar{\lambda}) \end{bmatrix}, \tag{2.9}$$

donde K_+ es el cono dual de K definido en (2.6).

Para las condiciones suficientes de segundo orden, definimos los siguientes conjuntos de índices

$$I_0 = \{i \in \{k+1, \ldots, l, \ldots, l+2m\} : G_i(\bar{u}) = 0\},$$
$$I_+ = \{i \in \{k+1, \ldots, l, \ldots, l+2m\} : \bar{\lambda}_i > 0\} \cap \mathcal{A}(\bar{u}).$$

DEFINICIÓN 2.32 (Cono crítico)**.** *El cono crítico asociado a un vector factible $\bar{u}$ para el problema (NP'), y denotado por $C_{\bar{u}}$, es el conjunto de todos los vectores v en $\mathbb{R}^m$ que satisfacen*

$$\nabla G_i(\bar{u})^T v = 0, \quad \forall i \in \{1, \ldots, k\} \cup I_+,$$
$$\nabla G_i(\bar{u})^T v \leq 0, \quad \forall i \in I_0 \backslash I_+.$$

DEFINICIÓN 2.33. *Decimos que $(\bar{u}, \bar{\lambda})$ satisface la propiedad de coercividad, si se cumple*

$$u^T \nabla_u^2 \mathcal{L}(\bar{u}, \bar{\lambda}) u > 0,$$

para todo $u \in C_{\bar{u}} \backslash \{0\}$.

PROPOSICIÓN 2.20. *Sea $\bar{u}$ un vector factible del problema (NP'), tal que las condiciones de primer orden (2.7)-(2.8) y la propiedad de coercividad se satisface en $\bar{u}$. Entonces existen constantes positivas ω y ε, tales que*

$$f(u) - f(\bar{u}) \geq \omega ||u - \bar{u}||^2$$

se cumple para todo u factible con $||u - \bar{u}|| \leq \varepsilon$.

Para la demostración ver [22], páginas 333-335, Teorema 12.6.

DEFINICIÓN 2.34. *Decimos que $(\bar{u}, \bar{\lambda})$ satisface la condición suficiente de segundo orden, si se cumple*

$$v^T \nabla_u^2 \mathcal{L}(\bar{u}, \bar{\lambda}) v > 0, \quad \forall v \in C_{\bar{u}},\ v \neq 0,$$

donde

$$C_{\bar{u}} = \{v \in \mathbb{R}^m : \nabla G_i(\bar{u})^T v = 0,\ \forall i \in \{1,\dots,k\} \cup \{i \in \{k+1,\dots,l+2m\} : \bar{\lambda}_i > 0\}\}.$$

PROPOSICIÓN 2.21. *Sea f y G_i para $i = 1,\dots,l+2m$ dado en (NP'), los cuales son dos veces diferenciables en el punto $\bar{u} \in \mathcal{U}_{ad}$. Suponemos que $\bar{u}$, junto con $\bar{\lambda} \in \mathbb{R}^{l+2m}$ resuelve (2.9). Si $(\bar{u}, \bar{\lambda})$ satisface la condición suficiente de segundo orden junto con la condición (LICQ), entonces (2.9) es fuertemente regular en $(\bar{u}, \bar{\lambda})$.*

La demostración puede ser encontrada en [26], página 56, Teorema 4.1.

2.5. Un resultado abstracto de la existencia del multiplicador de Lagrange

Los siguiente resultados fueron tomados de [3]. Para lo cual realizamos las siguientes notaciones

- X, W espacios Banach.
- K_1 es un subconjunto convexo, cerrado, no vacío de X.
- K_2 es un subconjunto convexo, cerrado, de interior no vacío de W.
- f una función de X en $\mathbb{R}$.
- g una función de X en W.
- h una función de X en $\mathbb{R}^n$, para $n \in \mathbb{N}$ dado.

Con esto, vamos a considerar el siguiente problema:

$$\begin{cases} \min\limits_{x \in K_1} f(x) \\ \text{sujeto a:} \\ g(x) \in K_2, \\ h(x) = 0. \end{cases} \tag{2.10}$$

La proposición que sigue es muy importante, pues con esta vamos a obtener nuestras condiciones de optimalidad para nuestro problema en estudio.

PROPOSICIÓN 2.22. *Sea $\bar{x}$ una solución de (2.10). Suponemos que f y h son funciones Lipschitz cerca $\bar{x}$ y g una función estrictamente derivable en una vecindad de $\bar{x}$. Entonces existe $\lambda \geq 0$, $\varphi \in W^*$ y $\xi \in \mathbb{R}^n$ tales que*

$$\lambda + ||\varphi||_{W^*} + \sum_{i=1}^{n} |\xi_i| > 0, \tag{2.11}$$

$$\langle \varphi, w - g(\bar{x}) \rangle_{W^*W} \leq 0, \qquad \forall w \in K_2, \tag{2.12}$$

$$0 \in \lambda \partial_c f(\bar{x}) + [D_s g(\bar{x})]^* \varphi + \sum_{i=1}^{n} \xi_i \partial_c h_i(\bar{x}) + \partial I_{K_1}(\bar{x}) \qquad \text{en } X^*. \tag{2.13}$$

OBSERVACIÓN. En la ausencia de las restricciones de igualdad, los resultados anteriores son válidos si se retira los términos correspondientes a dicha restricción.

Para la demostración ver [3], páginas 71-73, Teorema 2.1.

PROPOSICIÓN 2.23. *Si h es estrictamente derivable, $\{\nabla h_i(\bar{x}) : i \in \{1, \ldots, n\}\}$ es linealmente independiente y si existe $x_0 \in \overset{\circ}{K_1}$ tal que*

1. $g(\bar{x}) + D_s g(\bar{x})(x_0 - \bar{x}) \in \overset{\circ}{K_2}$,
2. $\langle \nabla h_i(\bar{x}), x_0 - \bar{x} \rangle = 0$, $\forall i = 1, \ldots, n$,

entonces la conclusión de la Proposición 2.22 es verdadera para $\lambda = 1$.

Para la demostración ver [3], páginas 73-75, Teorema 2.2.

Capítulo 3

Análisis del problema de control óptimo

En este capítulo se plantea el problema de control óptimo gobernado por una ecuación diferencial parcial lineal elíptica con condiciones de frontera Dirichlet homogénea y restricciones finitas sobre el estado, analizamos la existencia y unicidad de la solución óptima y se realizarán estudios sobre las condiciones de optimalidad necesarias y suficientes. Para esto, utilizaremos algunos resultados descritos en el capítulo anterior.

3.1. Planteamiento del problema

Planteamos el siguiente problema de control óptimo:

$$(P) \qquad \begin{cases} \min\limits_{(y,u)} J(y,u) = \dfrac{1}{2}||y - y_d||^2 + \dfrac{\alpha}{2}||u||_2^2 + \beta||u||_1 \\ \text{sujeto a:} \\ Ay(x) = \sum\limits_{i=1}^{M} u_i e_i(x), \quad \text{en } \Omega, \\ y(x) = 0, \qquad\qquad\qquad \text{sobre } \Gamma, \\ y(x_j) \leq b_j, \qquad\qquad\quad \forall j = 1, \ldots, l, \\ u \in \mathcal{U}_{ad}, \end{cases}$$

donde $\Omega \subset \mathbb{R}^2$ es un conjunto convexo, abierto y acotado con frontera Lipschitz poligonal Γ. $y_d \in L^2(\Omega)$ es un estado deseado, con parámetro de Tikhonov $\alpha > 0$, $\beta > 0$, las funciones $e_i \in C^{0,\gamma}(\Omega)$, $i = 1, 2, ..., M$, para algún $0 < \gamma < 1$, se define el

conjunto de los controles admisibles por

$$\mathcal{U}_{ad} = \{u \in \mathbb{R}^M : u_{a,i} \leq u_i \leq u_{b,i}, \quad \forall i = 1, \ldots, M\},$$

con $u_a, u_b \in \mathbb{R}$ con $u_a < u_b$. La ecuación diferencial se basa en un operador diferencial uniformemente elíptico y simétrico definido por

$$Ay(x) := -\sum_{i,j=1}^{2} \partial_j(a_{ij}(x)\partial_i y(x)) + a_0(x)y(x),$$

y cuyos coeficientes $a_{ij} \in C^{1+\delta}(\Omega)$, $0 < \delta < 1$, y $a_0 \in C^{0,\gamma}(\Omega)$ tal que $a_0(x) \geq 0$ en Ω, y la restricción sobre el estado se impone en los puntos $x_j \in \Omega$ con $j = 1, \ldots, l$, cuya cota superior viene dado por el vector $b \in \mathbb{R}^l$.

3.2. Estudio de la ecuación de estado

En esta sección se analizará la existencia, unicidad y regularidad de la solución débil de la ecuación de estado. Para esto nos referimos a [4], [14], [20], tomando algunas definiciones y resultados fundamentales sobre Ecuaciones Diferenciales Parciales elípticas.

DEFINICIÓN 3.1. *Sea $e \in L^2(\Omega)$, decimos que una función $y_e \in H_0^1(\Omega)$ es una solución débil de la ecuación*

$$\begin{cases} Ay(x) = e(x), & \text{en } \Omega, \\ y(x) = 0, & \text{sobre } \Gamma, \end{cases} \tag{3.1}$$

si y_e cumple la ecuación variacional:

$$\sum_{i,j=1}^{2} (a_{ij}\partial_i y_e, \partial_j \phi) + (a_0 y_e, \phi) = (e, \phi)$$

para todo $\phi \in H_0^1(\Omega)$, donde $(\cdot, \cdot)$ representa el producto escalar en $L^2(\Omega)$.

PROPOSICIÓN 3.1. *Para cada función $e \in L^2(\Omega)$, existe una única solución débil $y_e \in H_0^1(\Omega) \cap H^2(\Omega)$ de (3.1) y la función $F : L^2(\Omega) \to H_0^1(\Omega)$ definida por*

$$F(e) = y_e, \quad \forall e \in L^2(\Omega),$$

es lineal y continua. Más aún, si $e \in C^{0,\gamma}(\Omega)$, entonces $y_e \in C^{2,\gamma}(\Omega)$.

Demostración. Para la existencia y unicidad ver [14] [theorem 3, p. 301] y para el resultado de regularidad ver [15],[16], pues Ω es convexo. □

El siguiente resultado muestra la existencia, unicidad y regularidad de la solución débil para la ecuación de estado.

PROPOSICIÓN 3.2. *Para cada $u \in \mathbb{R}^M$, la ecuación de estado tiene una única solución débil $y_u \in H_0^1(\Omega) \cap H^2(\Omega) \cap C^{2,\gamma}(\Omega)$, es decir y_u satisface*

$$\sum_{i,j=1}^{2} (a_{ij}\partial_i y_u, \partial_j \phi) + (a_0 y_u, \phi) = (\sum_{i=1}^{M} u_i e_i, \phi), \tag{3.2}$$

para todo $\phi \in H_0^1(\Omega)$.

Demostración. Dividimos en M subproblemas definidos como en (3.1), es decir para $i = 1, \ldots, M$, consideramos la ecuación

$$\begin{cases} Ay(x) = e_i(x), & \text{en } \Omega, \\ y(x) = 0, & \text{sobre } \Gamma, \end{cases}$$

por la Proposición 3.1 existe una única solución débil $y_i \in H_0^1(\Omega) \cap H^2(\Omega) \cap C^{2,\gamma}(\Omega)$ para cada uno de estos subproblemas. Entonces si aplicamos el principio de superposición se sigue que

$$y_u = \sum_{i=1}^{M} u_i y_i$$

es una única solución débil de la ecuación de estado, para todo $u \in \mathbb{R}^M$. □

3.3. Existencia de una solución para el problema (P)

Para el análisis de la existencia de una solución óptima, vamos a replantear nuestro problema de control óptimo (P) como un problema de optimización solamente en términos de u. Para esto definiremos el operador control-estado.

DEFINICIÓN 3.2. *Sea S el operador control-estado definido por*

$$\begin{array}{rccl} S: & \mathbb{R}^M & \to & C(\bar{\Omega}) \\ & u & \mapsto & y_u = Su = \sum_{i=1}^{M} u_i y_i, \end{array}$$

donde y_i es la solución débil de (3.1) con $e = e_i$, para $i = 1, \ldots, M$.

Claramente, de la Proposición 3.2, Su es la solución de la ecuación de estado, para $u \in \mathbb{R}^M$.

OBSERVACIÓN. El operador S es lineal y continuo.

Observemos que si reemplazamos $y = Su$ en el funcional de costo J y denotamos $f(u) = J(Su, u)$, nos queda el siguiente problema

$$\begin{cases} \min\limits_{u \in \mathcal{U}_{ad}} f(u) = \dfrac{1}{2} \left|\left| \sum\limits_{i=1}^{M} u_i y_i - y_d \right|\right|^2 + \dfrac{\alpha}{2} ||u||_2^2 + \beta ||u||_1 \\ \text{sujeto a:} \\ \sum\limits_{i=1}^{M} u_i y_i(x_j) \leq b_j, \quad \forall j = 1, \ldots, l. \end{cases} \tag{3.3}$$

Para tomar en cuenta las restricciones de estado definimos el conjunto factible

$$\mathcal{U}_{feas} = \left\{ u \in \mathcal{U}_{ad} : \sum_{i=1}^{M} u_i y_i(x_j) \leq b_j, \ \forall j = 1, \ldots, l \right\}.$$

Además, para nuestro estudio hacemos la siguiente suposición

HIPÓTESIS 3.3. *$\mathcal{U}_{feas}$ es un conjunto no vacío.*

Verifiquemos ahora, los siguientes resultados

PROPOSICIÓN 3.4. *$\mathcal{U}_{ad}$ es un conjunto convexo cerrado de $\mathbb{R}^M$.*

Demostración. Dados $u, v \in \mathcal{U}_{ad}$ y $t \in [0, 1]$, se tiene que

$$\begin{gathered} t u_{a,i} \leq t u_i \leq t u_{b,i}, \quad \text{y} \\ (1-t) u_{a,i} \leq (1-t) v_i \leq (1-t) u_{b,i}, \end{gathered}$$

para todo $i = 1, \ldots, M$. Al sumar estas dos desigualdades, obtenemos la siguiente desigualdad

$$u_{a,i} \leq t u_i + (1-t) v_i \leq u_{b,i}, \quad \forall i = 1, \ldots, M,$$

por tanto $tu + (1-t)v \in \mathcal{U}_{ad}$, es decir $\mathcal{U}_{ad}$ es un conjunto convexo.

Ahora demostremos que $\mathcal{U}_{ad}$ es un conjunto cerrado, sea $(u_n)_{n \in \mathbb{N}}$ en $\mathcal{U}_{ad}$ tal que $u_n \to u$, verifiquemos que $u \in \mathcal{U}_{ad}$. Puesto que $u_n \in \mathcal{U}_{ad}$ para todo $n \in \mathbb{N}$, se sigue que

$$u_{a,i} \leq u_{i,n} \leq u_{b,i}, \quad \forall i = 1, \ldots, M,$$

y tomando el límite cuando $n \to +\infty$, se obtiene que

$$u_{a,i} \leq u_i \leq u_{b,i}, \quad \forall i = 1, \ldots, M.$$

Esto implica que $u \in \mathcal{U}_{ad}$. □

PROPOSICIÓN 3.5. *$\mathcal{U}_{feas}$ es un subconjunto cerrado, convexo y acotado de $\mathbb{R}^M$.*

Demostración. Sea $(u_n)_{n\in\mathbb{N}}$ una sucesión de $\mathcal{U}_{feas}$ tal que $u_n \to u$ en $\mathbb{R}^M$, puesto que $\mathcal{U}_{ad}$ es cerrado se tiene que $u \in \mathcal{U}_{ad}$ y para $i = 1, ..., M$ se tiene que $u_{i,n} \to u_i$ en $\mathbb{R}$. Por otra parte, puesto que $u_n \in \mathcal{U}_{feas}$ para todo $n \in \mathbb{N}$ se sigue que

$$\sum_{i=1}^{M} u_{i,n} y_i(x_j) \leq b_j, \quad \forall j = 1, \ldots, M,$$

por lo cual al tomar el límite cuando $n \to +\infty$, por continuidad obtenemos

$$\sum_{i=1}^{M} u_i y_i(x_j) \leq b_j, \quad \forall j = 1, \ldots, M.$$

Es decir $u \in \mathcal{U}_{feas}$ y por consiguiente hemos verificado que $\mathcal{U}_{feas}$ es un conjunto cerrado. Análogo a la Proposición 3.4 se demuestra la convexidad de $\mathcal{U}_{feas}$.

Puesto que $\mathcal{U}_{ad}$ es acotado se concluye que $\mathcal{U}_{feas}$ también lo es. □

A continuación demostramos la existencia y unicidad de la solución óptima para (P) utilizando su forma reducida.

PROPOSICIÓN 3.6. *Existe una única solución óptima para el problema (P).*

Demostración. Se tiene que f es una función continua definida en un espacio de dimensión finita y que $\mathcal{U}_{feas}$ es cerrado y acotado. Entonces, por el teorema de Weierstrass, existe $\bar{u} \in \mathcal{U}_{ad}$ control óptimo para el problema (3.3) y de la Proposición 2.15 se sigue que $\bar{u}$ es único, pues f es estrictamente convexa. Luego, gracias al operador control estado se concluye que $(S\bar{u}, \bar{u})$ es una solución óptima para (P). □

OBSERVACIÓN. Si $\bar{u}$ es el control óptimo de (3.3), su estado asociado viene dado por

$$\bar{y} = \sum_{i=1}^{M} \bar{u}_i y_i.$$

3.4. Condiciones de optimalidad

El siguiente paso en nuestro análisis consiste en derivar las condiciones necesarias y suficientes del problema (P), para lo cual aplicaremos la Proposición 2.22 y 2.23 a nuestro caso. Empecemos realizando la siguiente hipótesis

$$\begin{aligned} g_j(u) &= \sum_{i=1}^{M} u_i y_i(x_j) - b_j, && \text{para } j = 1, \dots, l, \\ g_{l+i}(u) &= u_{a,i} - u_i, && \text{para } i = 1, \dots, M, \\ g_{l+M+i}(u) &= u_i - u_{b,i}, && \text{para } i = 1, \dots, M, \end{aligned} \tag{3.4}$$

y denotamos g a la función cuyas componentes son g_i para $i = 1, \dots, l + 2M$.

Si consideramos $\bar{u}$ la solución óptima de (3.3), y para el posterior análisis realizamos las siguientes hipótesis.

HIPÓTESIS 3.7. *Para las restricciones dadas por g, la condición (LICQ) se satisface en $\bar{u}$.*

HIPÓTESIS 3.8 (Condición de Slater)**.** *Existe un control $u^\circ \in \operatorname{int} \mathcal{U}_{ad}$ tal que*

$$\sum_{i=1}^{M} u_i^\circ y_i(x_j) < b_j, \quad \forall j = 1, \dots, l.$$

DEFINICIÓN 3.3. *La parte positiva a^+ y la parte negativa a^- de un número real a son dos números reales no negativos definidos por*

$$\begin{aligned} a^+ &= \max(a, 0), \\ a^- &= -\min(a, 0). \end{aligned}$$

La siguiente Proposición describe las condiciones que satisface la solución óptima de nuestro problema.

PROPOSICIÓN 3.9. *Sea $(\bar{u}, \bar{y})$ la solución óptima para el problema (P) entonces existen $v \in \mathbb{R}^l$ y $\mathbf{w} \in \mathbb{R}^{2M}_+$ tales que*

$$\begin{aligned} \sum_{j=1}^{l} v_j(\bar{y}(x_j) - b_j) &= 0, \\ \mathbf{w}_i(u_{a,i} - \bar{u}_i) &= 0, \\ \mathbf{w}_{i+M}(\bar{u}_i - u_{b,i}) &= 0, \end{aligned}$$

$$\int_\Omega (\bar{y} - y_d) y_i \, dx + \alpha \bar{u}_i + \sum_{j=1}^{l} v_j y_i(x_j) - (\mathbf{w}_i - \mathbf{w}_{i+M}) = -\beta, \quad si\ \bar{u}_i > 0,$$

$$\int_\Omega (\bar{y} - y_d) y_i \, dx + \alpha \bar{u}_i + \sum_{j=1}^{l} v_j y_i(x_j) - (\mathbf{w}_i - \mathbf{w}_{i+M}) = \beta, \quad si\ \bar{u}_i < 0,$$

$$\left| \int_\Omega (\bar{y} - y_d) y_i \, dx + \alpha \bar{u}_i + \sum_{j=1}^{l} v_j y_i(x_j) - (\mathbf{w}_i - \mathbf{w}_{i+M}) \right| \leq \beta, \quad si\ \bar{u}_i = 0,$$

para todo $i = 1, \ldots, M$.

Demostración. Primero procedemos a verificar las hipótesis de la Proposición 2.22. Para esto, replanteamos nuestro problema de la siguiente manera:

$$(P) \qquad \begin{cases} \min\limits_{u \in \mathcal{U}_{ad}} f(u) := \dfrac{1}{2} \left|\left| \sum\limits_{i=1}^{M} u_i y_i - y_d \right|\right|^2 + \dfrac{\alpha}{2} ||u||_2^2 + \beta ||u||_1 \\ \text{sujeto a:} \\ \sum\limits_{i=1}^{M} u_i y_i(x_j) - b_j \leq 0, \quad \forall j = 1, \ldots, l. \end{cases}$$

Ahora definimos la función $g : \mathbb{R}^M \to \mathbb{R}^l$, como

$$u \mapsto g(u) = \left[\sum_{i=1}^{M} u_i y_i(x_j) - b_j \right]_{j=1}^{l}$$

cuya derivada viene representada por la matriz Jacobiana:

$$Dg(\bar{u}) = \begin{pmatrix} y_1(x_1) & y_2(x_1) & \cdots & y_M(x_1) \\ y_1(x_2) & y_2(x_2) & \cdots & y_M(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ y_1(x_l) & y_2(x_l) & \cdots & y_M(x_l) \end{pmatrix}.$$

Puesto que la derivada no depende de u, se sigue que g es estrictamente diferenciable en una vecindad $\bar{u}$ y se cumple que $D_s g(\bar{u}) = Dg(\bar{u})$ (ver la Proposición 2.11). Por otra parte, expresamos la función objetivo de la forma $f(u) = F(u) + \beta f_1(u)$, donde

$$F(u) = \frac{1}{2} \left|\left| \sum_{i=1}^{M} u_i y_i - y_d \right|\right|^2 + \frac{\alpha}{2} ||u||_2^2 \quad \text{y} \quad f_1(u) = ||u||_1.$$

Usando el hecho de que F y f_1 son funciones convexas, se sigue que f es una función convexa y gracias a la Proposición 2.10, se tiene que f es Lipschitz cerca de todo punto $u \in \mathbb{R}^M$, en particular de $\bar{u}$. Por tanto, cumplimos con las hipótesis de la

Proposición 2.22 y puesto que no tenemos restricciones de igualdad, se tiene que existe $\lambda \geq 0$ y $v \in \mathbb{R}^l$ tal que

$$\lambda + ||v||_2 > 0, \tag{3.5}$$

$$\langle v, z - g(\bar{u})\rangle \leq 0, \qquad \forall z \in \mathbb{R}^{l-}, \tag{3.6}$$

$$0 \in \lambda \partial_c f(\bar{u}) + [D_s g(\bar{u})]^* v + \partial I_{\mathcal{U}_{ad}}(\bar{u}), \qquad \text{en } \mathbb{R}^M. \tag{3.7}$$

Puesto que $\mathcal{U}_{ad}$ es un conjunto convexo cerrado, por la Proposición 2.6 se sigue que $\partial I_{\mathcal{U}_{ad}}(\bar{u}) = N_{\mathcal{U}_{ad}}(\bar{u}) = \{\omega \in \mathbb{R}^M : \langle \omega, u - \bar{u}\rangle \leq 0,\ \forall u \in \mathcal{U}_{ad}\}$. Adicionalmente, por la Proposición 2.9 se obtiene que $\partial_c f(\bar{u}) = \partial f(\bar{u})$, por lo tanto (3.7) puede ser expresada como

$$0 \in \lambda \partial f(\bar{u}) + [Dg(\bar{u})]^* v + N_{\mathcal{U}_{ad}}(\bar{u}),$$

además, por las Proposiciones 2.5, 2.7 y 2.8 se tiene que $\partial f(\bar{u}) = \nabla F(\bar{u}) + \beta \partial f_1(\bar{u})$ de donde obtenemos

$$0 \in \lambda \left(\nabla F(\bar{u}) + \beta \partial f_1(\bar{u})\right) + [Dg(\bar{u})]^* v + N_{\mathcal{U}_{ad}}(\bar{u}). \tag{3.8}$$

Luego, gracias a la Hipótesis 3.8 se cumple la Proposición 2.23, en consecuencia $\lambda = 1$ y reemplazando en (3.5), (3.6) y (3.8) se obtiene las siguientes condiciones:

$$\langle v, z - g(\bar{u})\rangle \leq 0, \qquad \forall z \in \mathbb{R}^{l-}, \tag{3.9}$$

$$0 \in \nabla F(\bar{u}) + \beta \partial f_1(\bar{u}) + [D_s g(\bar{u})]^* v + N_{\mathcal{U}_{ad}}(\bar{u}), \qquad \text{en } \mathbb{R}^M. \tag{3.10}$$

De (3.9) se tiene

$$\sum_{j=1}^{l} v_j (z_j - \bar{y}(x_j) + b_j) \leq 0, \quad \forall z \in \mathbb{R}^{l-},$$

en particular tomando $z = 0$ y $z = [2(\bar{y}(x_j) - b_j)]_{j=1}^{l}$, se obtiene que

$$\sum_{j=1}^{l} v_j (\bar{y}(x_j) - b_j) = 0. \tag{3.11}$$

La condición (3.10) es equivalente a

$$-\nabla F(\bar{u}) - [D_G g(\bar{u})]^* v - \omega \in \beta \partial f_1(\bar{u}),$$

para algún $\omega \in N_{\mathcal{U}_{ad}}(\bar{u})$. Por otra parte, sabemos que

$$\nabla F(\bar{u}) = [(\bar{y} - y_d, y_i) + \alpha \bar{u}_i]_{i=1}^{M},$$

y que

$$[D_G g(\bar{u})]^* v = [D_G g(\bar{u})]^T v = \left[\sum_{j=1}^{l} v_j, y_i(x_j) \right]_{i=1}^{M},$$

entonces

$$-[(\bar{y} - y_d, y_i) + \alpha \bar{u}_i]_{i=1}^{M} - \left[\sum_{j=1}^{l} v_j y_i(x_j) \right]_{i=1}^{M} - \omega \in \beta \partial f_1(\bar{u}) \quad \text{en } \mathbb{R}^M.$$

Luego, por la definición de subgradiente, tenemos

$$-(\bar{y} - y_d, y_i) - \alpha \bar{u}_i - \sum_{j=1}^{l} v_j y_i(x_j) - \omega_i \in \begin{cases} \{\beta\}, & \text{si } \bar{u}_i > 0, \\ \{-\beta\}, & \text{si } \bar{u}_i < 0, \\ [-\beta, \beta], & \text{si } \bar{u}_i = 0, \end{cases}$$

de lo cual se obtienen las relaciones:

$$\begin{aligned} (\bar{y} - y_d, y_i) + \alpha \bar{u}_i + \sum_{j=1}^{l} v_j y_i(x_j) + \omega_i &= -\beta, \quad \text{si } \bar{u}_i > 0, \\ (\bar{y} - y_d, y_i) + \alpha \bar{u}_i + \sum_{j=1}^{l} v_j y_i(x_j) + \omega_i &= \beta, \quad \text{si } \bar{u}_i < 0, \\ \left| (\bar{y} - y_d, y_i) + \alpha \bar{u}_i + \sum_{j=1}^{l} v_j y_i(x_j) + \omega_i \right| &\leq \beta, \quad \text{si } \bar{u}_i = 0, \end{aligned} \tag{3.12}$$

para todo $i = 1, \ldots, M$, donde $\omega \in \mathbb{R}^M$ tal que

$$\sum_{i=1}^{M} \omega_i (u_i - \bar{u}_i) \leq 0, \quad \forall u \in \mathcal{U}_{ad},$$

en particular, para todo $u_{a,i} \leq u_i \leq u_{b,i}$ se tiene que

$$\omega_i (u_i - \bar{u}_i) \leq 0, \tag{3.13}$$

para todo $i = 1, \ldots, M$. Si consideramos la parte positiva y la parte negativa de ω, podemos definir el vector $\mathbf{w} \in \mathbb{R}^{2M}_+$ como

$$\begin{aligned} \mathbf{w}_i &= \omega_i^-, \quad \forall i = 1, \ldots, M, \\ \mathbf{w}_{i+M} &= \omega_i^+, \quad \forall i = 1, \ldots, M. \end{aligned}$$

Con esto probaremos que

$$\begin{aligned} \mathbf{w}_i (u_{a,i} - \bar{u}_i) &= 0, \\ \mathbf{w}_{i+M} (\bar{u}_i - u_{b,i}) &= 0. \end{aligned} \tag{3.14}$$

En efecto, analizaremos los siguientes tres casos:

- Si $u_{a,i} < \bar{u}_i < u_{b,i}$, entonces de (3.13) se sigue que

  $$\omega_i(u_i - \bar{u}_i) = 0,$$

  para todo $u_{a,i} < u_i < u_{b,i}$. En particular, si tomamos $u_i \neq \bar{u}_i$ obtenemos que $\omega_i = 0$, en consecuencia se tiene que $\mathbf{w}_i = 0$ y $\mathbf{w}_{i+M} = 0$.

