FACULTAD DE MATEMÁTICAS

DEPARTAMENTO DE ECUACIONES DIFERENCIALES Y ANÁLISIS NUMÉRICO

ANÁLISIS Y CONTROL DE ALGUNAS EDPS NO LINEALES

Irene Marín Gayte

Dirigido por:
Enrique Fernández Cara

Abstract

This work is framed within the field Theory of Partial Differential Equations. The study is divided into three main parts. In the first part, a work related to the theoretical analysis of a PDE is provided, specifically, the existence of a kind of solution for the Navier-Stokes equation is studied. The second part includes three works related to Optimal Control Theory. Here, two bi-objective problems associated with some PDEs and a minimal time control problem are added. The work ends including a last part with two works related to controllability problems. Then, it has included a null control problem associated with a nonlinear equation and with a semilinear heat equation. All the numerical implementations have been carried out with MatLab and FreeFem ++ . The conclusions are detailed separately in each chapter.

Agradecimientos

Esta memoria surge fruto de varios años en los que he podido iniciarme y formarme en el mundo de la investigación matemática. Pero como todo proceso de formación, ésta no habría llegado a su culmen si no fuera gracias a la ayuda de personas que, de manera directa o indirecta, me han ido guiando.

En primer lugar, me gustaría dar las gracias a Dios. Él me ha guiado a lo largo de estos años de vida y me ha regalado el don de ciencia para poder adentrarme en este maravilloso mundo de las matemáticas, las cuales nos ayudan a entender su creación. También me gustaría agradecerle que me haya dado a mi madre Inmaculada Gayte Delgado, a mi familia, a mi novio Álvaro y a mis amigos, entre ellos al Padre Álvaro Pereira. Ellos son mi apoyo desde las sombras y los que me han formado y me seguirán formando personal y profesionalmente.

Cada vez que echo la vista atrás, agradezco a mi profesor de secundaria, Don Antonio Ruiz, esa conversación en la que me animaba a estudiar la carrera de Matemáticas en esos momentos en los que una no sabía qué elegir. A él le doy las gracias por enseñarme mis primeras matemáticas y mostrarme lo apasionante que es su estudio.

Pero, sin duda alguna, la persona a la que debo agradecer esta memoria es a mi director, Enrique Fernández Cara. Él, con su paciencia y sus horas de dedicación exclusivamente para mí, me ha ayudado a iniciarme en el campo de la investigación y me ha trasmitido su pasión por las ecuaciones. También agradezco al departamento de EDAN, que me ha acogido estos años, a mis compañeros profesores, becarios y personal de administración y servicios.

Por último, y no por ello menos importante, me gustaría dar las gracias al "Laboratoire de mathématiques Blaise Pascal" de Clermont-Ferrand que me acogió durante mi estancia internacional. En concreto, quería darle las gracias a los Profesores Arnaud Münch y Jérôme Lémoine por guiarme y permitirme trabajar con ellos durante mis días en Francia, sin su ayuda no habría sido posible realizar el trabajo relativo al Capítulo 6 de esta memoria.

A todos ellos y por todos ellos, GRACIAS.

Índice general

Introducción

En este trabajo se presentan varios resultados de tipo teórico y numérico para diversos problemas diferenciales. Comenzaremos analizando un resultado de existencia de solución "admisible" para las EDPs de Navier-Stokes con condiciones de contorno de tipo Dirichlet. A continuación, trataremos dos problemas de control óptimo multi-objetivo asociados a EDPs estacionarias (elíptica lineal y semi-lineal y Navier-Stokes). Por otro lado, estudiaremos problemas de tiempo mínimo para EDOs y para la ecuación del calor. También, obtendremos y analizaremos resultados de controlabilidad nula para una EDP parabólica quasi-lineal. Finalmente, presentaremos métodos de aproximación numérica de controles nulos parabólicos semi-lineales basados en mínimos cuadrados.

Aspectos históricos:

A lo largo de la Historia el hombre siempre ha estado en constante búsqueda de leyes y principios que rijan formas o fenómenos naturales, con el propósito de explicar o entender el comportamiento de la Naturaleza. Así, el francés Pierre Louis Moreau de Maupertuis (1698-1759) enunció, en 1744, el principio de Mínima Acción, por el cual se establece que:

> "En todo cambio que se produzca en la Naturaleza, la cantidad de acción necesaria ha de ser la mínima posible".

Este es un principio físico que posteriormente las Matemáticas han fundamentado rigurosamente. En particular, se han desarrollado las técnicas necesarias para dar respuesta a una amplia clase de problemas de optimización.

Desde la época romana se han estudiado problemas relacionados con la optimización y el control de fenómenos naturales, como por ejemplo los que aparecen en relación con la construcción de los acueductos. De este modo, mediante un adecuado sistema de válvulas, chimeneas, bifurcaciones, etc. se pretendió y se consiguió mantener un nivel constante de agua, algo esencial para el sumnistro diario de una pobalción.

También, en el antiguo Egipto existía el oficio de "estirador de cuerdas", que tenía como objetivo producir largos segmentos rectos que ayudasen a la construcción de las Pirámides. Se considera que esto es una evidencia de que ya por entonces se había comprendido que la distancia más corta entre dos puntos es la línea recta. De esta manera, se conseguía minimizar la acción o energía empleada (el principio clásico y fundamental de la Optimización y del Cálculo de Variaciones).

Más tarde, a finales del siglo XVII, aparecen los trabajos sobre el péndulo de Ch. Huygens y R. Hooke, en los que se intenta de medir de forma precisa el tiempo y, en ellos, vemos claros de nuevo elementos de lo que hoy llamamos la Teoría de Control.

El primer análisis riguroso de los mecanismos asociados al control de sistemas fue realizado por J.C. Maxwell en 1868. Para ello usó herramientas de la teoría cualitativa de las EDOs.

Gracias a los avances de la revolución industrial, fueron tomando forma la Teoría de Control y, de este modo, la Ingeniería del Control empezó a ser reconocida como una disciplina científico-tecnológica con relevancia.

Durante la Segunda Guerra Mundial y los años que la siguieron, los ingenieros y científicos tuvieron que afinar su experiencia en los mecanismos de control de seguimiento de aviones y proyectiles y en el diseño de baterías antiaéreas.

A partir de 1960, todo lo que se había desarrollado hasta ese momento empezó a conocerse como Teoría Clásica de Control. Comenzó una nueva era en la que se pretendía desarrollar nuevos modelos con los que se pudiera representar la complejidad del mundo real, dando cuenta del comportamiento no lineal y no determinista.

Gracias a las contribuciones de R. Bellman (programación dinámica), R. Kalman (filtrado y análisis algebraico de problemas de control) y L. Pontryagin (principio del máximo para problemas de control óptimo no-lineal) quedó fundamentada la investigación en este campo.

En este sentido también cabe destacar el papel jugado por D.E. Russel y J.-L. Lions, ilustres matemáticos, que influyeron de manera decisiva en el desarrollo de esta disciplina, el segundo en particular en sus conexiones con las EDPs, el Análisis Numérico y las aplicaciones industriales.

En lo que se refiere al control óptimo, los primeros trabajos se remontan a 1937 con Valentine, 1939 con McShane y 1947 con Hestenes. Años más tarde, en la década de los 50, un grupo de ingenieros especialmente motivados por los problemas aeroespaciales se interesó por el control de sistemas gobernados por ecuaciones diferenciales. En dichos casos era natural querer controlar de forma que se minimizara una variable dada. Se comprendió que, con una pequeña mejora en el rendimiento, se podía obtener grandes ahorros en el coste. Sin embargo, el tema no atrajo gran atención hasta que el matemático ruso Pontryagin y su equipo, Boltyanskii, Gamkrelidze y Mishchenko, desarrollaron con rigor matemático lo que hoy se conoce como Principio del Máximo (un resultado que proporciona condiciones necesarias para la optimalidad de un problema planteado).

Mencionamos que otra herrramienta muy útil en Teoría de Control fue descubierta en torno a 1962, cuando Dubovitskii y Milyoutin, consiguen establecer condiciones de optimalidad para problemas con restricciones determinadas mediante la intersección de varios conjuntos, de interior no vacío (restricciones de desigualdad) y/o de interior vacío (restricciones de igualdad).

El control de ecuaciones y sisemas diferenciales ha recibido mucha atención en los últimos tiempos. En particular, el estudio de problemas asociados a EDOs y EDPs no lineales ha generado una amplia y profunda área de investigación, dando informaciones cruciales sobre un número creciente de fenómenos de las distintas ramas de la Ciencia como Física, Biología, Economía, Medicina, etc. e Ingeniería. Si el lector desea conocer más detalles sobre lo que precede, puede consultar [13].

Muchos son los campos donde se presentan retos para la Teoría de Control. En algunos casos se confía en ser capaces de resolver éstos mediante avances tecnológicos que permitan la implementación de controles más eficientes; es el caso, por ejemplo, del control molecular mediante tecnología láser; véase [4]. También, la Robótica es una de las áreas de la Tecnología que presenta los retos más estimulantes para los próximos años. En este caso, la Teoría de Control está también muy implicada en el avance, puesto que el desarrollo de los robots depende de manera fundamental de la eficiencia y robustez de los correspondientes algoritmos computacionales. A este respecto, no

resulta difícil imaginar la complejidad del proceso de control que hace que un robot pueda imitar comportamientos humanos tales como mover las extremidades, agarrar objetos, etc. (véase [32]).

Por otro lado, el control de fluidos presenta grandes retos debido a su poder para evitar desastres medioambientales, como las inundaciones. Es el caso por ejemplo de la barrera del Támesis, en el sur de Inglaterra, con la que se consigue controlar el nivel del agua, evitando así que una eventual crecida cause daños irreparables. Para tomar la decisión de cerrar la barrera se usan y se resuelven numéricamente modelos de EDPs que predicen y simulan las condiciones meteorológicas. A pesar de que la barrera responde a las necesidades de hoy, el problema no está resuelto a largo plazo, puesto que el nivel medio del río sube 75 centímetros cada siglo de modo que, con el tiempo, este método dejará de ser eficiente.

El control de fluidos puede aplicarse también al campo aeroespacial, en el que se busca optimizar la forma o perfil del ala de una aeronave, de manera que se gobierne el flujo de aire que hay a su alrededor, véase [41]. También tienen relevancia aplicaciones en Medicina, donde se pretende controlar el comportamiento de los fluidos que invaden células o moléculas que se incorporan a la sangre; por ejemplo, en personas diabéticas debemos controlar la concentración de glucosa e insulina.

En todas estas aplicaciones se necesitan importantes avances teóricos, como por ejemplo en el caso del control de células cancerígenas de un tumor, véase [23]. Aunque en los últimos años se ha progresado considerablemente en el estudio de la Teoría de Control, dando lugar a numerosos trabajos como son [6, 14–16, 19, 27, 40], aún quedan multitud de preguntas y retos sin resolver, véase [28, 35].

Control óptimo y controlabilidad:

En términos generales, desde el punto de vista de la Teoría Clásica de Control, el objetivo es hallar un dato que lleve una ecuación o sistema a un estado deseado o (al menos) cerca de él. Así, el sistema físico puede venir descrito por una ecuación de estado

$$A(u) = b(f), \tag{1}$$

donde u es la variable de estado (que da información sobre la situación del sistema) y f es un dato (el control) que se puede elegir libremente dentro de una familia (un conjunto de controles admisibles) que varía en función del problema y de las restricciones que se impongan.

En la práctica, (1) es una ecuación o sistema algebraico o funcional (integral, diferencial ordinario, en derivadas parciales, etc.), que eventualmente debe ser completado con condiciones iniciales, de contorno u otras.

No siempre existen controles que lleven el estado a una situación deseada. Por ello, cuando esta propiedad se cumple, se dice que el sistema es controlable. Pero nada impide que en dicho caso pueda existir más de un control que satisface el objetivo.

Frecuentemente, aunque no siempre, hacemos el problema un poco más sencillo marcando como objetivo minimizar un determinado funcional dentro de una familia de controles admisibles. Éstos son los llamados problemas de control óptimo y tienen multitud de aplicaciones.

Una situación diferente pero también interesante, es la que surge cuando se pretende minimizar varios funcionales al mismo tiempo. Llegamos así a los llamados problemas de control óptimo multi-objetivo.

Los problemas de control óptimo son muy interesantes desde el punto de vista teórico y numérico y también por su gran conexión con diversas aplicaciones. La motivación principal de estos problemas es el deseo de actuar sobre el sistema de manera que no sólo su evolución sea la mejor posible sino que, además, esto ocurra con mínimo esfuerzo.

En nuestra memoria analizaremos problemas de control óptimo en los que la ecuación de estado (1) que liga las variables a minimizar es una EDO o una EDP (o un sistema de EDOs y/o EDPs). Así, un problema de control óptimo estándar toma la forma:

$$\begin{cases} \text{Mín } J(f) \\ \text{Sujeto a } \ f \in U_{ad}, \ \ A(u) = b(f), \end{cases} \tag{2}$$

donde U_{ad} es un conjunto del espacio de controles en el que queremos encontrar la solución.

Queremos establecer condiciones de existencia (y posible unicidad) de solución (denominado "control óptimo"), caracterizar dicha solución mediante lo que se denomina "sistema de optimalidad" y, finalmente, proporcionar algoritmos o métodos de cálculo.

Las principales herramientas para el estudio y caracterización del control óptimo son el llamado Principio del Máximo de Pontryagin (1962), el Principio del Máximo Local y las condiciones de optimalidad de primer orden de Karush-Kuhn-Tucker (KKT, 1939).

El Principio del Máximo de Pontryagin proporciona condiciones necesarias de optimalidad para conjuntos U_{ad} arbitrarios. En cambio, el Principio del Máximo Local es más restrictivo y corresponde al caso en que U_{ad} es convexo con interior no vacío.

Las condiciones necesarias de optimalidad de primer orden de Karush-Kuhn-Tucker, véase [22], para problemas de optimización matemática y las establecidas por el Principio del Máximo de Pontryagin, véase [31], para problemas de control óptimo, no son suficientes, en general, para decidir si el punto analizado es de hecho un óptimo global. Consecuentemente, tenemos dos caminos a seguir para decidir si el candidato es óptimo: usar condiciones de segundo orden o alguna estructura especial de los funcionales involucrados en el problema (convexidad, deformación para problemas más simples, etc).

Si las funciones involucradas en el problema son convexas, las condiciones necesarias de optimalidad son también suficientes. Pero hay una amplia clase de problemas no convexos de gran interés.

Con el fin de debilitar la hipótesis de convexidad, surgieron en la literatura varias corrientes que pretendieron dar respuestas a estos casos. Una de estas obras surge en 1962 con Dubovitskii y Milyoutin. En ella, consiguen establecer un formalismo que proporciona condiciones de optimalidad, para problemas con restricciones determinadas por la intersección de varios conjuntos en términos de funcionales lineales y continuos, donde cada uno de estos funcionales se relaciona con un conjunto de restricciones, véase [17].

Dado el enfoque universal y unificador del formalismo de Dubovitskii-Milyoutin, éste permite determinar, con el lenguaje del Análisis Funcional, condiciones necesarias de optimalidad para una amplia clase de problemas de control, tanto con ecuaciones lineales como no lineales.

En este trabajo ampliaremos esta Teoría de Control Óptimo para problemas multi-objetivo, con varios funcionales a minimizar y uno o varios controles. Nos serviremos principalmente del formalismo de Dubovitskii-Milyoutin para demostrar la existencia y caracterización de solución. Es importante recalcar que, en un problema de optimización multi-objetivo, a diferencia de lo que

ocurre en problemas estándar podemos seguir diferentes estrategias en función de la definición que tomemos de solución.

Obviamente, en estos problemas no existe un mínimo como tal sino más bien lo que se conoce como "equilibrio". Dependiendo de qué definición demos de equilibrio necesitaremos razonar de un modo u otro. Por ejemplo, podemos definir equilibrios cooperativos, en los que todos los controles "cooperan" para minimizar en cierto sentido todos los funcionales; éstos son por ejemplo los equilibrios de Pareto, véase [30]. Pero también podemos considerar equilibrios en los que cada control se ocupe de alguna manera de un solo funcional; en tal caso, estaríamos hablando de equilibrios no cooperativos como son los equilibrios de Nash, véase [29], o Stackelberg, véase [37]. Todos estos conceptos tienen su origen en la Teoría de Juegos y están motivados principalmente por problemas de la Economía.

Otro ámbito de estudio en la Teoría de Control son los llamados problemas de controlabilidad. Así, podemos buscar resolver problemas de controlabilidad nula, de control exacto, de control aproximado, de control frontera, de control a trayectorias, etc. Todos estos problemas tienen en común que lo que se pretende es resolver una ecuación o sistema en la que gracias a la presencia de un control, el estado final del sistema coincide o al menos se parece a un estado dado.

La resolución de este tipo de problemas depende en gran medida de la naturaleza del sistema considerado. En particular, las siguientes características pueden jugar un papel crucial: reversibilidad temporal, regularidad del estado, estructura del conjunto de controles admisibles, etc.

La controlabilidad de las EDPs ha sido objeto de una intensa investigación desde hace más de 40 años. En 1978, Russell [33], hizo un estudio bastante completo de los resultados más relevantes que estaban disponibles en la literatura en ese momento. En ese artículo, el autor describió una serie de herramientas diferentes que fueron desarrolladas para abordar problemas de controlabilidad, a menudo inspiradas y relacionadas con otros temas próximos a las EDPs: método de los multiplicadores, problemas y métodos de los momentos, series de Fourier no armónicas, etc. Poco después, J.-L. Lions introdujo el llamado Método de Unicidad de Hilbert (H.U.M.; ver [26]), que permitió conectar los conceptos de controlabilidad, observabilidad, continuación única, etc..

En un problema de controlabilidad habitual, la ecuación de estado es de la forma

$$\begin{cases} y_t + A(y) = B(v) \quad t \in (0,T), \\ y(0) = y_0. \end{cases} \tag{3}$$

En este punto suponemos que y es el estado (que identifica las propiedades físicas del sistema), v es el control (que determina la acción que ejercemos), A y B son operadores diferenciales o algebraicos y T es el tiempo final de resolución del problema.

Si el sistema (3) está bien planteado, es decir, para cada dato y_0 y cada control v en determinados espacios existe exactamente una solución de la ecuación (3), entonces podemos enunciar diversos problemas de controlabilidad:

- *Controlabilidad nula:* Para cada y_0 encontrar v tal que la correspondiente función y verifica $y(T) = 0$.
- *Controlabilidad exacta:* Para cada y_0 encontrar v tal que $y(T)$ sea igual a un dato dado.
- *Controlabilidad aproximada:* Para cada y_0, cada y_d y cada $\delta > 0$ encontrar v tal que $\|y(T) - y_d\| \leq \delta$.

- *Controlabilidad exacta a una trayectoria:* Para cada y_0 y cada trayectoria libre y^* del sistema, es decir, solución de la ecuación
$$y_t + A(y) = 0,$$
hallar v tal que $y(T) = y^*(T)$.

Nótese que la controlabilidad exacta a trayectorias es una propiedad muy útil desde el punto de vista de las aplicaciones: si podemos encontrar un control tal que $y(T) = y^*(T)$, después de un tiempo, podemos suprimir el control y dejar que el sistema "siga" la trayectoria de manera autónoma. Para cada sistema de la forma (3), estos problemas conducen a varias preguntas interesantes, como son aquéllas relativas a la existencia de v, la caracterización del control de norma mínima o su cálculo.

En el contexto de las EDPs, una de las principales herramientas que permite abordar problemas de controlabilidad son las llamadas desigualdades globales de Carleman. Estas desigualdades han sido aplicadas en muchos trabajos, como son [7, 19–21, 38].

Las desigualdades de Carleman son estimaciones ponderadas mediante funciones de peso adecuadas que permiten acotar integrales globales de las soluciones. Supongamos que (3) es una EDP lineal y debe cumplirse en $\Omega \times (0,T)$, donde $\Omega \subset \mathbb{R}^N$ es un abierto no vacío. En su forma más simple, la desigualdad de Carleman asociada tiene la forma

$$\iint_{\Omega\times(0,T)} \rho_1^{-2}|\varphi|^2 \, dx \, dt \leq C \iint_{\omega\times(0,T)} \rho_2^{-2}|\varphi|^2 \, dx \, dt,$$

donde $\omega \subset \Omega$ es un abierto no vacío, φ es solución del problema adjunto asociado a (3)

$$\begin{cases} -\varphi_t + A^*(\varphi) = 0, \\ \varphi(T) = \varphi_T \end{cases}$$

y $C > 0$ es una constante independiente de φ. En este caso, las funciones de peso o de Carleman son ρ_1 y ρ_2, funciones positivas que "explotan" al menos cuando $t \nearrow T$.

Gracias a las desigualdades de Carleman, el Análisis de Fourier o los argumentos Hilbertianos, se puede conseguir con frecuencia resolver un problema de controlabilidad (aunque, como es natural, la complejidad de los argumentos dependerá de las propiedades que tenga la ecuación de estado).

En este trabajo, $\Omega \subset \mathbb{R}^N$ ($N \leq 3$) denotará un abierto acotado, $\omega, \omega_1, \omega_2 \subset \Omega$ serán los dominios de control, abiertos y no vacíos, 1_U denotará la función característica del conjunto U y $T > 0$ será el tiempo final de evolución del sistema. Para las experiencias numéricas que se incluyen en el trabajo, hemos usado los softwares MatLab y FreeFem++.

Descripción de la memoria:

La memoria está dividida en seis capítulos. Cada uno de ellos aborda un problema distinto.

En el primer capítulo, se da una nueva prueba del resultado presentado por Caffarelli-Kohn-Nirenberg en 1982, véase [5]. Este trabajo es de corte más teórico y se basa en un análisis detallado de las propiedades de las aproximaciones de las ecuaciones de Navier-Stokes. En este sentido,

somos capaces de demostrar que el límite de las sucesiones dadas por un esquema de Euler semi-implícito, tanto en el caso semi-discretizado como completamente discretizado, aplicado a la ecuación de Navier-Stokes en dimensión 3 con condiciones de Dirichlet, es una solución "admisible" en el sentido de Scheffer, véase [34]. El interés práctico de esta nueva prueba reside principalmente en que esta demostración ayuda en la comprobación de los criterios de Caffarelli-Kohn-Nirenberg y, también, podría ayudar a localizar los puntos singulares, gracias a que las soluciones son límites de sistemas discretos. Las técnicas aquí utilizadas se pueden aplicar a muchos otros esquemas de aproximación que conducen a desigualdades de energía análogas.

En el segundo capítulo abordamos un problema de control multi-objetivo. En este caso estudiaremos la existencia, caracterización, aproximación y simulación numérica de los equilibrios de Pareto asociados a problemas de control óptimo para una EDP de Poisson, una EDP elíptica semilineal y las ecuaciones de Navier-Stokes estacionaria. Análogamente, en el capítulo tercero estudiaremos para estos mismos problemas los equilibrios de Nash tanto desde el punto de vista teórico como numérico. En estos dos trabajos hemos utilizado el formalismo de Dubovitskii-Milyoutin en los casos en que los argumentos clásicos de convexidad fallan como es en el caso de las ecuaciones de Navier-Stokes estacionaria. El interés de estos trabajos es que proporcionan una nueva visión de los equilibrios de Pareto y de Nash asociados a problemas de control óptimo, algo que se puede extender y aplicar a muchas otras ecuaciones y sistemas.

En el cuarto capítulo estudiamos problemas de tiempo mínimo. Éstos se corresponden con problemas de control óptimo en el que dentro del funcional que se desea minimizar se incluye también la variable temporal, de modo que no sólo se desea resolver el problema con mínimo "esfuerzo" posible sino también en el menor tiempo posible. En este caso, comenzaremos el estudio viendo qué ocurre cuando nos encontramos con el problema asociado a EDOs lineales y no lineales. Así, terminaremos el trabajo presentando resultados para la EDP del calor. Se dan varios algoritmos que permiten calcular el óptimo y además se incluyen varias simulaciones numéricas.

En el quinto y sexto capítulo abordaremos cuestiones relacionadas con la controlabilidad. El primero de ellos, en el Capítulo 5 es el problema de controlabilidad nula asociada a una EDP de tipo parabólica quasi-lineal. Así, estudiaremos la existencia y caracterización de solución para este problema y terminaremos introduciendo un esquema numérico de aproximación y su posterior simulación en el ordenador. Para la demostración de que el sistema es controlable a cero, usaremos las estimaciones de Carleman y nos serviremos de un sistema auxiliar linealizado. También, el algoritmo de aproximación que usaremos es de tipo quasi-Newton. Este trabajo se ha realizado en colaboración con Juan Límaco.

En el sexto capítulo terminaremos abordando el problema de controlabilidad nula para la EDP del calor semi-lineal. En este caso, para demostrar que el sistema es controlable a cero, usaremos de nuevo las estimaciones de Carleman pero nos basaremos también en una aproximación de tipo mínimos cuadrados, es decir, buscaremos la solución como el mínimo de un funcional cuadrático. El método aquí aportado nos ayuda a caracterizar la solución y poder luego incluir resultados y simulaciones. El contenido de este capítulo lo hemos desarrollado en colaboración con Arnaüd Münch y Jérôme Lémoine como fruto de mi estancia en Clermont-Ferrand.

A continuación incluimos un resumen un poco más detallado de cada capítulo en el que explicaremos de forma precisa los resultados y técnicas que se presentan.

Nueva demostración de la existencia de soluciones admisibles y otras observaciones para la ecuación de Navier-Stokes

A new proof of the existence of suitable weak solutions and other remarks for the Navier-Stokes equations

En esta primera parte de la memoria vamos a analizar desde el punto de vista teórico la existencia de un tipo de solución característica para la ecuación de Navier-Stokes. En este trabajo, presentamos una nueva prueba del teorema proporcionado por Caffarelli-Kohn-Nirenberg en [5]; esto ha dado lugar a la publicación [9].

Comenzaremos haciendo un breve repaso de los resultados fundamentales de [5], consideraremos la ecuación de Navier-Stokes

$$\begin{cases} u_t + u \cdot \nabla u - \Delta u + \nabla p = f, \ (x,t) \in Q, \\ \nabla \cdot u = 0, \ (x,t) \in Q, \\ u(x,t) = 0, \ (x,t) \in \Sigma, \\ u(x,0) = u_0(x), \ x \in \Omega, \end{cases} \tag{4}$$

donde $\Omega \subset \mathbb{R}^3$ es un abierto, convexo y acotado de frontera regular $\partial\Omega$, $T > 0$, $Q := \Omega \times (0,T)$, $\Sigma := \partial\Omega$, $f \in L^q(Q)^3$ con $q \geq 2$ y $\nabla \cdot f = 0$ y, por último, $u_0 \in H_0^1(\Omega)$ con $\nabla \cdot u_0 = 0$.

Para esta ecuación, Caffarelli, Kohn y Nirenberg son capaces de demostrar el siguiente resultado:

Teorema 1. *Sea (u,p) una solución "admisible" para la ecuación de Navier-Stokes en un cilindro $D = G \times (a,b) \subset \mathbb{R}^3 \times \mathbb{R}$. Entonces el conjunto de puntos singulares (puntos en los que la solución no es L^∞ en ningún entorno) satisface que $\mathcal{P}^1(S) = 0$, donde $\mathcal{P}^1$ es una medida en $\mathbb{R}^3 \times \mathbb{R}$ que está mayorada por la medida de Hausdorff uno-dimensional.*

En este trabajo decimos que (u,p) es una solución débil "admisible" de (4) en el cilindro D si satisface:

1. $u \in L^\infty(0,T;L^2(G)^3)$, $\partial_i u \in L^2(D)$, $p \in L^{5/3}(D)$, $u = 0$ sobre $\partial G \times (0,T)$,
2. (u,p) verifica las dos primeras igualdades de (4) en el sentido de las distribuciones en D.
3. Para cada $\phi \in C_0^\infty(D)$ con $\phi \geq 0$, se cumple que

$$2 \iint_D |\nabla u|^2 \phi \leq \iint_D \Big(|u|^2(\phi_t + \Delta\phi) + (|u|^2 + 2p) u \cdot \nabla\phi + 2(u \cdot f)\phi \Big). \tag{5}$$

Observemos que (5) es una desigualdad de energía local. El primer miembro corresponde a la verdadera energía que se va perdiendo a lo largo del tiempo debido al rozamiento de las partículas, también llamada energía viscosa. Por tanto, podemos ver la desigualdad (5) como una estimación local en la que se muestra que, en cada punto del espacio, la energía viscosa que se pierde está acotada por términos que dependen únicamente de la velocidad, la presión y las fuerzas a las que

está sometido el fluido. Es notable ver también que en (5) no aparecen derivadas de u ni de p a la derecha.

Caffarelli, Kohn y Nirenberg introdujeron esta desigualdad local en la demostración del teorema por dos motivos. El primero de ellos fue para expresar de manera rigurosa que la energía debida al rozamiento se va perdiendo de manera local, es decir, que leste proceso de pérdida es un fenómeno local. El otro motivo por el que se introdujo (5) fue porque ellos esperaban que con dicha desigualdad se podría demostrar la unicidad de solución admisible; pero hasta el momento nadie ha conseguido este objetivo. De hecho, esto podría deberse a que la desigualdad (5) se hace desde el punto de vista euleriano, es decir, la desigualdad se tiene en cada punto del sistema de referencia y en cada uno de ellos se observa el movimiento de las partículas. Por el contrario, se podría intentar trabajar con una desigualdad análoga a (5) tomando como sistema de referencia cada una de las partículas, es decir, con la función test ϕ anulándose en un entorno de cada partícula e integrales móviles en función del movimiento descrito. Pero aún no se sabe bien cómo escribir esto.

Para la demostración del Teorema 1, Caffarelli-Kohn-Nirenberg se apoyan en dos proposiciones auxiliares en las que consiguen demostrar bajo qué condiciones un punto es no singular. Estas proposiciones son:

Proposición 2. *Supongamos que (u, p) es una solución "admisible" para la ecuación de Navier-Stokes en Q_1. Entonces existen constantes $C_1, \epsilon_1, \epsilon_2 > 0$ tales que, si*

$$\iint_{Q_1} (|u|^3 + |u||p|) + \int_{-1}^{0} \Big(\int_{|x|<1} |p| \, dx \Big)^{5/4} dt \leq \epsilon_1 \tag{6}$$

y

$$\iint_{Q_1} |f|^q \leq \epsilon_2, \tag{7}$$

entonces $|u| \leq C_1$ en $Q_{1/2}$. Aquí, Q_r es el cilindro parabólico en torno a $(0, 0)$ de radio r. En particular, el punto $(0, 0)$ es no singular.

Proposición 3. *Supongamos que (u, p) es una solución "admisible" para la ecuación de Navier-Stokes en un entorno de un punto (x, t). Entonces existe ϵ_3 tal que, si*

$$\limsup_{r \to 0} \frac{1}{r} \iint_{Q_r^*(x,t)} |\nabla u|^2 \leq \epsilon_3, \tag{8}$$

el punto (x, t) es no singular. De nuevo, $Q_r^*(x, t)$ es un cilindro parabólico en torno a (x, t) de radio r.

En este punto, nos gustaría mencionar también una prueba alternativa del Teorema 1 proporcionada por Lin en [25], más simple que la que llevan a cabo Caffarelli, Kohn y Nirenberg en [5]. Dicha demostración se basa principalmente en las estimaciones proporcionadas por Sohr y Wahl en [36]; haciendo uso de ellas, consigue probar el teorema utilizando únicamente la Proposición 3.

Por otro lado, en [5] también se prueba bajo una serie de condiciones un teorema de existencia de soluciones "admisibles" para la ecuación de Navier-Stokes. Esta demostración se basa en un un esquema de aproximación en tiempo en el que cada etapa utiliza la solución en un tiempo anterior.

Para este esquema se prueba gracias a determinadas estimaciones que se puede pasar al límite y que dicho límite es una solución "admisible".

En nuestro trabajo, presentamos una nueva prueba de este último teorema de existencia. En ella se usa un esquema de aproximación de tipo Galerkin. Esta nueva demostración aporta un argumento que puede ser muy útil desde el punto de vista del análisis numérico y de la computación de la solución. Así, utilizamos el siguiente esquema numérico:

$$\begin{cases} \dfrac{u^{m+1}-u^m}{\tau} + (u^m \cdot \nabla)u^{m+1} - \Delta u^{m+1} + \nabla p^{m+1} = f^{m+1}\ x \in \Omega, \\ \nabla \cdot u^{m+1} = 0\ x \in \Omega,\ \displaystyle\int_\Omega p^{m+1} = 0, \\ u^{m+1} = 0,\ x \in \partial\Omega, \end{cases} \tag{9}$$

donde N es un número natural suficientemente grande, $\tau := T/N$, $t^m = m\tau$ y

$$f^m := \frac{1}{\tau}\int_{t^{m-1}}^{t^m} f(x,t)dt,$$

$u^m \simeq u(\cdot, t^m)$ y $p^m \simeq p(\cdot, t^m)$, con $u^0 = u_0$ para $m = 0, 1, ..., N-1$.

Por un razonamiento de inducción demostramos que (9) está bien definido, es decir, existe una única solución (u^{m+1}, p^{m+1}) para cada m. Y con ayuda de este esquema construímos una sucesión de la cual podemos extraer una subsucesión que converge a una solución "admisible" de Navier-Stokes, como vemos en el resultado siguiente:

Teorema 4. *Denotemos:*

- $u_N : [0,T] \mapsto V$, *como la única función continua lineal a trozos que satisface:*

$$u_N(t^m) = u^m,\ \textit{para}\ m = 0, 1, \ldots, N.$$

- $u_N^*[0,T] \mapsto V$, *como la función constante a trozos que satisface:*

$$u_N^*(t) = u^{m+1},\ \textit{en}\ t \in (t^m, t^{m+1}],\ m = 0, 1, \ldots, N-1.$$

De manera similar introducimos la aproximación de la presión p_N^* y del segundo miembro f_N^*, de nuevo como funciones constantes a trozos. Entonces dichas funciones verifican:

$$\begin{cases} u_{N,t} + (u_N^*(t-\tau)\cdot\nabla)u_N^* - \Delta u_N^* + \nabla p_N^* = f_N^*, \\ \nabla \cdot u_N^* = 0. \end{cases} \tag{10}$$

para cualquier N y en casi todo $t \in (0,T)$. Además, existe una subsucesión de funciones u_N^* que converge débilmente en $L^2(0,T;V)$, débil- en $L^\infty(0,T;H)$ y fuertemente en $L^2(Q)^3$ hacia una solución "admisible" de la ecuación de Navier-Stokes.*

Aquí, hemos usado la notación clásica:

$$V := \{v \in H_0^1(\Omega)^3 :\ \nabla \cdot v = 0 \text{ en } \Omega\},$$

$$H := \{v \in L^2(\Omega)^3 : \ \nabla \cdot v = 0 \text{ en } \Omega, \ v \cdot \overrightarrow{n} = 0 \text{ sobre } \partial\Omega\}.$$

Este es uno de los principales resultados del trabajo y, como hemos dicho, consigue dar una nueva prueba del resultado de existencia de Caffarelli, Kohn y Nirenberg. La demostración se basa principalmente en las estimaciones clásicas de energía y utiliza, también, las estimaciones proporcionadas por Sohr y Wahl en [36] y una versión discreta de la desigualdad Young. Es importante recalcar que el argumento necesita una buena estimación para la presión. De hecho, no se conoce mejora del resultado debido principalmente a que la mejor estimación que se tiene para la presión está en $L^{5/3}$ que da la clave para demostrar el Teorema 1. Para acotar la presión en nuestro trabajo usamos una versión discreta de la desigualdad de Young, de tipo convolución. De esto modo, trasladamos esta buena estimación de la presión a nuestro esquema numérico.

Por otra parte, en el trabajo también se proporciona un esquema de Galerkin totalmente discreto, tanto en tiempo como en espacio, para el cual somos capaces de probar resultados análogos a los ya expuestos para el esquema (9). El segundo esquema considerado es

$$\begin{cases} \left(\frac{u_h^{m+1}-u_h^m}{\tau}, v_h\right) + \left((u_h^m \cdot \nabla) u_h^{m+1}, v_h\right) \\ \qquad + \left(\nabla u_h^{m+1}, \nabla v_h\right) + \left(\nabla p_h^{m+1}, v_h\right) = \left(f_h^{m+1}, v_h\right) \quad \forall v_h \in X_h, \\ \left(q_h, \nabla \cdot u_h^{m+1}\right) = 0 \ \forall q_h \in P_h, \\ (u_h^{m+1}, p_h^{m+1}) \in (X_h, P_h), \end{cases} \tag{11}$$

para $m = 0, 1, ..., N_1$ y donde los (X_h, P_h) son espacios uniformemente compatibles, es decir, espacios para los que se verifica la condición *inf-sup* uniforme

$$\inf_{q_h \in P_h \setminus \{0\}} \sup_{v_h \in X_h \setminus \{0\}} \frac{(\nabla q_h, v_h)}{\|v_h\|_{H^{1-s}} \|q_h\|_{H^s}} \geq c \quad \forall h \in [0, 1]. \tag{12}$$

El trabajo termina indicando, que como ya hemos mencionado, igual que podemos extender el Teorema 4 al caso completamente discreto, también podemos hacerlo para otros esquemas numéricos con estimaciones de energía análogas.

También podemos considerar otros sistemas, tales como el sistema de Boussinesq

$$\begin{cases} u_t - \Delta u + (u \cdot \nabla) u + \nabla p = f + \theta k, \ (x, t) \in Q, \\ \nabla u = 0, \ (x, t) \in Q, \\ \theta_t + u \cdot \nabla \theta - \Delta \theta = g, \ (x, t) \in Q, \\ u(x, t) = 0, \ \theta(x, t) = 0, \ (x, t) \in \Sigma, \\ u(x, 0) = u_0(x), \ \theta(x, 0) = \theta_0(x), \ x \in \Omega \end{cases} \tag{13}$$

o la ecuación de Navier-Stokes con densidad variable

$$\begin{cases} \rho_t + \nabla \cdot (\rho u) = 0, \ (x, t) \in Q, \\ \rho(u_t + (u \cdot \nabla) u) - \Delta u + \nabla p = \rho f, \ (x, t) \in Q, \\ \nabla \cdot u = 0, \ (x, t) \in Q, \\ u(x, t) = 0, \ (x, t) \in \Sigma, \\ u(x, 0) = u_0(x), \ \rho(x, 0) = \rho_0(x), \ x \in \Omega. \end{cases} \tag{14}$$

En el caso del sistema de Boussinesq, podemos probar resultados análogos usando un esquema de Galerkin semidiscreto o totalmente discreto. Por ejemplo,

$$\begin{cases} \dfrac{u^{m+1}-u^m}{\tau} - \Delta u^{m+1} + (u^m \cdot \nabla)u^{m+1} + \nabla p^{m+1} = f^{m+1} + \theta^{m+1}k, \\ \nabla u^{m+1} = 0, \\ \dfrac{\theta^{m+1}-\theta^m}{\tau} + u^m \cdot \nabla \theta^{m+1} - \Delta \theta^{m+1} = g^{m+1}. \end{cases} \tag{15}$$

En cambio, para la ecuación de Navier-Stokes con densidad variable el estudio no es tan sencillo de extender debido principalmente a la falta de regularidad de la densidad de masa ρ. De hecho, para este sistema no está clara la definición de solución "admisible" ya que obtener una desigualdad local de energía análoga a (5) parece díficil de nuevo, debido principalmente al término relacionado con la presión.

Relacionado con este trabajo, también existen en la literatura otros donde se dan nuevas pruebas del resultado de Caffarelli-Kohn-Nirenberg, véase [18, 39]. En el caso de Guermond, [18], se usa un esquema de tipo Galerkin distinto que conduce a la misma conclusión.

Por otro lado, en el trabajo de Da Veiga, véase [39], la prueba reposa sobre un método de regularización de cuarto orden.

Resultados teóricos y numéricos para problemas de control óptimo bi-objetivo

Theoretical and numerical results for some bi-objective optimal control problems

La segunda parte de esta memoria está formada por tres trabajos relacionados con la Teoría de Control Óptimo. En el primero de ellos, estudiamos los equilibrios de Pareto asociados a diversos problemas de control bi-objetivo. Ha dado lugar al artículo [12].

Más precisamente, nos centramos en problemas bi-objetivos en los que tenemos dos funcionales a minimizar de la forma:

$$J_i(f) := \frac{a}{2}\int_{\mathcal{O}_i} |u - u_{id}|^2 + \frac{\mu}{2}\int_{\omega} |f|^2, \quad i = 1, 2, \tag{16}$$

donde las u_{id} son funciones dadas y μ y a dos constantes positivas. A lo largo de este estudio analizaremos el caso en el que las variables u y f están ligadas por una EDP de Poisson

$$\begin{cases} -\Delta u = f1_\omega, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \end{cases} \tag{17}$$

una EDP elíptica semi-lineal

$$\begin{cases} -\Delta u + \phi(u) = f1_\omega, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \end{cases} \tag{18}$$

donde ϕ es una función a la que le imponemos las siguientes condiciones (que garantizan la existencia de solución)

$$\begin{cases} \phi : \mathbb{R} \mapsto \mathbb{R} \text{ is continuously differentiable,} \\ \phi'(s) \geq 0 \text{ and } |\phi(s)| \leq C + C|s| \quad \forall s \in \mathbb{R} \end{cases} \tag{19}$$

y, por último, las EDPs estacionarias de Navier-Stokes

$$\begin{cases} -\nu\Delta u + (u \cdot \nabla)u + \nabla p = f1_\omega, \quad x \in \Omega, \\ \nabla \cdot u = 0, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega. \end{cases} \tag{20}$$

Como hemos mencionado anteriormente, el análisis multi-objetivo el problema varía en función de la definición de equilibrio que pretendamos estudiar. En este primer caso, vamos a trabajar con equilibrios de Pareto.

Diremos que $\hat{f}$ es un equilibrio de Pareto si no existe ningún otro control f que verifique

$$J_1(f) \leq J_1(\hat{f}) \quad \text{y} \quad J_2(f) \leq J_2(\hat{f}), \tag{21}$$

siendo alguna de las desigualdades estrictas. Dicho de otro modo $\hat{f}$ es un equilibrio de Pareto si no existe ningún otro control f que mejore simultáneamente los dos criterios, siendo esta mejora estricta al menos en un caso.

Es importante recalcar que si $\hat{f}$ es un equilibrio de Pareto para dos funcionales J_1 y J_2 (por ejemplo) de clase $\mathcal{C}^1$, entonces existe un número $\alpha \in [0, 1]$ tal que se verifica:

$$\alpha J_1'(\hat{f}) + (1 - \alpha)J_2'(\hat{f}) = 0.$$

Por lo tanto, si J_1 y J_2 son convexas, como ocurre cuando el estado está dado por (17), tenemos el siguiente resultado:

Teorema 5. *Sean J_1 y J_2 los funcionales dados en* (16) *y sea $\hat{f}$ un equilibrio de Pareto para* (16)*,* (17)*. Entonces existe $\alpha \in [0, 1]$ tal que $\hat{f}$ es un mínimo del funcional:*

$$J_{(\alpha)} := \alpha J_1 + (1 - \alpha)J_2.$$

Por otra parte, hemos estudiado bajo qué condiciones un mínimo del funcional $J_{(\alpha)}$ es un equilibrio de Pareto. Para este estudio ha sido importante reescribir un mínimo del funcional de $J_{(\alpha)}$ como la solución de un sistema de optimalidad para el control $\hat{f}$, el estado asociado y un estado adjunto apropiado.

Así, diremos que f es un quasi-equilibrio de Pareto si es solución del sistema de optimalidad junto con la variable estado u, el estado adjunto φ. Veamos ahora el sistema de optimalidad para cada caso.

- Quasi-equilibrio de Pareto para la EDP de Poisson. En este caso, f debe verificar, junto con u y φ, el sistema de optimalidad:

$$\begin{cases} -\Delta u = f1_\omega, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \\ -\Delta\varphi = \alpha(u - u_{1d})1_{\mathcal{O}_1} + (1-\alpha)(u - u_{2d})1_{\mathcal{O}_2}, \quad x \in \Omega, \\ \varphi = 0, \quad x \in \partial\Omega, \\ f = -\dfrac{a}{\mu}\,\varphi|_\omega\,. \end{cases} \tag{22}$$

- Quasi-equilibrio de Pareto para la EDP elíptica semi-lineal. Ahora f, u y φ verifican el sistema de optimalidad:

$$\begin{cases} -\Delta u + \phi(u) = f1_\omega, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \\ -\Delta\varphi + \phi'(u)\varphi = \alpha(u - u_{1d})1_{\mathcal{O}_1} + (1-\alpha)(u - u_{2d})1_{\mathcal{O}_2}, \quad x \in \Omega, \\ \varphi = 0, \quad x \in \partial\Omega, \\ f = -\dfrac{a}{\mu}\,\varphi|\,\omega. \end{cases} \tag{23}$$

- Quasi-equilibrio de Pareto para la EDP de Navier-Stokes. Tenemos que (f, u, p, φ, q) es solución de

$$\begin{cases} -\nu\Delta u + (u\cdot\nabla)u + \nabla p = f|_\omega\,, \quad x \in \Omega, \\ \nabla\cdot u = 0, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \\ -\nu\Delta\varphi + (u\cdot\nabla)\varphi + (\nabla u)^t\varphi + \nabla q \\ \qquad = \alpha(u - u_{1d})1_{\mathcal{O}_1} + (1-\alpha)(u - u_{2d})1_{\mathcal{O}_2}, \quad x \in \Omega, \\ \nabla\cdot\varphi = 0, \quad x \in \Omega, \\ \varphi = 0, \quad x \in \partial\Omega, \\ f = -\dfrac{a}{\mu}\varphi 1_\omega. \end{cases} \tag{24}$$

Una vez establecidas las definiciones de equilibrio y quasi-equilibrio de Pareto, analizamos y deducimos resultados de existencia, unicidad y equivalencia entre ambos conceptos; llegamos a la conclusión de que, en el caso lineal, ambos conceptos son equivalentes y podemos garantizar existencia y unicidad.

En el caso semi-lineal, ocurre que la existencia se puede garantizar sin imponer nuevas condiciones pero para conseguir la unicidad para cada α y establecer la equivalencia, necesitamos imponer nuevas condiciones sobre la función ϕ y sus derivadas primera y segunda. En el estudio del caso semi-lineal ya podemos observar que la complejidad del problema aumenta a medida que aumenta el tamaño de a/μ.

Por último, en el caso de las EDPs de Navier-Stokes, los resultados son más complejos debido principalmente a que trabajamos con un sistema fuertemente no lineal y, además, no está garantizado que podamos asociar a cada control un único estado. En cuanto a los objetivos logrados, indiquemos que, en el contexto de Navier-Stokes, hemos conseguido demostrar la existencia de equilibrios de Pareto y, también, que un equilibrio de Pareto es un quasi-equilibrio:

Teorema 6. *Sea $\hat{f}$ un equilibrio de Pareto para* (16)*,* (20)*. Entonces $\hat{f}$ es un quasi-equilibrio de Pareto, es decir, es solución de* (24) *junto con las variables asociadas.*

Para las demostraciones de todos estos resultados hemos usado resultados clásicos del Cálculo de Variaciones, aprovechando las propiedades de los funcionales: coercitividad, convexidad, semicontinuidad inferior débil, etc. En el caso de las EDPs de Navier-Stokes, para probar el Teorema 6 necesitamos usar el formalismo de Dubovitsky-Milyoutin y, en particular, el siguiente resultado técnico:

Lema 7. *Sean $K_1, \dots, K_n$ conos convexos en el espacio de Banach X. Si para cada i asumimos que K_i es abierto o bien un subespacio cerrado, las siguientes condiciones son equivalentes:*

- $\bigcap_{i=1}^{n} K_i = \emptyset$.
- *Existen formas lineales $f_i \in K_i^*$, con $i = 1, \dots, n$, no todos cero, tales que*

$$\sum_{i=1}^{n} f_i = 0.$$

Las ideas básicas del formalismo se pueden explicar de la siguiente manera. En un mínimo local, el cono de direcciones de descenso asociado a la función de coste debe ser disjunto de la intersección de los conos de direcciones factibles y tangentes, respectivamente determinadas por la familia de controles admisibles y la ecuación, en este caso (20). De hecho, no podemos "movernos" del mínimo a otro punto admisible en una dirección que mejore las funciones objetivo. Como consecuencia del Teorema de Hahn-Banach y de algunos argumentos adicionales, se deduce que deben existir elementos en los conos duales asociados, no todos ellos cero, que sumen cero. Esta condición algebraica es el sistema de Euler-Lagrange del problema extremal que nos ocupa. Cuando es posible identificar los conos primarios y duales anteriores, este sistema proporciona las condiciones de optimalidad de primer orden de forma sistemática. En el caso de un problema de control óptimo estándar (mono-objetivo), también conduce al Principio del Máximo de Pontryagin correspondiente.

Terminamos el estudio de cada ecuación añadiendo una serie de algoritmos que aproximan los quasi-equilibrios. Aparte de los clásicos algoritmos de punto fijo, gradiente con paso óptimo y gradiente conjugado, para la ecuación de Navier-Stokes incluimos, también, un método de tipo Newton. Este algoritmo iterativo se basa en producir aproximaciones de una solución del sistema de optimalidad, viéndolo como una ecuación a la que le buscamos su cero.

Por otro lado, en el caso de Navier-Stokes, incluimos también simulaciones numéricas como vemos a continuación. Éstas están realizadas con el software *FreeFem++*. Como método de aproximación numérica usamos el método de elementos finitos mixtos, es decir, aproximamos al mismo tiempo las velocidades u y la presión p en cada iteración.

Presentaremos ahora brevemente una de las experiencias numéricas detalladas en este trabajo.

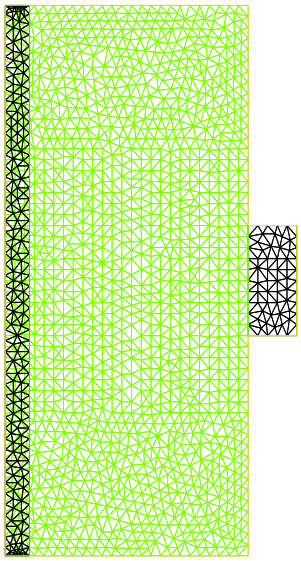


Figura 1: Dominio escogido. Número de nodos: 1519 . Número de triángulos: 2876.

Hemos considerado un dominio formado por dos rectángulos unidos en un lateral. Suponemos que el control se va a aplicar en una banda lateral, de manera que haciendo cada vez más estrecha la banda, podamos aproximarnos a resolver un problema de control frontera.

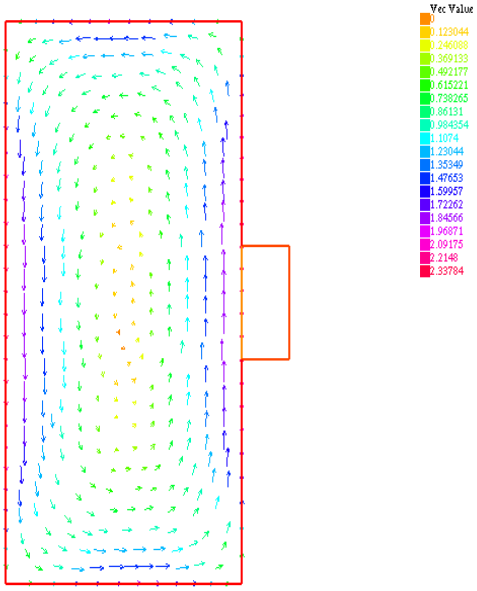
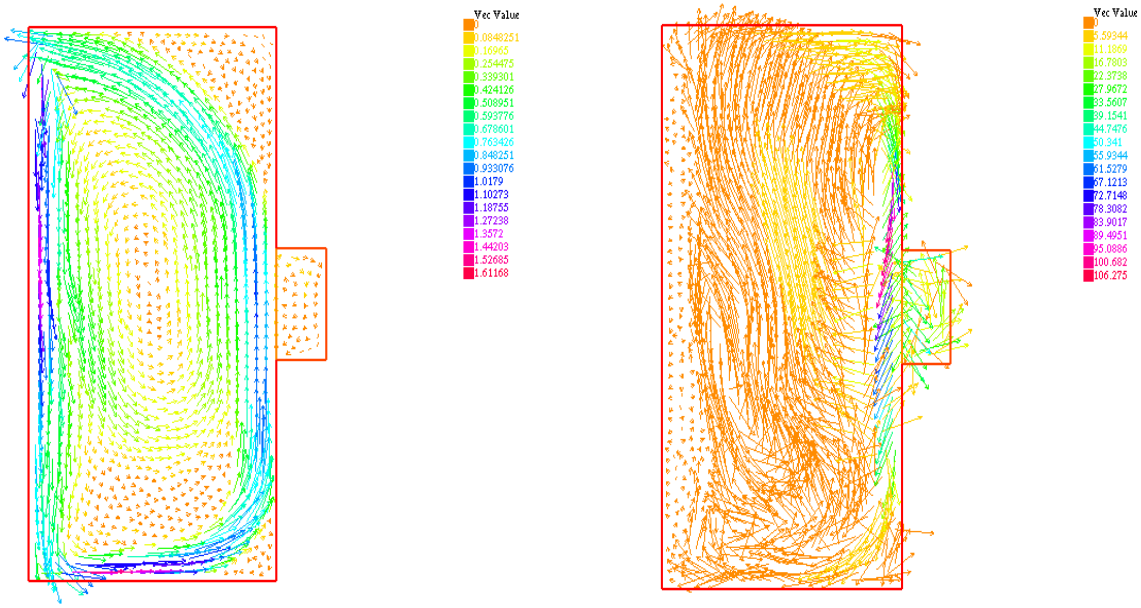


Figura 2: Las funciones deseadas.

El problema consiste en hallar un control $\hat{f}$ de manera que el fluido rote en el rectángulo grande y permanezca en reposo en el rectángulo pequeño. En este momento, podemos observar cómo es

muy importante el valor que le demos al parámetro α; éste va a determinar en gran manera el peso que damos a cada solución deseada en cada rectángulo. Así, si hacemos α muy próximo a 0, priorizamos que la solución permanezca en reposo por encima de que rote; por el contrario, si α es próximo a 1, estaremos buscando que la solución rote en el rectángulo grande aunque esto ocasione movimientos en el pequeño.

Así, gracias al método de Newton implementado, obtenemos el estado y el estado adjunto que podemos ver en la Figura 3.



(a) Velocity (b) Adjount State

Figura 3: Velocidad final y estado adjunto final para un fluido con número de Reynolds de 3500 y $\alpha = 0{,}5$.

Como punto final a este trabajo nos gustaría hacer algunos comentarios con respecto a futuras líneas de investigación. En este sentido, queda aún saber bajo qué condiciones que no impliquen unicidad de solución se tiene que un quasi-equilibrio de Pareto es un equilibrio.

Otro aspecto importante que queremos tratar en trabajos futuros es la adaptación del análisis y los resultados a las EDPs de Navier-Stokes de evolución y otras ecuaciones y sistemas evolutivos.

Por último, debemos indicar que el estudio se podría extender al caso de tres o más funcionales, funcionales de otro tipo (no cuadráticos y/o con estructuras diferentes), problemas de control frontera, etc.

Teoría y análisis numérico de los problemas de control óptimo bi-objetivo: equilibrios de Nash

Theoretical and numerical bi-objective optimal control: Nash equilibria

El tercer trabajo de esta memoria está relacionado también con la Teoría de Control Óptimo y ha dado lugar al artículo [11]. En este caso, volvemos a estudiar un problema de control multi-objetivo pero ahora desde el punto de vista de los equilibrios de Nash.

Supondremos ahora que la ecuación de estado considerada está siendo controlada por dos controles (este número va coincidir con el número de funcionales de coste). Diremos que $(\widehat{f_1}, \widehat{f_2})$ es un equilibrio de Nash si se verifica

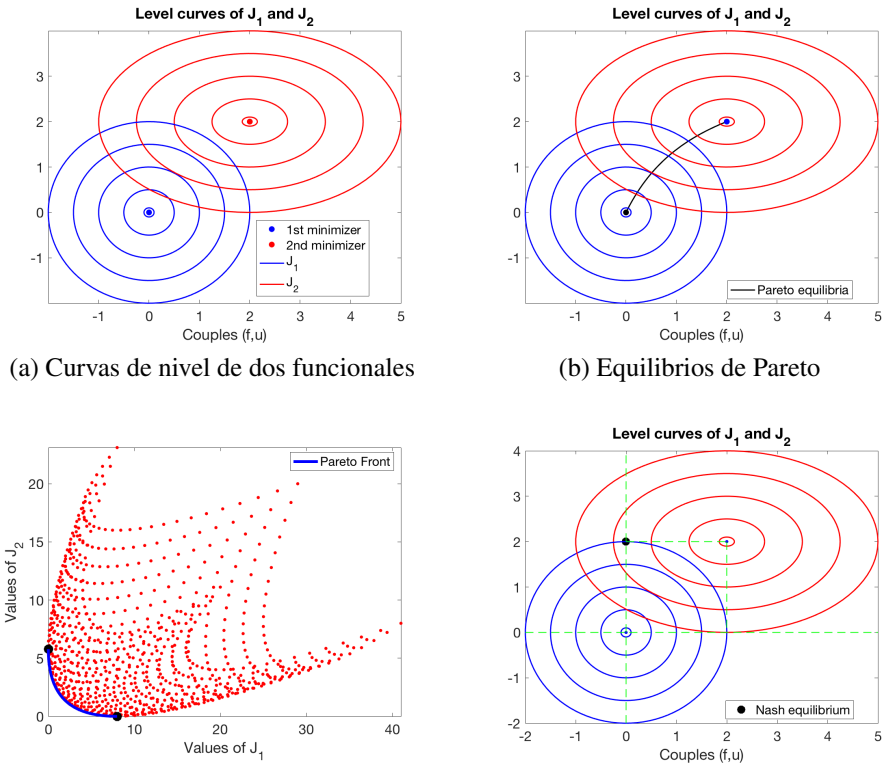
$$J_1(\widehat{f_1}, \widehat{f_2}) \leq J_1(f_1, \widehat{f_2}) \quad \text{y} \quad J_2(\widehat{f_1}, \widehat{f_2}) \leq J_2(\widehat{f_1}, f_2) \tag{25}$$

para cualquier par de controles admisibles (f_1, f_2).

En nuestro caso, tendremos

$$J_i(f_1, f_2) := \frac{a}{2} \int_{\mathcal{O}_i} |u - u_{id}|^2 + \frac{\mu}{2} \int_{\omega_i} |f_i|^2, \quad i = 1, 2, \tag{26}$$

donde u es el estado correspondienre al par (f_1, f_2).



(a) Curvas de nivel de dos funcionales

(b) Equilibrios de Pareto

(c) Frontera de Pareto

(d) Equilibrios de Nash

Figura 4: Relación entre equilibrios de Nash y de Pareto.

Observemos que en el caso de los equilibrios de Nash cada control se dedica exclusivamente a minimizar un funcional en una dirección, mientras que en el caso de los equilibrios de Pareto, el control minimiza en cierto sentido (con un peso α) los dos funcionales al mismo tiempo (es decir, el funcional $J_{(\alpha)}$). Esta diferencia la podemos ver gráficamente en la Figura 4. Aquí, el equilibrio de Nash es único; en cambio, los equilibrios de Pareto forman una curva parametrizada por $\alpha \in [0,1]$. También hemos representado lo que en Economía se conoce como Frontera de Pareto, que no es más que la curva en el plano (J_1, J_2) formada por los valores que toman los equilibrios de Pareto.

Al igual que en el anterior trabajo, en éste vamos a estudiar los equilibrios asociados a la EDP de Poisson

$$\begin{cases} -\Delta u = f_1 1_{\omega_1} + f_2 1_{\omega_2}, & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \end{cases} \tag{27}$$

a la EDP elíptica semi-lineal

$$\begin{cases} -\Delta u + \phi(u) = f_1 1_{\omega_1} + f_2 1_{\omega_2}, & x \in \Omega, \\ u = 0, & x \in \partial\Omega \end{cases} \tag{28}$$

(donde ϕ vuelve a verificar las condiciones (19)) y el sistema estacionario de Navier-Stokes

$$\begin{cases} -\nu\Delta u + (u \cdot \nabla)u + \nabla p = f_1 1_{\omega_1} + f_2 1_{\omega_2}, & x \in \Omega, \\ \nabla \cdot u = 0, & x \in \Omega, \\ u = 0, & x \in \partial\Omega. \end{cases} \tag{29}$$

A diferencia de lo que ocurría en el caso de los equilibrios de Pareto, para los equilibrios de Nash no podemos reducir el estudio a un problema mono-objetivo. Por ello, todas las demostraciones y situaciones se complican y gana aún más relevancia el tamaño de a/μ.

Introducimos ahora los distintos sistemas de optimalidad para cada ecuación. En cada caso, la solución del sistema se denomina quasi-equilibrio de Nash.

- Quasi-equilibrio de Nash en el caso de la EDP de Poisson (27): es toda solución $(f_1, f_2, u, \varphi_1, \varphi_2)$ del sistema de optimalidad:

$$\begin{cases} -\Delta u = f_1 1_{\omega_1} + f_2 1_{\omega_2}, & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \\ -\Delta \varphi_i = (u - u_{id}) 1_{\mathcal{O}_i}, & x \in \Omega, \\ \varphi = 0, & x \in \partial\Omega, \\ f_i = -\dfrac{a}{\mu}\, \varphi_i|_{\omega_i}, & i = 1, 2. \end{cases} \tag{30}$$

- Quasi-equilibrio de Nash en el caso de la EDP elíptica semi-lineal (28): se trata de toda solución $(f_1, f_2, u, \varphi_1, \varphi_2)$ de

$$\begin{cases} -\Delta u + \phi(u) = f_1 1_{\omega_1} + f_2 1_{\omega_2}, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \\ -\Delta\varphi_i + \phi'(u)\varphi_i = (u - u_{id}) 1_{\mathcal{O}_i}, \quad x \in \Omega, \\ \varphi_i = 0, \quad x \in \partial\Omega, \\ f_i = -\dfrac{a}{\mu}\, \varphi_i|_{\omega_i} \quad i = 1, 2. \end{cases} \tag{31}$$

- Quasi-equilibrio de Nash para las EDPs de Navier-Stokes (29): se llama así a toda solución $(f_1, f_2, u, p, \varphi_1, q_1, \varphi_2, q_2)$ del sistema

$$\begin{cases} -\nu\Delta u + (u \cdot \nabla)u + \nabla p = f_1 1_{\omega_1} + f_2 1_{\omega_2}, \quad x \in \Omega, \\ \nabla \cdot u = 0, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \\ -\nu\Delta\varphi_i + (u \cdot \nabla)\varphi_i + (\nabla u)^t \varphi_i + \nabla q_i = (u - u_{id}) 1_{\mathcal{O}_i}, \quad x \in \Omega, \\ \nabla \cdot \varphi_i = 0, \quad x \in \Omega, \\ \varphi_i = 0, \quad x \in \partial\Omega, \\ f_i = -\dfrac{a}{\mu}\varphi_i 1_{\omega_i} \quad i = 1, 2. \end{cases} \tag{32}$$

En este trabajo, establecemos en cada caso condiciones bajo las cuales los conceptos de equilibrio y quasi-equilibrio de Nash son equivalentes. También, probamos varios resultados relativos a la existencia y unicidad.

De manera análoga a lo que ocurría en el caso de los equilibrios de Pareto, podemos deducir que, si un par $(\widehat{f}_1, \widehat{f}_2)$ es un equilibrio de Nash para los funcionales J_1 y J_2, entonces

$$\frac{\partial J_1}{\partial f_1}(\widehat{f}_1, \widehat{f}_2) = 0 \quad \frac{\partial J_2}{\partial f_2}(\widehat{f}_1, \widehat{f}_2) = 0.$$

Mirando estas igualdades, resulta razonable pensar que, en general, un equilibrio de Nash será, junto con el estado asociado y los correspondientes estados adjuntos, solución del sistema de optimalidad, es decir, un quasi-equilibrio. Notemos que esta afirmación no está muy clara en todos los casos.

Así, en el caso del problema (25), (26), (27), establecemos que si, a/μ es suficientemente pequeño, existe un único equilibrio. También probamos que, para cualesquiera $a > 0$ y $\mu > 0$, existe un único equilibrio en el caso en que los dominios de observación $\mathcal{O}_1$ y $\mathcal{O}_2$ son iguales. En relación con la equivalencia de los dos conceptos, vemos que, independientemente del valor de a/μ, un equilibrio de Nash es un quasi-equilibrio y viceversa.

Para (25), (26), (28), debemos imponer condiciones sobre la función ϕ, sus derivadas primera y segunda y el tamaño de a/μ para poder probar la equivalencia entre equilibrio y quasi-equilibrio y, también, para probar resultados de existencia y unicidad. Además, presentamos un resultado

relativo a la existencia de quasi-equilibrios de Nash en el que rebajamos las condiciones sobre ϕ y eliminamos la restricción sobre el tamaño de a/μ. El resultado es el siguiente:

Teorema 8. *Si existe $\delta > 0$ tal que $\phi'(s) \geq \delta$ para cualquier $s \in \mathbb{R}$ y $\phi(0) = 0$, entonces existen quasi-equilibrios de Nash.*

La prueba de este teorema reposa sobre un lema de Brézis y Nirenberg, véase [3], que establece condiciones bajo las cuales la suma de un operador monótono y otro compacto es una aplicación sobreyectiva.

Por otro lado, para las EDPs de Navier-Stokes estacionarias, conseguimos demostrar que, si el tamaño de a/μ es suficientemente pequeño entonces existen quasi-equilibrios de Nash y que, si éste es aún más pequeño, el quasi-equilibrio es único. El resultado de equivalencia es desconocido. No obstante, probamos que todo equilibrio de Nash es un quasi-equilibrio, gracias de nuevo al formalismo de Dubovitsky-Milyoutin:

Teorema 9. *Sea (f_1, f_2) un equilibrio de Nash para* (25)*,* (26)*,* (29)*. Entonces (f_1, f_2) es un quasi-equilibrio, es decir, es solución, junto con el estado (u, p) y los estados adjuntos (φ_1, q_1) y (φ_2, q_2), de* (32)*.*

Al igual que antes, terminamos el estudio de cada ecuación añadiendo una serie de algoritmos que aproximan los quasi-equilibrios. También, en el caso de las EDPs de Navier-Stokes, incluimos los resultados de varias experiencias numéricas, de nuevo realizadas con *FreeFem++*.

Para la aproximación numérica, hemos usado el método de elementos finitos mixtos ($\mathbb{P}_2 - \mathbb{P}_1$), es decir, aproximamos al mismo tiempo las velocidades y las presiones en cada iteración. En uno de los tests realizados, el dominio y el mallado son como en la Fig. 5.

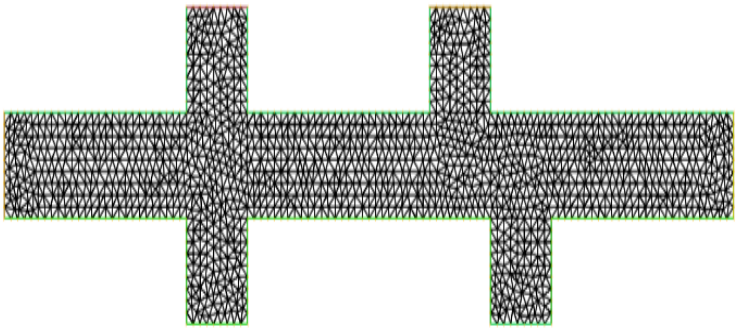


Figura 5: Dominio escogido. Número de nodos: 1519 . Número de triángulos: 2876.

En este caso los dos primeros rectángulos pequeños (tanto el superior como el inferior), se corresponden con los dominios de control. En cada uno de ellos, gracias a la acción de los controles, permitimos que entre o salga fluido. Por otro lado, buscamos que la solución gire en el rectángulo superior siguiente y permanezca en reposo en el inferior, como vemos en la Fig. 6. De nuevo podemos ver cómo la definición de equilibrio tiene pleno sentido (ya que nunca seremos capaces

de conseguir rotación en el rectángulo superior y a la vez reposo en el inferior). Al buscar equilibrios de Nash, procuraremos que el primer control se ocupe del giro mientras que el segundo se ocupe de que la solución esté lo más quieta posible en el rectángulo inferior.

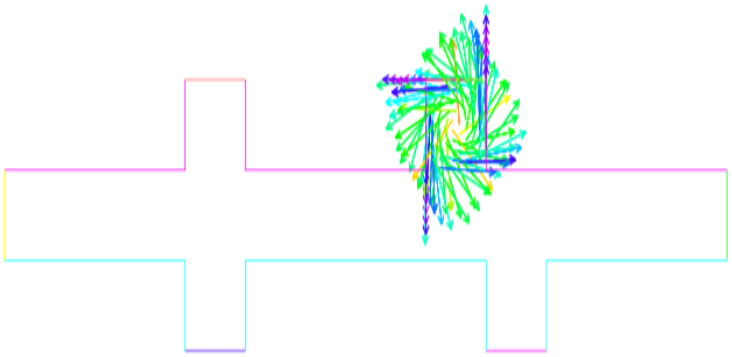


Figura 6: Las funciones deseadas.

Para esta experiencia numérica hemos considerado que el fluido entra con un perfil de tipo parabólico: $u_0 = (q_0(y - y_2)(y_3 - y)$ con $y_2 < y_3 \in \mathbb{R}$ y $q_0 = 100$.

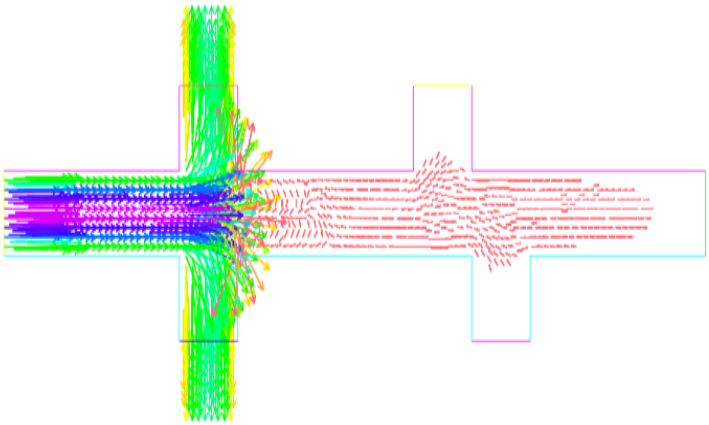


Figura 7: Perfil parabólico de entrada.

El estado calculado se puede ver en la Fig. 8.

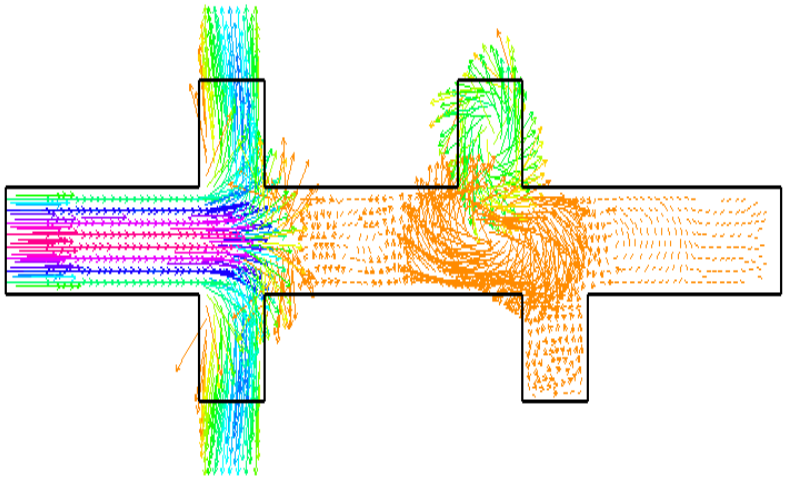


Figura 8: Velocidad final para un fluido con número de Reynolds de 1625, aproximadamente y $a = 1{,}99$.

Vemos que, en general, es más complicado calcular equilibrios de Nash que equilibrios de Pareto. Esto tiene pleno sentido ya que, al fin y al cabo, en los equilibrios de Pareto, hemos conseguido reducir la cuestión a un problema mono-objetivo (gracias al funcional $J_{(\alpha)}$). Contrariamente, en los equilibrios de Nash tenemos forzosamente dos funcionales a minimizar. Esto provoca también que para equilibrios de Nash para (25), (26), (29), no podamos aumentar tanto el número de Reynolds sobre todo en el caso en que a/μ es grande, como ocurre en la Fig. 8.

Aún hay preguntas abiertas en este trabajo, como son: saber bajo qué condiciones que no impliquen unicidad de solución para Navier-Stokes se tiene que un quasi-equilibrio de Nash es un equilibrio. También queda por demostrar un resultado relativo a la existencia de los equilibrios de Nash ya que en esta ocasión nos enfrentamos a estudiar aplicaciones multivaluadas y no queda claro cómo usar resultados clásicos, similares al Teorema de Kakutani.

Otro aspecto importante que estamos tratando para trabajos futuros es el de adaptar el análisis y los resultados a las EDPs de Navier-Stokes de evolución y a otras ecuaciones y sistemas evolutivos.

Por último, observemos que, como en el capítulo precedente, puede ser interesante extender los argumentos, técnicas y resultados a muchos otros problemas de control multi-objetivo.

Análisis y resolución numérica de algunos problemas de control de tiempo mínimo

Analysis and numerical solution of some minimal time control problems

Este es el tercer trabajo relativo a la Teoría de Control Óptimo y ha dado lugar al artículo [8]. En este nuevo capítulo estudiamos problemas de control de tiempo mínimo. Como ya hemos mencionado, se trata de problemas en los que la variable tiempo es uno de los argumentos del funcional a minimizar, es decir, juega el papel de un control adicional. Como en los dos trabajos anteriores,

vamos a analizar problemas asociados a diversas ecuaciones, en esta ocasión una EDO lineal, una EDO no lineal y la EDP del calor.

Así, el primer problema al que nos enfrentamos es el siguiente:

$$\begin{cases} \text{Minimizar } \phi(T,h) := \dfrac{T^2}{2} + \dfrac{b}{2}\displaystyle\int_0^{+\infty} |h|^2 \, dt, \\ \text{Sujeto a: } (T,h) \in \mathbb{R}_+ \times L^2(0,+\infty), \\ \qquad\qquad (y,h) \text{ resuelve (34)}, \\ \qquad\qquad |y(T) - y_d| = \delta, \end{cases} \tag{33}$$

con $b \geq 0$, $\delta > 0$ e $y_d \in \mathbb{R}$ dada y un estado $y = y(t)$ dado por

$$\begin{cases} y_t + ay = h(t), \quad t \in (0,T), \\ y(0) = y_0, \end{cases} \tag{34}$$

con $a > 0$ e $y_0 \in \mathbb{R}$ dada.

Somos capaces de demostrar que existe una única solución del problema (33), (34) por argumentos clásicos de convexidad. Además, demostramos que dicha solución (T,h) satisface, junto con el estado asociado y, un estado adjunto ψ y una constante $\lambda > 0$, el sistema de optimalidad

$$\begin{cases} y_t + ay = h, \quad t \in (0,T), \quad y(0) = y_0, \\ -\psi_t + a\psi = 0, \quad t \in (0,T), \quad \psi(T) = y(T) - y_d, \\ |y(T) - y_d| = \delta, \\ h = -\dfrac{1}{\lambda b}\psi \quad \text{in } (0,T), \\ T = -\dfrac{1}{\lambda}(y(T) - y_d) y_t(T). \end{cases} \tag{35}$$

Para la demostración, volvemos a hacer uso del formalismo de Dubovitsky-Milyoutin. Además, conseguimos probar que, para cualesquiera $y_0, y_d \in \mathbb{R}$ y para cualquier $a > 0$, el sistema (35) posee una única solución.

El segundo caso que nos ocupa se corresponde con una EDO no lineal. Así, el problema considerado es el siguiente:

$$\begin{cases} \text{Minimizar } \phi(T,h) := \dfrac{T^2}{2} + \dfrac{b}{2}\displaystyle\int_0^{+\infty} |h|^2 \, dt, \\ \text{Sujeto a: } (T,h) \in \mathbb{R}_+ \times L^2(0,+\infty), \\ \qquad\qquad (y,h) \text{ resuelve (37)}, \\ \qquad\qquad |y(T) - y_d| = \delta, \end{cases} \tag{36}$$

siendo $b \geq 0$, $\delta > 0$ e $y_d \in \mathbb{R}$ dada y estando dado en este caso el estado $y = y(t)$ por

$$\begin{cases} y_t + H(y) = h(t), \quad t \in (0,T), \\ y(0) = y_0, \end{cases} \tag{37}$$

donde $y_0 \in \mathbb{R}$ y suponemos que H verifica

$$\begin{cases} H : \mathbb{R} \mapsto \mathbb{R} \text{ es de clase } \mathcal{C}^1, \\ 0 \leq H'(s) \leq C \quad \forall s \in \mathbb{R}. \end{cases} \tag{38}$$

Para este problema volvemos a demostrar un resultado de existencia de solución (T, h). Se prueba además que dicha solución verifica junto con el estado y, el estado adjunto asociado ψ y una constante $\lambda > 0$, el sistema:

$$\begin{cases} y_t + H(y) = h, \quad t \in (0, T), \\ y(0) = y_0, \\ -\psi_t + H'(y)\psi = 0, \quad t \in (0, T), \\ \psi(T) = y(T) - y_d, \\ |y(T) - y_d| = \delta, \\ h = -\dfrac{1}{\lambda b}\psi \quad \text{in } (0, T), \\ T = -\dfrac{1}{\lambda}(y(T) - y_d)y_t(T). \end{cases} \tag{39}$$

De nuevo, la prueba de esta implicación se basa en el formalismo de Dubovitsky-Milyoutin. En este caso, el sistema (39) ha adquirido complejidad, debido principalmente al fuerte acoplamiento de las ecuaciones. Una consecuencia importante es que no somos capaces de probar la unicidad de solución.

El tercer y último problema estudiado está relacionado con la EDP del calor:

$$\begin{cases} \text{Minimizar } \phi(T, h) := \dfrac{T^2}{2} + \dfrac{b}{2} \displaystyle\iint_{\omega \times (0, +\infty)} |h|^2 \, dx \, dt, \\ \text{Sujeto a: } (T, h) \in \mathbb{R}_+ \times L^2(\omega \times (0, +\infty)), \\ \qquad\qquad (\theta, h) \text{ resuelve (41)}, \\ \qquad\qquad \|\theta(T) - \theta_d\| = \delta, \end{cases} \tag{40}$$

donde $b > 0$, $\delta > 0$ y $\theta_d \in L^2(\Omega)$ está dado y el estado $\theta = \theta(x, t)$ es la solución de

$$\begin{cases} \theta_t - \Delta\theta = h1_\omega, \quad (x, t) \in Q_T := \Omega \times (0, T), \\ \theta = 0, \quad (x, t) \in \Sigma_T := \partial\Omega \times (0, T), \\ \theta(0) = \theta_0, \end{cases} \tag{41}$$

donde $\theta_0 \in L^2(\Omega)$ es un dato inicial dado.

Al igual que en los casos anteriores, probamos que existe solución (T, h) del problema (40) y que ésta verifica, junto con el estado θ, el estado adjunto asociado ψ y una constante positiva λ, el sistema de optimalidad

$$\begin{cases} \theta_t - \Delta\theta = h1_\omega, \quad (x,t) \in Q_T, \\ \theta = 0, \quad (x,t) \in \Sigma_T, \quad \theta(0) = \theta_0, \\ -\psi_t - \Delta\psi = 0, \quad (x,t) \in Q_T, \\ \psi = 0, \quad (x,t) \in \Sigma, \quad \psi(T) = \theta(T) - \theta_d, \\ \|\theta(T) - \theta_d\| = \delta \\ h = -\dfrac{1}{\lambda b}\, \psi|_{\omega\times(0,T)}\,, \\ T = -\dfrac{1}{\lambda}\Big((\theta(T) - \theta_d), \theta_t(T)\Big). \end{cases} \tag{42}$$

Una vez más, la prueba utiliza el formalismo de Dubovitsky-Milyoutin. El estado adjunto ψ y la constante λ aparecen como multiplicadores ligados, respectivamente, a la segunda y tercera restricción de (40).

Como ya ocurría en el caso de una EDO no lineal, en esta ocasión el acoplamiento del sistema (42) hace difícil probar un resultado de unicidad, una cuestión que queda abierta.

Desde el punto de vista práctico, lo más novedoso de este trabajo es la parte final de cada sección, en la que se incluyen algoritmos de resolución y se añaden algunas experiencias numéricas para cada una de las ecuaciones.

Para las ecuaciones (41) y (40) incluímos, entre otros, un algoritmo de penalización y probamos varios resultados relacionados con la convergencia del método.

A grandes rasgos podemos decir que el método de penalización se basa en minimizar funcionales análogos al original, donde las restricciones del problema de partida se incluyen multiplicadas por un parámetro grande:

$$\tilde{\phi}(T, h; \mu^n) := \frac{T^2}{2} + \frac{b}{2} \iint_{\omega\times(0,+\infty)} |h|^2 + \frac{1}{2\mu^n}\Big((\|\overline{\theta}(T) - \theta_d\| - \delta)^2 + T_-^2\Big), \tag{43}$$

donde μ^n es el parámetro de penalización elegido en la n-ésima iteración y $\mu^n \to 0^+$.

Este nuevo funcional se minimiza usando métodos clásicos, como el Gradiente o el Gradiente Conjugado con paso óptimo. Y los resultados de convergencia que probamos van encaminados a demostrar que, efectivamente, una solución de este problema es solución de (40) para μ suficientemente pequeño.

A la hora de la programación con *FreeFem++*, hemos implementado un método numérico basado en evaluar en cada T y cada valor de $\mu = 1/\lambda$ las expresiones

$$\|\theta^k(T^k) - \theta_d\| - \delta \quad \text{y} \quad -\mu^j(\theta^k(T) - \theta_d, \theta_t^k(T^k)). \tag{44}$$

donde θ^k verifica junto con un control h^k y un estado adjunto ψ^k, las igualdades 1 a 4 y 6 de (42).

Una vez calculados estos valores, resolvemos mediante un método de Newton las ecuaciones

$$T = -\mu(\theta(T) - \theta_d, \theta_t(T)), \quad \|\theta(T) - \theta_d\| = \delta$$

y, para los valores calculados, hallamos el estado θ, el estado adjunto ψ y el control h asociados.

Ahora vamos a incluir una experiencia numérica que hemos implementado con este método.

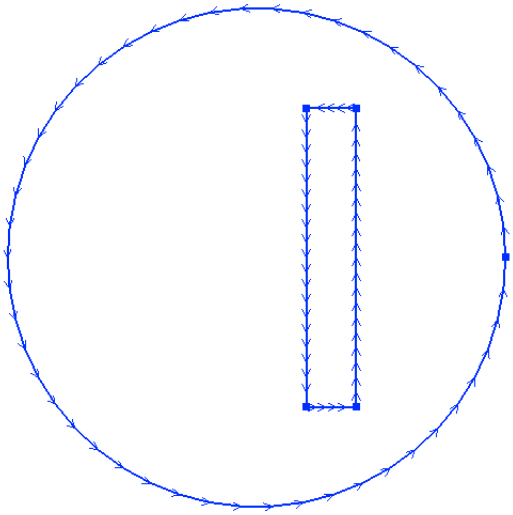


Figura 9: Dominio espacial.

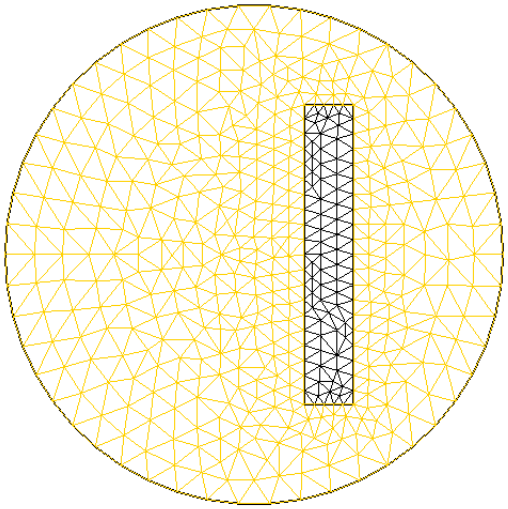


Figura 10: Mallado.

Para esta simulación hemos elegido un dominio circular con una banda en su interior que se corresponde con el dominio de control, véase Figura 9.

Nuestro problema consiste en calcular el control h que lleve la ecuación (41) a un estado deseado en el menor tiempo T^* posible. El estado deseado que hemos elegido es el correspondiente a la solución de (41) con $h \equiv 0$ y $T = 20$.

En la Fig. 11 representamos los dos miembros de la primera ecuación de (44), donde μ está calculada de manera que se tenga la segunda igualdad.

Para mayor claridad, en la Fig. 4.12 quedan ilustrados los valores de $-\mu(\theta(T) - \theta_d, \theta_t(T))$ en función de μ y T.

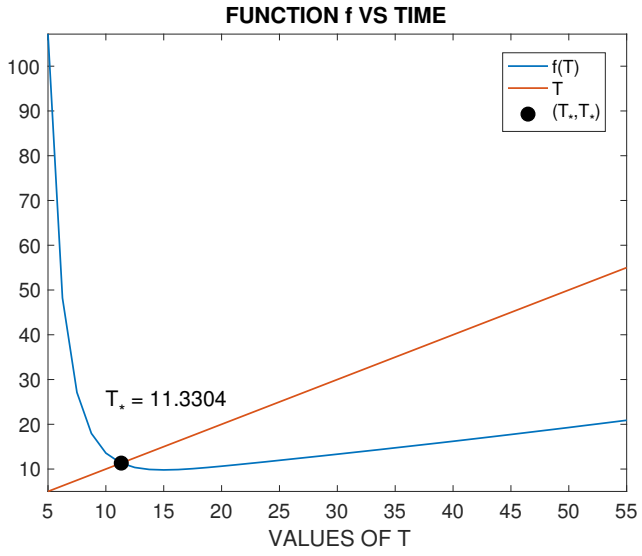


Figura 11: Tiempo mínimo T_*.

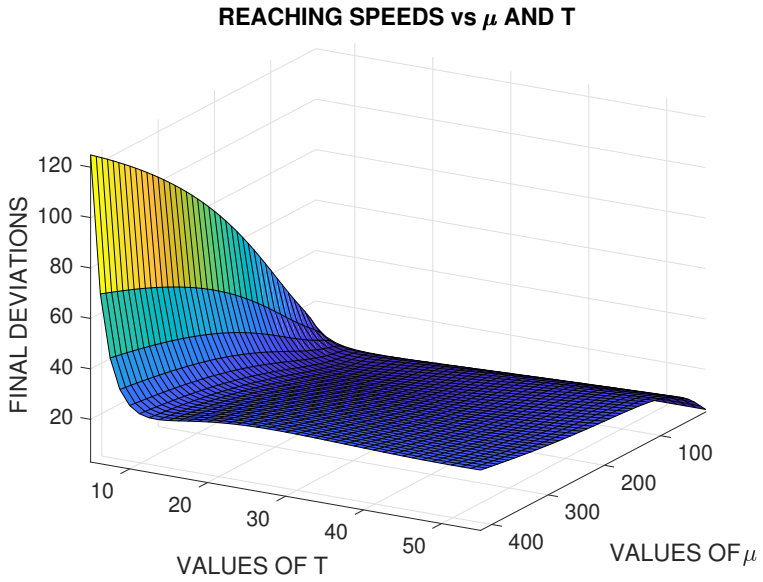


Figura 12: Valores de $-\mu(\theta(T) - \theta_d)\theta_t(T)$ frente a μ y a T.

Como líneas futuras de este trabajo nos planteamos implementar y analizar la convergencia de un método de tipo Lagrangiano aumentado.

Otra línea de investigación futura es realizar este estudio para otras EDPs más complejas, como pueden ser: una EDP del calor no lineal, la EDP de ondas (lineal y semi-lineal) y la EDP de Stokes y de Navier-Stokes. Así, podríamos considerar por ejemplo, el problema siguiente:

$$\begin{cases} \text{Minimizar } \phi(T,h) := \dfrac{T^2}{2} + \dfrac{b}{2}\displaystyle\iint_{\omega\times(0,+\infty)} |f|^2\,dx\,dt, \\ \text{Sujeto a: } (T,f) \in \mathbb{R}_+ \times L^2(\omega\times(0,+\infty)), \\ \qquad\qquad (u,p,f) \text{ resuelve (46)}, \\ \qquad\qquad \|u(T) - u_d\| = \delta, \end{cases} \tag{45}$$

con

$$\begin{cases} u_t - \nu\Delta u + (u\cdot\nabla)u + \nabla p = f1_\omega, \quad x\in\Omega\times(0,+\infty), \\ \nabla\cdot u = 0, \quad x\in\Omega\times(0,+\infty), \\ u = 0, \quad x\in\partial\Omega\times(0,+\infty), \\ u(0) = u_0, \quad x\in\Omega. \end{cases} \tag{46}$$

De nuevo tiene sentido estudiar si existe solución (y si es única). También debemos encontrar una caracterización análogo a la que encontramos para la EDP del calor y diseñar e implementar algoritmos y experiencias numéricas. Observemos que este salto no es trivial, ya que el cambio a una EDP más compleja dificulta mucho el estudio, sobre todo cuando pasamos a un marco no lineal. Por otro lado, los aspectos numéricos también adquieren gran complejidad, ya que las ecuaciones de (42) están fuertemente acopladas y no es fácil encontrar nuevos algoritmos que converjan a la solución y que, además, conduzcan a cálculos factibles.

También podemos plantearnos nuevos problemas de tiempo mínimo ligados, por ejemplo, a la búsqueda del tiempo necesario para que el estado final se aleje de un estado "deseado". En este caso deberíamos tener $\|\theta_0 - \theta_d\| < \delta$ y la restricción pasaría a ser $\|\theta(T) - \theta_d\| \geq \delta$.

Análisis teórico y numérico de la controlabilidad local nula para una ecuación parabólica quasi-lineal en dimensión 2 y 3

Theoretical and numerical local null controllability of a quasi-linear parabolic equation in dimensions 2 and 3

Los dos últimos trabajos incluidos en esta memoria se enmarcan dentro de la Teoría de Control, más concretamente dentro del campo de la controlabilidad nula. Comenzamos esta parte con un trabajo recogido en el artículo [10].

En este nuevo capítulo analizamos desde un punto de vista teórico y numérico la controlabilidad nula para una EDP parabólica quasi-lineal en dimensión 2 y 3. En otras palabras, pretendemos llevar a cero en el tiempo final $t = T$ la solución de

$$\begin{cases} y_t - \nabla\cdot(a(y)\nabla y) = v\tilde{1}_\omega, \quad (x,t)\in Q := \Omega\times(0,T), \\ y = 0, \quad (x,t)\in\Sigma := \partial\Omega\times(0,T), \\ y(x,0) = y_0(x), \quad x\in\Omega \end{cases} \tag{47}$$

donde $\tilde{1}_\omega \in C_0^\infty(\Omega)$ verifica $0 < \tilde{1}_\omega \leq 1$ en ω y $\tilde{1}_\omega = 0$ fuera de ω (esto es, $\tilde{1}_\omega$ es una función

característica regularizada) y $a \in C^3(\mathbb{R})$ cumple las condiciones

$$0 < m \leq a(r) \leq M, \quad |a'(r)| + |a''(r)| + |a'''(r)| \leq M \quad \forall r \in \mathbb{R}.$$

En este contexto hay trabajos anteriores en los que se consigue probar la controlabilidad local nula de (47) en dimensión 1. La novedad que aporta este trabajo reside en extender el resultado a una mayor dimensión espacial, algo que, como veremos, no es trivial. El resultado principal es el siguiente:

Teorema 10. *Bajo las condiciones ya mencionadas sobre a, existe* $\varepsilon > 0$ *tal que, si* $y_0 \in H^3(\Omega) \cap H_0^1(\Omega)$ *y además*

$$\|y_0\|_{H^3} \leq \varepsilon,$$

entonces existe un control $v \in L^2(\omega \times (0,T))$ *tal que*(47) *posee una solución que verifica*

$$y(x,T) = 0 \text{ en } \Omega.$$

Para la demostración del Teorema 10 hemos usado técnicas usuales, que reposan sobre el Teorema de la Función Inversa de Liusternik (véase [1]) y las desigualdades de Carleman. El origen de estas ideas está en el tabajo de Fursikov e Imanuvilov [16]. Han sido aplicadas en varios contextos distintos, véase las referencias de [9]. La dificultad del problema reside principalmente en el tratamiento del término no lineal. Por ello, en un primer paso de la demostración del teorema, trabajamos con el sistema linealizado

$$\begin{cases} y_t - a(0)\Delta y = v\tilde{1}_\omega + h(x,t), \quad (x,t) \in Q, \\ y = 0, \quad (x,t) \in \Sigma, \\ y(x,0) = y_0(x), \quad x \in \Omega \end{cases} \tag{48}$$

y probamos, gracias a las desigualdades de Carleman, que este sistema es exactamente controlable a cero, es decir, que existe un control v capaz de llevar a cero la solución en el tiempo final.

Un segundo paso de la demostración consiste en escribir el problema de control nulo asociado a (47) como una ecuación a resolver en un espacio de Hilbert adecuado, constituido por pares de estados-controles que se anulan para $t = T$. La ecuación tiene la forma

$$\mathcal{H}(y,v) = (0,y_0), \quad (y,v) \in Y. \tag{49}$$

Llegados a este punto, hacemos uso del Teorema de Liusternik, probamos que (49) pose solución cuando el dato inicial y_0 es suficientemente pequeño en norma H^3 y así concluimos la demostración. Para poder aplicar el Teorema de Liusternik, es preciso comprobar que $\mathcal{H}$ está bien definida y es $\mathcal{C}^1$ en un entorno de $(0,0)$ y que $\mathcal{H}'(0,0)$ es sobreyectiva. Aquí, necesitamos probar determinadas estimaciones no triviales de la solución del problema linealizado. Esta tarea es complicada desde el punto de vista técnico y constituye la contribución teórica más importante del trabajo.

Para terminar el estudio, añadimos un algoritmo de tipo quasi-Newton, dado como sigue:

ALG 1:

1. Elegir $(y^0, v^0) \in Y$.
2. Entonces, dados $n \geq 0$ y $(y^n, v^n) \in Y$, calcular

$$(y^{n+1}, v^{n+1}) = (y^n, v^n) - \mathcal{H}'(0,0)^{-1}(\mathcal{H}(y^n, v^n) - (0, y_0)). \tag{50}$$

Observemos que para este algoritmo mantenemos fija la evaluación de la inversa de $\mathcal{H}'$ en $(0,0)$. Incluimos un resultado relacionado con la convergencia del algoritmo (Teorema 11) y varios experimentos numéricos:

Teorema 11. *Sea $y_0 \in H^3(\Omega) \cap H_0^1(\Omega)$ dada, con $\|y_0\|_{H^3} \leq \varepsilon$ y sea (y, v) una solución de* (49) *proporcionada por el Teorema 10. Existe $\kappa \in (0,1)$ tal que, si $(y^0, v^0) \in Y$ y*

$$\|(y^0, v^0) - (y, v)\|_Y \leq \kappa,$$

entonces las (y^n, v^n) que genera ALG 1 convergen hacia (y, v) y verifican

$$\|(y^{n+1}, v^{n+1}) - (y, v)\|_Y \leq \theta \|(y^n, v^n) - (y, v)\|_Y \tag{51}$$

para todo $n \geq 0$, con $\theta \in (0,1)$.

De cara a las experiencias numéricas debemos ser capaces de identificar la inversa de $\mathcal{H}'$ en $(0,0)$. Para ello, se comprueba que $\mathcal{H}'(0,0)^{-1}(h, y_0) = (y, v)$ para cualesquiera (h, y_0) del espacio "admisible", con

$$y = \rho^{-2} L^* p, \quad v = -\rho_0^2 \, p|_{\omega \times (0,T)}\,, \tag{52}$$

siendo p la única solución del problema de Lax-Milgram

$$\pi(p, p') = \iint_Q h p' \, dx \, dt + \int_\Omega y_0(x) p'(x,0) \, dx \quad \forall p' \in P, \; p \in P. \tag{53}$$

En la formulación de este problema, aparece el espacio de Hilbert P que, por definición, es el completado de

$$P_0 := \{\varphi \in C^2(\overline{Q}) : \; \varphi = 0 \text{ on } \Sigma\}$$

$$\pi(\varphi, \tilde{\varphi}) := \iint_Q \left(\rho^{-2} L^* \varphi L^* \tilde{\varphi} + \tilde{1}_\omega \rho_3^{-2} \varphi \tilde{\varphi}\right) dx \, dt,$$

para el producto escalar donde $Ly := y_t - a(0)\Delta y$, $L^*\varphi := -\varphi_t - a(0)\Delta\varphi$.

Para resolver numéricamente (53), en principio es suficiente construir explícitamente espacios de dimensión finita $P_h \subset P$ que aproximen P. Sin embargo, esto requiere un trabajo considerable ya que las funciones de P deben satisfacer $L^* p \in L^2_{loc}(Q)$. Por ello, una aproximación basada en una triangulación estándar requiere que los espacios P_h sean espacios de funciones globalmente $\mathcal{C}^0$ en todas las variables y globalmente $\mathcal{C}^1$ en las variables espaciales. Esta construcción es compleja y demasiado trabajosa y costosa numéricamente. En consecuencia, hemos optado por resolver (53) mediante una formulación mixta y la correspondiente aproximación numérica.

Indiquemos brevemente cómo se puede hacer esto. Para ello, introducimos las nuevas variables:

$$z = \rho^{-1} L^* p \quad m = \rho_3^{-1} p,$$

los espacios: $Z := \{(z,m) \in L^2(Q) \times L^2(Q) : (\rho_3 m)_t \in L^2(Q),\ \nabla(\rho_3 m) \in L^2(Q)^N\}$ y $\Lambda := \{\lambda : \rho\lambda \in L^2(Q),\ \nabla(\rho\lambda) \in L^2(Q)\}$, las formas bilineales:

$$\alpha((z,m),(z',m')) := \iint_Q zz'\,dx\,dt + \iint_{\omega\times(0,T)} mm'\,dx\,dt,$$

$$\beta((z,m),\lambda) := \iint_Q \Big[\lambda\Big(z + \rho^{-1}((\rho_3 m)_t)\Big) - \nabla(\rho^{-1}\lambda)\cdot\nabla(\rho_3 m)\Big]\,dx\,dt$$

y la forma lineal

$$\langle \ell, (z,m)\rangle := \iint_Q \rho h m\,dx\,dt + \int_\Omega \rho_3(x,0) y_0(x) m(x,0)\,dx.$$

No es difícil comprobar que $\alpha(\cdot\,,\cdot)$, $\beta(\cdot\,,\cdot)$ y ℓ están bien definidas y son continuas, respectivamente, en $Z\times Z$, $Z\times\Lambda$ y Z.

Entonces, una formulación mixta apropiada de (53) es la siguiente:

Encuentra $(z,m) \in Z$ *y* $\lambda \in \Lambda$ *tal que*

$$\begin{cases} \alpha((z,m),(z',m')) + \beta((z',m'),\lambda) = \langle \ell,(z',m')\rangle \quad \forall (z',m') \in Z, \\ \beta((z,m),\lambda') = 0 \quad \forall \lambda' \in \Lambda. \end{cases} \tag{54}$$

Así, lo que tenemos que hacer es resolver numéricamente (54) y luego tomar

$$y = \rho^{-1} z, \quad v = -\rho_3^{-1}\, m|_{\omega\times(0,T)}\,.$$

Al contrario de P, no es difícil construir subespacios de dimensión finita de Z y Λ. Esto conduce a aproximaciones mixtas "naturales" y permite cálculos eficientes y computacionalmente razonables.

Vemos por tanto que, en cada paso de **ALG 1**, el problema se puede reescribir en la forma (54) y, después, aproximar en el sentido de los elementos finitos mixtos (54). Los cálculos se han realizado con el paquete *FreeFem++*.

Para las simulaciones numéricas hemos considerado los datos siguientes:

- $N = 2$, $\Omega = (0,1)\times(0,1)$, $\omega = (0{,}2, 0{,}8)\times(0{,}2, 0{,}8)$, $T = 0{,}5$.
- $y_0(x_1,x_2) = \sin(\pi x_1)\sin(\pi x_2)$.
- $a(s) = \exp(-2\exp(-0{,}3s))$.

El dominio elegido ha sido un rectángulo con otro rectángulo más pequeño en su interior que se corresponde con el dominio de control. El dominio espacio-temporal y su mallado aparecen en la Fig. 13. Por otra parte, el control y el estado calculados han sido detallados en la Fig. 14.

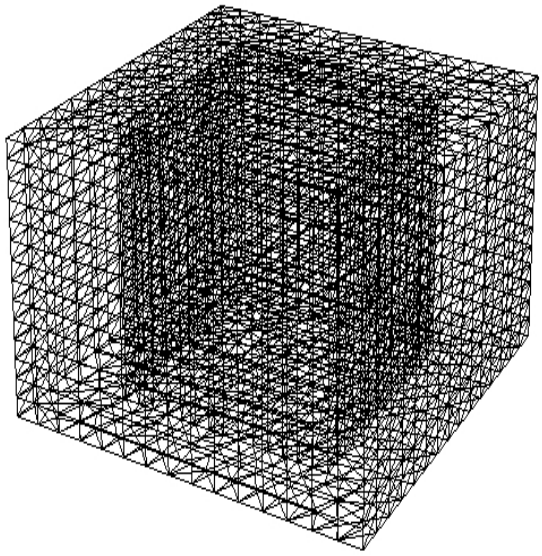
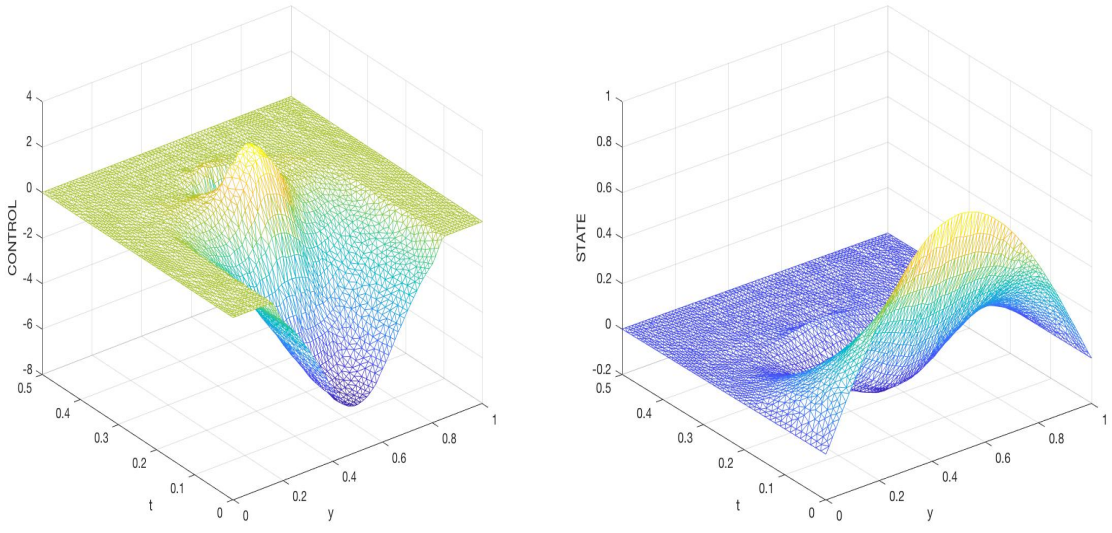


Figura 13: Mallado en 3D, con la tercera componente identificada a la variable temporal. Número de vértices: 7425. Número de tetraedros: 38976.



(a) Control (b) Estado asociado

Figura 14: Control y estado asociado en $x_1 = 0{,}68$.

Una línea futura de este trabajo podría ser ampliar el estudio a un problema de control frontera. Observemos que el resultado teórico demostrado en este caso conduce a otro análogo con control frontera. También podemos afirmar que conduce a otro de control nulo para T grande. No obstante, estas extensiones necesitan un tratamiento numérico específico que no es trivial. Como tema pendiente queda establecer un resultado análogo de control exacto a trayectorias y un resultado de control nulo global. Estas cuestiones parecen complicadas.

Aproximación a los controles nulos para una ecuación del calor semi-lineal usando aproximaciones por mínimos cuadrados

Approximation of null controls for semilinear heat equations using a least-squares approach

La memoria concluye con un último trabajo dedicado a la resolución de un problema de controlabilidad nula utilizando el método de mínimos cuadrados, véase [24].

Pretendemos llevar a cero el estado, esto es, la solución del problema semi-lineal.

$$\begin{cases} y_t - \Delta y + g(y) = f1_\omega \quad \text{en } Q_T := \Omega \times (0,T), \\ y = 0 \quad \text{sobre } \Sigma_T := \partial\Omega \times (0,T), \\ y(\cdot,0) = u_0 \quad \text{en } \Omega, \end{cases} \tag{55}$$

en el tiempo $t = T$. Aquí, $u_0 \in L^2(\Omega)$ y g es una función (al menos) localmente Lipschitz-continua que verifica

$$\begin{cases} g(0) = 0, \quad g' \in L^\infty(\mathbb{R}), \\ \sup\limits_{a,b\in\mathbb{R}, a\neq b} \dfrac{|g'(a) - g'(b)|}{|a-b|^s} < \infty \quad \text{con } s \in [0,1]. \end{cases} \tag{56}$$

Bajo estas hipótesis, se cumplen las condiciones impuestas en [14] para que (55) sea exactamente controlable a cero. Nuestro trabajo aporta sobre todo un nuevo esquema numérico de aproximación del control e incluye algunos ejemplos numéricos que corroboran la robustez del método implementado.

Comenzamos el trabajo introduciendo el método de mínimos cuadrados. Para ello, partimos de la formulación siguiente:

$$\begin{cases} \text{Minimizar } E(y,f) := \dfrac{1}{2}\Big\|\rho_2\big(y_t - \Delta y + g(y) - f1_\omega\big)\Big\|^2_{L^2(0,T;H^{-1}(\Omega)} \\ \text{Sujeto a: } (y,f) \in \mathcal{A} \end{cases} \tag{57}$$

donde $\mathcal{A}$ es un espacio "adecuado" y ρ_2 es un peso de tipo Carleman.

Más precisamente, tomaremos

$$\mathcal{A} = \Big\{ (y,f) : \rho\, y \in L^2(Q_T),\ \rho_1 \nabla y \in L^2(Q_T)^d,\ \rho_0 f \in L^2(q_T),$$
$$\rho_2(y_t - \Delta y - f\,1_\omega) \in L^2(0,T;H^{-1}(\Omega)),\ y(\cdot,0) = u_0 \text{ en } \Omega,\ y = 0 \text{ sobre } \Sigma_T \Big\},$$

donde, de nuevo, ρ, ρ_1 y ρ_0 son pesos de tipo Carleman. Esencialmente, esto quiere decir que se trata de funciones continuas y estrictamente positivas en $\Omega \times [0,T)$ que tienden a $+\infty$ exponencialmente cuando $t \nearrow T$.

Seguidamente, probamos que para cada $(y,f) \in \mathcal{A}$

$$\|(Y^1, F^1)\| \leq C\sqrt{E(y,f)} \tag{58}$$

y también

$$E'(y,f)\cdot(Y^1,F^1) = 2E(y,f) \tag{59}$$

para una norma y un producto escalar apropiados, donde (Y^1, F^1) es solución del sistema lineal auxiliar

$$\begin{cases} Y^1_t - \Delta Y^1 + g'(y)Y^1 = F^1 1_\omega + (y_t - \Delta y + g(y) - f1_\omega) \quad \text{en } Q_T, \\ Y^1 = 0 \quad \text{sobre } \Sigma_T, \quad Y^1(\cdot,0) = 0 \quad \text{en } \Omega. \end{cases} \tag{60}$$

Específicamente, en (58) y (59), la norma y el producto escalar están dados por

$$\begin{aligned} \big((y,f),(\overline{y},\overline{f})\big)_{\mathcal{A}} = \big(\rho y, \rho\overline{y}\big)_2 + \big(\rho_1\nabla y, \rho_1\nabla\overline{y}\big)_2 + \big(\rho_0 f, \rho_0\overline{f}\big)_2 \\ + \big(\rho_2(y_t - \Delta y - f\,1_\omega), \rho_2(\overline{y}_t - \Delta\overline{y} - \overline{f}\,1_\omega)\big)_{L^2(0,T;H^{-1}(\Omega))} \end{aligned}$$

y

$$\|(Y,F)\|_{\mathcal{A}} = \sqrt{((Y,F),(Y,F))_{\mathcal{A}}}.$$

Estas propiedades implican que resolver el problema de control nulo para (55) equivale a encontrar un mínimo para el funcional E donde E valga 0.

En una segunda parte del trabajo, consideramos el algoritmo:

$$\begin{cases} (y_0,f_0) \text{ dado en } \mathcal{A}, \\ (y_{k+1},f_{k+1}) = (y_k,f_k) - \lambda_k(Y^1_k, F^1_k) \quad \text{para } k \geq 0, \\ \lambda_k = \text{ arg mín } (E((y_k,f_k) - \lambda(Y^1_k,F^1_k)). \end{cases} \tag{61}$$

Y obtenemos el resultado siguiente:

Teorema 12. *Supongamos que g cumple las condiciones* (56) *para $s = 1$. Sea $\{(y_k,f_k)\}_{k\in\mathbb{N}}$ una sucesión dada por* (61)*. Entonces (y_k,f_k) converge en el sentido de la norma $\|\cdot\|_{\mathcal{A}}$ hacia un mínimo de E. Además, la convergencia es cuadrática después de un determinado número de iteraciones.*

En la tercera parte del trabajo conseguimos probar un teorema de convergencia análogo al Teorema 12 en el caso en que $s \in (0,1)$:

Teorema 13. *Supongamos que g cumple las condiciones* (56) *con $s \in (0,1)$. Sea $\{(y_k,f_k)\}_{k\in\mathbb{N}}$ una sucesión dada por* (61)*. Entonces (y_k,f_k) converge a un mínimo de E en el sentido de $\|\cdot\|$. Además la convergencia es de orden $1+s$ después de un determinado número de iteraciones.*

Las demostraciones de estos dos teoremas son muy técnicas y usan explícitamente las propiedades de la función g; de ahí que se distingan los casos $s = 1$ y $s < 1$.

En el caso en que $s = 0$, debemos imponer una condición adicional sobre la derivada de g para poder garantizar que la sucesión converge. Este caso se estudia por separado y conduce a sucesiones que convergen linealmente.

Para terminar el trabajo, se añade una serie de experiencias numéricas que corroboran la robustez del método.

Bibliografía

[1] V. M. Alekseev, V. M. Tikhomorov, S. V. Formin, *Optimal Control,* Consultants Bureau, New York, 1987.

[2] G. Allaire and A. Craig, *"Numerical Analysis and Optimization. An Introduction to Mathematical Modelling and Numerical Simulation",* Oxford, London, 2007.

[3] H. Brézis, L. Nirenberg, *"Characterizations of the ranges of some nonlinear operators and applications to boundary value problems"*, Annali della Scuola Normale Superiore di Pisa, Classe di Scienze, IV, **5**, 1978, 225-326.

[4] P. Brumer, M. Shapiro, *Laser control of chemical reactions*, Scientific American, 1995, 34-39.

[5] L. Caffarelli, R. Kohn, L. Nirenberg, *Partial regularity of suitable weak solutions of the Navier-Stokes equations,* Communications on Pure and Applied Mathematics, Vol. 35, 1982, 771-831.

[6] C. Fabre, *"Uniqueness results for Stokes equations and their consequences in linear and nonlinear control problems,"* Control, Optimisation and Calculus of Variations, **1**, 1996, 267-302.

[7] C. Fabre, G. Lebeau *"Prolongement unique des solutions de l'équation de Stokes,"* Comm. Partial Differential Equations, **21**, 1996, 573-596.

[8] E. Fernández-Cara, I. Marín-Gayte, *Analysis and numerical solution of some minimal time control problems*, submited.

[9] E. Fernández- Cara, I. Marín-Gayte, *A New Proof of the Existence of Suitable Weak Solutions and Other Remarks for the Navier-Stokes Equations*, Applied Mathematics, 9, 2018, 383-402.

[10] Fernández-Cara E., Límaco J., Marín-Gayte I., *Null controllability of a non-linear parabolic equation*, submited.

[11] E. Fernández-Cara, I. Marín-Gayte, *Theoretical and numerical bi-objective optimal control: Nash equilibria*, submited.

[12] E. Fernández-Cara, I. Marín-Gayte, *Theoretical and numerical results for some bi-objective optimal control problems*, Communications on Pure and Applied Analysis, 2020, 19, 4, 210-2126.

[13] E. Férnandez-Cara, E. Zuazua, *"Control theory: history, mathematical achievements and perspectives"*, Bol. SeMA, 26, 2003, 79-140.

[14] E. Férnandez-Cara, E. Zuazua, *"Null and approximate controllability for weakly blowing up semilinear heat equations"*, Ann. Inst. H. Poincaré, Anal. non linéaire, **17**, 5, 2000, 583-616.

[15] A. V. Fursikov, *"Optimal Control of Distributed Systems: Theory and Applications"*, American Mathematical Society Boston, MA, USA 2000.

[16] A. V. Fursikov, O. Y. Imanuvilov *"Controllability of Evolution Equations"*, Lecture Notes, Seoul National University, Korea, **34**, 1996.

[17] I. V. Girsanov, "Lectures on mathematical theory of extremum problems", Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, **67**, 1972.

[18] J. L. Guermond, *Faedo-Galerkin weak solutions of the Navier-Stokes equations with Dirichlet boundary conditions are suitable,* J. Math. Pure Appl., Vol. 88, 2007, 87-106.

[19] O. Y. Imanuvilov, *"Controllability of parabolic equations"*, Mat. Sbornik. Novaya Seriya, **186**, 1995, 109-132.

[20] O. Y. Imanuvilov, J.-P. Puel, *"Global Carleman estimates for weak elliptic non homogeneous Dirichlet problem"*, Int. Math. Research Notices, **16**, 2003, 883-913.

[21] O. Y. Imanuvilov, M. Yamamoto, *"Carleman estimates for a parabolic equation in a Sobolev space of negative order and its applications"*, Lecture Notes in Pure and Appl. Math, **218**, Dekker, New York, 2001.

[22] W. Karush, *Tesis: Minimal of functions of several variables with inequalities as side conditions*, Departamento de Matemáticas, Universidad de Chicago, 1939.

[23] U. Ledzewicz, H. Schättler, *Antiangiogenic therapy in cancer treatment as an optimal control problem*, SIAM Journal on Control and Optimization, **46**, (3) 2007 1052-1079.

[24] J. Lemoine, I. Marín-Gayte, A. Münch, *Approximation of null controls for semilinear heat equations using a least-squares approach*, submited.

[25] F. H. Lin, *A new proof of the Caffarelli-Kohn-Nirenberg theorem,* Comm. Pure Appl. Math., 51, 1998, 241-257.

[26] J. L. Lions, *Exact controllability, stabilizability and perturbations for distributed systems*, SIAM Review, 30, 1988, 1–68.

[27] J. L. Lions, E. Zuazua, *"A generic uniqueness result for the Stokes system and its control theoretical consequences"*, Partial Differential Equations and Applications, **177**, 1996, 221-235.

[28] R. M. Murray, ed., *Control in an information Rrich World: Report of the panel on future directions in control, dynamics, and systems*, SIAM, 2003.

[29] J. F. Nash, *"Noncooperative games",* Ann. Math., **54** (1951), 286-295.

[30] V. Pareto, *"Cours d'économie politique",* Rouge, Laussane, Switzerland, 1896.

[31] L. S. Pontryagin, V. G. Boltyanskii, R. V. Gamkrelidze, E. F. Mischenko, *The Mathematical Theory of Optimal Processes*, Interscience, 1962.

[32] http://www.pim.tsinghua.edu.cn/units/me/robot/en/home.html Robotics and Automation Laboratory.

[33] D. L. Russell, *Controllability and stabilizability theory for linear partial differential equations. Recent progress and open questions*, SIAM Review 20, 1978, 639–739.

[34] V. Scheffer, *Hausdorff measure and the Navier–Stokes equations,* Comm. Math. Phys. Vol. 55, 1977, 97-112.

[35] Lecture Notes in Control and Information Sciences, *New Directions and Applications in Control Theory*, Springer-Verlag Berlin Heidelberg, 321, 2008.

[36] H. Sohr, W. Wahl, *On the regularity of the pressure of weak solutions of Navier- Stokes equations,* Arch. Math., Vol. 46, 1986, 428-439.

[37] H. Von Stalckelberg, *"Marktform und gleichgewicht",* Springer, Berlin, Germany, 1934.

[38] D. Tataru, *"Carleman estimates and unique continuation for solutions to boundary value problems",* J. Math. Pures Appl., **75**, 1996, 367-408.

[39] H. B. Da Veiga, *On the suitable weak solution to the Navier-Stokes equations in the whole space,* J. Math. pures Appl., Vol. 64, 1985, 77-86.

[40] E. Zuazua, *"Exact boundary controllability for the semilinear wave equation"*, in Nonlinear Partial Differential Equations and their Applications, H. Brezis and J.L. Lions eds., Pitman **X**, 1991, 357-391.

[41] http://www.engr.uky.edu/˜jdjacob/fml/research/adaptive/index.html

Capítulo 1

A new proof of the existence of suitable weak solutions and other remarks for the Navier-Stokes equations

In this chapter, we prove that the limits of the semi-discrete and the discrete semi-implicit Euler schemes for the three-dimensional Navier–Stokes equations supplemented with Dirichlet boundary conditions are suitable in the sense of Scheffer [23]. This provides a new proof of the existence of suitable weak solutions, first established by Caffarelli, Kohn and Nirenberg [4]. These schemes have a discrete commutator property and satisfy a proper inf-sup condition (for space approximation). Finite element and wavelet spaces appear for this purpose. Our results are similar to the main result in [15]. We also present some additional remarks and open question on suitable solutions. This chapter is based on the work [11].

1.1. Introduction

The main objective of this work is to provide a new proof of the existence of suitable weak solutions to the Navier-Stokes equations.

We will be mainly concerned with the convergence of the semi-implicit Euler scheme with Diriclet boundary conditions in bounded domains $\Omega \times (0,T)$ (as usual, Ω is the spatial domain, a regular, bounded, connected open set in $\mathbb{R}^3$ "filled" by the fluid particles; on the other hand, $(0,T)$ is the time observation interval).

The practical interest of this new proof is that it may help to check wether the Caffarelli-Kohn-Nirenberg criteria are satisfied and locate singular points.

The techniques extend those generally used for the Navier-Stokes equations and can be applied to many other approximation schemes that lead to energy inequalities, as those in [7, 8, 14, 22].

More precisely, we first use the well known energy estimates, together with appropiate interpolation results and see that the approximate solutions converge to a weak solution (u,p). Then, we analyze the role of the associated pressure p. This reduces in fact to a detailed study of the behavior of the time derivate of the velocity field. This way, we are able to prove that we can take (lower) limits in local energy identities and deduce that (u,p) is suitable.

The plan of the chapter is the following:

- In Section 2, we review the main results in the papers [4] and [17]. In particular, we explain why suitable solutions are relevant in the context of the regularity problem.
- In Section 3, we recall the Euler approximation schemes and we establish the convergence to a suitable solution of the Navier-Stokes equations.
- Finally, Section 4 is devoted to some additional comments and open questions.

1.2. Background: the basic results by Caffarelli, Kohn, Nirenberg

1.2.1. The main properties of suitable solutions

In this section, we will recall the main results of Caffarelli, Kohn and Nirenberg, see [4]. In this reference, the best results known to date in relation to the regularity of the Navier-Stokes equations are established.

We will consider the Navier-Stokes equations in three dimensions with data and boundary conditions as follows,

$$\begin{cases} u_t + u \cdot \nabla u - \Delta u + \nabla p = f, \ (x,t) \in Q, \\ \nabla \cdot u = 0, \ (x,t) \in Q, \\ u(x,t) = 0, \ (x,t) \in \Sigma, \\ u(x,0) = u_0(x), \ x \in \Omega. \end{cases} \tag{1.1}$$

Here $\Omega \subset \mathbb{R}^3$ is a regular, bounded, connected open set, $T > 0$, $Q := \Omega \times (0,T)$, $\Sigma := \partial\Omega \times (0,T)$, $f = (f^1, f^2, f^3)$ verifies $f \in L^q(\Omega \times (0,T))^3$ with $q \geq 2$ and $\nabla \cdot f = 0$ and $u_0 \in H^1_0(\Omega)^3$ with $\nabla \cdot u_0 = 0$.

For convenience, let us introduce the (frequently used) spaces H and V, with

$$H := \{v \in L^2(\Omega)^3 : \ \nabla \cdot v = 0 \text{ in } \Omega, \ v \cdot \overline{n} \text{ on } \partial\Omega\},$$
$$V := \{v \in H^1_0(\Omega)^3 : \ \nabla \cdot v = 0 \text{ in } \Omega\}.$$

Definition 1.2.1. *It will be said that the couple (u,p) is a **weak solution** to* (1.1) *if the following holds:*

1. $u \in L^2(0,T;V) \cap L^\infty(0,T;H)$, $p \in L^{5/3}(Q)$,
2. u *and* p *satisfy the Navier-Stokes equations in* (1.1) *in the distributional sense in* Q.
3. $u|_{t=0} = u_0$ *a.e. in* Ω.

It is well known that any couple (u, p) satisfying the previous first and second points also verifies

$$u \in L^{10/3}(Q)^3 \cap C^0_w([0,T];H),\ u_t \in L^{4/3}(0,T;V'). \tag{1.2}$$

In particular, u can be viewed as a well defined H-valued function and the third assertion in Definition 1.2.1 has a sense as an equality in H.

In order to understand the situation, it is convenient to associate a dimension to each variable in (1.1). Note that, if the pair (u, p) is a solution to the Navier-Stokes problem, then for each $\lambda > 0$ the functions

$$u_\lambda(x,t) := \lambda u(\lambda x, \lambda^2 t)$$

and

$$p_\lambda := \lambda^2 p(\lambda x, \lambda^2 t)$$

solve a similar problem in $Q_\lambda := \lambda^{-1}\Omega \times (0,T)$, with force $f_\lambda := \lambda^3 f(\lambda x, \lambda^2 t)$.

Thus, we say that a variable or a linear differential operator is of dimension k, with k an integer, if it is adimensionalized when multiplied by λ^{-k}, where λ is a characteristic length. We can affirm that

- x_i has dimension 1 and t is of dimension 2,
- u^i has dimension -1 and p has dimension -2,
- f has dimension -3,
- ∂_i has dimension -1 and ∂_t has dimension -2,

so that, each term of the equation has dimension -3.

The analysis of the existence of a weak solution to (1.1) can be found for instance in [26] and [20]. Now, we will speak of the regularity problem. So, we start with the following definition:

Definition 1.2.2. *Let $(x,t) \in \Omega \times (0,T)$ be given. It will be said that (x,t) is a **singular point** if the solution u is not L^∞ in any neighborhood of (x,t), that is, there are no r and C such that $|u(x,t)| \leq C$ for $(x,t) \in B((x,t);r)$. The remaining points, where u is locally bounded, will be called **regular points**.*

In [4], the authors were able to estimate a Hausdorff-type dimension of the set of singular points for a class of weak solutions. This estimate shows that this set is "small" and therefore leads to a partial regularity theorem.