- Si $\bar{u}_i = u_{a,i}$, entonces de (3.13) se tiene que

  $$\omega_i(u_{b,i} - \bar{u}_i) \leq 0,$$

  esto implica que $\omega_i \leq 0$, por lo tanto se sigue que $\mathbf{w}_i = -\omega_i$ y $\mathbf{w}_{i+M} = 0$, demostrando así (3.14).

- Si $\bar{u}_i = u_{b,i}$, entonces de (3.13) se sigue que

  $$\omega_i(u_{a,i} - \bar{u}_i) \leq 0,$$

  lo que implica que $\omega_i \geq 0$, en consecuencia se obtiene que $\mathbf{w}_i = 0$ y $\mathbf{w}_{i+M} = \omega_i$, con lo cual se prueba (3.14).

Luego, gracias a (3.12) y puesto que $\omega_i = -(\mathbf{w}_i - \mathbf{w}_{i+M})$ se obtiene las siguientes relaciones

$$\begin{aligned}
&(\bar{y} - y_d, y_i) + \alpha\bar{u}_i + \sum_{j=1}^{l} v_j y_i(x_j) - (\mathbf{w}_i - \mathbf{w}_{i+M}) = -\beta, \quad \text{si } \bar{u}_i > 0, \\
&(\bar{y} - y_d, y_i) + \alpha\bar{u}_i + \sum_{j=1}^{l} v_j y_i(x_j) - (\mathbf{w}_i - \mathbf{w}_{i+M}) = \beta, \qquad \text{si } \bar{u}_i < 0, \\
&\left| (\bar{y} - y_d, y_i) + \alpha\bar{u}_i + \sum_{j=1}^{l} v_j y_i(x_j) - (\mathbf{w}_i - \mathbf{w}_{i+M}) \right| \leq \beta, \quad \text{si } \bar{u}_i = 0,
\end{aligned} \tag{3.15}$$

Con la deducción de (3.11), (3.14) y (3.15) hemos obtenido las condiciones de optimalidad necesarias para el problema (P). □

OBSERVACIÓN. Hay que enfatizar que dichas condiciones también son suficientes, en virtud de la convexidad de (P).

Capítulo 4

Estimación de errores

El objetivo de este capítulo es obtener el orden de la estimación del error para la aproximación del problema de control óptimo. Para esto, primero es necesario estimar el error para la solución aproximada de la ecuación discreta por el método de elementos finitos. Luego, mediante la discretización de la ecuación de estado se define el problema de control discreto, el cual es formulado como un problema equivalente de control, cuyo funcional de costo es diferenciable. Este análisis es fundamental, pues nos servirá como herramienta para alcanzar el objetivo propuesto, en vista de que dicha equivalencia nos permitirá aplicar resultados conocidos para problemas diferenciables.

4.1. Discretización de la ecuación de estado

En esta sección realizamos la aproximación numérica de la ecuación de estado utilizando el método de elementos finitos. El siguiente paso es el de estimar los errores de la solución de dicha ecuación. Las definiciones y resultados presentados a continuación son tomados de [8] y [24].

Consideramos una familia de mallas $(\mathcal{T}_h)_{h>0}$ de $\bar{\Omega}$ que consisten de triángulos $T \in \mathcal{T}_h$ tales que

- $\bigcup_{T\in\mathcal{T}_h} T = \bar{\Omega}$ y
- 2 triángulos T_i y T_j, $i \neq j$ comparten un vértice, un lado o son disjuntos.

Para cada triángulo $T \in \mathcal{T}_h$, denotamos $\rho(T)$ al diámetro del conjunto T, y $\sigma(T)$ al diámetro de la bola más grande contenida en T. El tamaño de la malla h lo definimos

como

$$h = \max_{T \in \mathcal{T}_h} \rho(T),$$

y asumimos la siguiente condición de regularidad sobre la malla:

HIPÓTESIS 4.1. *Existen dos constantes positivas ρ y σ tales que*

$$\frac{\rho(T)}{\sigma(T)} \leq \sigma \quad y \quad \frac{h}{\rho(T)} \leq \rho, \qquad \forall T \in \mathcal{T}_h,$$

para todo $h > 0$.

DEFINICIÓN 4.1. *Definimos el espacio (asociado con la triangulación T_h)*

$$Y_h = \{y_h \in C(\bar{\Omega}) : y_h|_T \in P_1(T)\ \forall T \in \mathcal{T}_h,\ y_h = 0 \text{ sobre } \Gamma\},$$

donde $P_1(T)$ denota el conjunto de las funciones afines a valores reales definidas sobre T.

Adicionalmente, para cada $i = 1, \ldots, M$, se define el estado discreto y_i^h, como el elemento de Y_h que satisface la ecuación

$$\sum_{j,k=1}^{2} (a_{jk} \partial_j y_i^h, \partial_k \phi_h) + (a_0 y_i^h, \phi_h) = (e_i, \phi_h), \tag{4.1}$$

para todo $\phi_h \in Y_h$.

OBSERVACIÓN. Y_h es un subespacio de dimensión finita de $H_0^1(\Omega)$.

PROPOSICIÓN 4.2. *Para cada $i = 1, \ldots, M$, la ecuación (4.1) tiene una única solución $y_i^h \in Y_h$.*

Demostración. Sea $\{\varphi_1, \ldots, \varphi_N\}$ una base del subespacio Y_h. Se tiene que si $y_i^h \in Y_h$ existen $u_s^i,\ s = 1, \ldots, N$, tales que

$$y_i^h(x) = \sum_{s=1}^{N} u_s^i \varphi_s(x).$$

Entonces si tomamos $\phi_h = \varphi_l$ en (4.1), el problema se reduce a encontrar N escalares $\{u_1^i, \ldots, u_N^i\}$ tales que

$$\sum_{s=1}^{N} u_s^i a(\varphi_s, \varphi_l) = (e_i, \varphi_l), \quad \forall l = 1, \ldots, N, \tag{4.2}$$

donde $a(\cdot,\cdot)$ es el operador bilineal continuo y coercivo, definido por

$$a(\varphi_s, \varphi_l) = \sum_{j,k=1}^{2} (a_{jk}\partial_j\varphi_s, \partial_k\varphi_l) + (a_0\varphi_s, \varphi_l).$$

Entonces el problema (4.2) es equivalente a resolver el sistema lineal

$$\mathbf{A}u^i = \mathbf{e}_i,$$

donde $\mathbf{A} = (a_{sl})_{N\times N}$, con $a_{sl} = a(\varphi_s, \varphi_l)$, y el lado derecho $\mathbf{e}_i = [(e_i, \varphi_l)]_{l=1}^N$. Luego, puesto que $a(\cdot,\cdot)$ es bilineal y coercivo se tiene que $\mathbf{A}$ es una matriz definida positiva por tanto es invertible, lo cual implica que el sistema tenga una única solución. □

Notemos por Ω_0 un subdominio de Ω que contiene a x_j para todo $j = 1, \ldots, l$. Con el cual se presenta el siguiente resultado de aproximación.

PROPOSICIÓN 4.3. *Existe una constante $c > 0$ que no depende de h tal que*

$$||y_i - y_i^h|| \leq ch^2, \tag{4.3}$$

$$||y_i - y_i^h||_{L^\infty(\Omega)} \leq ch, \tag{4.4}$$

$$||y_i - y_i^h||_{L^\infty(\Omega_0)} \leq ch^2|\log h|, \tag{4.5}$$

para $i = 1, \ldots, M$.

Para la demostración de (4.3) y (4.5) ver [20][Proposición 3.3], y la estimación (4.4) se encuentra en [21][Teorema 3.1].

Y para nuestra ecuación de estado, el estado discreto está definido como el único elemento y_u^h de Y_h que satisface la ecuación

$$\sum_{j,k=1}^{2} (a_{jk}\partial_j y_u^h, \partial_k\phi_h) + (a_0 y_u^h, \phi_h) = \left(\sum_{i=1}^{M} u_i e_i, \phi_h\right), \quad \forall \phi \in Y_h, \tag{4.6}$$

fijado cualquier $u \in \mathbb{R}^M$.

PROPOSICIÓN 4.4. *El estado discreto de (4.6), viene dado por*

$$y_u^h = \sum_{i=1}^{M} u_i y_i^h, \tag{4.7}$$

para todo $u \in \mathcal{U}_{ad}$.

Demostración. De la Proposición 4.2 se tiene existencia y unicidad de y_i^h solución de

la ecuación (4.1), para cada $i = 1, \ldots, M$. Por lo tanto el resultado es inmediato de la linealidad de la ecuación (4.6). □

Denotamos $y_u = \sum_{i=1}^{M} u_i y_i$ a la solución débil de la ecuación de estado (3.2) asociado al control $u \in \mathcal{U}_{ad}$. Luego, si a las estimaciones obtenidas en la Proposición 4.3 aplicamos la desigualdad triangular y utilizamos la acotación de $\mathcal{U}_{ad}$, obtenemos el siguiente resultado

PROPOSICIÓN 4.5. *Para todo $u \in \mathcal{U}_{ad}$, se cumple que*

$$||y_u - y_u^h|| \leq Ch^2, \tag{4.8}$$

$$||y_u - y_u^h||_{L^\infty(\Omega)} \leq Ch, \tag{4.9}$$

$$||y_u - y_u^h||_{L^\infty(\Omega_0)} \leq Ch^2|\log h|, \tag{4.10}$$

para alguna un constante $C > 0$ independiente de h y u.

Puesto que $y_u - y_u^h$ es una función continua en todo Ω, se tiene entonces que la norma $L^\infty(\Omega)$ coincide con la norma $C(\bar{\Omega})$, por lo tanto

$$||y_i - y_i^h||_{C(\bar{\Omega})} \leq ch, \tag{4.11}$$

$$||y_i - y_i^h||_{C(\bar{\Omega}_0)} \leq ch^2|\log h|, \tag{4.12}$$

$$||y_u - y_u^h||_{C(\bar{\Omega})} \leq Ch, \tag{4.13}$$

$$||y_u - y_u^h||_{C(\bar{\Omega}_0)} \leq Ch^2|\log h|. \tag{4.14}$$

De (4.11), se tiene que

$$||y_i^h||_{C(\bar{\Omega})} - ||y_i||_{C(\bar{\Omega})} \leq ch,$$

lo que implica

$$||y_i^h||_{C(\bar{\Omega})} \leq C, \tag{4.15}$$

para h suficientemente pequeño.

PROPOSICIÓN 4.6. *Existe una constante $C > 0$ tal que*

$$||y_u||_{C(\bar{\Omega})} \leq C, \tag{4.16}$$

para todo $u \in \mathcal{U}_{ad}$.

Demostración. Utilizando la desigualdad triangular y puesto que $\mathcal{U}_{ad}$ es acotado se tiene el resultado. □

PROPOSICIÓN 4.7. *Para todo $u \in \mathcal{U}_{ad}$ y h suficientemente pequeño, existe una constante $c > 0$ independiente de h, tal que*

$$||y_u^h||_{C(\bar{\Omega})} \leq c. \tag{4.17}$$

Demostración. Usando la desigualdad triangular en (4.13) se tiene

$$||y_u^h||_{C(\bar{\Omega})} - ||y_u||_{C(\bar{\Omega})} \leq Ch,$$

entonces

$$\begin{aligned} ||y_u^h||_{C(\bar{\Omega})} &\leq Ch + ||y_u||_{C(\bar{\Omega})}, \\ &\leq c. \end{aligned}$$

□

4.2. Problema de control equivalente

En vista de que la norma ℓ_1 no es diferenciable, las condiciones de primer orden resultan ser muy abstractas, más aún si hablamos de las condiciones de segundo orden, pues de estas no poseemos información alguna. Esto representa una dificultad en el análisis de la estimación del error para nuestro problema y de aquí la importancia de esta sección.

El objetivo de esta sección es plantear un problema de control óptimo cuyo funcional de costo sea diferenciable y que a su vez sea equivalente a nuestro problema reducido (3.3). El estudio que se realizará para este nuevo problema nos permitirá demostrar que la nueva solución óptima satisface condiciones de optimalidad de primer y segundo orden, lo cual es crucial en la estimación del error para la aproximación del control óptimo del problema estudiado en las secciones anteriores. Aplicaremos la técnica de descomposición descrita en [29][Página 9] en el contexto del control óptimo de ecuaciones diferenciales parciales, que consiste en separar las partes positiva y negativa del control.

Tomando en cuenta la Definición 3.3, empecemos esta sección indicando la relación que existe entre un número y sus partes positiva y negativa, la cual está dada por:

$$a = a^+ - a^-, \tag{4.18}$$

$$|a| = a^+ + a^-. \tag{4.19}$$

Además, satisfacen la siguiente propiedad

$$a^+ a^- = 0. \tag{4.20}$$

Es decir a^+ y a^- son ortogonales. Tenemos la siguiente caracterización.

PROPOSICIÓN 4.8. *Sean $a \in \mathbb{R}$ y a_1, a_2 dos números no negativos, tales que*

$$a = a_1 - a_2 \quad y \quad a_1 a_2 = 0,$$

entonces $a_1 = a^+$ y $a_2 = a^-$.

Demostración. Puesto que $a_1 a_2 = 0$ se tiene que $a_1 = 0$ o $a_2 = 0$, y analizamos los siguientes casos:

1. Si $a = 0$ entonces $a_1 = a_2 = 0$ y $a^+ = a^- = 0$, por lo tanto $a_1 = a^+$ y $a_2 = a^-$.
2. Si $a > 0$ se tiene que $a^+ = a = a_1 - a_2$, entonces la única opción es que $a_2 = 0$, caso contrario $a^+ < 0$, por lo tanto $a^+ = a_1$ y $a_2 = a^- = 0$.
3. Si $a < 0$ se tiene que $-a^- = a = a_1 - a_2$, entonces la única opción es que $a_1 = 0$, caso contrario $-a^- > 0$, por lo tanto $-a^- = -a_2$ y $a_1 = a^+ = 0$, es decir $a^- = a_2$ y $a_1 = a^+$. □

DEFINICIÓN 4.2. *La parte positiva u^+ de un vector $u \in \mathbb{R}^M$ está definido por el vector*

$$u^+ = [u_i^+]_{i=1}^M,$$

similarmente, la parte negativa de u está definida como

$$u^- = [u_i^-]_{i=1}^M.$$

De (4.18) se tiene que cualquier $u \in \mathbb{R}^M$ puede ser expresado en términos de u^+ y u^- como

$$u = u^+ - u^-. \tag{4.21}$$

Además, por (4.19) se obtiene que

$$||u||_1 = \sum_{i=1}^{M} (u_i^+ + u_i^-).$$

Por otra parte, de la propiedad de ortogonalidad (4.20) se concluye

$$\begin{aligned} ||u||_2^2 &= \sum_{i=1}^{M} u_i^2 \\ &= \sum_{i=1}^{M} (u_i^+ - u_i^-)^2 \\ &= \sum_{i=1}^{M} \left[(u_i^+)^2 + (u_i^-)^2 - 2u_i^+ u_i^- \right] \\ &= \sum_{i=1}^{M} \left[u_i^{+2} + u_i^{-2} \right]. \end{aligned}$$

Ahora, recordemos que el conjunto de los controles admisibles estaba definido por

$$\mathcal{U}_{ad} = \{u \in \mathbb{R}^M : u_{a,i} \leq u_i \leq u_{b,i}, \quad \forall i = 1, \ldots, M\},$$

con lo cual definimos los siguiente conjuntos

$$\mathcal{U}_{ad}^+ = \{u \in \mathbb{R}^M : u_{a,i}^+ \leq u_i \leq u_{b,i}^+, \quad \forall i = 1, \ldots, M\}$$

y

$$\mathcal{U}_{ad}^- = \{u \in \mathbb{R}^M : -u_{a,i}^- \leq -u_i \leq -u_{b,i}^-, \quad \forall i = 1, \ldots, M\}.$$

La relación que existe entre estos dos conjuntos y $\mathcal{U}_{ad}$, se explica en la siguiente proposición.

PROPOSICIÓN 4.9. *La siguiente identidad de conjuntos es cierta:*

$$\mathcal{U}_{ad} = \mathcal{U}_{ad}^+ - \mathcal{U}_{ad}^-.$$

Para la demostración de la Proposición 4.9 vamos a utilizar el siguiente Lema.

LEMA 4.10. *Sean $a, b \in \mathbb{R}$ tal que $a \leq b$, entonces $a^+ \leq b^+$ y $-a^- \leq -b^-$.*

Demostración. Analizamos los siguiente casos:

1. Si $a \geq 0$ y $b \geq 0$ entonces $a = a^+$ y $b = b^+$ y por hipótesis se obtiene que

$$a^+ \leq b^+.$$

Además, tenemos que $a^- = 0$ y $b^- = 0$, con lo cual se cumple que

$$-a^- \leq -b^-.$$

2. Si $a \leq 0$ y $b \leq 0$ se tiene que $a = -a^-, a^+ = 0, b = -b^-$ y $b^+ = 0$, por lo tanto

$$-a^- \leq -b^- \quad \text{y} \quad a^+ \leq b^+.$$

3. Si $a \leq 0$ y $b \geq 0$ entonces $a = -a^-, a^+ = 0, b^+ = b$ y $b^- = 0$, con lo cual se obtiene que

$$a^+ \leq b^+ \quad \text{y} \quad -a^- \leq -b^-. \qquad \square$$

Demostración de la Proposición 4.9. Sea $u \in \mathcal{U}_{ad}$. Por (4.21) podemos expresar u como $u = u^+ - u^-$, probaremos entonces que $u^+ \in \mathcal{U}_{ad}^+$ y $u^- \in \mathcal{U}_{ad}^-$. Para ello, sabemos que para todo $i = 1, \ldots, M$ se tiene que

$$u_{a,i} \leq u_i \leq u_{b,i},$$

lo cual es equivalente a

$$u_{a,i} \leq u_i \quad \text{y} \quad u_i \leq u_{b,i}, \tag{4.22}$$

luego, del Lema 4.10 se sigue que

$$u_{a,i}^+ \leq u_i^+ \quad \text{y} \quad u_i^+ \leq u_{b,i}^+,$$

y que

$$-u_{a,i}^- \leq -u_i^- \quad \text{y} \quad -u_i^- \leq -u_{b,i}^-,$$

juntando las desigualdades anteriores se sigue que

$$u_{a,i}^+ \leq u_i^+ \leq u_{b,i}^+,$$

y que

$$-u_{a,i}^- \leq -u_i^- \leq -u_{b,i}^-,$$

es decir $u^+ \in \mathcal{U}_{ad}^+$ y $u^- \in \mathcal{U}_{ad}^-$.

Recíprocamente, sea $v \in \mathcal{U}_{ad}^+$ y $w \in \mathcal{U}_{ad}^-$ probaremos que $v - w \in \mathcal{U}_{ad}$. Para ello, sabemos que para todo $i = 1, \ldots, M$ se cumple que

$$u_{a,i}^+ \leq v_i \leq u_{b,i}^+$$

y que

$$-u_{a,i}^- \leq -w_i \leq -u_{b,i}^-.$$

Sumando estas desigualdades se obtiene que

$$u_{a,i}^+ - u_{a,i}^- \leq v_i - w_i \leq u_{b,i}^+ - u_{b,i}^-,$$

es decir

$$u_{a,i} \leq v_i - w_i \leq u_{b,i},$$

por lo tanto $v - w \in \mathcal{U}_{ad}$. □

COROLARIO 4.11. *Sea $u \in \mathbb{R}^M$, entonces $u \in \mathcal{U}_{ad}$ si y solo si $u^+ \in \mathcal{U}_{ad}^+$ y $u^- \in \mathcal{U}_{ad}^-$.*

Estamos en condiciones de reformular nuestro problema (3.3), el cual puede ser escrito como

$$\begin{cases} \min\limits_{u \in \mathcal{U}_{ad}} f(u) = \dfrac{1}{2} \left|\left| \sum\limits_{i=1}^{M} (u_i^+ - u_i^-) y_i - y_d \right|\right|^2 + \dfrac{\alpha}{2} \sum\limits_{i=1}^{M} \left[u_i^{+2} + u_i^{-2} \right] + \beta \sum\limits_{i=1}^{M} (u_i^+ + u_i^-) \\ \text{sujeto a:} \\ u = u^+ - u^-, \\ \sum\limits_{i=1}^{M} (u_i^+ - u_i^-) y_i(x_j) \leq b_j, \quad \forall j = 1, \ldots, l. \end{cases}$$

Este planteamiento nos sirve para reformular el siguiente problema de control óptimo:

$$\begin{cases} \min\limits_{\mathbf{u} \in \mathcal{V}_{ad}} F(\mathbf{u}) = \dfrac{1}{2} \left|\left| \sum\limits_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i - y_d \right|\right|^2 + \dfrac{\alpha}{2} \sum\limits_{i=1}^{2M} \mathbf{u}_i^2 + \beta \sum\limits_{i=1}^{2M} \mathbf{u}_i \\ \text{sujeto a:} \\ \sum\limits_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i(x_j) - b_j \leq 0, \quad \forall j = 1, \ldots, l, \end{cases} \tag{4.23}$$

donde $\mathcal{V}_{ad} \subset \mathbb{R}^{2M}$ es el conjunto de los controles admisibles definido como

$$\mathcal{V}_{ad} = \mathcal{U}_{ad}^+ \times \mathcal{U}_{ad}^-. \tag{4.24}$$

OBSERVACIÓN. Note que realizando la separación de u en sus partes positiva y negativa definimos un problema que tiene el doble de variables del problema (3.3).

OBSERVACIÓN. Nótese que si $\mathbf{u} \in \mathcal{V}_{ad}$ entonces $\mathbf{u} \geq 0$, es decir $\mathcal{V}_{ad} \subset \mathbb{R}_+^{2M}$.

Realizando un análisis análogo al problema (3.3), se define el conjunto

$$\mathcal{V}_{feas} = \left\{ \mathbf{u} \in \ \mathcal{V}_{ad} : \sum_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i(x_j) - b_j \leq 0, \ \forall j = 1, \ldots, l \right\} \subset \mathbb{R}^{2M}.$$

Similarmente a la Proposición 3.5, se tiene que $\mathcal{V}_{feas}$ es un conjunto cerrado, convexo y acotado, además, por la Hipótesis 3.3 es distinto del vacío. De la misma manera,

F es una función continua, por lo tanto se concluye que existe un control óptimo $\bar{\mathbf{u}} \in \mathcal{V}_{ad}$ para (4.23) y que es único, pues F es estrictamente convexa.

A continuación se demuestra la condición de ortogonalidad que satisface un control óptimo de (4.23). Está propiedad es importante ya que nos permitirá demostrar la equivalencia entre (4.23) y nuestro problema reducido (3.3).

PROPOSICIÓN 4.12. *Si* $\bar{\mathbf{u}}$ *es el control óptimo de (4.23) entonces* $\bar{\mathbf{u}}_i \bar{\mathbf{u}}_{i+M} = 0$ *para todo* $i = 1, \ldots, M$.