Before presenting this estimates let us recall a previous result from Scheffer [23] on partial regularity:

Theorem 1.2.3. *If $f \equiv 0$, there exists a weak solution of (1.1) whose singular set S satisfies:*

$$\mathcal{H}^{5/3}(S) < +\infty \ \text{ and } \ \mathcal{H}^1(S \cap (\Omega \times \{t\})) < +\infty \ \text{uniformly in } t. \tag{1.3}$$

Here, $\mathcal{H}^k$ denotes the usual Hausdorff's $k-$dimensional measure in $\mathbb{R}^4$.

The main result in [4] is a local partial regularity theorem for a particular class of weak solutions, denoted **suitable weak solutions** or simply suitable solutions. It is shown that, for any suitable solution the set of singular points has Hausdorff's one-dimensional measure equal to zero. In fact, the authors prove that $\mathcal{P}^1(S) = 0$, where $\mathcal{P}^1$ is a measure analogous to the Hausdorff one-dimensional measure $\mathcal{H}^1$ where we use parabolic cylinders instead of balls. In particular, $\mathcal{H}^1 \leq C\mathcal{P}^1$ and one has $\mathcal{H}^1(S) = 0$.

The definition of a suitable solution is the following:

Definition 1.2.4. *Let $D = G \times (a, b)$ be a cylinder in $\mathbb{R}^3 \times \mathbb{R}$. It is said that (u, p) is a **suitable weak solution** to the Navier-Stokes equations in D if it satisfies points 1. and 2. of Definition 1.2.1 and, furthermore, the following generalized energy inequality:*

For each $\phi \in C_0^\infty(D)$ with $\phi \geq 0$,

$$2 \iint_D |\nabla u|^2 \phi \leq \iint_D \Big(|u|^2(\phi_t + \Delta\phi) + (|u|^2 + 2p)u \cdot \nabla\phi + 2(u \cdot f)\phi \Big). \tag{1.4}$$

The main result of [4] is the following:

Theorem 1.2.5. *Let (u, p) be a suitable solution to the Navier-Stokes equation in D. Then the associated singular set satisfies $\mathcal{P}^1(S) = 0$.*

In particular, this results hold for any suitable weak solution to (1.1).

This result improves Theorem 1.2.3 in several aspects: first, it has local character; then it gives a better estimate of the Haussdorff dimension of S and, finally, it allows a rather general force term f.

In the following, for each (x, t) and $r > 0$, we set

$$Q_r(x, t) := \{(y, \tau) : \ |y - x| < r, \ t - r^2 < \tau < t\}.$$

It will be said that $Q_r(x, t)$ is a **parabolic cylinder** around (x, t). For the proof of Theorem 1.2.5, we need several results. The first one is the following:

Proposition 1.2.6. *Suppose that (u, p) is a suitable weak solution to the Navier-Stokes equations in $Q_1 = Q_1(0, 0)$ and $f \in L^q(Q_1)^3$, $q > \frac{5}{2}$. There exist $\epsilon_1, C_1 > 0$ and $\epsilon_2 = \epsilon_2(q) > 0$ such that, if*

$$\iint_{Q_1} (|u|^3 + |u||p|) + \int_{-1}^0 \Big(\int_{|x|<1} |p| \, dx \Big)^{5/4} dt \leq \epsilon_1 \tag{1.5a}$$

and

$$\iint_{Q_1} |f|^q \leq \epsilon_2, \tag{1.5b}$$

then

$$|u(x, t)| \leq C_1 \text{ a.e. in } Q_{1/2} := Q_{1/2}(0, 0). \tag{1.5c}$$

In particular, the point $(0, 0)$ is regular.

Proposition 1.2.6 shows that the sizes of the data and the suitable solutions are not independent of their regularity.

Now, if we introduce

$$M(r) := \frac{1}{r^2}\iint_{Q_r}(|u|^3+|u||p|) + r^{-13/4}\int_{t-r^2}^{t}\Big(\int_{|y-x|<r}|p|dy\Big)^{5/4}d\tau \tag{1.6a}$$

and

$$F_q(r) := r^{3q-5}\iint_{Q_r}|f|^q \tag{1.6b}$$

we can deduce the following:

Corollary 1.2.7. *Suppose that (u,p) is a suitable solution to the Navier-Stokes equations in the cylinder $Q_r(x,t)$ and $f \in L^q(Q_r(x,t))^3$, $q > \frac{5}{2}$. Then, if $M(r) \le \epsilon_1$ and $F_q(r) \le \epsilon_2$, one has*

$$|u| \le C_1 r^{-1} \text{ a.e. in } Q_{r/2}(x,t). \tag{1.7}$$

In particular, every point of the cylinder $Q_{r/2}(x,t)$ is regular.

Let's put

$$Q_r^*(x,t) := \{(y,\tau) :\ |y-x| < r,\ t - \frac{7}{8}r^2 < \tau < t + \frac{1}{8}r^2\}.$$

Note that $Q^*r(x,t) = Q_r(x, t+\frac{1}{8}r^2)$. The second fundamental tool for the proof of Theorem 1.2.5 reads as follows:

Proposition 1.2.8. *Let (u,p) be a suitable solution to the Navier-Stokes equations in a neighborhood of (x,t). There exists $\epsilon_3 > 0$ such that, if*

$$\limsup_{r\to 0}\frac{1}{r}\iint_{Q_r^*(x,t)}|\nabla u|^2 \le \epsilon_3, \tag{1.8}$$

then (x,t) is a regular point.

In order to use the generalized energy inequality, Caffarelli, Kohn, Nirenberg bounded the integral in the right

$$\iint(|u|^2+2p)u\cdot\nabla\phi$$

in terms of

$$\iint|u|^2\phi \text{ and } \iint|\nabla u|^2\phi.$$

Bounds of this kind play a fundamental role in the proofs of both Propositions 1.2.6 and 1.2.8. In all cases, they amount to interpolation inequalities for u, combined with estimates for p. The methods used in [4] to prove Proposition 1.2.8 play a relevant role here: they bounded u in terms of ∇u by interpolation and then p in terms of u by solving $\Delta p = -\nabla_{ij}(u^i,u^j)$ at each time. They also need information on how $\int|u|^2$ changes in time; this is controlled using the generalized energy inequality.

However, in [15], Guermond provided a new proof of the existence of suitable solutions using Faedo-Galerkin method. He proved that solutions obtained by the Faedo–Galerkin method verify the local energy estimate.

The two main stumbling blocks for proving the local energy estimate are in the passage to the limit in the nonlinear terms $\nabla\cdot(u^2u)$ and $\nabla\cdot(pu)$. While the discrete commutator property together with standard a priori estimates is just what to take care of $\nabla\cdot(u^2u)$, passing to the limit on $\nabla\cdot(pu)$ requires nontrivial estimates on the pressure. Thus, Guermond try to reproduce for the discrete pressure a priori estimates that are similar to the estimates of Sohr and Von Wahl, see [25]. This is achieved using of the fractional powers of the discrete Stokes operator and deriving estimates in the $H^\tau(0,T;H^{-\alpha}(\Omega))$-norm.

1.2.2. Sketch of the proof of Theorem 1.2.5

Theorem 1.2.5 is a consequence of Proposition 1.2.8. The argument is explained below.

Consider first the proof of the fact that S has Hausdorff dimension less than or equal to $5/3$, that is, Theorem 1.2.3.

Using Corollary 1.2.7 and a covering lemma, we can easily see that, for each $\delta>0$, S can be covered by a family of mutually disjoint parabolic cylinders $\{Q^*_{r_i}(x_i,t_i)\}$ such that $r_i<\delta$ and

$$\frac{1}{r_i^2}\iint_{Q^*_{r_i/5}(x_i,t_i)}(|u|^3+|u||p|)+r^{-13/4}\int_{t_i-\frac{7}{8}r_i^2}^{t_i+\frac{1}{8}r_i^2}\Big(\int_{|x-x_i|<r_i}|p|dx\Big)^{5/4}dt>\epsilon_1 \tag{1.9}$$

for all i. Using Hölder's inequality, we deduce that

$$r_i^{-5/3}\iint_{Q^*_{r_i/5}(x_i,t_i)}(|u|^{10/3}+|p|^{5/3})\geq C(\epsilon_1)$$

and therefore

$$\sum r_i^{5/3}\leq C\iint_{\cup Q^*_{r_i/5}(x_i,t_i)}(|u|^{10/3}+|p|^{5/3})\leq C.$$

Taking $\delta\to 0$, we find that $\mathcal{P}^{5/3}(S)=0$, whence we see, in particular, that the Hausdorff dimension of S is at most $5/3$.

To show that $\mathcal{P}^1(S)=0$ by a similar method, instead of the integral of $|u|^{10/3}+|p|^{5/3}$, we need a global quantity of dimension 1. This is furnished by Proposition 1.2.8. Indeed, this result allows to replace (1.9) by

$$\frac{1}{r_i}\iint_{Q^*_{r_i/5}(x_i,t_i)}|\nabla u|^2>\epsilon_3 \tag{1.10}$$

and, this way, we are led to the estimate

$$\sum r_i\leq C\iint_{\cup Q^*_{r_i/5}(x_i,t_i)}|\nabla u|^2$$

whence we conclude that $\mathcal{P}^1(S)=0$.

It is natural to ask if we can get a better estimate of the dimension of S. In other words, can we find $k < 1$ such that $\mathcal{P}^k(S) = 0$? This question has not been answered up to now. Actually, the answer does not seem simple and is related to the possibility of demonstrating an additional estimate of the (suitable) weak solutions of order less than 1.

It is important to note that the assumption $f \in L^q(Q)^3$ with $q > 5/2$ is mainly needed to prove Proposition 1.2.6. On the other hand, note that in Theorem 1.2.5, Caffarelli, Kohn and Nirenberg chose to estimate the measure $\mathcal{P}^1$ of the set S, instead the standard measure $\mathcal{H}^1$. Both definitions are special cases of a construction made by Carathéodory that is detailed in [9].

The argument used by Caffarelli, Kohn and Nirenberg is valid for any suitable solution. In the Appendix of [4], they prove the existence of such a solution. Thus, the following holds:

Theorem 1.2.9. *Suppose that $u_0 \in V$, $f \in L^q(Q)^3$ with $q > \frac{5}{2}$ and $\nabla \cdot f = 0$ in Q. . Then, there exists at least one suitable weak solution (u, p) to the Navier-Stokes equations in $Q := \Omega \times (0, T)$ satisfying*

$$u(t) \rightharpoonup u_0 \text{ in } H \text{ as } t \to 0.$$

In addition, one has:

$$\int_{\Omega\times\{t\}} |u|^2\phi + 2\int_0^t\int_\Omega |\nabla u|^2\phi \le \int_\Omega |u_0|^2\phi(x,0) + \int_0^t\int_\Omega \Big(|u|^2(\phi_t + \Delta\phi) + (|u|^2 + 2p)u\cdot\nabla\phi + 2(u\cdot f)\phi\Big) \tag{1.11}$$

for all functions $\phi \in \mathcal{D}(\Omega \times [0, T])$ with $\phi \ge 0$ and $\phi = 0$ near $\partial\Omega \times (0, T)$.

1.3. Some convergence results

In the sequel, we will denote by $|\cdot|$ and $(\cdot,\cdot)$ the usual L^2 norm and scalar product, respectively.

1.3.1. The convergence of the semi-approximate problems

In this section, we will give a new proof of Theorem 1.2.9 different from in [4]. To do this, we will apply the semi-implicit Euler semi-discrete time scheme to the Navier-Stokes equations.

The scheme is the following. We take N large enough (the number of time steps) and define the time step $\tau := T/N$, the instants $t^m = m\tau$ and the approximations

$$f^m := \frac{1}{\tau}\int_{t^{m-1}}^{t^m} f(x,t)dt, \tag{1.12}$$

$u^m \simeq u(\cdot, t^m)$ and $p^m \simeq p(\cdot, t^m)$, with $u^0 = u_0$ and

$$\begin{cases} \dfrac{u^{m+1} - u^m}{\tau} + (u^m\cdot\nabla)u^{m+1} - \Delta u^{m+1} + \nabla p^{m+1} = f^{m+1}\ x \in \Omega, \\ \nabla\cdot u^{m+1} = 0\ x\in\Omega,\ \displaystyle\int_\Omega p^{m+1} = 0, \\ u^{m+1} = 0,\ x \in \partial\Omega, \end{cases} \tag{1.13}$$

for $m = 0, 1, ..., N-1$.

First of all, let us check that the u^m are well defined:

Lemma 1.3.1. *The previous Euler scheme is well defined, that is, for every $m \geq 0$ there exists a unique solution (u^{m+1}, p^{m+1}) to* (1.13).

The proof is inmediate by induction. We only need to note that, for each m, (1.13) is a Dirichlet problem for a linear PDE system that ca be written in the form

$$\begin{cases} \text{Find } u \in V \text{ such that} \\ \\ \dfrac{1}{\tau}\Big(u, v\Big) + \Big((w \cdot \nabla)u, v\Big) + \Big(\nabla u, \nabla v\Big) = \Big(g, v\Big) \quad \forall v \in V, \end{cases} \tag{1.14}$$

where $w \in V$ and $g \in L^2(\Omega)^3$ are given.

Now, let us see that the u^m are uniformly bounded in the norm $|\cdot|$. We have:

$$\Big(\frac{u^{m+1} - u^m}{\tau}, u^{m+1}\Big) + \Big((u^m \cdot \nabla)u^{m+1}, u^{m+1}\Big) + \Big(\nabla u^{m+1}, \nabla u^{m+1}\Big) = \Big(f^{m+1}, u^{m+1}\Big), \tag{1.15}$$

which can be rewritten in the form

$$\frac{1}{2}(|u^{m+1}|^2 - |u^m|^2) + \frac{1}{2}|u^{m+1} - u^m|^2 + \tau|\nabla u^{m+1}|^2 = \tau(f^{m+1}, u^{m+1}). \tag{1.16}$$

Using the Cauchy-Schwarz and Young inequalities, we easily get

$$\frac{1}{2}(|u^{m+1}|^2 - |u^m|^2) + \tau|\nabla u^{m+1}|^2 \leq C|f^{m+1}|^2 + \frac{\tau}{2}|\nabla u^{m+1}|^2, \tag{1.17}$$

whence

$$|u^{n+1}|^2 + \tau \sum_{m=0}^{n} |\nabla u^{m+1}|^2 \leq C \sum_{m=0}^{n} |f^{m+1}|^2 + |u^0|^2 \leq C \tag{1.18}$$

for all n and, certainly, u^m is uniformly bounded in H.

Using this Euler scheme, we can construct the approximate solutions of the Navier-Stokes system. More precisely, let us define the following functions:

- $u_N : [0, T] \mapsto V$, the unique continuous piecewise linear function satisfying

$$u_N(t^m) = u^m, \text{ for } m = 0, 1, \ldots, N.$$

- $u_N^*[0, T] \mapsto V$, the piecewise constant function characterized by

$$u_N^*(t) = u^{m+1}, \text{ in } t \in (t^m, t^{m+1}], \ m = 0, 1, \ldots, N-1.$$

In a similar way, we can introduce the approximate pressure p_N^* and force f_N^* (again piecewise constant). The following holds:

Lemma 1.3.2. *For any N and almost every $t \in (0,T)$, one has*

$$\begin{cases} u_{N,t} + (u_N^*(t-\tau)\cdot\nabla)u_N^* - \Delta u_N^* + \nabla p_N^* = f_N^*, \\ \nabla \cdot u_N^* = 0. \end{cases} \tag{1.19}$$

We can now present the main result of this section, that is related to the convergence of u_N and u_N^* towards a suitable weak solution of the Navier-Stokes equation:

Theorem 1.3.3. *After eventual extraction of a subsequence, the functions u_N^* converge weakly in $L^2(0,T;V)$, weakly-$*$ in $L^\infty(0,T;H)$ and strongly in $L^2(Q)^3$ and a.e. towards a suitable weak solution to* (1.1) *as $N \to +\infty$.*

For the proof of Theorem 1.3.3, it will be convenient to recall the following Lemma, whose proof is given in [26]:

Lemma 1.3.4. *Let u a function of $L^2(0,T;V)$ and $u_t \in L^2(0,T;V')$. Then u is equal a.e. to a continuous function from $[0,T]$ into H. In addition, the function $t \mapsto |u(t)|^2$ is absolutely continuous and*

$$\frac{d}{dt}|u(t)|^2 = 2\langle u_t(t), u(t)\rangle \text{ a.e. in } (0,T). \tag{1.20}$$

Proof of Theorem 1.3.3:

Let us first try to find the spaces where the functions u_N^*, u_N and p_N^* are uniformly bounded.

Consider the identities (1.16). Let us fix N, take n so that $0 \le n \le N-1$ and let us add in m, from 0 to n. The following is obtained:

$$\frac{1}{2}|u^{n+1}|^2 + \frac{1}{2}\sum_{m=0}^{n}|u^{m+1}-u^m|^2 + \tau\sum_{m=0}^{n}|\nabla u^{m+1}|^2 = \tau\sum_{m=0}^{n}(f^{m+1},u^{m+1}) + \frac{1}{2}|u^0|^2. \tag{1.21}$$

Obviously, this can also be written in the form

$$\frac{1}{2}|u_N^*(t)|^2 + \frac{1}{2}\sum_{m=0}^{n}|u^{m+1}-u^m|^2 + \sum_{m=0}^{n}\int_{t^m}^{t^{m+1}}|\nabla u_N^*(t)|^2dt$$
$$= \sum_{m=0}^{n}\int_{t^m}^{t^{m+1}}(f_N^*(t),u_N^*(t))dt + \frac{1}{2}|u^0|^2 \text{ for all } t \in (t^m,t^{m+1}].$$

Therefore,

$$\frac{1}{2}|u_N^*(t)|^2 + \int_0^{t^{n+1}}|\nabla u_N^*(t)|^2dt \le \int_0^{t^{n+1}}(f_N^*(t),u_N^*(t))dt + \frac{1}{2}|u^0|^2 \tag{1.22}$$

and, from the Cauchy-Schwarz and Young inequalities, we easily see that

$$|u_N^*(t)|^2 + \int_0^T|\nabla u_N^*(t)|^2dt \le |u^0|^2 + C\|f_N^*\|^2_{L^2(Q)} \ \forall t \in [0,T]. \tag{1.23}$$

This means that

$$u_N^* \text{ is uniformly bounded in } L^2(0,T;V) \text{ and } L^\infty(0,T;H). \tag{1.24}$$

On the other hand, it can also be deduced from (1.13) that

$$\int_0^T |u_N(t) - u_N^*(t)|^2 dt \leq \tau \sum_{m=0}^{N-1} |u^{m+1} - u^m|^2 \leq C\tau. \tag{1.25}$$

where $\|u_N^* - u_N\|^2_{L^2(Q)} \leq C\tau$.

To estimate u_N, we use its definition and the fact that, for any $t \in (t^m, t^{m+1})$,

$$|u_N(t)| \leq |u^m| + |u^{m+1}| \text{ and } |\nabla u_N(t)| \leq |\nabla u^m| + |\nabla u^{m+1}|.$$

Accordingly, we also have that

$$u_N \text{ is uniformly bounded in } L^2(0,T;V) \text{ and } L^\infty(0,T;H). \tag{1.26}$$

Now, from classical interpolation results, we deduce that u_N^* and u_N are uniformly bounded in the spaces
$L^r(0,T;L^{6r/(3r-4)}(\Omega))$ for $r \in [2,+\infty]$.

It is well known that these estimates allow us to prove that the u_N belong and are uniformly bounded in the Sobolev spaces of fractional order to a dimension of $H^\gamma(0,T;H)$ for $0 < \gamma < 1/4$. Therefore, as a consequence of the well known Aubin-Lion's principle, the u_N belong to a compact set of $L^2(Q)$, see for example [26].

Hence, at least for a subsequence (again denoted $\{u_N\}$), we must have:

$$\begin{cases} u_N \to u \text{ weakly in } L^2(0,T;V) \text{ and weakly-} * \text{ in } L^\infty(0,T;H), \\ u_N \to u \text{ strongly in } L^2(Q) \text{ and a.e. in } Q. \end{cases} \tag{1.27}$$

This is enough to pass to the limit in (1.19) and deduce that u is a weak solution of (1.1).

Note that it can also be assumed that

$$u_N \to u \text{ strongly in } L^r(0,T;L^q(\Omega)^3) \text{ for } 2 < r < +\infty,\ 1 \leq q < 6r/(3r-4). \tag{1.28}$$

To show that u is suitable, we have to complete the previous estimates. To this purpose, we will use some regularity results that play the role of the Sohr and Wahl [25] estimates in the results in [17].

For $0 < s < 1$, the space $H^s(\Omega) := [H^1(\Omega), L^2(\Omega)]_s$ is defined by the method of real interpolation between $H^1(\Omega)$ and $L^2(\Omega)$, i.e. the so-called K-method of Lions and Peetre [19], see also [18] or [1]. We will denote by $H_0^s(\Omega)$ the closure of $\mathcal{D}(\Omega)$ in $H^s(\Omega)$ For any $s < 0$, the space $H^{-s}(\Omega)$ is defined by duality and, in particular,

$$\|v\|_{H^{-s}} = \sup_{w \in \mathcal{D}(\Omega)\setminus\{0\}} \frac{(v,w)}{\|w\|_{H^s}} \quad \forall v \in L^2(\Omega).$$

We will look for a uniform estimate of $u_{N,t}$ in a space of the form $L^a(0,T;H^{-\sigma}(\Omega))$. This way, applying De-Rham's Lemma (see [24]), we will get a bound of p_N^* in $L^a(0,T;H^{1-\sigma}(\Omega))$ and we will be able to take limits in the generalized energy inequality.

Note that, for all m, one has $u^m = w^m + z^m$, where the w^m and the z^m are respectively given by

$$\begin{cases} \dfrac{w^{m+1}-w^m}{\tau} + Aw^{m+1} = 0 \\ w^0 = u_0 \end{cases} \tag{1.29}$$

and

$$\begin{cases} \dfrac{z^{m+1}-z^m}{\tau} + Az^{m+1} = F^{m+1} \\ z^0 = 0, \end{cases} \tag{1.30}$$

with $F^{m+1} = f^{m+1} - (u^m \cdot \nabla)u^{m+1}$, A being the Stoke operator. Recall that $A : D(A) \subset H \mapsto H$, with

$$D(A) = H^2(\Omega)^3 \cap V,\ Av = P(-\Delta v) \quad \forall v \in V$$

(here, $P : L^2(\Omega)^3 \mapsto H$ is the orthogonal projector). Also, recall that there exists an orthogonal basis of V formed by eigenfunctions

$$A\xi_j = \lambda_j \xi_j,\ \ \xi_j \in V,\ \ |\xi_j| = 1,\ \ \lambda_j \nearrow +\infty \tag{1.31a}$$

and

$$D(A^r) = \Big\{ v \in H :\ \sum_{j\geq 1} \lambda_j^{2r} |(v,\xi_j)|^2 < +\infty \Big\} \tag{1.31b}$$

for all $r \geq 0$.

In the sequel, we will consider the functions w_N, w_N^*, z_N and z_N^*, whose definitions are similar to the definitions of u_N and u_N^*.

First, note that

$$w^m = (Id + \tau A)^{-m} u_0 \ \forall m = 0, 1, \ldots, N,$$

for all m, whence

$$\begin{aligned} \|w_{N,t}\|^2_{L^2(Q)} &= \tau \sum_{m=0}^{N-1} |A(Id+\tau A)^{-(m+1)} u_0|^2 \\ &= \tau \sum_{m=0}^{N-1} \sum_{j\geq 1} \frac{\lambda_j^2}{(1+\tau\lambda_j)^{2(m+1)}} |(u_0,\xi_j)|^2 \\ &= \tau \sum_{j\geq 1} \Big(\sum_{m=0}^{N-1} \frac{1}{(1+\tau\lambda_j)^{2(m+1)}} \Big) \lambda_j^2 |(u_0,\xi_j)|^2 \\ &\leq \sum_{j\geq 1} \frac{\lambda_j^2}{(1+\tau\lambda_j)^2} |(u_0,\xi_j)|^2 = \|u_0\|^2_{L^2(Q)}. \end{aligned} \tag{1.32}$$

Therefore

$$w_{N,t} \text{ and } Aw_N^* \text{ are uniformly bounded in } L^2(Q). \tag{1.33}$$

Let us see now what can be said of $z_{N,t}$ and Az_N^*. For all $m \geq 1$, we have

$$z^m = \sum_{l=1}^{m} (Id + \tau A)^{-(m+1-l)} F^l. \tag{1.34}$$

Let $s, \sigma \in (0,1)$ be such that with $\sigma > s$. Then

$$\|Az^{m+1}\|_{H^{-\sigma}} \leq \tau \sum_{l=1}^{m+1} \|A(Id + \tau A)^{-(m+2-l)}\|_{\mathcal{L}(H^{-s};H^{-\sigma})} \|F^l\|_{H^{-s}} = \tau \sum_{l=1}^{m} a_{m-l} b_l, \tag{1.35}$$

where the a_n and the b_l are given by

$$a_n = \|A(Id + \tau A)^{-(n+2)}\|_{\mathcal{L}(H^{-s};H^{-\sigma})}, \;\; b_l = \|F^l\|_{H^{-s}}. \tag{1.36}$$

We will apply the following result that must be viewed as a discrete version of the well known Young inequality for convolution products:

Lemma 1.3.5. *Let us assume that $k \geq 1$, $a \in l^p$ and $b \in l^q$. Then, if $r \in [1, +\infty]$ and*

$$\frac{1}{p} + \frac{1}{q} = \frac{1}{r} + 1, \tag{1.37}$$

one has

$$\Big(\sum_{n=1}^{k} \Big| \sum_{l=1}^{n} a_{n-l} b_l \Big|^r \Big)^{1/r} \leq \Big(\sum_{n=1}^{k} a_n^p \Big)^{1/p} \Big(\sum_{n=1}^{k} b_n^q \Big)^{1/q} \tag{1.38}$$

for all $k \geq 1$.

The proof of this result can be found in [13].
Using Lemma 1.38 with $r = a$, $p = 1$ and $q = a$, we find that

$$\Big(\tau \sum_{n=1}^{N} \|Az^{m+1}\|_{H^{-\sigma}}^a \Big)^{1/a} \leq \Big(\tau^{1+a} \sum_{n=1}^{N} \Big| \sum_{l=1}^{n} a_{n-l} b_l \Big|^a \Big)^{1/a} \leq \Big(\sum_{n=1}^{N} a_n \Big) \Big(\tau \sum_{n=1}^{N} b_n^a \Big)^{1/a}. \tag{1.39}$$

From the estimates (1.24) already obtained for u_N^*, it is immediate that, for any $a \in [1,2]$, F_N^* is uniformly bounded in $L^a(0,T; L^{3a/(4a-2)}(\Omega))$ and, consequently, also in $L^a(0,T; H^{-(5a-4)/(2a)}(\Omega))$. Thus choosing a in $[1,2]$ and taking $s = (5a-4)/(2a)$, we get:

$$\|F_N^*\|_{L^a(0,T;H^{-s}(\Omega))} = \Big(\tau \sum_{m=1}^{N} b_m^a \Big)^{1/a} \leq C(a). \tag{1.40}$$

On the other hand, for any smooth z, one has

$$\begin{aligned}\|A(Id+\tau A)^{-(n+2)}z\|^2_{H^{-\sigma}} &= \sum_{j\geq 1}\lambda_j^{-\sigma}\frac{\lambda_j^2}{(1+\tau\lambda_j)^{2(n+2)}}|(z,\xi_j)|^2\\ &\leq \Big[\sup_j \frac{\lambda_j^{2(1-\epsilon)}}{(1+\tau\lambda_j)^{2(n+2)}}\Big]\|z\|^2_{H^{-s}},\end{aligned} \tag{1.41}$$

where $\epsilon=(\sigma-s)/2$. Therefore, recalling the definition of the a_n, we deduce that

$$a_n \leq \frac{C(\epsilon)}{(n\tau)^{1-\epsilon}},\ \ \tau\sum_{n=1}^m a_n \leq C(\epsilon)\int_0^T \frac{ds}{s^{1-\epsilon}} \leq C(\epsilon) \tag{1.42}$$

and finally,

$$Az_N^*\|_{L^a(0,T;H^{-\sigma})} \leq C\Big(\sum_{m=1}^N \tau a_m\Big)\Big(\tau\sum_{m=1}^N \|F^m\|^a_{H^{-s}}\Big)^{1/a} \leq C\|F_N^*\|_{L^a(0,T;H^{-s})}. \tag{1.43}$$

Note that this estimate is valid for all $a\in[1,2]$, with $s=(5a-4)/(2a)$ and $\sigma\in(0,1)$, $\sigma>s$. Obviously, the same estimate is valid for $\|z_{N,t}\|_{L^a(0,T;H^{-\sigma})}$. This prove that

$$\begin{aligned}&z_{N,t} \text{ and } Az_N^* \text{ are uniformly bounded in}\\ &\qquad L^a(0,T;H^{-\sigma}(\Omega)^3)\ \forall a\in[0,1]\ \forall\sigma>s=(5a-4)/(2a).\end{aligned} \tag{1.44}$$

In view of (1.33) and (1.44), it follows that Au_N^* and $u_{N,t}^*$ are also uniformly bounded in $L^a(0,T;H^{-\sigma}(\Omega)^3)$. From De-Rham's Lemma [24], we see that p_N^* is uniformly bounded in $L^a(0,T;H^{1-\sigma}(\Omega))$ which is continuously embedded in $L^a(0,T;L^{6/(1+2\sigma)}(\Omega))$. In particular, for $a=5/3$, we have $s=13/10$ and we can take σ as close as desired to s, which gives $6/(1+2\sigma)$ as close as desired to $5/3$.

After extracting a new squence (if that is needed), we can assume that p_N^* converges weakly in $L^{5/3}(0,T;L^\beta(\Omega))$ for all $\beta>5/3$. Let us check that the local energy inequality holds for u and p.

If we multiply the equation (1.19) by the function $u_N^*\phi$, where $\phi\in C_0^\infty(\Omega\times[0,T])$ is nonnegative and we integrate in space, we have:

$$\int_\Omega u_{N,t}\cdot u_N^*\phi+\int_\Omega (u_N^*(t-\tau)\cdot\nabla)|u_N^*|^2\phi-\int_\Omega \Delta u_N^*\cdot u_N^*\phi+\int_\Omega \nabla p_N^*\cdot u_N^*\phi=\int_\Omega f_N^*\cdot u_N^*\phi, \tag{1.45}$$

If $t\in[0,T]$, there exists n such that $t\in(t^n,t^{n+1}]$ and then

- $\displaystyle\int_\Omega u_{N,t}u_N^*\phi=\int_\Omega u_{N,t}u_N\phi+\int_\Omega u_{N,t}(u_N^*-u_N)\phi$
  $\displaystyle=\frac12\frac{d}{dt}\int_\Omega|u_N|^2\phi-\frac12\int_\Omega|u_N|^2\phi_t+\int_\Omega u_{N,t}(u_N^*-u_N)\phi$, using Lemma 1.3.4.

  Note moreover that $\displaystyle\int_\Omega u_{N,t}(u_N^*-u_N)\phi\geq 0$ because $u_N-u_N^*$ by definition is equal to $(t^{n+1}-t)u_{N,t}$.

- $\displaystyle\int_\Omega (u_N^*(t-\tau)\cdot\nabla)|u_N^*|^2\phi = -\frac{1}{2}\int_\Omega u_N^*(t-\tau)|u_N^*|^2\cdot\nabla\phi,$
- $\displaystyle -\int_\Omega \Delta u_N^*\cdot u_N^*\phi = \int_\Omega |\nabla u_N^*|^2\phi - \frac{1}{2}\int_\Omega |u_N^*|^2\Delta\phi,$
- $\displaystyle\int_\Omega \nabla p_N^*\cdot u_N^*\phi = -\int_\Omega p_N^*\cdot u_N^*\nabla\phi$

Consequently, we can write the following inequality holds:

$$\begin{aligned}\frac{1}{2}\frac{d}{dt}\int_\Omega |u_N|^2\phi + \int_\Omega |\nabla u_N^*|^2\phi \le \\ \frac{1}{2}\int_\Omega |u_N|^2\phi_t + \int_\Omega \Big[\Big(\frac{1}{2}u_N^*(t-\tau)|u_N^*|^2 + p_N^*u_N^*\Big)\cdot\nabla\phi + \frac{1}{2}|u_N^*|^2\Delta\phi + f_N^*\cdot u_N^*\phi\Big]\end{aligned} \tag{1.46}$$

If we integrate in time, we see that

$$\begin{aligned}\int_\Omega |u_N|^2\phi + 2\iint_{\Omega\times(0,t)} |\nabla u_N^*|^2\phi \\ \le \int_\Omega |u_0|^2\phi + \iint_{\Omega\times(0,t)} |u_N|^2\phi_t \\ + \iint_{\Omega\times(0,t)} \Big[\Big(u_N^*(t-\tau)|u_N^*|^2 + 2p_N^*u_N^*\Big)\cdot\nabla\phi + |u_N^*|^2\Delta\phi + 2f_N^*\cdot u_N^*\phi\Big].\end{aligned} \tag{1.47}$$

Thanks to the previous estimates, we can take the lower limit in the left hand side and the limit in the terms in the right. In all of them this is possible; for example, since u_N^* converges strongly (for example) in $L^{5/2}(0,T;L^{19/5}(\Omega))$ and p_N^* converges weakly in $L^{5/3}(0,T;L^{14/9}(\Omega))$, $p_N^*u_N^*\cdot\nabla\phi$ converges weakly in $L^1(Q)^3$ towards $pu\cdot\nabla\phi$.

The final result is

$$2\iint_{\Omega\times(0,t)} |\nabla u|^2\phi \le \int_\Omega |u_0|^2\phi + \iint_{\Omega\times(0,t)} |u|^2\Big(\phi_t+\Delta\phi\Big) + \Big(u|u|^2+2pu\Big)\cdot\nabla\phi + 2fu\phi, \tag{1.48}$$

and this shows that (u,p) is a suitable weak solution of (1.1). □

1.3.2. The convergence of the fully approximate problems

In this section, we will argue as in [15] and we will check that the approximate solutions obtained via the semi-implicit Euler discrete scheme, used together with an appropiate approximation in space, converge to a suitable solution to (1.1).

As before, let us introduce N, $\tau := T/N$ and the $t^m = m\tau$. We will also consider two families of finite dimensional spaces $\{X_h\}_{h>0}$ and $\{P_h\}_{h>0}$ with $X_h \subset H_0^1(\Omega)^3$ and $P_h \subset L^2(\Omega)$ such that

$$\begin{cases} \inf_{v_h\in X_h} \|v-v_h\|_{H^1} \to 0 \ \ \forall v\in H_0^1(\Omega)^3, \\ \inf_{q_h\in P_h} \|q-q_h\|_{L^2} \to 0 \ \ \forall q\in L^2(\Omega)\end{cases} \tag{1.49}$$

and the (X_h, P_h) are uniformly compatible in the sense that there exists a constant c independent of h such that the following $inf - sup$ conditions are satisfied:

$$\inf_{q_h \in P_h \setminus \{0\}} \sup_{v_h \in X_h \setminus \{0\}} \frac{(\nabla q_h, v_h)}{\|v_h\|_{H^{1-s}} \|q_h\|_{H^s}} \geq c \quad \forall h \in [0,1]. \tag{1.50}$$

Now, we consider the approximations $u_h^m = u(\cdot, t^m) \in X_h$ and $p_h^m = p(\cdot, t^m) \in P_h$, with $u_h^0 = u_{0h}$ (the orthogonal projection of u_0 on X_h) and

$$\begin{cases} \left(\frac{u_h^{m+1} - u_h^m}{\tau}, v_h\right) + \left((u_h^m \cdot \nabla) u_h^{m+1}, v_h\right) \\ \qquad + \left(\nabla u_h^{m+1}, \nabla v_h\right) + \left(\nabla p_h^{m+1}, v_h\right) = \left(f_h^{m+1}, v_h\right) \quad \forall v_h \in X_h, \\ \left(q_h, \nabla \cdot u_h^{m+1}\right) = 0 \ \ \forall q_h \in P_h, \\ (u_h^{m+1}, p_h^{m+1}) \in (X_h, P_h), \end{cases} \tag{1.51}$$

for $m = 0, 1, ..., N_1$.

The following result, which is an inmediate consequence of (1.50), gives coherence to our scheme:

Lemma 1.3.6. *The previous discrete scheme in time and space is well defined, that is, for every $m \geq 0$ and every $h > 0$, there exists a unique solution (u_h^{m+1}, p_h^{m+1}) to* (1.51)*.*

As before, the u_h^m and p_h^m serve to construct approximate solutions to the Navier-Stokes system. More precisely, we define functions $u_{N,h}, u_{N,h}^*, p_{N,h}^*$, etc. respectively similar to u_N, u_N^*, p_N^*, etc.

The main result of this section is the following:

Theorem 1.3.7. *After eventual extraction of a subsequence, the functions $u_{N,h}^*$ converge weakly in $L^2(0,T;V)$, weakly-$*$ en $L^\infty(0,T;H)$ and strongly in $L^2(Q)^3$ towards a suitable weak solution to* (1.1) *as $N \to +\infty$ and $h \to 0$.*

Sketch of the proof:

Arguing as in the proof of the Theorem 1.3.3, it can be proved that the $u_{N,h}$ and the $u_{N,h}^*$ are uniformly bounded in $L^2(0,T;V)$ and $L^\infty(0,T;H)$ and, furthermore, $\|u_{N,h} - u_{N,h}^*\|_{L^2(Q)}^2 \leq C\tau$. As in [26], we can also prove that the $u_{N,h}$ are uniformly bounded in $H^\sigma(0,T;H)$ for any $\gamma \in (0, 1/4)$. Consequently, at least for a subsequence (still indexed with N and h), one has:

$$\begin{cases} u_{N,h} \to u \text{ weakly in } L^2(0,T;V) \text{ and weakly-} * \text{ in } L^\infty(0,T;H), \\ u_{N,h} \to u \text{ strongly in } L^2(Q) \text{ and a.e. in } Q \end{cases} \tag{1.52}$$

as $N \to +\infty$ and $h \to 0$. Again, this is enough to pass to the limit and deduce that u is a weak solution of (1.1).

Note that it can also be assumed that

$$u_{N,h} \to u \text{ strongly in } L^r(0,T;L^q(\Omega)^3) \text{ for } 2 < r < +\infty, \ 1 \leq q < 6r/(3r-4).$$

To show that u is suitable, we need to improve these estimates as Guermond did in [15]. To this end, he used regularity results that play the same role played by the Sohr and Wahl [25] estimates

in the proof of Theorem 2.8 in [17]. In this context, we need the spaces $\tilde{H}_0^s(\Omega) := [L^2(\Omega), H_0^1(\Omega)]_s$ for $s \in (0,1)$ and their dual spaces $\tilde{H}^{-s}(\Omega)$.

The following estimates are established in [15]:

- For any $\alpha \in [1/4, 1/2)$ and any $\delta < \overline{\delta}(\alpha) = 2(1+\alpha)/5$, one has

$$\|u_{N,h,t}\|_{H^{\delta-1}(0,T;\tilde{H}^{-\alpha})} + \|u^*_{N,h}\|_{H^{\delta}(0,T;\tilde{H}^{-\alpha})} \leq C(\alpha), \tag{1.53}$$

- For any $s \in [1/2, 7/10]$ and any $r > \overline{r}(s) = (3-2s)/4$, one also has

$$\|u_{N,h,t}\|_{H^{-r}(0,T;H^{-s})} + \|p^*_{N,h}\|_{H^{-r}(0,T;H^{1-s})} \leq C(s), \tag{1.54}$$

As a consequence, it can be assumed that the $p^*_{N,h}$ converge weakly (for instance) in $H^{-r}(0,T;H^{3/8}(\Omega))$ for all $r > 7/16$ and the $u^*_{N,h}$ converge strongly in $H^{\delta}(0,T;\tilde{H}^{-\alpha}(\Omega)^3)$ for all $\alpha < 3/8$ and $\delta < 11/20$. This is sufficient to ensure that $p^*_{N,h}u^*_{N,h}$ converges weakly in $L^1(Q)^3$ towards pu.

Now, arguing as in the final part of the proof of Theorem 1.3.3, it is not difficult to check that the limit (u,p) of the $(u^*_{N,h}, p^*_{N,h})$ is a suitable weak solution to (1.1). This ends the proof. □

1.4. Some additional coments and questions

1.4.1. The same results hold for the Boussinesq system

The Boussinesq systemis the following:

$$\begin{cases} u_t - \Delta u + (u\cdot\nabla)u + \nabla p = f + \theta k,\ (x,t) \in Q, \\ \nabla u = 0,\ (x,t) \in Q, \\ \theta_t + u\cdot\nabla\theta - \Delta\theta = g,\ (x,t) \in Q, \\ u(x,t) = 0,\ \theta(x,t) = 0,\ (x,t) \in \Sigma, \\ u(x,0) = u_0(x),\ \theta(x,0) = \theta_0(x),\ x \in \Omega. \end{cases} \tag{1.55}$$

We assume here that

$$u_0 \in V,\ \ \theta_0 \in H_0^1(\Omega)\ \ f \in L^2(0,T,H^{-1}(\mathbb{R}^3)^3), k \in \mathbb{R}^3 \ \text{ and } g \in L^2(0,T;L^2(\Omega)). \tag{1.56}$$

As in [3], we can speak of weak solutions to (1.55) and, also, of suitable weak solutions to the previous PDEs in any set of the form $D = G \times (a,b)$, with $G \subset \mathbb{R}^3$ a connected open set.

The results in Section 3 can be extended to this framework. Thus, we can for instance consider the semi-implicit Euler scheme

$$\begin{cases} \dfrac{u^{m+1}-u^m}{\tau} - \Delta u^{m+1} + (u^m \cdot \nabla)u^{m+1} + \nabla p^{m+1} = f^{m+1} + \theta^{m+1}k, \\ \nabla u^{m+1} = 0, \\ \dfrac{\theta^{m+1}-\theta^m}{\tau} + u^m \cdot \nabla \theta^{m+1} - \Delta \theta^{m+1} = g^{m+1} \end{cases} \tag{1.57}$$

and prove that, at least for a subsequence, the associated $u_n, u_N^*, p_N^*, \theta_N$ and θ_N^* converge, in a appropiate sense, to a suitable weak solution (u, p, θ).

1.4.2. Possible extensions to other systems

It would be interesting to prove similar results to Theorems 1.3.3 and 1.3.7 for the solutions to the variable-density Navier-Stokes equations:

$$\begin{cases} \rho_t + \nabla \cdot (\rho u) = 0, \ (x,t) \in Q, \\ \rho(u_t + (u \cdot \nabla)u) - \Delta u + \nabla p = \rho f, \ (x,t) \in Q, \\ \nabla \cdot u = 0, \ (x,t) \in Q, \\ u(x,t) = 0, \ (x,t) \in \Sigma, \\ u(x,0) = u_0(x), \ \rho(x,0) = \rho_0(x), \ x \in \Omega. \end{cases} \tag{1.58}$$

However, this is not clear at present. Note that the "reasonable" definition of a suitable weak solution should involve the following property: for any $\phi \in \mathcal{D}(Q)$ with $\phi \geq 0$,

$$\iint_D |\nabla u|^2 \phi \leq \iint_D \Big(|u|^2(\rho\phi_t + \Delta\phi) + (\rho|u|^2 + 2p)(u \cdot \nabla\phi) + 2\rho(u \cdot f)\phi \Big). \tag{1.59}$$

But, unfortunately, the apparent lack of regularity of p makes it difficult to prove this.

1.4.3. Extensions to other approximation schemes for the Navier-Stokes equations

As we already said, Theorems 1.3.3 and 1.3.7 can be adapted to many other time approximation schemes. Among then, let us simply recall the following:

- Crank-Nicholson scheme:

$$\frac{u^{m+1}-u^m}{\tau} + (u^m \cdot \nabla)u^{m+1} - \Delta\Big(\frac{u^{m+1}+u^m}{2}\Big) + \nabla p^{m+1} = f^{m+1}, \quad \nabla \cdot u^{m+1} = 0. \tag{1.60}$$

- Gear scheme:

$$\frac{3u^{m+1}-4u^m+u^{m-1}}{2\tau} + (u^m \cdot \nabla)u^{m+1} - \Delta u^{m+1} + \nabla p^{m+1} = f^{m+1} \quad \nabla \cdot u^{m+1} = 0. \tag{1.61}$$

- $\theta-$scheme: For α and β such that $0 < \alpha, \beta < 1$ and $\alpha + \beta = 1$, we compute $(u^{n+\theta}, p^{n+\theta})$, then $u^{n+1-\theta}$ and finally (u^{n+1}, p^{n+1}) as follows:

$$\frac{u^{n+\theta} - u^n}{\theta \Delta t} - \alpha\nu\Delta u^{n+\theta} + \nabla p^{n+\theta} = f^{n+\theta} + \beta\nu\Delta u^n - (u^n \cdot \nabla)u^n, \quad \nabla \cdot u^{n+\theta} = 0, \tag{1.62a}$$

$$\frac{u^{n+1-\theta} - u^{n+\theta}}{(1-2\theta)\Delta t} - \beta\nu\Delta u^{n+1-\theta} + (u^{n+1-\theta} \cdot \nabla)u^{n+1-\theta} = f^{n+\theta} + \alpha\nu\Delta u^{n+\theta} - \nabla p^{n+\theta}. \tag{1.62b}$$

$$\begin{aligned} \frac{u^{n+1} - u^{n+1-\theta}}{\theta \Delta t} - \alpha\nu\Delta u^{n+1} + \nabla p^{n+1} &= f^{n+1} \\ +\beta\nu\Delta u^{n+1-\theta} - (u^{n+1-\theta} \cdot \nabla)u^{n+1-\theta}, \nabla \cdot u^{n+1} &= 0. \end{aligned} \tag{1.62c}$$

It would be interesting: to find the analog of Propositions 1.2.6 and 1.2.8 for a family of approximated solutions. This can help to detect or discard the occurence of singular points.

Bibliography

[1] R. A. Adams, J. J. F. Fournier, *Sobolev Spaces, second ed.,* J. Pure and Applied Maths, Vol. 140, 2003, pp. 305.

[2] W. Briggs, H. Van Emden *The DFT, An owner's manual for the Discrete Fourier Transform,* Society for Industrial and Applied Mathematics, 1995.

[3] G. Boling, Y. Guangwei, *On the suitable weak solutions for the Cauchy-Problem of the Boussinesq equations*, Nonlinear Analysis Theory, Methods and Applications, Vol. 26, 1996, pp. 1367-1385.

[4] L. Caffarelli, R. Kohn, L. Nirenberg, *Partial regularity of suitable weak solutions of the Navier-Stokes equations,* Communications on Pure and Applied Mathematics, Vol. 35,1982, pp. 771-831.

[5] A. J. Chorin, J. E. Marsden, *A mathematical introduction to fluid mechanics,* Springer-Verlag, New-York, 1979.

[6] D. Cioranescu, *Sur une classe de fluides non-newtoniens,* Appl. Math. Optimiz, Vol. 3, 1977, pp. 263-282.

[7] E. J. Dean, R. Glowinski, *On some finite element methods for the numerical simulation for incompressible viscous flow,* Cambridge University Press, New York,1993, pp. 109-150.

[8] J. Douglas, H. H. Rachford, *On the solution of the heat conduction problem in 2 and 3 space variables,* Trans. Amer. MAth. Soc., Vol. 82, 1956, pp. 421-439.

[9] H. Federer, *Geometric measure theory,* Springer-Verlag, New York, 1969.

[10] E. Fernández-Cara, F. Guillén, R. R. Ortega, *Mathematical modeling and analysis of viscoelastic fluids of the Oldroyd kind,* Handbooks of Numerical Analysis, North-Holland, Amsterdam, Part 2, Vol. 8, 2002, pp. 543-661.

[11] E. Fernández-Cara, I. Marín-Gayte, *A New Proof of the Existence of Suitable Weak Solutions and Other Remarks for the Navier-Stokes Equations*, Applied Mathematics, 2018, 9, 383-402.

[12] D. Fujiwara, H. Morimoto, *An L, theorem on the Helmholtz decomposition of vector fields,* Tokyo Univ. Fac. Sciences J., Vol. 24, 1977, pp. 685-700.

[13] Y. Galperin, *Young's convolution inequalities for weighted mixed (quasi-)norm spaces,* Journal of Inequalities and Special Functions, Vol. 5, 2014, pp. 1-12.

[14] R. Glowinski, *Numerical methods for fluids (Part 3),* North-Holland, Vol. IX, 2013.

[15] J. L. Guermond, *Faedo-Galerkin weak solutions of the Navier-Stokes equations with Dirichlet boundary conditions are suitable,* J. Math. Pure Appl., Vol. 88, 2007, pp. 87-106.

[16] D. D. Joseph, *Fluid dynamics os viscoelastic liquids,* Applied Mathematical Sciences, Springer-Verlag, New-York, Vol. 84, 1990.

[17] F. H. Lin, *A new proof of the Caffarelli-Kohn-Nirenberg theorem,* Comm. Pure Appl. Math., 51, 1998, pp. 241-257.

[18] J. L. Lions, E. Magenes, *Problèmes aux limites non homogénes et applications,*Vol.1, Dunod, Paris, France, 1968.

[19] J. L. Lions, J. Peetre, *Sur une classe d'espaces d'interpolation,* Inst. Hautes Études Sci. Publ.Math.19, 1964, pp.5-68.

[20] P. L. Lions, *Mathematical topics in fluid mechanics, Vol. 1: Incompresible models,* Clarendon Press, Oxford, 1996.

[21] R. L. Panton, *Incompressible flow,* Wiley Interscience, New York, 1984.

[22] D. H. Peaceman, H. H. Rachford, *The numerical solution of parabolic and elliptic differential equations,* J. Soc. Ind. Appl. Math., Vol. 3, 1955, pp. 28-41.

[23] V. Scheffer, *Hausdorff measure and the Navier–Stokes equations,* Comm. Math. Phys. Vol. 55, 1977, pp. 97–112.

[24] J. Simon, *Non-homogeneous viscous incompressible fluids: Existence of velocity, density and pressure,* SIAM J. Math. Anal., Vol 21, n°5, 1990.

[25] H. Sohr, W. Wahl, *On the regularity of the pressure of weak solutions of Navier- Stokes equations,* Arch. Math., Vol. 46, pp. 428-439,1986.

[26] R. Temam, *Navier-Stokes equations. Theory and numerical analysis,* Studies in Mathematics and Applications, North-Holland Publishing Co., Amsterdm- New York- Oxford, Vol. 2, 1977.

[27] H. B. Da Veiga, *On the suitable weak solution to the Navier-Stokes equations in the whole space,* J. Math. pures Appl., Vol. 64, 1985, pp. 77-86.

Capítulo 2

Theoretical and numerical results for some bi-objective optimal control problems

This chapter deals with the solution of some multi-objective optimal control problems for several PDEs: linear and semilinear elliptic equations and stationary Navier-Stokes systems. More precisely, we look for Pareto equilibria associated to standard cost functionals. First, we study the linear and semilinear cases. We prove the existence of equilibria, we deduce appropriate optimality systems, we present some iterative algorithms and we establish convergence results. Then, we analyze the existence and characterization of Pareto equilibria for the Navier-Stokes equations. Here, we use the formalism of Dubovitskii and Milyoutin. In this framework, we also present a finite element approximation of the bi-objective problem and we illustrate the techniques with several numerical experiments. The work is based on [12].

2.1. Introduction

In this work we consider bi-objective optimal control problems for various PDEs and systems. First, an introductory problem corresponding to a linear elliptic PDE is analyzed with detail. Then, we deal with a semilinear elliptic PDE. Finally, we focus on the stationary Navier-Stokes system, that is, the equations satisfied by the velocity field and the pressure of a steady viscous incompressible fluid.

Our aims are to prove existence, characterize efficiently the equilibria and, also, compute the solutions to these multi-objective control problems. They are very important from the theoretical and practical viewpoints and appear frequently in the applications. For some previous works on the subject, see for instance [3].

In classical control theory, we usually find a state equation or system and one control with the mission of achieving a predetermined goal. Frequently (but not always), the goal is to minimize a cost functional within a prescribed family of admissible controls. A different and interesting situation arises when several (in general, conflictive or contradictory) objectives are considered. This may happen, for example, if the cost function is the sum of several terms and it is not clear that an average provides a reasonable criterion. Furthermore, it can also be expectable to have more than one control acting on the equation. In these cases, we are led to consider multi-objective

control problems. In contrast with the mono-objective case, various strategies for the choice of good controls can appear, depending of the characteristics of the problem. Moreover, these strategies can be cooperative or noncooperative (depending on whether or not several controls mutually cooperate in order to achieve prescribed goal).

There exist several equilibrium concepts for multi-objective problems, with origin in game theory and mainly motivated by economics. Each of them determines a strategy. Thus, let us mention the noncooperative optimization strategy proposed by Nash [22], the Pareto cooperative strategy [23] and the Stackelberg hierarchical strategy [25]. In the context of the control of PDEs, a relevant question is whether one is able to steer the system to a desired state (exactly or approximately) by applying controls that correspond to one of these strategies. Up to date, there have been some works on the subject like the seminal papers by Lions [19, 21] and other more recent contributions, like [4, 5, 7].

In this work, we will be concerned with Pareto equilibria associated to standard cost functionals. Let us give the details in the case of the stationary Navier-Stokes equations.

Thus, let the fluid domain be a bounded open set $\Omega \subset \mathbb{R}^N$, with $N = 2$ or 3. Let us introduce three nonempty open subsets, $\mathcal{O}_1, \mathcal{O}_2$ and ω (which is the control domain) and let us assume that a velocity field u_{id} defined on $\mathcal{O}_i$ is given for $i = 1, 2$.

In this context, we want to find suitable forces f (the controls) in $L^2(\omega)^N$ with the following property: there exists an associated state (u, p), that is, a weak solution to the system

$$\begin{cases} -\nu\Delta u + (u \cdot \nabla)u + \nabla p = f1_\omega, \quad x \in \Omega, \\ \nabla \cdot u = 0, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \end{cases} \tag{2.1}$$

such that (f, u) is a Pareto equilibrium for the functionals

$$J_i(f, u) := \frac{a}{2} \int_{\mathcal{O}_i} |u - u_{id}|^2 + \frac{\mu}{2} \int_\omega |f|^2, \quad i = 1, 2, \tag{2.2}$$

where $a, \mu > 0$ (for the definition of Pareto equilibrium, see below).

Our first main goal will be to show that at least one optimal Pareto solution exists. The second one will be to characterize such equilibria in terms of first order optimality conditions, i.e. to deduce a system of PDEs that any optimal solution (together with an associated adjoint state) must satisfy. The third one will be to indicate how Pareto equilibria can be computed, to present some related algorithms and illustrate the results with numerical experiments.

The proof of the existence of Pareto equilibria is (more or less) standard. It relies on suitable and well known a priori estimates for the solutions to (2.1).

In what concerns optimality conditions, the situation is more delicate. Indeed, due to the lack of uniqueness, the techniques usually employed for distributed control problems (see for instance [1,8,9,20]) cannot be applied in this case. For this reason, we will use an alternative technique that relies on the so called formalism of Dubovitskii and Milyoutin. This approach was introduced in the context of mathematical programming and has been succesfully applied to the solution of optimal control problems for differential equations since the 70's. A good presentation of its applications to these areas can be found in Girsanov [16]; see also Flett [13]. In particular, these techniques have been applied in a very promising way to some distributed control problems; see for instance [4,5,7].

The basic ideas of the formalism can be explained as follows. At a local minimizer, the cone of descent directions associated to a cost functional must be disjoint of the intersection of the cones of feasible and tangent directions, respectively determined by the admissible control set and (2.1). Indeed, we cannot "move" from the minimizer to another admissible pair in a direction that improves the objective functions. Consequently, from Hahn-Banach Theorem and some additional arguments, it follows that there must exist elements in the associated dual cones, not all them zero, that add up to zero. This algebraic condition is just the Euler-Lagrange system of the extremal problem at hand. When it is possible to identify the previous primal and dual cones, this system provides the first order optimality conditions in a systematic way. In the case of a standard (mono-objective) optimal control problem, it also leads to the corresponding Pontryagin minimum (or maximum) principle.

Thus, a major task in our problem is the identification of the cones mentioned above in terms of the involved PDEs and functionals. Note that, in the particular case of (2.1)-(2.2), the main difficulties are related to the highly nonlinear behavior of (2.1) and the possible nonuniqueness of u.

The plan of this chapter is the following:

Section 2: A relatively simple problem: linear elliptic PDE and quadratic functionals.
Definitions, existence and characterization.
Algorithms and convergence.

Section 3: A slightly more complex problem: semilinear elliptic PDE and quadratic functionals.
New definitions, existence and optimality systems.
Algorithms and convergence.

Section 4: The stationary Navier-Stokes system (a more difficult problem).
The difficulties: nonlinearity, lack of uniqueness, lack of regularity of the functionals.
Definitions, existence and characterization.
Algorithms and numerical experiments.

2.2. Introductory problem: a linear elliptic PDE

This section aims to be an introducction to the study of Pareto equilibria for linear PDEs. Specifically, we will consider a multi-objective optimal control problem for a linear elliptic equation, we will present the definition of Pareto equilibrium, we will prove its existence, we will describe its characterization and, finally, we will formulate some related algorithms. The linearity of the problem will greatly facilitate the study.

In the sequel, we denote by $\|\cdot\|$ and $(\cdot\,,\cdot)$ the usual L^2 norm and scalar product, respectively. The symbol $\mathbb{1}_D$ will be used to denote the characteristic function of the set D. For simplicity, we will assume that only two functionals are considered but very similar considerations hold for systems with a higher number of functionals.

2.2.1. Definition of Pareto equilibria

To fix ideas, we will consider systems with only one control acting on a (small) subset of the domain.

Let $\Omega \subset \mathbb{R}^N$ be a nonempty regular, bounded and connected open set and let us assume that $\omega \subset \Omega$ is a nonempty open subset.

We will consider the problem

$$\begin{cases} -\Delta u = f 1_\omega, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \end{cases} \tag{2.3}$$

where $f \in L^2(\omega)^N$ is the control and $\mathbb{1}_\omega$ is the characteristic function of ω.

Let $\mathcal{O}_1$ and $\mathcal{O}_2$ be open sets that represent prescribed observations domains and let the J_i given by

$$J_i(f) := \frac{a}{2}\int_{\mathcal{O}_i} |u - u_{id}|^2 + \frac{\mu}{2}\int_\omega |f|^2, \quad i = 1, 2, \tag{2.4}$$

where the $u_{id} \in L^2(\mathcal{O}_i)^N$ are given functions and μ and a are positive constants.

The first multi-objective control problem considered in this work is the following:

> Find a *Pareto equilibrium* associated to (2.3) and (2.4), that is, a control $\hat{f} \in L^2(\omega)^N$ such that there is no f satisfying
>
> $$\begin{cases} f \in L^2(\omega)^N,\ J_1(f) \leq J_1(\hat{f}) \text{ and } J_2(f) \leq J_2(\hat{f}), \\ \text{with strict inequality for at least one } J_i. \end{cases} \tag{2.5}$$

In this framework, since the control-to-state mapping is linear and the cost functionals J_i are quadratic and strictly convex, it is not difficult to prove that $\hat{f}$ is a Pareto equilibria if and only if

$$\exists \alpha \in [0, 1] \text{ such that } \ \alpha J_1'(\hat{f}) + (1 - \alpha) J_2'(\hat{f}) = 0. \tag{2.6}$$

This is a consequence of the Karash-Kuhn-Tucker Theorem; see for instance [2].

In the sequel, the following notation will be used:

$$J_{(\alpha)} := \alpha J_1 + (1 - \alpha) J_2 \quad \text{for any } \alpha \in [0, 1].$$

2.2.2. Existence and characterization of Pareto equilibria

For future purposes, note that the J_i are $\mathcal{C}^1$ and, also,

$$(J_i'(f), g) = \int_\omega (a\varphi_i + \mu f) g \quad \forall f, g \in L^2(\omega)^N, \tag{2.7a}$$

where φ_i is the i-th adjoint state for f, i.e. the solution to

$$\begin{cases} -\Delta \varphi_i = (u - u_{id}) 1_{\mathcal{O}_i}, \quad x \in \Omega, \\ \varphi_i = 0, \quad x \in \partial\Omega. \end{cases} \tag{2.7b}$$

Here, u is the state associated to f.

We can now present the main result of this section. It is related to the existence and characterization of Pareto equilibria.

Theorem 2.2.1. *Let us assume that $\hat{f} \in L^2(\omega)^N$. Then*

1. *$\hat{f}$ is a Pareto equilibria if only if there exists $\alpha \in [0,1]$, $\hat{u}$ and $\hat{\varphi}$ such that*

$$\begin{cases} -\Delta \hat{u} = \hat{f} 1_\omega, \quad x \in \Omega, \\ \hat{u} = 0, \quad x \in \partial\Omega, \\ -\Delta \hat{\varphi} = \alpha(\hat{u} - u_{1d}) 1_{\mathcal{O}_1} + (1-\alpha)(\hat{u} - u_{2d}) 1_{\mathcal{O}_2}, \quad x \in \Omega, \\ \hat{\varphi} = 0, \quad x \in \partial\Omega, \\ \hat{f} = -\dfrac{a}{\mu}\, \hat{\varphi}|_\omega \,. \end{cases} \tag{2.8}$$

2. *For all $\alpha \in [0,1]$ there exists exactly one solution to* (2.8)*. Consequently, there exists a family $\{\hat{f}_\alpha\}_{\alpha\in[0,1]}$ of Pareto equilibria associated to* (2.3) *and* (2.4)*.*

Proof:

Let us first assume that $\hat{f} \in L^2(\omega)^N$ is a Pareto equilibria. Then, (2.6) holds. In view of (2.7a) and (2.7b), one must have

$$\hat{f} = -\frac{a}{\mu} \left(\alpha \hat{\varphi}_1 + (1-\alpha)\hat{\varphi}_2 \right) \Big|_\omega$$

for some $\alpha \in [0,1]$ (here $\hat{\varphi}_i$ solves (2.7b) for $u = \hat{u}$). Consequently, (2.8) is satisfied with $\hat{\varphi} = \alpha\hat{\varphi}_1 + (1-\alpha)\hat{\varphi}_2$.

Conversely, if $\hat{f}, \hat{u}$ and $\hat{\varphi}$ satisfy (2.8) for some $\alpha \in [0,1]$, then $J'_{(\alpha)}(\hat{f}) = 0$. Indeed, it is clear from (2.7a) that, for any $f, g \in L^2(\omega)^N$, one has

$$(J'_{(\alpha)}(f), g) = \int_\omega (a\varphi + \mu f) g,$$

where φ solves

$$\begin{cases} -\Delta \varphi = \alpha(u - u_{1d}) 1_{\mathcal{O}_1} + (1-\alpha)(u - u_{2d}) 1_{\mathcal{O}_2}, \quad x \in \Omega, \\ \varphi = 0, \quad x \in \partial\Omega. \end{cases}$$

Therefore, (2.6) holds and $\hat{f}$ is a Pareto equilibrium.

For each $\alpha \in [0,1]$, (2.6) and (2.8) posesses a unique solution, since this system is equivalent to the identity $J'_{(\alpha)}(\hat{f}) = 0$ and $J_{(\alpha)} : L^2(\omega) \mapsto \mathbb{R}$ is strictly convex, $\mathcal{C}^1$ and coercive. □

2.2.3. Algorithms and convergence

We will recall in this section three standard algorithms that can be used for the computation of Pareto equilibria.

In the sequel, we assume that α is fixed in $[0,1]$ and we try to solve (2.8); equivalently, we try to find the unique minimizer of $J_{(\alpha)}$ in $L^2(\omega)$.

ALG 1: Fixed-Point

(a) Choose $f^0 \in L^2(\omega)^N$.

(b) Then, for given $n \geq 0$ and $f^n \in L^2(\omega)^N$, compute the solution u^n to

$$\begin{cases} -\Delta u^n = f^n 1_\omega, \quad x \in \Omega, \\ u^n = 0, \quad x \in \partial\Omega \end{cases} \tag{2.9}$$

and the solution φ^n to the system

$$\begin{cases} -\Delta \varphi^n = \alpha(u^n - u_{1d})1_{\mathcal{O}_1} + (1-\alpha)(u^n - u_{2d})1_{\mathcal{O}_2}, \quad x \in \Omega, \\ \varphi^n = 0, \quad x \in \partial\Omega \end{cases} \tag{2.10}$$

and, finally, take

$$f^{n+1} = -\frac{a}{\mu}\, \varphi^n|_\omega\,. \tag{2.11}$$

ALG 2: Optimal Step Gradient Method

(a) Choose $f^0 \in L^2(\omega)^N$.

(b) Then, for given $n \geq 0$ and $f^n \in L^2(\omega)^N$, compute the solution u^n to (2.9) and the solution φ^n to (2.10) and take

$$f^{n+1} = f^n - \rho^n g^n, \tag{2.12}$$

where

$$g^n = a\, \varphi^n|_\omega + \mu f^n \tag{2.13}$$

and

$$\rho^n = \arg\Big(\min_{\rho \geq 0} J_{(\alpha)}(f^n - \rho g^n)\Big). \tag{2.14}$$

ALG 3: Optimal Step Conjugate Gradient Method

(a) Choose $f^0 \in L^2(\omega)^N$.

(b) For $n = 0$, perform one step of **ALG 2** and take $d^0 = g^0$.

(c) Then, for given $n \geq 1$ and $f^n \in L^2(\omega)^N$ compute the solution u^n to (2.9) and the solution φ^n to (2.10) and take

$$f^{n+1} = f^n - \rho^n d^n, \tag{2.15}$$

where

$$\begin{cases} d^n = g^n + \gamma^n d^{n-1}, \gamma^n = \dfrac{\|g^n\|^2}{\|g^{n-1}\|^2}, \\ g^n = a\, \varphi^n|_\omega + \mu f^n \end{cases} \tag{2.16}$$

and

$$\rho^n = \arg\Big(\min_{\rho \geq 0} J_{(\alpha)}(f^n - \rho d^n)\Big). \tag{2.17}$$

The following convergence results hold:

Theorem 2.2.2. *Let $\alpha \in (0,1)$ be given and let us denote by $\hat{f}$ the associated Pareto equilibrium. There exists $\varepsilon_0 = \varepsilon_0(\Omega,\omega) > 0$ such that, if $a/\mu \leq \varepsilon_0$, the controls f^n furnished by* **ALG 1** *satisfy $f^n \to \hat{f}$ as $n \to +\infty$. Furthermore, in such case, the speed of convergence is at least linear.*

This proof is very easy. It suffices to observe that, if ε_0 is small enough and $a/\mu \leq \varepsilon_0$, then the mapping $\Psi : L^2(\omega) \mapsto L^2(\omega)$ given by $\Psi(f^n) = f^{n+1}$ is a well defined contraction.

Theorem 2.2.3. *Let α and $\hat{f}$ be as in Theorem 2.2.2 and let the f^n be controls furnished by* **ALG 2**. *Then, $f^n \to \hat{f}$ as $n \to +\infty$.*

Theorem 2.2.4. *The assertion in Theorem 2.2.3 also holds for the controls f^n furnished by* **ALG 3**.

Note that the controls f^n furnished by **ALG 2** and **ALG 3** converge independently of the size of a/μ. Theorems 2.2.3 and 2.2.4 are consequences of classical convergence properties of the optimal step gradient and conjugate gradient algorithms (since the functional $J_{(\alpha)}$ is elliptic and satisfies Polak's condition; see [2]).

2.3. The case of a semilinear elliptic PDE

2.3.1. Pareto equilibria and quasi-equilibria

Let us assume that

$$\begin{cases} \phi : \mathbb{R} \mapsto \mathbb{R} \text{ is continuously differentiable,} \\ \phi'(s) \geq 0 \text{ and } |\phi(s)| \leq C + C|s| \quad \forall s \in \mathbb{R}. \end{cases} \tag{2.18}$$

In this section, the state equation will be the following:

$$\begin{cases} -\Delta u + \phi(u) = f 1_\omega, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega. \end{cases} \tag{2.19}$$

It is well known that, for each $f \in L^2(\omega)^N$, there exists exactly one solution u to (2.19). As in Section 2, we will consider the cost functionals J_i in (2.4), where the $u_{id} \in L^2(\mathcal{O}_i)^N$ and $a, \mu > 0$.

This section is devoted to introduce Pareto optima in this semilinear context. Now, we will have to distinguish equilibria from quasi-equilibria and take into account the particularities of each of them.

We will begin with some definitions:

Definition 2.3.1. *It will be said that $\hat{f}$ is a Pareto equilibrium for* (2.19) *and* (2.4) *if there is no $f \in L^2(\omega)^N$ satisfying*

$$\begin{cases} f \in L^2(\omega)^N,\ J_1(f) \leq J_1(\hat{f}) \text{ and } J_2(f) \leq J_2(\hat{f}), \\ \text{with strict inequality for at least one } J_i. \end{cases} \tag{2.20}$$

It is not difficult to prove (and in fact it is well known in control theory, see for instance [15,20]) that, under the previous assumptions on ϕ, the cost functionals J_i are $\mathcal{C}^1$ and satisfy

$$(J_i'(f), g) = \int_\omega (a\varphi_i + \mu f)g \quad \forall f, g \in L^2(\omega)^N,$$

where φ_i is the unique solution to

$$\begin{cases} -\Delta\varphi_i + \phi'(u)\varphi_i = (u - u_{id})1_{\mathcal{O}_i}, & x \in \Omega, \\ \varphi = 0, & x \in \partial\Omega \end{cases}$$

(that is, the i-th adjoint state corresponding to f) and u is the solution of (2.19).

Definition 2.3.2. *It will be said that $\hat{f}$ is a Pareto quasi-equilibrium for* (2.19) *and* (2.4) *if $\hat{f}$ satisfies* (2.6)*, that is to say, there exists $\alpha \in [0,1]$ such that $\hat{f}$ solves, together with $\hat{u}$ and $\hat{\varphi}$, the optimality system*

$$\begin{cases} -\Delta\hat{u} + \phi(\hat{u}) = \hat{f}1_\omega, & x \in \Omega, \\ \hat{u} = 0, & x \in \partial\Omega, \\ -\Delta\hat{\varphi} + \phi'(\hat{u})\hat{\varphi} = \alpha(\hat{u} - u_{1d})1_{\mathcal{O}_1} + (1-\alpha)(\hat{u} - u_{2d})1_{\mathcal{O}_2}, & x \in \Omega, \\ \hat{\varphi} = 0, & x \in \partial\Omega, \\ \hat{f} = -\dfrac{a}{\mu}\,\hat{\varphi}\,|\,\omega. \end{cases} \tag{2.21}$$

Note that, if $\hat{f}$ is a Pareto equilibrium, then $\hat{f}$ is also a Pareto quasi-equilibrium, in view of Karush-Kuhn-Tucker Theorem. However, the converse is not necessarily true.

Theorem 2.3.3. *Let us assume that $N \leq 8$ and, together with* (2.18)*, one has $\phi \in \mathcal{C}^2(\mathbb{R})$, with $|\phi'| + |\phi''| \leq C$. There exists ε, only depending on $\Omega, u_{1d}, u_{2d}, \|\phi\|_{W^{2,\infty}}$ and $\|\hat{f}\|$, such that, if $a/\mu \leq \varepsilon$, then the following assertions are equivalent:*

(a) $\hat{f}$ is a Pareto equilibrium for (2.19) *and* (2.4)*.*

(b) $\hat{f}$ is a Pareto quasi-equilibrium for (2.19) *and* (2.4)*.*

Proof:

We have to prove that, under the previous conditions, if $\hat{f}$ is a Pareto quasi-equilibrium, there is no other control $f \in L^2(\omega)^N$ satisfying (2.21). Thus, let us assume that (2.6) holds.

Let $\alpha \in [0,1]$ be given. The function $J_{(\alpha)} := \alpha J_1 + (1-\alpha)J_2$ is now twice continuously differentiable. Let us see that, if a/μ is sufficiently small, then $J''_{(\alpha)}(\hat{f}; g, g) > 0$ for all nonzero $g \in L^2(\omega)^N$.

We know that

$$(J'_{(\alpha)}(f), g)_{L^2(\omega)} = \int_\omega (a\varphi + \mu f)g \quad \forall f, g \in L^2(\omega)^N, \tag{2.22}$$

where φ is, together with u, the solution to

$$\begin{cases} -\Delta u + \phi(u) = f1_\omega, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \\ -\Delta\varphi + \phi'(u)\varphi = \alpha(u - u_{1d})1_{\mathcal{O}_1} + (1-\alpha)(u - u_{2d})1_{\mathcal{O}_2}, \quad x \in \Omega, \\ \varphi = 0, \quad x \in \partial\Omega. \end{cases} \tag{2.23a}$$

Let $g \in L^2(\omega)^N$ be given. For any small $\varepsilon > 0$, let us introduce u^ε and φ^ε with

$$\begin{cases} -\Delta u^\varepsilon + \phi(u^\varepsilon) = (\hat{f} + \varepsilon g)1_\omega, \quad x \in \Omega, \\ u^\varepsilon = 0, \quad x \in \partial\Omega, \\ -\Delta\varphi^\varepsilon + \phi'(u^\varepsilon)\varphi^\varepsilon = \alpha(u^\varepsilon - u_{1d})1_{\mathcal{O}_1} + (1-\alpha)(u^\varepsilon - u_{2d})1_{\mathcal{O}_2}, \quad x \in \Omega, \\ \varphi^\varepsilon = 0, \quad x \in \partial\Omega. \end{cases} \tag{2.23b}$$

Let us put

$$z := \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}(u^\varepsilon - \hat{u}) \quad \text{and} \quad \psi := \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}(\varphi^\varepsilon - \hat{\varphi})$$

(recall that $(\hat{f}, \hat{u}, \hat{\varphi})$ solves (2.21)). Note that these limits exist in $H_0^1(\Omega)$. This is easy to see by substracting (2.23a) written for $f = \hat{f}$ from (2.23b), dividing by ε and letting $\varepsilon \to 0$ (here, we must use that $\phi \in \mathcal{C}^2(\mathbb{R})$ and ϕ' and ϕ'' are uniformly bounded). Furthermore, one has

$$\begin{cases} -\Delta z + \phi'(\hat{u})z = g1_\omega, \quad x \in \Omega, \\ z = 0, \quad x \in \partial\Omega, \\ -\Delta\psi + \phi'(\hat{u})\psi + \phi''(\hat{u})z\hat{\varphi} = z(\alpha 1_{\mathcal{O}_1} + (1-\alpha)1_{\mathcal{O}_2}), \quad x \in \Omega, \\ \psi = 0, \quad x \in \partial\Omega. \end{cases} \tag{2.24}$$

Therefore,

$$J''_{(\alpha)}(\hat{f}; g, g) = \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\Big(J'_{(\alpha)}(\hat{f} + \varepsilon g) - J'_{(\alpha)}(\hat{f}), g\Big) = \int_\omega (a\psi + \mu g)g. \tag{2.25}$$

Observe that, from elliptic regularity, one has

$$\|\hat{\varphi}\|_{H^2} \le C_1(1 + \|\hat{u}\|) \quad \text{and} \quad \|\nabla\hat{u}\| \le C_2\|\hat{f}\|, \tag{2.26}$$

for some constants $C_1 = C_1(u_{1d}, u_{2d}, \Omega)$ and $C_2 = C_2(\Omega)$. On the other hand, from the PDEs satisfied by ψ and z, one also has

$$\|\nabla\psi\| \le C_3(1 + \|\hat{\varphi}\|_{L^{N/2}})\|\nabla z\| \quad \text{and} \quad \|\nabla z\| \le C_2\|g\|, \tag{2.27}$$

where $C_3 = C_3(\|\phi\|_{W^{2,\infty}}, \Omega)$. Consequently, taking into account that $H^2(\Omega) \hookrightarrow L^{N/2}(\Omega)$ for $N \le 8$ with continuous embedding, we see from (2.26) and (2.27) that

$$J''_{(\alpha)}(\hat{f}; g, g) \ge \mu\|g\|^2 - aC_4(1 + \|\hat{f}\|)\|g\|^2$$

for some $C_4 = C_4(u_{1d}, u_{2d}, \|\phi\|_{W^{2,\infty}}, \Omega)$.

Clearly, this proves that, if $N \le 8$ and a/μ is sufficiently small, $J''_{(\alpha)}(\hat{f}, g, g) > 0$ for all $g \ne 0$, whence $J_{(\alpha)}$ possesses a unique global minimum at $\hat{f}$.

In other words, $\hat{f}$ is a Pareto equilibrium for (2.19) and (2.4). □

2.3.2. Existence of Pareto equilibria

We can now prove the existence of Pareto equilibria for (2.19) and (2.4).

Theorem 2.3.4. *Let us assume that* (2.18) *is satisfied. There exists a family* $\{f_\alpha\}_{\alpha\in(0,1)}$ *of Pareto equilibria for* (2.19) *and* (2.4).

The proof is not difficult. It suffices to note that, for each $\alpha \in (0,1)$, there exists at least one minimizer of $J_{(\alpha)}$ in $L^2(\omega)$, in view of the properties of ϕ. Any such minimizer is clearly a Pareto optimum. Note however that this cannot be ensured if $\alpha = 0$ or $\alpha = 1$.

Let us remark that, in this result, the uniqueness of the minimizer f_α is not guaranteed. To ensure that there is at most one solution we must impose more conditions to ϕ.

Theorem 2.3.5. *Let us assume that ϕ is as in Theorem 2.3.4, $N \le 8$ and, moreover, $\phi \in W^{2,\infty}(\mathbb{R})$ and $\phi' \ge 0$. There exists $\chi = \chi(\Omega)$ such that, if $a/\mu \le \chi$, for each $\alpha \in (0,1)$, the minimizer of $J_{(\alpha)}$ furnished by Theorem 2.3.4 is unique.*

Proof:

In order to fix ideas, let us assume that $3 \le N \le 8$ (the case $N = 2$ is similar and even easier).

Let $\alpha \in (0,1)$ be given and let us assume that there exists two minimizers f^1 and f^2 of $J_{(\alpha)}$ in $L^2(\omega)$. Then they solve the following systems for $j = 1$ and 2:

$$\begin{cases} -\Delta u^j + \phi(u^j) = f^j 1_\omega, \quad x \in \Omega, \\ u^j = 0, \quad x \in \partial\Omega, \\ -\Delta \varphi_i^j + \phi'(u^j)\varphi_i^j = (u^j - u_{id})1_{\mathcal{O}_i}, \quad x \in \Omega, \\ \varphi_i^j = 0, \quad x \in \partial\Omega. \\ f^j = -\dfrac{a}{\mu}\Big(\alpha\varphi_1^j + (1-\alpha)\varphi_2^j\Big)1_\omega \end{cases}$$

Let us set $f := f^1 - f^2$, $\varphi_i := \varphi_i^1 - \varphi_i^2$ and $u := u^1 - u^2$. Then, we have

$$\begin{cases} -\Delta u + \phi'(\tilde{u})u = f1_\omega, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \\ -\Delta \varphi_i + \phi'(u^1)\varphi_i + \phi''(\overline{u})\varphi_i^2 u = u1_{\mathcal{O}_i}, \quad x \in \Omega, \\ \varphi_i = 0, \quad x \in \partial\Omega. \\ f = -\dfrac{a}{\mu}\Big(\alpha\varphi_1 + (1-\alpha)\varphi_2\Big)1_\omega, \end{cases} \tag{2.28}$$

where, for each x, one has $\tilde{u}(x) = \beta(x)u^1(x) + (1-\beta(x))u^2(x)$ and $\overline{u}(x) = \lambda(x)u^1(x) + (1-\lambda(x))u^2(x)$ for some $\beta(x), \lambda(x) \in (0,1)$. From the properties of ϕ, we deduce that

$$\int_\Omega \Big(|\nabla u|^2 + \phi'(\tilde{u})|u|^2\Big)dx \ge (f,u).$$

Therefore, from the Cauchy-Schwarz and Young inequalities, we see that

$$\|\nabla u\|^2 \le C\frac{a^2}{\mu^2}(\|\varphi_1\|^2 + \|\varphi_2\|^2), \tag{2.29}$$

for some positive $C = C(\Omega)$.

By a similar reason, denoting by 2^* the Sobolev embedding exponent of $H^1(\Omega)$ and $(2^*)'$ its conjugate, that is, $2^* = 2N/(N-2)$ and $(2^*)' = 2N/(N+2)$, we also have:

$$\begin{aligned}\|\nabla\varphi_i\|^2 &\leq C\|\varphi_i^2 u\|_{L^{(2^*)'}}\|\varphi_i\|_{L^{(2^*)}} + C\|u\|\|\varphi_i\| \\ &\leq \frac{1}{2}\|\nabla\varphi_i\|^2 + C\|u\|^2 + C\|\varphi_i^2\|^2_{L^{N/2}}\|u\|^2_{L^{2^*}} \\ &\leq \frac{1}{2}\|\nabla\varphi_i\|^2_{L^{N/2}} + C\Big(1+\|\varphi_i^2\|^2\Big)\|\nabla u\|^2,\end{aligned}$$

where again $C = C(\Omega)$.

Consequently,

$$\|\nabla u\|^2 \leq C\frac{a^2}{\mu^2}(1 + \|\varphi_1^2\|^2_{L^{N/2}} + \|\varphi_2^2\|^2_{L^{N/2}})\|\nabla u\|^2. \tag{2.30}$$

For $N \leq 8$, one has $N/2 \leq 2N/(N-4)$. Accordingly, $L^{N/2}(\Omega) \hookrightarrow H^2(\Omega)$ and, from the usual elliptic estimates, the following is found:

$$\|\varphi_i^2\|^2_{L^{N/2}} \leq C\|(u^2 - u_{id})1_{\mathcal{O}_i}\|^2 \leq C(1+\|u^2\|^2) \leq C\Big(1+\frac{a^2}{\mu^2}(\|\varphi_1^2\|^2+\|\varphi_2^2\|^2)\Big), \quad i = 1, 2. \tag{2.31}$$

This indicates that, if a/μ is sufficiently small, $\|\varphi_1^2\|^2_{L^{N/2}} + \|\varphi_2^2\|^2_{L^{N/2}} \leq C$ and, coming back to (2.30), $u = 0$. Thus, in this case, we necessarily have $f = 0$ and the proof is done. □

2.3.3. Algorithms and convergence

In this section, we present some iterative algorithms, similar to those considered in the linear case. To fix ideas, it will be assumed that $\alpha \in (0, 1)$ and $\mu > 0$ are fixed and we will seach for a sequence $\{f^n\}$ of approximations to a minimizer of $J_{(\alpha)}$.

ALG 4: Fixed point

(a) Choose $f^0 \in L^2(\omega)^N$.

(b) Then, for given $n \geq 0$ and $f^n \in L^2(\omega)^N$, compute the solution u^n to

$$\begin{cases} -\Delta u^n + \phi(u^n) = f^n 1_\omega, & x \in \Omega, \\ u^n = 0, & x \in \partial\Omega, \end{cases} \tag{2.32}$$

the solution φ^n to the system

$$\begin{cases} -\Delta\varphi^n + \phi'(u^n)\varphi^n = \alpha(u^n - u_{1d})1_{\mathcal{O}_1} + (1-\alpha)(u^n - u_{2d})1_{\mathcal{O}_2}, & x \in \Omega, \\ \varphi^n = 0, & x \in \partial\Omega \end{cases} \tag{2.33}$$

and, finally, take

$$f^{n+1} = -\frac{a}{\mu}\,\varphi^n|_\omega\,. \tag{2.34}$$

ALG 5: Optimal Step Gradient Method

(a) Choose $f^0 \in L^2(\omega)^N$.

(b) Then, for given $n \geq 0$ and $f^n \in L^2(\omega)^N$, compute the solution u^n to (2.32) and the solution φ^n to (2.33) and take

$$f^{n+1} = f^n - \rho^n g^n, \tag{2.35}$$

where

$$g^n = a\ \varphi^n|_\omega + \mu f^n \tag{2.36}$$

(g^n is the gradient of $J_{(\alpha)}$ at u^n) and

$$\rho^n = \arg\Big(\min_{\rho\geq 0} J_{(\alpha)}(f^n - \rho g^n)\Big). \tag{2.37}$$

ALG 6: Optimal Step Conjugate Gradient Method

(a) Choose $f^0 \in L^2(\omega)^N$.

(b) For $n = 0$, perform one step of **ALG 5** and take $d^0 = g^0$.

(c) Then, for given $n \geq 1$, $f^n \in L^2(\omega)^N$, $g^{n-1}, d^{n-1} \in L^2(\omega)^N$, compute the solution u^n to (2.32) and the solution φ^n to (2.33) and take

$$f^{n+1} = f^n - \rho^n d^n, \tag{2.38}$$

where

$$\begin{cases} d^n = g^n + \gamma^n d^{n-1}, \quad \gamma^n = \dfrac{(g^n - g^{n-1}, g^n)}{\|g^{n-1}\|^2}, \\ g^n = a\ \varphi^n|_\omega + \mu f^n \end{cases} \tag{2.39}$$

and

$$\rho^n = \arg\Big(\min_{\rho\geq 0} J_{(\alpha)}(f^n - \rho d^n)\Big). \tag{2.40}$$

Note that in **ALG 3** and **ALG 6**, the coefficient γ^n is given by different expressions. The reason is that, now, the system is nonlinear and we must impose Polak condition to ensure convergence.

Theorem 2.3.6. *Let the assumptions in Theorem 2.3.5 be satisfied and let the controls f^n be furnished by* **ALG 4**. *Then, $f^n \to \hat{f}$ as $n \to +\infty$.*

This proof is easy. Indeed, as in the linear case, if a/μ is sufficiently small, then we can write that $f^{n+1} = \Psi(f^n)$ for all $n \geq 0$, where Ψ is a contraction.

Theorem 2.3.7. *Let the assumptions in Theorem 2.3.5 be satisfied, let a/μ small enough and let the controls f^n be furnished by* **ALG 5**. *Then, $f^n \to \hat{f}$ as $n \to +\infty$.*

The proof can be obtained arguing as in the proof of Theorem 8.4-3 in [10]. Indeed, from the proof of Theorem 2.3.5, we deduce that, if a/μ is sufficiently small, then $J_{(\alpha)}$ is elliptic, that is,

$$(J'_{(\alpha)}(f_1) - J'_{(\alpha)}(f_2), f_1 - f_2) \geq c\|f_1 - f_2\|^2_{L^2(\omega)} \quad \forall f_1, f_2 \in L^2(\omega)$$

for some $c > 0$. Consequently,

$$J_{(\alpha)}(f^n) - J_{(\alpha)}(f^{n+1}) \geq \frac{c}{2}\|f^n - f^{n+1}\|^2_{L^2(\omega)} \quad \text{and} \quad \|J'_{(\alpha)}(f^n)\|_{L^2(\omega)} \leq \|J'_{(\alpha)}(f^n) - J'_{(\alpha)}(f^{n+1})\|_{L^2(\omega)}$$

for all $n \geq 1$, whence in particular we have that $\|f^n - f^{n+1}\|_{L^2(\omega)} \to 0$ as $n \to +\infty$. Taking into account the expression of $J'_{(\alpha)}$, we also have that $\|J'_{(\alpha)}(f^n) - J'_{(\alpha)}(f^{n+1})\|_{L^2(\omega)} \to 0$, whence $J'_{(\alpha)}(f^n) \to 0$ and at least a subsequence of $\{f^n\}$ converges weakly towards the unique minimizer $\hat{f}$ of $J_{(\alpha)}$. Since

$$\|f^n - \hat{f}\|_{L^2(\omega)} \leq \frac{1}{c}\|J'_{(\alpha)}(f^n)\|_{L^2(\omega)} \quad \forall n \geq 1,$$

we see that, in fact, the whole sequence converges strongly to $\hat{f}$.

Theorem 2.3.8. *The assertion in Theorem 2.3.7 also holds for the controls f^n furnished by* **ALG 6**.

For the proof, we can use the arguments in p.96–98 in [24] (Theorems 1.5.8 and 1.5.9). More precisely, note first that there exists $\beta > 0$ such that

$$\left(J'_{(\alpha)}(f^n), d^n\right) \geq \beta\|J'_{(\alpha)}(f^n)\|_{L^2(\omega)}\|d^n\|_{L^2(\omega)} \quad \forall n \geq 1.$$

Therefore,

$$J_{(\alpha)}(f^{n+1}) - J_{(\alpha)}(f^n) \leq -C\|J'_{(\alpha)}(f^n)\|_{L^2(\omega)}$$

and $J'_{(\alpha)}(f^n) \to 0$ as $n \to +\infty$. Thus, we again have that subsequence of f^n converges weakly to $\hat{f}$ and arguing as before, we are led to the strong convergence of the whole sequence.

Note that, now, we must impose for the three algorithms conditions of the same kind to prove convergence. Of course, this is due to the fact that (2.19) is nonlinear.

2.4. The stationary Navier-Stokes system

This section is devoted to the existence and characterization of Pareto equilibria and quasi-equilibria for the stationary Navier-Stokes equations. As expected, in view of the features of the state system and, in particular, the possible lack of uniqueness, this will be more complicated than in Sections 2 and 3.

The stationary Navier-Stokes equations are the following:

$$\begin{cases} -\nu\Delta u + (u \cdot \nabla)u + \nabla p = f1_\omega, \quad x \in \Omega, \\ \nabla \cdot u = 0, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega. \end{cases} \tag{2.41}$$

The state variables are $u = (u_1, \ldots, u_N)$ and p. They can be viewed as the velocity field and the pressure of a viscous Newtonian fluid. The control is $f1_\omega$ and can viewed as a field of external forces applied at the points in ω.

2.4.1. Pareto equilibria and quasi-equilibria

The positive constant ν is the kinematic viscosity of the fluid. It must be regarded as a measure of "thickness", that is, tendency to favor friction.

As in the previous sections, let us introduce the functionals

$$J_i(f,u) := \frac{a}{2}\int_{\mathcal{O}_i} |u-u_{id}|^2 + \frac{\mu}{2}\int_\omega |f|^2, \quad i=1,2, \tag{2.42}$$

where the $u_{id} \in L^2(\mathcal{O}_i)^N$ and $a, \mu > 0$.

Note that, here, contrarily to the cost functionals in Sections 2 and 3, we assume that the J_i depend not only on the control f but also on the state u. This is due to the possible nonuniqueness of solution to (2.41), that may appear when ν is not sufficiently large.

Definition 2.4.1. *It will be said that $f \in L^2(\omega)^N$ is a Pareto equilibrium for* (2.41) *and* (2.42) *if there exists an associated state (u,p) such that there is no triplet (f',u',p'), where $f' \in L^2(\omega)^N$ is a control and (u',p') is an associated state, satisfying*

$$\begin{cases} J_1(f',u') \le J_1(f,u) \text{ and } J_2(f',u') \le J_2(f,u), \\ \text{with strict inequality for at least one } J_i. \end{cases} \tag{2.43}$$

Definition 2.4.2. *It will be said that $f \in L^2(\omega)^N$ is a Pareto quasi-equilibrium for* (2.41) *and* (2.42) *if there exists $\alpha \in [0,1]$, such that f solves, together with some (u,p) and (φ,q), the following coupled system*

$$\begin{cases} -\nu\Delta u + (u\cdot\nabla)u + \nabla p = f|_\omega\,, \quad x\in\Omega, \\ \nabla\cdot u = 0, \quad x\in\Omega, \\ u = 0, \quad x\in\partial\Omega, \\ -\nu\Delta\varphi + (u\cdot\nabla)\varphi + (\nabla u)^t\varphi + \nabla q = \alpha(u-u_{1d})1_{\mathcal{O}_1} + (1-\alpha)(u-u_{2d})1_{\mathcal{O}_2}, \quad x\in\Omega, \\ \nabla\cdot\varphi = 0, \quad x\in\Omega, \\ \varphi = 0, \quad x\in\partial\Omega, \\ f = -\dfrac{a}{\mu}\varphi 1_\omega. \end{cases} \tag{2.44}$$

2.4.2. Existence of Pareto equilibria and quasi-equilibria

As already said, it is natural to expect that the proofs of existence of Pareto equilibria and quasi-equilibria be now more complicate.

Let us recall the definitions of some classical spaces, usual for the analysis of the Navier-Stokes equations:

$$H := \{v \in L^2(\Omega)^N : \ \nabla\cdot v = 0 \text{ in } \Omega,\ v\cdot n = 0 \text{ on } \partial\Omega\},$$
$$V := \{v \in H_0^1(\Omega)^N : \ \nabla\cdot v = 0 \text{ in } \Omega\}.$$

They are closed subspaces of $L^2(\Omega)^N$ and $H_0^1(\Omega)^N$, respectively; accordingly, they are Hilbert spaces for the scalar products $(\cdot\,,\cdot)$ and $(\cdot\,,\cdot)_{H_0^1}$. Also, we have the canonical compact embeddings $V \hookrightarrow H \equiv H' \hookrightarrow V'$ where X' denotes the dual space of X.

Let us consider the trilinear continuous forms $b(\cdot,\cdot,\cdot)$ and $c(\cdot,\cdot,\cdot)$, with

$$b(u,v,w) := \sum_{i,j=1}^{N} \int_\Omega u_i D_i v_j w_j \, dx \quad \forall u,v,w \in V$$

and

$$c(u,v,w) := b(v,u,w) = \sum_{i,j=1}^{N} \int_\Omega D_i u_j v_i w_j \, dx \quad \forall u,v,w \in V.$$

Note that there exist bilinear continuous mappings B and C, such that

$$\langle B(u,v), w\rangle = b(u,v,w) \quad \text{and} \quad \langle C(u,v), w\rangle = c(u,v,w) \quad \forall u,v,w \in V.$$

Recall that, for every $f \in L^2(\Omega)^N$, the nonlinear system (2.41) possesses at least one weak solution $(u,p) \in V \times L^2(\Omega)$, see for instance [26]. This solution is in fact strong, that is, $(u,p) \in H^2(\Omega)^N \times H^1(\Omega)$ and the PDEs in (2.41) are satisfied a.e. in Ω.

As in the previous sections, for any $\alpha \in [0,1]$, we will use the notation $J_{(\alpha)} := \alpha J_1 + (1-\alpha) J_2$.

Theorem 2.4.3. *For each $\alpha \in [0,1]$, there exists at least one solution to the following extremal problem:*

$$(P_\alpha) \begin{cases} \textit{Minimize } J_{(\alpha)}(f,u) \\ \textit{Subject to } f \in L^2(\omega)^N,\ (u,p) \textit{ solves } (2.41). \end{cases}$$

Consequently, there exists a whole family $\{f_\alpha\}_{\alpha\in(0,1)}$ of Pareto equilibria for (2.41) *and* (2.42)*.*

The proof is easy, since the J_i are coercive and lower semicontinuous for the weak convergence in $L^2(\omega)^N \times L^2(\Omega)^N$ and the set $\{(f,u) : f \in L^2(\omega)^N,\ (u,p) \text{ solves } (2.41)\}$ is nonempty and sequentially weakly closed in the same space.

Obviously, if (f,u) is a solution to (P_α) for some $\alpha \in (0,1)$, then f is a Pareto equilibrium. However, as in Section 3, this cannot be ensured if $\alpha = 0$ or $\alpha = 1$.

The characterization of Pareto equilibria is furnished by the following result:

Theorem 2.4.4. *Let f be a Pareto equilibrium for* (2.41) *and* (2.42)*. Then, f is a Pareto quasi-equilibrium.*

We will give a proof of this result that relies on the Dubovitsky-Milyoutin formalism (see [16]). To this purpose, we will first recall some technical results.

Lemma 2.4.5. *Let $K_1,\dots,K_n$ be convex cones in a Banach space X with apex at 0. For each i, we assume that either K_i is open or it is a closed subspace. Then the following conditions are equivalent:*

- $\bigcap_{i=1}^{n} K_i = \emptyset$.

- *There exist linear functionals $f_i \in K_i^*$ with $i = 1, \dots, n$, not all them zero, such that $\sum_{\ell=1}^n f_\ell = 0$.*

Here, for any i, we have denoted by K_i^* the dual cone to K_i, that is, $K_i^* := \{f \in X' : f(e) \geq 0 \ \forall e \in K_i\}$. For the proof, see for instance Lemma 5.11 of [16].

Lemma 2.4.6. *Let $f \in L^2(\omega)^N$ be given and let $u \in V$ be (together with p) an associated solution to* (2.41)*. Let $R : V \mapsto V$ be the linear mapping defined by $R\varphi := (\nu A)^{-1}(-D\varphi \cdot u)$, where A is the Stokes operator in V, that is,*

$$\langle Av, w\rangle = (v, w)_{H_0^1} \quad \forall v, w \in V.$$

Then,

$$\big[\varphi\big] := \| \varphi|_\omega \|_{L^2(\omega)} \tag{2.45}$$

is a norm in Ker$(Id + R)$.

Proof:

As already said, one has $u \in H^2(\Omega)^N, \nabla u \in L^6(\Omega)^{N\times N}$ and $u \in L^\infty(\Omega)^N$. Consequently, R is well defined and compact. We only need to prove that, for every $\varphi \in \text{Ker}(Id + R)$ with $\varphi|_\omega = 0$, one has $\varphi \equiv 0$.

Thus, let us assume that $\varphi \in V$, $\varphi + (\nu A)^{-1}(-D\varphi \cdot u)$ vanishes, that is,

$$\begin{cases} -\nu\Delta\varphi - D\varphi \cdot u + \nabla q = 0, \quad x \in \Omega \\ \nabla \cdot \varphi = 0, \quad x \in \Omega \\ \varphi = 0, \quad x \in \partial\Omega \end{cases}$$

for some q and $\varphi = 0$ a.e. in ω. Then, we can use the unique continuation property of the Stokes system with coefficients in L^∞ (see [11]) and deduce that, certainly, φ vanishes identically. □

Lemma 2.4.7. *Let f and u be as in Lemma 2.4.6. Let the (φ_n, ψ_n) be given in $V \times V$ with $\varphi_n|_\omega \to \varphi|_\omega$ in $L^2(\omega)$, $\psi_n := \varphi_n + R\varphi_n$ and $\psi_n \to \psi \in V$. Then $\|\varphi_n\|_V \leq C$ for some positive constant C independent of n.*

Proof:

First, note that, since R is a compact operator, $\dim(\text{Ker}(Id + R)) < +\infty$ and $\text{Rank}(Id + R)$ is closed, in view of Fredholm's Alternative Theorem (see for instance [6]).

Now, let $\tilde\varphi_n$ be, for each n, the unique function in $\text{Ker}(Id + R)$ satisfying

$$\|\varphi_n - \tilde\varphi_n\|_V = \inf_{\tilde\varphi \in \text{Ker}(Id+R)} \|\varphi_n - \tilde\varphi\|_V.$$

Then, $\psi_n = (\varphi_n - \tilde\varphi_n) + R(\varphi_n - \tilde\varphi_n)$.

Also, $\|\varphi_n - \tilde\varphi_n\|_V$ is bounded by a constant C. Indeed, if this is not the case, we can assume that

$$\|\varphi_n - \tilde\varphi_n\|_V = \text{dist}(\varphi_n, \text{Ker}(Id + R)) \to +\infty.$$

Let us introduce

$$\zeta_n := \frac{\varphi_n - \tilde\varphi_n}{\|\varphi_n - \tilde\varphi_n\|_V}.$$

Then

$$\zeta_n + R\zeta_n = \frac{\psi_n}{\|\varphi_n - \tilde{\varphi}_n\|_V}. \tag{2.46}$$

Since the $\|\zeta_n\|_V = 1$ and $\psi_n \to \psi$ in V, we see from (2.46) that $\zeta_n \to \zeta$ for some $\zeta \in \text{Ker}(Id + R)$. So,

$$\left\|\varphi_n - \left(\tilde{\varphi}_n + \zeta\|\varphi_n - \tilde{\varphi}_n\|_V\right)\right\|_V = \|\varphi_n - \tilde{\varphi}_n\|_V\|\zeta_n - \zeta\|_V \geq \|\varphi_n - \tilde{\varphi}_n\|_V$$

for all n. But this is an absurd, since $\zeta_n \to \zeta$ in V.

Finally, let us deduce that $\|\varphi_n\|_V \leq C$. Indeed, we can write that $\varphi_n = \tilde{\varphi}_n + \eta_n$, with

- $\|\eta_n\|_V = \|\varphi_n - \tilde{\varphi}_n\|_v \leq C$.
- $\|\tilde{\varphi}_n\|_V \leq C[\tilde{\varphi}_n]$ (because the $\tilde{\varphi}_n$ belong to a finite dimensional space) and $[\tilde{\varphi}_n] \leq [\varphi_n] + [\eta_n] \leq [\varphi_n] + C\|\eta_n\|_V \leq C$.

This ends the proof. □

Now, we can prove the main result in this section.

Proof of Theorem 2.4.4:

As announced, we will use the Dubovitsky-Milyoutin formalism (and, more precisely, Lemma 2.4.5). Thus, let f be a Pareto equilibrium for (2.41) and (2.42). There exists a weak solution (u, p) to (2.41) such that the couple (f, u) solves following problem:

$$\begin{cases} J_1(f,u) \leq J_1(f',u') \\ \forall (f',u') \in \mathcal{F} \text{ with } J_2(f',u') \leq E_2, \end{cases} \tag{2.47}$$

where $E_2 := J_2(f,u)$ and $\mathcal{F} := \{(f',u') \in L^2(\omega)^N \times V : (u',p') \text{ solves (2.41) with } f = f'\}$.

Now, we introduce some cones in $L^2(\omega)^N \times V$ associated to (2.47):

$$D_i := \{(h,w) :\ J_i'(f,u)(h,w) < 0\} \quad (i = 1,2) \quad \text{and} \quad T := \text{Ker}(M'(f,u)),$$

where $M'(f,u)$ is the derivative at (f,u) of the nonlinear mapping $M : L^2(\omega)^N \times V \mapsto V'$, given by $M(f,u) := \nu Au + B(u,u) - f\mathbb{1}_\omega$.

Clearly, M is continuosly differentiable,

$$M'(f,u)(h,w) = \nu Aw + B(u,w) + B(w,u) - h\mathbb{1}_\omega \quad \forall (h,w) \in L^2(\omega)^N \times V$$

and

$$M'(f,u)^*\varphi = (-\varphi|_\omega, \nu A\varphi - B(u,\varphi) + C(u,\varphi)) \quad \forall \varphi \in V.$$

Let us prove that $\text{Rank}(M'(u,u)^*)$ is closed. Indeed, if $(h,w) \in \overline{\text{Rank}(M'(f,u)^*)}$, there exist fields $\varphi_n \in V$ such that

$$-\varphi_n|_\omega \to h \quad \text{in } L^2(\omega),$$
$$\nu A\varphi_n - B(u,\varphi_n) + C(u,\varphi_n) \to w \quad \text{in } V'.$$

Let us introduce the linear mapping $S : V \mapsto V'$, with $S\varphi := \nu A\varphi - B(u,\varphi) + C(u,\varphi)$. It is not difficult to see that $S = \nu A \cdot (Id + R)$, where R is the linear operator introduced in Lemma

2.4.6. Therefore, by Fredholm's Alternative Theorem, we have that $\text{Rank}(S)$ is closed and, from Lemma 2.4.7, we find that the φ_n are uniformly bounded in V. As an inmediate consequence, we see that, at least for a subsequence, $\varphi_n \to \varphi$ weakly in V, $-\varphi|_\omega = h$ and $S\varphi = w$. In other words, $(h, w) \in \text{Rank}(M'(f,u)^*)$.

At this moment, we apply the Dubovitskiy-Milyoutin formalism to (2.47) and we deduce that the descent cone D_1 must be disjoint of the descent cone D_2 and the tangent space T (see [16]):

$$D_1 \cap D_2 \cap T = \emptyset.$$

In view of Lemma 2.4.5, there exist $(h_1', w_1') \in D_1^*, (h_2', w_2') \in D_2^*$ and $(h', w') \in T^*$, not all zero, such that

$$(h_1', w_1') + (h_2', w_2') + (h', w') = (0, 0). \tag{2.48}$$

Taking into account the definitions of D_1, D_2 and T, we see at once that

$$D_i^* = \{-\lambda J_i'(f,u) : \lambda \geq 0\}, \quad i = 1, 2,$$
$$T^* = (\text{Ker}(M'(f,u)))^* = \overline{\text{Rank}(M'(f,u)^*)} = \text{Rank}(M'(f,u)^*).$$

Hence, there must exist nonnegative λ_1 and λ_2, not both zero, such that

$$\langle (h_i', w_i'), (h, w) \rangle_{L^2 \times V', L^2 \times V} = \lambda_i \Big(a \int_{\mathcal{O}_i} (u - u_{id}) w + \mu \int_\omega f h \Big) \quad \forall (h, w) \in L^2(\omega)^N \times V, \quad i = 1, 2$$

and there must exist $\varphi \in V$ such that

$$h' = -\varphi|_\omega, \quad w' = \nu A\varphi - B(u, \varphi) + C(u, \varphi). \tag{2.49}$$

Consequently, dividing by $\lambda_1 + \lambda_2$ and redefining (h', w'), we see that (2.48) can be rewritten in the form

$$\begin{cases} a\Big(\alpha \displaystyle\int_{\mathcal{O}_1} (u - u_{1d}) w + (1 - \alpha) \int_{\mathcal{O}_2} (u - u_{2d}) w\Big) + \mu \displaystyle\int_\omega f h = \int_\omega h' h \; + \langle w', w \rangle_{V', V} \\ \forall (h, w) \in L^2(\omega)^N \times V \end{cases} \tag{2.50}$$

for some $\alpha \in [0, 1]$.

Now, taking $w = 0$ we find that $h' = \mu f$.

On the other hand, taking $h = 0$ and recalling (2.49), we see that $\varphi|_\omega = -\mu f$ and

$$\int_\Omega \Big(\nu \nabla\varphi \cdot \nabla w - (u \cdot \nabla)\varphi \cdot w + (\nabla u)^t \varphi \cdot w \Big) = a \int_\Omega \Big(\alpha(u - u_{1d}) 1_{\mathcal{O}_1} + (1 - \alpha)(u - u_{2d}) 1_{\mathcal{O}_2} \Big) w$$

for all $w \in V$. Therefore, φ solves, together with some $q \in L^2(\Omega)$, the linear system

$$\begin{cases} -\nu\Delta\varphi - (u \cdot \nabla)\varphi + (\nabla u)^t \varphi + \nabla q = a\Big(\alpha(u - u_{1d}) 1_{\mathcal{O}_1} + (1 - \alpha)(u - u_{2d}) 1_{\mathcal{O}_2} \Big), \quad x \in \Omega, \\ \nabla \cdot \varphi = 0, \quad x \in \Omega, \\ \varphi = 0, \quad x \in \partial\Omega \end{cases} \tag{2.51a}$$

and, furthermore,

$$f = -\frac{1}{\mu} \varphi|_\omega \tag{2.51b}$$

From (2.41), (2.51a) and (2.51b), we deduce that f is a Pareto quasi-equilibrium and the proof is done. □

2.4.3. Algorithms and numerical experiments

This section is devoted to present three iterative algorithms (similar to those above) for the computation of Pareto equilibria for (2.41) and (2.42). Additionally, we will present a new algorithm based on Newton's method.

As before, we fix $\alpha \in (0,1)$ and we look for a solution to (P_α).

ALG 7: Fixed-Point

(a) Choose $f^0 \in L^2(\omega)^N$ and $u^0 \in V$.

(b) Then, for given $n \geq 0$, $f^n \in L^2(\omega)^N$ and $u^n \in V$, compute the solution (u^{n+1}, p^{n+1}) to

$$\begin{cases} -\nu\Delta u^{n+1} + (u^n \cdot \nabla)u^{n+1} + \nabla p^{n+1} = f^n 1_\omega, \quad x \in \Omega, \\ \nabla \cdot u^{n+1} = 0, \quad x \in \Omega, \\ u^{n+1} = 0, \quad x \in \partial\Omega. \end{cases} \tag{2.52}$$

the solution (φ^{n+1}, q^{n+1}) to the system

$$\begin{cases} -\nu\Delta\varphi^{n+1} - D\varphi^{n+1} \cdot u^{n+1} + \nabla q^{n+1} \\ \qquad\qquad = \alpha\Big((u^{n+1} - u_{1d})1_{\mathcal{O}_1}\Big) + (1-\alpha)\Big((u^{n+1} - u_{2d})1_{\mathcal{O}_2}\Big), \quad x \in \Omega, \\ \nabla \cdot \varphi^{n+1} = 0, \quad x \in \Omega, \\ \varphi^{n+1} = 0, \quad x \in \partial\Omega \end{cases} \tag{2.53}$$

and, finally, set

$$f^{n+1} = -\frac{a}{\mu}\,\varphi^{n+1}\big|_\omega\,. \tag{2.54}$$

ALG 8: Optimal Step Gradient Method

(a) Choose $f^0 \in L^2(\omega)^N$.

(b) Then, for given $n \geq 0$ and $f^n \in L^2(\omega)^N$, compute the solution (u^{n+1}, p^{n+1}) to (2.52) and the solution (φ^{n+1}, q^{n+1}) to (2.53) and set

$$f^{n+1} = f^n - \rho^n g^{n+1}, \tag{2.55}$$

where

$$g^{n+1} = a\,\varphi^{n+1}\big|_\omega + \mu f^n \tag{2.56}$$

and

$$\rho^n = \arg\Big(\min_{\rho>0} J_{(\alpha)}(f^n - \rho g^{n+1})\Big). \tag{2.57}$$

ALG 9: Optimal Step Conjugate Gradient Method

(a) Choose $f^0 \in L^2(\omega)^N$.

(b) Then, for $n = 0$, perform one step of **ALG 8** and set $d^0 = g^0$.

(c) Then, for given $n \geq 1$ and $f^n \in L^2(\omega)^N$, compute the solution (u^{n+1}, p^{n+1}) to (2.52), the solution (φ^{n+1}, q^{n+1}) to (2.53) and then set

$$f^{n+1} = f^n - \rho^n d^n, \tag{2.58}$$

where

$$\begin{cases} d^n = g^{n+1} + \gamma^n d^{n-1}, \quad \gamma^n = \dfrac{(g^n - g^{n-1}, g^n)}{\|g^{n-1}\|^2} \\ g^{n+1} = a\ \varphi^{n+1}|_\omega + \mu f^n, \end{cases} \tag{2.59}$$

and

$$\rho^n = \arg\Big(\min_{\rho>0} J_{(\alpha)}(f^n - \rho d^n)\Big). \tag{2.60}$$

Before presenting the results of some simulations, we will consider another algorithm. It is based on Newton's method and aims to compute a solution to the optimality system (2.44). In practice, this method is much faster than the **ALG 8** and **ALG 9** but, as is usual for Newton methods and variants, it needs a nontrivial starting process (see below).

ALG 10: Newton Method

We want to solve the problem (2.44) with $\nu = \tilde{\nu}$. We fix a decreasing factor $a \in (0,1)$ and we do as follows.

(a) Choose $f^0 \in L^2(\omega)^N$ and $\nu^0 \in \mathbb{R}^+$ and compute the solution (u^0, p^0) to

$$\begin{cases} -\nu^0 \Delta u^0 + \nabla p^0 = f^0 1_\omega, \quad x \in \Omega, \\ \nabla \cdot u^0 = 0, \quad x \in \Omega, \\ u^0 = 0, \quad x \in \partial\Omega, \end{cases} \tag{2.61}$$

and the solution (φ^0, q^0) to

$$\begin{cases} -\nu^0 \Delta \varphi^0 + \nabla q^0 = \alpha\Big((u^0 - u_{1d})1_{\mathcal{O}_1}\Big) + (1-\alpha)\Big((u^0 - u_{2d})1_{\mathcal{O}_2}\Big), \quad x \in \Omega, \\ \nabla \cdot \varphi^0 = 0, \quad x \in \Omega, \\ \varphi^0 = 0, \quad x \in \partial\Omega \end{cases} \tag{2.62}$$

and take

$$f^0 = -\frac{a}{\mu}\ \varphi^0\big|_\omega \quad \text{ and } \quad \nu^1 = \max\{\tilde{\nu}, \nu^* \nu^0\}.$$

(b) For given $n \geq 0$, ν^n and $f^n \in L^2(\omega)^N$, (u^n, p^n) and (φ^n, q^n), do the following:

(b.1) Take $f^{n,0} = -\dfrac{a}{\mu}\varphi^n\ |_\omega$, $u^{n,0} = u^n$, $\varphi^{n,0} = \varphi^n$ and $\nu^{n+1} = \max(a\nu^n, \tilde{\nu})$.

(b.2) Then, for given $k \geq 0$, $f^{n,k}, u^{n,k}, \varphi^{n,k}$, set

$$F^{n,k} := -\nu^{n+1}\Delta u^{n,k} + (u^{n,k} \cdot \nabla)u^{n,k} - f^{n,k} 1_\omega$$

and

$$G^{n,k} := -\nu^{n+1}\Delta\varphi^{n,k} - (u^{n,k}\cdot\nabla)\varphi^{n,k} + (\nabla u^{n,k})^t\varphi^{n,k} \\ - \alpha(u^{n,k}-u_{1d})1_{\mathcal{O}_1} - (1-\alpha)(u^{n,k}-u_{2d})1_{\mathcal{O}_2},$$

compute the solution $(v^k, h^k, \psi^k, \eta^k)$ to

$$\begin{cases} -\nu^{n+1}\Delta v^k + (u^{n,k}\cdot\nabla)v^k + (v^k\cdot\nabla)u^{n,k} + \nabla h^k = F^{n,k}, \quad x\in\Omega, \\ \nabla\cdot v^k = 0, \quad x\in\Omega, \\ v^k = 0, \quad x\in\partial\Omega, \\ -\nu^{n+1}\Delta\psi^k - (u^{n,k}\cdot\nabla)\psi^k - (v^k\cdot\nabla)\varphi^{n,k} \\ \qquad + (\nabla u^{n,k})^t\psi^k + (\nabla v^k)^t\varphi^{n,k} + \nabla\eta^k = G^{n,k}, \quad x\in\Omega, \\ \nabla\cdot\psi^k = 0, \quad x\in\Omega, \\ \psi^k = 0, \quad x\in\partial\Omega \end{cases} \tag{2.63}$$

and take:

$$u^{n,k+1} = u^{n,k} - v^k, \quad \varphi^{n,k+1} = \varphi^{n,k} - \psi^k. \tag{2.64}$$

Note that **ALG 7** and **ALG 10** are conceived to compute a solution to the optimality system (2.44) that, maybe, is not a minimizer of $J_{(\alpha)}$. Thus, they are expected to furnish numerical approximations of Pareto quasi-equilibria. From the viewpoint of the Calculus of Variations,**ALG 7** and **ALG 10** are related to the so called "indirect method". Contrarily, **ALG 8** and **ALG 9** intend to provide (numerical approximation of) a minimizing sequence of $J_{(\alpha)}$. Accordingly, they correspond to realizations of the "direct method" of Calculus of Variations.

Now, in order to illustrate the behavior of the previous algorithms, we discuss some numerical experiments. Specifically, we will try to compute a minimizer of the functional

$$J_{(\alpha)}(f,u) = \frac{a\alpha}{2}\int_{\mathcal{O}_1}|u-u_{1d}|^2 + \frac{a(1-\alpha)}{2}\int_{\mathcal{O}_2}|u-u_{2d}|^2 + \mu\int_\omega |f|^2, \tag{2.65}$$

where $\alpha = 0.5$, u_{1d}, u_{2d} are given functions and $a, \mu \in (0,2)$ are fixed parameters, with $\mu = 2-a$.

Our domain is composed by two rectangles $\mathcal{O}_1$ and $\mathcal{O}_2$ and we assume that the controls act on a narrow band ω. In order to solve numerically the systems (2.52), (2.53), (2.61), (2.62) and (2.63), we have to fix a mesh and a finite element method. We have used the mesh depicted in Fig. 1 and a mixed finite element formulation with continuous piecewise $\mathbb{P}_1$-bubble and $\mathbb{P}_1$ functions respectively for the velocity field and the pressure; for details, see [14, 17].

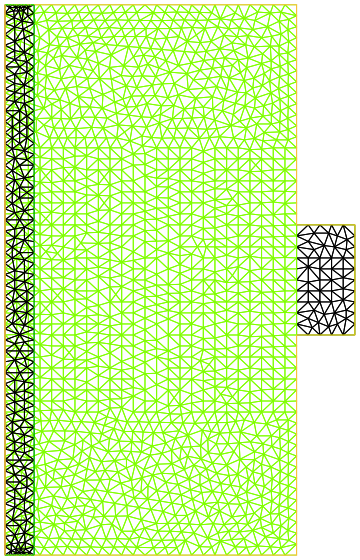


Figure 2.1: The domain and the "rough" mesh; Ω is composed of the band ω, the large rectangle $\mathcal{O}_1$ and the small rectangle $\mathcal{O}_2$. Number of nodes: 1519 . Number of triangles: 2876.

The data u_{id} are the following: $u_{1d} = \nabla \times \psi_{1d}$, where ψ_{1d} is the solution to the problem

$$\begin{cases} -\Delta\psi_{1d} = 1, & x \in \mathcal{O}_1, \\ \psi_{1d} = 0, & x \in \partial\mathcal{O}_1 \end{cases}$$

and $u_{2d} \equiv 0$. That means that the "desired" configuration corresponds to a uniformly rotating flow in $\mathcal{O}_1$ and a fluid at rest in $\mathcal{O}_2$ (see Fig. 2).

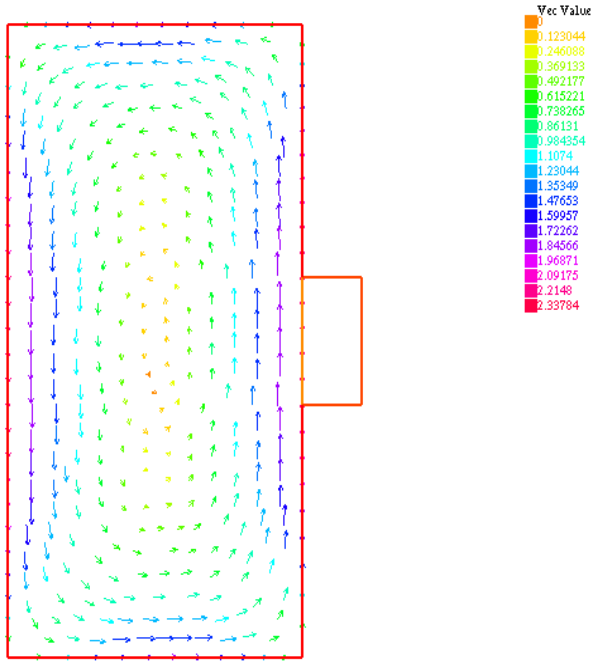


Figure 2.2: The function u_{1d}.

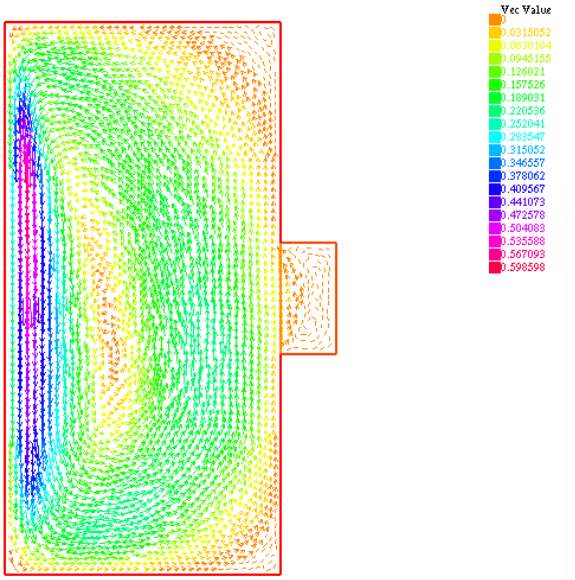
For **ALG 7**, **ALG 8** and **ALG 9**, the stopping test has been

$$\|u^{n+1} - u^n\|_{L^\infty} + \|p^{n+1} - p^n\|_{L^\infty} + \|q^{n+1} - q^n\|_{L^\infty} \leq \varepsilon,$$

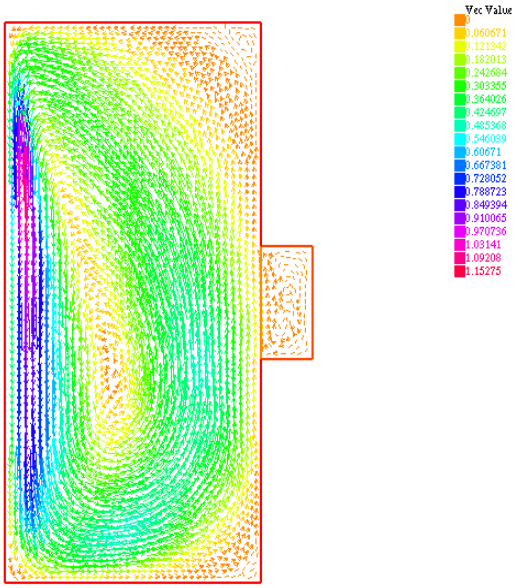
with $\varepsilon = 10^{-6}$. This has also been the stopping criterion for the external iterates in **ALG 10**. For the internal loops (indexed by k), the stopping test has been

$$\|u^{n,k+1} - u^{n,k}\|_{L^\infty} + \|\varphi^{n,k+1} - \varphi^{n,k}\|_{L^\infty} \leq \varepsilon.$$

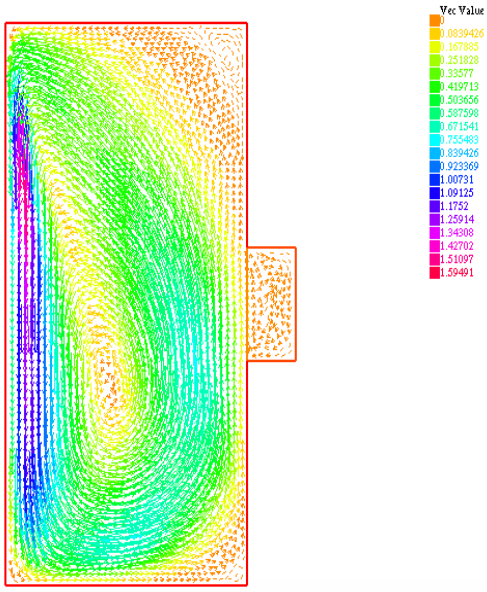
The computations have been performed with the FreeFem++ package (see [18]). We have used three different meshes: a "rough" mesh with 1519 nodes, a "reasonable " mesh with 3449 nodes and, also, a "fine" mesh with 6003 nodes. Some results are depicted in Fig. 3–5.