Demostración. La demostración se consigue por contradicción. Suponemos, sin pérdida de generalidad, que $\bar{\mathbf{u}}_1 \bar{\mathbf{u}}_{1+M} > 0$, es decir $\bar{\mathbf{u}}_1 > 0$ y $\bar{\mathbf{u}}_{1+M} > 0$, y consideramos el vector $\widetilde{\mathbf{u}} \in \mathbb{R}^M$, definido como

$$\begin{aligned}
&\widetilde{\mathbf{u}}_1 = \bar{\mathbf{u}}_1 - \text{mín}(\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_{1+M}) \geq 0, \\
&\widetilde{\mathbf{u}}_{1+M} = \bar{\mathbf{u}}_{1+M} - \text{mín}(\bar{\mathbf{u}}_1, \bar{\mathbf{u}}_{1+M}) \geq 0, \\
&\widetilde{\mathbf{u}}_i = \bar{\mathbf{u}}_i, \qquad\qquad \forall i = 2, \ldots, M, \\
&\widetilde{\mathbf{u}}_{i+M} = \bar{\mathbf{u}}_{i+M}, \qquad\qquad \forall i = 2, \ldots, M.
\end{aligned}$$

Se tiene que $\widetilde{\mathbf{u}} \in \mathcal{V}_{ad}$. En efecto, puesto que $\bar{\mathbf{u}} \in \mathcal{V}_{ad}$ se tiene que

$$\begin{aligned}
u_{a,i}^{+} \leq \bar{\mathbf{u}}_i \leq u_{b,i}^{+}, \qquad &\text{y que} \\
-u_{a,i}^{-} \leq -\bar{\mathbf{u}}_{i+M} \leq -u_{b,i}^{-}, &
\end{aligned} \tag{4.25}$$

para todo $i = 1, \ldots, M$. Por otra parte, por como se encuentra definido $\widetilde{\mathbf{u}}$, es suficiente probar que

$$\begin{aligned}
u_{a,1}^{+} \leq \widetilde{\mathbf{u}}_1 \leq u_{b,1}^{+}, \qquad &\text{y que} \\
-u_{a,1}^{-} \leq -\widetilde{\mathbf{u}}_{1+M} \leq -u_{b,1}^{-}. &
\end{aligned}$$

Para ello, primero sumamos las desigualdades de (4.25) con $i = 1$, obteniendo así

$$u_{a,1} \leq \bar{\mathbf{u}}_1 - \bar{\mathbf{u}}_{1+M} \leq u_{b,1}, \tag{4.26}$$

y analizamos los siguientes casos:

1. Si $u_{a,1}, u_{b,1} \geq 0$, de (4.26) se tiene que $0 \leq \bar{\mathbf{u}}_1 - \bar{\mathbf{u}}_{1+M}$, entonces $\bar{\mathbf{u}}_{1+M} \leq \bar{\mathbf{u}}_1$, y con la definición de $\widetilde{\mathbf{u}}_1, \widetilde{\mathbf{u}}_{1+M}$ se obtiene que

$$\begin{aligned}
\widetilde{\mathbf{u}}_1 = \bar{\mathbf{u}}_1 - \bar{\mathbf{u}}_{1+M} \geq 0, \quad &\text{y además} \\
\widetilde{\mathbf{u}}_{1+M} = 0, &
\end{aligned}$$

y utilizando nuevamente (4.26), obtenemos que

$$\begin{aligned} u_{a,1}^{+} \leq \widetilde{\mathbf{u}}_1 \leq u_{b,1}^{+}, \qquad & \text{y que} \\ -u_{a,1}^{-} \leq -\widetilde{\mathbf{u}}_{1+M} \leq -u_{b,1}^{-}. & \end{aligned}$$

2. Si $u_{a,1}, u_{b,1} \leq 0$, de (4.26) se tiene que $\bar{\mathbf{u}}_1 - \bar{\mathbf{u}}_{1+M} \leq 0$, entonces $\bar{\mathbf{u}}_1 \leq \bar{\mathbf{u}}_{1+M}$, con lo cual se obtiene que

$$\begin{aligned} \widetilde{\mathbf{u}}_1 = 0, \qquad & \text{y además} \\ \widetilde{\mathbf{u}}_{1+M} = \bar{\mathbf{u}}_{1+M} - \bar{\mathbf{u}}_1 \geq 0, & \end{aligned}$$

y utilizando nuevamente (4.26), obtenemos que

$$\begin{aligned} -u_{a,1}^{-} \leq -\widetilde{\mathbf{u}}_{1+M} \leq -u_{b,1}^{-}, \quad & \text{y que} \\ u_{a,1}^{+} \leq \widetilde{\mathbf{u}}_1 \leq u_{b,1}^{+}. & \end{aligned}$$

3. Si $u_{a,1} \leq 0$ y $u_{b,1} \geq 0$, entonces $u_a = -u_a^{-}$, $u_a^{+} = 0$, $u_b = u_b^{+}$ y $u_b^{-} = 0$, luego analizamos los siguientes casos:

   - Si $\bar{\mathbf{u}}_1 \geq \bar{\mathbf{u}}_{1+M}$, se obtiene que

   $$\begin{aligned} \widetilde{\mathbf{u}}_1 = \bar{\mathbf{u}}_1 - \bar{\mathbf{u}}_{1+M} \geq 0, \quad & \text{y además} \\ \widetilde{\mathbf{u}}_{1+M} = 0, & \end{aligned}$$

   y utilizando nuevamente (4.26) obtenemos que

   $$\begin{aligned} u_{a,1}^{+} \leq \widetilde{\mathbf{u}}_1 \leq u_{b,1}^{+}, \qquad & \text{y que} \\ -u_{a,1}^{-} \leq -\widetilde{\mathbf{u}}_{1+M} \leq -u_{b,1}^{-}. & \end{aligned}$$

   - Si $\bar{\mathbf{u}}_1 < \bar{\mathbf{u}}_{1+M}$, se obtiene que

   $$\begin{aligned} \widetilde{\mathbf{u}}_1 = 0, \qquad & \text{y además} \\ \widetilde{\mathbf{u}}_{1+M} = \bar{\mathbf{u}}_{1+M} - \bar{\mathbf{u}}_1 > 0, & \end{aligned}$$

   y utilizando nuevamente (4.26) obtenemos que

   $$\begin{aligned} u_{a,1}^{+} \leq \widetilde{\mathbf{u}}_1 \leq u_{b,1}^{+}, \qquad & \text{y que} \\ -u_{a,1}^{-} \leq -\widetilde{\mathbf{u}}_{1+M} \leq -u_{b,1}^{-}. & \end{aligned}$$

Por otra parte, puesto que $\bar{\mathbf{u}}_1 > 0$ y $\bar{\mathbf{u}}_{1+M} > 0$ se obtiene que $0 \leq \widetilde{\mathbf{u}}_1 < \bar{\mathbf{u}}_1$ y $0 \leq \widetilde{\mathbf{u}}_{1+M} < \bar{\mathbf{u}}_{1+M}$, de lo cual se sigue que

$$\bar{\mathbf{u}}_1 + \bar{\mathbf{u}}_{1+M} > \widetilde{\mathbf{u}}_1 + \widetilde{\mathbf{u}}_{1+M}, \tag{4.27}$$

y por monotonía

$$\bar{\mathbf{u}}_1^2 + \bar{\mathbf{u}}_{1+M}^2 > \widetilde{\mathbf{u}}_1^2 + \widetilde{\mathbf{u}}_{1+M}^2, \tag{4.28}$$

además

$$\bar{\mathbf{u}}_1 - \bar{\mathbf{u}}_{1+M} = \widetilde{\mathbf{u}}_1 - \widetilde{\mathbf{u}}_{1+M},$$

esta igualdad, en particular implica que

$$\sum_{i=1}^{M} (\widetilde{\mathbf{u}}_i - \widetilde{\mathbf{u}}_{i+M}) y_i(x_j) - b_j \leq 0, \quad \forall j = 1, \ldots, l,$$

es decir $\widetilde{\mathbf{u}}$ satisface las restricciones de (4.23). De las dos desigualdades (4.27) y (4.28), se obtiene la siguiente acotación

$$\begin{aligned}
F(\bar{\mathbf{u}}) &= \frac{1}{2} \left\| \sum_{i=1}^{M} (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_{i+M}) y_i - y_d \right\|^2 + \frac{\alpha}{2} \sum_{i=1}^{2M} \bar{\mathbf{u}}_i^2 + \beta \sum_{i=1}^{2M} \bar{\mathbf{u}}_i \\
&= \frac{1}{2} \left\| \sum_{i=1}^{M} (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_{i+M}) y_i - y_d \right\|^2 + \frac{\alpha}{2} \sum_{i=1}^{M} (\bar{\mathbf{u}}_i^2 + \bar{\mathbf{u}}_{i+M}^2) + \beta \sum_{i=1}^{M} (\bar{\mathbf{u}}_i + \bar{\mathbf{u}}_{i+M}) \\
&= \frac{1}{2} \left\| \sum_{i=1}^{M} (\widetilde{\mathbf{u}}_i - \widetilde{\mathbf{u}}_{i+M}) y_i - y_d \right\|^2 + \frac{\alpha}{2} \sum_{i=1}^{M} (\bar{\mathbf{u}}_i^2 + \bar{\mathbf{u}}_{i+M}^2) + \beta \sum_{i=1}^{M} (\bar{\mathbf{u}}_i + \bar{\mathbf{u}}_{i+M}) \\
&> \frac{1}{2} \left\| \sum_{i=1}^{M} (\widetilde{\mathbf{u}}_i - \widetilde{\mathbf{u}}_{i+M}) y_i - y_d \right\|^2 + \frac{\alpha}{2} \sum_{i=1}^{M} (\widetilde{\mathbf{u}}_i^2 + \widetilde{\mathbf{u}}_{i+M}^2) + \beta \sum_{i=1}^{M} (\widetilde{\mathbf{u}}_i + \widetilde{\mathbf{u}}_{i+M}) \\
&= F(\widetilde{\mathbf{u}}),
\end{aligned}$$

lo cual es una contradicción debido a la optimalidad de $\bar{\mathbf{u}}$. □

PROPOSICIÓN 4.13. *Sea* $\bar{\mathbf{u}}$ *el control óptimo de (4.23) y* $\bar{w} \in \mathbb{R}^M$ *el vector definido por*

$$\bar{w} = [\bar{\mathbf{u}}_i]_{i=1}^{M} - [\bar{\mathbf{u}}_{i+M}]_{i=1}^{M},$$

entonces

$$[\bar{\mathbf{u}}_i]_{i=1}^{M} = \bar{w}^+ \quad y \quad [\bar{\mathbf{u}}_{i+M}]_{i=1}^{M} = \bar{w}^-.$$

Demostración. De la definición de $\bar{w}$ se tiene que

$$\bar{w}_i = \bar{\mathbf{u}}_i - \bar{\mathbf{u}}_{i+M},$$

para todo $i = 1 \ldots, M$. Sabemos que $\bar{\mathbf{u}}_i \geq 0, \bar{\mathbf{u}}_{i+M} \geq 0$ y de la Proposición 4.12 se sigue que $\bar{\mathbf{u}}_i \bar{\mathbf{u}}_{i+M} = 0$. Así, la Proposición 4.8 implica que

$$\bar{\mathbf{u}}_i = \bar{w}_i^+ \quad \text{y} \quad \bar{\mathbf{u}}_{i+M} = \bar{w}_i^-, \quad \forall i = 1 \ldots, M,$$

esto nos permite concluir que

$$[\bar{\mathbf{u}}_i]_{i=1}^{M} = \bar{w}^{+} \quad \text{y} \quad [\bar{\mathbf{u}}_{i+M}]_{i=1}^{M} = \bar{w}^{-}. \qquad \square$$

PROPOSICIÓN 4.14. *Si $\bar{u}$ y $\bar{\mathbf{u}}$ son soluciones de los problemas (3.3) y (4.23) respectivamente, entonces*

$$[\bar{\mathbf{u}}_i]_{i=1}^{M} = \bar{u}^{+} \quad y \quad [\bar{\mathbf{u}}_{i+M}]_{i=1}^{M} = \bar{u}^{-},$$

es decir $\bar{u} = [\bar{\mathbf{u}}_i]_{i=1}^{M} - [\bar{\mathbf{u}}_{i+M}]_{i=1}^{M}$. En otras palabras, los problemas (3.3) y (4.23) son equivalentes.

Demostración. Si $\bar{u}$ es la solución óptima de (3.3), al ser descompuesto de la forma $\bar{u} = \bar{u}^{+} - \bar{u}^{-}$, podemos definir el siguiente vector

$$\bar{\mathbf{w}} = \begin{pmatrix} \bar{u}^{+} \\ \bar{u}^{-} \end{pmatrix} \in \mathbb{R}^{2M}.$$

Puesto que $\bar{u} \in \mathcal{U}_{ad}$ se obtiene que $\bar{u}^{+} \in \mathcal{U}_{ad}^{+}$ y $\bar{u}^{-} \in \mathcal{U}_{ad}^{-}$, en consecuencia $\bar{\mathbf{w}} \in \mathcal{V}_{ad}$. Además, como $\bar{u}$ cumple las restricciones de (3.3), se sigue que

$$\sum_{i=1}^{M} \bar{u}_i y_i(x_j) - b_j \leq 0, \quad \forall j = 1, \ldots, l,$$

entonces

$$\sum_{i=1}^{M} (\bar{u}_i^{+} - \bar{u}_i^{-}) y_i(x_j) - b_j \leq 0, \quad \forall j = 1, \ldots, l,$$

y al ser escrito en términos de $\bar{\mathbf{w}}$, obtenemos que

$$\sum_{i=1}^{M} (\bar{\mathbf{w}}_i - \bar{\mathbf{w}}_{i+M}) y_i(x_j) - b_j \leq 0, \quad \forall j = 1, \ldots, l,$$

es decir $\bar{\mathbf{w}} \in \mathcal{V}_{feas}$. Además, tenemos también que

$$\begin{aligned} f(\bar{u}) &= f(\bar{u}^{+} - \bar{u}^{-}) \\ &= \frac{1}{2} \left|\left| \sum_{i=1}^{M} (\bar{u}_i^{+} - \bar{u}_i^{-}) y_i - y_d \right|\right|^2 + \frac{\alpha}{2} \sum_{i=1}^{M} ((\bar{u}_i^{+})^2 + (\bar{u}_i^{-})^2) + \beta \sum_{i=1}^{M} (\bar{u}_i^{+} + \bar{u}_i^{-}) \\ &= \frac{1}{2} \left|\left| \sum_{i=1}^{M} (\bar{\mathbf{w}}_i - \bar{\mathbf{w}}_{i+M}) y_i - y_d \right|\right|^2 + \frac{\alpha}{2} \sum_{i=1}^{M} (\bar{\mathbf{w}}_i^2 + \bar{\mathbf{w}}_{i+M}^2) + \beta \sum_{i=1}^{M} (\bar{\mathbf{w}}_i + \bar{\mathbf{w}}_{i+M}) \\ &= F(\bar{\mathbf{w}}). \end{aligned}$$

Por otro lado, para $\bar{\mathbf{u}} \in \mathcal{V}_{ad}$ la solución del problema (4.23), definimos el vector

$\bar{w} \in \mathbb{R}^M$ tal que

$$\bar{w} = [\bar{\mathbf{u}}_i]_{i=1}^M - [\bar{\mathbf{u}}_{i+M}]_{i=1}^M,$$

y puesto que $[\bar{\mathbf{u}}_i]_{i=1}^M \in \mathcal{U}_{ad}^+$ y $[\bar{\mathbf{u}}_{i+M}]_{i=1}^M \in \mathcal{U}_{ad}^-$ se tiene que $\bar{w} \in \mathcal{U}_{ad}$. Además, como $\bar{\mathbf{u}}$ satisface las restricciones de (4.23), se sigue que

$$\sum_{i=1}^{M} (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_{i+M}) y_i(x_j) - b_j \leq 0, \quad \forall j = 1, \ldots, l,$$

de lo cual se obtiene que

$$\sum_{i=1}^{M} \bar{w}_i y_i(x_j) - b_j \leq 0, \quad \forall j = 1, \ldots, l,$$

es decir $\bar{w} \in \mathcal{U}_{feas}$. Adicionalmente, por la Proposición 4.13, tenemos

$$\begin{aligned}
F(\bar{\mathbf{u}}) &= \frac{1}{2} \left|\left| \sum_{i=1}^{M} (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_{i+M}) y_i - y_d \right|\right|^2 + \frac{\alpha}{2} \sum_{i=1}^{M} (\bar{\mathbf{u}}_i^2 + \bar{\mathbf{u}}_{i+M}^2) + \beta \sum_{i=1}^{M} (\bar{\mathbf{u}}_i + \bar{\mathbf{u}}_{i+M}) \\
&= \frac{1}{2} \left|\left| \sum_{i=1}^{M} (\bar{w}_i^+ - \bar{w}_i^-) y_i - y_d \right|\right|^2 + \frac{\alpha}{2} \sum_{i=1}^{M} ((\bar{w}_i^+)^2 + (\bar{w}_i^-)^2) + \beta \sum_{i=1}^{M} (\bar{w}_i^+ + \bar{w}_i^-) \\
&= \frac{1}{2} \left|\left| \sum_{i=1}^{M} \bar{w}_i y_i - y_d \right|\right|^2 + \frac{\alpha}{2} ||\bar{w}||_2^2 + \beta ||\bar{w}||_1 \\
&= f(\bar{w}).
\end{aligned}$$

Puesto que $\bar{\mathbf{u}}$ es la solución óptima del problema (4.23) se obtiene que

$$f(\bar{u}) = F(\bar{\mathbf{w}}) \geq F(\bar{\mathbf{u}}) = f(\bar{w}),$$

y por unicidad del control óptimo para el problema (3.3), se concluye que $\bar{w} = \bar{u}$, lo cual es equivalente a $\bar{\mathbf{w}} = \bar{\mathbf{u}}$. □

Siendo (4.23) un problema de dimensión finita, obtenemos las condiciones de optimalidad aplicando la teoría de optimización en $\mathbb{R}^n$. Por tanto, introducimos el Lagrangiano $\mathcal{L} : \mathbb{R}^{2M} \times \mathbb{R}^{l+4M} \to \mathbb{R}$ asociado a (4.23), definido por

$$\mathcal{L}(\mathbf{u}, \mu, \nu) = F(\mathbf{u}) + \sum_{j=i}^{l} \mu_j G_j(\mathbf{u}) + \sum_{i=i}^{4M} \nu_i G_{l+i}(\mathbf{u}), \tag{4.29}$$

donde

$$\begin{aligned}
G_j(\mathbf{u}) &= \sum_{i=1}^{M}(\mathbf{u}_i - \mathbf{u}_{i+M})y_i(x_j) - b_j, && \text{para } j = 1,\dots,l,\\
G_{l+i}(\mathbf{u}) &= u_{a,i}^{+} - \mathbf{u}_i, && \text{para } i = 1,\dots,M,\\
G_{l+M+i}(\mathbf{u}) &= u_{b,i}^{-} - \mathbf{u}_{i+M}, && \text{para } i = 1,\dots,M,\\
G_{l+2M+i}(\mathbf{u}) &= \mathbf{u}_i - u_{b,i}^{+}, && \text{para } i = 1,\dots,M,\\
G_{l+3M+i}(\mathbf{u}) &= \mathbf{u}_{i+M} - u_{a,i}^{-}, && \text{para } i = 1,\dots,M,
\end{aligned} \tag{4.30}$$

y denotamos G a la función cuyas componentes son G_i para $i = 1,\dots,l+4M$.

En lo que sigue vamos a denotar $\bar{\mathbf{u}}$ a la solución óptima de (4.23) y por la Hipótesis 3.7 $\bar{\mathbf{u}}$ satisface la condición (LICQ) para G. Para la obtención de las condiciones necesarias de primer orden utilizamos la Proposición 2.19, la cual se cumple gracias a que la condición (LICQ) se satisface en $\bar{\mathbf{u}}$. Por lo tanto, existen $\bar{\mu} \in \mathbb{R}^l_+$ y $\bar{v} \in \mathbb{R}^{4M}_+$ tales que

$$\begin{aligned}
\nabla F(\bar{\mathbf{u}}) + \sum_{j=i}^{l}\bar{\mu}_j\nabla G_j(\bar{\mathbf{u}}) + \sum_{i=i}^{4M}\bar{v}_i\nabla G_{l+i}(\bar{\mathbf{u}}) &= 0,\\
G_j(\bar{\mathbf{u}}) \le 0, \quad \forall j &= 1,\dots,l+4M,\\
\bar{\mu}_j G_j(\bar{\mathbf{u}}) = 0, \quad \forall j &= 1,\dots,l,\\
\bar{v}_i G_{l+i}(\bar{\mathbf{u}}) = 0, \quad \forall i &= 1,\dots,4M.
\end{aligned} \tag{4.31}$$

Para detallar adecuadamente estas condiciones calculamos las siguientes derivadas, recordando que $(\cdot,\cdot)$ denota el producto escalar en $L^2(\Omega)$.

$$\nabla F(\mathbf{u}) = \begin{pmatrix}
(y_u - y_d, y_1) + \alpha\mathbf{u}_1 + \beta\\
\vdots\\
(y_u - y_d, y_i) + \alpha\mathbf{u}_i + \beta\\
\vdots\\
(y_u - y_d, y_M) + \alpha\mathbf{u}_M + \beta\\
-(y_u - y_d, y_1) + \alpha\mathbf{u}_{1+M} + \beta\\
\vdots\\
-(y_u - y_d, y_i) + \alpha\mathbf{u}_{i+M} + \beta\\
\vdots\\
-(y_u - y_d, y_M) + \alpha\mathbf{u}_{2M} + \beta
\end{pmatrix}, \tag{4.32}$$

donde $y_u = \sum_{k=1}^{M} (\mathbf{u}_k - \mathbf{u}_{k+M}) y_k$. Por otra parte tenemos

$$\nabla G_j(\mathbf{u}) = \begin{pmatrix} y_1(x_j) \\ \vdots \\ y_i(x_j) \\ \vdots \\ y_M(x_j) \\ -y_1(x_j) \\ \vdots \\ -y_i(x_j) \\ \vdots \\ -y_M(x_j) \end{pmatrix}, \quad \forall j = 1, \ldots, l. \tag{4.33}$$

Por lo tanto, nuestras condiciones (4.31) quedan expresadas por

$$\begin{aligned} (\bar{y} - y_d, y_i) + \alpha \bar{\mathbf{u}}_i + \beta + \sum_{j=1}^{l} \bar{\mu}_j y_i(x_j) - \bar{v}_i + \bar{v}_{i+2M} = 0, \\ -(\bar{y} - y_d, y_i) + \alpha \bar{\mathbf{u}}_{i+M} + \beta - \sum_{j=1}^{l} \bar{\mu}_j y_i(x_j) - \bar{v}_{i+M} + \bar{v}_{i+3M} = 0, \end{aligned} \tag{4.34}$$

para $i = 1, \ldots, M$, donde $\bar{y} = \sum_{i=1}^{M} (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_{i+M}) y_i$, junto con

$$G_j(\bar{\mathbf{u}}) \leq 0, \quad \forall j = 1, \ldots, l + 4M. \tag{4.35}$$

Además, satisface las condiciones de complementariedad

$$\begin{aligned} \bar{\mu}_j \left(\bar{y}(x_j) - b_j \right) &= 0, \quad \forall j = 1, \ldots, l, \\ \bar{v}_i (u_{a,i}^{+} - \bar{\mathbf{u}}_i) &= 0, \quad \forall i = 1, \ldots, M, \\ \bar{v}_{i+M} (u_{b,i}^{-} - \bar{\mathbf{u}}_{i+M}) &= 0, \quad \forall i = 1, \ldots, M, \\ \bar{v}_{i+2M} (\bar{\mathbf{u}}_i - u_{b,i}^{+}) &= 0, \quad \forall i = 1, \ldots, M, \\ \bar{v}_{i+3M} (\bar{\mathbf{u}}_{i+M} - u_{a,i}^{-}) &= 0, \quad \forall i = 1, \ldots, M. \end{aligned} \tag{4.36}$$

Para encontrar las condiciones necesarias de segundo orden, es necesario calcular la matriz Hessiana del Lagrangiano con respecto a $\mathbf{u}$. Para ello, por la linealidad de las restricciones, notamos que

$$\nabla^2 \mathcal{L}(\bar{\mathbf{u}}, \mu, \nu) = \nabla^2 F(\bar{\mathbf{u}}),$$

donde $\nabla^2 F(\bar{\mathbf{u}})$ está dado por:

$$\nabla^2 F(\bar{\mathbf{u}}) = \left(\begin{array}{cccc|cccc} (y_1,y_1)+\alpha & (y_2,y_1) & \cdots & (y_M,y_1) & -(y_1,y_1) & -(y_2,y_1) & \cdots & -(y_M,y_1) \\ (y_1,y_2) & (y_2,y_2)+\alpha & \cdots & (y_M,y_2) & -(y_1,y_2) & -(y_2,y_2) & \cdots & -(y_M,y_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (y_1,y_M) & (y_2,y_M) & \cdots & (y_M,y_M)+\alpha & -(y_1,y_M) & -(y_2,y_M) & \cdots & -(y_M,y_M) \\ \hline -(y_1,y_1) & -(y_2,y_1) & \cdots & -(y_M,y_1) & (y_1,y_1)+\alpha & (y_2,y_1) & \cdots & (y_M,y_1) \\ -(y_1,y_2) & -(y_2,y_2) & \cdots & -(y_M,y_2) & (y_1,y_2) & (y_2,y_2)+\alpha & \cdots & (y_M,y_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -(y_1,y_M) & -(y_2,y_M) & \cdots & -(y_M,y_M) & (y_1,y_M) & (y_2,y_M) & \cdots & (y_M,y_M)+\alpha \end{array}\right).$$

Visto de otro modo, se tiene que

$$\nabla^2 \mathcal{L}(\bar{\mathbf{u}},\bar{\mu},\bar{v}) = \nabla^2 F(\bar{\mathbf{u}}) = \left(\begin{array}{c|c} A+\alpha I_M & -A \\ \hline -A & A+\alpha I_M \end{array}\right), \tag{4.37}$$

donde la matriz A es simétrica semi-definida positiva y está dada por

$$A = \begin{pmatrix} (y_1,y_1) & (y_2,y_1) & \cdots & (y_M,y_1) \\ (y_1,y_2) & (y_2,y_2) & \cdots & (y_M,y_2) \\ \vdots & \vdots & \ddots & \vdots \\ (y_1,y_M) & (y_2,y_M) & \cdots & (y_M,y_M) \end{pmatrix}.$$

PROPOSICIÓN 4.15. *La matriz $\nabla^2 \mathcal{L}(\bar{\mathbf{u}},\bar{\mu},\bar{v})$ es definida positiva.*

Demostración. Para $v \neq 0$ y puesto que A es simétrica semi-definida positiva y $\alpha > 0$ se tiene que

$$v^T \nabla^2 \mathcal{L}(\bar{\mathbf{u}},\bar{\mu},\bar{v}) v = \sum_{i,j=1}^{M} (v_j - v_{j+M})(y_i,y_j)(v_i - v_{i+M}) + \alpha \sum_{j=1}^{2M} v_j^2 > 0. \qquad \square$$

Finalmente concluimos esta sección con el siguiente resultado.