(a) $\alpha = 0.25$

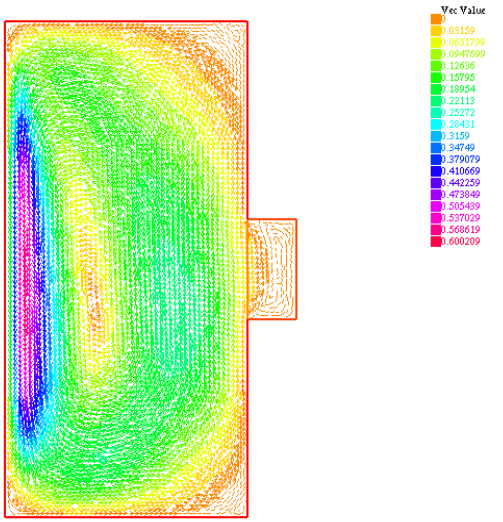


(b) $\alpha = 0.5$

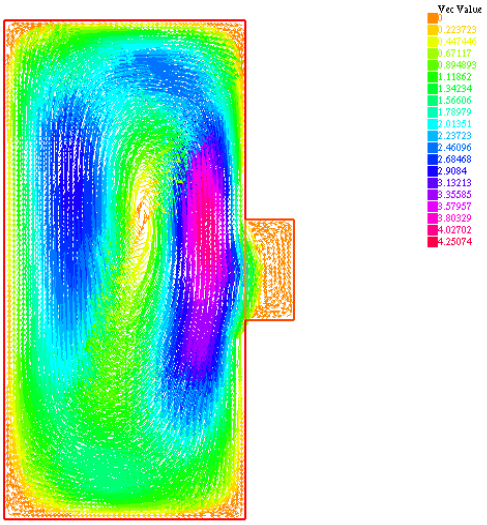


(c) $\alpha = 0.75$

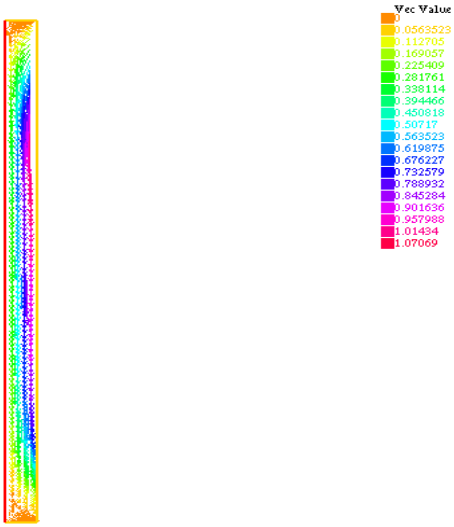
Figure 2.3: The final velocity fields computed with **ALG 8** for various α in the case $a = 1.5$ and $\nu = 0.06$. Number of nodes: 3449. Number of triangles: 6658.



(a) Velocity

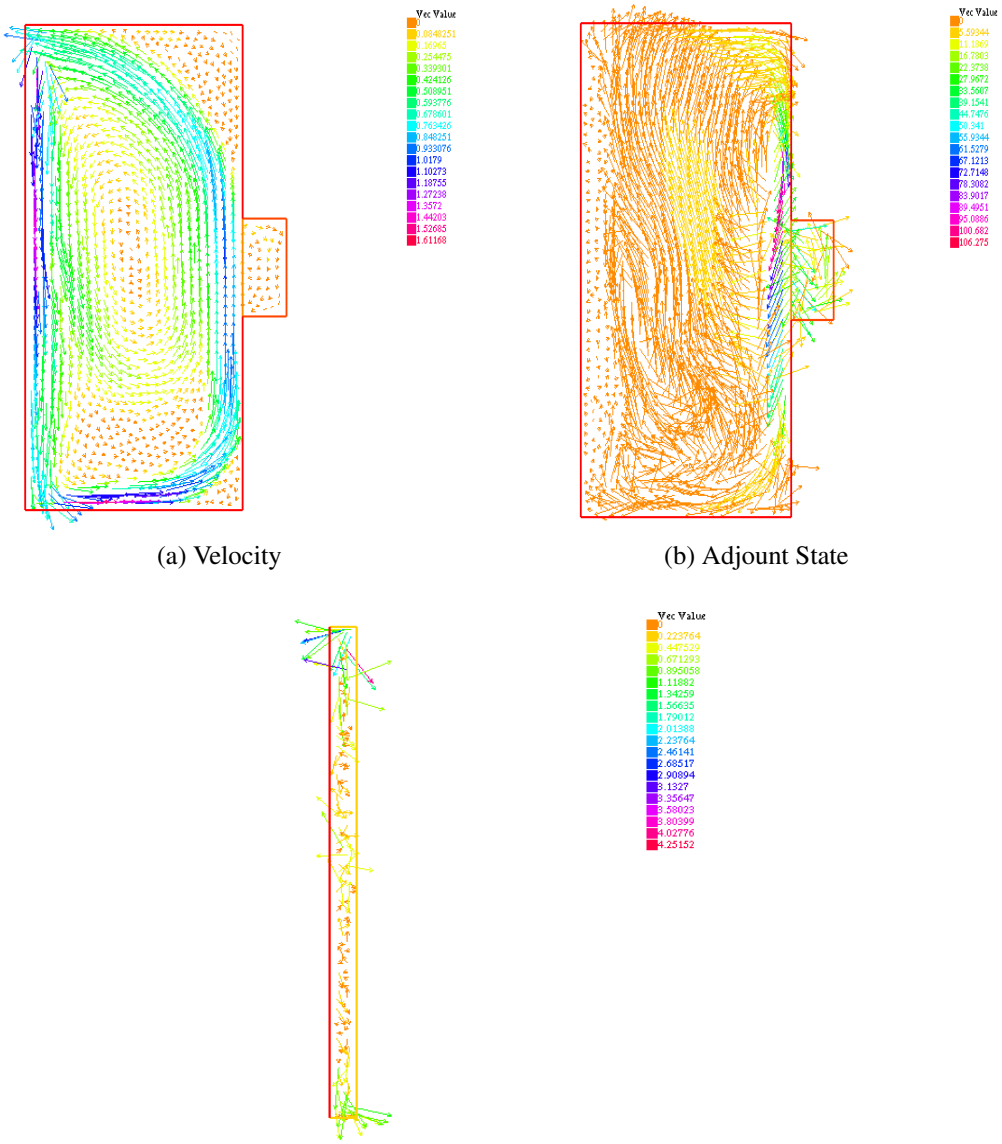


(b) Adjount State



(c) Control

Figure 2.4: The final velocity field, the adjoint and the control computed with **ALG 9** for $\alpha = 0.5$ in the case $a = 0.8$ and $\nu = 0.1$. Number of nodes: 6003. Number of triangles: 11684.



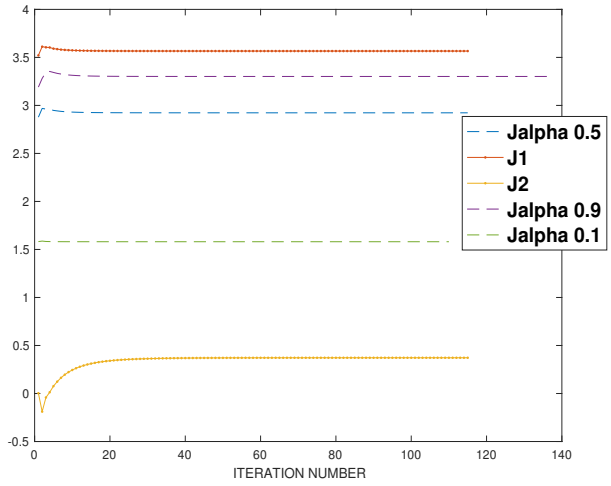
(a) Velocity

(b) Adjount State

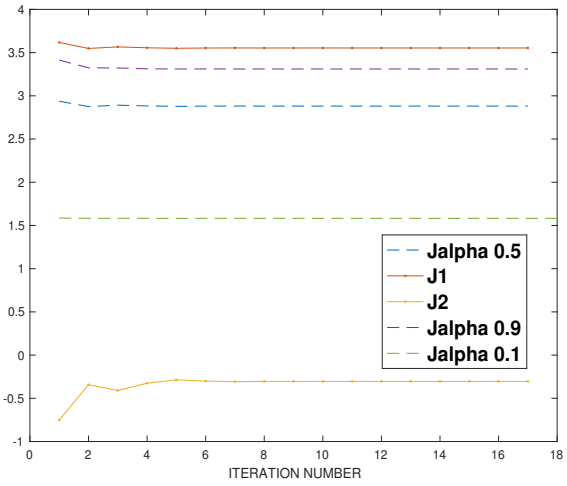
(c) Control

Figure 2.5: The final velocity field, the adjoint and the control computed with **ALG 10** for $\alpha = 0.5$ in the case $a = 1.8$ and $\nu = 0.00204$. Number of nodes: 1519. Number of triangles: 2876.

On the other hand, we have compared the values of the functionals J_1, J_2 and $J_{(\alpha)}$ for some values of α. See Fig. 2.6, where this is done for **ALG 8** and **ALG 9** (corresponding to "direct" methods).



(a) Gradient iteration



(b) Gradient Conjugate iteration

Figure 2.6: The logarithms of the functionals J_1, J_2 for $\alpha = 0.5$ and $J_{(\alpha)}$ for $\alpha = 0.1, 0.5$ and 0.9, with $a = 1.5$ and $\nu = 0.06$. Number of nodes: 1519. Number of triangles: 2876.

In order to compare the behavior of the gradient, the conjugate gradient and Newton methods, we present numerical values in the tables in Fig. 2.7 to 2.9. More precisely, we gather in Fig. 2.7 the number of iterates needed by each method to fulfill the stopping test and produce an approximation with error less or equal than 10^{-6}.

υ \ a	0.1	0.8	1.5	1.8
1	6	6	6	6
0.5	6	6	6	10
0.06	8	34	115	250
0.04	13	71	274	830

(a) Gradient (**ALG 8**)

υ \ a	0.1	0.8	1.5	1.8
1	2	3	3	4
0.5	2	3	4	6
0.06	5	12	17	28
0.04	8	13	24	39

(b) Conjugate Gradient (**ALG 9**)

υ \ a	0.1	0.8	1.5	1.8
1	2	3	3	3
0.5	2	3	3	3
0.06	8	9	14	16
0.04	9	10	16	19
0.00203	794	801	810	811

(c) Newton (**ALG 10**)

Figure 2.7: Number of iterates with $\alpha = 0.5, 1519$ nodes and $\varepsilon = 10^{-6}$ for various values of a and ν.

The tables in Fig. 2.7 illustrate the convergence properties of these algorithms. We see that the Newton method converges for small values of ν (the smallest value corresponds to a Reynolds number of 3500 approximately). However, remember that we cannot ensure in principle that the corresponding computed solution minimizes $J_{(\alpha)}$.

The tables in Fig. 2.8 and 2.9 show the errors corresponding to the 50^{th} iterate for each of the previous methods and (again) various parameters and data.

υ	NP		
	1519	3449	6003
1	6.64e-8	4.38e-8	5.65e-8
0.5	5.32e-7	5.77e-7	4.98e-7
0.06	6.37e-4	6.64e-4	8.96e-4
0.04	1.03e-2	1.06e-2	1.18e-2

(a) Gradient (**ALG 8**)

υ	NP		
	1519	3449	6003
1	4.28e-8	1.06e-8	7.80e-8
0.5	5.32e-8	1.04e-8	1.53e-8
0.06	7.85e-7	6.40e-7	3.46e-7
0.04	5.96e-7	2.38e-7	7.12e-7

(b) Conjugate Gradient (**ALG 9**)

υ	NP		
	1519	3449	6003
1	1.47e-10	3.36e-10	1.03e-10
0.5	5.81e-10	2.92e-10	1.11e-9
0.06	7.57e-10	1.49e-9	1.31e-9
0.04	2.35e-8	2.47e-8	2.23e-8

(c) Newton (**ALG 10**)

Figure 2.8: Precision in iteration 50 for $\alpha = 0.5$ and $a = 1.5$ (NP is the number of nodes).

a	NP		
	1519	3449	6003
0.1	1.52e-7	1.29e-7	2.68e-8
0.8	3.17e-7	2.84e-7	3.22e-7
1.5	6.37e-4	6.64e-4	8.96e-4
1.8	6.18e-3	6.19e-3	7.02e-3

(a) Gradient (**ALG 8**)

a	NP		
	1519	3449	6003
0.1	3.66e-7	6.9e-7	8.37e-7
0.8	1.36e-7	2.53e-5	3.22e-7
1.5	6.37e-7	6.64e-7	8.96e-7
1.8	6.18e-7	6.19e-7	7.02e-7

(b) Conjugate Gradient (**ALG 9**)

a	NP		
	1519	3449	6003
0.1	6.87e-9	2.97e-9	7.52e-9
0.8	6.82e-10	1.35e-9	6.72e-9
1.5	7.57e-10	1.49e-9	1.31e-9
1.8	2.02e-9	1.83e-9	4.14e-9

(c) Newton (**ALG 10**)

Figure 2.9: Precision in iteration 50 for $\alpha = 0.5$ and $\nu = 0.06$.

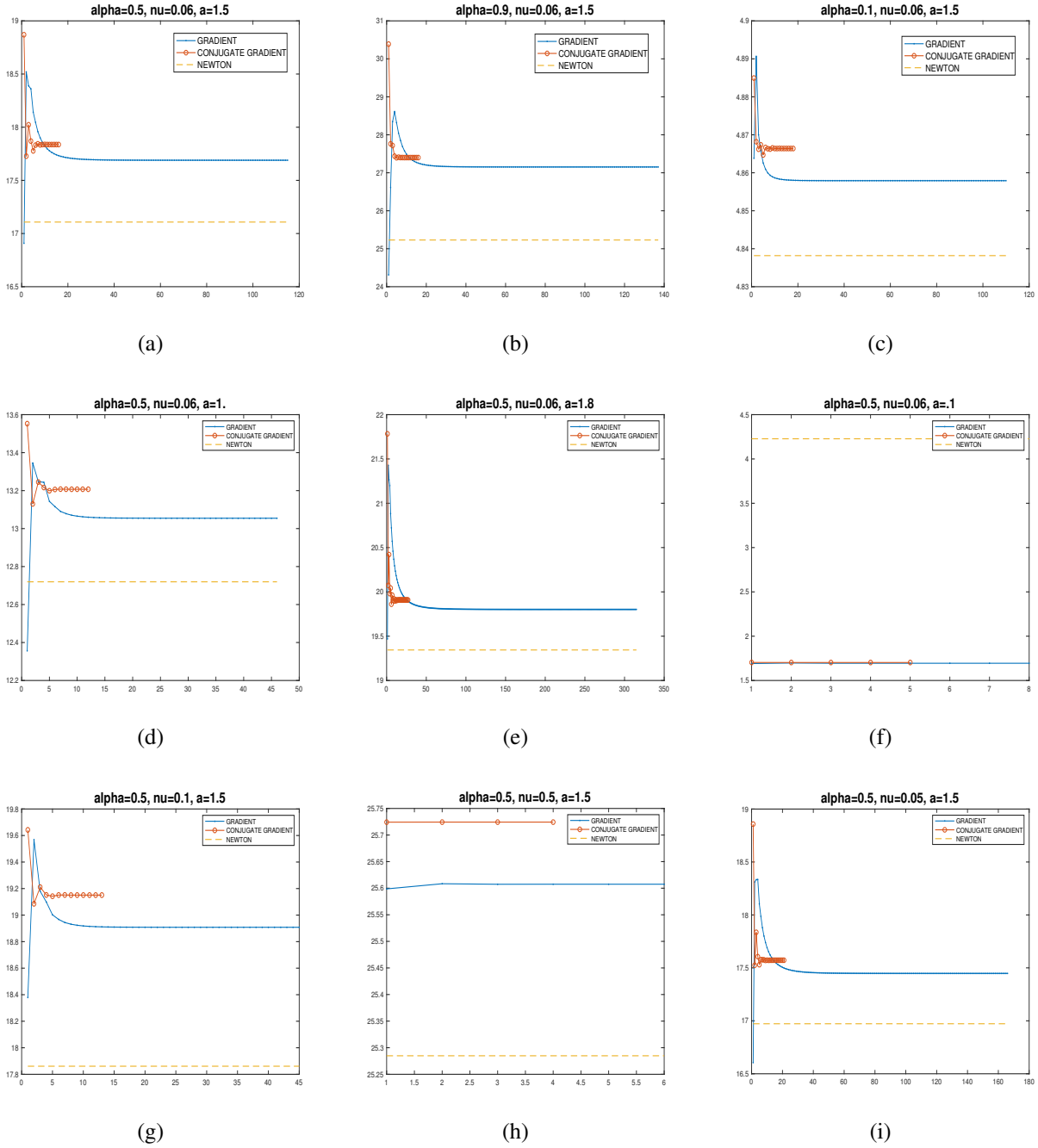
The results exhibited in these tables as we increase the number of nodes show that the behavior of **ALG 8**, **ALG 9** and **ALG 10** is consistent, in the sense that they remain approximately constant.

Also, we have included in Fig. 2.10 a table with a comparison of the computation times and required number of iterates of each method.

Finally, the functional $J_{(\alpha)}$ has been depicted in Fig. 2.11 for several values of the parameters.

NP	GRADIENT		CONJ.GRADIENT		NEWTON	
	CPU	#iter	CPU	#iter	CPU	#iter
1519	189.1	115	202.333	17	13.4	14
3449	470.7	117	617.855	20	36.2	15
6003	876.5	127	980.916	19	62.9	15

Figure 2.10: Computation times (in seconds) and numbers of iterates to reach an error less than $\varepsilon = 10^{-6}$, for $\alpha = 0.5$, $a = 1.5$ and $\nu = 0.06$.



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

Figure 2.11: The cost $J_{(\alpha)}$ for various methods and parameters.

Bibliography

[1] F. Abergel and R. Temam, *"On some control problems in fluid mechanics,"* Theoret. Comput. Fluid Dyn., **1** (1990), 303–325.

[2] G. Allaire and A. Craig, *"Numerical Analysis and Optimization. An Introduction to Mathematical Modelling and Numerical Simulation",* Oxford, London, 2007.

[3] L. J. Álvarez Vázquez, N. García Chan and A. Martínez, M. E. Vázquez Méndez, *"Multiobjective Pareto optimal control: an application to wastewater management",* Comput Optim Appl, Berlin, **46**, (2010), 135–157.

[4] J. L. Boldrini, B.M. Calsavara Caretta and E. Fernández-Cara, *"Some optimal control problems for a two-phase field model of solidification",* Rev Mat Complut, **23** (2010), 49–75.

[5] J.L. Boldrini, E. Fernández-Cara and M. A. Rojas-Medar, *"An Optimal Control Problem for a Generalized Boussinesq Model: The Time Dependent Case",* Rev. Mat. Complut., **20** (2007), 339–366.

[6] H. Brézis, *"Functional Analysis, Sobolev Sapces and Partial Differential Equations",* Springer, London, 2011.

[7] P.P. Carvalho and E. Fernández-Cara, *"On the Computation of Nash and Pareto Equilibria for some Bi-Objective Control Problems",* submited.

[8] E. Casas, *"The Navier-Stokes equations coupled with the heat equation: analysis and control",* Control Cybernet., **23** (1994), 605–620.

[9] E. Casas, J. P. Raymond and H. Zidani, *"Pontryagin principle for local solutions of control problems with mixed control-state constraints",* SIAM J. Control Optim., **39** (2000), 1182–1203.

[10] Ph.G. Ciarlet, *"Introduction to Numerical Linear Algebra and Optimisation",*, Cambridge University Press, Cambridge, 1989.

[11] C. Fabre, *"Uniqueness results for Stokes equations and their consequences in linear and nonlinear control problems",* ESAIM: Control, Optimisation and Calculus of Variations, **1** (1996), 267–302.

[12] E. Fernández-Cara, I. Marín-Gayte, *Theoretical and numerical results for some bi-objective optimal control problems*, Communications on Pure and Applied Analysis, 2020, 19, 4, 2101- 2126.

[13] T.M. Flett, *"Differential Analysis: Differentiation, Differential Equations and Differential Inequalities",* Cambridge University Press, Cambridge, 1980.

[14] A. V. Fursikov and O. Pironneau, *"Finite Element Methods for Navier-Stokes Equations",* Annual Review of Fluid Mechanics, **24** (1992), 167–204.

[15] A. V. Fursikov, *"Optimal Control of Distributed Systems: Theory and Applications",* American Mathematical Society, Boston, MA, 2000.

[16] I. V. Girsanov, "Lectures on mathematical theory of extremum problems", Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, **67** (1972).

[17] R. Glowinski, "Finite element methods for incompressible viscous flow", Handbook of Numerical Analysis, **9** (2003).

[18] F. Hecht, http://www.freefem.org

[19] J.-L. Lions, *"Contrôle de Pareto de systèmes distribués. Le cas d'évolution",* C.R. Acad. Sci. Paris, Série I, **302** (1986), 413–417.

[20] J.-L. Lions, *"Optimal control of systems governed by partial differential equations",* Springer-Verlag, New York, 1971.

[21] J. L. Lions, *"Some remarks on Stackelbergâs optimization",* Math. Models Methods Appl. Sci., **4** (1994), 477–487.

[22] J. F. Nash, *"Noncooperative games",* Ann. Math., **54** (1951), 286–295.

[23] V. Pareto, *"Cours d'économie politique",* Rouge, Laussane, Switzerland, 1896.

[24] E. Polak, *"Optimization. Algorithm and Consistent Approximation",* Springer-Verlag, New York, 1997.

[25] H. Von Stalckelberg, *"Marktform und gleichgewicht",* Springer, Berlin, Germany, 1934.

[26] R. Témam, *"Navier-Stokes Equations. Theory and Numerical Analysis",* Studies in Mathematics and Applications, North-Holland Publishing Co., Amsterdam, **2** (1977).

Capítulo 3

Theoretical and numerical bi-objective optimal control: Nash equilibria

This chapter deals with the solution of some multi-objective optimal control problems for several PDEs: linear and semilinear elliptic equations and stationary Navier-Stokes systems. More precisely, we look for Nash equilibria associated to standard cost functionals. We prove the existence of equilibria, we deduce appropiate optimality systems, we present some iterative algorithms and, in some cases, we establish convergence results. For the existence and characterization of Nash equilibria in the Navier-Stokes case, we use the formalism of Dubovitskii and Milyoutin. In this framework, we also present a finite element approximation of the bi-objective problem and we illustrate the techniques with several numerical experiments. It is based on [13].

3.1. Introduction

We consider bi-objective optimal control problems for various PDEs and systems. First, an introductory problem corresponding to a linear elliptic PDE is analyzed with detail. Then, we deal with a similar semilinear elliptic PDE. Finally, we deal with the stationary Navier-Stokes system, that is, the equations satisfied by the velocity field u and the pressure p of a viscous incompressible fluid.

Our aims are to prove existence, to characterize efficiently the equilibria and, also, to compute numerical solutions to these multi-objective control problems. They are very important from the mathematical viewpoint and appear frequently in the applications; for some previous works on the subject, see for instance [3].

In classical control theory, we usually find a state equation or system and one control with the mission of achieving a predetermined goal. Frequently (but not always), the goal is to minimize a cost functional within a prescribed family of admissible controls. A different and interesting situation arises when several (in general, conflictive or contradictory) objectives are considered. This may happen, for example, if the cost function is the sum of several terms and it is not clear that an average provides a reasonable criterion. Also, it can be expectable to have more than one control acting on the equation. In these cases, we are led to consider multi-objective control problems. In contrast with the mono-objective case, various strategies for the choice of good or the best controls

can appear, depending on the characteristics of the problem. Moreover, these strategies can be cooperative or noncooperative (depending on whether or not several controls mutually cooperate in order to achieve prescribed goals).

There exist several equilibrium concepts for multi-objective problems, with origin in game theory. Each of them determines a strategy. Thus, let us mention the non-cooperative optimization strategy proposed by Nash [24], the Pareto cooperative strategy [25] and the Stackelberg hierarchical-cooperative strategy [27]. In the context of the control of PDEs, a relevant question is whether one is able to steer the system to a desired state (exactly or approximately). Up to date, there have been some works on the subject like the seminal papers by Lions [21,23] and other more recent contributions, like [5,6,8].

In this paper, we will be concerned with Nash equilibria associated to standard cost functionals. To be more precise, let us give some details in the case of the stationary Navier-Stokes equations. Thus, let the fluid domain be a bounded open set $\Omega \subset \mathbb{R}^N$, with $N = 2$ or 3. Let us introduce four nonempty open subsets, $\mathcal{O}_1, \mathcal{O}_2, \omega_1$ and ω_2 and let us assume that a velocity field u_{id} defined on $\mathcal{O}_i$ is given for $i = 1, 2$.

In this context, we want to find a couple $(f_1, f_2) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$ (the control pair) with the following property: there exists an associated state (u, p), that is, a weak solution to the system

$$\begin{cases} -\nu\Delta u + (u \cdot \nabla)u + \nabla p = f_1 1_{\omega_1} + f_2 1_{\omega_2} \quad x \in \Omega, \\ \nabla \cdot u = 0 \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \end{cases} \tag{3.1}$$

such that (f_1, f_2, u) is a Nash equilibrium for the functionals

$$J_i(f_1, f_2, u) := \frac{a}{2} \int_{\mathcal{O}_i} |u - u_{id}|^2 + \frac{\mu}{2} \int_{\omega_i} |f_i|^2, \quad i = 1, 2, \tag{3.2}$$

where $a, \mu > 0$ (see Definition 3.4.1 in Section 4).

Our first main goal will be to find conditions under which at least one Nash equilibrium exists. The second one will be to characterize these equilibria in terms of first order optimality conditions, i.e. to deduce a system of PDEs that the optimal solution and some associated adjoint states must satisfy. The third one will be to indicate how Nash equilibria can be computed, to present some related algorithms and illustrate the results with numerical experiments.

The proof of the existence of Nash equilibria is (more or less) standard. It relies on suitable well known a priori estimates for the solutions to (3.1), that holds when a/μ us sufficiently small. In what concerns optimality conditions, the situation is in general more delicate. Indeed, due to the possible lack uniqueness, the techniques usually employed for distributed control problems (see for instance [1,9,10,22]) cannot be applied in this case.

In this work we will use an alternative technique that relies on the so called formalism of Dubovitskii and Milyoutin. This approach was introduced in the context of mathematical programming and has been succesfully applied to the solution of many optimal control problems for ODEs since the 70's. A good presentation of its applications to these areas can be found in Girsanov [17]; see also Flett [14]. Later, these techniques have been applied successfully to some distributed control problems; see [5,6,8].

Some basic ideas that can be used to explain the formalism are the following. At a local minimizer, the cone of descent directions associated to a cost functional must be disjoint of the intersection of the cones of feasible and tangent directions, respectively determined by the admissible control set and (3.1). Consequently, from Hahn-Banach Theorem and some additional arguments, it follows that there must exist elements in the associated dual cones, not all them zero, that add up to zero. This algebraic condition is just the Euler-Lagrange system of the extremal problem at hand. When it is possible to identify the previous primal and dual cones, this system provides the first order optimality conditions in a systematic way. In the case of a standard (mono-objective) optimal control problem, it also leads to the corresponding Pontryagin minimum (or maximum) principle. Thus, a major task in our problem is the identification of the cones mentioned above. Note that, in the particular case of (3.1)-(3.2), the main difficulties are related to the highly nonlinear behavior of (3.1) and the possible nonuniqueness of u.

The plan of this chapter is the following

In Section 2, we consider a relatively simple problem: a linear elliptic PDE, together with quadratic functionals. We prove the existence of Nash equilibria, we furnish an optimality system and we present some iterative algorithms for their computation.

In Section 3, a more complex problem is analyzed: a semilinear elliptic PDE together with functionals of the same kind. Here, in view of the nonlinearity, we must work with equilibria and quasi-equilibria (see below). Again, existence, characterization and computation-oriented results are established.

Finally, Section 4 deals with the stationary Navier-Stokes system.

New difficulties are found: nonlinearity, lack of uniqueness, lack of regularity of the functionals, etc. We also provide the existence and optimality of Nash equilibria and quasi-equilibria. Additionally, we present some iterative algorithms and the results of several numerical experiments.

3.2. Introductory problem: a linear elliptic PDE

In the sequel, we denote by $\|\cdot\|$ and $(\cdot\,,\cdot)$ the usual L^2 norm and scalar product, respectively. The symbol 1_D will be used to denote the characteristic function of the set D and C will stand for a generic positive constant. For simplicity, we will assume that only two controls act on the system and two functionals are minimized but very similar considerations hold for systems with a higher number of controls and functionals.

3.2.1. Definition of Nash equilibria

Let $\Omega \subset \mathbb{R}^N$ be a nonempty bounded connected open set with regular boundary $\partial\Omega$ and let us assume that ω_1 and ω_2 are nonempty disjoint open subsets of Ω.

We will consider the problems

$$\begin{cases} -\Delta u = f_1 1_{\omega_1} + f_2 1_{\omega_2}, & x \in \Omega, \\ u = 0, & x \in \partial\Omega, \end{cases} \tag{3.3}$$

where the $f_i \in L^2(\omega_i)$ are the controls and u is the state.

Let $\mathcal{O}_1$ an $\mathcal{O}_2$ be open sets, representing prescribed observation domains and let the J_i be given by

$$J_i(f_1, f_2) := \frac{a}{2}\int_{\mathcal{O}_i} |u - u_{id}|^2 + \frac{\mu}{2}\int_{\omega_i} |f_i|^2, \quad i = 1, 2, \tag{3.4}$$

where the $u_{id} \in L^2(\mathcal{O}_i)$ are given functions and a and μ are positive constants.

The first bi-objective control problem considered in this work is the following:

Find a *Nash equilibria* associated to (3.3) and (3.4), that is, a pair of controls $(\widehat{f}_1, \widehat{f}_2) \in L^2(\omega_1) \times L^2(\omega_2)$ such that

$$\begin{cases} J_1(\widehat{f}_1, \widehat{f}_2) \leq J_1(f_1, \widehat{f}_2) & \forall f_1 \in L^2(\omega_1), \\ J_2(\widehat{f}_1, \widehat{f}_2) \leq J_2(\widehat{f}_1, f_2) & \forall f_2 \in L^2(\omega_2). \end{cases} \tag{3.5}$$

In this case, since the control-to-state mapping is well-defined, linear and continuous and the cost functionals J_i are quadratic and strictly convex and $\mathcal{C}^1$, it is not difficult to prove that $(\widehat{f}_1, \widehat{f}_2)$ is a Nash equilibria if and only if

$$\frac{\partial J_1}{\partial f_1}(\widehat{f}_1, \widehat{f}_2) = 0, \quad \frac{\partial J_2}{\partial f_2}(\widehat{f}_1, \widehat{f}_2) = 0. \tag{3.6}$$

3.2.2. Existence and characterization of Nash equilibria

For future purposes, note that

$$\Big(\frac{\partial J_i}{\partial f_i}(f_1, f_2), g_i\Big) = \int_{\omega_i} (a\varphi_i + \mu f_i) g_i \quad \forall f_i, g_i \in L^2(\omega_i), \tag{3.7a}$$

where φ_i is the i-th adjoint state associated to (f_1, f_2), i.e. the solution to

$$\begin{cases} -\Delta\varphi_i = (u - u_{id})1_{\mathcal{O}_i}, \quad x \in \Omega, \\ \varphi_i = 0, \quad x \in \partial\Omega. \end{cases} \tag{3.7b}$$

Here, u is the state associated to (f_1, f_2).

We can now present the main result of this section. It deals with the existence and characterization of Nash equilibria.

Theorem 3.2.1. *Let us assume that $(\widehat{f}_1, \widehat{f}_2) \in L^2(\omega_1) \times L^2(\omega_2)$. Then*

1. *$(\widehat{f}_1, \widehat{f}_2)$ is a Nash equilibrium if only if there exist $\widehat{u}$, $\widehat{\varphi}_1$ and $\widehat{\varphi}_2$ such that*

$$\begin{cases} -\Delta\widehat{u} = \widehat{f}_1 1_{\omega_1} + \widehat{f}_2 1_{\omega_2}, \quad x \in \Omega, \\ \widehat{u} = 0, \quad x \in \partial\Omega, \\ -\Delta\widehat{\varphi}_i = (\widehat{u} - u_{id})1_{\mathcal{O}_i}, \quad x \in \Omega, \\ \widehat{\varphi}_i = 0, \quad x \in \partial\Omega, \\ \hat{f}_i = -\dfrac{a}{\mu}\,\widehat{\varphi}_i|_{\omega_i}, \quad i = 1, 2. \end{cases} \tag{3.8}$$

2. *There exists $\chi = \chi(\Omega, \mathcal{O}_1, \mathcal{O}_2, \omega_1, \omega_2)$ such that, if $a/\mu \leq \chi$, then* (3.8) *possesses exactly one solution. Consequently, if $a > 0$, $\mu > 0$ and a/μ is sufficiently small, there exists a unique Nash equilibrium for J_1 and J_2.*

3. *Let us assume that $\mathcal{O}_1 = \mathcal{O}_2$. Then, for any $a > 0$ and $\mu > 0$, there exists exactly one solution to* (3.8)*, that is, a Nash equilibrium associated to* (3.3) *and* (3.4)*.*

Proof:

1. Let us first assume that $(\widehat{f_1}, \widehat{f_2})$ is a Nash equilibrium. Then, (3.6) holds. In view of (3.7a)-(3.7b), one must have

$$\hat{f}_i = -\frac{a}{\mu}\, \widehat{\varphi_i}|_{\omega_i} \quad \text{for } i = 1, 2$$

(here $\widehat{\varphi_i}$ solves (3.7b) for $u = \widehat{u}$). Hence, (3.8) is satisfied.

Conversely, if $(\widehat{f_1}, \widehat{f_2}), \widehat{u}$ and the $\widehat{\varphi_i}$ satisfy (3.8), then we see from (3.7a) that $\dfrac{\partial J_i}{\partial f_i}(\widehat{f_1}, \widehat{f_2}) = 0$.

Therefore, (3.6) holds and $(\widehat{f_1}, \widehat{f_2})$ is a Nash equilibrium.

So, the problem is to find (f_1, f_2) such that:

$$\Big(Id + \frac{a}{\mu}\Lambda_0\Big)(f_1, f_2) = -\frac{a}{\mu}\Big(z_1, z_2\Big).$$

Note that Λ_0 is a lineal, continuous, positive and self-adjoint operator by we consider $\mathcal{O}_1 = \mathcal{O}_2 = \mathcal{O}$. So, we ensure that the functional

$$K(f_1, f_2) := \frac{1}{2}\Big((Id + \frac{a}{\mu}\Lambda_0)(f_1, f_2), (f_1, f_2)\Big) + \Big(\frac{a}{\mu}(z_1, z_2), (f_1, f_2)\Big)$$

has a unique minimum (f_1, f_2) which is a Nash equilibrium. □

Remark that in this case, the Nash equilibrium can be viewed as the minimal of the functional K. And its existence and unicity is independent on the size of a/μ.

In the case $\mathcal{O}_1 \neq \mathcal{O}_2$, we only can say that there exists at most $a_1/\mu_1, a_2/\mu_2, a_3/\mu_3, \ldots$ with $a_n/\mu_n \to 0$ such that for all $a/\mu \neq a_n/\mu_n$ there exists a unique Nash equilibrium.

3.2.3. Algorithms and convergence

We will recall in this section three standard algorithms that can be used for the computation of Nash equilibria.

ALG 1: Fixed-Point.

(a) Choose $f_i^0 \in L^2(\omega_i)$, $i = 1, 2$.

(b) Then, for given $n \geq 0$ and $f_i^n \in L^2(\omega_i)$, compute the solution u^n to

$$\begin{cases} -\Delta u^n = f_1^n 1_{\omega_1} + f_2^n 1_{\omega_2} & x \in \Omega, \\ u^n = 0, & x \in \partial\Omega \end{cases} \tag{3.9}$$

and the solutions φ_i^n to the systems

$$\begin{cases} -\Delta\varphi_i^n = (u^n - u_{id})1_{\mathcal{O}_i} & x \in \Omega, \\ \varphi_i^n = 0, & x \in \partial\Omega \end{cases} \tag{3.10}$$

and, finally, take

$$f_i^{n+1} = -\frac{a}{\mu}\,\varphi_i^n|_{\omega_i}\,, \quad i = 1, 2. \tag{3.11}$$

ALG 2: Optimal Step Gradient Method.

(a) Choose $f_i^0 \in L^2(\omega_i)$, $i = 1, 2$.

(b) Then, for given $n \geq 0$ and $f_i^n \in L^2(\omega_i)$, compute the solution u^n to (3.9) and the solution φ_i^n to (3.10) and take

$$f_i^{n+1} = f_i^n - \rho_i^n g_i^n, \quad i = 1, 2, \tag{3.12}$$

where

$$g_i^n = a\,\varphi_i^n|_{\omega_i} + \mu f_i^n \tag{3.13}$$

and

$$\rho_1^n = \arg\Big(\min_{\rho\geq 0} J_1(f_1^n - \rho g_1^n, f_2^n)\Big), \qquad \rho_2^n = \arg\Big(\min_{\rho\geq 0} J_2(f_1^n, f_2^n - \rho g_2^n)\Big). \tag{3.14}$$

ALG 3: Optimal Step Conjugate Gradient Method.

(a) Choose $f_i^0 \in L^2(\omega_i)$, $i = 1, 2$.

(b) For $n = 0$, perform one step of **ALG 2** and take $d_i^0 = g_i^0$, $i = 1, 2$.

(c) Then, for given $n \geq 1$ and $f_i^n \in L^2(\omega_i)$ compute the solution u^n to to (3.9) and the solution φ_i^n to (3.10) and take

$$f_i^{n+1} = f_i^n - \rho_i^n d_i^n, \tag{3.15}$$

where

$$\begin{cases} d_i^n = g_i^n + \gamma_i^n d_i^{n-1}, \quad \gamma_i^n = \dfrac{\|g_i^n\|^2_{L^2(\omega_i)}}{\|g_i^{n-1}\|^2_{L^2(\omega_i)}}, \\ g_i^n = a\,\varphi_i^n|_{\omega_i} + \mu f_i^n \end{cases} \tag{3.16}$$

and

$$\rho_1^n = \arg\Big(\min_{\rho\geq 0} J_1(f_1^n - \rho d_1^n, f_2^n)\Big), \qquad \rho_2^n = \arg\Big(\min_{\rho\geq 0} J_2(f_1^n, f_2^n - \rho d_2^n)\Big). \tag{3.17}$$

The following convergence results hold:

Theorem 3.2.2. *Let us suppose that $a/\mu < \chi$, where χ is the constant furnished by Theorem 3.2.1. Then, the controls (f_1^n, f_2^n) furnished by ALG 1 satisfy $(f_1^n, f_2^n) \to (\widehat{f}_1, \widehat{f}_2)$ as $n \to +\infty$, where $(\widehat{f}_1, \widehat{f}_2)$, is the unique Nash equilibrium associated to J_1 and J_2. Furthermore, the speed of convergence is at least linear.*

The proof is inmediate: it suffices to argue as in the proof of Theorem 3.2.1 and note that ALG 1 is the usual fixed-point iteration method for Λ.

Regard to the convergence of the algorithms ALG 2 and ALG 3, we cannot guarantee that it checks since we are not considering an algorithm of the usual gradient. We are limiting ourselves to going in the directions that mark the partial derivatives, not the gradient.

Now, let's consider, as before, that the two open sets $\mathcal{O}_1$ and $\mathcal{O}_2$ are the same $\mathcal{O}$. In this case, we can ensure the convergence of these algorithms since in these case, the Nash equilibrium corresponds to the minimum of the functional K.

So, we consider the following algorithms:

ALG 2': Optimal Step Gradient Method.

(a) Solve the systems:

$$\begin{cases} -\Delta Z_i = u_{id} 1_{\mathcal{O}} & x \in \Omega, \\ Z_i = 0, & x \in \partial\Omega \end{cases} \tag{3.18}$$

for $i = 1$ and 2.

(b) Choose $f_i^0 \in L^2(\omega_i)$, $i = 1, 2$.

(c) Then, for given $n \geq 0$ and $f_i^n \in L^2(\omega_i)$, compute the solution u^n to (3.9) and the solution ψ^n to

$$\begin{cases} -\Delta \psi^n = u^n 1_{\mathcal{O}} & x \in \Omega, \\ \psi^n = 0, & x \in \partial\Omega. \end{cases} \tag{3.19}$$

and take

$$f_i^{n+1} = f_i^n - \rho^n d_i^n, \quad i = 1, 2, \tag{3.20}$$

where

$$d_i^n = f_i^n + \frac{a}{\mu} \psi^n|_{\omega_i} + \frac{a}{\mu} Z_i \tag{3.21}$$

and

$$\rho^n = \arg \Big(\min_{\rho \geq 0} K(f_1^n - \rho d_1^n, f_2^n - \rho d_2^n) \Big), \tag{3.22}$$

ALG 3': Optimal Step Conjugate Gradient Method.

(a) Solve the systems:

$$\begin{cases} -\Delta Z_i = u_{id} 1_{\mathcal{O}} & x \in \Omega, \\ Z_i = 0, & x \in \partial\Omega \end{cases} \tag{3.23}$$

for $i = 1$ and 2.

(b) Choose $f_i^0 \in L^2(\omega_i)$, $i = 1, 2$.

(c) For $n = 0$, perform one step of **ALG 2'** and take d_i^0, $i = 1, 2$.

(c) Then, for given $n \geq 1$ and $f_i^n \in L^2(\omega_i)$ compute the solution u^n to to (3.9) and the solution ψ^n to (3.19) and take

$$f_i^{n+1} = f_i^n - \rho^n d_i^n, \tag{3.24}$$

where

$$\begin{cases} d_i^n = g_i^n + \gamma^n d_i^{n-1}, \quad \gamma^n = \dfrac{\|(g_1^n, g_2^n)\|^2_{L^2(\omega_1)\times L^2(\omega_2)}}{\|(g_1^{n-1}, g_2^{n-1})\|^2_{L^2(\omega_1)\times L^2(\omega_2)}}, \\ g_i^n = f_i^n + \dfrac{a}{\mu}\,\psi^n|_{\omega_i} + \dfrac{a}{\mu} Z_i \end{cases} \tag{3.25}$$

and

$$\rho^n = \arg\Big(\min_{\rho\geq 0} K(f_1^n - \rho d_1^n, f_2^n - \rho d_2^n)\Big). \tag{3.26}$$

Note that the operator K is symmetric and is a cuadratic operator. So, ρ is characterized for each n as the minimum of a quadratic function, i.e. ρ is the vertex.

Theorem 3.2.3. *The controls (f_1^n, f_2^n) furnished by ALG 2' and ALG 3' satisfy $(f_1^n, f_2^n) \to (\widehat{f}_1, \widehat{f}_2)$ as $n \to +\infty$, respectively, where $(\widehat{f}_1, \widehat{f}_2)$, is the unique Nash equilibrium associated to J_1 and J_2 (or K).*

The proof of this theorem is immediate by the properties of Λ_0 when $\mathcal{O}_1 = \mathcal{O}_2$.

Also, it is remarkable that if we have the same set $\mathcal{O}$, ALG 2 and ALG 2' and ALG3 and ALG 3 can be, respectively, equivalent. In fact, the gradient of K is μ divided by g_i^n of the ALG 2 and ALG 3. And if we take the same ρ_i^n and the same γ_i^n for each functional J_i as in ALG 2' and ALG 3' we can ensure the convergence of the ALG 2 and ALG 3.

3.3. The case of a semilinear elliptic PDE

3.3.1. Nash equilibria and quasi-equilibria

This section is devoted to introduce Nash optima in the semilinear case. Now, we must distinguish equilibria from quasi-equilibria and take into account the particularities of each of them.

Let us assume that

$$\begin{cases} \phi : \mathbb{R} \mapsto \mathbb{R} \text{ is } \mathcal{C}^1(\mathbb{R}), \\ 0 \leq \phi'(s) \leq C \quad \forall s \in \mathbb{R}. \end{cases} \tag{3.27}$$

The state equation is now the following:

$$\begin{cases} -\Delta u + \phi(u) = f_1 1_{\omega_1} + f_2 1_{\omega_2}, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega. \end{cases} \tag{3.28}$$

It is well known that, for each $(f_1, f_2) \in L^2(\omega_1) \times L^2(\omega_2)$, there exists exactly one solution u to (3.28). As in Section 2, we will consider the cost functionals J_i in (3.4), where the $u_{id} \in L^2(\mathcal{O}_i)$ and $a, \mu > 0$. Again, it will be said that $(\widehat{f}_1, \widehat{f}_2)$ is a Nash equilibrium for (3.28) and (3.4) if (3.6) is satisfied.

It is known that, under the previous assumptions on ϕ, the cost functionals J_i are $\mathcal{C}^1$ and satisfy

$$\Big(\frac{\partial J_i}{\partial f_i}(f_1,f_2),g_i\Big)=\int_{\omega_i}(a\varphi_i+\mu f_i)g_i\quad\forall f_i,g_i\in L^2(\omega_i),$$

where φ_i is the unique solution to

$$\begin{cases}-\Delta\varphi_i+\phi'(u)\varphi_i=(u-u_{id})1_{\mathcal{O}_i}\quad x\in\Omega,\\ \varphi_i=0,\quad x\in\partial\Omega\end{cases}\tag{3.29}$$

(that is, φ_i is the i-th adjoint state corresponding to f_1 and f_2) and u is the solution to (3.28); see for instance [16, 22].

Definition 3.3.1. *It will be said that $(\widehat{f}_1,\widehat{f}_2)$ is a **Nash quasi-equilibrium** for* (3.28) *and* (3.4) *if $(\widehat{f}_1,\widehat{f}_2)$ satisfies* (3.6)*, that is, $(\widehat{f}_1,\widehat{f}_2)$ solves, together with $\widehat{u}$ and the $\widehat{\varphi}_i$, the optimality system*

$$\begin{cases}-\Delta\widehat{u}+\phi(\widehat{u})=\widehat{f}_1 1_{\omega_1}+\widehat{f}_2 1_{\omega_2},\quad x\in\Omega,\\ \widehat{u}=0,\quad x\in\partial\Omega,\\ -\Delta\widehat{\varphi}_i+\phi'(\widehat{u})\widehat{\varphi}_i=(\widehat{u}-u_{id})1_{\mathcal{O}_i},\quad x\in\Omega,\\ \widehat{\varphi}_i=0,\quad x\in\partial\Omega,\\ \hat{f}_i=-\dfrac{a}{\mu}\,\widehat{\varphi}_i|_{\omega_i}.\end{cases}\tag{3.30}$$

Note that, if $(\widehat{f}_1,\widehat{f}_2)$ is a Nash equilibrium, then $(\widehat{f}_1,\widehat{f}_2)$ is also a Nash quasi-equilibrium. However, the converse is not necessarily true.

Theorem 3.3.2. *Let us assume that $N\le 8$ and, besides* (3.27)*, one has $\phi\in\mathcal{C}^2(\mathbb{R})$, with $|\phi'|+|\phi''|\le C$. There exists ε, only depending on Ω,u_{1d},u_{2d} and $\|\phi\|_{W^{2,\infty}}$, such that, if $a/\mu\le\varepsilon$, then the following assertions are equivalent:*

(a) $(\widehat{f}_1,\widehat{f}_2)$ is a Nash equilibrium for (3.28) *and* (3.4)*.*

(b) $(\widehat{f}_1,\widehat{f}_2)$ is a Nash quasi-equilibrium for (3.28) *and* (3.4)*.*

Proof:

We have to prove that, under the previous conditions, if the couple $(\widehat{f}_1,\widehat{f}_2)$ is a Nash quasi-equilibrium, then it satisfies (3.30). Thus, let us assume that (3.6) holds and let the functionals $\tilde{J}_i$ be given, with

$$\tilde{J}_1(f_1):=J_1(f_1,\widehat{f}_2)\ \forall f_1\in L^2(\omega_1)\quad\text{ and }\quad\tilde{J}_2(f_2):=J_2(\widehat{f}_1,f_2)\ \forall f_2\in L^2(\omega_2).$$

First, we must to prove that there exists ε_0 and C_0 positives constants such that if $a/\mu\le\varepsilon_0$, then any quasi-equilibrium $(\widehat{f}_1,\widehat{f}_2)$ satisfies that

$$\|\hat{f}_i\|_{L^2(\omega_i)}\le C_0(\Omega,u_{1d},u_{2d},\|\phi\|_{W^{2,\infty}}).$$

Effectively, it is true because if we take the equation for u and for φ_i and we multiply by u and φ_i, respectively, and we integrate we have that:

$$\|\nabla u\|^2 + \int_\Omega \phi(u)u + \frac{a}{\mu}\Big(\int_{\omega_1} u\varphi_1 + \int_{\omega_2} u\varphi_2\Big) = 0$$

and

$$\|\nabla \varphi_i\|^2 + \int_\Omega \phi'(u)|\varphi_i|^2 - \int_{\mathcal{O}_i} u\varphi_i = -\int_{\mathcal{O}_i} u_{id}\varphi_i.$$

Now, by using the properties of ϕ and Hölder and Young inequalities, we get

$$\|\nabla u\|^2 \le C\Big(\frac{a}{\mu}\Big)\Big(\|\nabla\varphi_1\|^2 + \|\nabla\varphi_2\|^2\Big)$$

and

$$\|\nabla \varphi_i\|^2 \le C + C\Big(\frac{a}{\mu}\Big)\Big(\|\nabla\varphi_1\|^2 + \|\nabla\varphi_2\|^2\Big).$$

So, if $a/\mu \le \varepsilon_0$ there exists a constant $C_0 = C_0(\Omega, u_{1d}, u_{2d}, \|_{W^{2,\infty}})$ such that

$$\|\hat{f}_i\|_{L^2(\omega_i)} \le C_0.$$

Now, note that, the $\tilde{J}_i$ are twice continuosly differentiable. Let us see that, if a/μ is sufficiently small, $\tilde{J}_1''(\widehat{f}_1; g_1, g_1) > 0$ and $\tilde{J}_2''(\widehat{f}_2; g_2, g_2) > 0$ for all nonzero $g_i \in L^2(\omega_i)$.

We know that

$$(\tilde{J}_i'(f_i), g_i)_{L^2(\omega_i)} = \int_{\omega_i} \Big(a\varphi_i + \mu f_i\Big) g_i\, dx \quad \forall f_i, g_i \in L^2(\omega_i), \tag{3.31}$$

where φ_i is, together with u, the solution to

$$\begin{cases} -\Delta u + \phi(u) = f_1 1_{\omega_1} + f_2 1_{\omega_2} \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \\ -\Delta\varphi_i + \phi'(u)\varphi_i = (u - u_{id})1_{\mathcal{O}_i} \quad x \in \Omega, \\ \varphi_i = 0, \quad x \in \partial\Omega. \end{cases} \tag{3.32a}$$

For any small $\varepsilon > 0$, let us introduce u^ε and φ_i^ε, with

$$\begin{cases} -\Delta u^\varepsilon + \phi(u^\varepsilon) = (\widehat{f}_1 + \varepsilon g_1)1_{\omega_1} + (\widehat{f}_2 + \varepsilon g_2)1_{\omega_2}, \quad x \in \Omega, \\ u^\varepsilon = 0, \quad x \in \partial\Omega, \\ -\Delta\varphi_i^\varepsilon + \phi'(u^\varepsilon)\varphi_i^\varepsilon = (u^\varepsilon - u_{id})1_{\mathcal{O}_i} \quad x \in \Omega, \\ \varphi_i^\varepsilon = 0, \quad x \in \partial\Omega, \end{cases} \tag{3.32b}$$

where $g_i \in L^2(\omega_i)^N$. Let us put

$$z := \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}(u^\varepsilon - \widehat{u}) \quad \text{and} \quad \psi_i := \lim_{\varepsilon\to 0} \frac{1}{\varepsilon}(\varphi_i^\varepsilon - \widehat{\varphi}_i).$$

Note that these limits exist in $H_0^1(\Omega)$. This is easy to see by substracting the system (3.32a) from (3.32b), dividing by ε and letting $\varepsilon \to 0$ (here, we must use that $\phi_i \in \mathcal{C}^2(\mathbb{R})$ and ϕ_i' and ϕ_i'' are uniformly bounded). Furthermore, one has

$$\begin{cases} -\Delta z + \phi'(\widehat{u})z = g_1 1_{\omega_1} + g_2 1_{\omega_2} \quad x \in \Omega, \\ z = 0, \quad x \in \partial\Omega, \\ -\Delta\psi_i + \phi'(\widehat{u})\psi_i + \phi''(\widehat{u})z\widehat{\varphi}_i = z1_{\mathcal{O}_i} \quad x \in \Omega, \\ \psi_i = 0, \quad x \in \partial\Omega. \end{cases} \tag{3.33}$$

Therefore,

$$\begin{cases} \tilde{J}_1''(\widehat{f}_1; g_1, g_1) = \lim\limits_{\varepsilon\to 0} \dfrac{1}{\varepsilon}\Big(\tilde{J}_1'(\widehat{f}_1 + \varepsilon g_1) - \tilde{J}_1'(\widehat{f}_1), g_1\Big) = \displaystyle\int_{\omega_1} (a\psi_1 + \mu g_1)g_1, \\ \tilde{J}_2''(\widehat{f}_2; g_2, g_2) = \lim\limits_{\varepsilon\to 0} \dfrac{1}{\varepsilon}\Big(\tilde{J}_2'(\widehat{f}_2 + \varepsilon g_2) - \tilde{J}_2'(\widehat{f}_2), g_2\Big) = \displaystyle\int_{\omega_2} (a\psi_2 + \mu g_2)g_2. \end{cases} \tag{3.34}$$

Observe that, from elliptic regularity, one has

$$\|\widehat{\varphi}_i\|_{H^2} \leq C_1(1 + \|\widehat{u}\|) \quad \text{and} \quad \|\nabla\widehat{u}\| \leq C_2(\|\widehat{f}_1\| + \|\widehat{f}_2\|), \tag{3.35}$$

for some constants $C_1 = C_1(\Omega, u_{1d}, u_{2d})$ and $C_2 = C_2(\Omega)$. On the other hand, from the PDEs satisfied by ψ_i and z, one has

$$\|\nabla\psi_i\| \leq C_3(1 + \|\widehat{\varphi}_i\|_{L^{N/2}})\|\nabla z\| \quad \text{and} \quad \|\nabla z\| \leq C_2(\|g_1\| + \|g_2\|), \tag{3.36}$$

where $C_3 = C_3(\Omega, \|\phi\|_{W^{2,\infty}})$. Consequently, taking into account that, for $N \leq 8$, $H^2(\Omega) \hookrightarrow L^{N/2}(\Omega)$ with continuous embedding, we see from (3.35) and (3.36) that

$$\tilde{J}_1''(\widehat{f}_1; g_1, g_1) \geq \mu\|g_1\|^2 - aC_4(1 + \|\widehat{f}_1\|)\|g_1\|^2$$
$$\tilde{J}_2''(\widehat{f}_2; g_2, g_2) \geq \mu\|g_2\|^2 - aC_4(1 + \|\widehat{f}_2\|)\|g_2\|^2$$

for some $C_4 = C_4(\Omega, u_{1d}, u_{2d}, \|\phi\|_{W^{2,\infty}})$.

Clearly, this proves that, if $N \leq 8$ and a/μ is sufficiently small, $\tilde{J}_i''(\hat{f}_i, g_i, g_i) > 0$ for all $g_i \neq 0$ and, consequently, $\tilde{J}_i$ possesses a unique global minimum at $\hat{f}_i$, for $i = 1, 2$.

In other words, $(\widehat{f}_1, \widehat{f}_2)$ is a Nash equilibrium for (3.28) and (3.4). □

Remark that also, we can prove this results by using the definition of the Nash equilibrium. If we consider the applications $\Psi_1 : f_2 \mapsto \widehat{f}_1(f_2) = \min J_1(\cdot\,, f_2)$ and $\Psi_2 : f_1 \mapsto \widehat{f}_2(f_1) = \min J_2(f_1\,, \cdot)$ then Nash equilibrium is the fixed point of $\Psi_2 \circ \Psi_1$ and it exists because $J_1(\cdot\,, f_2)$ and $J_2(\widehat{f}_1(f_2)\,, \cdot)$ are strictly convex.

3.3.2. Existence of Nash equilibria and quasi-equilibria

We can now prove the existence of Nash equilibria for (3.28) and (3.4).

Theorem 3.3.3. *Let the assumptions in Theorem 3.3.2 be satisfied.*

1- There exists $\chi_0 \leq \varepsilon$, only depending on u_{1d}, u_{2d} and $\|\phi\|_{W^{2,\infty}}$ such that, if $a/\mu \leq \chi_0$, then there exists at least one Nash equilibrium $(\widehat{f}_1, \widehat{f}_2)$ for (3.28) *and* (3.4)*.*

2- There exists $\chi_1 \leq \chi_0$ such that, whenever $a/\mu \leq \chi_1$, the Nash equilibrium is unique.

Proof:

To prove the existence of Nash equilibrium, we will check that, if a/μ is sufficiently small, the optimality system (3.30) possesses at least one solution. To this purpose, we will use *Schauder's Fixed-Point Theorem*.

The argument is similar to the proof of Theorem 3.3.2. Thus, let us consider the mapping $\Psi : L^2(\omega_1) \times L^2(\omega_2) \mapsto L^2(\omega_1) \times L^2(\omega_2)$ defined as follows: $(f_1, f_2) = \Psi(\tilde{f}_1, \tilde{f}_2)$ if and only if

$$f_i = -\frac{a}{\mu}\, \varphi_i|_{\omega_i}\,, \quad i = 1, 2,$$

where φ_i is the solution to (3.29) and u is the state associated to $\tilde{f}_1$ and $\tilde{f}_2$. Obviously, Ψ is well-defined, continuous and compact. Furthermore, if a/μ is small enough, Ψ maps the whole space $L^2(\omega_1) \times L^2(\omega_2)$ into a ball. This can be seen from the following estimates:

(a) First, from (3.28) and the properties satisfied by ϕ, one has

$$\|u\|^2_{H^1_0} + \int_\Omega \big(\phi(u) - \phi(0)\big)u = -\frac{a}{\mu}\int_\Omega \Big(\varphi_1 1_{\omega_1} + \varphi_2 1_{\omega_2}\Big)u - \phi(0)\int_\Omega u,$$

whence

$$\|u\|^2_{H^1_0} \leq C\Big(\Big(\frac{a}{\mu}\Big)^2\Big(\|\varphi_1\|^2_{H^1_0} + \|\varphi_2\|^2_{H^1_0}\Big) + 1\Big). \tag{3.37a}$$

(b) Then, taking into account (3.29), we deduce that

$$\|\varphi_i\|^2_{H^1_0} + \int_\Omega \phi'(u)|\varphi_i|^2 = \int_{\mathcal{O}_i} (u - u_{id})\varphi_i,$$

which yields

$$\|\varphi_i\|^2_{H^1_0} \leq C(\|u\|^2_{H^1_0} + 1). \tag{3.37b}$$

Form (3.37a) and (3.37b), our assertion follows.

Therefore, we can apply Schauder's Theorem to Ψ and the existence of a Nash equilibrium is ensured.

This ends the proof of existence.

Let us now see that, if a/μ is still smaller, the solution to (3.30) is unique.

In order to fix ideas, let us assume that $3 \leq N \leq 8$ (the case $N = 2$ is similar and even easier).

Let us assume that there exists two Nash equilibria (f_1^1, f_2^1) and (f_1^2, f_2^2) in $L^2(\omega_1) \times L^2(\omega_2)$. Then they solve the following systems for $j = 1$ and 2:

$$\begin{cases} -\Delta u^j + \phi(u^j) = f_1^j 1_{\omega_1} + f_2^j 1_{\omega_2} \quad x \in \Omega, \\ u^j = 0, \quad x \in \partial\Omega, \\ -\Delta \varphi_i^j + \phi'(u^j)\varphi_i^j = (u^j - u_{id}) 1_{\mathcal{O}_i} \quad x \in \Omega, \\ \varphi_i^j = 0, \quad x \in \partial\Omega. \\ f_i^j = -\dfrac{a}{\mu}\varphi_i^j 1_{\omega_i} \end{cases}$$

with $i = 1, 2$.

Let us set $f_i := f_i^1 - f_i^2$, $u := u^1 - u^2$ and $\varphi_i := \varphi_i^1 - \varphi_i^2$ for $i = 1, 2$. Then, we have

$$\begin{cases} -\Delta u + \phi'(\tilde{u})u = f_1 1_{\omega_1} + f_2 1_{\omega_2} \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \\ -\Delta \varphi_i + \phi'(u^1)\varphi_i + \phi''(\overline{u})\varphi_i^2 u = u 1_{\mathcal{O}_i} \quad x \in \Omega, \\ \varphi_i = 0, \quad x \in \partial\Omega. \\ f_i = -\dfrac{a}{\mu}\varphi_i 1_{\omega_i}, \end{cases} \tag{3.38}$$

where, for each x, one has $\tilde{u}(x) = \beta(x)u^1(x) + (1 - \beta(x))u^2(x)$ and $\overline{u}(x) = \lambda(x)u^1(x) + (1 - \lambda(x))u^2(x)$ for some $\beta(x), \lambda(x) \in (0, 1)$. We deduce that

$$\int_\Omega \Big(|\nabla u|^2 + \phi'(\tilde{u})|u|^2\Big) dx = (f_1 1_{\omega_1}, u) + (f_2 1_{\omega_2}, u)$$

and, therefore,

$$\|\nabla u\|^2 \leq C\frac{a^2}{\mu^2}(\|\varphi_1\|^2 + \|\varphi_2\|^2). \tag{3.39}$$

By a similar reason, denoting by 2^* the Sobolev embedding exponent of $H^1(\Omega)$ and $(2^*)'$ its conjugate, that is, $2^* = (2N)/(N-2)$ and $(2^*)' = 2N/(N+2)$, we also have:

$$\begin{aligned} \|\nabla\varphi_i\|^2 &\leq C\|\varphi_i^2 u\|_{L^{(2^*)'}}\|\varphi_i\|_{L^{2^*}} + C\|u\|\|\varphi_i\| \\ &\leq \frac{1}{2}\|\nabla\varphi_i\|^2 + C\|u\|^2 + C\|\varphi_i^2\|_{L^{N/2}}^2\|u\|_{L^{2^*}}^2 \\ &\leq \frac{1}{2}\|\nabla\varphi_i\|_{L^{N/2}}^2 + C\Big(1 + \|\varphi_i^2\|^2\Big)\|\nabla u\|^2, \end{aligned}$$

whence

$$\|\nabla u\|^2 \leq C\frac{a^2}{\mu^2}(1 + \|\varphi_1^2\|_{L^{N/2}}^2 + \|\varphi_2^2\|_{L^{N/2}}^2)\|\nabla u\|^2. \tag{3.40}$$

For $N \leq 8$, one has $N/2 \leq 2N/(N-4)$. Accordingly, $L^{N/2}(\Omega) \hookrightarrow H^2(\Omega)$ and, from the usual elliptic estimates, the following is found:

$$\|\varphi_i^2\|_{L^{N/2}}^2 \leq C\|(u^2 - u_{id})1_{\mathcal{O}_i}\|^2 \leq C(1 + \|u^2\|^2) \leq C\Big(1 + \frac{a^2}{\mu^2}(\|\varphi_1^2\|^2 + \|\varphi_2^2\|^2)\Big), \quad i = 1, 2. \tag{3.41}$$

This indicates that, if a/μ is sufficiently small, $\|\varphi_1^2\|^2_{L^{N/2}} + \|\varphi_2^2\|^2_{L^{N/2}} \leq C$ and, from (3.40), we necessarily have $u = 0$. Thus, in this case, we necessarily have $f_i = 0$ for $i = 1, 2$ and the proof is done. □

The following theorem show us that for any $a, \mu > 0$ there exists a Nash quasi-equilibrium.

Theorem 3.3.4. *If there exits $\delta > 0$ such that $\phi'(s) \geq \delta$ for any $s \in \mathbb{R}$ and $\phi(0) = 0$, then there exists a solution to the system* (3.30)*, i.e., there exists a Nash quasi-equilibrium.*

The proof of this theorem needs the following Lemma to the Brézis and Nirenberg's article,

Lemma 3.3.5. *Suppose that A is a closed linear operator with $N(A) = N(A^*)$ and A^{-1} is compact. Assume that B is a nonlinear and monotone demicontinuous operator satisfying*

$$(Bu - Bw, u) \geq \frac{1}{\gamma}|Bu|^2 - C(w), \quad \forall u, w$$

where $C(w)$ depends only on w and with $\gamma > 0$.

If $N(A) \subset R(B)$ then $A + B$ is onto.

This result can be viewed in [4]. Now, we pass to prove Theorem 3.3.4.

Proof:

The proof of this theorem also use the Lemma 3.3.5. Here we consider the operator A, i.e.,

$$A\begin{pmatrix} u \\ \varphi_1 \\ \varphi_2 \end{pmatrix} := \begin{pmatrix} -\Delta u + \dfrac{a}{\mu}\varphi_1 1_{\omega_1} + \dfrac{a}{\mu}\varphi_2 \\ -\Delta\varphi_1 - u1_{\mathcal{O}_1} \\ -\Delta\varphi_2 - u1_{\mathcal{O}_2} \end{pmatrix}$$

but the operator $B : L^2(\Omega)^3 \mapsto L^2(\Omega)^3$ is

$$B\begin{pmatrix} u \\ \varphi_1 \\ \varphi_2 \end{pmatrix} := \begin{pmatrix} \phi(u) \\ \phi'(u)\varphi_1 \\ \phi'(u)\varphi_2 \end{pmatrix}.$$

As similar way that the linear case, we can show that A satisfies the properties of the Lemma 3.3.5. So, only we must to prove that B is demicontinuous operator and that B satisfies

$$\left(B\begin{pmatrix} u \\ \varphi_1 \\ \varphi_2 \end{pmatrix} - B\begin{pmatrix} v \\ \psi_1 \\ \psi_2 \end{pmatrix}, \begin{pmatrix} u \\ \varphi_1 \\ \varphi_2 \end{pmatrix} \right) \geq \frac{1}{\gamma}\left\| B\begin{pmatrix} u \\ \varphi_1 \\ \varphi_2 \end{pmatrix} \right\|^2 - C\begin{pmatrix} v \\ \psi_1 \\ \psi_2 \end{pmatrix}$$

for all $(u, \varphi_1, \varphi_2)$, $(w, \psi_1, \psi_2) \in L^2(\Omega)^3$ and C depends only on w, ψ_1, ψ_2 and $\gamma > 0$.

It is easy to see that B is a continuous operator due to the properties of ϕ so we pass to see the proof of the other propertie.

Note that, for every $\gamma > 0$

$$\left(B\begin{pmatrix} u \\ \varphi_1 \\ \varphi_2 \end{pmatrix} - B\begin{pmatrix} v \\ \psi_1 \\ \psi_2 \end{pmatrix}, \begin{pmatrix} u \\ \varphi_1 \\ \varphi_2 \end{pmatrix} \right) = \frac{1}{\gamma}\Big(\int_\Omega |\phi(u)|^2 + \int_\Omega |\phi'(u)|^2|\varphi_1|^2 + + \int_\Omega |\phi'(u)|^2|\varphi_2|^2 \Big)$$

$$+ \int_\Omega \Big((\phi(u) - \phi(v))u - \frac{1}{\gamma}|\phi(u)|^2 \Big)$$

$$+ \int_\Omega \Big((\phi'(u)\varphi_1 - \phi'(v)\psi_1)\varphi_1 - \frac{1}{\gamma}|\phi'(u)|^2|\varphi_1|^2 \Big) + \int_\Omega \Big((\phi'(u)\varphi_2 - \phi'(v)\psi_2)\varphi_2 - \frac{1}{\gamma}|\phi'(u)|^2|\varphi_2|^2 \Big).$$

By Young's inequatlity we can write that

$$(\phi(u)-\phi(v))u - \frac{1}{\gamma}|\phi(u)|^2 \geq \phi'(\tilde{u})|u|^2 - \varepsilon|u|^2 - C_\varepsilon|\phi(v)|^2 - \frac{1}{\gamma}|\phi(u)|^2 \geq (\delta - \frac{C}{\gamma} - \varepsilon)|u|^2 - C_\varepsilon|\phi(v)|^2$$

with $\varepsilon > 0$, $C_\varepsilon = C(\varepsilon)$ the Young's constant and C and δ the constan of ϕ and $\tilde{u} = \lambda u$ with $\lambda \in (0, 1)$.

On the other hand, by Young inequality we have

$$(\phi'(u)\varphi_i - \phi'(v)\psi_i)\varphi_i - \frac{1}{\gamma}|\phi'(u)|^2|\varphi_i|^2 \geq \delta|\varphi_i|^2 - C\varphi_i\psi_i - \frac{C^2}{\gamma}|\varphi_i|^2 \geq \Big(\delta - \frac{C^2}{\gamma} - \varepsilon\Big)|\varphi_i|^2 - C_\varepsilon|\psi_i|^2$$

with $\varepsilon > 0$, $C_\varepsilon = C(\varepsilon)$ the Young's constant and C and δ the constan of ϕ for $i = 1, 2$.

So, if $\delta > \max\Big\{\frac{C}{\gamma} + \varepsilon, \frac{C^2}{\gamma} + \varepsilon\Big\}$ we can apply the Lemma 3.3.5, so $A + B$ is onto then there exists a Nash quasi-equilibrium. □

3.3.3. Algorithms and convergence

In this section, we present some iterative algorithms, similar to those considered in the linear case.

ALG 4: Fixed point.

(a) Choose $(f_1^0, f_2^0) \in L^2(\omega_1) \times L^2(\omega_2)$.

(b) Then, for given $n \geq 0$ and $(f_1^n, f_2^n) \in L^2(\omega_1) \times L^2(\omega_2)$, compute the solution u^n to

$$\begin{cases} -\Delta u^n + \phi(u^n) = f_1^n 1_{\omega_1} + f_2^n 1_{\omega_2} & x \in \Omega, \\ u^n = 0, & x \in \partial\Omega, \end{cases} \tag{3.42}$$

the solution φ_i^n to the system

$$\begin{cases} -\Delta\varphi_i^n + \phi'(u^n)\varphi_i^n = (u^n - u_{id})1_{\mathcal{O}_i} & x \in \Omega, \\ \varphi_i^n = 0, & x \in \partial\Omega \end{cases} \tag{3.43}$$

and, finally, take

$$f_i^{n+1} = -\frac{a}{\mu}\,\varphi_i^n|_{\omega_i} \tag{3.44}$$

for $i = 1, 2$.

ALG 5: Optimal Step Gradient Method.

(a) Choose $(f_1^0, f_2^0) \in L^2(\omega_1) \times L^2(\omega_2)$.

(b) Then, for given $n \geq 0$ and $(f_1^n, f_2^n) \in L^2(\omega_1) \times L^2(\omega_2)$, compute the solution u^n to (3.42) and the solution φ_i^n to (3.43) and take

$$f_i^{n+1} = f_i^n - \rho_i^n g_i^n, \tag{3.45}$$

where

$$g_i^n = a\ \varphi_i^n|_{\omega_i} + \mu f_i^n \tag{3.46}$$

and

$$\rho_1^n = \arg\Big(\min_{\rho\geq 0} J_1(f_1^n - \rho g_1^n, f_2^n)\Big), \qquad \rho_2^n = \arg\Big(\min_{\rho\geq 0} J_2(f_1^n, f_2^n - \rho g_2^n)\Big). \tag{3.47}$$

ALG 6: Optimal Step Conjugate Gradient Method.

(a) Choose $(f_1^0, f_2^0) \in L^2(\omega_1) \times L^2(\omega_2)$.

(b) For $n = 0$, perform one step of ALG 5 and take $d^0 = g^0$.

(c) Then, for given $n \geq 1$, $(f_1^n, f_2^n) \in L^2(\omega_1) \times L^2(\omega_2)$, $g_i^{n-1} \in L^2(\omega_i)$ and $d_i^{n-1} \in L^2(\omega_i)$, compute the solution u^n to (3.42) and the solution φ_i^n to (3.43) and take

$$f_i^{n+1} = f_i^n - \rho_i^n d_i^n, \tag{3.48}$$

where

$$\begin{cases} d_i^n = g_i^n + \gamma_i^n d_i^{n-1}, \quad \gamma_i^n = \dfrac{(g_i^n - g_i^{n-1}, g_i^n)_{L^2(\omega_i)\times L^2(\omega_i)}}{\|g_i^{n-1}\|^2_{L^2(\omega_i)}}, \\ g_i^n = a\ \varphi_i^n|_{\omega_i} + \mu f_i^n \end{cases} \tag{3.49}$$

and

$$\rho_1^n = \arg\Big(\min_{\rho\geq 0} J_1(f_1^n - \rho g_1^n, f_2^n)\Big), \qquad \rho_2^n = \arg\Big(\min_{\rho\geq 0} J_2(f_1^n, f_2^n - \rho g_2^n)\Big). \tag{3.50}$$

Note that in **ALG 3** and **ALG 6**, the coefficient γ_i^n is given by different expressions. The reason is that, now, the system is nonlinear and we must impose Polak condition to ensure convergence.

Theorem 3.3.6. *Let the assumptions in Theorem 3.3.3 be satisfied and let us assume that $a/\mu \leq \chi_0$. Then, the couples (f_1^n, f_2^n) furnished by **ALG 4** satisfy $(f_1^n, f_2^n) \to (\widehat{f}_1, \widehat{f}_2)$ as $n \to +\infty$ where $(\widehat{f}_1, \widehat{f}_2)$ is the unique Nash equilibrium.*

This proof is easy. Indeed, as already shown, we can write the iterates in the form $(f_1^{n+1}, f_2^{n+1}) = \Psi(f_1^n, f_2^n)$ for all $n \geq 0$. If a/μ is sufficiently small, arguing as in the proof of Theorem 3.3.3, it is easy to prove that Ψ is a contraction.

Theorem 3.3.7. *Let the assumptions in Theorem 3.3.3 be satisfied, let a/μ small enough and let the pair of controls (f_1^n, f_2^n) furnished by ALG 5. Then, $(f_1^n, f_2^n) \to (\widehat{f}_1, \widehat{f}_2)$ as $n \to +\infty$ with $(\widehat{f}_1, \widehat{f}_2)$ a Nash equilibrium.*

The proof can be obtained arguing as in the proof of Theorem 8.4-3 in [11]. Indeed, from the proof of Theorem 3.3.3, we deduce that, if a/μ is sufficiently small, then $\tilde{J}_i$ is elliptic, that is,

$$\left(\tilde{J}_i'(f_i) - \tilde{J}_i'(f_i'), f_i - f_i'\right) \geq c\|f_i - f_i'\|^2_{L^2(\omega_i)} \quad \forall f_i, f_i' \in L^2(\omega_i)^N$$

for some $c > 0$. Consequently,

$$\tilde{J}_i(f_i^n) - J_\alpha(f_i^{n+1}) \geq \frac{c}{2}\|f_i^n - f_i^{n+1}\|^2_{L^2(\omega_i)} \quad \text{and} \quad \|\tilde{J}_i'(f_i^n)\|_{L^2(\omega_i)} \leq \|\tilde{J}_i'(f_i^n) - \tilde{J}_i'(f_i^{n+1})\|_{L^2(\omega_i)}$$

for all $n \geq 1$ and $i = 1, 2$, whence in particular we have that $\|f_i^n - f_i^{n+1}\|_{L^2(\omega_i)} \to 0$ as $n \to +\infty$. Taking into account the expression of $\tilde{J}_i'$, we also have that $\|\tilde{J}_i'(f_i^n) - \tilde{J}_i'(f_i^{n+1})\|_{L^2(\omega_i)} \to 0$, whence $\tilde{J}_i'(f^n) \to 0$ and at least a subsequence of $\{f_i^n\}$ converges weakly towards the unique minimizer $\hat{f}_i$ of $\tilde{J}_i$. Since

$$\|f_i^n - \hat{f}_i\|_{L^2(\omega_i)} \leq \frac{1}{c}\|\tilde{J}_i'(f_i^n)\|_{L^2(\omega_i)} \quad \forall n \geq 1,$$

we see that, in fact, the whole sequence converges strongly to $\hat{f}_i$.

Theorem 3.3.8. *The assertion in Theorem 3.3.7 also holds for the pair of controls (f_1^n, f_2^n) furnished by ALG 6.*

Note that here we must be impose similiar condition to prove the convergence results for the three algorithms. It is due to the problem is not linear so to ensure the unicity we need more condition, not only that a/μ is sufficiently small.

For the proof, we can use the arguments in 96-98 in [26] (Theorems 1.5.8 and 1.5.9). More precisely, note first that there exists $\beta > 0$ such that

$$\left(\tilde{J}_i'(f_i^n), d_i^n\right) \geq \beta\|\tilde{J}_i'(f_i^n)\|_{L^2(\omega_i)}\|d_i^n\|_{L^2(\omega_i)} \quad \forall n \geq 1.$$

Therefore,

$$\tilde{J}_i(f_i^{n+1}) - \tilde{J}_i(f_i^n) \leq -C\|\tilde{J}_i'(f_i^n)\|_{L^2(\omega_i)}$$

and $\tilde{J}_i'(f_i^n) \to 0$ as $n \to +\infty$. Thus, we again have that subsequence of f_i^n converges weakly to $\hat{f}_i$ and arguing as before, we are led to the strong convergence of the whole sequence.

3.4. The stationary Navier-Stokes system

This section is devoted to the existence and characterization of Nash equilibria and quasi-equilibria for the stationary Navier-Stokes equations. In view of the properties of the state system and, in particular, the possible lack of uniqueness, this will be more complicated than in Section 2 and 3.

The stationary Navier-Stokes equations are the following:

$$\begin{cases} -\nu\Delta u + (u\cdot\nabla)u + \nabla p = f_1 1_{\omega_1} + f_2 1_{\omega_2}, \quad x\in\Omega, \\ \nabla\cdot u = 0, \quad x\in\Omega, \\ u = 0, \quad x\in\partial\Omega. \end{cases} \tag{3.51}$$

The state variables are $u = (u_1, \dots, u_N)$ and p. They can be interpreted as the velocity field and the pressure of a steady viscous Newtonian fluid. The controls are $f_1 1_{\omega_1}$ and $f_2 1_{\omega_2}$ and can viewed as external fields of forces respectively applied at the points $x \in \omega_1$ and $x \in \omega_2$.

3.4.1. Nash equilibria and quasi-equilibria

The positive constant ν is the kinematic viscosity of the fluid. It must be regarded as a measure of "thickness", i.e. tendency to favor friction.

As in the previous sections, let us introduce the functionals J_i with

$$J_i(f_1, f_2, u) := \frac{a}{2}\int_{\mathcal{O}_i} |u - u_{id}|^2 + \frac{\mu}{2}\int_{\omega_i} |f_i|^2, \quad i = 1, 2, \tag{3.52}$$

where $u_{id} \in L^2(\mathcal{O}_i)^N$ and $a, \mu > 0$.

Note that, here, we assume that the J_i depend not only on the controls f_i but also on the associated state$u$. This is due to the possible non-uniqueness of solution to (3.51), that can take place when ν is not sufficiently large.

Definition 3.4.1. *It will be said that $(\widehat{f_1}, \widehat{f_2}) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$ is a **Nash equilibrium** for* (3.51) *and* (3.52) *if there exists an associated state $(\widehat{u}, \widehat{p})$ satisfying*

$$\begin{cases} J_1(\widehat{f_1}, \widehat{f_2}, \widehat{u}) \le J_1(f_1, \widehat{f_2}, u) \quad \forall f_1 \in L^2(\omega_1)^N \text{ and any associated state } u \text{ to } (f_1, \widehat{f_2}), \\ J_2(\widehat{f_1}, \widehat{f_2}, \widehat{u}) \le J_2(\widehat{f_1}, f_2, u) \quad \forall f_2 \in L^2(\omega_2)^N \text{ and any associated state } u \text{ to } (\widehat{f_1}, f_2). \end{cases} \tag{3.53}$$

Definition 3.4.2. *It will be said that $(\widehat{f_1}, \widehat{f_2}) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$ is a **Nash quasi-equilibrium** for* (3.51) *and* (3.52) *if there exists a solution $(\widehat{u}, \widehat{p}, \widehat{\varphi}_1, \widehat{q}_1, \widehat{\varphi}_2, \widehat{q}_2)$ to the following coupled system for $i = 1, 2$*

$$\begin{cases} -\nu\Delta\widehat{u} + (\widehat{u}\cdot\nabla)\widehat{u} + \nabla\widehat{p} = \widehat{f_1} 1_{\omega_1} + \widehat{f_2} 1_{\omega_2} \quad x\in\Omega, \\ \nabla\cdot\widehat{u} = 0, \quad x\in\Omega, \\ \widehat{u} = 0, \quad x\in\partial\Omega, \\ -\nu\Delta\widehat{\varphi}_i + (\widehat{u}\cdot\nabla)\widehat{\varphi}_i + (\nabla\widehat{u})^t\widehat{\varphi}_i + \nabla\widehat{q}_i = (\widehat{u} - u_{id}) 1_{\mathcal{O}_i} \quad x\in\Omega, \\ \nabla\cdot\widehat{\varphi}_i = 0, \quad x\in\Omega, \\ \widehat{\varphi}_i = 0, \quad x\in\partial\Omega, \\ \hat{f}_i = -\dfrac{a}{\mu}\widehat{\varphi}_i 1_{\omega_i}. \end{cases} \tag{3.54}$$

3.4.2. Existence of Nash quasi-equilibria

Let us recall some classical spaces, usual for the analysis of the Navier Stokes equations:

$$H := \{v \in L^2(\Omega)^N : \nabla \cdot v = 0 \text{ in } \Omega,\ v \cdot n = 0 \text{ on } \partial\Omega\},$$

$$V := \{v \in H_0^1(\Omega)^N : \nabla \cdot v = 0 \text{ in } \Omega\}.$$

They are closed subspaces of $L^2(\Omega)^N$ and $H_0^1(\Omega)^N$, respectively; accordingly, they are Hilbert spaces for $(\cdot\,,\cdot)$ and $(\cdot\,,\cdot)_{H_0^1}$. Also, we have the compact embeddings $V \hookrightarrow H \equiv H' \hookrightarrow V'$ where X' denotes the dual space of X.

In the sequel we will need the Stokes operatot $A : D(A) \subset H \mapsto H$, where $D(A) = H^2(\Omega)^N \cap V$ and

$$Av = P(-\Delta v) \quad v \in V,$$

where $P : L^2(\Omega)^N \mapsto H$ is the usual orthogonal projector. It is known that A can be uniquely extended to a bounded linear operator in $\mathcal{L}(V; V')$, again denoted by A.

Then, A is self-adjoint and one has

$$\langle Av, w\rangle = (v, w)_{H_0^1} \quad \forall v, w \in V.$$

Let us consider the trilinear continuous forms $b(\cdot\,,\cdot\,,\cdot)$ and $\widehat{b}(\cdot\,,\cdot\,,\cdot)$, with

$$b(u, v, w) := \sum_{i,j=1}^{N} \int_\Omega u_i \partial w_j \, dx \quad \forall u, v, w \in V$$

and

$$\widehat{b}(u, v, w) := b(v, u, w) = \sum_{i,j=1}^{N} \int_\Omega \partial v_i w_j \, dx \quad \forall u, v, w \in V.$$

Note that there exist bilinear continuous mappings B and $\widehat{B}$, such that

$$\langle B(u, v), w\rangle = b(u, v, w) \quad \text{and} \quad \langle \widehat{B}(u, v), w\rangle = \widehat{b}(u, v, w) \quad \forall u, v, w \in V.$$

Theorem 3.4.3. *Let the optimallity sysmtem associated to a Nash equilibrium to* (3.51)*:*

$$\begin{cases} -\nu\Delta u + (u \cdot \nabla)u + \nabla p = f_1 1_{\omega_1} + f_2 1_{\omega_2} \quad x \in \Omega, \\ \nabla \cdot u = 0, \quad x \in \Omega, \\ u = 0, \quad x \in \partial\Omega, \\ -\nu\Delta\varphi_i + (u \cdot \nabla)\varphi_i + (\nabla u)^t \varphi_i + \nabla q_i = (u - u_{id}) 1_{\mathcal{O}_i} \quad x \in \Omega, \\ \nabla \cdot \varphi_i = 0, \quad x \in \Omega, \\ \varphi_i = 0, \quad x \in \partial\Omega, \\ f_i = -\dfrac{a}{\mu}\varphi_i 1_{\omega_i}. \end{cases} \tag{3.55}$$

Then, there exists $\varepsilon = \varepsilon(\Omega, \omega_1, \omega_2, u_{1d}, u_{2d}, \mathcal{O}_1, \mathcal{O}_2) > 0$ such that if $a/\mu \leq \varepsilon$ then there exists $(\widehat{f_1}, \widehat{f_2})$ a Nash quasi-equilibrium. In addition, there exists $\varepsilon_0 > 0$ with $\varepsilon_0 \leq \varepsilon$ such that if $a/\mu \leq \varepsilon_0$ then there exists a unique Nash quasi-equilibrium.

Proof:
To this proof we rewrite the system (3.55) as the following system

$$\begin{pmatrix} u \\ \varphi_1 \\ \varphi_2 \end{pmatrix} = \Lambda \begin{pmatrix} u \\ \varphi_1 \\ \varphi_2 \end{pmatrix}$$

with $\Lambda : V \times V \times V \mapsto V \times V \times V$ continuous and compact application defined by

$$\Lambda \begin{pmatrix} u \\ \varphi_1 \\ \varphi_2 \end{pmatrix} := \begin{pmatrix} \frac{1}{\nu} A^{-1}\big(-(u\cdot\nabla)u - \frac{a}{\mu}(\varphi_1 1_{\omega_1} + \varphi_2 1_{\omega_2}\big) \\ \frac{1}{\nu} A^{-1}\big(-(u\cdot\nabla)\varphi_1 - (\nabla u)^t\varphi_1 + (u - u_{1d})1_{\mathcal{O}_1}\big) \\ \frac{1}{\nu} A^{-1}\big(-(u\cdot\nabla)\varphi_2 - (\nabla u)^t\varphi_2 + (u - u_{2d})1_{\mathcal{O}_2}\big) \end{pmatrix}.$$

Now, we want to find a fixed point of Λ. It is easy to see that if a/μ is less than ε by Leray-Schauder Theorem's we can ensure that there exists $(u, \varphi_1, \varphi_2)$ fixed point of Λ. In addition, if $a/\mu \leq \varepsilon_0$ we can ensure that the fixed point is unique. □

Note that, the proof of the existence of Nash equilibrium, in this case, it is not easy and we don't know how proof that because we don't have convexity of the functionals restricted to a one variable.

Theorem 3.4.4. *Let $(\widehat{f_1}, \widehat{f_2})$ be a Nash equilibrium for* (3.51) *and* (3.52)*. Then, $(\widehat{f_1}, \widehat{f_2})$ is a Nash quasi-equilibrium.*

We will give a proof of this result that relies on the Dubovitsky-Milyoutin formalism (see [17]). To this purpose, we have to recall some technical results.

Lemma 3.4.5. *Let $K_1, \ldots, K_n$ be convex cones in a Banach space X with apex at 0. For each i, we assume that either K_i is open or it is a closed subspace. Then the following conditions are equivalent:*

- $\bigcap_{i=1}^{n} K_i = \emptyset$.
- *There exist linear functionals $f_i \in K_i^*$ with $i = 1, \ldots, n$, not all zero, such that $f_1 + f_2 + \cdots + f_n = 0$.*

Here, for any i, we have denoted by K_i^* the dual cone to K_i, that is, $K_i^* := \{f \in X' : f(e) \geq 0 \ \ \forall e \in K_i\}$. For the proof, see for instance Lemma 5.11 of [17].

Note that the adjoint system

$$\begin{cases} -\nu\Delta\varphi_i + (u\cdot\nabla)\varphi_i + (\nabla u)^t\varphi_i + \nabla q_i = (u - u_{id})1_{\mathcal{O}_i} \quad x \in \Omega, \\ \nabla\cdot\varphi_i = 0, \quad x \in \Omega, \\ \varphi_i = 0, \quad x \in \partial\Omega, \end{cases}$$

can be equivalently rewritten in the form

$$\begin{cases} -\nu\Delta\varphi_i - D\varphi_i \cdot u + \nabla q_i = (u - u_{id})1_{\mathcal{O}_i} \quad x \in \Omega, \\ \nabla \cdot \varphi_i = 0, \quad x \in \Omega, \\ \varphi_i = 0, \quad x \in \partial\Omega, \end{cases}$$

where $D\varphi_i = \dfrac{1}{2}(\nabla\varphi_i + (\nabla\varphi_i)^t)$ and the pressure has been redefined. This observation will be frequently used in the following Lemma and in the following Section.

Lemma 3.4.6. *Let $(f_1, f_2) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$ be given and let $u \in H^2(\Omega)^N \cap V$ be (together with p) an associated solution to* (3.51)*. Let $R : V \mapsto V$ be the linear mapping defined by $R\varphi := (\nu A)^{-1}(-D\varphi \cdot u)$, where A is the (extended) Stokes operator in V. Then, for any non-empty open set $\omega \subset \Omega$,*

$$\left[\varphi\right]_\omega := \| \varphi|_\omega \|_{L^2(\omega)} \tag{3.56}$$

is a norm in $N(Id + R)$.

Proof:

As already said, one has $u \in H^2(\Omega)^N, \nabla u \in L^6(\Omega)^{N\times N}$ and $u \in L^\infty(\Omega)^N$. Consequently, R is well defined and compact. We only have to prove that, for every $\varphi \in N(Id + R)$ with $\varphi|_\omega = 0$, one has $\varphi \equiv 0$.

Thus, let us assume that $\varphi \in V$ and $\varphi + (\nu A)^{-1}(-D\varphi \cdot u)$, that is,

$$\begin{cases} -\nu\Delta\varphi - D\varphi \cdot u + \nabla q = 0 \quad x \in \Omega, \\ \nabla \cdot \varphi = 0 \quad x \in \Omega, \\ \varphi = 0 \quad x \in \partial\Omega \end{cases}$$

and $\varphi = 0$ a.e. in ω. Then, we can use the unique continuation property of the Stokes system with coefficients in L^∞ (see [12]) and deduce that, certainly, φ vanishes identically. □

Lemma 3.4.7. *Let (f_1, f_2), (u, p) and ω be as in Lemma 3.4.6. Let the (φ_n, ψ_n) be given in $V \times V$ with $\varphi_n|_\omega \to \varphi|_\omega$ in $L^2(\omega)$, $\psi_n := \varphi_n + R\varphi_n$ and $\psi_n \to \psi \in V$. Then $\|\varphi_n\|_{H_0^1} \leq C$ for some positive constant C independent on n.*

Proof:

First, note that, since R is a compact operator, $\dim(N(Id + R)) < +\infty$ and $R(Id + R)$ is closed, in view of Fredholm's Alternative Theorem (see for instance [7]).

Now, let $\tilde{\varphi}_n$ be, for each n, the unique function in $N(Id + R)$ satisfying

$$\|\varphi_n - \tilde{\varphi}_n\|_{H_0^1} = \inf_{\tilde{\varphi}\in N(Id+R)} \|\varphi_n - \tilde{\varphi}\|_{H_0^1}.$$

Then, $\psi_n = (\varphi_n - \tilde{\varphi}_n) + R(\varphi_n - \tilde{\varphi}_n)$.

Also, $\|\varphi_n - \tilde{\varphi}_n\|_{H_0^1}$ is bounded by a constant C. Indeed, if this were not the case, we could assume that

$$\|\varphi_n - \tilde{\varphi}_n\|_V = \text{dist}(\varphi_n, N(Id + R)) \to +\infty.$$

Let us introduce

$$\zeta_n := \frac{\varphi_n - \tilde{\varphi}_n}{\|\varphi_n - \tilde{\varphi}_n\|_{H_0^1}}.$$

Then

$$\zeta_n + R\zeta_n = \frac{\psi_n}{\|\varphi_n - \tilde{\varphi}_n\|_{H_0^1}}. \tag{3.57}$$

Since the $\|\zeta_n\|_V = 1$ and $\psi_n \to \psi$ in V, we see from (3.57) that $\zeta_n \to \zeta$ for some $\zeta \in N(Id+R)$. So,

$$\left\|\varphi_n - \left(\tilde{\varphi}_n + \|\varphi_n - \tilde{\varphi}_n\|_{H_0^1}\right)\right\|_{H_0^1} = \|\varphi_n - \tilde{\varphi}_n\|_{H_0^1}\|\zeta_n - \zeta\|_{H_0^1} \geq \|\varphi_n - \tilde{\varphi}_n\|_{H_0^1}$$

for all n. But this is an absurd, since $\zeta_n \to \zeta$ in V.

Finally, let us deduce that $\|\varphi_n\|_{H_0^1} \leq C$. We can write that $\varphi_n = \tilde{\varphi}_n + \eta_n$, with

- $\|\eta_n\|_{H_0^1} = \|\varphi_n - \tilde{\varphi}_n\|_{H_0^1} \leq C$.
- $\|\tilde{\varphi}_n\|_{H_0^1} \leq C[\tilde{\varphi}_n]_\omega$ (note that the $\tilde{\varphi}_n$ belong to a finite dimensional space) and

$$[\tilde{\varphi}_n]_\omega \leq [\varphi_n]_\omega + [\eta_n]_\omega \leq [\varphi_n]_\omega + C\|\eta_n\|_{H_0^1} \leq C.$$

This ends the proof. □

Now, we can prove the main result in this section.

Proof of Theorem 3.4.4:

As already mentioned, in order to prove this result we will use the Dubovitsky-Milyoutin formalism (and, more precisely, Lemma 3.4.5) for each functional. Thus, let $(\widehat{f_1}, \widehat{f_2})$ be a Nash equilibrium for (3.51) and (3.52). Let us introduce the sets

$$\mathcal{F}_1 = \{(f_1, u) \in L^2(\omega_1)^N \times D(A) : \ (f_1, \widehat{f_2}, u) \ \text{ solves (3.51) } \text{ with some } p \text{ for } f_2 = \widehat{f_2}\}$$

and

$$\mathcal{F}_2 = \{(f_2, u) \in L^2(\omega_2)^N \times D(A) : \ (\widehat{f_1}, f_2, u) \ \text{ solves (3.51) } \text{ with some } p \text{ for } f_1 = \widehat{f_1}\}.$$

There exists a strong solution $(\widehat{u}, \widehat{p})$ to (3.51) such that $(\widehat{f_1}, \widehat{u})$ and $(fd, \widehat{u})$ respectively solve the following extremal problems:

$$\begin{cases} \text{Minimize } \widehat{J_1}(f_1, u) = J_1(f_1, \widehat{f_2}, u) \\ \text{Subject to } \ (f_1, u) \in \mathcal{F}_1 \end{cases} \tag{3.58a}$$

and

$$\begin{cases} \text{Minimize } \widehat{J_2}(f_2, u) = J_1(\widehat{f_1}, f_2, u) \\ \text{Subject to } \ (f_2, u) \in \mathcal{F}_2. \end{cases} \tag{3.58b}$$

Now, we introduce some cones associated to (3.58a) and (3.58b). To this end, let us set

$$M_1(f_1, u) = -\nu Au + B(u, u) - f_1 1_{\omega_1} - \widehat{f_2} 1_{\omega_2} \quad \forall (f_1, u) \in L^2(\omega_1)^N \times V$$

and

$$M_2(f_2,u) = -\nu Au + B(u,u) - \widehat{f}_1 1_{\omega_1} - f_2 1_{\omega_2} \quad \forall (f_2,u) \in L^2(\omega_2)^N \times V.$$

It is then clear that M_1 and M_2 are V'- valued continuosly differentiable mappings, with

$$M_1'(f_1,u)(h_1,w) = \nu Aw + B(u,w) + B(w,u) - h_1 1_{\omega_1} \quad \forall (h_1,w) \in L^2(\omega_1)^N \times V$$

and

$$M_2'(f_2,u)(h_2,w) = \nu Aw + B(u,w) + B(w,u) - h_2 1_{\omega_2} \quad \forall (h_2,w) \in L^2(\omega_2)^N \times V.$$

We also have

$$M_1'(\widehat{f}_1,\widehat{u})^* \varphi_1 = (-\varphi_1|_{\omega_1}, \nu A\varphi_1 - B(\widehat{u},\varphi_1) + C(\widehat{u},\varphi_1)) \quad \forall \varphi_1 \in V$$

and

$$M_2'(\widehat{f}_2,\widehat{u})^* \varphi_2 = (-\varphi_2|_{\omega_2}, \nu A\varphi_2 - B(\widehat{u},\varphi_2) + C(\widehat{u},\varphi_2)) \quad \forall \varphi_2 \in V.$$

In connection with (3.58a) and (3.58b), we define the cones of descent directions

$$D_i = \{(h_i,v) \in L^2(\omega_i)^N \times V : \ \langle \widehat{J}_i'(\hat{f}_i,\widehat{u}), (h_i,v)\rangle < 0\}$$

and the spaces of tangent directions

$$T_i = \{(h_i,v) \in L^2(\omega_i)^N \times V : \ M_i'(\hat{f}_i,\widehat{u})(h_i,v) = 0\} = N(M_i'(\hat{f}_i,\widehat{u}))$$

for $i = 1, 2$.

Let us prove that $R(M_i'(f,u)^*)$ is closed in $L^2(\omega_i)^N \times V$ for each $i = 1, 2$. Indeed, if $(h,w) \in \overline{R(M_i'(f_i,u)^*)}$, there exists fields $\varphi_n \in V$ such that

$$-\varphi_n|_{\omega_i} \to h \quad \text{in } L^2(\omega_i)^N,$$

$$\nu A\varphi_n - B(\widehat{u},\varphi_n) + C(\widehat{u},\varphi_n) \to w \quad \text{in } V'.$$

Let us introduce the linear mapping $S : V \mapsto V'$, with $S\varphi := \nu A\varphi - B(\widehat{u},\varphi) + C(\widehat{u},\varphi)$. It is not difficult to see that $S = \nu A(Id + R)$, where R is the linear operator introduced in Lemma 3.4.6. Therefore, by Fredholm's Alternative Theorem, we have that $R(S)$ is closed and, from Lemma 3.4.7, we find that the φ_n are uniformly bounded in V. As an inmediate consequence, we see that, at least for a subsequence, $\varphi_n \to \varphi$ weakly in V, $-\varphi|_{\omega_i} = h$ and $S\varphi = w$. In other words, $(h,w) \in R(M_i'(\hat{f}_i,\widehat{u})^*)$.

At this moment, we can apply the Dubovitskiy-Milyoutin formalism to (3.58a) and (3.58b) and deduce that the descent cone D_i must be disjoint of the tangent space T_i for each $i = 1, 2$ (see [17]):

$$D_i \cap T_i = \emptyset \quad \text{for } i = 1, 2.$$

In view of Lemma 3.4.5, there exist $(h_i', w_i') \in D_i^*$ and $(h_i', w_i') \in T_i^*$, not both zero, such that

$$(h_i', w_i') + (h_i', w_i') = (0,0). \tag{3.59}$$

Taking into account the definitions of D_i and T_i and the fact that $R(M_i'(\hat{f}_i,\widehat{u})^*)$ is closed, we see at once that

$$D_i^* = \{-\lambda \widehat{J}_i'(\hat{f}_i,\widehat{u}) : \lambda \geq 0\} \quad \text{and}$$
$$T_i^* = (N(M_i'(\hat{f}_i\widehat{u})))^* = R(M_i'(f_i,u)^*).$$

Hence, there exists $\varphi_1, \varphi_2 \in V$ such that

$$h_i' = -\left.\varphi_i\right|_{\omega_i}, \quad w_i' = \nu A\varphi_i - B(\widehat{u},\varphi_i) + C(\widehat{u},\varphi_i) \tag{3.60}$$

and (3.59) can be rewritten in the form

$$\int_{\mathcal{O}_i} (\widehat{u} - u_{id})w_i + \mu \int_{\omega_i} \hat{f}_i h_i = \int_{\omega_i} h_i' h_i \; + \langle w_i', w_i\rangle \quad \forall (h_i,w_i) \in L^2(\omega_i)^N \times V. \tag{3.61}$$

Now, taking $w_i = 0$ we find that $h_i' = \mu f_i$.

On the other hand, taking $h_i = 0$ and recalling (3.60), we see that $\left.\varphi_i\right|_{\omega_i} = -\mu\hat{f}_i$ for each $i = 1, 2$ and

$$\int_\Omega \Big(\nu\nabla\varphi_i\cdot\nabla w_i - (\widehat{u}\cdot\nabla)\varphi_i\cdot w_i + (\nabla\widehat{u})^t\varphi_i\cdot w_i\Big) = a\int_{\mathcal{O}_i}(\widehat{u}-u_{id})w_i \quad \forall w_i \in V.$$

Therefore, φ_i solves, together with some $q_i \in L^2(\Omega)$, the following linear system for each $i = 1, 2$

$$\begin{cases} -\nu\Delta\varphi_i - (\widehat{u}\cdot\nabla)\varphi_i + (\nabla\widehat{u})^t\varphi_i + \nabla q_i = a(\widehat{u}-u_{id})1_{\mathcal{O}_i} \quad x\in\Omega, \\ \nabla\cdot\varphi_i = 0 \quad x\in\Omega, \\ \varphi_i = 0 \quad x\in\partial\Omega \end{cases} \tag{3.62a}$$

and, furthermore,

$$\hat{f}_i = -\frac{1}{\mu}\left.\varphi_i\right|_{\omega_i} \tag{3.62b}$$

From (3.51), (3.62a) and (3.62b), we deduce that $\widehat{f_1}, \widehat{f_2})$ is a Nash quasi-equilibrium and the proof is done. □

3.4.3. Algorithms

This section is devoted to present three iterative algorithms (similar to those in Sections 3.2.3 and 3.3.3) for the computation of Nash equilibria for (3.51) and (3.52). Additionally, we will present a new algorithm based on Newton's method.

ALG 7: Fixed-Point-like Method

(a) Choose $(f_1^0, f_2^0) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$ and $u^0 \in V$.

(b) Then, for given $n \geq 0$, $(f_1^n, f_2^n) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$ and $u^n \in V$, compute a solution (u^{n+1}, p^{n+1}) to

$$\begin{cases} -\nu\Delta u^{n+1} + (u^n \cdot \nabla)u^{n+1} + \nabla p^{n+1} = f_1^n 1_{\omega_1} + f_2^n 1_{\omega_2} \quad x \in \Omega, \\ \nabla \cdot u^{n+1} = 0 \quad x \in \Omega, \\ u^{n+1} = 0 \quad x \in \partial\Omega, \end{cases} \tag{3.63}$$

a solution $(\varphi_i^{n+1}, q_i^{n+1})$ to the system

$$\begin{cases} -\nu\Delta\varphi_i^{n+1} - D\varphi_i^{n+1} \cdot u^{n+1} + \nabla q_i^{n+1} = (u^{n+1} - u_{id})1_{\mathcal{O}_i} \quad x \in \Omega, \\ \nabla \cdot \varphi_i^{n+1} = 0 \quad x \in \Omega, \\ \varphi_i^{n+1} = 0 \quad x \in \partial\Omega \end{cases} \tag{3.64}$$

and, finally, set

$$f_i^{n+1} = -\frac{a}{\mu}\, \varphi_i^{n+1}\big|_{\omega_i}. \tag{3.65}$$

ALG 8: Optimal-Step-Gradient-like Method

(a) Choose $(f_1^0, f_2^0) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$.

(b) Then, for given $n \geq 0$ and $(f_1^n, f_2^n) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$, compute the solution (u^{n+1}, p^{n+1}) to (3.63) and the solution $(\varphi_i^{n+1}, q_i^{n+1})$ to (3.64) and set

$$f_i^{n+1} = f_i^n - \rho_i^n g_i^{n+1}, \tag{3.66}$$

where

$$g_i^{n+1} = a\, \varphi_i^{n+1}\big|_{\omega_i} + \mu f_i^n \tag{3.67}$$

and

$$\begin{cases} \rho_1^n = \arg\Big(\min\limits_{\rho\geq 0} J_1(f_1^n - \rho g_1^n, f_2^n)\Big), \\ \rho_2^n = \arg\Big(\min\limits_{\rho\geq 0} J_2(f_1^n, f_2^n - \rho g_2^n)\Big). \end{cases} \tag{3.68}$$

ALG 9: Optimal-Step-Conjugate-Gradient-like Method

(a) Choose $(f_1^0, f_2^0) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$.

(b) Then, for $n = 0$, perform one step of ALG 8 and set $d_i^0 = g_i^0$.

(c) Then, for given $n \geq 1$ and $(f_1^n, f_2^n) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$, compute the solution (u^{n+1}, p^{n+1}) to (3.63), the solution $(\varphi_i^{n+1}, q_i^{n+1})$ to (3.64) and then set

$$f_i^{n+1} = f_i^n - \rho_i^n d_i^n, \tag{3.69}$$

where

$$\begin{cases} d_i^n = g_i^n + \gamma_i^n d_i^{n-1}, \quad \gamma_i^n = \dfrac{(g_i^n - g_i^{n-1}, g_i^n)}{\|g_i^{n-1}\|^2}, \\ g_i^n = a\, \varphi_i^n|_{\omega_i} + \mu f_i^n \end{cases} \tag{3.70}$$

and

$$\begin{cases} \rho_1^n = \arg\Big(\min_{\rho\geq 0} J_1(f_1^n - \rho g_1^n, f_2^n)\Big), \\ \rho_2^n = \arg\Big(\min_{\rho\geq 0} J_2(f_1^n, f_2^n - \rho g_2^n)\Big). \end{cases} \tag{3.71}$$

Before presenting the results of some simulations, let us present another algorithm. It is based on Newton's iterates and aims to compute directly a solution to the optimality system (3.54). In practice, this method is much faster than ALG 8 and ALG 9 but, as is usual for Newton methods and variants, it needs a nontrivial starting process (see below).

ALG 10: Newton Method

We want to solve the problem (3.54) with $\nu = \tilde{\nu}$. We fix a decreasing factor $\tilde{a} \in (0,1)$ and we do as follows.

(a) Choose$(f_1^0, f_2^0) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$ and $\nu^0 \in \mathbb{R}^+$ and compute a solution (u^0, p^0) to

$$\begin{cases} -\nu^0 \Delta u^0 + \nabla p^0 = f_1^0 1_{\omega_1} + f_2^0 1_{\omega_2} \quad x \in \Omega, \\ \nabla \cdot u^0 = 0 \quad x \in \Omega, \\ u^0 = 0, \quad x \in \partial\Omega, \end{cases} \tag{3.72}$$

and a solution (φ_i^0, q_i^0) to

$$\begin{cases} -\nu^0 \Delta \varphi_i^0 + \nabla q_i^0 = (u^0 - u_{id}) 1_{\mathcal{O}_i} \quad x \in \Omega, \\ \nabla \cdot \varphi_i^0 = 0 \quad x \in \Omega, \\ \varphi_i^0 = 0, \quad x \in \partial\Omega \end{cases} \tag{3.73}$$

and take

$$f_i^0 = -\frac{a}{\mu}\, \varphi_i^0\big|_{\omega_i} \quad \text{and} \quad \nu^1 = \max\{\tilde{\nu}, \tilde{a}\nu^0\}.$$

(b) Then, for given $n \geq 0$, ν^n and $(f_1^n, f_2^n) \in L^2(\omega_1)^N \times L^2(\omega_2)^N$, (u^n, p^n) and (φ_i^n, q_i^n), do the following

(b.1) Take $f_i^{n,0} = -\frac{a}{\mu}\varphi_i^n\,|_{\omega_i}$, $u^{n,0} = u^n$, $\varphi_i^{n,0} = \varphi_i^n$ and $\nu^{n+1} = \max(\tilde{a}\nu^n, \tilde{\nu})$.

(b.2) For given $k \geq 0$, $f_i^{n,k}$, $u^{n,k}$ and $\varphi_i^{n,k}$, set

$$F^{n,k} := -\nu^{n+1}\Delta u^{n,k} + (u^{n,k}\cdot\nabla)u^{n,k} - f_1^{n,k}1_{\omega_1} - f_2^{n,k}1_{\omega_2}$$

and

$$G_i^{n,k} := -\nu^{n+1}\Delta\varphi_i^{n,k} - (u^{n,k}\cdot\nabla)\varphi_i^{n,k} + (\nabla u^{n,k})^t\varphi_i^{n,k} - (u^{n,k} - u_{id})1_{\mathcal{O}_i},$$

compute the solution $(v^k, h^k, \psi_i^k, \eta_i^k)$ to

$$\begin{cases} -\nu^{n+1}\Delta v^k + (u^{n,k}\cdot\nabla)v^k + (v^k\cdot\nabla)u^{n,k} + \nabla h^k = F^{n,k} \quad x\in\Omega, \\ \nabla\cdot v^k = 0 \quad x\in\Omega, \\ v^k = 0, \quad x\in\partial\Omega, \\ -\nu^{n+1}\Delta\psi_i^k - (u^{n,k}\cdot\nabla)\psi_i^k - (v^k\cdot\nabla)\varphi_i^{n,k} + (\nabla u^{n,k})^t\psi_i^k + (\nabla v^k)^t\varphi_i^{n,k} + \nabla\eta_i^k = G_i^{n,k} \quad x\in\Omega, \\ \nabla\cdot\psi_i^k = 0 \quad x\in\Omega, \\ \psi_i^k = 0, \quad x\in\partial\Omega \end{cases} \tag{3.74}$$

and take:

$$u^{n,k+1} = u^{n,k} - v^k, \quad \varphi_i^{n,k+1} = \varphi_i^{n,k} - \psi_i^k. \tag{3.75}$$

Note that ALG 7 and ALG 10 are conceived to compute a solution to the optimality system (3.54) that, possibly, is not be a minimizer of $\tilde{J}_i$. Thus, they are expected to furnish numerical approximations of Nash quasi-equilibria. From the viewpoint of the Calculus of Variations, ALG 7 and ALG 10 are related to the so called "indirect method". Contrarily, ALG 8 and ALG 9 intend to provide (numerical approximations of) controls in minimizing sequences. Accordingly, they correspond to realizations of the "direct method" of Calculus of Variations.

3.4.4. Numerical experiments

Now, in order to illustrate the behavior of the previous algorithms, we will present the results of some numerical experiments. Very precisely, we want to compute numerical approximations of Nash equilibria for (3.51) and (3.52).

The computations have been performed with the FreeFem++ package (see [19]).

Test 1

In this first test, our domain is composed of two rectangles $\mathcal{O}_1$ and $\mathcal{O}_2$ and we assume that the controls act on the narrow band ω_1 and ω_2. In order to solve numerically the systems (3.63), (3.64), (3.72), (3.73) and (3.74), we have to fix a mesh and a finite element method. We have used the mesh depicted in Fig. 3.1 and a mixed finite element formulation with continuous piecewise $\mathbb{P}_1$-bubble and $\mathbb{P}_1$ functions respectively for the velocity field and the pressure; for details, see [15, 18].

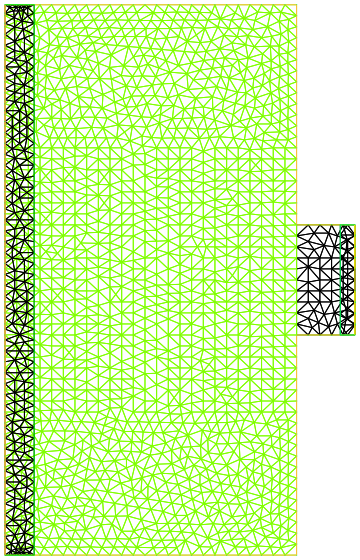


Figure 3.1: The domain and the "rough" mesh; Ω is composed of the band ω, the large rectangle $\mathcal{O}_1$ and the small rectangle $\mathcal{O}_2$. Number of nodes: 1519 . Number of triangles: 2876.

The data u_{id} are the following: $u_{1d} = \nabla \times \psi_{1d}$, where ψ_{1d} is the solution to the problem

$$\begin{cases} -\Delta\psi_{1d} = 1, & x \in \mathcal{O}_1, \\ \psi_{1d} = 0, & x \in \partial\mathcal{O}_1 \end{cases}$$

and $u_{2d} \equiv 0$. That means that the "desired" (ideal) configurations correspond to a uniformly rotating flow in $\mathcal{O}_1$ and a fluid at rest in $\mathcal{O}_2$ (see Fig. 3.2).

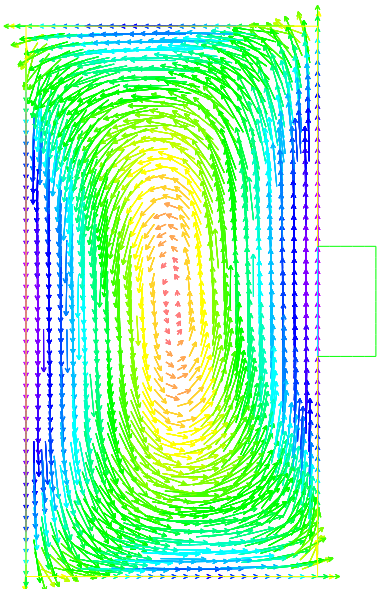


Figure 3.2: The function u_{1d}.

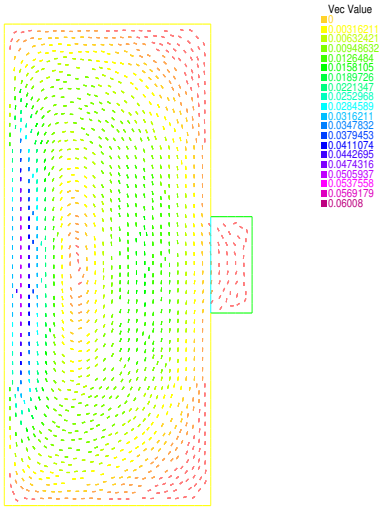
For **ALG 7**, **ALG 8** and **ALG 9**, the stopping test has been

$$\|u^{n+1} - u^n\|_{L^\infty} + \|p^{n+1} - p^n\|_{L^\infty} + \|q^{n+1} - q^n\|_{L^\infty} \leq \varepsilon,$$

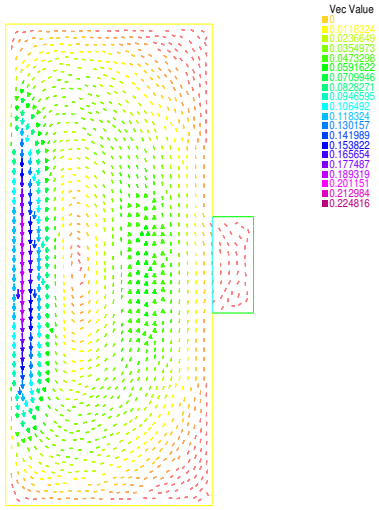
with $\varepsilon = 10^{-6}$. This has also been the stopping criterion for the external iterates in **ALG 10**. For the internal loops (indexed by k), the stopping test has been

$$\|u^{n,k+1} - u^{n,k}\|_{L^\infty} + \|\varphi^{n,k+1} - \varphi^{n,k}\|_{L^\infty} \leq \varepsilon.$$

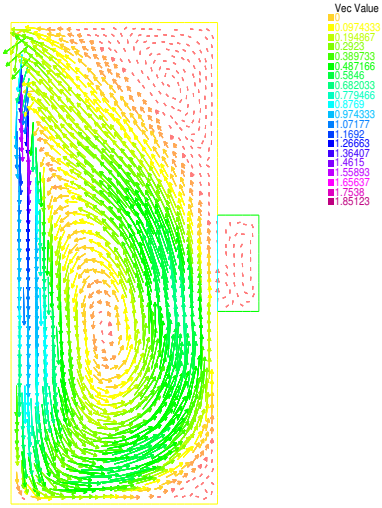
Some results are depicted in Fig. 3.3–3.4.



(a) $\nu = 1$.

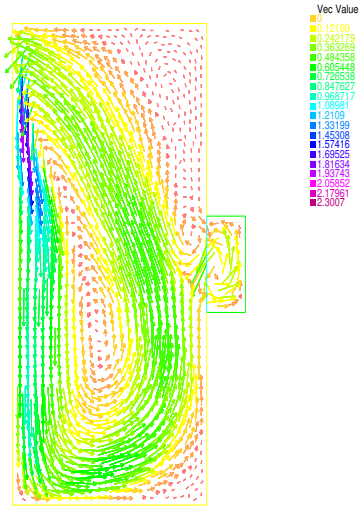


(b) $\nu = 0.5$

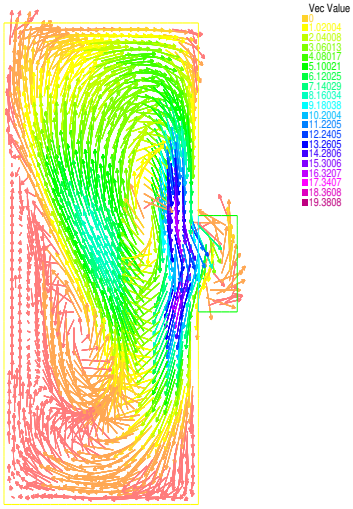


(c) $\nu = 0.16$

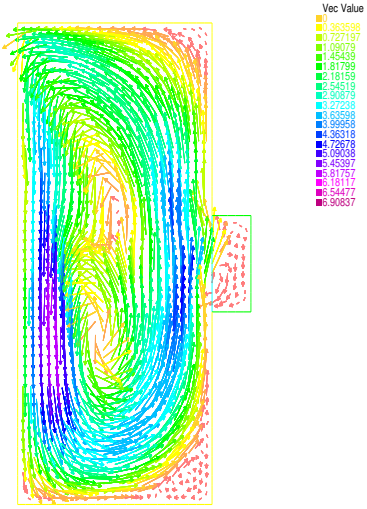
Figure 3.3: The final velocity fields computed with **ALG 8** for various ν. Here, $a = 1.2$, $\mu = 2 - a = 0.8$.



(a) Velocity



(b) Adjount State of u_{1d}



(c) Adjount State of u_{2d}

Figure 3.4: The final velocity field and the adjoint states computed with **ALG 10** for $\nu = 0.05$. Now, $a = 1.5$ and $\mu = 2 - a = 0.5$.

In order to compare the behavior of the Gradient, the Gonjugate Gradient and Newton methods, we present some related numerical values in the Tables in Fig. 3.5. Specifically, we have indicated there the numbers of iterates needed by each method to fulfill the stopping test and produce an approximation with error less or equal than 10^{-6}.

ν \ a	0.1	0.8	1.5	1.8
1	3	3	5	5
0.5	3	5	9	9
0.1	7	17	51	129
0.05	12	55	2206	4997

(a) Gradient (**ALG 8**)

ν \ a	0.1	0.8	1.5	1.8
1	2	3	3	4
0.5	2	4	4	5
0.1	4	9	13	17
0.05	8	15	21	37

(b) Conjugate Gradient (**ALG 9**)

ν \ a	0.1	0.8	1.5	1.8
1	2	3	3	3
0.5	3	3	3	4
0.1	27	31	33	35
0.05	234	247	259	--

(c) Newton (**ALG 10**)

Figure 3.5: Number of iterates for various values of ν, a and $\mu = 2 - a$.

Also, we have included in Fig. 3.6 a Table with a comparison of the computation times and required numbers of iterates for each method.

	CPU TIME	ITERATES
GRADIENT	28305.9	2206
CONJUGATE GRADIENT	3920.53	21
NEWTON	848.034	259

Figure 3.6: Computation times (in seconds) and numbers of iterates to reach an error less than $\varepsilon = 10^{-6}$ for $\nu = 0.05$, $a = 1.5$ and $\mu = 2 - a = 0.5$.

Test 2

In this second test, our aims is to control the flow of a fluid in a pipe. We assume that the fluid enters the domains with a parabolic profile (as we can see in Figure 3.9). In which, at one point, the pipe bifurcates in two. Before this bifurcation at the upper and lower end of the pipe there are two control regions (ω_1 and ω_2, respectively) in which we can interfere or remove fluid so that once the fluid forks is as close as possible to the desired situation in the two bifurcations ($\mathcal{O}_1$ and $\mathcal{O}_2$), as we see in Figure 3.8. To get an idea of how the domain is we can see Figure 3.7.

Our goal is to solve a control problem which we approximate a border cotrol problem being the border control the solution of our problem restricted to the desired border of our control region.

In order to solve numerically the systems (3.63), (3.64), (3.72), (3.73) and (3.74), we have to fix a mesh and a finite element method. We have used the mesh depicted in Fig. 3.7 and a mixed finite element formulation with continuous piecewise $\mathbb{P}_1$-bubble and $\mathbb{P}_1$ functions respectively for the velocity field and the pressure.

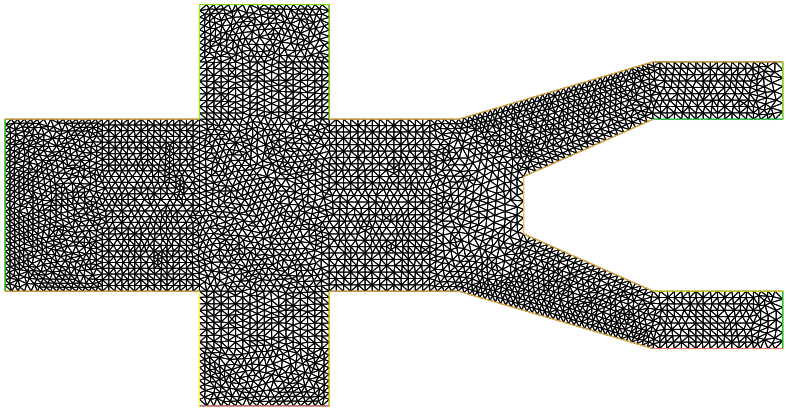


Figure 3.7: The domain and the "rough" mesh; Ω is composed of the two rectangles ω_1, ω_2, the upper bifurcation is $\mathcal{O}_1$ and the lower is $\mathcal{O}_2$. Number of nodes: 1993 . Number of triangles: 3685.

The data u_{id} are the following: $u_{1d} = (q_1(y - y_0)(y_1 - y), 0)$ with $y_0 < y_1, q_1 \in \mathbb{R}$ and $u_{2d} \equiv 0$. That means that the "desired" configuration corresponds to a parabolic profile flow in $\mathcal{O}_1$ and a fluid at rest in $\mathcal{O}_2$ (see Fig. 3.8).

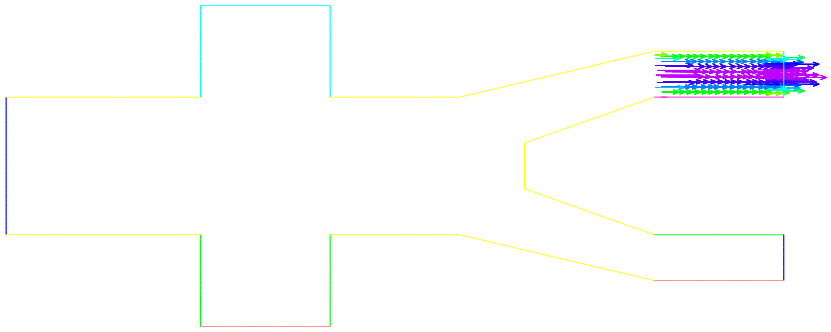


Figure 3.8: The function u_{1d}.

We consider that the fluid enter to the pipe with a parabolic profile as $u_0 = (q_0(y - y_2)(y_3 - y)$ with $y_2 < y_3, q_0 \in \mathbb{R}$. Also we suppose that we want all the fluid that enters the pipe then to come out after the bifurcation, for this it must happen that $q_0 = 1/27q_1$ and in our case we have taken $q_1 = 2700$.

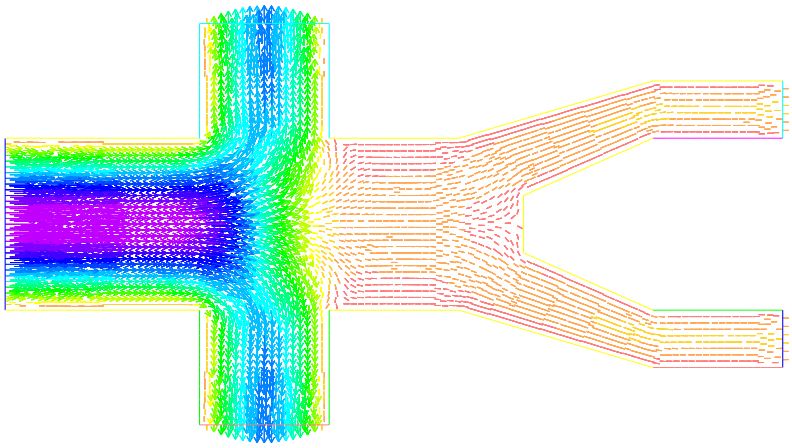


Figure 3.9: The parabolic profile which enters by the pipe.