PROPOSICIÓN 4.16. *Existen constantes positivas ω y ε tales que*

$$F(\mathbf{u}) - F(\bar{\mathbf{u}}) \geq \omega ||\mathbf{u} - \bar{\mathbf{u}}||^2,$$

para todo $\mathbf{u} \in \mathcal{V}_{ad}$ con $||\mathbf{u} - \bar{\mathbf{u}}|| \leq \varepsilon$.

Demostración. La demostración es inmediata pues se satisfacen las hipótesis de la Proposición 2.20. $\square$

4.3. Discretización del problema de control

En esta sección consideramos la discretización de la ecuación de estado estudiada en la sección 4.1, la cual nos permite definir el siguiente problema de control discreto reemplazando el estado discreto y_u^h definido en (4.7):

$$\begin{cases} \min\limits_{u\in\mathcal{U}_{ad}} f_h(u) := \dfrac{1}{2}\left|\left|\sum\limits_{i=1}^{M} u_i y_i^h - y_d\right|\right|^2 + \dfrac{\alpha}{2}||u||_2^2 + \beta||u||_1 \\ \text{sujeto a:} \\ \sum\limits_{i=1}^{M} u_i y_i^h(x_j) - b_j \leq 0, \qquad \forall j = 1,\ldots,l. \end{cases} \tag{4.38}$$

Luego, de manera similar al problema continuo, usamos la descomposición de cualquier vector en $\mathbb{R}^M$ como la diferencia entre su parte positiva y su parte negativa. Así definimos el problema

$$\begin{cases} \min\limits_{\mathbf{u}\in\mathcal{V}_{ad}} F_h(\mathbf{u}) := \dfrac{1}{2}\left|\left|\sum\limits_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i^h - y_d\right|\right|^2 + \dfrac{\alpha}{2}\sum\limits_{i=1}^{2M} \mathbf{u}_i^2 + \beta\sum\limits_{i=1}^{2M} \mathbf{u}_i \\ \text{sujeto a:} \\ \sum\limits_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i^h(x_j) - b_j \leq 0, \qquad \forall j = 1,\ldots,l. \end{cases} \tag{4.39}$$

Para demostrar la existencia de una solución óptima para (4.38) se debe asegurar que el conjunto

$$\mathcal{U}_{feas}^h = \left\{ u \in \mathcal{U}_{ad} : \sum_{i=1}^{M} u_i y_i^h(x_j) - b_j \leq 0,\ \forall j = 1,\ldots,l \right\},$$

sea distinto del vacío. De la misma forma, (4.39) tiene una solución óptima si el conjunto

$$\mathcal{V}_{feas}^h = \left\{ \mathbf{u} \in \mathcal{V}_{ad} : \sum_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i^h(x_j) - b_j \leq 0,\ \forall j = 1,\ldots,l \right\},$$

es diferente del vacío.

OBSERVACIÓN. $\mathcal{V}_{feas}^h$ es distinto del vacío si y solo si $\mathcal{U}_{feas}^h$ es distinto del vacío.

Por el momento para nuestro análisis vamos asumir que (4.38) y (4.39) poseen una solución óptima. En la Sección 4.4 se discutirá la existencia de única solución para el problema (4.39).

PROPOSICIÓN 4.17. *(Condición de ortogonalidad) Si $\bar{\mathbf{u}}^h$ es un control óptimo para el problema (4.39), entonces $\bar{\mathbf{u}}_i^h \bar{\mathbf{u}}_{i+M}^h = 0$ para todo $i = 1, \ldots, M$.*

Demostración. La demostración es análoga a la Proposición 4.12. □

PROPOSICIÓN 4.18. *Si $\bar{u}^h$ y $\bar{\mathbf{u}}^h$ son soluciones de los problemas (4.38) y (4.39) respectivamente, entonces*

$$\left[\bar{\mathbf{u}}_i^h\right]_{i=1}^M = \bar{u}^{h+} \quad y \quad \left[\bar{\mathbf{u}}_{i+M}^h\right]_{i=1}^M = \bar{u}^{h-},$$

es decir $\bar{u}^h = \left[\bar{\mathbf{u}}_i^h\right]_{i=1}^M - \left[\bar{\mathbf{u}}_{i+M}^h\right]_{i=1}^M$. En otras palabras, los problemas (4.38) y (4.39) son equivalentes.

Demostración. La prueba es similar a la Proposición 4.14. □

Establecemos ahora las condiciones necesarias de primer orden para el problema (4.39). De manera análoga introducimos el Lagrangiano $\mathcal{L}_h : \mathbb{R}^{2M} \times \mathbb{R}^{l+4M} \to \mathbb{R}$ definido por

$$\mathcal{L}_h(\mathbf{u}, \mu, \nu) = F_h(\mathbf{u}) + \sum_{j=1}^{l} \mu_j G_j^h(\mathbf{u}) + \sum_{i=i}^{4M} \nu_i G_{l+i}^h(\mathbf{u}), \tag{4.40}$$

donde

$$\begin{aligned}
&G_j^h(\mathbf{u}) = \sum_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i^h(x_j) - b_j, && \text{para } j = 1, \ldots, l,\\
&G_{l+i}^h(\mathbf{u}) = u_{a,i}^+ - \mathbf{u}_i, && \text{para } i = 1, \ldots, M,\\
&G_{l+M+i}^h(\mathbf{u}) = u_{b,i}^- - \mathbf{u}_{i+M}, && \text{para } i = 1, \ldots, M,\\
&G_{l+2M+i}^h(\mathbf{u}) = \mathbf{u}_i - u_{b,i}^+, && \text{para } i = 1, \ldots, M,\\
&G_{l+3M+i}^h(\mathbf{u}) = \mathbf{u}_{i+M} - u_{a,i}^-, && \text{para } i = 1, \ldots, M,
\end{aligned} \tag{4.41}$$

y denotamos G_h a la función cuyas componentes son G_i^h para $i = 1, \ldots, l + 4M$.

Si $\bar{\mathbf{u}}^h$ es la solución óptima de (4.39) y satisface la condición (LICQ), se tiene entonces que existe un $\bar{\mu}^h \in \mathbb{R}_+^l$ y $\bar{\nu}^h \in \mathbb{R}_+^{4M}$ tal que

$$\nabla F_h(\bar{\mathbf{u}}^h) + \sum_{j=i}^{l} \bar{\mu}_j^h \nabla G_j^h(\bar{\mathbf{u}}^h) + \sum_{i=i}^{4M} \bar{\nu}_i^h \nabla G_{l+i}^h(\bar{\mathbf{u}}^h) = 0,$$

y

$$
\begin{aligned}
&G_j^h(\bar{\mathbf{u}}^h) \leq 0, && \forall j = 1, \ldots, l + 4M, \\
&\bar{\mu}_j^h G_j^h(\bar{\mathbf{u}}^h) = 0, && \forall j = 1, \ldots, l, \\
&\bar{v}_i^h G_{l+i}^h(\bar{\mathbf{u}}^h) = 0, && \forall i = 1, \ldots, 4M.
\end{aligned}
$$

Calculando las derivadas correspondientes tenemos:

$$
\nabla F_h(\mathbf{u}) = \begin{pmatrix}
(y_u^h - y_d, y_1^h) + \alpha \mathbf{u}_1 + \beta \\
\vdots \\
(y_u^h - y_d, y_i^h) + \alpha \mathbf{u}_i + \beta \\
\vdots \\
(y_u^h - y_d, y_M^h) + \alpha \mathbf{u}_M + \beta \\
-(y_u^h - y_d, y_1^h) + \alpha \mathbf{u}_{1+M} + \beta \\
\vdots \\
-(y_u^h - y_d, y_i^h) + \alpha \mathbf{u}_{i+M} + \beta \\
\vdots \\
-(y_u^h - y_d, y_M^h) + \alpha \mathbf{u}_{2M} + \beta
\end{pmatrix}, \tag{4.42}
$$

donde $y_u^h = \sum_{k=1}^{M} (\mathbf{u}_k - \mathbf{u}_{k+M}) y_k^h$. Por otra parte, se tiene

$$
\nabla G_j^h(\mathbf{u}) = \begin{pmatrix}
y_1^h(x_j) \\
\vdots \\
y_i^h(x_j) \\
\vdots \\
y_M^h(x_j) \\
-y_1^h(x_j) \\
\vdots \\
-y_i^h(x_j) \\
\vdots \\
-y_M^h(x_j)
\end{pmatrix}, \quad \forall j = 1, \ldots, l. \tag{4.43}
$$

Así, nuestras condiciones quedan expresadas por

$$
\begin{aligned}
(\bar{y}^h - y_d, y_i^h) + \alpha \bar{\mathbf{u}}_i^h + \beta + \sum_{j=1}^{l} \bar{\mu}_j^h y_i^h(x_j) - \bar{v}_i^h + \bar{v}_{i+2M}^h = 0, \\
-(\bar{y}^h - y_d, y_i^h) + \alpha \bar{\mathbf{u}}_{i+M}^h + \beta - \sum_{j=1}^{l} \bar{\mu}_j^h y_i^h(x_j) - \bar{v}_{i+M}^h + \bar{v}_{i+3M}^h = 0,
\end{aligned} \tag{4.44}
$$

para $i = 1, \ldots, M$, donde $\bar{y}^h = \sum_{i=1}^{M} (\bar{\mathbf{u}}_i^h - \bar{\mathbf{u}}_{i+M}^h) y_i^h$, junto con

$$G_j^h(\bar{\mathbf{u}}^h) \leq 0, \quad \forall j = 1, \ldots, l + 4M.$$

Adicionalmente, se satisfacen las condiciones de complementariedad

$$\begin{aligned}
\bar{\mu}_j^h \left(\bar{y}^h(x_j) - b_j \right) &= 0, \quad \forall j = 1, \ldots, l, \\
\bar{v}_i^h (u_{a,i}^+ - \bar{\mathbf{u}}_i^h) &= 0, \quad \forall i = 1, \ldots, M, \\
\bar{v}_{i+M}^h (u_{b,i}^- - \bar{\mathbf{u}}_{i+M}^h) &= 0, \quad \forall i = 1, \ldots, M, \\
\bar{v}_{i+2M}^h (\bar{\mathbf{u}}_i^h - u_{b,i}^+) &= 0, \quad \forall i = 1, \ldots, M, \\
\bar{v}_{i+3M}^h (\bar{\mathbf{u}}_{i+M}^h - u_{a,i}^-) &= 0, \quad \forall i = 1, \ldots, M.
\end{aligned} \tag{4.45}$$

4.4. Estimación del error para el control

En esta sección presentamos el resultado principal de este proyecto: una estimación del error para los controles óptimos del problema de control discreto (4.38). Para llegar a esta estimación, por conveniencia de la diferenciabilidad en la función de costo, trabajamos con los problemas equivalentes (4.23) y (4.39), para los cuales estamos interesados en estimar la diferencia entre sus respectivas soluciones. Esta interrogante nos lleva a realizar un análisis de sensibilidad de problemas de programación no lineal con respecto a sus perturbaciones para aplicarlo a nuestros problemas de control (4.23) y (4.39).

Empecemos nuestro estudio observando que el problema (4.23) se puede formular en la forma:

$$\begin{cases} \text{mín } F(\mathbf{u}) \\ G_j(\mathbf{u}) \leq 0, \quad j = 1, \ldots, l, \\ \mathbf{u} \in \mathcal{V}_{ad}, \end{cases} \tag{4.46}$$

donde G_j definido como (4.30), mientras que el problema discreto (4.39) tiene la forma

$$\begin{cases} \text{mín } F_h(\mathbf{u}) \\ G_j^h(\mathbf{u}) \leq 0, \quad j = 1, \ldots, l, \\ \mathbf{u} \in \mathcal{V}_{ad}, \end{cases} \tag{4.47}$$

con G_j^h está definido como (4.41).

En lo que sigue de esta sección denotamos:

$$y_u = \sum_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i,$$
$$y_u^h = \sum_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i^h,$$

para cualquier $\mathbf{u} \in \mathcal{V}_{ad}$.

A continuación, se demuestran algunos resultados técnicos de estimación, las cuales serán utilizados en los resultados subsiguientes.

PROPOSICIÓN 4.19. *Existe una constante $C > 0$, independiente de h, tal que*

$$\sum_{j=1}^{l} \left(|G_j(\mathbf{u}) - G_j^h(\mathbf{w})| + ||\nabla G_j(\mathbf{u}) - \nabla G_j^h(\mathbf{w})|| + ||\nabla^2 G_j(\mathbf{u}) - \nabla^2 G_j^h(\mathbf{w})|| \right) \leq C(||\mathbf{u} - \mathbf{w}|| + h^2 |\log h|)$$

se cumple para todo $\mathbf{u}, \mathbf{w} \in \mathcal{V}_{ad}$.

Demostración. Directamente de la definición de G_j y G_j^h, se tiene que

$$\begin{aligned} |G_j(\mathbf{u}) - G_j^h(\mathbf{w})| &= \left| \sum_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i(x_j) - \sum_{i=1}^{M} (\mathbf{w}_i - \mathbf{w}_{i+M}) y_i^h(x_j) \right| \\ &= \left| \sum_{i=1}^{M} \left[(\mathbf{u}_i - \mathbf{u}_{i+M}) - (\mathbf{w}_i - \mathbf{w}_{i+M}) \right] y_i(x_j) \right. \\ &\qquad \left. + \sum_{i=1}^{M} (\mathbf{w}_i - \mathbf{w}_{i+M}) \left[y_i(x_j) - y_i^h(x_j) \right] \right|, \end{aligned}$$

utilizando la desigualdad triangular, tenemos que

$$\begin{aligned} |G_j(\mathbf{u}) - G_j^h(\mathbf{w})| &\leq \sum_{i=1}^{M} \left(|(\mathbf{u}_i - \mathbf{u}_{i+M}) - (\mathbf{w}_i - \mathbf{w}_{i+M})| \, |y_i(x_j)| + |y_w(x_j) - y_w^h(x_j)| \right) \\ &\leq \sum_{i=1}^{M} \left[\left(|\mathbf{u}_i - \mathbf{w}_i| + |\mathbf{u}_{i+M} - \mathbf{w}_{i+M}| \right) |y_i(x_j)| + |y_w(x_j) - y_w^h(x_j)| \right]. \end{aligned}$$

Puesto que Ω_0 contiene x_j para todo $j = 1, \ldots, l$, consideramos la norma $||\cdot||_{C(\bar{\Omega}_0)}$ junto con el $\max_{i,j} |y_i(x_j)|$, obteniendo así

$$|G_j(\mathbf{u}) - G_j^h(\mathbf{w})| \leq C \left(\sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| + ||y_w - y_w^h||_{C(\bar{\Omega}_0)} \right),$$

por la estimación (4.14), se tiene que

$$|G_j(\mathbf{u}) - G_j^h(\mathbf{w})| \leq C(||\mathbf{u} - \mathbf{w}|| + h^2|\log h|). \tag{4.48}$$

Por otra parte, de (4.33) y (4.43) se obtiene la siguiente estimación

$$\begin{aligned} ||\nabla G_j(\mathbf{u}) - \nabla G_j^h(\mathbf{w})|| &\leq C\sum_{i=1}^{M} |y_i(x_j) - y_i^h(x_j)| \\ &\leq C\sum_{i=1}^{M} ||y_i - y_i^h||_{C(\bar{\Omega}_0)}, \end{aligned}$$

de la estimación (4.12), se sigue que

$$||\nabla G_j(\mathbf{u}) - \nabla G_j^h(\mathbf{w})|| \leq Ch^2|\log h|. \tag{4.49}$$

Finalmente, puesto que $\nabla^2 G_j(\mathbf{u})$ y $\nabla^2 G_j^h$ se anulan ya que las restricciones son lineales, el resultado se concluye sumando (4.48) y (4.49). □

PROPOSICIÓN 4.20. *Existe una constante $C > 0$, independiente de h, tal que*

$$|F(\mathbf{u}) - F(\mathbf{w})| + ||\nabla F(\mathbf{u}) - \nabla F(\mathbf{w})|| + ||\nabla^2 F(\mathbf{u}) - \nabla^2 F(\mathbf{w})|| \leq C||\mathbf{u} - \mathbf{w}|| \tag{4.50}$$

se cumple para todo $\mathbf{u}, \mathbf{w} \in \mathcal{V}_{ad}$.

Demostración. De la definición de F, se tiene que

$$\begin{aligned} |F(\mathbf{u}) - F(\mathbf{w})| = \Bigg| &\frac{1}{2}\left|\left|\sum_{i=1}^{M}(\mathbf{u}_i - \mathbf{u}_{i+M})y_i - y_d\right|\right|^2 - \frac{1}{2}\left|\left|\sum_{i=1}^{M}(\mathbf{w}_i - \mathbf{w}_{i+M})y_i - y_d\right|\right|^2 \\ &+\frac{\alpha}{2}\sum_{i=1}^{2M}(\mathbf{u}_i^2 - \mathbf{w}_i^2) + \beta\sum_{i=1}^{2M}(\mathbf{u}_i - \mathbf{w}_i)\Bigg| \\ = \Bigg| &\frac{1}{2}\left(||y_u - y_d||^2 - ||y_w - y_d||^2\right) + \frac{\alpha}{2}\sum_{i=1}^{2M}(\mathbf{u}_i - \mathbf{w}_i)(\mathbf{u}_i + \mathbf{w}_i) \\ &+\beta\sum_{i=1}^{2M}(\mathbf{u}_i - \mathbf{w}_i)\Bigg| \\ = \Bigg| &\frac{1}{2}\Big(||y_u - y_d|| - ||y_w - y_d||\Big)\Big(||y_u - y_d|| + ||y_w - y_d||\Big) \\ &+\frac{\alpha}{2}\sum_{i=1}^{2M}(\mathbf{u}_i - \mathbf{w}_i)(\mathbf{u}_i + \mathbf{w}_i) + \beta\sum_{i=1}^{2M}(\mathbf{u}_i - \mathbf{w}_i)\Bigg|, \end{aligned}$$

por la desigualdad triangular, se sigue que

$$|F(\mathbf{u}) - F(\mathbf{w})| \leq \frac{1}{2}\Big|\, ||y_u - y_d|| - ||y_w - y_d||\,\Big|\,\Big|\, ||y_u - y_d|| + ||y_w - y_d||\,\Big| + \frac{\alpha}{2}\sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i||\mathbf{u}_i + \mathbf{w}_i| + \beta \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i|\,.$$

Puesto $\mathbf{u}, \mathbf{w} \in \mathcal{V}_{ad}$ se tiene que $|\mathbf{u}_i + \mathbf{w}_i|$ es acotado para todo $i = 1, \ldots, 2M$, y gracias a la estimación (4.16) se sigue que $||y_u - y_d|| + ||y_w - y_d||$ también es acotado, con lo cual tenemos que

$$|F(\mathbf{u}) - F(\mathbf{w})| \leq \frac{1}{2}C_1\Big|\, ||y_u - y_d|| - ||y_w - y_d||\,\Big| + \frac{\alpha}{2}C_2 \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| + \beta \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i|\,,$$

con $C_1, C_2 > 0$ y usando nuevamente la desigualdad triangular, se sigue que

$$\begin{aligned}
|F(\mathbf{u}) - F(\mathbf{w})| &\leq c_1||y_u - y_d - (y_w - y_d)|| + c_2 \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| \\
&\leq c\left(||y_u - y_w|| + \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| \right) \\
&\leq c\left(\sum_{i=1}^{M} \big|(\mathbf{u}_i - \mathbf{u}_{i+M}) - (\mathbf{w}_i - \mathbf{w}_{i+M})\big|\, ||y_i|| + \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| \right) \\
&\leq c\left(\sum_{i=1}^{M} \big(|\mathbf{u}_i - \mathbf{w}_i| + |\mathbf{u}_{i+M} - \mathbf{w}_{i+M}|\big)\, ||y_i|| + \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| \right) \\
&\leq c\left(\max_{i=1,\ldots,M}\big\{||y_i||\big\} \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| + \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| \right),
\end{aligned}$$

con $c_1, c_2, c > 0$. Puesto que $\sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i|$ puede ser acotada por $||\mathbf{u} - \mathbf{w}||$, se obtiene que

$$|F(\mathbf{u}) - F(\mathbf{w})| \leq C||\mathbf{u} - \mathbf{w}||.$$

Por otra parte, de (4.32) se obtiene que

$$\begin{aligned}
||\nabla F(\mathbf{u}) - \nabla F(\mathbf{w})|| &\leq c \sum_{i=1}^{M} \Big(\big|(y_u - y_d, y_i) + \alpha\mathbf{u}_i + \beta - (y_w - y_d, y_i) - \alpha\mathbf{w}_i - \beta\big| \\
&\qquad + \big|-(y_u - y_d, y_i) + \alpha\mathbf{u}_{i+M} + \beta + (y_w - y_d, y_i) - \alpha\mathbf{w}_{i+M} - \beta\big| \Big) \\
&= c \sum_{i=1}^{M} \Big(\big|(y_u, y_i) + \alpha\mathbf{u}_i - (y_w, y_i) - \alpha\mathbf{w}_i\big| \\
&\qquad + \big|-(y_u, y_i) + \alpha\mathbf{u}_{i+M} + (y_w, y_i) - \alpha\mathbf{w}_{i+M}\big| \Big)
\end{aligned}$$

de lo cual se sigue

$$||\nabla F(\mathbf{u}) - \nabla F(\mathbf{w})|| \leq c \sum_{i=1}^{M} \Big(\big| (y_u - y_w, y_i) + \alpha(\mathbf{u}_i - \mathbf{w}_i) \big| + \big| -(y_u - y_w, y_i) + \alpha(\mathbf{u}_{i+M} - \mathbf{w}_{i+M}) \big| \Big),$$

con $c > 0$. Por la desigualdad triangular, se ve que

$$\begin{aligned} ||\nabla F(\mathbf{u}) - \nabla F(\mathbf{w})|| &\leq c \sum_{i=1}^{M} \Big(\big| (y_u - y_w, y_i) \big| + \alpha |\mathbf{u}_i - \mathbf{w}_i| \\ &\qquad + \big| -(y_u - y_w, y_i) \big| + \alpha |\mathbf{u}_{i+M} - \mathbf{w}_{i+M}| \Big) \\ &= c \left(2 \sum_{i=1}^{M} \big| (y_u - y_w, y_i) \big| + \alpha \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| \right), \end{aligned}$$

usando la desigualdad de Cauchy-Schwarz, tenemos

$$\begin{aligned} ||\nabla F(\mathbf{u}) - \nabla F(\mathbf{w})|| &\leq c \left(2 \sum_{i=1}^{M} ||y_u - y_w||\, ||y_i|| + \alpha \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| \right) \\ &\leq c \Big(c_1 ||y_u - y_w|| + \alpha \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| \Big), \end{aligned}$$

con $c_1 > 0$ y utilizando la desigualdad triangular, se sigue que

$$\begin{aligned} ||\nabla F(\mathbf{u}) - \nabla F(\mathbf{w})|| &\leq \widetilde{C} \left(\sum_{i=1}^{M} \big| (\mathbf{u}_i - \mathbf{u}_{i+M}) - (\mathbf{w}_i - \mathbf{w}_{i+M}) \big|\, ||y_i|| + \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| \right) \\ &\leq \widetilde{C} \left(\sum_{i=1}^{M} \Big(|\mathbf{u}_i - \mathbf{w}_i| + |\mathbf{u}_{i+M} - \mathbf{w}_{i+M}| \Big)\, ||y_i|| + \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| \right) \\ &\leq \widetilde{C} \left(\max_{i=1,\ldots,M} \{ ||y_i|| \} \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| + \sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i| \right), \end{aligned}$$

con $\widetilde{C} > 0$. Puesto que $\sum_{i=1}^{2M} |\mathbf{u}_i - \mathbf{w}_i|$ puede ser acotada por $||\mathbf{u} - \mathbf{w}||$, se obtiene que

$$||\nabla F(\mathbf{u}) - \nabla F(\mathbf{w})|| \leq C ||\mathbf{u} - \mathbf{w}||.$$

Finalmente, como $\nabla^2 F(\mathbf{u})$ es constante respecto a $\mathbf{u}$, se tiene que

$$||\nabla^2 F(\mathbf{u}) - \nabla^2 F(\mathbf{w})|| = 0 \leq C ||\mathbf{u} - \mathbf{w}||.$$

Sumando estás tres desigualdades se concluye (4.50). □

PROPOSICIÓN 4.21. *Existe una constante $C > 0$, independiente de h, tal que*

$$|F_h(\mathbf{u}) - F_h(\mathbf{w})| + ||\nabla F_h(\mathbf{u}) - \nabla F_h(\mathbf{w})|| + ||\nabla^2 F_h(\mathbf{u}) - \nabla^2 F_h(\mathbf{w})|| \leq C||\mathbf{u} - \mathbf{w}|| \tag{4.51}$$

se cumple para todo $\mathbf{u}, \mathbf{w} \in \mathcal{V}_{ad}$.

La demostración es similar a la demostración de la anterior proposición, usando además la acotación de y_i^h visto en (4.15).