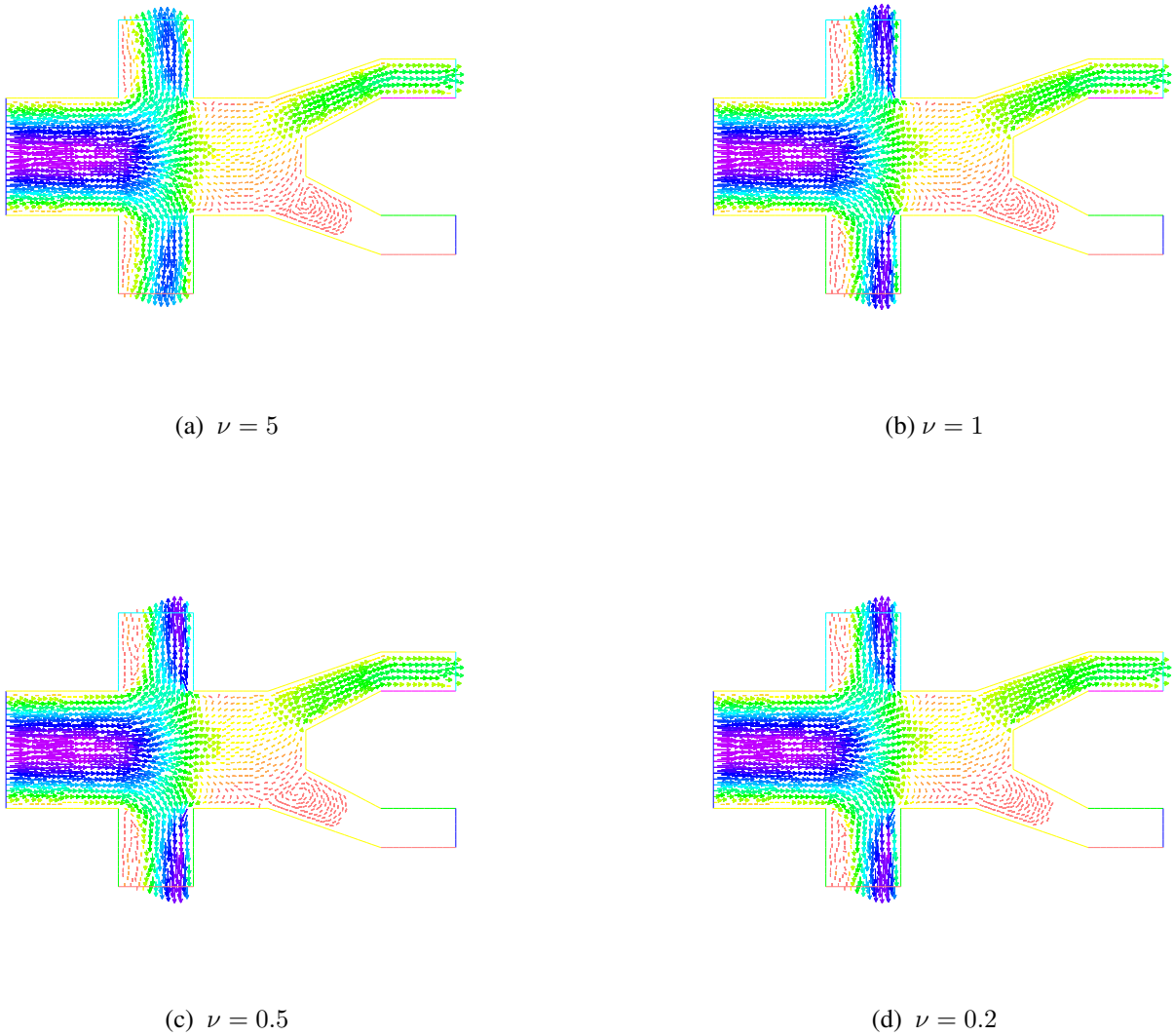
For **ALG 10**, the stopping test has been

$$\|u^{n+1} - u^n\|_{L^\infty} + \|p^{n+1} - p^n\|_{L^\infty} + \|q^{n+1} - q^n\|_{L^\infty} \leq \varepsilon,$$

with $\varepsilon = 10^{-12}$. This has also been the stopping criterion for the external iterates in **ALG 10**. For the internal loops (indexed by k), the stopping test has been

$$\|u^{n,k+1} - u^{n,k}\|_{L^\infty} + \|\varphi^{n,k+1} - \varphi^{n,k}\|_{L^\infty} \leq \varepsilon.$$

Some results are depicted in Fig. 3.10.



(a) $\nu = 5$ (b) $\nu = 1$

(c) $\nu = 0.5$ (d) $\nu = 0.2$

Figure 3.10: The final velocity fields computed with **ALG 10** for various ν in the case $a = 1.95$.

Test 3

In this new test we have a new pipe on which a fluid circulates. This fluid enters with a parabolic profile (as we can see in Figure 3.13) and exits through two areas before the areas where we look for a desired state ($\mathcal{O}_1$ and $\mathcal{O}_2$) . Our objectives are to know how we have to act in the two specific control zones (ω_1 and ω_2, respectively) to get the fluid reaches a desired state, as we see in Figure 3.12. To get an idea of how the domain is we can see Figure 3.11.

In order to solve numerically the systems (3.63), (3.64), (3.72), (3.73) and (3.74), we have to fix a mesh and a finite element method. We have used the mesh depicted in Fig. 3.11 and a mixed finite element formulation with continuous piecewise $\mathbb{P}_1$-bubble and $\mathbb{P}_1$ functions respectively for

the velocity field and the pressure.

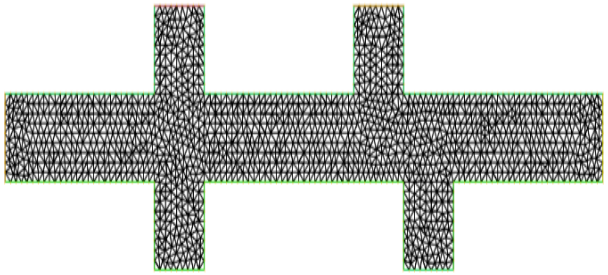


Figure 3.11: The domain and the "rough" mesh; Ω is composed of the two first rectangles ω_1, ω_2, the second upper rectangle $\mathcal{O}_1$ and the second lower rectangle $\mathcal{O}_2$. Number of nodes: 1541 . Number of triangles: 2774.

The data u_{id} are the following: $u_{1d} = \nabla \times \psi_{1d}$, where ψ_{1d} is the solution to the problem

$$\begin{cases} -\Delta\psi_{1d} = 1, & x \in \mathcal{O}_1, \\ \psi_{1d} = 0, & x \in \partial\mathcal{O}_1 \end{cases}$$

and $u_{2d} \equiv 0$. That means that the "desired" (ideal) configurations correspond to a uniformly rotating flow in $\mathcal{O}_1$ and a fluid at rest in $\mathcal{O}_2$ (see Fig. 3.12).

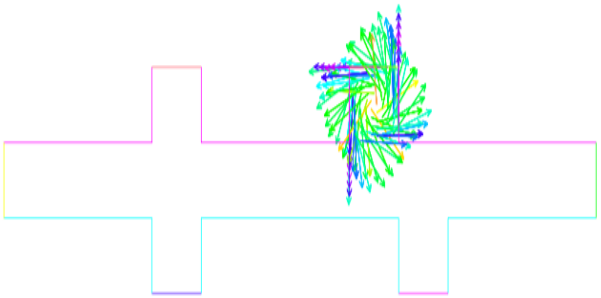


Figure 3.12: The function u_{1d}.

We consider that the fluid enter to the pipe with a parabolic profile as $u_0 = (q_0(y - y_2)(y_3 - y)$ with $y_2 < y_3 \in \mathbb{R}$ and we take $q_0 = 100$.

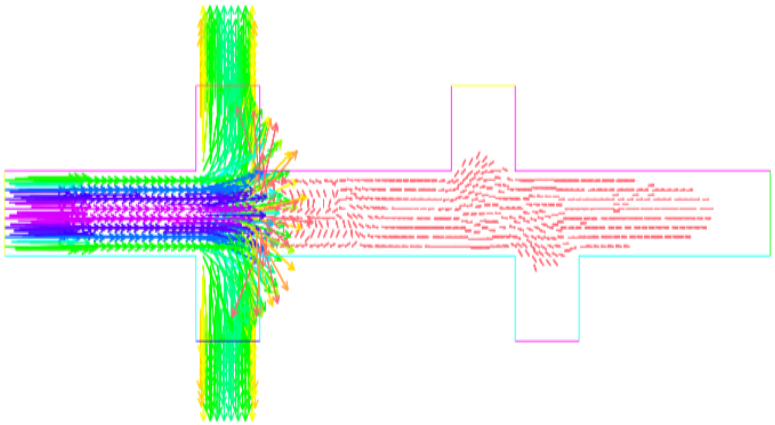


Figure 3.13: The parabolic profile which enters by the pipe.

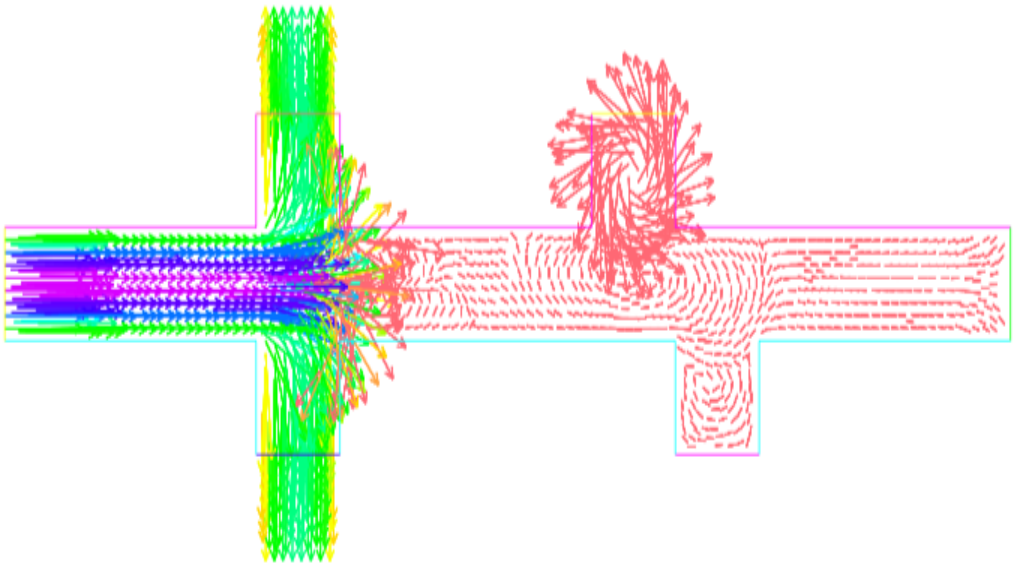
For **ALG 10** the stopping test has been

$$\|u^{n+1} - u^n\|_{L^\infty} + \|p^{n+1} - p^n\|_{L^\infty} + \|q^{n+1} - q^n\|_{L^\infty} \leq \varepsilon,$$

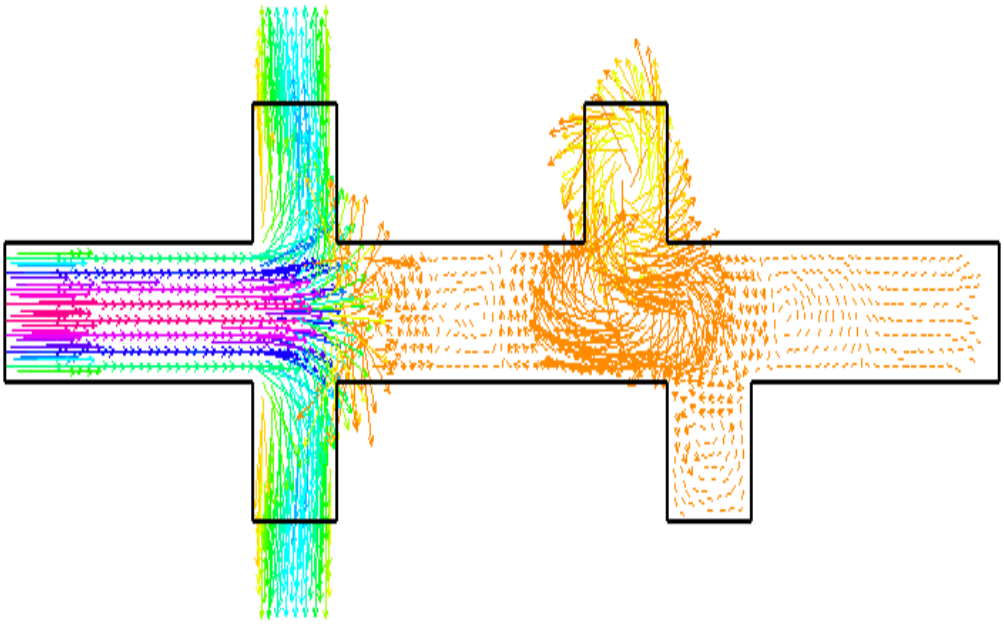
with $\varepsilon = 10^{-12}$. For the internal loops (indexed by k), the stopping test has been

$$\|u^{n,k+1} - u^{n,k}\|_{L^\infty} + \|\varphi^{n,k+1} - \varphi^{n,k}\|_{L^\infty} \leq \varepsilon.$$

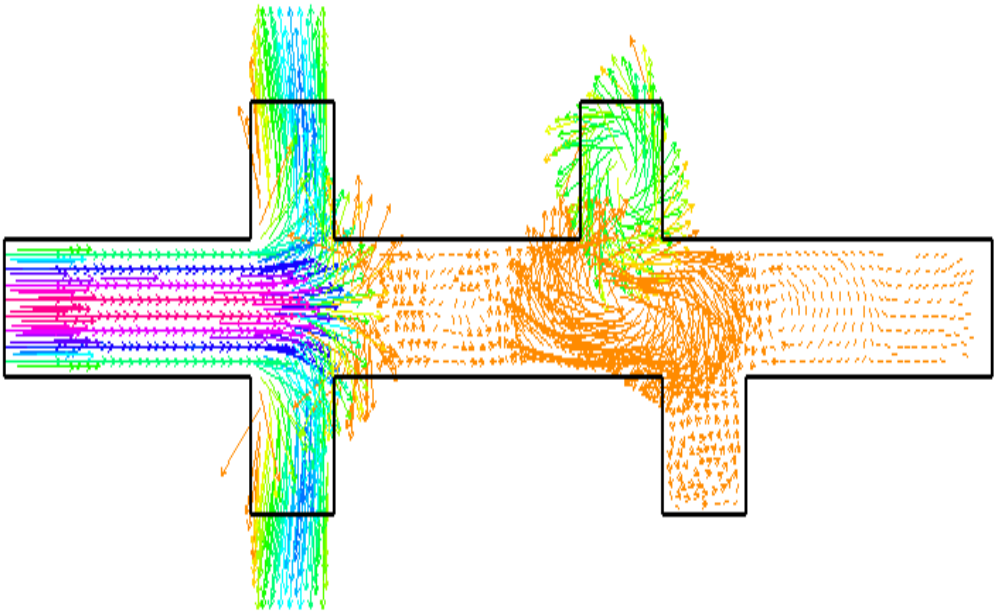
Some results are depicted in Fig. 3.14.



(a) $\nu = 5$



(b) $\nu = 1$



(c) $\nu = 0.5$

Figure 3.14: The final velocity fields computed with **ALG 10** for various ν in the case $a = 1.99$.

Conclusions

To conclude the work we would like to highlight the objectives we set and that we have been able to answer and expose some questions that are still open.

First, we posed a problem of linear parabolic control and looked for two Nash equilibria for this problem. In this case our theoretical study is finished and we propose different approximation algorithms although we have not been able to test convergence results for the algorithms.

In a second part, we pose the same problem but now in a semi-linear case. Here, we use the Lemma 3.3.5 to be able to demonstrate the existence of quasi-equilibria while a result on the existence of Nash equilibria without using quasi-equilibria we have not been able to prove it. It is important to note that the existence of quasi-equilibria of Nash is independent of the size of a/μ. Also, the algorithms, as in the linear case, we have not tested any convergence results.

Finally, we have extended the previous study by applying it now to the Navier-Stokes case. In this case the theoretical study is not as extensive as the previous cases due to the great nonlinearity of the equation. An important result here is that a Nash equilibrium is a quasi-equilibrium which, thanks to Dubovitskii-Milyoutin Lemma. Here, in the numerical part, we have presented the algorithms and added several simulations for various situations and domains. These experiments are very visual and in them we can see how effectively we are able to control a fluid to take it to the desired situation from the beginning.

This concludes our work although there are still many questions and open questions.

Bibliography

[1] F. Abergel and R. Temam, *"On some control problems in fluid mechanics,"* Theoret. Comput. Fluid Dyn., **1** (1990), 303–325.

[2] G. Allaire and A. Craig, *"Numerical Analysis and Optimization. An Introduction to Mathematical Modelling and Numerical Simulation",* Oxford, London, 2007.

[3] L. J. Álvarez-Vázquez, N. García-Chan and A. Martínez, M. E. Vázquez-Méndez, *"Multiobjective Pareto-optimal control: an application to wastewater management,"* Comput Optim Appl, Berlin, **46** (2010), 135–157.

[4] H. Brézis, L. Nirenberg, *"Characterizations of the ranges of some nonlinear operators and applications to boundary value problems"*, Annali della Scuola Normale Superiore di Pisa, Classe di Scienze, IV, **5**, (1978), 225-326.

[5] J. L. Boldrini, B. M. Calsavara Caretta and E. Fernández-Cara, *"Some optimal control problems for a two-phase field model of solidification,"* Rev Mat Complut, **23** (2010), 49–75.

[6] J. L. Boldrini, E. Fernández-Cara and M. A. Rojas-Medar, *"An Optimal Control Problem for a Generalized Boussinesq Model: The Time Dependent Case,"* Rev. Mat. Complut., **20** (2007), 339–366.

[7] H. Brézis, *"Functional Analysis, Sobolev Sapces and Partial Differential Equations,"* Springer New York Dordrecht Heidelberg London, 2011.

[8] P. P. Carvalho and E. Fernández-Cara, *"On the Computation of Nash and Pareto Equilibria for some Bi-Objective Control Problems"*.

[9] E. Casas,*"The Navier-Stokes equations coupled with the heat equation: analysis and control"*, Control Cybernet., **23** (1994), 605–620.

[10] E. Casas, J. P. Raymond and H. Zidani, *"Pontryagin's principle for local solutions of control problems with mixed control-state constraints"*, SIAM J. Control Optim., **39** (2000), 1182–1203.

[11] Ph. E. Ciarlet, *"Introduction to numerical linear algebra and optimisation"*, Cambridge University Press, Cambridge, 1989.

[12] C. Fabre, *"Uniqueness results for Stokes equations and their consequences in linear and nonlinear control problems,"* Control, Optimisation and Calculus of Variations, **1** (1996), 267-302.

[13] E. Fernández-Cara, I. Marín-Gayte, *Theoretical and numerical bi-objective optimal control: Nash equilibria*, submited.

[14] T. M. Flett, *"Differential analysis: Differentiation, differential equations and differential inequalities"*, Cambridge University Press, Cambridge, 1980.

[15] A. V. Fursikov and O. Pironneau, *"Finite Element Methods for Navier-Stokes Equations"*, Annual Review of Fluid Mechanics, **24** (1992), 167-204.

[16] A. V. Fursikov, *"Optimal Control of Distributed Systems: Theory and Applications"*, American Mathematical Society Boston, MA, USA 2000.

[17] I. V. Girsanov, *"Lectures on mathematical theory of extremum problems,"* Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, **67** (1972).

[18] R. Glowinski, *"Finite element methods for incompressible viscous flow,"* Handbook of Numerical Analysis, **9** (2003).

[19] F. Hecht, `http://www.freefem.org`

[20] S. Kakutani, *"A generalization of Brouwer's fixed point theorem"* , Duke Mathematical Journal 8 (3), 1941, 457–459.

[21] J. L. Lions, *"Contrôle de Pareto de systèmes distribués. Le cas d'évolution"*, C.R. Acad. Sci. Paris, Sér. I, **302** (1986), 413–417.

[22] J. L. Lions, *"Optimal control of systems governed by partial differential equations"*, Springer-Verlag, New York, 1971.

[23] J. L. Lions, *"Some remarks on Stackelberg's optimization"*, Math. Models Methods Appl. Sci., **4** (1994), 477–487.

[24] J. F. Nash, *"Noncooperative games"*, Ann. Math., **54** (1951), 286–295.

[25] V. Pareto, *"Cours d'économie politique"*, Rouge, Laussane, Switzerland, 1896.

[26] E. Polak, *"Optimization. Algorithm and consistent approximation"*, Springer-Verlag, NewvYork, 1997.

[27] H. Von Stalckelberg, *"Marktform und gleichgewicht,"* Springer, Berlin, Germany, 1934.

[28] R. Temam, *"Navier-Stokes equations. Theory and numerical analysis,"* Studies in Mathematics and Applications, North-Holland Publishing Co., Amsterdam, **2** (1977).

Capítulo 4

Analysis and numerical solution of some minimal time control problems

This chapter is devoted to the theoretical and numerical analysis of some minimal time control problems associated to linear and nonlinear differential equations. We start by studying simple cases concerning linear and nonlinear ODEs. Then, we deal with the heat equation. In all these situations, we analyze the existence of solution, we deduce optimality results and we present several algorithms for the computation of optimal controls. Finally, we illustrate the results with several numerical experiments. It is based on the paper [8].

4.1. Introduction

In this paper, we consider some minimal time control problems where the state is given by the solution to a differenttial equation. We will be concerned first with simple situations (corresponding to linear and nonlinear ODEs) and then with a linear PDE (more precisely, the classical heat equation). Similar problems have already been considered in [6, 13].

Optimal time control problems have a lot of interest from the theoretical and numerical viewpoints and also in connection with many applications; see to this respect for instance, [11, 14]. Frequently, the motivation is the need to act on the system in such a way that not only the evolution be as good as possible with a minimal effort but, also, this takes place in a shortest period of time. Some related works are [3, 7].

Here, we will present several theoretical and numerical results in a rather academic framework. More precisely, we will establish existence and optimality results, we will formulate somel iterative algorithms and, finally, we will illustrate the proposed iterates with numerical experiments.

The plan of the paper is the following. Section 2 concerns the case where the state is given by the solution to a linear ODE. The results are then extended in Section 3 to systems governed by nonlinear ODEs. Finally, PDE control system (with distributed and locally supported in space controls) will be the objetive of Section 4.

4.2. A minimal time problem asssociated to a linear ODE

In this section we want to illustrate the study of minimal time control problems. To this purpose, we will begin by considering a very simple situation: the problem associated to a linear ODE.

So, let us assume that the state is given by

$$\begin{cases} y_t + ay = h(t), \quad t \in (0,T), \\ y(0) = y_0, \end{cases} \tag{4.1}$$

where y_0 and $a > 0$ are given. The function $h = h(t)$ can be viewed as a control.

Our goal is to drive the state to a desired y_d in the shortest possible time with a minimal effort (in an appropiate sense). In other words, we want to find a solution to the following problem:

$$\begin{cases} \text{Minimize } \phi(T,h) := \dfrac{T^2}{2} + \dfrac{b}{2}\displaystyle\int_0^{+\infty} |h|^2 \, dt, \\ \text{Subject to: } (T,h) \in \mathbb{R}_+ \times L^2(0,+\infty), \\ \qquad\qquad (y,h) \text{ solves (4.1)}, \\ \qquad\qquad |y(T) - y_d| = \delta, \end{cases} \tag{4.2}$$

where $b \geq 0$, $\delta > 0$ and $y_d \in \mathbb{R}$.

Obviously, if (T,h) is a minimizer, one has $h \equiv 0$ for $t > T$. Furthermore, in order to discard trivial situations, we will assume that $|y_0 - y_d| > \delta$. Note that this implies that any solution to (4.2) satisfies $T > 0$.

At this point, it makes sense to investigate if (4.2) has a solution, if this solution is unique, how can we get a characterization and, finally, which solution methods can be applied.

4.2.1. Existence and characterization

In order to study the existence of the solution of (4.2) we will define the admissible solution set:

$$\mathcal{H}_{ad} := \{(T,h) : \ T \geq 0, \ h \in L^2(0,+\infty), \ |y_h(T) - y_d| = \delta\}$$

where $y_h(t)$ is the solution to (4.1) associated to h.

Note that $\mathcal{H}_{ad} \subset \mathbb{R} \times L^2(0,+\infty)$ and the solution y_h is explicitly given by

$$y_h(t) := e^{-at} y_0 + \int_0^t e^{-a(t-s)} h(s) \, ds \quad \forall t \in [0,T].$$

Theorem 4.2.1. *Let $a, \delta > 0$, $b > 0$ and $y_0, y_d \in \mathbb{R}$ with $|y_0.y_d| > \delta$. Then, there exists at least one solution $(T,h) \in \mathcal{H}_{ad}$ to* (4.2).

Proof:

To prove the existence of a solution to (4.2), we first check that $\mathcal{H}_{ad}$ is nonempty and sequentially weakly closed. Note that there exist functions $h_0 \in L^2(0,+\infty)$ with $h_0 \equiv 0$ for $t \geq 1$ such that

$$\int_0^1 e^{as} h_0(s) \, ds = 1.$$

Let us take $h = (e^a(y_d - \delta) - y_d)h_0$. Then, $|y_h(1) - y_d| = \delta$ and, consequently, $(1, h) \in \mathcal{H}_{ad}$. This proves that $\mathcal{H}_{ad}$ is not the empty set.

On the other hand, if $(T_n, h_n) \in \mathcal{H}_{ad}$ for all n, $T_n \to T$ and $h_n \to h$ weakly in $L^2(0, +\infty)$, it is clear that $T \geq 0$ and $y_{h_n}(T_n) \to y_h(T)$, whence $(T, h) \in \mathcal{H}_{ad}$. Therefore, $\mathcal{H}_{ad}$ is sequentially weakly closed.

Since $\phi : \mathcal{H}_{ad} \mapsto \mathbb{R}$ is strictly convex, continuous and coercive, (4.2) is solvable and the proof is done. □

In the next result, we characterize the solutions to (4.2):

Theorem 4.2.2. *Let $(T, h) \in \mathcal{H}_{ad}$ be a solution to* (4.2)*. Then, there exists $\lambda > 0$ and $\psi = \psi(t)$ such the following optimality system is satisfied:*

$$\begin{cases} y_t + ay = h, \quad t \in (0, T), \quad y(0) = y_0, \\ -\psi_t + a\psi = 0, \quad t \in (0, T), \quad \psi(T) = y(T) - y_d, \\ |y(T) - y_d| = \delta, \\ h = -\dfrac{1}{\lambda b}\psi \quad \textit{in } (0, T), \\ T = -\dfrac{1}{\lambda}(y(T) - y_d)y_t(T). \end{cases} \tag{4.3}$$

We will give a proof of this result that relies on the Dubovitsky-Milyoutin formalism (see [9]). This will be useful to understand the argument not only in the case of (4.1), but also for other more complex state systems. We will need the following well known result:

Lemma 4.2.3. *Let $K_1, \ldots, K_n$ be convex cones in a Banach space X with apex at 0. For each i, we assume that either K_i is open or it is a closed subspace. Then the following conditions are equivalent:*

- $\bigcap_{i=1}^{n} K_i = \emptyset$.
- *There exist linear functionals $f_i \in K_i^*$ with $i = 1, \ldots, n$, not all zero, such that $f_1 + f_2 + \cdots + f_n = 0$.*

Here, for any i, we have denoted by K_i^* the dual cone to K_i, that is,

$$K_i^* := \{f \in X' : f(e) \geq 0 \ \ \forall e \in K_i\}.$$

For the proof, see for instance Lemma 5.11 of [9].

Proof of Theorem 4.2.2:

Let (T, h) be a solution to (4.2) with $T > 0$. Let us introduce a cone in $\mathbb{R} \times L^2(0, +\infty)$ associated to (4.2):

$$D_\phi(T, h) := \{(S, g) \in \mathbb{R} \times L^2(0, +\infty) : \langle \phi'(T, h), (S, g) \rangle < 0.\}$$

This is the descent (open) cone of ϕ at (T,h). Its dual cone is

$$D^* = \{-\lambda\phi'(T,h) : \ \lambda \geq 0\}.$$

Note that we can write $\mathcal{H}_{ad} = \mathcal{G} \cap \mathcal{K}$, with

$$\mathcal{G} := \{(S,g) : S \geq 0, \ g \in L^2(0,+\infty)\}$$

and

$$\mathcal{K} = \Big\{(S,g) : S \in \mathbb{R}, \ \frac{1}{2}\Big(e^{-aS}y_0 + \int_0^S e^{-a(S-s)}g(s)\,ds - y_d\Big)^2 = \frac{\delta^2}{2}\Big\}.$$

Now, let us introduce the feasible (or admissible) descent cones of $\mathcal{G}$ and $\mathcal{K}$ at (T,h):

$$\mathcal{A}_{\mathcal{G}}(T,h) := \mathbb{R} \times L^2(0,+\infty) \quad \text{and} \quad \mathcal{A}_{\mathcal{K}}(T,h) := \mathrm{N}(\Lambda_{(T,h)}),$$

where $\Lambda_{(T,h)}$ is the linear operator given by $\Lambda_{(T,h)}(S,g) := (y_h(T) - y_d)(y_h)_t(T) \cdot S + (y_h(T) - y_d)z_g(T)$ and z_g is the solution to

$$\begin{cases} (z_g)_t + az_g = g & t \in (0,T), \\ z_g(0) = 0. \end{cases}$$

The corresponding dual cones are

$$\mathcal{A}_{\mathcal{G}}(T,h)^* = \{(\mu,\varphi) : \mu = 0, \ \varphi(t) = 0 \ \ \forall t \in (0,T)\} \quad \text{and} \quad \mathcal{A}_{\mathcal{K}}(T,h)^* = \mathrm{R}(\Lambda^*_{(T,h)}),$$

since $\Lambda^*_{(T,h)}$ is a closed-rank operator. Consequently, we can apply the Dubovitsky-Milyoutin formalism and more precisely Lemma 4.2.3 and deduce that there exist $\lambda \geq 0$ and $z \in \mathbb{R}$, not both zero, such that

$$-\lambda\phi'(T,h) - \big((y_h(T) - y_d)(y_h)_t(T), \psi\big)z = (0,0). \tag{4.4}$$

It is easy to check that we must have $\lambda > 0$ and $z \neq 0$. This yields:

$$T = -\frac{z}{\lambda}(y_h(T) - y_d)(y_h)_t(T), \quad h = -\frac{z}{\lambda b}\psi.$$

Note that necessarily $z > 0$; indeed, if $z < 0$, then $(y_h(T) - y_d)(y_h)_t(T) > 0$ and this means that, for some $T' < T$ one has $|y_h(T') - y_d| < \delta$, which is an absurd.

This ends the proof. □

We are going to prove now that the optimality system (4.3) posseses exactly one solution:

Theorem 4.2.4. *For any $y_0, y_d \in \mathbb{R}$ and any $a > 0$, the system* (4.3) *is uniquely solvable.*

Proof:

For simplicity, we will suppose that $y_0 = 0$ (a similar argument works with any other $y_0 \in \mathbb{R}$). Recall that

$$y(t) = e^{-at}\int_0^t e^{as}h(s)\,ds \quad \text{and} \quad \psi(t) = e^{-a(T-t)}\Big(\int_0^T e^{-a(T-s)}h(s)\,ds - y_d\Big) \quad \forall t \in [0,T].$$

We can write that $h = Ae^{at}$ with $A := -\frac{1}{\lambda b}\Big(e^{-2aT}\int_0^T e^{as}h(s)\,ds - y_d e^{-aT}\Big)$. But, since $h = -\frac{1}{\lambda b}\psi$, we also have that

$$\Big(1 + \frac{1}{2ab\lambda}(1 - e^{-2aT}\Big)A = \frac{1}{\lambda b}y_d e^{-aT}. \tag{4.5}$$

Since $\lambda > 0$, one has $2ab\lambda + 1 - e^{-2aT} \neq 0$ and there exists a unique solution to (4.5). After some computations, we find that

$$A(\lambda, T) = \frac{a}{\sinh(aT)}(y_d \pm \delta)$$

and, if we return to the equation for h, we see that

$$\lambda(y_d \mp \delta e^{-aT}) = \pm\frac{\delta e^{-aT}}{ab\sinh(aT)}.$$

This argument shows that, for every $T > 0$, there exists a unique $\lambda > 0$ such that all the equalities in (4.3) except the last one are satisfied.

Finally, using that $T = -\frac{1}{\lambda}(y(T) - y_d)y_t(T)$ and recalling the explicit expression of y, we get that

$$T = ba^2(y_d - \delta)^2\frac{e^{aT}(e^{aT} + e^{-aT})}{b(e^{aT} - e^{-aT})^2} \tag{4.6}$$

and this has exactly one solution $T > 0$. □

As a consequence of the previous results, it can be affirmed that there exists a unique solution to (4.2) satisfying (4.3) for some λ and ψ.

4.2.2. Some algorithms

In this section, we will present some algorithms which can be used to compute the solution to (4.3). This will help to clarify the ideas and understand better other more complex problems in the following sections.

ALG 1: First Fixed-Point Method

(a) Choose (h^0, λ^0, T^0) with $(T^0, h^0) \in \mathcal{H}_{ad}$ and $\lambda^0 > 0$.

(b) Then for $n \geq 0$ with (h^n, λ^n, T^n) given, we compute y^{n+1} with

$$y_t^{n+1} + ay^{n+1} = h^n, \quad t \in (0, T^n), \quad y^{n+1}(0) = y_0$$

and ψ^{n+1} with

$$-\psi_t^{n+1} + a\psi^{n+1} = 0, \quad t \in (0, T^n), \quad \psi^{n+1}(T^n) = y^{n+1}(T^n) - y_d.$$

(c) Finally, we update the values of h, λ and T:

$$\lambda^{n+1} \text{ such that } |y^{n+1}(T^n) - y_d| = \delta,$$

$$h^{n+1} = -\frac{1}{\lambda^{n+1} b}\psi^{n+1},$$

$$T^{n+1} = \frac{\delta}{\lambda^{n+1}} y_t^{n+1}(T^{n+1}).$$

ALG 2: Second Fixed-Point Method

(a) Choose $(h^0, \lambda^0, T^{`0})$ with $(T^0, h^0) \in \mathcal{H}_{ad}$ and $\lambda^0 > 0$.

(b) Then, for $n \geq 0$ with (h^n, λ^n, T^n) given, we compute y^{n+1} with

$$y_t^{n+1} + a y^{n+1} = h^n \quad t \in (0, T^n), \quad y^{n+1}(0) = y_0$$

and ψ^{n+1} with

$$-\psi_t^{n+1} + a\psi^{n+1} = 0 \quad t \in (0, T^n), \quad \psi^{n+1}(T^n) = y^{n+1}(T^n) - y_d.$$

(c) Finally, we update the values of h, λ and T:

$$T^{n+1} \text{ such that } T^{n+1} = \frac{\delta}{\lambda^n} y_t^n(T^{n+1}),$$

$$\lambda^{n+1} \text{ such that } |y^{n+1}(T^{n+1}) - y_d| = \delta,$$

$$h^{n+1} = -\frac{1}{\lambda^{n+1} b}\psi^{n+1}.$$

ALG 3: Third Fixed-Point Method

(a) Choose (y^0, ψ^0) with $y^0(0) = y_0$ and $\psi^0(T) = y^0(T) - y_d$.

(b) Then, for $n \geq 0$ with (y^n, ψ^n) given, we compute the solution (λ^{n+1}, T^{n+1}) to the equation

$$K^n(\lambda, T) = (0, 0), \tag{4.7}$$

where $K^n(\lambda, T) := (y[\lambda, T^n](T) - y_d \pm \delta, T - \frac{\delta}{\lambda} y_t[\lambda, T^n](T))$ and $y[\lambda, T^n]$ is the state associated to $h = -\frac{1}{\lambda b}\psi^n$.

(c) Finally, we compute the new control

$$h^{n+1} = -\frac{1}{\lambda^{n+1} b}\psi^{n+1}.$$

Note that, in order to solve (4.7), we can use, for example, a Newton algorithm (see for instance [4]).

For the problem considered in this section, since the state can be made explicit, we also have a direct (non-iterative) solution method. It is the following:

ALG 4: Direct Method

(a) First, we compute the optimal time T^* solving the equation

$$T = ba^2(y_d - \delta)^2 \frac{e^{aT}(e^{aT} + e^{-aT})}{b(e^{aT} - e^{-aT})^2}.$$

(b) Then, we compute λ^* by imposing

$$\|y(T^*) - y_d\| = \delta,$$

where y is the state associated to $h = Ae^{at}$ and A is given by (4.5) for $T = T^*$. Finally, we use T^* and λ^* to compute the assiciated state-control pair (y^*, h^*).

4.2.3. A numerical experiment

In order to illustrate the behavior of ALG 4, we will exhibit in this section the results of some numerical experiments.

The computations have been performed with MatLab.

The following data have been fixed in (4.3):

$$\delta = 10^{-5}, \quad y_d = 10, \quad b = 0.5 \quad \text{and} \quad a = 0.5.$$

Some results are depicted in Figures 4.1 to 4.3. More precisely, the left and right hand sides in (4.5) are displayed in Figure 4.1. The computation of λ^* is illustrated in Figure 4.2, where the values of $y(T^*) - y_d - \delta$ are shown as a function of $1/\lambda$. Finally, the state and control given by T^* and λ^* are presented in Figure 4.3.

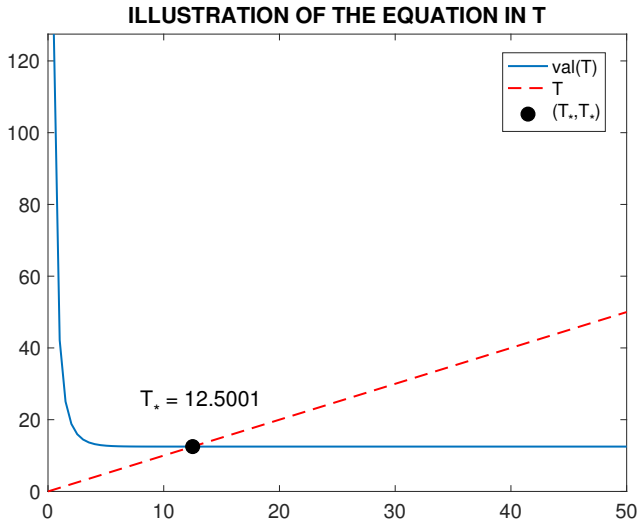


Figure 4.1: Minimal time T_* obtained with ALG 4. By definition, $val(T)$ is the right hand side in (4.3).

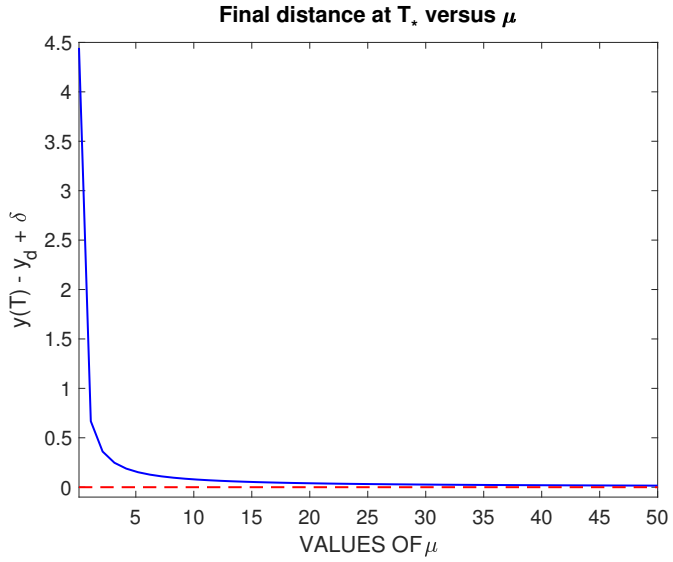


Figure 4.2: Value to $y(T^*) - y_d - \delta$ versus $1/\lambda$.

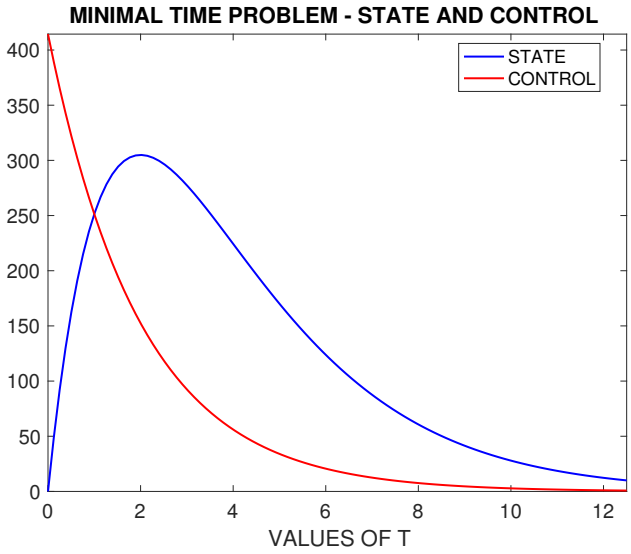


Figure 4.3: State and control corresponding to the computed solution to (4.3).

4.3. A minimal time problem asssociated to a nonlinear ODE

This section is devoted to analyze a little more difficult case, where the state is the solution to a nonlinear ODE. Now, in general, we do not have explicit expressions of the solutions. However, we can get results similar to those in the previous section

Let us consider the following system:

$$\begin{cases} y_t + H(y) = h(t), \quad t \in (0,T), \\ y(0) = y_0, \end{cases} \tag{4.8}$$

where y_0 is given. For simplicity, we will assume that

$$\begin{cases} H : \mathbb{R} \mapsto \mathbb{R} \text{ is of class } \mathcal{C}^1, \\ 0 \leq H'(s) \leq C \quad \forall s \in \mathbb{R}. \end{cases} \tag{4.9}$$

It is obvious that, for each $h \in L^2(0,+\infty)$, there exists exactly one solution y to (4.8) defined for all $t \in (0,T)$. As in Section 4.2, we can consider the minimal problem associated to (4.8):

$$\begin{cases} \text{Minimize } \phi(T,h) := \dfrac{T^2}{2} + \dfrac{b}{2}\displaystyle\int_0^{+\infty} |h|^2 \, dt, \\ \text{Subject to: } (T,h) \in \mathbb{R}_+ \times L^2(0,+\infty), \\ \qquad\qquad (y,h) \text{ solves (4.8)}, \\ \qquad\qquad |y(T) - y_d| = \delta, \end{cases} \tag{4.10}$$

where again we have $b \geq 0$, $\delta > 0$ and $y_d \in \mathbb{R}$ is the desired state. As in Section 1, it will be assumed that $|y_0 - y_d| > \delta$ in order to avoid trivial situations.

4.3.1. Existence and characterization results

Let us introduce again the admissible solution set

$$\mathcal{H}_{ad} := \{(T,h):\ T \geq 0,\ h \in L^2(0,+\infty),\ |y_h(T) - y_d| = \delta\},$$

where $y_h(t)$ is now the solution to (4.8) associated to h.

Theorem 4.3.1. *Let $a, \delta > 0$, $b > 0$ and $y_0, y_d \in \mathbb{R}$ be given with $|y_0 - y_d| > \delta$. Then, there exists at least one solution $(T,h) \in \mathcal{H}_{ad}$ to* (4.10).

Proof:

As before, it can be easily proved that $\mathcal{H}_{ad}$ is non empty. In fact, for any $T > 0$ and any $y_0, y_T \in \mathbb{R}$, there exist couples $(\overline{h}, \overline{y})$ satisfying (4.8) and $|y(T) - y_d| = \delta$. Indeed, it suffices to take, for instance,

$$\overline{y}(t) = \Big(1 - \frac{t}{T}\Big)(y_0 - y_T) + y_d - \delta \ \text{ and } \ \overline{h}(t) = \overline{y}_t(t) + H(\overline{y}(t)) \ \ \forall t \in (0,T).$$

On the other hand, $\mathcal{H}_{ad}$ is sequentially weakly closed: obviously, if the $(T_n, h_n) \in \mathcal{H}_{ad}$, $T_n \to T$ and $h_n \to h$ weakly in $L^2(0,+\infty)$, then $T \geq 0$ and $|y(T) - y_d| = \lim_{n\to+\infty} |y_n(T) - y_d| = \delta$.

Taking into account that ϕ is convex, continuous and coercive, we deduce that there exists at least one solution to (4.10). □

Unfortunately, $\mathcal{H}_{ad}$ is not convex and, in principle, we do not know whether the solution to (4.10) is unique.

Let us now characterize the solutions to (4.10) with an appropiate optimality system:

Theorem 4.3.2. *Let $(T,h) \in \mathcal{H}_{ad}$ be a solution to* (4.10). *Then, there exist $\lambda > 0$ and $\psi = \psi(t)$ such that the following holds:*

$$\begin{cases} y_t + H(y) = h, \quad t \in (0,T), \\ y(0) = y_0, \\ -\psi_t + H'(y)\psi = 0, \quad t \in (0,T), \\ \psi(T) = y(T) - y_d, \\ |y(T) - y_d| = \delta, \\ h = -\dfrac{1}{\lambda b}\psi \ \ \textit{in } (0,T), \\ T = -\dfrac{1}{\lambda}(y(T) - y_d)y_t(T). \end{cases} \tag{4.11}$$

The proof is completely analogous to the proof of Theorem 4.2.2. Again, we can apply the Dubovitsky-Milyoutin formalism and deduce a property similar to (4.4). Note that, thanks to (4.9), we have that the mappings $\Lambda_{(T,h)}$ and $\Lambda^*_{(T,h)}$ continue to have the same good properties.

4.3.2. Algorithm and numerical experiments

It is possible to adapt the algorithms in Section 4.2.2 to this case. For brevity, we omit the details. We only mention explicitly the analog of ALG 4:

ALG 5: Direct Method

(a) First, for each $T > 0$, we compute the solution $\lambda(T)$ to the equation $|y(T) - y_d| = \delta$, where y is, together with ψ and h, the solution to the solution to the reduced optimality system

$$\begin{cases} y_t + H(y) = h, \quad t \in (0,T), \\ y(0) = y_0, \\ -\psi_t + H'(y)\psi = 0, \quad t \in (0,T), \\ \psi(T) = y(T) - y_d, \\ |y(T) - y_d| = \delta, \\ h = -\dfrac{1}{\lambda b}\psi \quad \text{in } (0,T). \end{cases}$$

(b) Then, we compute T^* by solving

$$T = -\frac{1}{\lambda(T)}(y(T) - y_d)y_t(T). \tag{4.12}$$

Finally, we use T^* and $\lambda(T^*)$ to compute the associated state-control pair (y^*, h^*).

In the sequel, we present a related numerical experiment. In this case, we have taken

$$\delta = 10^{-5}, \quad y_d = 10, \quad b = 0.5 \quad \text{and} \quad H(y) \equiv -y + 0.5\sin(y).$$

The results are given in Figures 4.4 to 4.6. Thus, Figure 4.4 provides a graphical explanation of the solution to the equation (4.12). In Figure 4.5, the function $T \mapsto 1/\lambda(T)$ has beem depicted. Finally, the computed optimal state and control have been shown in Figure 4.6.

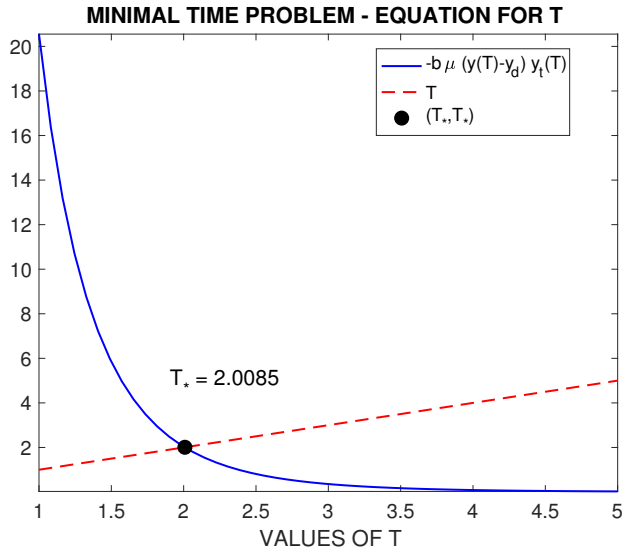


Figure 4.4: Minimal time T_* obtained with ALG 5.

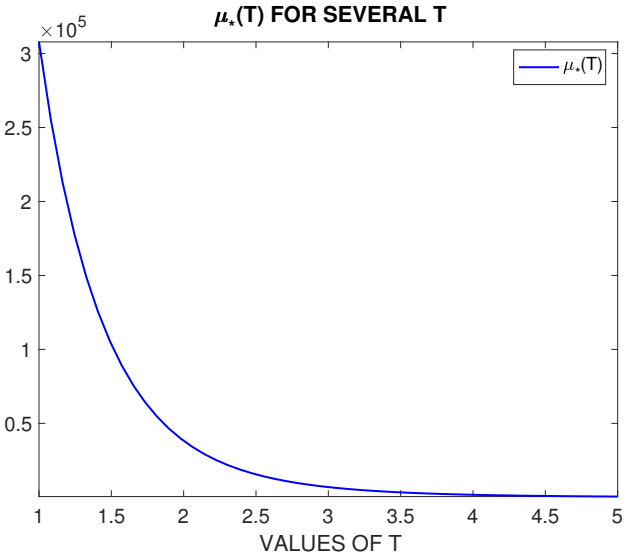


Figure 4.5: The values of $\mu(T) = 1/\lambda(T)$ versus T.

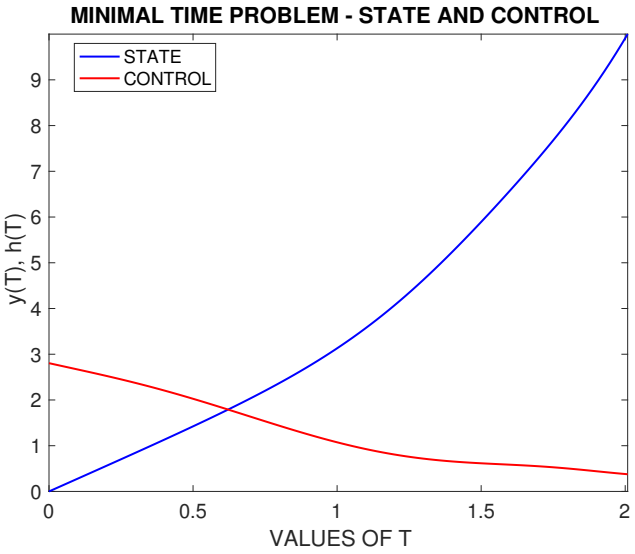


Figure 4.6: State and control corresponding to the computed solution to (4.11).

4.4. The minimal time problem associated to the heat PDE

In this section, we are going to study a minimal time problem related to the heat equation. We will see that, in this case, new difficulties appear and, in particular, the computation of the solution will need much more work.

The state equation is in this case

$$\begin{cases} \theta_t - \Delta\theta = h1_\omega, \quad (x,t) \in Q_T := \Omega \times (0,T), \\ \theta = 0, \quad (x,t) \in \Sigma_T := \partial\Omega \times (0,T), \\ \theta(0) = \theta_0, \end{cases} \tag{4.13}$$

where $\Omega \subset \mathbb{R}^N$ is a non-empty bounded connected open set ($N = 2$ or $N = 3$), $\omega \subset \Omega$ is also non-empty and open (the control subdomain) and $\theta_0 \in L^2(\Omega)$.

As before, the goal is to drive the state θ to a desired θ_d in the shortest possible time with a minimal effort. Accordingly, we want to solve the problem

$$\begin{cases} \text{Minimize } \phi(T,h) := \dfrac{T^2}{2} + \dfrac{b}{2} \displaystyle\iint_{\omega\times(0,+\infty)} |h|^2 \, dx \, dt, \\ \text{Subject to: } (T,h) \in \mathbb{R}_+ \times L^2(\omega \times (0,+\infty)), \\ \qquad\qquad (\theta,h) \text{ solves (4.13)}, \\ \qquad\qquad \|\theta(T) - \theta_d\| = \delta, \end{cases} \tag{4.14}$$

where $b > 0$, $\delta > 0$ and $\theta_d \in L^2(\Omega)$ are given. As in Section 4.2, we can make an assumption to discard trivial situations: $\|\theta_0 - \theta_d\| > \delta$. Again, the minimum is searched with $h \equiv 0$ for $t \geq T$.

Note that we use $\|\cdot\|$ and $(\cdot,\cdot)$ to denote the $L^2(\Omega)$ norm and the inner product, respectively.

4.4.1. Existence and characterization of the solution

In order to study the existence of a solution to (4.14), let us introduce the admissible set

$$\mathcal{H}_{ad} := \{(T,h) : T \geq 0,\ h \in L^2(\omega \times (0,+\infty)),\ \|\theta(T) - \theta_d\| = \delta\},$$

where $\theta = \theta(x,t)$ is the solution to (4.13) associated to the control h.

Theorem 4.4.1. *Let $a, \delta > 0$, $b > 0$ and $\theta_0, \theta_d \in L^2(\Omega)$ be given with $|y_0 - y_d| > \delta$. Then, there exists at least one solution $(T,h) \in \mathcal{H}_{ad}$ to* (4.14)*.*

Proof:

First, we note that $\mathcal{H}_{ad}$ is nonempty. Indeed, it is well known that (4.13) is approximately controllable and, consequently, for every $T > 0$ we can find controls $h \in L^2(\omega \times (0,T))$ such that the corresponding solutions to (4.13) satisfy $\|\theta(T) - \theta_d\| = \delta$, see for instance [5].

On the other hand, $\mathcal{H}_{ad}$ is sequentially weakly closed. Indeed, if the $(T_n, h_n) \in \mathcal{H}_{ad}$, $T_n \to T$ and $h_n \to h$ weakly in $L^2(0,+\infty)$, then $T \geq 0$ and, from well known parabolic regularity results, we deduce that the associated states θ_n satisfy $\theta_n(T_n) \to \theta(T)$ strongly in $L^2(\Omega)$, whence $\|\theta(T) - \theta_d\| = \delta$.

Since the function ϕ is convex, continuous and coercive, (4.14) is solvable and the proof is done. □

Again, uniqueness is open, since $\mathcal{H}_{ad}$ is not convex.

Theorem 4.4.2. *Let $(T,h) \in \mathcal{H}_{ad}$ be a solution to* (4.14)*. Then there exist $\lambda > 0$ and $\psi = \psi(x,t)$ such that the following optimality system is satisfied:*

$$\begin{cases} \theta_t - \Delta\theta = h1_\omega, \quad (x,t) \in Q_T, \\ \theta = 0, \quad (x,t) \in \Sigma_T, \quad \theta(0) = \theta_0, \\ -\psi_t - \Delta\psi = 0, \quad (x,t) \in Q_T, \\ \psi = 0, \quad (x,t) \in \Sigma, \quad \psi(T) = \theta(T) - \theta_d, \\ \|\theta(T) - \theta_d\| = \delta, \\ h = -\dfrac{1}{\lambda b}\,\psi|_{\omega\times(0,T)}\,, \\ T = -\dfrac{1}{\lambda}\Big((\theta(T) - \theta_d), \theta_t(T)\Big). \end{cases} \tag{4.15}$$

Proof:

Once more, we can apply the Dubovitsky-Milyoutin formalism. The argument is a follows. We introduce the descent cone of ϕ at (T,h)

$$D_\phi(T,h) := \{(S,g) \in \mathbb{R} \times L^2(\omega \times (0,+\infty)) : \langle \phi'(T,h), (S,g)\rangle < 0\},$$

where $\langle\cdot,\cdot\rangle$ stands for the scalar product in $\mathbb{R} \times L^2(\omega \times (0,T))$. The associated dual cone is given by

$$D^* = \{-\lambda\phi'(T,h) :\ \lambda > 0\}.$$

Again, we put $\mathcal{H}_{ad} = \mathcal{G} \cap \mathcal{K}$, where

$$\mathcal{G} := \{(T,h) :\ T \geq 0,\ h \in L^2(\omega \times (0,+\infty))\}$$

and

$$\mathcal{K} = \{(T,h) :\ T \in \mathbb{R},\ \frac{1}{2}\|\theta(T) - \theta_d\|^2 = \frac{\delta^2}{2}\}.$$

Let us introduce the feasible cones

$$\mathcal{A}_{\mathcal{G}}(T,h) := \{(S,g) :\ S \in \mathbb{R},\ g(t) = 0\ t \geq T\} \quad \text{and} \quad \mathcal{A}_{\mathcal{K}}(T,h) := \mathrm{N}(\Lambda_{(T,h)}),$$

where $\Lambda_{(T,h)}$ is given by $\Lambda_{(T,h)}(S,g) := \Big(\theta(T) - \theta_d, \theta_t(T)\Big) \cdot S + \Big(\theta(T) - \theta_d, \eta(T)\Big)$, and η is the solution to

$$\begin{cases} \eta_t - \Delta\eta = g1_\omega, \quad (x,t) \in Q_T, \\ \eta = 0, \quad (x,t) \in \Sigma_T, \\ \eta(0) = 0. \end{cases}$$

The associated dual cones are respectively

$$\mathcal{A}_{\mathcal{G}}(T,h)^* = \{(\mu,\varphi) :\ \mu = 0,\ \varphi(t) = 0\ t \in (0,T)\} \quad \text{and} \quad \mathcal{A}_{\mathcal{K}}(T,h)^* = \mathrm{R}(\Lambda^*_{(T,h)}),$$

since, again, $\Lambda^*_{(T,h)}$ is a closed-rank operator.

Then, as in the proofs of Theorems 4.2.2 and 4.3.2, we can apply Lemma 4.2.3 and deduce that

$$-\lambda(T,bh) - ((\theta(T) - \theta_d, \theta_t(T)), \psi)z = (0,0)$$

for some $\lambda > 0$ and some $z > 0$ and we obtain the optimality system (4.15). □

In view of Theorems 4.4.1 and 4.4.2, it is clear that, for any a, b, δ, θ_0 and θ_d, (4.15) is solvable. For completeness, we will present in the sequel an additional argument that may shed some light on the solution method.

Thus, let $\theta_0, \theta_d \in L^2(\Omega)$ and $a, b, \delta > 0$ be given and let us assume that $\|\theta_d\| > \delta$. We want to prove that there exists (T, λ) satisfying

$$\begin{cases} \|\theta(T) - \theta_d\| = \delta, \\ T = -\dfrac{1}{\lambda}(\theta(T) - \theta_d, \theta_t(T)), \end{cases}$$

where θ solves, together with h and ψ, the system

$$\begin{cases} \theta_t - \Delta\theta = h1_\omega, \quad (x,t) \in Q_T, \\ -\psi_t - \Delta\psi = 0, \quad (x,t) \in Q_T, \\ \theta = 0,\ \psi = 0, \quad (x,t) \in \Sigma, \\ \theta(0) = \theta_0,\ \psi(T) = \theta(T) - \theta_d, \\ h = -\dfrac{1}{\lambda b}\ \psi|_{\omega\times(0,T)}\,. \end{cases} \tag{4.16}$$

By using the semigroup theory, we can write the solution to the system

$$\begin{cases} \theta_t - \Delta\theta = 0, \quad (x,t) \in Q_T, \\ \theta(0) = 0, \quad (x,t) \in \Sigma_T, \\ \theta(0) = \theta_0 \end{cases}$$

in the form $\theta(t) = S(t)\theta_0$, where $S(t)$ is a linear contration on $L^2(\Omega)$ for all $t \geq 0$, see for instance [2].