PROPOSICIÓN 4.22. *La siguiente estimación*

$$|F(\mathbf{u}) - F_h(\mathbf{u})| + ||\nabla F(\mathbf{u}) - \nabla F_h(\mathbf{u})|| + ||\nabla^2 F(\mathbf{u}) - \nabla^2 F_h(\mathbf{u})|| \leq Ch^2, \quad \forall \mathbf{u} \in \mathcal{V}_{ad}$$

se cumple para alguna constante $C > 0$ independiente de h y $\mathbf{u}$.

Demostración. De la definición de F y F_h, se obtiene que

$$\begin{aligned} |F(\mathbf{u}) - F_h(\mathbf{u})| &= \left| \frac{1}{2} ||y_u - y_d||^2 - \frac{1}{2}||y_u^h - y_d||^2 \right| \\ &= \frac{1}{2} \left| ||y_u - y_d|| - ||y_u^h - y_d|| \right| \left| ||y_u - y_d|| + ||y_u^h - y_d|| \right|, \end{aligned}$$

usando la desigualdad triangular, se sigue que

$$\begin{aligned} |F(\mathbf{u}) - F_h(\mathbf{u})| &\leq \frac{1}{2}||y_u - y_d - y_u^h + y_d|| \Big(||y_u|| + ||y_u^h|| + 2||y_d|| \Big) \\ &= \frac{1}{2} \Big(||y_u|| + ||y_u^h|| + 2||y_d|| \Big) ||y_u - y_u^h||. \end{aligned}$$

Luego, de (4.16) y (4.17) se tiene que $||y_u|| + ||y_u^h|| + 2||y_d||$ es acotado, obteniendo así

$$|F(\mathbf{u}) - F_h(\mathbf{u})| \leq c||y_u - y_u^h||,$$

con $c > 0$. De la estimación (4.8) tenemos

$$|F(\mathbf{u}) - F_h(\mathbf{u})| \leq Ch^2.$$

Por otra parte, de (4.32) y (4.42) obtenemos la siguiente desigualdad

$$\begin{aligned}||\nabla F(\mathbf{u}) - \nabla F_h(\mathbf{u})|| &\leq c\sum_{i=1}^{M}\Big(\Big|(y_u - y_d, y_i) + \alpha\mathbf{u}_i + \beta - (y_u^h - y_d, y_i^h) - \alpha\mathbf{u}_i - \beta\Big| \\ &\quad + \Big|-(y_u - y_d, y_i) + \alpha\mathbf{u}_{i+M} + \beta + (y_u^h - y_d, y_i^h) - \alpha\mathbf{u}_{i+M} - \beta\Big|\Big) \\ &= c\sum_{i=1}^{M}\Big(\Big|(y_u - y_d, y_i) - (y_u^h - y_d, y_i^h)\Big| \\ &\quad + \Big|-(y_u - y_d, y_i) + (y_u^h - y_d, y_i^h)\Big|\Big) \\ &= 2c\sum_{i=1}^{M}\Big|(y_u - y_d, y_i) - (y_u^h - y_d, y_i^h)\Big|,\end{aligned}$$

con $c > 0$. Por la desigualdad triangular, se sigue que

$$\begin{aligned}||\nabla F(\mathbf{u}) - \nabla F_h(\mathbf{u})|| &\leq 2c\sum_{i=1}^{M}\Big(\Big|(y_u, y_i) - (y_u^h, y_i^h)\Big| + \Big|(y_d, y_i) - (y_d, y_i^h)\Big|\Big) \\ &= 2c\sum_{i=1}^{M}\Big(\Big|(y_u, y_i) - (y_u^h, y_i) - (y_u^h, y_i^h) + (y_u^h, y_i)\Big| \\ &\quad + \Big|(y_d, y_i - y_i^h)\Big|\Big) \\ &= 2c\sum_{i=1}^{M}\Big(\Big|(y_u - y_u^h, y_i) + (y_u^h, y_i - y_i^h)\Big| + \Big|(y_d, y_i - y_i^h)\Big|\Big) \\ &\leq 2c\sum_{i=1}^{M}\Big(\Big|(y_u - y_u^h, y_i)\Big| + \Big|(y_u^h, y_i - y_i^h)\Big| + \Big|(y_d, y_i - y_i^h)\Big|\Big),\end{aligned}$$

usando la desigualdad de Cauchy-Schwarz, tenemos

$$\begin{aligned}||\nabla F(\mathbf{u}) - \nabla F_h(\mathbf{u})|| &\leq 2c\sum_{i=1}^{M}\Big(||y_i||\,||y_u - y_u^h|| + ||y_u^h||\,||y_i - y_i^h|| + ||y_d||\,||y_i - y_i^h||\Big) \\ &= 2c\sum_{i=1}^{M}\Big(||y_i||\,||y_u - y_u^h|| + (||y_u^h|| + ||y_d||)||y_i - y_i^h||\Big),\end{aligned}$$

gracias a (4.17) se tiene que $||y_u^h|| + ||y_d||$ es acotado, con lo cual tenemos

$$||\nabla F(\mathbf{u}) - \nabla F_h(\mathbf{u})|| \leq 2c\left(c_1||y_u - y_u^h|| + c_2\sum_{i=1}^{M}||y_i - y_i^h||\right),$$

utilizando las estimaciones (4.3) y (4.8) se obtiene que

$$||\nabla F(\mathbf{u}) - \nabla F_h(\mathbf{u})|| \leq Ch^2.$$

Para la última estimación es necesario calcular $\nabla^2 F_h(\bar{\mathbf{u}})$, la cual está dada por

$$\nabla^2 F_h(\bar{\mathbf{u}}) = \left(\begin{array}{cccc|cccc} (y_1^h,y_1^h)+\alpha & (y_2^h,y_1^h) & \cdots & (y_M^h,y_1^h) & -(y_1^h,y_1^h) & -(y_2^h,y_1^h) & \cdots & -(y_M^h,y_1^h) \\ (y_1^h,y_2^h) & (y_2^h,y_2^h)+\alpha & \cdots & (y_M^h,y_2^h) & -(y_1^h,y_2^h) & -(y_2^h,y_2^h) & \cdots & -(y_M^h,y_2^h) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (y_1^h,y_M^h) & (y_2^h,y_M^h) & \cdots & (y_M^h,y_M^h)+\alpha & -(y_1^h,y_M^h) & -(y_2^h,y_M^h) & \cdots & -(y_M^h,y_M^h) \\ \hline -(y_1^h,y_1^h) & -(y_2^h,y_1^h) & \cdots & -(y_M^h,y_1^h) & (y_1^h,y_1^h)+\alpha & (y_2^h,y_1^h) & \cdots & (y_M^h,y_1^h) \\ -(y_1^h,y_2^h) & -(y_2^h,y_2^h) & \cdots & -(y_M^h,y_2^h) & (y_1^h,y_2^h) & (y_2^h,y_2^h)+\alpha & \cdots & (y_M^h,y_2^h) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -(y_1^h,y_M^h) & -(y_2^h,y_M^h) & \cdots & -(y_M^h,y_M^h) & (y_1^h,y_M^h) & (y_2^h,y_M^h) & \cdots & (y_M^h,y_M^h)+\alpha \end{array}\right),$$

por tanto, para todo $i,j=1,\ldots,M$ se tiene que

$$\begin{aligned} \left|(y_i,y_j)-(y_i^h,y_j^h)\right| &= \left|(y_i,y_j)-(y_i^h,y_j)-(y_i^h,y_j^h)+(y_i^h,y_j)\right| \\ &\leq \left|(y_i-y_i^h,y_j)\right|+\left|(y_i^h,y_j-y_j^h)\right| \\ &\leq ||y_j||\,||y_i-y_i^h||+||y_i^h||\,||y_j-y_j^h|| \\ &\leq c_1h^2+c_2||y_i^h||_{C(\bar{\Omega})}\,h^2 \\ &\leq Ch^2, \end{aligned}$$

lo que nos permite concluir

$$||\nabla^2 F(\mathbf{u})-\nabla^2 F_h(\mathbf{u})||\leq Ch^2.$$

Sumando estas desigualdades se obtiene la estimación deseada. □

PROPOSICIÓN 4.23. *Existe una constante $C>0$, independiente de h, tal que*

$$\begin{aligned} |F(\mathbf{u})-F_h(\mathbf{w})|+||\nabla F(\mathbf{u})-\nabla F_h(\mathbf{w})||+||\nabla^2 F(\mathbf{u})-\nabla^2 F_h(\mathbf{w})|| \\ +\sum_{j=1}^{l}\Big(|G_j(\mathbf{u})-G_j^h(\mathbf{w})|+||\nabla G_j(\mathbf{u})-\nabla G_j^h(\mathbf{w})|| \\ +||\nabla^2 G_j(\mathbf{u})-\nabla^2 G_j^h(\mathbf{w})||\Big)\leq C(||\mathbf{u}-\mathbf{w}||+h^2|\log h|) \end{aligned} \tag{4.52}$$

se cumple para todo $\mathbf{u},\mathbf{w}\in\mathcal{V}_{ad}$ y h suficientemente pequeño.

Demostración. De la desigualdad triangular se tiene que

$$\begin{aligned} |F(\mathbf{u})-F_h(\mathbf{w})|+||\nabla F(\mathbf{u})-\nabla F_h(\mathbf{w})||+||\nabla^2 F(\mathbf{u})-\nabla^2 F_h(\mathbf{w})|| &\leq |F(\mathbf{u})-F(\mathbf{w})| \\ &+|F(\mathbf{w})-F_h(\mathbf{w})|+||\nabla F(\mathbf{u})-\nabla F(\mathbf{w})||+||\nabla F(\mathbf{w})-\nabla F_h(\mathbf{w})|| \\ &+||\nabla^2 F(\mathbf{u})-\nabla^2 F(\mathbf{w})||+||\nabla^2 F(\mathbf{w})-\nabla^2 F_h(\mathbf{w})||, \end{aligned}$$

luego, por la Proposición 4.20 y 4.22 se obtiene que

$$|F(\mathbf{u}) - F_h(\mathbf{w})| + ||\nabla F(\mathbf{u}) - \nabla F_h(\mathbf{w})|| \\ + ||\nabla^2 F(\mathbf{u}) - \nabla^2 F_h(\mathbf{w})|| \leq C(||\mathbf{u} - \mathbf{w}|| + h^2).$$

Así, al sumar la desigualdad de la Proposición 4.19, y puesto que $h^2 \leq h^2|\log h|$ para h suficientemente pequeño, se concluye la estimación (4.52). □

Para continuar con nuestro análisis tomamos en cuenta el conjunto $K = \mathbb{R}^l_+$. Por tanto es posible reescribir (4.46) y (4.47) como

$$\begin{cases} \text{mín } F(\mathbf{u}) \\ G(\mathbf{u}) \leq_K 0, \\ \mathbf{u} \in \mathcal{V}_{ad}, \end{cases} \tag{4.53}$$

y

$$\begin{cases} \text{mín } F_h(\mathbf{u}) \\ G_h(\mathbf{u}) \leq_K 0, \\ \mathbf{u} \in \mathcal{V}_{ad}, \end{cases} \tag{4.54}$$

respectivamente, considerando las primeras l componentes de G y G_h.

Nuestro objetivo es mostrar que existe una sucesión $\bar{\mathbf{u}}^h$ de soluciones para el problema (4.39) que convergen a $\bar{\mathbf{u}}$ cuando h tiende a 0. Para esto, se considerará el problema auxiliar

$$\begin{cases} \text{mín } F_h(\mathbf{u}) \\ G_h(\mathbf{u}) \leq_K 0, \\ \mathbf{u} \in \mathcal{V}_{ad} \cap \bar{B}(\bar{\mathbf{u}}, \rho), \end{cases}$$

para algún $\rho > 0$ que será especificado en lo posterior. Se probará que este problema admite una única solución $\bar{\mathbf{u}}^h_\rho$ y satisface cierta estimación con respecto a $\bar{\mathbf{u}}$. Esto nos permitirá demostrar que:

- Si ρ es suficientemente pequeño entonces $\bar{\mathbf{u}}^h_\rho \in B(\bar{\mathbf{u}}, \rho)$, es decir $\bar{\mathbf{u}}^h_\rho$ no está en la frontera de $B(\bar{\mathbf{u}}, \rho)$, por lo tanto $\bar{\mathbf{u}}^h_\rho$ es una solución local de (4.39) que por ser un problema estrictamente convexo implica que $\bar{\mathbf{u}}^h_\rho$ es la única solución óptima.

- Bajo las mismas hipótesis se cumple la estimación $||\bar{\mathbf{u}} - \bar{\mathbf{u}}^h_\rho|| \leq h^2|\log h|$, mediante el uso de ecuaciones generalizadas.

OBSERVACIÓN. $\mathbf{u} \in \mathcal{V}_{feas}$ si y solo si $G(\mathbf{u}) \leq_K 0$ y $\mathbf{u} \in \mathcal{V}_{ad}$. De igual manera, $\mathbf{u} \in \mathcal{V}_{feas}^h$ es equivalente a $\mathbf{u} \in \mathcal{V}_{ad}$ y $G_h(\mathbf{u}) \leq_K 0$.

Recordemos que $\bar{\mathbf{u}}$ denota la solución óptima de (4.23), por tanto $\bar{\mathbf{u}}$ es solución de (4.53). Además, satisface la condición (LICQ). Con esto procedemos a seguir con nuestro análisis de perturbación.

OBSERVACIÓN. La Proposición 2.18 implica que $\bar{\mathbf{u}}$ satisface la condición de regularidad de Robinson.

Para la demostración del siguiente resultado usaremos las definiciones y resultados de la Sección 2.3 referente a ecuaciones generalizadas.

PROPOSICIÓN 4.24. *Existen constantes positivas C y h_0, tal que para cada $h \in (0, h_0)$ existe $\mathbf{u}^h \in \mathcal{V}_{feas}^h$ y cumple la estimación*

$$||\bar{\mathbf{u}} - \mathbf{u}^h|| \leq Ch^2 |\log h|.$$

Demostración. Consideremos la función $\mathbf{G} : \mathbb{R}_+ \times \mathbb{R}^{2M} \to \mathbb{R}^l$ definida por

$$\mathbf{G}(h, \mathbf{u}) = \begin{cases} G(\mathbf{u}), & \text{si } h = 0, \\ G_h(\mathbf{u}), & \text{si } h \neq 0. \end{cases}$$

y la tripleta $\{\mathbb{R}_+, 0, \mathbf{G}(\cdot, \cdot)\}$, de acuerdo a la Definición 2.19, es una perturbación admisible para el sistema $G(\mathbf{u}) \leq_k 0$ en $\bar{\mathbf{u}}$. En efecto, tenemos que $\mathbf{G}(0, \mathbf{u}) = G(\mathbf{u})$ para todo $\mathbf{u} \in \mathbb{R}^{2M}$, y de la Proposición 4.19 se deduce que $\mathbf{G}$ y $\partial \mathbf{G}/\partial \mathbf{u}$ son continuas en $(0, \bar{\mathbf{u}})$. Además, tenemos que la condición de regularidad de Robinson se cumple en $\bar{\mathbf{u}}$, por lo cual cumplimos con las hipótesis de la Proposición 2.12, de la cual obtenemos que existen $h_0 > 0$, $c > 0$ y una vecindad $\mathcal{O}$ de $\bar{\mathbf{u}}$, tal que para todo $h \in (0, h_0)$ el sistema $\mathbf{G}(h, \mathbf{u}) \leq_K 0$ tiene solución y se cumple que

$$\text{dist}[\mathbf{w}, \Sigma(h)] \leq c \, \text{dist}[0, \mathcal{G}(h, \mathbf{w})], \quad \forall h \in (0, h_0), \forall \mathbf{w} \in \mathcal{O}, \tag{4.55}$$

donde $\Sigma(h) = \{\mathbf{u} \in \mathcal{V}_{ad} : \mathbf{G}(h, \mathbf{u}) \leq_K 0\}$ y $\mathcal{G}(h, \mathbf{w})$ está definido por

$$\mathcal{G}(h, \mathbf{w}) = \begin{cases} \mathbf{G}(h, \mathbf{w}) + K, & \text{si } \mathbf{w} \in \mathcal{V}_{ad}, \\ \emptyset, & \text{si } \mathbf{w} \notin \mathcal{V}_{ad}. \end{cases}$$

En particular, si tomamos $\mathbf{w} = \bar{\mathbf{u}}$ se tiene que

$$\begin{aligned}\operatorname{dist}[\bar{\mathbf{u}}, \Sigma(h)] &\leq c \operatorname{dist}[0, \mathcal{G}(h, \bar{\mathbf{u}})] \\ &= c \operatorname{dist}[0, \mathcal{G}(0, \bar{\mathbf{u}}) + \mathcal{G}(h, \bar{\mathbf{u}}) - \mathcal{G}(0, \bar{\mathbf{u}})],\end{aligned}$$

y por tratarse de una distancia se sigue que

$$\begin{aligned}\operatorname{dist}[\bar{\mathbf{u}}, \Sigma(h)] &\leq c(\operatorname{dist}[0, \mathcal{G}(0, \bar{\mathbf{u}})] + \operatorname{dist}[0, \mathcal{G}(h, \bar{\mathbf{u}}) - \mathcal{G}(0, \bar{\mathbf{u}})]) \\ &= c(\operatorname{dist}[0, \mathbf{G}(0, \bar{\mathbf{u}}) + K] + \operatorname{dist}[0, \mathbf{G}(h, \bar{\mathbf{u}}) + K - \mathbf{G}(0, \bar{\mathbf{u}}) - K]) \\ &= c(\operatorname{dist}[0, G(\bar{\mathbf{u}}) + K] + \operatorname{dist}[0, G_h(\bar{\mathbf{u}}) + K - G(\bar{\mathbf{u}}) - K]) \\ &\leq c(0 + ||G_h(\bar{\mathbf{u}}) - G(\bar{\mathbf{u}})||) \\ &\leq Ch^2 |\log h|.\end{aligned}$$

Luego, por la caracterización de ínfimo existe $\mathbf{u}^h \in \Sigma(h)$ tal que

$$||\bar{\mathbf{u}} - \mathbf{u}^h|| \leq Ch^2 |\log h|,$$

y puesto que $\Sigma(h) = \mathcal{V}^h_{feas}$, se concluye que $\mathbf{u}^h \in \mathcal{V}^h_{feas}$. □

PROPOSICIÓN 4.25. *Para $\rho > 0$, el problema auxiliar*

$$\begin{cases} \text{mín } F_h(\mathbf{u}) \\ G_h(\mathbf{u}) \leq_K 0, \\ \mathbf{u} \in \mathcal{V}_{ad} \cap \bar{B}(\bar{\mathbf{u}}, \rho), \end{cases} \tag{4.56}$$

tiene una única solución $\bar{\mathbf{u}}^h$, para h suficientemente pequeño. Además, existe un elemento $\mathbf{v}^h \in \mathcal{V}_{feas}$ que cumple la estimación

$$||\bar{\mathbf{u}}^h - \mathbf{v}^h|| \leq Ch^2 |\log h|.$$

con $C > 0$ independiente de h.

Demostración. De la Proposición 4.24 se tiene que existe algún $h'_0 > 0$, tal que para todo $h \in (0, h'_0)$ existe $\mathbf{u}^h \in \mathcal{V}^h_{feas}$ que satisface $||\bar{\mathbf{u}} - \mathbf{u}^h|| \leq Ch^2 |\log h|$. Con esto, se sigue que existe $h_0(\rho)$ tal que para todo $h \in (0, h_0(\rho))$ implica que $||\bar{\mathbf{u}} - \mathbf{u}^h|| \leq \rho$. Por lo tanto $\mathbf{u}^h \in \mathcal{V}^h_{feas} \cap \bar{B}(\bar{\mathbf{u}}, \rho)$ por tanto, el conjunto factible de (4.56) es distinto del vacío. Gracias al teorema de Weierstrass podemos concluir que (4.56) tiene una única solución óptima $\bar{\mathbf{u}}^h$.

Ahora, para encontrar un $\mathbf{v}^h \in \mathcal{V}_{feas}$ tal que $||\bar{\mathbf{u}}^h - \mathbf{v}^h|| \leq Ch^2 |\log h|$, vamos a construir una ecuación cuya solución sea $\mathbf{v}^h$.

Construcción de la ecuación para $\mathbf{v}^h$*:* Denotamos por $I(\bar{\mathbf{u}})$ el conjunto de los índices de las componentes inactivas de G en $\bar{\mathbf{u}}$. Probaremos que si ρ es suficientemente pequeño entonces todas las componentes inactivas de G en $\bar{\mathbf{u}}$ son también inactivas en $\bar{\mathbf{u}}^h$. En efecto, si $i \in I(\bar{\mathbf{u}})$ entonces $G_i(\bar{\mathbf{u}}) < 0$ y como G_i es continua, existe una vecindad V de $\bar{\mathbf{u}}$ tal que para todo $\mathbf{u} \in V$ se cumple $G_i(\mathbf{u}) < 0$. Por otro lado sabemos que $\bar{\mathbf{u}}^h \in \bar{B}(\bar{\mathbf{u}}, \rho)$ para todo $h < h_0$, por lo cual, si tomamos ρ lo suficientemente pequeño se concluye que $G_i(\bar{\mathbf{u}}^h) < 0$.

Luego, suponemos que existen r componentes de G que son activas en $\bar{\mathbf{u}}$ y reordenándolas si es necesario, podemos asumir que $\mathcal{A}(\bar{\mathbf{u}}) = \{1, 2, \ldots, r\}$, lo que significa $G_1(\bar{\mathbf{u}}) = \cdots = G_r(\bar{\mathbf{u}}) = 0$. Además, de (4.33) se sabe que ∇G_i es una función constante, entonces $\nabla G_1(\bar{\mathbf{u}}) = \nabla G_1(\bar{\mathbf{u}}^h), \ldots, \nabla G_r(\bar{\mathbf{u}}) = \nabla G_r(\bar{\mathbf{u}}^h)$ y por la condición (LICQ) en $\bar{\mathbf{u}}$ se obtiene que $\nabla G_1(\bar{\mathbf{u}}^h), \ldots, \nabla G_r(\bar{\mathbf{u}}^h)$ son linealmente independientes. De lo anterior, la matriz

$$B_h = [\nabla G_1(\bar{\mathbf{u}}^h), \ldots, \nabla G_r(\bar{\mathbf{u}}^h)]$$

tiene rango r, y reordenado las filas, si es necesario, podemos encontrar una submatriz invertible D_h de orden $r \times r$, obteniendo así la siguiente descomposición

$$B_h = \begin{bmatrix} D_h \\ E_h \end{bmatrix},$$

con E_h una submatriz de orden $(2M - r) \times r$. Consideremos ahora $\psi_h : \mathbb{R}^r \to \mathbb{R}^r$ la función definida por

$$\psi_{h,i}(w) = G_i(w, \bar{\mathbf{u}}^h_{r+1}, \ldots, \bar{\mathbf{u}}^h_{2M}) - G^h_i(\bar{\mathbf{u}}^h), \quad \forall i = 1, \ldots, r.$$

Así, para encontrar $\mathbf{v}^h$ fijamos sus $2M - r$ últimas componentes por

$$\mathbf{v}^h_i = \bar{\mathbf{u}}^h_i, \quad \forall i = r + 1, \ldots, 2M.$$

Resta determinar sus primeras r componentes, las cuales van a ser la solución del siguiente sistema

$$\psi_h(w) = 0, \tag{4.57}$$

es decir, las primeras r componentes de $\mathbf{v}^h$ van a estar definidas por

$$\mathbf{v}^h_i = w_i, \quad \forall i = 1, \ldots, r.$$

Solución de (4.57): Definimos $\bar{w}^h = (\bar{\mathbf{u}}^h_1, \ldots, \bar{\mathbf{u}}^h_r)^T$, $\bar{w} = (\bar{\mathbf{u}}_1, \ldots, \bar{\mathbf{u}}_r)^T$. Por otro lado

de la Proposición 4.19 se tiene que

$$||\psi_h(\bar{w}^h)|| = ||G(\bar{\mathbf{u}}^h) - G_h(\bar{\mathbf{u}}^h)|| \leq ch^2|\log h|. \tag{4.58}$$

Tomando en cuenta $\psi_h'(w) = [\nabla_w G_1(w, \bar{\mathbf{u}}_{r+1}^h, \ldots, \bar{\mathbf{u}}_{2M}^h), \ldots, \nabla_w G_r(w, \bar{\mathbf{u}}_{r+1}^h, \ldots, \bar{\mathbf{u}}_{2M}^h)]$, donde ∇_w denota el gradiente respecto a w, se tiene de (4.33), que $\nabla_w G_i$ es constante por lo tanto ψ_h' es constante, obteniendo así

$$||\psi_h'(w_1) - \psi_h'(w_2)|| = 0 \leq \gamma||w_1 - w_2||, \qquad \forall w_1, w_2 \in B(\bar{w}, \rho),$$

puesto que $\psi_h'(w)$ es de rango completo, se sigue que $(\psi_h'(w))^{-1}$ existe y cumple

$$||(\psi_h'(w))^{-1}|| \leq \beta, \qquad \forall w \in B(\bar{w}, \rho),$$

para todo $0 \leq h < h_0$, si ρ es tomado suficientemente pequeño. De aquí obtenemos que

$$||(\psi_h'(w))^{-1}\psi_h(\bar{w}_h)|| \leq ||(\psi_h'(w))^{-1}||\ ||\psi_h(\bar{w}_h)|| \leq \beta||\psi_h(\bar{w}_h)||,$$

si definimos $\eta := \beta|\psi_h(\bar{w}_h)|$ se sigue que

$$||(\psi_h'(w))^{-1}\psi_h(\bar{w}_h)|| \leq \eta,$$

y de (4.58) se tiene que

$$\frac{1}{2}\beta\gamma\eta = \frac{1}{2}\beta^2\gamma||\psi_h(\bar{w}_h)|| \leq \frac{1}{2}\beta^2\gamma ch^2|\log h| < 1,$$

si h es suficientemente pequeño. Entonces estamos en las hipótesis del Teorema Newton-Mysovskii (página 412, de [23]), el cual nos asegura que el método de Newton con punto inicial $w_0 = \bar{w}_h$ genera una solución w de (4.57) en $\bar{B}(\bar{w}_h, c_0\eta)$ (donde c_0 es una cierta constante) y por ser solución de (4.57), se sigue que

$$\psi_h(w) = G_i(\mathbf{v}^h) - G_i^h(\bar{\mathbf{u}}^h) = G_i(w, \bar{\mathbf{u}}_{r+1}^h, \ldots, \bar{\mathbf{u}}_{2M}^h) - G_i^h(\bar{\mathbf{u}}^h) = 0,$$

para todo $i = 1, \ldots, r$. Lo anterior implica que

$$G_i(\mathbf{v}^h) = G_i^h(\bar{\mathbf{u}}^h) \leq 0, \quad \forall i = 1, \ldots, r,$$

pues $\bar{\mathbf{u}}^h$ es solución de (4.56). Finalmente, puesto que $w \in \bar{B}(w_h, c_0\eta)$ se sigue que $||w - w_h|| \leq c_0\eta \leq ch^2|\log h|$ lo que implica $||\mathbf{v}^h - \bar{\mathbf{u}}^h|| \leq ch^2|\log h|$. Además, para h suficientemente pequeño se tiene que $G_j(\mathbf{v}^h) < 0$ para todo $j = r+1, \ldots, l+4M$, pues son los índices de inactividad, por lo tanto se concluye que $\mathbf{v}^h \in \mathcal{V}_{feas}$. □

PROPOSICIÓN 4.26. *Si $\rho > 0$ suficientemente pequeño y $h \in (0, h_0(\rho))$ entonces la*

solución $\bar{\mathbf{u}}^h$ del problema auxiliar (4.56) pertenece a $B(\bar{\mathbf{u}}, \rho)$, es decir $\bar{\mathbf{u}}^h$ es la solución óptima del problema (4.39).

Demostración. Sea $\bar{\mathbf{u}}^h$ solución de (4.56), de la Proposición 4.24 sabemos que existe $\mathbf{u}^h \in \mathcal{V}^h_{feas} \cap \bar{B}(\bar{\mathbf{u}}, \rho)$ que se aproxima a $\bar{\mathbf{u}}$ con orden $h^2|\log h|$. Por la optimalidad de $\bar{\mathbf{u}}^h$ se sigue que

$$F_h(\bar{\mathbf{u}}^h) \leq F_h(\mathbf{u}^h) \leq |F_h(\mathbf{u}^h) - F_h(\bar{\mathbf{u}})| + |F_h(\bar{\mathbf{u}}) - F(\bar{\mathbf{u}})| + F(\bar{\mathbf{u}}),$$

sumando los términos $F(\bar{\mathbf{u}}^h) - F_h(\bar{\mathbf{u}}^h)$, se obtiene que

$$F(\bar{\mathbf{u}}^h) - F_h(\bar{\mathbf{u}}^h) + F_h(\bar{\mathbf{u}}^h) \leq |F_h(\mathbf{u}^h) - F_h(\bar{\mathbf{u}})| + |F_h(\bar{\mathbf{u}}) - F(\bar{\mathbf{u}})| + F(\bar{\mathbf{u}}^h) - F_h(\bar{\mathbf{u}}^h) + F(\bar{\mathbf{u}}),$$

por lo tanto

$$F(\bar{\mathbf{u}}^h) \leq |F_h(\mathbf{u}^h) - F_h(\bar{\mathbf{u}})| + |F_h(\bar{\mathbf{u}}) - F(\bar{\mathbf{u}})| + |F(\bar{\mathbf{u}}^h) - F_h(\bar{\mathbf{u}}^h)| + F(\bar{\mathbf{u}}).$$

Ahora, de la Proposición 4.21, se tiene que

$$F(\bar{\mathbf{u}}^h) \leq ||\mathbf{u}^h - \bar{\mathbf{u}}|| + |F_h(\bar{\mathbf{u}}) - F(\bar{\mathbf{u}})| + |F(\bar{\mathbf{u}}^h) - F_h(\bar{\mathbf{u}}^h)| + F(\bar{\mathbf{u}}),$$

y de las Proposiciones 4.23 y 4.24, obtenemos

$$F(\bar{\mathbf{u}}^h) \leq c_1\, h^2|\log h| + F(\bar{\mathbf{u}}). \tag{4.59}$$

Por otra parte, de la Proposición 4.25 existe $\mathbf{v}^h \in \mathcal{V}_{feas}$ que se aproxima a $\bar{\mathbf{u}}^h$ con orden $h^2|\log h|$. Luego, puesto que $\mathbf{v}^h$ es tan cercano a $\bar{\mathbf{u}}$, para ρ y h suficientemente pequeños, se sigue de la Proposición 4.16 que

$$\omega||\bar{\mathbf{u}} - \mathbf{v}^h||^2 \leq F(\mathbf{v}^h) - F(\bar{\mathbf{u}}),$$

por lo tanto

$$F(\bar{\mathbf{u}}) + \omega||\bar{\mathbf{u}} - \mathbf{v}^h||^2 \leq F(\mathbf{v}^h),$$

por la desigualdad triangular, se ve que

$$F(\bar{\mathbf{u}}) + \omega||\bar{\mathbf{u}} - \mathbf{v}^h||^2 \leq |F(\mathbf{v}^h) - F(\bar{\mathbf{u}}^h)| + F(\bar{\mathbf{u}}^h).$$

Usando la Proposición 4.20, obtenemos que

$$F(\bar{\mathbf{u}}) + \omega||\bar{\mathbf{u}} - \mathbf{v}^h||^2 \leq ||\mathbf{v}^h - \bar{\mathbf{u}}^h|| + F(\bar{\mathbf{u}}^h),$$

así, de la Proposición 4.25

$$F(\bar{\mathbf{u}}) + \omega||\bar{\mathbf{u}} - \mathbf{v}^h||^2 \leq c_2\, h^2|\log h| + F(\bar{\mathbf{u}}^h), \tag{4.60}$$

combinando (4.59) y (4.60), se obtiene que

$$F(\bar{\mathbf{u}}) + \omega||\bar{\mathbf{u}} - \mathbf{v}^h||^2 \leq c\, h^2|\log h| + F(\bar{\mathbf{u}}),$$

y por consiguiente

$$||\bar{\mathbf{u}} - \mathbf{v}^h|| \leq c\sqrt{h^2|\log h|}. \tag{4.61}$$

Por otro lado, se tiene que

$$||\bar{\mathbf{u}} - \bar{\mathbf{u}}^h|| \leq ||\bar{\mathbf{u}} - \mathbf{v}^h|| + ||\bar{\mathbf{u}}^h - \mathbf{v}^h|| \leq ||\bar{\mathbf{u}} - \mathbf{v}^h|| + c\, h^2|\log h|,$$

y de (4.61), se concluye que

$$||\bar{\mathbf{u}} - \bar{\mathbf{u}}^h|| \leq c\left(\sqrt{h^2|\log h|} + \ h^2|\log h|\right). \tag{4.62}$$

Finalmente, para h suficientemente pequeño se tiene que $||\bar{\mathbf{u}} - \bar{\mathbf{u}}^h|| < \rho$, por lo tanto $\bar{\mathbf{u}}^h \in B(\bar{\mathbf{u}}, \rho)$. Con lo cual, $\bar{\mathbf{u}}^h$ es la solución óptimo de (4.39). □

OBSERVACIÓN. La estimación (4.62) de la Proposición 4.26 nos permite concluir la existencia de una sucesión $\bar{\mathbf{u}}_h$ de soluciones del problema (4.39) que convergen a $\bar{\mathbf{u}}$, cuando $h \to 0$.

En lo que sigue denotamos por $\bar{\mathbf{u}}^h$ la solución óptima de (4.39). Para proseguir con nuestro estudio, consideramos nuevamente todas las componentes de G, con lo cual $K = \mathbb{R}^{l+4M}_+$ es el nuevo cono asociado a nuestras restricciones. Entonces el cono dual K_+ asociado a K está definido por

$$K_+ = \{z \in \mathbb{R}^{l+4M} : z^T v \geq 0, \quad \forall v \in K\},$$

y puesto que $K = \mathbb{R}^{l+4M}_+$, se sigue que

$$K_+ = \{z \in \mathbb{R}^{l+4M} : z_i \geq 0, \quad i = 1, \ldots, l + 4M\} = \mathbb{R}^{l+4M}_+.$$

Recordemos ahora que el cono normal $\partial\psi_E(x)$ de un conjunto convexo $E \subset \mathbb{R}^{2M}$ en un punto $x \in \mathbb{R}^{2M}$ está definido por

$$\partial\psi_E(x) = \begin{cases} z \in \mathbb{R}^{2M} \text{ con } z^T(e - x) \leq 0 \quad \forall e \in E, & \text{si } x \in E \\ \emptyset, & \text{si } x \notin E. \end{cases}$$

Por facilidad, consideramos los multiplicadores de Lagrange estructurados de la siguiente forma

$$\bar{\lambda} = \begin{pmatrix} \bar{\mu} \\ \bar{\nu} \end{pmatrix} \in \mathbb{R}_+^{l+4M}.$$

Es claro que las condiciones de complementariedad (4.36) pueden ser escritas como

$$G(\bar{\mathbf{u}}) \leq_K 0, \bar{\lambda} \in K_+ \text{ y } G(\bar{\mathbf{u}})^T\bar{\lambda} = 0, \tag{4.63}$$

las cuales son equivalentes a $G(\bar{\mathbf{u}}) \in \partial\psi_{K_+}(\bar{\lambda})$. En efecto, suponemos (4.63) entonces $G(\bar{\mathbf{u}}) \leq_K 0$ implica $G(\bar{\mathbf{u}})^T\eta \leq 0$ para todo $\eta \in K_+$, de lo cual $G(\bar{\mathbf{u}})^T(\eta - \bar{\lambda}) \leq 0$ pues $G(\bar{\mathbf{u}})^T\bar{\lambda} = 0$ y puesto que $\bar{\lambda} \in K_+$ se concluye $G(\bar{\mathbf{u}}) \in \partial\psi_{K_+}(\bar{\lambda})$. Recíprocamente, si $G(\bar{\mathbf{u}}) \in \partial\psi_{K_+}(\bar{\lambda})$ entonces $\bar{\lambda} \in K_+$, con lo cual obtenemos que $G(\bar{\mathbf{u}})^T(\eta - \bar{\lambda}) \leq 0$ para todo $\eta \in K_+$, en particular si tomamos $\eta = 0$ y $\eta = 2\bar{\lambda}$ se sigue que $G(\bar{\mathbf{u}})^T\bar{\lambda} = 0$, luego si tomamos $\eta = \nu + \bar{\lambda}$, con $\nu \in K_+$ arbitrario, obtenemos que $G(\bar{\mathbf{u}})^T\nu \leq 0$ y puesto que ν es arbitrario en K_+ se concluye que $G_i(\bar{\mathbf{u}}) \leq 0$ con $i = 1, \ldots, l$ es decir $G(\bar{\mathbf{u}}) \leq_K 0$.

Ahora, vamos a expresar las condiciones necesarias de primer orden para el problema (4.23), en la forma de una ecuación generalizada. Para ello, observemos que las condiciones necesarias de primer orden pueden ser expresadas como

$$0 \in \mathcal{F}(\bar{\mathbf{u}}, \bar{\lambda}) + \mathcal{T}(\bar{\mathbf{u}}, \bar{\lambda}),$$

donde $\mathcal{F}$ es una función definida en $\mathbb{R}^{2M} \times \mathbb{R}^{l+4M}$ tal que

$$\mathcal{F}(\bar{\mathbf{u}}, \bar{\lambda}) = \begin{bmatrix} \nabla_{\mathbf{u}}\mathcal{L}(\bar{\mathbf{u}}, \bar{\lambda}) \\ -G(\bar{\mathbf{u}}) \end{bmatrix},$$

mientras que $\mathcal{T}$ es una función cuyas imagenes son conjuntos, definida por

$$\mathcal{T}(\bar{\mathbf{u}}, \bar{\lambda}) = \{0\} \times \partial\psi_{K_+}(\bar{\lambda}).$$

De la definición del cono normal en $\mathbb{R}^{2M}$ se tiene que $\{0\} = \partial\psi_{\mathbb{R}^{2M}}(\bar{\mathbf{u}})$ y usando la Proposición 2.13, se obtiene que

$$\mathcal{T}(\bar{\mathbf{u}}, \bar{\lambda}) = \partial\psi_{\mathbb{R}^{2M}}(\bar{\mathbf{u}}) \times \partial\psi_{K_+}(\bar{\lambda}) = \partial\psi_{\mathbb{R}^{2M}\times K_+}(\bar{\mathbf{u}}, \bar{\lambda}).$$

Por lo tanto, la solución de la ecuación generalizada

$$0 \in \mathcal{F}(\mathbf{u}, \lambda) + \partial\psi_{\mathbb{R}^{2M}\times K_+}(\mathbf{u}, \lambda) \tag{4.64}$$

expresa las condiciones necesarias de primer orden para el problema (4.23).

De manera análoga, el sistema de optimalidad para el problema discretizado (4.39) es equivalente a la solución de la ecuación generalizada

$$0 \in \mathcal{F}_h(\mathbf{u}^h, \lambda^h) + \partial \psi_{\mathbb{R}^{2M} \times K_+}(\mathbf{u}^h, \lambda^h),$$

donde $\mathcal{F}_h$ está definida por

$$\mathcal{F}_h(\mathbf{u}, \lambda) = \begin{bmatrix} \nabla_{\mathbf{u}} \mathcal{L}_h(\mathbf{u}, \lambda) \\ -G_h(\mathbf{u}) \end{bmatrix}.$$

PROPOSICIÓN 4.27. *Si h es suficientemente pequeño, $\bar{\mathbf{u}}^h$ satisface la condición (LICQ) para el problema (4.39).*

Demostración. De la Proposición 4.19 y la estimación (4.62), se tiene que

$$||\nabla G_j(\bar{\mathbf{u}}) - \nabla G_j^h(\bar{\mathbf{u}}^h)|| \leq C(||\bar{\mathbf{u}} - \bar{\mathbf{u}}^h|| + h^2|\log h|) \leq C\left(\sqrt{h^2|\log h|} + h^2|\log h|\right),$$

por lo tanto, si $G_j(\bar{\mathbf{u}}) < 0$ entonces $G_j^h(\bar{\mathbf{u}}^h) < 0$, para h suficientemente pequeño. Es decir, si j es un índice inactivo del problema (4.53) en $\bar{\mathbf{u}}$ implica que j es inactivo para el problema (4.54) en $\bar{\mathbf{u}}^h$.

Resta analizar los índices actividad del problema (4.53) en $\bar{\mathbf{u}}$, puesto que $\bar{\mathbf{u}}$ satisface la condición (LICQ) para el problema (4.23). Para todo $j \in \mathcal{A}(\bar{\mathbf{u}})$ los $\nabla G_j(\bar{\mathbf{u}})$ son linealmente independientes, esto implica que para todo $j \in \mathcal{A}(\bar{\mathbf{u}})$ los $\nabla G_j^h(\bar{\mathbf{u}}^h)$ también son linealmente independientes, siempre que h sea suficientemente pequeño. Por consiguiente en todos los índices de actividad del problema (4.54) en $\bar{\mathbf{u}}^h$ los $\nabla G_j^h(\bar{\mathbf{u}}^h)$ son linealmente independiente, es decir $\bar{\mathbf{u}}^h$ satisface la condición (LICQ) para el problema (4.39). □

OBSERVACIÓN. De este resultado y la Proposición 4.44, existe

$$\bar{\lambda}^h = \begin{pmatrix} \bar{\mu}^h \\ \bar{\nu}^h \end{pmatrix} \in \mathbb{R}_+^{l+4M},$$

que satisface las condiciones de optimalidad de primer orden para el problema (4.39), siempre que h sea suficientemente pequeño.

PROPOSICIÓN 4.28. *Sea $\bar{\mathbf{u}}^h$ la sucesión de soluciones del problema (4.39) que convergen a $\bar{\mathbf{u}}$ cuando $h \to 0$, entonces los multiplicadores $\bar{\lambda}^h$ asociado a $\bar{\mathbf{u}}^h$ son uniformente acotados para todo $h > 0$ suficientemente pequeño.*

Demostración. Puesto que $\bar{\mathbf{u}}^h \to \bar{\mathbf{u}}$ cuando $h \to 0$, se tiene que $\bar{\mathbf{u}}^h$ es una sucesión acotada. Luego, si h es suficientemente pequeño se sigue que todas las componentes inactivas de G en $\bar{\mathbf{u}}$ son también componentes inactivas de G_h en $\bar{\mathbf{u}}^h$. En consecuencia, $(\bar{\mathbf{u}}^h, \bar{\lambda}^h)$ satisface

$$\nabla_u \mathcal{L}_h(\bar{\mathbf{u}}^h, \bar{\lambda}^h) = \nabla F_h(\bar{\mathbf{u}}^h) + \sum_{i \in \mathcal{A}(\bar{\mathbf{u}})} \bar{\lambda}_i^h \nabla G_i^h(\bar{\mathbf{u}}^h) = 0. \tag{4.65}$$

Reordenando si es necesario, suponemos que $\mathcal{A}(\bar{\mathbf{u}}) = \{1, \ldots, r\}$, entonces la matriz

$$B_h = [\nabla G_1^h(\bar{\mathbf{u}}^h), \ldots, \nabla G_r^h(\bar{\mathbf{u}}^h)]$$

tiene rango r, y reordenado las filas, si es necesario, podemos encontrar una submatriz invertible D_h de orden $r \times r$, obteniendo así la siguiente descomposición

$$B_h = \begin{bmatrix} D_h \\ E_h \end{bmatrix},$$

con E_h una submatriz de orden $(2M - r) \times r$. Entonces de (4.65) se tiene que

$$\bar{\lambda}_h = -D_h^{-1} \left[\nabla F_h(\bar{\mathbf{u}}^h) \right]_{i=1}^r, \tag{4.66}$$

en la demostración de la Proposición 4.25 se obtuvo que D_h^{-1} es una matriz acotada, para h suficientemente pequeño. Verificamos que $\left[\nabla F_h(\bar{\mathbf{u}}^h)\right]_{i=1}^r$ es acotada. De la Proposición 4.23 y 4.26 se obtiene que

$$\begin{aligned} ||\nabla F_h(\bar{\mathbf{u}}^h) - \nabla F(\bar{\mathbf{u}})|| &\leq c(||\bar{\mathbf{u}}^h - \bar{\mathbf{u}}|| + h^2 |\log h|) \\ &\leq c_1 \left(\sqrt{h^2 |\log h|} + h^2 |\log h| \right) \\ &\leq C, \end{aligned}$$

y aplicando la desigualdad triangular, se sigue que

$$||\nabla F_h(\bar{\mathbf{u}}^h)|| \leq C + ||\nabla F(\bar{\mathbf{u}})|| \leq c,$$

por lo tanto $\left[\nabla F_h(\bar{\mathbf{u}}^h)\right]_{i=1}^r$ es acotada. Entonces de (4.66) se obtiene que $[\bar{\lambda}^h]_{i=1}^r$ son acotados para h suficientemente pequeño y para $i = r+1, \ldots, l+4M$ se tiene que $\bar{\lambda}_i^h = 0$, pues en estos índices G_h es inactivo en $\bar{\mathbf{u}}^h$. □

OBSERVACIÓN. La igualdad (4.66) de la Proposición 4.28, nos permite concluir que $\bar{\lambda}^h$, la sucesión de los mutiplicadores de Lagrange asociados a $\bar{\mathbf{u}}^h$, converge a $\bar{\lambda}$, el multiplicador de Lagrange asociado a $\bar{\mathbf{u}}$, cuando $h \to 0$.

Consideramos $\bar{\mathbf{u}}^h$ la solución óptima del problema aproximado (4.39), obtenido de la Proposición 4.26, con $\bar{\lambda}^h$ el vector de los multiplicadores de Lagrange asociado a $\bar{\mathbf{u}}^h$. Por lo visto anteriormente, se tiene que

$$0 \in \mathcal{F}_h(\bar{\mathbf{u}}^h, \bar{\lambda}^h) + \partial\psi_{\mathbb{R}^{2M}\times K_+}(\bar{\mathbf{u}}^h, \bar{\lambda}^h), \tag{4.67}$$

vamos a probar que $(\bar{\mathbf{u}}^h, \bar{\lambda}^h)$ resuelve la ecuación generalizada perturbada

$$\delta_h \in \mathcal{F}(\mathbf{u}^h, \lambda^h) + \partial\psi_{\mathbb{R}^{2M}\times K_+}(\mathbf{u}^h, \lambda^h), \tag{4.68}$$

con δ_h una perturbación de orden $h^2|\log h|$.

PROPOSICIÓN 4.29. *La ecuación generalizada (4.68) tiene solución en* $(\bar{\mathbf{u}}^h, \bar{\lambda}^h)$.

Demostración. De la ecuación generalizada (4.67), se tiene que

$$0 \in \begin{bmatrix} \nabla_{\mathbf{u}}\mathcal{L}_h(\bar{\mathbf{u}}^h, \bar{\lambda}^h) \\ -G_h(\bar{\mathbf{u}}^h) \end{bmatrix} + \begin{bmatrix} \{0\} \\ \partial\psi_{K_+}(\bar{\lambda}^h) \end{bmatrix},$$

de la primera componente de esta ecuación, obtenemos que

$$\begin{aligned}
0 &= \nabla_{\mathbf{u}}\mathcal{L}_h(\bar{\mathbf{u}}^h, \bar{\lambda}^h) \\
&= \nabla F_h(\bar{\mathbf{u}}^h) + \sum_{j=i}^{l+4M} \bar{\lambda}_j^h \nabla G_j^h(\bar{\mathbf{u}}^h) \\
&= \nabla F(\bar{\mathbf{u}}^h) + \nabla F_h(\bar{\mathbf{u}}^h) - \nabla F(\bar{\mathbf{u}}^h) + \sum_{j=i}^{l+4M} \bar{\lambda}_j^h \left(\nabla G_j(\bar{\mathbf{u}}^h) + \nabla G_j^h(\bar{\mathbf{u}}^h) - \nabla G_j(\bar{\mathbf{u}}^h)\right) \\
&= \nabla F(\bar{\mathbf{u}}^h) + \nabla F_h(\bar{\mathbf{u}}^h) - \nabla F(\bar{\mathbf{u}}^h) + \sum_{j=i}^{l+4M} \bar{\lambda}_j^h \nabla G_j(\bar{\mathbf{u}}^h) \\
&\qquad + \sum_{j=i}^{l+4M} \bar{\lambda}_j^h \left(\nabla G_j^h(\bar{\mathbf{u}}^h) - \nabla G_j(\bar{\mathbf{u}}^h)\right),
\end{aligned}$$

si definimos $r_{h,1} = \nabla F_h(\bar{\mathbf{u}}^h) - \nabla F(\bar{\mathbf{u}}^h)$ y $r_{h,2} = \sum_{j=i}^{l+4M} \bar{\lambda}_j^h(\nabla G_j^h(\bar{\mathbf{u}}^h) - \nabla G_j(\bar{\mathbf{u}}^h))$, se sigue que

$$\begin{aligned}
0 &= \nabla F(\bar{\mathbf{u}}^h) + \sum_{j=i}^{l+4M} \bar{\lambda}_j^h \nabla G_j(\bar{\mathbf{u}}^h) + r_{h,1} + r_{h,2} \\
&= \nabla_{\mathbf{u}}\mathcal{L}(\bar{\mathbf{u}}^h, \bar{\lambda}^h) - \delta_{h,1},
\end{aligned}$$

donde $\delta_{h,1} = -(r_{h,1} + r_{h,2})$. Probaremos que $||\delta_{h,1}|| \leq c\ h^2|\log h|$. En efecto, de la

desigualdad triangular se tiene que

$$\begin{aligned}||\delta_{h,1}|| &\leq ||r_{h,1}|| + ||r_{h,2}|| \\ &= ||\nabla F_h(\bar{\mathbf{u}}^h) - \nabla F(\bar{\mathbf{u}}^h)|| + ||\sum_{j=i}^{l+4M} \bar{\lambda}_j^h (\nabla G_j^h(\bar{\mathbf{u}}^h) - \nabla G_j(\bar{\mathbf{u}}^h))|| \\ &\leq ||\nabla F_h(\bar{\mathbf{u}}^h) - \nabla F(\bar{\mathbf{u}}^h)|| + \sum_{j=i}^{l+4M} |\bar{\lambda}_j^h|\, ||\nabla G_j^h(\bar{\mathbf{u}}^h) - \nabla G_j(\bar{\mathbf{u}}^h)||,\end{aligned}$$

gracias a la Proposición 4.28 se tiene que los multiplicadores $\bar{\lambda}^h$ son uniformemente acotados, obteniendo así

$$||\delta_{h,1}|| \leq ||\nabla F_h(\bar{\mathbf{u}}^h) - \nabla F(\bar{\mathbf{u}}^h)|| + c \sum_{j=i}^{l+4M} ||\nabla G_j^h(\bar{\mathbf{u}}^h) - \nabla G_j(\bar{\mathbf{u}}^h)||,$$

con $c > 0$. Luego, de la estimación (4.52), se sigue que

$$||\delta_{h,1}|| \leq c\, h^2 |\log h|.$$

De forma análoga, analizando la segunda componente se obtiene que

$$\begin{aligned}0 &\in -G_h(\bar{\mathbf{u}}^h) + \partial\psi_{K_+}(\bar{\lambda}^h) \\ 0 &\in -G(\bar{\mathbf{u}}^h) + \Big(G(\bar{\mathbf{u}}^h) - G_h(\bar{\mathbf{u}}^h)\Big) + \partial\psi_{K_+}(\bar{\lambda}^h) \\ 0 &\in -G(\bar{\mathbf{u}}^h) - \delta_{h,2} + \partial\psi_{K_+}(\bar{\lambda}^h),\end{aligned}$$

donde $\delta_{h,2} = -(G(\bar{\mathbf{u}}^h) - G_h(\bar{\mathbf{u}}^h))$ y de la estimación (4.52) se tiene que $||\delta_{h,2}|| \leq c\, h^2 |\log h|$. Entonces de estos dos resultados, se concluye que $(\bar{\mathbf{u}}^h, \bar{\lambda}^h)$ resuelve la ecuación generalizada (4.68), con una perturbación $\delta_h^T = (\delta_{h,1}, \delta_{h,2})^T$ que es de orden $h^2 |\log h|$. □

De manera análoga que en [19] obtenemos el siguiente resultado.

PROPOSICIÓN 4.30. *Existe una constante $C > 0$ independiente de h, tal que*

$$||\bar{\mathbf{u}} - \bar{\mathbf{u}}^h|| + ||\bar{\lambda} - \bar{\lambda}^h|| \leq Ch^2 |\log h|,$$

siempre que h sea suficientemente pequeño.

Demostración. Definimos la función $\mathbf{F} : \mathbb{R}^{2M} \times \mathbb{R}^{l+4M} \times \mathbb{R}^{2M+l+4M} \to \mathbb{R}^{2M} \times \mathbb{R}^{l+4M}$, tal que

$$\mathbf{F}(\mathbf{u}, \lambda, \delta) = \mathcal{F}(\mathbf{u}, \lambda) - \delta,$$

se tiene que $\mathbf{F}(\mathbf{u}, \lambda, 0) = \mathcal{F}(\mathbf{u}, \lambda)$. Así, la ecuación generalizada (4.64) puede ser expresado como

$$0 \in \mathbf{F}(\mathbf{u}, \lambda, 0) + \partial\psi_{\mathbb{R}^{2M} \times K_+}(\mathbf{u}, \lambda), \tag{4.69}$$

y puesto que la condición fuerte de segundo orden se satisface en $(\bar{\mathbf{u}}, \bar{\lambda})$ junto con la condición (LICQ), y gracias a la Proposición 2.21, se concluye que la ecuación generalizada (4.64) satisface la condición de regularidad fuerte en $(\bar{\mathbf{u}}, \bar{\lambda})$. Por consiguiente (4.69) es fuertemente regular en $(\bar{\mathbf{u}}, \bar{\lambda})$, con lo cual estamos en las hipótesis del teorema de la función implícita de Robinson (Proposición 2.14) por lo tanto existen vecindades $\mathcal{O}$ de 0 y W de $(\bar{\mathbf{u}}, \bar{\lambda})$, tal que para todo $\delta \in \mathcal{O}$ existe un único $(\widetilde{\mathbf{u}}, \widetilde{\mu}) \in W$ que resuelve

$$0 \in \mathbf{F}(\widetilde{\mathbf{u}}, \widetilde{\lambda}, \delta) + \partial\psi_{\mathbb{R}^{2M} \times K_+}(\widetilde{\mathbf{u}}, \widetilde{\lambda}),$$

que es lo mismo

$$\delta \in \mathcal{F}(\widetilde{\mathbf{u}}, \widetilde{\lambda}) + \partial\psi_{\mathbb{R}^{2M} \times K_+}(\widetilde{\mathbf{u}}, \widetilde{\lambda}). \tag{4.70}$$

Adicionalmente del teorema de la función implícita, se cumple que

$$\begin{aligned} ||(\bar{\mathbf{u}}, \bar{\lambda}) - (\widetilde{\mathbf{u}}, \widetilde{\lambda})|| &\leq C||\mathbf{F}(\widetilde{\mathbf{u}}, \widetilde{\lambda}, 0) - \mathbf{F}(\widetilde{\mathbf{u}}, \widetilde{\lambda}, \delta)|| \\ &= C||\mathcal{F}(\widetilde{\mathbf{u}}, \widetilde{\lambda}) - \mathcal{F}(\widetilde{\mathbf{u}}, \widetilde{\lambda}) + \delta|| \\ &= C||\delta||, \end{aligned}$$

por lo tanto

$$||\bar{\mathbf{u}} - \widetilde{\mathbf{u}}|| + ||\bar{\lambda} - \widetilde{\lambda}|| \leq C||\delta||. \tag{4.71}$$

De la Proposición 4.29, sabemos que $\delta_h \to 0$ cuando $h \to 0$, por lo tanto $\delta_h \in \mathcal{O}$ para $h > 0$ suficientemente pequeño y de (4.70) obtenemos

$$\delta_h \in \mathcal{F}(\widetilde{\mathbf{u}}, \widetilde{\lambda}) + \partial\psi_{\mathbb{R}^{2M} \times K_+}(\widetilde{\mathbf{u}}, \widetilde{\lambda}). \tag{4.72}$$

Luego, de las observaciones realizadas en la Proposición 4.26 y 4.28, respectivamente, se obtiene que para $h > 0$ suficientemente pequeño $(\bar{\mathbf{u}}^h, \bar{\lambda}^h) \in W$ y puesto que $(\bar{\mathbf{u}}^h, \bar{\lambda}^h)$ satisface (4.68), se sigue por la unicidad de (4.72) que $(\bar{\mathbf{u}}^h, \bar{\lambda}^h) = (\widetilde{\mathbf{u}}, \widetilde{\lambda})$. Finalmente de (4.71), se concluye que

$$||\bar{\mathbf{u}} - \bar{\mathbf{u}}^h|| + ||\bar{\lambda} - \bar{\lambda}^h|| \leq C||\delta_h|| \leq Ch^2|\log h|. \qquad \square$$

Finalizamos esta sección con el resultado más importante de este proyecto, el cual es la estimación del error para la aproximación del problema de control óptimo.

PROPOSICIÓN 4.31. *Si $\bar{u}$ es una solución del problema (3.3) entonces existe una su-*

cesión $\bar{u}^h$ de soluciones óptimas del problema (4.38) y una constante $C > 0$ que no depende de h, tal que la estimación

$$||\bar{u} - \bar{u}^h|| \leq Ch^2 |\log h|,$$

se cumple para $h > 0$ suficientemente pequeño.

Demostración. De la Proposición 4.14, sabemos que $\bar{u} = [\bar{\mathbf{u}}_i]_{i=1}^M - [\bar{\mathbf{u}}_{i+M}]_{i=1}^M$. Luego, de la Proposición 4.30 deducimos la existencia de una sucesión de soluciones óptimas del problema (4.39) y de la Proposición 4.18 se tiene que la solución del problema (4.38) es $\bar{u}^h = [\bar{\mathbf{u}}_i^h]_{i=1}^M - [\bar{\mathbf{u}}_{i+M}^h]_{i=1}^M$. Por lo tanto, obtenemos la siguiente acotación

$$\begin{aligned} ||\bar{u} - \bar{u}^h|| &= || \left(\left[\bar{\mathbf{u}}_i\right]_{i=1}^M - \left[\bar{\mathbf{u}}_{i+M}\right]_{i=1}^M \right) - \left(\left[\bar{\mathbf{u}}_i^h\right]_{i=1}^M - \left[\bar{\mathbf{u}}_{i+M}^h\right]_{i=1}^M \right) || \\ &\leq || \left[\bar{\mathbf{u}}_i\right]_{i=1}^M - \left[\bar{\mathbf{u}}_i^h\right]_{i=1}^M || + || \left[\bar{\mathbf{u}}_{i+M}\right]_{i=1}^M - \left[\bar{\mathbf{u}}_{i+M}^h\right]_{i=1}^M || \\ &\leq C||\bar{\mathbf{u}} - \bar{\mathbf{u}}^h|| \\ &\leq Ch^2 |\log h|, \end{aligned}$$

siempre que $h > 0$ sea suficientemente pequeño. □

Capítulo 5

Notas sobre el problema sin regularización de Tikhonov ($\alpha = 0$)

El estudio realizado a nuestro problema (P) en los capítulos anteriores no contemplan el caso en que $\alpha = 0$. El objetivo de este capítulo es abordar un estudio del problema sin regularización de Tikhonov ($\alpha = 0$), para lo cual es necesario realizar una suposición sobre las funciones bases y_i y con ello hacer un análisis análogo a los Capítulos 3 y 4.

El problema de control óptimo a analizar está planteado de la siguiente manera :

$$(P') \qquad \begin{cases} \min\limits_{(y,u)} J(y,u) = \dfrac{1}{2}||y - y_d||^2 + \beta||u||_1 \\ \text{sujeto a:} \\ Ay(x) = \displaystyle\sum_{i=1}^{M} u_i e_i(x), & \text{en } \Omega, \\ y(x) = 0, & \text{sobre } \Gamma, \\ y(x_j) \leq b_j, & \forall j = 1, \ldots, l, \\ u \in \mathcal{U}_{ad}, \end{cases}$$

con las mismas suposiciones hechas en la Sección 3.1 para el problema (P).

Realizaremos el análisis de existencia y unicidad de una solución óptima del problema (P'), para lo cual es necesario asumir la siguiente hipótesis.

HIPÓTESIS 5.1. *La matriz A, definida por*

$$A = \begin{pmatrix} (y_1, y_1) & (y_2, y_1) & \cdots & (y_M, y_1) \\ (y_1, y_2) & (y_2, y_2) & \cdots & (y_M, y_2) \\ \vdots & \vdots & \ddots & \vdots \\ (y_1, y_M) & (y_2, y_M) & \cdots & (y_M, y_M) \end{pmatrix}$$

es definida positiva.

OBSERVACIÓN. Es fácil notar que la Hipótesis 5.1 puede ser satisfecha si tomamos $\{y_1, \ldots, y_M\}$ una base ortonormal de $L^2(\Omega)$. Así, $A = I_M$, y se sabe que la matriz identidad es definida positiva.

Haciendo un análisis análogo a la Sección 3.3, concluimos que nuestro problema reducido está dado por:

$$\begin{cases} \min\limits_{u \in \mathcal{U}_{ad}} f(u) = \dfrac{1}{2} \left|\left| \sum\limits_{i=1}^{M} u_i y_i - y_d \right|\right|^2 + \beta ||u||_1 \\ \text{sujeto a:} \\ \sum\limits_{i=1}^{M} u_i y_i(x_j) \leq b_j, \quad \forall j = 1, \ldots, l, \end{cases} \tag{5.1}$$

donde y_i es la solución débil de (3.1) con $e = e_i$, para $i = 1, \ldots, M$.

OBSERVACIÓN. f es una función estrictamente convexa.

Demostración. Es suficiente probar que la función

$$f_1(u) = \frac{1}{2} \left|\left| \sum_{i=1}^{M} u_i y_i - y_d \right|\right|^2$$

es estrictamente convexa. Para ello, calculamos la Hessiana de f_1, la cual está dada por

$$\nabla^2 f_1(u) = \begin{pmatrix} (y_1, y_1) & (y_2, y_1) & \cdots & (y_M, y_1) \\ (y_1, y_2) & (y_2, y_2) & \cdots & (y_M, y_2) \\ \vdots & \vdots & \ddots & \vdots \\ (y_1, y_M) & (y_2, y_M) & \cdots & (y_M, y_M) \end{pmatrix},$$

gracias a la Hipótesis 5.1 se tiene que $\nabla^2 f_1(u)$ es definida positiva, por tanto f_1 es estrictamente convexa y por consiguiente f es estrictamente convexa. □

La demostración del siguiente resultado es similar a la Proposición 3.6.

PROPOSICIÓN 5.2. *Existe una única solución óptima $(\bar{y}, \bar{u})$ para el problema (P').*

Para derivar las condiciones de optimalidad de (P') procedemos similarmente a la Sección 3.4, obteniendo así el siguiente resultado.

PROPOSICIÓN 5.3. *Sea $(\bar{u}, \bar{y})$ la solución óptima para el problema (P'), bajo la Hipótesis 3.7, existen $v \in \mathbb{R}^l$ y $w \in \mathbb{R}^{2M}_+$ tales que*

$$\begin{aligned}
\sum_{j=1}^{l} v_j(\bar{y}(x_j) - b_j) &= 0,\\
w_i(u_{a,i} - \bar{u}_i) &= 0,\\
w_{i+M}(\bar{u}_i - u_{b,i}) &= 0,\\
\int_\Omega (\bar{y} - y_d)y_i\,dx + \sum_{j=1}^{l} v_j y_i(x_j) - (w_i - w_{i+M}) &= -\beta, \quad \textit{si } \bar{u}_i > 0,\\
\int_\Omega (\bar{y} - y_d)y_i\,dx + \sum_{j=1}^{l} v_j y_i(x_j) - (w_i - w_{i+M}) &= \beta, \quad \textit{si } \bar{u}_i < 0,\\
\left|\int_\Omega (\bar{y} - y_d)y_i\,dx + \sum_{j=1}^{l} v_j y_i(x_j) - (w_i - w_{i+M})\right| &\leq \beta, \quad \textit{si } \bar{u}_i = 0,
\end{aligned}$$

para todo $i = 1, \ldots, M$.

Ahora, para obtener el orden de la estimación del error para la aproximación del problema de control óptimo (P') desarrollamos el mismo análisis realizado en el Capítulo 4. Para esto, realizamos un análisis idéntico a la Sección 4.2, es decir, reformulamos el problema (5.1) en el siguiente problema de control óptimo:

$$\begin{cases}
\min\limits_{\mathbf{u}\in\mathcal{V}_{ad}} F(\mathbf{u}) := \dfrac{1}{2}\left|\left|\sum\limits_{i=1}^{M}(\mathbf{u}_i - \mathbf{u}_{i+M})y_i - y_d\right|\right|^2 + \beta\sum\limits_{i=1}^{2M}\mathbf{u}_i\\
\text{sujeto a:}\\
\sum\limits_{i=1}^{M}(\mathbf{u}_i - \mathbf{u}_{i+M})y_i(x_j) - b_j \leq 0, \quad \forall j = 1, \ldots, l,
\end{cases} \tag{5.2}$$

donde $\mathcal{V}_{ad} \subset \mathbb{R}^{2M}$ es el conjunto de los controles admisibles definido en (4.24). Puesto que F es continua existe al menos una solución óptima de (5.2).

Además, se tiene que F es una función convexa. En efecto, calculamos $\nabla^2 F(\bar{\mathbf{u}})$,

la cual está dada por

$$\nabla^2 F(\bar{\mathbf{u}}) = \left(\begin{array}{cccc|cccc} (y_1,y_1) & (y_2,y_1) & \cdots & (y_M,y_1) & -(y_1,y_1) & -(y_2,y_1) & \cdots & -(y_M,y_1) \\ (y_1,y_2) & (y_2,y_2) & \cdots & (y_M,y_2) & -(y_1,y_2) & -(y_2,y_2) & \cdots & -(y_M,y_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ (y_1,y_M) & (y_2,y_M) & \cdots & (y_M,y_M) & -(y_1,y_M) & -(y_2,y_M) & \cdots & -(y_M,y_M) \\ \hline -(y_1,y_1) & -(y_2,y_1) & \cdots & -(y_M,y_1) & (y_1,y_1) & (y_2,y_1) & \cdots & (y_M,y_1) \\ -(y_1,y_2) & -(y_2,y_2) & \cdots & -(y_M,y_2) & (y_1,y_2) & (y_2,y_2) & \cdots & (y_M,y_2) \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -(y_1,y_M) & -(y_2,y_M) & \cdots & -(y_M,y_M) & (y_1,y_M) & (y_2,y_M) & \cdots & (y_M,y_M) \end{array}\right),$$

visto de otro modo, se tiene que

$$\nabla^2 F(\bar{\mathbf{u}}) = \left(\begin{array}{c|c} A & -A \\ \hline -A & A \end{array}\right), \tag{5.3}$$

por la Hipótesis 5.1 la matriz A es simétrica definida positiva, entonces se tiene que

$$\begin{aligned} v^T \nabla^2 F(\bar{\mathbf{u}}) v &= \sum_{i,j=1}^{M} (v_j - v_{j+M})(y_i, y_j)(v_i - v_{i+M}) \\ &= \left([v_j - v_{j+M}]_{j=1}^{M}\right)^T A [v_i - v_{i+M}]_{i=1}^{M} \\ &\geq 0, \end{aligned}$$

por lo tanto F es convexa. En consecuencia toda solución óptima de (5.2) es una solución global.

Con esto y de forma análoga a la Proposición 4.12 y 4.14 se obtiene los siguientes resultados.

PROPOSICIÓN 5.4. *Si $\bar{\mathbf{u}}$ es una solución óptima de (5.2) entonces $\bar{\mathbf{u}}_i \bar{\mathbf{u}}_{i+M} = 0$ para todo $i = 1, \ldots, M$.*

PROPOSICIÓN 5.5. *Si $\bar{u}$ y $\bar{\mathbf{u}}$ son soluciones de los problemas (5.1) y (5.2) respectivamente, entonces*

$$\left[\bar{\mathbf{u}}_i\right]_{i=1}^{M} = \bar{u}^+ \quad y \quad \left[\bar{\mathbf{u}}_{i+M}\right]_{i=1}^{M} = \bar{u}^-,$$

es decir $\bar{u} = \left[\bar{\mathbf{u}}_i\right]_{i=1}^{M} - \left[\bar{\mathbf{u}}_{i+M}\right]_{i=1}^{M}$. En otras palabras, los problemas (5.1) y (5.2) son equivalentes.

OBSERVACIÓN. De la Proposición 5.5 y de la unicidad para la solución del problema (5.1) se concluye que la solución óptima del problema (5.2) es única.

Para la obtención de las condiciones de optimalidad para el problema (5.2) consideramos el Lagrangiano definido en (4.29) y procedemos de forma similar a la

obtención de (4.34), (4.35) y (4.36). Por lo tanto, existen $\bar{\mu} \in \mathbb{R}^l_+$ y $\bar{v} \in \mathbb{R}^{4M}_+$ tales que

$$\begin{aligned} (\bar{y} - y_d, y_i) + \beta + \sum_{j=1}^{l} \bar{\mu}_j y_i(x_j) - \bar{v}_i + \bar{v}_{i+2M} = 0, \\ -(\bar{y} - y_d, y_i) + \beta - \sum_{j=1}^{l} \bar{\mu}_j y_i(x_j) - \bar{v}_{i+M} + \bar{v}_{i+3M} = 0, \end{aligned} \tag{5.4}$$

para $i = 1, \ldots, M$, donde $\bar{y} = \sum_{i=1}^{M} (\bar{\mathbf{u}}_i - \bar{\mathbf{u}}_{i+M}) y_i$, junto con

$$G_j(\bar{\mathbf{u}}) \leq 0, \quad \forall j = 1, \ldots, l + 4M, \tag{5.5}$$

con G_j definido en (4.30). Además, satisface las condiciones de complementariedad

$$\begin{aligned} \bar{\mu}_j \left(\bar{y}(x_j) - b_j \right) &= 0, \quad \forall j = 1, \ldots, l, \\ \bar{v}_i (u^+_{a,i} - \bar{\mathbf{u}}_i) &= 0, \quad \forall i = 1, \ldots, M, \\ \bar{v}_{i+M} (u^-_{b,i} - \bar{\mathbf{u}}_{i+M}) &= 0, \quad \forall i = 1, \ldots, M, \\ \bar{v}_{i+2M} (\bar{\mathbf{u}}_i - u^+_{b,i}) &= 0, \quad \forall i = 1, \ldots, M, \\ \bar{v}_{i+3M} (\bar{\mathbf{u}}_{i+M} - u^-_{a,i}) &= 0, \quad \forall i = 1, \ldots, M. \end{aligned} \tag{5.6}$$

Para encontrar las condiciones necesarias de segundo orden asumimos que $\bar{u}$ satisface la condición de complementariedad estricta para el problema (5.1), por lo tanto el cono crítico para el problema (5.2) viene dado por

$$C_{\bar{\mathbf{u}}} = \{ \mathbf{v} \in \mathbb{R}^{2M} : \nabla G_j(\bar{\mathbf{u}})^T \mathbf{v} = 0, \quad \forall j \in \mathcal{A}(\bar{\mathbf{u}}) \},$$

donde $\mathcal{A}(\bar{\mathbf{u}}) = \{ j \in \{1, \ldots, l + 4M\} : G_j(\bar{\mathbf{u}}) = 0 \}$. El siguiente resultado es una caracterización del cono crítico.

PROPOSICIÓN 5.6. *Si $\mathbf{v} \in C_{\bar{\mathbf{u}}}$ y cumple que $\mathbf{v}_j = \mathbf{v}_{j+M}$ para todo $j = 1, \ldots, M$, entonces $\mathbf{v} = 0$.*

Demostración. Sea $i \in \{1, \ldots, M\}$, de la Proposición 5.4 se sabe que $\bar{\mathbf{u}}_i \bar{\mathbf{u}}_{i+M} = 0$, con lo cual podemos analizamos los siguiente casos:

- Si $\bar{\mathbf{u}}_i \neq 0$ entonces $\bar{\mathbf{u}}_{i+M} = 0$. Como $\bar{\mathbf{u}}$ satisface las restricciones de caja se tiene que necesariamente $u^-_{b,i} - \bar{\mathbf{u}}_{i+M} = 0$, considerando la definición de G se sigue que $G_{l+M+i}(\bar{\mathbf{u}}) = 0$ por tanto $l + M + i \in \mathcal{A}(\bar{\mathbf{u}})$ y puesto que $\mathbf{v} \in C_{\bar{\mathbf{u}}}$ se tiene que

$$0 = \nabla G_{l+M+i}(\bar{\mathbf{u}})^T \mathbf{v} = -\mathbf{v}_{i+M},$$

y por consiguiente $\mathbf{v}_{i+M} = 0$.

- Si $\bar{\mathbf{u}}_{i+M} \neq 0$ entonces $\bar{\mathbf{u}}_i = 0$, por las restricciones de caja se tiene que necesariamente $u^+_{a,i} - \bar{\mathbf{u}}_{i+M} = 0$ es decir $G_{l+i}(\bar{\mathbf{u}}) = 0$ por tanto $l + i \in \mathcal{A}(\bar{\mathbf{u}})$ y puesto que $\mathbf{v} \in C_{\bar{\mathbf{u}}}$ se tiene que

$$0 = \nabla G_{l+i}(\bar{\mathbf{u}})^T \mathbf{v} = -\mathbf{v}_i,$$

entonces $\mathbf{v}_i = 0$.

Con esto se tiene que $\mathbf{v}_i = 0$ o $\mathbf{v}_{i+M} = 0$ y puesto que $\mathbf{v}_i = \mathbf{v}_{i+M}$, se sigue que $\mathbf{v}_i = \mathbf{v}_{i+M} = 0$, para todo $i = 1, \ldots, M$, lo cual nos permite concluir que $\mathbf{v} = 0$. □

Para la siguiente proposición es necesario calcular la matriz Hessiana del Lagrangiano con respecto a $\mathbf{u}$, pero claramente se tiene que

$$\nabla^2 \mathcal{L}(\bar{\mathbf{u}}, \mu, \nu) = \nabla^2 F(\bar{\mathbf{u}}).$$

PROPOSICIÓN 5.7. *Para todo* $\mathbf{v} \in C_{\bar{\mathbf{u}}} \backslash \{0\}$ *se cumple que*

$$\mathbf{v}^T \nabla^2 \mathcal{L}(\bar{\mathbf{u}}, \bar{\mu}, \bar{\nu}) \mathbf{v} > 0.$$

Demostración. Para $\mathbf{v} \in C_{\bar{\mathbf{u}}} \backslash \{0\}$ obtenemos que

$$\begin{aligned} \mathbf{v}^T \nabla^2 \mathcal{L}(\bar{\mathbf{u}}, \bar{\mu}, \bar{\nu}) \mathbf{v} &= \sum_{i,j=1}^{M} (\mathbf{v}_j - \mathbf{v}_{j+M})(y_i, y_j)(\mathbf{v}_i - \mathbf{v}_{i+M}) \\ &= \left([\mathbf{v}_j - \mathbf{v}_{j+M}]_{j=1}^M \right)^T A \, [\mathbf{v}_i - \mathbf{v}_{i+M}]_{i=1}^M. \end{aligned}$$

Como $\mathbf{v} \neq 0$, de la Proposición 5.6 se tiene que existe un $j_0 \in \{1, \ldots, M\}$ tal que $\mathbf{v}_{j_0} \neq \mathbf{v}_{j_0+M}$ y puesto que A es definida positiva, se concluye que

$$\mathbf{v}^T \nabla^2 \mathcal{L}(\bar{\mathbf{u}}, \bar{\mu}, \bar{\nu}) \mathbf{v} > 0.$$ □

La demostración del siguiente resultado es consecuencia directa de la Proposición 2.20.

PROPOSICIÓN 5.8. *Existen constantes positivas* ω *y* ε *tales que*

$$F(\mathbf{u}) - F(\bar{\mathbf{u}}) \geq \omega ||\mathbf{u} - \bar{\mathbf{u}}||^2,$$

para todo $\mathbf{u} \in \mathcal{V}_{ad}$ *con* $||\mathbf{u} - \bar{\mathbf{u}}|| \leq \varepsilon$.

Si consideramos la discretización de la ecuación de estado estudiada en la sección 4.1, podemos definir el problema de control discreto reemplazando el estado discreto y_u^h definido en (4.7).

$$\begin{cases} \min\limits_{u\in\mathcal{U}_{ad}} f_h(u) := \dfrac{1}{2}\left|\left|\sum\limits_{i=1}^{M} u_i y_i^h - y_d\right|\right|^2 + \beta||u||_1 \\ \text{sujeto a:} \\ \sum\limits_{i=1}^{M} u_i y_i^h(x_j) - b_j \leq 0, \qquad \forall j = 1,\dots,l. \end{cases} \tag{5.7}$$

Reformulamos este problema en el siguiente problema de control óptimo discreto:

$$\begin{cases} \min\limits_{\mathbf{u}\in\mathcal{V}_{ad}} F_h(\mathbf{u}) := \dfrac{1}{2}\left|\left|\sum\limits_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i^h - y_d\right|\right|^2 + \beta\sum\limits_{i=1}^{2M} \mathbf{u}_i \\ \text{sujeto a:} \\ \sum\limits_{i=1}^{M} (\mathbf{u}_i - \mathbf{u}_{i+M}) y_i^h(x_j) - b_j \leq 0, \qquad \forall j = 1,\dots,l. \end{cases} \tag{5.8}$$

Realizando un estudio idéntico a la Sección 4.3 obtenemos la siguiente proposición.

PROPOSICIÓN 5.9. *Los problemas (5.7) y (5.8) son equivalentes.*

Tomando en cuenta el Lagrangiano definido en (4.40), obtenemos que las condiciones de optimalidad para (5.8) pueden ser expresadas por:

$$\begin{aligned} (\bar{y}^h - y_d, y_i^h) + \beta + \sum_{j=1}^{l} \bar{\mu}_j^h y_i^h(x_j) - \bar{v}_i^h + \bar{v}_{i+2M}^h = 0, \\ -(\bar{y}^h - y_d, y_i^h) + \beta - \sum_{j=1}^{l} \bar{\mu}_j^h y_i^h(x_j) - \bar{v}_{i+M}^h + \bar{v}_{i+3M}^h = 0, \end{aligned} \tag{5.9}$$

para $i = 1,\dots,M$, donde $\bar{y}^h = \sum_{i=1}^{M} (\bar{\mathbf{u}}_i^h - \bar{\mathbf{u}}_{i+M}^h) y_i^h$, junto con

$$G_j^h(\bar{\mathbf{u}}^h) \leq 0, \quad \forall j = 1,\dots,l+4M,$$

con G_j^h definido en (4.41). Además, se satisfacen las condiciones de complementa-

riedad

$$\begin{aligned}
\bar{\mu}_j^h \left(\bar{y}^h(x_j) - b_j \right) &= 0, \quad \forall j = 1, \ldots, l, \\
\bar{v}_i^h (u_{a,i}^+ - \bar{\mathbf{u}}_i^h) &= 0, \quad \forall i = 1, \ldots, M, \\
\bar{v}_{i+M}^h (u_{b,i}^- - \bar{\mathbf{u}}_{i+M}^h) &= 0, \quad \forall i = 1, \ldots, M, \\
\bar{v}_{i+2M}^h (\bar{\mathbf{u}}_i^h - u_{b,i}^+) &= 0, \quad \forall i = 1, \ldots, M, \\
\bar{v}_{i+3M}^h (\bar{\mathbf{u}}_{i+M}^h - u_{a,i}^-) &= 0, \quad \forall i = 1, \ldots, M.
\end{aligned} \tag{5.10}$$

Finalmente para obtener el orden de estimación del error para los controles óptimos del problema de control discreto (5.7), trabajamos con los problemas equivalentes (5.2) y (5.8). Luego realizando el mismo análisis que la Sección 4.4 concluimos el siguiente resultado:

PROPOSICIÓN 5.10. *Existe una constante $C > 0$ independiente de h, tal que*

$$||\bar{\mathbf{u}} - \bar{\mathbf{u}}^h|| + ||\bar{\lambda} - \bar{\lambda}^h|| \leq Ch^2 |\log h|,$$

siempre que h sea suficientemente pequeño.

Lo que deriva directamente en la siguiente proposición.

PROPOSICIÓN 5.11. *Si $\bar{u}$ es una solución del problema (5.1) entonces existe una sucesión $\bar{u}^h$ de soluciones óptimas del problema (5.7) y una constante $C > 0$ que no depende de h, tal que la estimación*

$$||\bar{u} - \bar{u}^h|| \leq Ch^2 |\log h|,$$

se cumple para $h > 0$ suficientemente pequeño.

Capítulo 6

Experimentos numéricos

En esta sección presentamos dos ejemplos de problemas de control óptimo de este trabajo, para los cuales verificamos numéricamente que el orden de la estimación del error es cercano a $\alpha(h) = h^2$. Esto no contradice nuestra teoría puesto el término $|\log h|$ resulta numéricamente difícil ser detectado cuando h es pequeño.

Los resultados numéricos de cada problema fueron obtenidos mediante la resolución aproximada de la ecuación de estado a partir del paquete IFEM en Matlab, para diferentes tamaños de malla h. Para el tratamiento numérico de las restricciones de estado hemos usado la regularización de Moreau-Yosida cf. [11][página 98]. Dicha regularización es una aproximación que consiste en penalizar las restricciones de estado por medio de la función

$$|| \, [\text{máx}(0, \gamma(y(x_j) - b_j))]_{i=1}^{l} ||^2,$$

entonces nuestro problema (P) se aproxima mediante el siguiente problema

$$(P_{reg}) \quad \begin{cases} \min\limits_{(y,u)} J(y,u) + \dfrac{1}{2\gamma} || \, [\text{máx}(0, \gamma(y(x_j) - b_j))]_{i=1}^{l} ||^2 \\ \text{sujeto a:} \\ Ay(x) = \sum\limits_{i=1}^{M} u_i e_i(x), \quad \text{en } \Omega, \\ y(x) = 0, \qquad\qquad\qquad \text{sobre } \Gamma, \\ u \in \mathcal{U}_{ad}, \end{cases}$$

para $\gamma \in \mathbb{R}$ lo suficientemente grande, para nuestros ejemplos vamos a utilizar el parámetro de regularización $\gamma = 1000000$. Así, obtenemos un problema discretizado que tiene la forma de un problema de optimización no diferenciable que es

resuelto usando el algoritmo OESOM cf. [12]. Ilustramos la convergencia para distintos tamaño de malla, y una tabla que muestra el error en la variable del control para diferentes valores de h y el error experimental de convergencia calculado por

$$EOC' = \frac{\log(||\bar{u}^{h_1} - \bar{u}_h^*||) - \log(||\bar{u}_h^* - \bar{u}^{h_2}||)}{\log(h_1) - \log(h_2)}, \tag{6.1}$$

para dos consecutivos tamaños de malla h_1 y h_2, y u_h^* la solución aproximada del problema.

EJEMPLO 1. Consideramos el siguiente problema con 5 puntos disjuntos para las restricciones de estado y 10 controles.

$$(P) \qquad \begin{cases} \min\limits_{u \in \mathcal{U}_{ad} \subset \mathbb{R}^{10}} J(y,u) = \dfrac{1}{2}||y - y_d||^2 + \dfrac{1}{2}||u||_2^2 + 10||u||_1 \\ \text{sujeto a:} \\ \begin{aligned} -\Delta y(x) + y(x) &= \textstyle\sum_{i=1}^{10} u_i e_i(x) \quad \text{en } \Omega = (0,1) \times (0,1), \\ y(x) &= 0 \qquad\qquad\quad \text{sobre } \Gamma, \end{aligned} \\ y(x_j) \leq -12, \quad \forall j = 1,2,3,4, \\ y(x_5) \leq 12, \end{cases}$$

con

$$x_1 = (0.25, 0.25),\ x_2 = (0.75, 0.25),\ x_3 = (0.75, 0.75),\ x_4 = (0.25, 0.75),$$
$$x_5 = (0.5, 0.5).$$

Las funciones prefijadas e_i, están dadas por

$$e_1(x) = (x_1 + x_2)^2,\ e_2(x) = x_1^2 + x_2^2,\ e_3(x) = (x_1 - x_2)^2,\ e_4(x) = x_1^3 + x_2^3,$$
$$e_5(x) = -4\pi^2 \cos(2\pi(x_1 - x_2)),\ e_6(x) = -4\pi^2 \cos(2\pi(x_1 + x_2))$$
$$e_7(x) = x_2^2 - x_1^2,\ e_8(x) = x_1 x_2^2 - 1,\ e_9(x) = \sin(2\pi x_1),\ e_{10}(x) = \sin(2\pi x_2).$$

Además consideramos el conjunto de los controles admisibles, definido por

$$\mathcal{U}_{ad} = \{u \in \mathbb{R}^{10} : -100 \leq u_i \leq 100, \quad \forall i = 1, 2, \ldots, 10\},$$

y el estado deseado es

$$y_d(x) = -\cos(2\pi x_1)\cos(2\pi x_2) + 1000.$$

Calculamos una solución aproximada $\bar{u}_h^* = \bar{u}^h$ con $h = 0.001953125$, la cual es considerada como solución "exacta" ya que la solución analítica es desconocida. Luego,

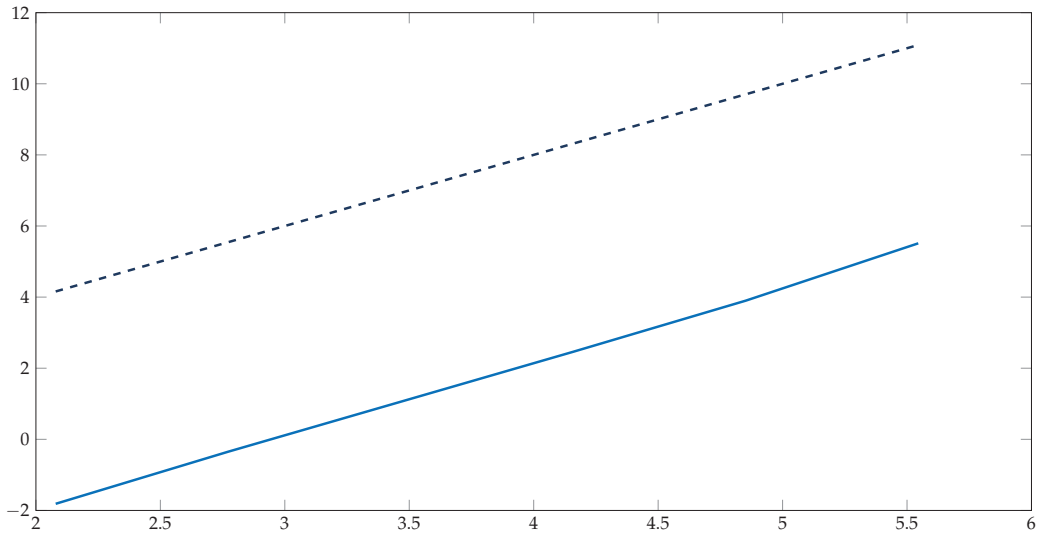

Figura 6.1: Ejemplo 1; $-\log(h)$ versus $-\log(||\bar{u}^h - \bar{u}^*_h||)$ (linea sólida) comparada con $-2\log(h)$ (linea entrecortada).

de referencia calculamos la solución para diferentes tamaños mallas, y obtenemos el error usando la solución de referencia $\bar{u}^*_h$. Además, medimos el orden del error de acuerdo a (6.1), obteniendo los siguientes resultados:

h	$\|\|\bar{u}^h - \bar{u}^*_h\|\|$	EOC'
0.125	6.12921317	-
0.0625	1.41605888	2.1138
0.015625	0.08520188	2.0274
0.0078125	0.02024204	2.0735
0.00390625	0.00403609	2.3263

Observamos que el estimador del error es 2 por lo que el orden es aproximadamente h^2. En la Figura 6.1 se compara el error numérico con la función h^2 en escala logarítmica. Donde se puede observar que aproximadamente tienen la misma pendiente.

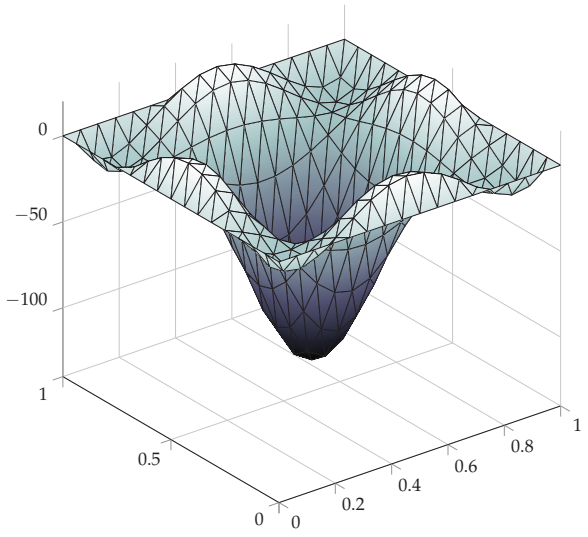

Figura 6.2: $\bar{y}^h$ calculado con $h = 0.0625$.

El control óptimo calculado para $h = 0.0625$, viene dado por:

$$\bar{u}^h = \begin{pmatrix} -25.2871 \\ -9.3130 \\ 0 \\ -3.0135 \\ 91.0788 \\ 89.2871 \\ 0 \\ 13.9271 \\ 0 \\ 0 \end{pmatrix},$$

observamos que los controles $\bar{u}_3^h, \bar{u}_7^h, \bar{u}_9^h$ y $\bar{u}_{10}^h$ satisfacen la propiedad de dispersión.

En la figura 6.2 se visualiza el estado óptimo $\bar{y}^h$ para $h = 0.0625$.

El vector de las restricciones de estado para $h = 0.0625$, es

$$\left[\bar{y}^h(x_j)\right]_{j=1}^{5} = \begin{pmatrix} -21.9103 \\ -18.5031 \\ -23.2232 \\ -18.5031 \\ -121.7314 \end{pmatrix}$$

esto refleja de que nuestro ejemplo no posee restricciones activas.

EJEMPLO 2. Consideramos el siguiente problema con 5 puntos disjuntos para las restricciones de estado y 8 controles.

$$(P) \qquad \begin{cases} \min\limits_{u \in \mathcal{U}_{ad} \subset \mathbb{R}^8} J(y,u) = \dfrac{1}{2}||y - y_d||^2 + \dfrac{1}{80}||u||_2^2 + 5||u||_1 \\ \text{sujeto a:} \\ \quad -\Delta y(x) + y(x) \;\; = \sum_{i=1}^{8} u_i e_i(x), \quad \text{en } \Omega = (0,1) \times (0,1), \\ \qquad\qquad\quad y(x) \;\; = 0, \qquad\qquad\quad \text{sobre } \Gamma, \\ y(x_j) \leq -10, \quad \forall j = 1,2,3,4,5, \end{cases}$$

con

$$x_1 = (0.08, 0.4),\ x_2 = (0.4, 0.4),\ x_3 = (0.84, 0.12),\ x_4 = (0.12, 0.44),\ x_5 = (0.2, 0.24).$$

Las funciones prefijadas e_i, están tomadas como sigue

$$\begin{aligned} &e_1(x) = x_1 + x_2,\ e_2(x) = 8\pi^2 \sin(2\pi x_1)\sin(2\pi x_2),\ e_3(x) = x_1 - x_2, \\ &e_4(x) = \cos^3(\pi x_1),\ e_5(x) = 4\pi^2 \cos(2\pi(x_1 + x_2)),\ e_6(x) = 4\pi^2 \cos(2\pi(x_1 - x_2)), \\ &e_7(x) = x_1^2 + x_2^2,\ e_8(x) = (x_1^2 - 1)(x_2^2 - 1)(x_1^2 + x_2^2). \end{aligned}$$

Además, el conjunto de los controles admisibles está dado por

$$\mathcal{U}_{ad} = \{u \in \mathbb{R}^8 : -500 \leq u_i \leq 500, \quad \forall i = 1, 2, \ldots, 8\},$$

y el estado deseado es

$$y_d(x) = -2\sin(2\pi x_1)\sin(2\pi x_2) + 5.$$

Calculamos una solución aproximada $\bar{u}_h^* = \bar{u}^h$ con $h = 0.00125$, la cual es considerada como solución "exacta" ya que la solución analítica es desconocida. Luego, de referencia calculamos la solución para diferentes mallas con tamaños descendentes, y calculamos el error usando la solución de referencia $\bar{u}_h^*$. Medimos el orden del error de acuerdo a (6.1), obteniendo los siguientes resultados:

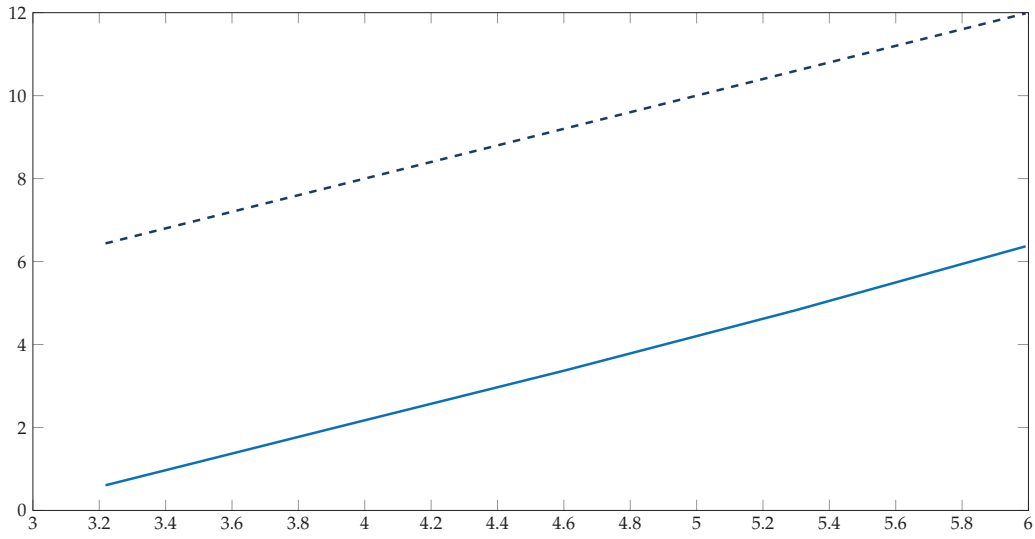

Figura 6.3: Ejemplo 2; $-\log(h)$ versus $-\log(||\bar{u}^h - \bar{u}_h^*||)$ (linea sólida) comparada con $-2\log(h)$ (linea entrecortada).

h	$\|\|\bar{u}^h - \bar{u}_h^*\|\|$	EOC'
0.04	0.54515935	-
0.02	0.13564030	2.0069
0.01	0.03410037	1.9919
0.005	0.00802171	2.0878
0.0025	0.00171625	2.2247

Observamos que el estimador del error es 2 por lo que el orden es aproximadamente h^2. En la Figura 6.3 se compara el error numérico con la función h^2 en escala logarítmica. Donde se puede observar que aproximadamente tienen la misma pendiente.

El control óptimo calculado para $h = 0.04$, está dado por:

$$\bar{u}^h = \begin{pmatrix} -455.9386 \\ 0 \\ 0 \\ 0 \\ 9.7017 \\ 13.1094 \\ -160.0708 \\ 0 \end{pmatrix},$$

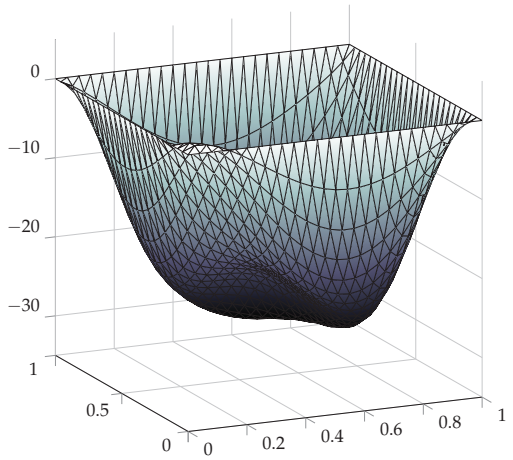

Figura 6.4: $\bar{y}^h$ calculado con $h = 0.04$.

vemos que la propiedad de dispersión se satisface en los controles $\bar{u}_2^h$, $\bar{u}_3^h$, $\bar{u}_4^h$ y $\bar{u}_8^h$.

En la Figura 6.4 se visualiza el estado óptimo $\bar{y}^h$ para $h = 0.04$.

De igual manera para $h = 0.04$, el vector de las restricciones de estado es

$$\left[\bar{y}^h(x_j)\right]_{j=1}^{5} = \begin{pmatrix} -9.9948 \\ -19.7645 \\ -9.9964 \\ -14.5492 \\ -9.9994 \end{pmatrix},$$

y esto en vista de que en los puntos x_1, x_3 y x_5 las restricciones son activas.

Capítulo 7

Conclusiones y comentarios

1. Al haber aproximado numéricamente el problema de control óptimo gobernado por una ecuación diferencial parcial elíptica de tipo Dirichlet con restricciones de control y estado, mediante el método de elementos finitos, se dedujo que el orden de las estimaciones del error para el estado óptimo es

$$||\bar{y} - \bar{y}^h||_{C(\Omega_0)} \leq C_1 h^2 |\log h|, \tag{7.1}$$

donde Ω_0 es un subdominio de Ω que contiene a x_j para todo $j = 1, \ldots, l$, y $C_1 > 0$ independiente de h. Así mismo, el orden de las estimaciones del error para el control óptimo es

$$||\bar{u} - \bar{u}^h|| \leq C_2 h^2 |\log h|, \tag{7.2}$$

para $C_2 > 0$ independiente de h. Por tanto, no hay disminución del orden de error para el control, en consecuencia es óptimo. Esto es relativamente esperado ya que no hay aproximación por elementos finitos para los controles.

2. Para el problema sin regularización de Tikhonov ($\alpha = 0$), al haber realizado la Hipótesis 5.1 se obtiene los mismos resultados que el problema (P), es decir el orden de las estimaciones del error para el control óptimo esta dado también por (7.2).

3. Los experimentos numéricos realizados en el Capítulo 6 verifican satisfactoriamente el orden del error dado por (7.2), para tal comprobación utilizamos el error experimental de convergencia EOC' definido en (6.1), es importante recalcar que en el caso de controles funcionales esta es una pregunta abierta, que para nuestro caso pudo ser respondida gracias a la estructura finita de los

controles.

4. Los resultados obtenidos del estudio de ecuaciones generalizadas fuertemente regulares resultaron ser una herramienta fundamental para la obtención del orden de estimación del error para el control.

5. La resolución numérica implica la aproximación por elementos finitos y luego la aplicación del algoritmo OESOM para la resolución de problemas dispersos en dimensión finita.

6. Una prolongación de este proyecto de investigación sería el realizar una estimación de errores para el caso en que se tenga restricciones puntuales de estado sobre todo el dominio en lugar de restricciones puntuales finitas. Otra variante interesante resultaría el analizar el mismo problema de control pero gobernado por una ecuación diferencial parcial semilineal elíptica de tipo Dirichlet o Neumann.

Bibliografía

[1] H. W. ALT, *Linear Functional Analysis An Application-Oriented Introduction*, Springer, Alemania, 2012.

[2] R. BARTLE, *The elements of integration and Lebesgue Measure*, Wiley Classics Library, Estados Unidos, 1995.

[3] J. BONNANS AND E. CASAS, *Contrôle de systèmes elliptiques similinéaires comportant des contraintes sur l'état* , (French. English summary) [Control of semilinear elliptic systems with state constraints] Nonlinear partial differential equations and their applications. Collège de France seminar, VIII (1988), pp. 69–86.

[4] H. BREZIS, *Functional analysis, sobolev spaces and partial differential equations*, Springer, Estados Unidos, 2010.

[5] E. CASAS, R. HERZOG, AND G. WACHSMUTH, *Optimality conditions and error analysis of semilinear elliptic control problems with L^1 cost functional*, SIAM J. Optim., 22 (2012), p. 795–820.

[6] E. CASAS AND K. KUNISCH, *Optimal control of semilinear elliptic equations in measure spaces*, SIAM J. Control Optim., 52 (2014), p. 339–364.

[7] E. CASAS AND F. TRÖLTZSCH, *Second-order and stability analysis for state-constrained elliptic optimal control problems with sparse controls*, SIAM J. Control Optim., 52 (2014), p. 1010–1033.

[8] P. CIARLET, *The finite element method for elliptic problems*, North-Holland, Amsterdam, 1978.

[9] F. CLARKE, *Functional Analysis, Calculus of Variations and Optimal Control*, Springer, Francia, 2013.

[10] F. H. CLARKE, *Optimization and Nonsmooth Analysis* , John Wiley & Sons, Estados Unidos, 1983.

[11] J. C. De Los Reyes, *Numerical PDE-Constrained Optimization*, Springer, Ecuador, 2015.

[12] J. C. De Los Reyes, E. Loayza, and P. Merino, *Second-order orthant-based methods with enriched Hessian information for sparse ℓ_1-optimization*, Comput. Optim. Appl., 67 (2017), p. 225–258.

[13] J. C. De Los Reyes, P. Merino, J. Rehberg, and F. Tröltzsch, *Optimality conditions for state-constrained PDE control problems with time-dependent controls*, Control Cybernet, 37 (2008), p. 5–38.

[14] L. Evans, *Partial Differential Equations*, AMS, Estados Unidos, 1998.

[15] D. Gilbarg and N. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Springer, Alemania, 1998.

[16] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman, Philadelphia, 1985.

[17] R. Henrion, *On constraint qualifications*, Optimization Theory and Applications, 72 (1992), pp. 187–197.

[18] K. Kunisch and D. Wachsmuth, *On time optimal control of the wave equation and its numerical realization as parametric optimization problem*, SIAM J. Control Optim., 51 (2013), p. 1232–1262.

[19] P. Merino, *Optimal Control Problems of Semiliear Partial Differential Equations with Finite-Dimensional Control Space*, PhD thesis, Escuela Politécnica Nacional, Ecuador, 2011.

[20] P. Merino, I. Neitzel, and F. Tröltzsch, *On linear-quadratic elliptic control problems of semi-infinite type* , Applicable Analysis, 90 (2011), pp. 1047–1074.

[21] P. Merino, F. Tröltzsch, and B. Vexler, *Error Estimates for the Finite Element Approximation of a Semilinear Elliptic Control Problem with State Constraints and Finite Dimensional Control Space* , Mathematical Modelling and Numerical Analysis, 44 (2010), pp. 167–188.

[22] J. Nocedal and S. J. Wright, *Numerical Optimization*, Springer, New York, 1999.

[23] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, SIAM, Estados Unidos, 2000.

[24] R. RANNACHER AND B. VEXLER, *A priori error estimates for the finite element discretization of elliptic parameter identification problems with pointwise measurements*, SIAM J. Control Optim., 44 (2005), pp. 1844 –1863.

[25] S. M. ROBINSON, *Stability Theory for Systems of Inequalities, part II: Differentiable Nonlinear Systems**, SIAM J. Numer. Anal., 13 (1976), pp. 497–513.

[26] ——, *Strongly Regular Generalized Equations*, Mathematics of Operations Research, 5 (1980), pp. 43–62.

[27] ——, *Generalized Equations and Their Solutions, Part II Applications to Nonlinear Programming* , Mathematical Programming Study, 19 (1982), pp. 200–221.

[28] G. STADLER, *Elliptic optimal control problems with L^1-control cost and applications for the placement of control devices*, Comput. Optim. Appl., 44 (2009), p. 159–181.

[29] H. TIANHONG, *Lasso and General ℓ_1 Regularized Regression under Linear Equality and Inequality Constraints*, PhD thesis, Purdue University, Estados Unidos, 2011.

[30] F. TRÖLTZSCH, *Optimal Control of Partial Differential Equations Theory, Methods and Applications*, AMS, Estados Unidos, 2010.

[31] V. VACA, *Control Óptimo y Resolución Numérica de Problema Parabólicos con Controles Finitos y Dispersos* . Proyecto de titulación previo a la obtención del título de Matemática, Escuela Politécnica Nacional, 2016.