Thus, the solution (θ, ψ) to (4.15) takes the form

$$\theta(t) = \int_0^t S(t-s)(h(s)1_\omega)\, ds, \quad \psi(t) = S(T-t)(\theta(T) - \theta_d).$$

and we can write that

$$\theta(T) = -\frac{1}{\lambda b}\int_0^T S(T-s)\Big(S(T-s)(\theta(T)-\theta_d)1_\omega\Big)\, ds.$$

Now, we fix $T > 0$, we consider the new variable $w := \theta(T)$ and we introduce the following linear operator:

$$M_T w := \frac{1}{b}\int_0^T S(T-s)\Big((S(T-s)w)1_\omega\Big)\, ds.$$

Our first goal is to find a couple (λ, w) such that

$$(M_T + \lambda)w = M_T\theta_d \quad \text{and} \quad \frac{1}{2}\|w - \theta_d\|^2 = \frac{\delta^2}{2}. \tag{4.17}$$

Note that $M_T : L^2(\Omega) \mapsto L^2(\Omega)$ is a compact operator, $(M_T w, w) \geq 0$ for all $w \in L^2(\Omega)$ and $(M_T w, w) = 0$ if only if $w = 0$. Therefore, we can apply the Hilbert-Schmidt Theorem (see for instance [2]) and deduce that there exists some eigenvalue-eigenfunction couples (μ_n, w_n) with $\mu_n > 0$ that

$$\begin{cases} M_T w_n = \mu_n w_n, \quad n \geq 1, \\ \|w_n\| = 1, \quad \mu_n \searrow 0^+, \\ \{w_n\} \text{ is an orthonormal basis of } L^2(\Omega). \end{cases}$$

Accordingly, for any $w \in L^2(\Omega)$, we can write

$$w = \sum_{n\geq 1} a_n w_n \quad \text{and} \quad M_T w = \sum_{n\geq 1} \mu_n a_n w_n,$$

with $a_n = (w, w_n)$ for all n. This means that the first equation in (4.17) is equivalent to

$$\sum_{n\geq 1}(\mu_n + \lambda)a_n w_n = \sum_{n\geq 1}\mu_n(\theta_d, w_n)w_n,$$

that is,

$$a_n = \frac{(\theta_d, w_n)}{1 + \lambda\lambda_n} \quad \text{with } \lambda_n = \frac{1}{\mu_n}\ \forall n \geq 1. \tag{4.18}$$

Thus, if w solves the first equation in (4.17), one has

$$\|w-\theta_d\|^2=\lambda^2\sum_{n\geq 1}\frac{\lambda_n^2}{(1+\lambda\lambda_n)^2}|(\theta_d,w_n)|^2\leq\|\theta_d\|^2.$$

For each $\lambda>0$, let us denote by $w[\lambda]$ the function in $L^2(\Omega)$ defined by (4.18) and let us set

$$G(\lambda):=\|w[\lambda]-\theta_d\|^2=\sum_{n\geq 1}b_n^2\Big(1-\frac{1}{1+\lambda\lambda_n}\Big)^2,$$

where $b_n=|(\theta_d,w_n)|$ for all $n\geq 1$. It is not difficult to prove that $G:\mathbb{R}_+\mapsto\mathbb{R}_+$ is well defined, G is analytic in $(0,+\infty)$, $G'(\lambda)>0$ for all $\lambda>0$, $G(\lambda)\searrow 0^+$ and $G(\lambda)\nearrow\|\theta_d\|^2$ as $\lambda\nearrow+\infty$.

Consequently, there exists a unique λ such that $G(\lambda)=\delta^2$ (recall that $0<\delta^2<\|\theta_d\|^2$).

This argument proves that, for each $T>0$, there exists exactly one $\lambda(T)$ such that the couple $(\lambda(T),w[\lambda(T)])$ satisfies (4.17).

Hence, our task will be achieved if we are able to find $T>0$ such that

$$T=-\frac{1}{\lambda(T)}(\theta^T(T)-\theta_d,\theta_t^T(T)),\tag{4.19}$$

where we have introduced the notation $\theta^T:=M_Tw[\lambda(T)]$.

Note that, in the similar (but much simpler) situation considered in Section 4.2, the analog of (4.19) is (4.6). There, we saw that the right hand side increases to $+\infty$ (respectively decreases to 0) as $T\searrow 0^+$ (resp. as $T\nearrow+\infty$). It is reasonable to conjecture that these properties also hold for the right hand side of (4.19). But, to our knowledge, this is unknown. Note that, if this were the case, the existence of a solution to (4.19) would be ensured and, consequently, we would have got a new proof of the solvability of (4.14).

4.4.2. Some algorithms

This section is devoted to present and analyze some iterative algorithms for the solution to (4.13).

First, observe that the algorithms in Section 4.2.2 can be adapted to the present framework. For brevity, we omit the details and only mention explicitly the analog of ALG 4. For comodity, we will work with the auxiliar parameter $\mu=1/\lambda$.

ALG 6:

(a) Fix $T_0,T_1,\mu_0,\mu_1\Delta t$ and $\Delta\mu$.

(b) Then, for each $T^k=T_0+k\Delta t$ with $k\geq 0$ and $T^k\leq T_1$ and each $\mu^j=\mu_0+j\Delta\mu$with $j\geq 0$ and $\mu^j\leq\mu_1$, compute:

$$\|\theta^{k,j}(T^k)-\theta_d\|-\delta\quad\text{ and }\quad-\mu^j(\theta^{k,j}(T)-\theta_d,\theta_t^{k,j}(T^k)),$$

where $\theta^{k,j}$ is the solution to (4.16) with $T=T^k$ and $\lambda=1/\mu^j$.

(c) Finally, we use this information to couple a solution (T_*, μ_*) to the system

$$T = -\mu(\theta(T) - \theta_d, \theta_t(T)), \quad \|\theta(T) - \theta_d\| = \delta \tag{4.20}$$

and then find the associated state-control pair (y_*, h_*) to (4.15).

ALG 7: Penalty

(a) Choose $\mu^0 > 0$, (T^0, h^0) with $T^0 > 0$ and $h^0 \in L^2(\omega \times (0, T^0))$.

(b) Then for $n \geq 0$ with μ^n, (T^n, h^n) given, we minimize in $\mathbb{R} \times L^2(\omega \times (0, +\infty))$ the functional:

$$\tilde{\phi}(T, h; \mu^n) := \frac{T^2}{2} + \frac{b}{2} \iint_{\omega \times (0,+\infty)} |h|^2 + \frac{1}{2\mu^n}\Big((\|\overline{\theta}(T) - \theta_d\| - \delta)^2 + T_-^2\Big). \tag{4.21}$$

To compute the minimum we can use different methods like the Optimal Step Gradient or Conjugate Gradient Method. We stop when the Gradient is less or equal to τ^n for some given $\tau^n > 0$.

(c) Now, we update the values of μ^n and τ^n. For instance, we can take $\mu^{n+1} = a\mu^n$ and $\tau^{n+1} = a\tau^n$ with $a < 1$.

Let us present some convergence results for ALG 7:

Theorem 4.4.3. *Let (T_k, h_k) be an exact global minimizer of $(T, h) \mapsto \tilde{\phi}(T, h; \mu^k)$ for each k and assume that $\mu_k \to 0$. Then, any weak limit point (T_*, h_*) of the sequence $\{(T_k, h_k)\}$ is a solution to* (4.14).

Proof:

First, remark that for each $k \geq 0$ there exists at least one minimizer (T_k, h_k) of $\tilde{\phi}(T, h; \mu_k)$. This can be checked arguing as in the proof of Theorem 4.4.1.

Let (T, h) the solution to (4.14) furnished by Theorem 4.4.1. Then $T > 0$. For each $k \geq 1$, one has

$$\tilde{\phi}(T_k, h_k; \mu_k) \leq \tilde{\phi}(T', h'; \mu_k) \quad \forall (T', h') \in \mathbb{R}_+ \times L^2(\omega \times (0, +\infty)).$$

In particular,

$$\phi(T_k, h_k) + \frac{1}{2\mu_k}\Big((\|\overline{\theta}_k(T_k) - \theta_d\| - \delta)^2 + (T_k)_-^2\Big) \leq \tilde{\phi}(T, h; \mu_k) = \phi(T, h), \tag{4.22}$$

where θ_k denotes the state associated to (T_k, h_k) and $\overline{\theta}_k(T_k)$ is defined accordingly. Therefore, the (T_k, h_k) are uniformly bounded in $\mathbb{R} \times L^2(\omega \times (0, +\infty))$ and

$$(\|\theta_k(T_k) - \theta_d\| - \delta)^2 + (T_k)_-^2 \leq 2\mu_k \phi(T, h). \tag{4.23}$$

A first consequence is that there exist weak limit points of the sequence $\{(T_k, h_k)\}$, that is, couples (T_*, h_*) such that, at least for a subsequence, one has

$$T_k \to T_* \quad \text{and} \quad h_k \to h_* \text{ weakly in } L^2(\omega \times (0, +\infty)).$$

From (4.23), we see that $T_* \geq 0$. Furthermore, if we had $T_* = 0$, we would also have $\overline{\theta}_k(T_k) \to \theta_0$ at least weakly in $L^2(\Omega)$ and from (4.23) we would get a contradiction. Therefore, $T_* > 0$.

Now, we can take limits in (4.23) as $k \to +\infty$ and get

$$(\|\theta_*(T_*) - \theta_d\| - \delta)^2 = 0,$$

where θ_* is the state associated to (T_*, h_*). Indeed, recall that the usual parabolic estimates imply that $\theta_k(T_k) \to \theta_*(T_*)$ strongly in $L^2(\Omega)$.

This way, we see that $\|\theta_*(T_*) - \theta_d\| = \delta$. Finally, taking limits in (4.22), we find that

$$\phi(T_*, h_*) \leq \phi(T, h),$$

whence the announced result holds. □

Theorem 4.4.4. *Let the tolerance and penalty parameters satisfy $\tau_k \to 0$ and $\mu_k \to 0$ and let us assume that the $(T_k, h_k) \in \mathbb{R} \times L^2(\omega \times (0, +\infty))$ verify*

$$\left\| \nabla_{(T,h)}\phi(T_k, h_k) + \frac{1}{2\mu_k}\nabla_{(T,h)}\big((\|\overline{\theta}_k(T_k) - \theta_d\| - \delta)^2 + (T_k)_-^2\big) \right\| \leq \tau_k$$

for all $k \geq 1$. Then, any associated weak limit point (T_*, h_*) with $T_* > 0$ solves (4.14) *and satisfies the usual Karush-Kuhn-Tucker (KKT) conditions for some $\lambda_*^1 \in \mathbb{R}$ and $\lambda_*^2 = 0$. Also,*

$$-(\delta - \|\theta^{k+1}(T^k) - \theta_d\|)/\mu^k \to \lambda_*^1 \quad \textit{and} \quad -(T^k)_-/\mu^k \to \lambda_*^2 = 0.$$

Proof:

Again, let (T, h) be a solution to (4.14). Then

$$\phi(T_k, h_k) + \frac{1}{2\mu_k}\Big((\|\theta_k(T_k) - \theta_d\| - \delta)^2 + (T_k)_-^2\Big) \leq \phi(T, h) + \tau_k\|(T_k, h_k) - (T, h)\|_{\mathbb{R}\times L^2(\omega\times(0,+\infty))}. \tag{4.24}$$

So, the (T_k, h_k) are uniformly bounded and there exists at least one weak limit point (T_*, h_*), with $\|\theta_*(T_*) - \theta_d\| = \delta$ and $T_* > 0$ (this can be deduced as in the proof of Theorem 4.4.3).

Note that (T_*, h_*) is a solution to (4.14). Then, we can also assume that $T_k > 0$ for all k. Let us set

$$\lambda_k^1 = -(\delta - \|\theta^{k+1}(T^k) - \theta_d\|)/\mu_k \quad \text{and} \quad \lambda_k^2 = 0$$

for all k. Taking into account (4.24), we can suppose that $\lambda_k^1 \to \lambda_*^1$. for some $\lambda_*^1 \in \mathbb{R}$.

Let us set

$$G_k := \nabla_{(T,h)}\phi(T_k, h_k) - \lambda_k^1\nabla_{(T,h)}(\delta - \|\theta_{k+1}(T_k) - \theta_d\|) - \lambda_k^2\nabla_{(T,h)}(T_k)_-.$$

By hipothesis, $G_k \to 0$ in $\mathbb{R} \times L^2(\omega \times (0, +\infty))$. But we also have $G_k \to G_*$ with $G_* = \nabla_{(T,h)}\phi(T_*, h_*) - \lambda_*^1\nabla_{(T,h)}(\delta - \|\theta_*(T_*) - \theta_d\|) - \lambda_*^2\nabla_{(T,h)}(T_k)_-$. Therefore, $G_* = 0$ and (T_*, h_*) is a solution to (4.14) that, together with λ_*^1 and λ_*^2, satisfies the KKT relations. □

The next algorithm is based on Augmented Lagrangian techniques, see for instance [12]. We start from the problem:

$$\begin{cases} \text{Minimize } \phi(T, h) - \lambda_1(\delta - \|\theta(T) - \theta_d\|) - \lambda_2(T - s) + \dfrac{1}{2\mu}\Big((\delta - \|\overline{\theta}(T) - \theta_d\|)^2 + (T - s)^2\Big) \\ \text{Subject to } (T, h) \in \mathbb{R} \times L^2(\omega \times (0, +\infty)), \;\; s \in \mathbb{R}_+, \end{cases}$$

ç

or, equivalently,

$$\begin{cases} \text{Minimize } \widehat{\phi}(T,h;\lambda_1,\lambda_2,\mu) \\ \text{Subject to } (T,h)\in\mathbb{R}\times L^2(\omega\times(0,+\infty)), \end{cases}$$

where we have introduced

$$\widehat{\phi}(T,h;\lambda_1,\lambda_2,\mu) := \phi(T,h) - \lambda_1(\delta - \|\overline{\theta}(T)-\theta_d\|) + \varphi(T;\lambda_2,\mu) + \frac{1}{2\mu}(\delta - \|\theta(T)-\theta_d\|)^2$$

and

$$\varphi(T;\lambda_2,\mu) := \begin{cases} -\lambda_2 T + \dfrac{1}{2\mu}T^2 & \text{if } T<\mu\lambda_2, \\ -\dfrac{\mu}{2}\lambda_2^2 & \text{otherwise.} \end{cases}$$

ALG 8: Augmented Lagragian

(a) Choose $\mu^0,\lambda_1^0,\lambda_2^0>0$, (T^0,h^0) with $T^0>0$ and $h^0\in L^2(\omega\times(0,T^0))$.

(b) Then for $n\geq 0$ with given $\mu^n,\lambda_1^n,\lambda_2^n$, (T^n,h^n), we solve the unconstrained problem

$$\begin{cases} \text{Minimize } \widehat{\phi}(T,h;\lambda_1^n,\lambda_2^n;\mu^n) \\ \text{Subject to } (T,h)\in\mathbb{R}\times L^2(\omega\times(0,+\infty)). \end{cases} \tag{4.25}$$

Again, we can use here various techniques, such as for instance Optimal Step Gradient or Conjugate Gradient Methods.

(c) Then, we update λ_1^n and λ_2^n:

$$\lambda_1^{n+1} = \lambda_1^n + (\|\overline{\theta}^{n+1}(T^{n+1})-\theta_d\| - \delta)/\mu^n,$$

$$\lambda_2^{n+1} = [\lambda_2^n - T^{n+1}/\mu^n]_+,$$

where θ^{n+1} is the state associated to the previously computed solution (T^{n+1},h^{n+1}) to (4.25).

(d) Finally, we update μ^n by taking $\mu^{n+1}=a\mu^n$, with $a<1$.

Arguing as in [1], it is possible to prove a convergence result for ALG 8 to a solution to (4.14), provided some appropiate conditions are satisfied. This will be analyzed in detail in a forthcoming paper.

4.4.3. Numerical experiments

This section is devoted to present the result of some numerical experiments for the solution of (4.13). We will employ ALG 6. The computations have benn performed with the FreeFem++ package (see [10]).

We have tested ALG 6 for (4.13) with

$$\delta = 0.07, \quad b = 100, \quad T_0 = 5, \quad T_1 = 20, \quad \Delta t = 1.25, \quad \mu_0 = 120, \quad \mu_1 = 160, \quad \Delta\mu = 10.$$

The desired state is the solution at time $T = 20$ of the heat equation with initial data $\theta_{0d} = 5$ and control $h \equiv 0$. Our domain is an open disk Ω inside which we find the rectangle ω (the control domain), as can be seen in Figure 4.7.

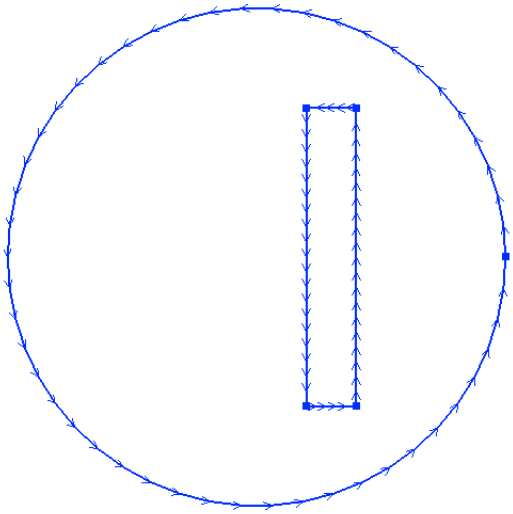


Figure 4.7: The spatial domain.

The numerical solutions of the state and adjoint state systems have been carried out with standard finite elements in space and finite differences in time. The mesh is displayed in Figure 4.8. For each T, the $\mu(T)$ that we can find solving 4.16 together with the identity $\|\theta(T) - \theta_d\| = \delta$ is depicted in Figure 4.9. Then, the functions in the left and the right of the first equality in (4.20) are displayed in Figure 4.10, together with the solution T_*. The results appear in Figure 4.10 to Figure 4.12.

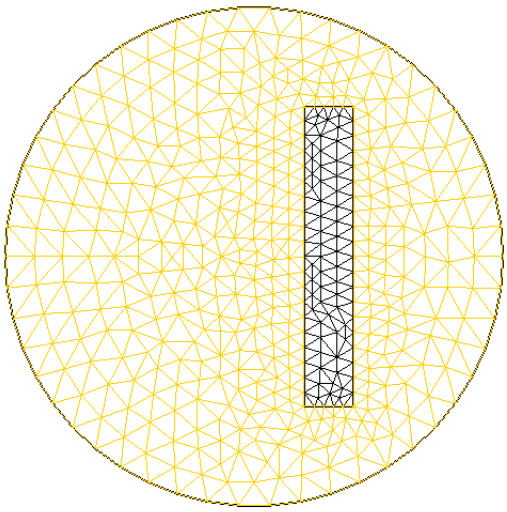


Figure 4.8: The mesh.

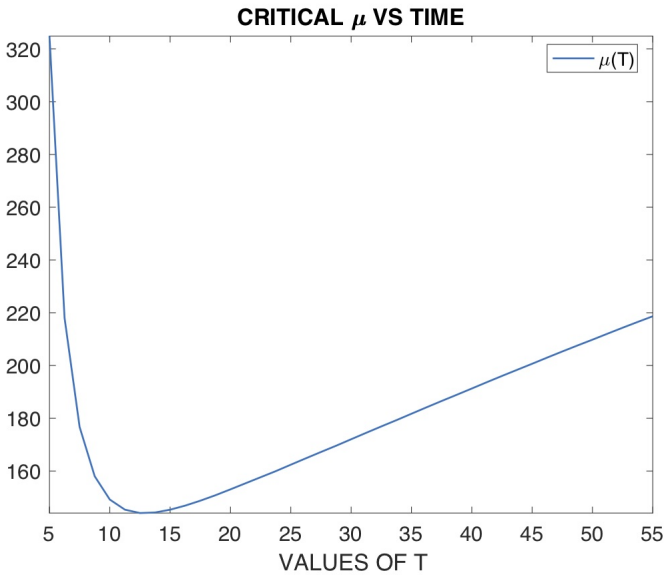


Figure 4.9: The values of μ that solve the equation $\|\theta(T) - \theta_d\| = \delta$.

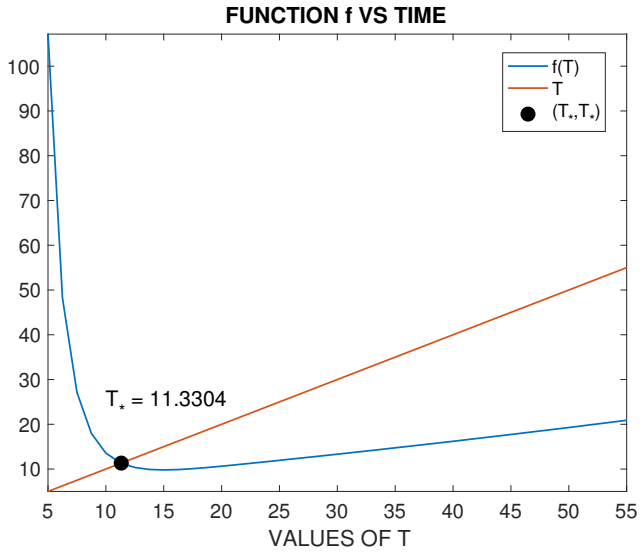


Figure 4.10: Minimal time T_* obtained with ALG 6 applied to (4.15).

Finally, we present in Figures 4.9 and 4.11 the values of $\|\theta(T)-\theta_d\|-\delta$ and $\frac{1}{\mu}(\theta(T)-\theta_d, \theta_t(T))$ associated to various T and μ.

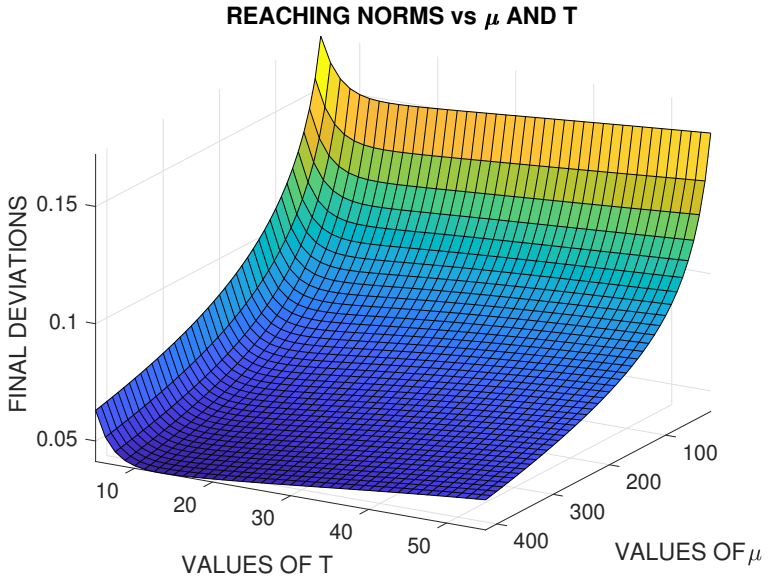


Figure 4.11: Values to the norm $\|\theta(T) - \theta_d\| - \delta$ versus μ and T.

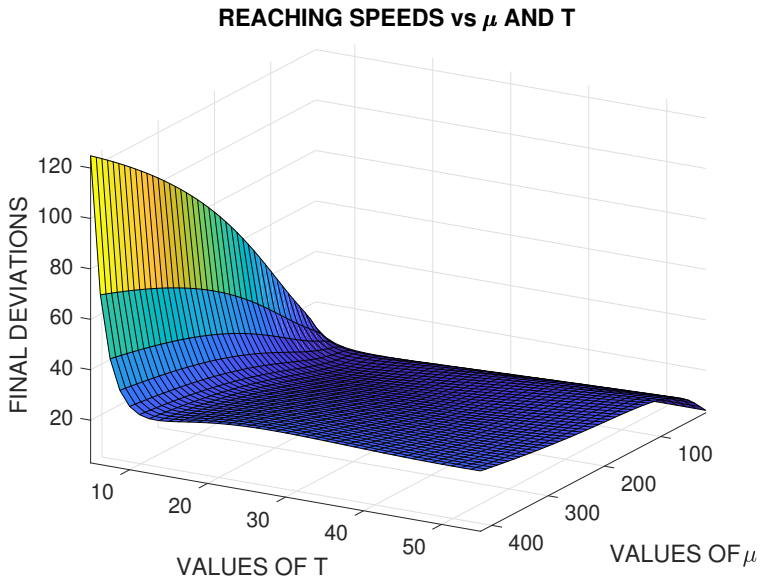


Figure 4.12: Values to $-\mu(\theta(T) - \theta_d)\theta_t(T)$ versus μ and T.

Bibliography

[1] D. P. Bertsekas, *"Constrained Optimization and Lagrange Multiplier Methods"*, Athena Scientific, Belmont, Massachusetts Institute of Technology,1996.

[2] H. Brézis, *"Functional Analysis, Sobolev Sapces and Partial Differential Equations",* Springer, London, 2011.

[3] R. C. Cabrales, G. Camacho, E. Fernández-Cara, *Analysis and optimal control of some solidification processes*, Discrete Contin. Dyn. Syst., 34, 10, 2014, 3985-4017.

[4] S. D. Conte, *" Elementary numerical analysis: An algorithmic approach"*, McGraw-Hill Book Co., New York-Toronto, Ont.-London, 1965.

[5] J. M. Coron, *" Control and nonlinearity. Mathematical Surveys and Monographs"*, 136. American Mathematical Society, Providence, RI, 2007.

[6] H. O. Fattorini, *Second order linear differential equations in Banach spaces*, North-Holland Mathematics Studies, 108. Notas de Matemática [Mathematical Notes], 99. North-Holland Publishing Co., Amsterdam, 1985.

[7] E. Fernández-Cara, *Motivation, analysis and control of the variable density Navier-Stokes equations* Discrete Contin. Dyn. Syst. Ser. S, 5, 6, 2012, 1021-1090.

[8] E. Fernández-Cara, I. Marín-Gayte, *Analysis and numerical solution of some minimal time control problems*, submited.

[9] I. V. Girsanov, *"Lectures on mathematical theory of extremum problems,"* Lecture Notes in Economics and Mathematical Systems, Springer-Verlag, Berlin, **67**, 1972.

[10] F. Hecht, `http://www.freefem.org`

[11] T. S. Ng, *Real time control engineering. Systems and automation,* Studies in Systems, Decision and Control, 65. Springer, [Singapore], 2016.

[12] J. Nocedal, S. J. Wright, *"Numerical Optimization "*, Springer Series in Operations Research, 1999.

[13] G. Wang, L. Wang, Y. Xu, Y. Zhang, *Time optimal control of evolution equations*, Progress in Nonlinear Differential Equations and their Applications, 92. Subseries in Control. Birkhäuser/Springer, Cham, 2018.

[14] Y. Xia, J. Zhang, K. Lu, N. Zhou, *Finite time and cooperative control of flight vehicles*, Advances in Industrial Control. Springer, Singapore, 2019.

Capítulo 5

Theoretical and numerical local null controllability of a quasi-linear parabolic equation in dimensions 2 and 3

This chapter is devoted to the theoretical and numerical analysis of the null controllability of a quasi-linear parabolic equation. First, we establish a local controllability result. The proof relies on an appropriate inverse function argument. Then, we formulate an iterative algorithm for the computation of the null control and we prove a convergence result. Finally, we illustrate the analysis with some numerical experiments. It is based on [9], in collaboration with J. Límaco.

5.1. Introduction and preliminaries

In this work, we analyze the theoretical and numerical null controllability in spatial dimensions 2 and 3 of the following PDE system

$$\begin{cases} y_t - \nabla \cdot (a(y)\nabla y) = v\tilde{1}_\omega, \quad (x,t) \in Q := \Omega \times (0,T), \\ y = 0, \quad (x,t) \in \Sigma := \partial\Omega \times (0,T), \\ y(x,0) = y_0(x), \quad x \in \Omega. \end{cases} \tag{5.1}$$

Here, $\Omega \subset \mathbb{R}^N$ is a bounded connected open set ($N \leq 3$), $\omega \subset\subset \Omega$ is a nonempty open set (the control domain), $\tilde{1}_\omega \in C_0^\infty(\Omega)$ satisfies $0 < \tilde{1}_\omega \leq 1$ in ω and $\tilde{1}_\omega = 0$ outside ω and $a \in C^3(\mathbb{R})$ possesses bounded derivatives of order ≤ 3 and satisfies

$$0 < m \leq a(r) \leq M \quad \forall r \in \mathbb{R}.$$

It will be said that (5.1) is (globally) null-controllable at time T if, for any $y_0 \in H_0^1(\Omega)$, there exists a control function $v \in L^2(\omega \times (0,T))$ and an associated state satisfying

$$y(x,T) = 0 \text{ in } \Omega. \tag{5.2}$$

On the other hand, it will be said that (5.1) is locally null-controllable at time T if there exists $\varepsilon > 0$ such that, for any $y_0 \in H_0^1(\Omega)$ with

$$\|y_0\|_{H_0^1} \leq \varepsilon,$$

there exists a control function $v \in L^2(\omega \times (0,T))$ and an associated state satisfying (5.2).

Recently, important progress has been made in the controllability analysis of linear and semilinear PDEs. We refer to the works [7, 8, 13, 16] and the references therein. For the controllability of equations with nonlocal terms, see [6]. For systems of the form (5.1) in spatial dimension 1, see [12].

Consequently, it is natural to try to extend the known results to equations like (5.1).

In this chapter, the first main result is the following:

Theorem 5.1.1. *Under the previous assumptions on the coefficient a, there exists $\varepsilon > 0$ such that, if $y_0 \in H^3(\Omega) \cap H_0^1(\Omega)$ and*

$$\|y_0\|_{H^3} \leq \varepsilon,$$

the nonlinear system (5.1) *possesses a solution satisfying* (5.2)*.*

Note that, in this result, the initial data is assumed to belong and be small in $H^3(\Omega) \cap H_0^1(\Omega)$. This is more restrictive than the condition required for the definition of a locally null-controllable system. As shown below, this is needed in the proof.

In order to prove Theorem 5.1.1, we will employ a technique relying on the so called *Liusternik's Inverse Function Theorem,* see [1]. Since the nonlinear term appears in the main part of the partial derivative operator, several nontrivial difficulties appear (among other things, we will have to work with regular and small initial data).

Thus, in a first step, we consider the linearized system at zero

$$\begin{cases} y_t - a(0)\Delta y = v\tilde{1}_\omega + h(x,t), \quad (x,t) \in Q, \\ y = 0, \quad (x,t) \in \Sigma, \\ y(x,0) = y_0(x), \quad x \in \Omega. \end{cases} \tag{5.3}$$

It is well known that, under some appropriate assumptions on h, (5.3) is null-controllable. More precisely, the adjoint of (5.3) is given by

$$\begin{cases} -\varphi_t - a(0)\Delta\varphi = F(x,t), \quad (x,t) \in Q, \\ \varphi = 0, \quad (x,t) \in \Sigma, \\ \varphi(x,T) = \varphi_T(x), \quad x \in \Omega, \end{cases} \tag{5.4}$$

where $\varphi_T \in L^2(\Omega)$; the announced null controllability property is implied by a well known Carleman inequality that can be established for any solution to a system of the form (5.4).

In a second step, we rewrite the null controllability problem for (5.1) as an equation in a well chosen space of "admissible" state-control pairs:

$$\mathcal{H}(y,v) = (0,y_0), \quad (y,v) \in Y. \tag{5.5}$$

Then, we apply Liusternik's Theorem and we deduce the (local) desired result. To this purpose, we previously have to establish some nontrivial estimates for the null controls and the associated states of (5.3).

This work is also devoted to the computation of a null control for (5.1). This is not a simple task; see [3–5, 15] for some achievements concerning the numerical controllability of linear and nonlinear PDEs. Here, we will argue as in [6, 12], taking advantage of the surjectivity of $\mathcal{H}'(0,0)$.

Thus, let Y be the Hilbert space where we can find a solution (y, v) to (5.5) (see (5.30a) in Section 5.3 for the precise definition of Y). We introduce the following iterative algorithm:

ALG 1:

1. Choose $(y^0, v^0) \in Y$.

2. Then, for given $n \geq 0$ and $(y^n, v^n) \in Y$, compute

$$(y^{n+1}, v^{n+1}) = (y^n, v^n) - \mathcal{H}'(0,0)^{-1}(\mathcal{H}(y^n, v^n) - (0, y_0)). \tag{5.6}$$

In these iterates, we use $\mathcal{H}'(0,0)^{-1}$, which is by definition an inverse to the left of $\mathcal{H}'(0,0)$; the precise definitions of $\mathcal{H}$ and $\mathcal{H}'(0,0)$ will be given in Section 5.4.

Note that **ALG 1** is an elementary quasi-Newton method and consequently has the following property: the finite dimensional approximations of the iterates lead to a set of algebraic systems whose coefficient matrices are always the same. This is very interesting and convenient from the numerical viewpoint, since it allows to perform just one factorization at the beginning and then compute quickly every (y^{n+1}, v^{n+1}).

In our second main result, we prove the convergence of **ALG 1** and we furnish some estimates:

Theorem 5.1.2. *Let $y_0 \in H^3(\Omega) \cap H_0^1(\Omega)$ be given with $\|y_0\|_{H^3} \leq \varepsilon$ (ε is furnished by Theorem 5.1.1). There exists $\kappa \in (0,1)$ such that, if $(y^0, v^0) \in Y$ and*

$$\|(y^0, v^0) - (y, v)\|_Y \leq \kappa,$$

then the (y^n, v^n) converge to (y, v) and satisfy

$$\|(y^{n+1}, v^{n+1}) - (y, v)\|_Y \leq \theta \|(y^n, v^n) - (y, v)\|_Y \tag{5.7}$$

for all $n \geq 0$ for some $\theta \in (0,1)$.

Remark 1. A natural question is whether Theorems 5.1.1 and 5.1.2 also hold for similar systems with PDEs of the form

$$y_t - \nabla \cdot (a(x,t;y)\nabla y) = v\tilde{1}_\omega,$$

that is, with a nonlinear diffusion coefficient nonhomogeneous in space and time. The answer is yes, provided $a : \overline{Q} \times \mathbb{R} \mapsto \mathbb{R}$ is regular enough. More precisely, if $a = a(x,t;r)$ satisfies (for instance)

$$0 < m \leq a(x,t;r) \leq M \quad \forall (x,t,r) \in \overline{Q} \times \mathbb{R}$$

and

$$\exists\, \frac{\partial^{n+m} a}{\partial x_i^n \partial r^m},\ \frac{\partial^{n+m} a}{\partial t^n \partial r^m} \in C_b^0(\overline{Q} \times \mathbb{R}) \ \text{ for } \ 0 \leq m \leq 3,\ \ 1 \leq i \leq N \ \text{ and } \ n = 0, 1,$$

the arguments in Sections 5.2–5.4 can be easily adapted to this situation. □

The chapter is organized as follows. Section 5.2 is devoted to recall some known results concerning the null controllability of the linearized system (5.3). In Section 5.3, we prove the first main result in this work (Theorem 5.1.1). Finally, Section 5.4 focuses on the proof of the second main result (Theorem 5.1.2), the numerical computation of the (y^n, v^n) and the presentation of the results of some numerical experiments.

In the sequel, we will denote by C a generic positive constant, usually depending on Ω, ω, T and sometimes the function a. Also, $\|\cdot\|$ and $(\cdot\,,\cdot)$ will stand for the usual norm and scalar product in $L^2(\Omega)$, respectively. Finally, we will use $\partial_i w$ to denote the partial derivative of w respect to x_i.

5.2. Carleman inequalities and the null controllability of (5.3)

Let $\omega_0 \subset\subset \omega$ be a non-empty open set. The following technical result, due to Fursikov and Imanuvilov [13], is fundamental.

Lemma 5.2.1. *There exists a function $\alpha_0 \in C^2(\overline{\Omega})$ satisfying:*

$$\begin{cases} \alpha_0(x) > 0 \quad \forall x \in \Omega, \quad \alpha_0(x) = 0 \quad \forall x \in \partial\Omega, \\ |\nabla \alpha_0(x)| > 0 \quad \forall x \in \overline{\Omega} \setminus \omega_0. \end{cases} \tag{5.8}$$

Let $m = m(t)$ be a function satisfying

$$m \in C^\infty([0,T]), \quad m \geq \frac{T^2}{8} \text{ in } [0, T/2], \quad m(t) = t(T-t) \text{ in } [T/2, T];$$

here we can consider, for example,

$$m(t) = \frac{T^2}{4} - \frac{T^2}{8} e^{-\frac{1}{(T/2-t)^2} + \frac{4}{T^2}} \text{ for } t \in [0, T/2].$$

which satisfies the aforementioned properties.

Let us set

$$\psi(x,t) := \frac{e^{\lambda\alpha_0(x)}}{t(T-t)}, \quad \beta(x,t) := \frac{\overline{\beta}(x)}{t(T-t)} := \frac{e^{R\lambda} - e^{\lambda\alpha_0(x)}}{t(T-t)},$$

where α_0 is as in Lemma 5.2.1, $R > \|\alpha_0\|_{L^\infty} + \log(2)$, $\lambda > 0$ and

$$\phi(x,t) := \frac{e^{\lambda\alpha_0(x)}}{m(t)}, \quad \alpha(x,t) := \frac{\overline{\alpha}(x)}{m(t)} := \frac{e^{R\lambda} - e^{\lambda\alpha_0(x)}}{m(t)}.$$

The following result from [13] is well known:

Lemma 5.2.2. *There exist positive constants λ_0, s_0 and C_0 only depending on Ω, ω and T such that, for any $s \geq s_0$ and $\lambda \geq \lambda_0$, any $F \in L^2(Q)$ and any $\varphi_T \in L^2(\Omega)$, the associated solution to* (5.4) *satisfies the Carleman estimate*

$$\begin{aligned} &\iint_Q e^{-2s\beta} \left[(s\psi)^{-1} \left(|\varphi_t|^2 + |\Delta\varphi|^2 \right) + \lambda^2 (s\psi) |\nabla\varphi|^2 + \lambda^4 (s\psi)^3 |\varphi|^2 \right] dx\, dt \\ &\quad \leq C_0 \left(\iint_Q e^{-2s\beta} |F|^2 \, dx\, dt + \iint_{\omega_0 \times (0,T)} e^{-2s\beta} \lambda^4 (s\psi)^3 |\varphi|^2 \, dx\, dt \right). \end{aligned} \tag{5.9}$$

From now on, we fix $s = s_0$ and $\lambda = \lambda_0$. The following result holds:

Proposition 1. *There exist a positive constant C only depending on Ω, ω and T such that, for any $F \in L^2(Q)$ and any $\varphi_T \in L^2(\Omega)$, the associated solution to* (5.4) *satisfies the Carleman estimate*

$$\begin{aligned} &\iint_Q e^{-2s\alpha} \left[\phi^{-1}\left(|\varphi_t|^2 + |\Delta\varphi|^2\right) + \phi|\nabla\varphi|^2 + \phi^3|\varphi|^2\right] dx\,dt \\ &\leq C\left(\iint_Q e^{-2s\alpha}\,|F|^2\,dx\,dt + \iint_{\omega_0\times(0,T)} e^{-2s\alpha}\phi^3|\varphi|^2\,dx\,dt\right), \end{aligned} \tag{5.10}$$

Proof: We want to get an estimate of the following integral:

$$\iint_Q e^{-2s\alpha} \left[\phi^{-1}\left(|\varphi_t|^2 + |\Delta\varphi|^2\right) + \phi|\nabla\varphi|^2 + \phi^3|\varphi|^2\right] dx\,dt.$$

First, we see that

$$\begin{aligned} &\iint_Q e^{-2s\alpha} \left[\phi^{-1}\left(|\varphi_t|^2 + |\Delta\varphi|^2\right) + \phi|\nabla\varphi|^2 + \phi^3|\varphi|^2\right] dx\,dt \\ &= \iint_{\Omega\times(0,T/2)} e^{-2s\alpha} \left[\phi^{-1}\left(|\varphi_t|^2 + |\Delta\varphi|^2\right) + \phi|\nabla\varphi|^2 + \phi^3|\varphi|^2\right] dx\,dt \\ &+ \iint_{\Omega\times(T/2,T)} e^{-2s\alpha} \left[\phi^{-1}\left(|\varphi_t|^2 + |\Delta\varphi|^2\right) + \phi|\nabla\varphi|^2 + \phi^3|\varphi|^2\right] dx\,dt. \end{aligned} \tag{5.11}$$

The last integral is bounded by the right hand side of (5.10) (for some appropriate C) thanks to Lemma 5.2.2.

Indeed, we have

$$\begin{aligned} &\iint_{\Omega\times(T/2,T)} e^{-2s\alpha} \left[\phi^{-1}\left(|\varphi_t|^2 + |\Delta\varphi|^2\right) + \phi|\nabla\varphi|^2 + \phi^3|\varphi|^2\right] dx\,dt \\ &= \iint_{\Omega\times(T/2,T)} e^{-2s\beta} \left[\psi^{-1}\left(|\varphi_t|^2 + |\Delta\varphi|^2\right) + \psi|\nabla\varphi|^2 + \psi^3|\varphi|^2\right] dx\,dt \\ &\leq C\left(\iint_Q e^{-2s\beta}\,|F|^2\,dx\,dt + \iint_{\omega_0\times(0,T)} e^{-2s\beta}\psi^3|\varphi|^2\,dx\,dt\right). \end{aligned}$$

In view of the definitions of α and β, one has $e^{-2s\beta} = e^{-2s\alpha}$ for $t \in (T/2, T)$; on the other hand, for $t \in (0, T/2)$, the following holds: $e^{-2s\beta} \leq 1 \leq K_0 e^{-2s\alpha}$ for $K_0 := e^{2sK_1}$ and $K_1 := e^{8R\lambda/T^2}$.

In a similar way, we also have $e^{-2s\beta}\psi^3 \leq Ce^{-2s\alpha}\phi^3$ in Q: one has $e^{-2s\beta}\psi^3 = e^{-2s\alpha}\phi^3$ for $t \in (T/2, T)$; also, there exist constants K_2 and K_3 (only depending on Ω, ω and T) such that $e^{-2s\beta}\psi^3 \leq K_2 \leq K_3 e^{-2s\alpha}\phi^3$ for $t \in (0, T/2)$.

As a consequence, we can bound the last two integrals above respectively by

$$C\int_Q e^{-2s\alpha}\,|F|^2 \quad \text{and} \quad C\int_{\omega_0\times(0,T)} e^{-2s\alpha}\phi^3|\varphi|^2$$

and our assertion follows.

On the other hand, in order to bound the first integral in the right hand side of (5.11), we can use energy estimates.

Indeed, by multiplying (5.4) by φ and integrating in space and then in time in $[t,s]$, with $t \in [0,T/2]$ and $s \in [T/2,3T/4]$, we can see that

$$\frac{1}{2}\|\varphi(\cdot,t)\|^2 + a(0)\iint_{\Omega\times(t,s)} |\nabla\varphi|^2\,dx\,d\sigma = \iint_{\Omega\times(t,s)} F\cdot\varphi\,dx\,d\sigma + \frac{1}{2}\|\varphi(\cdot,s)\|^2$$
$$\leq C\iint_{\Omega\times(0,3T/4)} |F|^2\,dx\,d\sigma + \frac{a(0)}{2}\iint_{\Omega\times(t,s)} |\nabla\varphi|^2\,dx\,d\sigma + \frac{1}{2}\|\varphi(\cdot,s)\|^2$$

and

$$\|\varphi(\cdot,t)\|^2 + a(0)\iint_{\Omega\times(0,T/2)} |\nabla\varphi|^2\,dx\,d\sigma \leq C\iint_{\Omega\times(0,3T/4)} |F|^2\,dx\,d\sigma + \|\varphi(\cdot,s)\|^2.$$

Now, integrating with respect to s in $[T/2,3T/4]$, one has

$$\begin{aligned}\frac{T}{4}\|\varphi(\cdot,t)\|^2 + \frac{a(0)T}{4}\iint_{\Omega\times(0,T/2)} |\nabla\varphi|^2\,dx\,d\sigma \\ \leq C\iint_{\Omega\times(0,3T/4)} |F|^2\,dx\,d\sigma + \iint_{\Omega\times(T/2,3T/4)} |\varphi|^2\,dx\,d\sigma.\end{aligned} \tag{5.12}$$

On the other hand, by multiplying (5.4) by $-\Delta\varphi$ and integrating again in space and time, the following holds:

$$\frac{1}{2}\|\nabla\varphi(\cdot,t)\|^2 + a(0)\iint_{\Omega\times(t,s)} |\Delta\varphi|^2\,dx\,d\sigma = \iint_{\Omega\times(t,s)} F(-\Delta\varphi)\,dx\,d\sigma + \frac{1}{2}\|\nabla\varphi(\cdot,s)\|^2,$$

whence

$$\|\nabla\varphi(\cdot,t)\|^2 + a(0)\iint_{\Omega\times(0,T/2)} |\Delta\varphi|^2\,dx\,d\sigma \leq C\iint_{\Omega\times(0,3T/4)} |F|^2\,dx\,d\sigma + \|\nabla\varphi(\cdot,s)\|^2$$

and, as before, integration with respect to s in $[T/2,3T/4]$ yields

$$\begin{aligned}\frac{T}{4}\|\nabla\varphi(\cdot,t)\|^2 + \frac{a(0)T}{4}\iint_{\Omega\times(0,T/2)} |\Delta\varphi|^2\,dx\,d\sigma \\ \leq C\iint_{\Omega\times(0,3T/4)} |F|^2\,dx\,d\sigma + \iint_{\Omega\times(T/2,3T/4)} |\nabla\varphi|^2\,dx\,d\sigma.\end{aligned} \tag{5.13}$$

Finally, we note from (5.4) that

$$|\varphi_t|^2 \leq \frac{1}{2}|\Delta\varphi|^2 + \frac{1}{2}|F|^2. \tag{5.14}$$

From (5.12)- (5.14), we obtain:

$$\iint_{\Omega\times(0,T/2)} \left[|\varphi_t|^2 + |\Delta\varphi|^2 + |\nabla\varphi|^2 + |\varphi|^2\right]\,dx\,dt$$
$$\leq C\left(\iint_{\Omega\times(0,3T/4)} |F|^2\,dx\,dt + \iint_{\Omega\times(T/2,3T/4)} \left[|\nabla\varphi|^2 + |\varphi|^2\right]\,dx\,dt\right).$$

Taking into account that α and ϕ are uniformly bounded from above and from below in $\Omega \times (0, 3T/4)$, we can deduce from this inequality the required estimate.

Indeed, let us first note that α is bounded from above in $\Omega \times (0, 3T/4)$ by a constant K_4 (again depending only on Ω, ω and T). Consequently,

$$\iint_{\Omega\times(0,3T/4)} |F|^2\,dx\,dt \le K_5 \iint_Q e^{-2s\alpha}|F|^2\,dx\,dt,$$

where $K_5 := e^{2sK_4}$. On the other hand, we also have $e^{2s\beta}\psi^{-1} \le K_6$ and $e^{2s\beta}\psi^{-3} \le K_7$ in $\Omega \times (T/2, 3T/4)$ for similar positive constants. Recalling Lemma 2.2 and arguing as above, we get:

$$\begin{aligned}
&\iint_{\Omega\times(T/2,3T/4)} \left[|\nabla\varphi|^2 + |\varphi|^2\right]\,dx\,dt \\
&\quad\le K_6 \iint_Q e^{-2s\beta}\psi|\nabla\varphi|^2\,dx\,dt + K_7 \iint_Q e^{-2s\beta}\psi^3|\varphi|^2\,dx\,dt \\
&\quad\le C\left(\iint_Q e^{-2s\beta}\,|F|^2\,dx\,dt + \iint_{\omega_0\times(0,T)} e^{-2s\beta}\psi^3|\varphi|^2\,dx\,dt\right) \\
&\quad\le C\left(\iint_Q e^{-2s\alpha}\,|F|^2\,dx\,dt + \iint_{\omega_0\times(0,T)} e^{-2s\alpha}\phi^3|\varphi|^2\,dx\,dt\right)
\end{aligned}$$

Hence, the first integral in the right hand side of (5.11) is also bounded by the right hand side of (5.10).

This ends the proof. □

Let us introduce the weights

$$\begin{aligned}
\rho := e^{s\alpha}, \quad \rho_3 := e^{s\alpha}\phi^{-3/2}, \quad \rho_5 := e^{s\alpha}m^{5/2}, \\
\rho_7 := e^{s\alpha}m^{7/2}, \quad \rho_9 := e^{s\alpha}m^{9/2}, \quad \rho_{11} := e^{s\alpha}m^{11/2}
\end{aligned} \tag{5.15}$$

and the constants

$$\alpha_1 := \min_{x\in\overline{\Omega}} \overline{\alpha}(x) = e^{R\lambda} - e^{\lambda\|\alpha_0\|_{L^\infty}}, \quad \alpha_2 := \max_{x\in\overline{\Omega}} \overline{\alpha}(x) = e^{R\lambda} - 1.$$

It is then clear that $2\alpha_1 > \alpha_2$ and the following inequalities hold:

$$e^{s\alpha_1/m(t)} < e^{s\alpha(x,t)} < e^{s\alpha_2/m(t)} < e^{2s\alpha_1/m(t)}.$$

The main result in this section is the following. It relies on the null controllability of (5.3), that is of course well known. The estimates in this result are also known in the literature and have been used with various purposes, although we believe that is is useful to recall them here, as well as their proofs.

Theorem 5.2.3. *Assume that the function h in* (5.3) *satisfies $\rho_3 h \in L^2(Q)$, $\rho_9 h_t \in L^2(Q)$ and $h(\cdot\,,0) \in H_0^1(\Omega)$. Then* (5.3) *is null-controllable. In fact, for each $y_0 \in H_0^1(\Omega)$, there exist null controls v with*

$$\rho_7 v \in L^2(0,T;H^2(\omega)) \cap C^0([0,T];H^1(\omega)), \;\; (\rho_7 v)_t \in L^2(\omega\times(0,T)), \;\; v(\cdot\,,0) \in H^1(\omega) \tag{5.16}$$

and associated states y satisfying (5.2). *Furthermore, if $y_0 \in H^3(\Omega)\cap H_0^1(\Omega)$, the associated states are such that*

$$N_1(y):=\iint_Q[\rho^2|y|^2+\rho_5^2|\nabla y|^2+\rho_7^2(|y_t|^2+|\Delta y|^2)+\rho_9^2|\nabla y_t|^2+\rho_{11}^2(|y_{tt}|^2+|\Delta y_t|^2)]\,dx\,dt<+\infty,$$
$$N_2(y):=\sup_{0\le t\le T}\int_\Omega[\rho_5|y|^2+\rho_7^2|\nabla y|^2+\rho_9^2(|y_t|^2+|\Delta y|^2)+\rho_{11}^2|\nabla y_t|^2]\,dx<+\infty. \tag{5.17}$$

Proof: Let us introduce some notation:

- $Ly := y_t - a(0)\Delta y, \quad L^*\varphi := -\varphi_t - a(0)\Delta\varphi,$
- $P_0 = \{\varphi \in C^2(\overline{Q}) :\ \varphi = 0 \text{ on } \Sigma\}, \quad \pi(\varphi,\tilde\varphi) := \iint_Q \left(\rho^{-2}L^*\varphi L^*\tilde\varphi + 1_\Omega\rho_3^{-2}\varphi\tilde\varphi\right)dx\,dt,$
- $P =$ the completion of P_0 for the scalar product $\pi(\cdot\,,\cdot)$.
- $b(\varphi) := \displaystyle\int_\Omega y_0(x)\varphi(x,0)\,dx + \iint_Q h(x,t)\varphi\,dx\,dt.$

Then, b is continuous linear form on the Hilbert space P (thanks to the Carleman inequality (5.10)).

From the results in [13], we know that, for any $y_0 \in L^2(\Omega)$ and any h with

$$\iint_Q \rho_3^2|h|^2 < +\infty,$$

there exist controls $v \in L^2(\omega\times(0,T))$ and associated solutions to (5.3) satisfying (5.2). The couples (y,v) can be found by minimizing

$$\iint_Q \rho^2|y|^2\,dx\,dt + \iint_{\omega\times(0,T)}\rho_3^2|v|^2\,dx\,dt$$

in the family of the state-control pairs (y,v) with $v \in L^2(\omega\times(0,T))$. Furthermore, we have

$$\iint_Q \rho^2|y|^2\,dx\,dt + \iint_{\omega\times(0,T)}\rho_3^2|v|^2\,dx\,dt \le C\Big(\|y_0\|^2 + \iint_Q \rho_3^2|h|^2\,dx\,dt\Big)$$

and

$$\iint_Q e^{-2s\alpha}\Big[\phi^{-1}(|\varphi_t|^2+|\Delta\varphi|^2)+\phi|\nabla\varphi|^2\Big]\,dx\,dt \le C\Big(\|y_0\|^2 + \iint_Q \rho_3^2|h|^2\,dx\,dt\Big).$$

Accordingly, we can write

$$y = \rho^{-2}L^*\varphi, \quad v = -\rho_3^{-2}\,\varphi|_{\omega\times(0,T)}\,, \tag{5.18}$$

where φ is the unique solution to the linear problem

$$\pi(\varphi,\tilde\varphi) = b(\tilde\varphi)\ \ \forall\tilde\varphi\in P,\ \ \varphi\in P.$$

Let us set $z := -\rho_3^{-2}\varphi$. Then $\rho_7 z = \rho m^{7/2} z = -\rho^{-1} m^{1/2} e^{3\lambda\alpha_0}\varphi$ and, after some computations, we see that

$$\begin{aligned} L^*(\rho_7 z) &= -\rho^{-1} m^{1/2} e^{3\lambda\alpha_0} L^*\varphi \\ &\quad -\partial_t(\rho^{-1} m^{1/2} e^{3\lambda\alpha_0})\varphi + 2\nabla(\rho^{-1} m^{1/2} e^{3\lambda\alpha_0})\cdot\nabla\varphi + \Delta(\rho^{-1} m^{1/2} e^{3\lambda\alpha_0})\varphi \\ &= J_1 + J_2 + 2J_3 + J_4. \end{aligned} \tag{5.19}$$

In view of (5.18), one has:

$$|J_1| \le \rho^{-1} m^{1/2} e^{3\lambda\alpha_0}\rho^2|y| \le C\rho|y|. \tag{5.20a}$$

Also, in view of the facts that $m^{-a} \le Cm^{-b}$ if $a < b$ and m' is bounded and the definition of ρ, we have that

$$|J_2| + |J_4| \le C\rho^{-1} m^{-3/2}|\varphi| \tag{5.20b}$$

and

$$|J_3| \le C\rho^{-1} m^{-1/2}|\nabla\varphi|. \tag{5.20c}$$

From the Carleman estimate (5.10) written for φ and the fact that $y = \rho^{-2}L^*\varphi$, we get:

$$\begin{aligned} &\iint_Q \rho^{-2}\lambda^2(s\phi)|\nabla\varphi|^2\,dx\,dt + \iint_Q \rho^{-2}\lambda^4(s\phi)^3|\varphi|^2\,dx\,dt \\ &\quad \le \iint_Q \rho^2|y|^2\,dx\,dt + \iint_{\omega\times(0,T)} \rho_3^2|v|^2\,dx\,dt < +\infty, \end{aligned}$$

whence

$$\iint_Q \rho^{-2} m^{-1}|\nabla\varphi|^2\,dx\,dt + \iint_Q \rho^{-2} m^{-3}|\varphi|^2\,dx\,dt < +\infty. \tag{5.20d}$$

From (5.20a)-(5.20d), we deduce that $J_1 + J_2 + 2J_3 + J_4 \in L^2(Q)$. Consequently, taking into account the PDE satisfied by $\rho_7 z$ and the fact that $(\rho_7 z)(\cdot\,, T) = 0$, we see that

$$\rho_7 z \in L^2(0,T; H^2(\Omega)) \cap C^0([0,T]; H_0^1(\Omega)),\ \ (\rho_7 z)_t \in L^2(Q). \tag{5.21}$$

In particular, (5.16) holds.

Notice that, up to now, we have only used that $\rho_3 h \in L^2(Q)$ and $y_0 \in L^2(\Omega)$. In order to get (5.17), we will establish several estimates:

Estimates I: Multiplying (5.3) by $\rho_5^2 y$ and integrating in Ω, we have that

$$\begin{aligned} &\frac{1}{2}\frac{d}{dt}\int_\Omega \rho_5^2|y|^2\,dx + a(0)\int_\Omega \rho_5^2|\nabla y|^2\,dx \\ &\quad \le C\left(\int_\omega \rho_3^2|v|^2\,dx + \int_\Omega \rho_3^2|h|^2\,dx + \int_\Omega \rho^2|y|^2\,dx\right) + \int_\Omega \rho_5^2|y|^2\,dx, \end{aligned}$$

due to $|\rho_{1,t}| \le C\rho^2$, $|\nabla\cdot(\rho_5\nabla\rho_5)| \le C\rho^2$ and $\rho_5^2 \le C\rho_3^2$. In view of Gronwall Lemma, we find that

$$\begin{aligned} &\sup_{0\le t\le T}\int_\Omega \rho_5^2|y|^2\,dx + \iint_Q \rho_5^2|\nabla y|^2\,dx\,dt \\ &\quad \le C\left(\iint_{\omega\times(0,T)} \rho_3^2|v|^2\,dx\,dt + \iint_Q \rho_3^2|h|^2\,dx\,dt + \|y_0\|^2\right). \end{aligned} \tag{5.22}$$

Now, multiplying (5.3) by $\rho_7^2 y_t$ and integrating in Ω, we see that

$$\begin{aligned}&\int_\Omega \rho_7^2|y_t|^2\,dx + a(0)\frac{d}{dt}\int_\Omega \rho_7^2|\nabla y|^2\,dx\\ &\quad\le C\left(\int_\omega \rho_3^2|v|^2\,dx + \int_\Omega \rho_3^2|h|^2\,dx + \int_\Omega \rho_5^2|\nabla y|^2\,dx\right)\end{aligned}$$

due to $|\rho_7|^2 \le C\rho_3^2$, $|\rho_7(\rho_7)_t| \le C\rho_5^2$ and $|\nabla\rho_7|^2 \le C\rho_3^2$, whence using (5.22) one has

$$\begin{aligned}&\iint_Q \rho_7^2|y_t|^2\,dx\,dt + \sup_{0\le t\le T}\int_\Omega \rho_7^2|\nabla y|^2\,dx\\ &\quad\le C\left(\iint_{\omega\times(0,T)} \rho_3^2|v|^2\,dx\,dt + \iint_Q \rho_3^2|h|^2\,dx\,dt + \|y_0\|^2_{H_0^1}\right).\end{aligned}\tag{5.23}$$

In a similar way, multiplying (5.3) by $-\rho_7^2\Delta y$ and integrating in $\Omega\times(0,T)$, the following is found:

$$\begin{aligned}&\sup_{0\le t\le T}\int_\Omega \rho_7^2|\nabla y|^2\,dx + \iint_Q \rho_7^2|\Delta y|^2\,dx\,dt\\ &\quad\le C\left(\iint_{\omega\times(0,T)} \rho_3^2|v|^2\,dx\,dt + \iint_Q \rho_3^2|h|^2\,dx\,dt + \|y_0\|^2_{H_0^1}\right).\end{aligned}\tag{5.24}$$

Observe that, in order to get (5.23) and (5.24), we just need $\rho_3 h \in L^2(Q)$ and $y_0 \in H_0^1(\Omega)$.

Estimates II: Let us assume that $y_0 \in H^2(\Omega)\cap H_0^1(\Omega))$. Differentiating with respect to time the PDE in (5.3), one has

$$y_{tt} - a(0)\Delta y_t = v_t\widetilde{1}_\omega + h_t.\tag{5.25}$$

Multiplying by $\rho_2^2 y_t$ and integrating in Ω, we now have

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\int_\Omega \rho_9^2|y_t|^2\,dx + a(0)\int_\Omega \rho_9^2|\nabla y_t|^2\,dx\\ &\quad= \int_\omega \rho_9^2 y_t v_t\,dx + \int_\Omega \rho_9^2 y_t h_t\,dx + \int_\Omega \rho_9\rho_{2,t}|y_t|^2\,dx + 2a(0)\int_\Omega \rho_9\nabla\rho_9\cdot\nabla y_t y_t\,dx.\end{aligned}$$

Arguing as in Estimates I, we arrive at the estimate

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\int_\Omega \rho_9^2|y_t|^2\,dx + a(0)\int_\Omega \rho_9^2|\nabla y_t|^2\,dx\\ &\quad\le C\Big(\int_\omega \rho_9^2|v_t|^2\,dx + \int_\Omega \rho_7^2|h_t|^2\,dx + \int_\Omega \rho_7^2|y_t|^2\,dx\Big).\end{aligned}$$

Now, using that $v = -\rho_3^{-2}\varphi|_{\omega\times(0,T)}$, $\rho_9^2\rho^{-4} \le Ce^{-2s\alpha}m^3 \le Ce^{-2s\alpha}\phi^{-1}$ and the weights are given by (5.15), we can easily show that

$$\begin{aligned}&\frac{1}{2}\frac{d}{dt}\int_\Omega \rho_9^2|y_t|^2\,dx + a(0)\int_\Omega \rho_9^2|\nabla y_t|^2\,dx\\ &\quad\le C\left(\int_\omega \rho_3^2|v|^2dx + \int_\Omega e^{-2s\alpha}\phi^{-1}|\varphi_t|^2\,dx + \int_\Omega \rho_7^2|h_t|^2\,dx + \int_\Omega \rho_7^2|y_t|^2\,dx\right).\end{aligned}$$

And, recalling (5.23) and the inequality to φ, we find that

$$\begin{aligned}&\sup_{0\le t\le T}\int_\Omega \rho_9^2|y_t|^2\,dx+\iint_Q \rho_9^2|\nabla y_t|^2\,dx\,dt\\&\quad\le C\Big(\iint_{\omega\times(0,T)}\rho_3^2|v|^2\,dx\,dt+\iint_Q\Big[\rho_7^2|h_t|^2+\rho_3^2|h|^2\Big]\,dx\,dt+\|y_0\|_{H^2}^2\Big).\end{aligned}\tag{5.26}$$

In the same way, multiplying (5.3) by $-\rho_{11}^2\Delta y_t$ and integrating in Ω, we have:

$$\begin{aligned}&\int_\Omega \rho_{11}^2|\nabla y_t|^2\,dx+\frac{a(0)}{2}\frac{d}{dt}\int_\Omega\rho_{11}^2|\Delta y|^2\,dx\\&\quad=-\int_\omega\rho_{11}^2\Delta y_t v\,dx-\int_\Omega\rho_{11}^2\Delta y_t h\,dx+\int_\Omega\nabla(\rho_{11}\cdot\nabla\rho_{11})|y_t|^2\,dx+a(0)\int_\Omega\rho_{11}(\rho_{11})_t|\Delta y|^2\end{aligned}$$

and

$$\begin{aligned}&\int_\Omega \rho_{11}^2|\nabla y_t|^2\,dx+\frac{a(0)}{2}\frac{d}{dt}\int_\Omega\rho_{11}^2|\Delta y|^2\,dx\\&\quad\le C\Big(\int_\omega\rho_3^2|v|^2dx+\int_\Omega e^{-2s\alpha}\phi^{-1}|\varphi_t|^2\,dx+\int_\Omega\left[\rho_7^2|h_t|^2\,dx+\rho_3^2|h|^2\right]\,dx\Big)+\int_\Omega\rho_{11}^2|\Delta y|^2\,dx.\end{aligned}$$

This gives

$$\begin{aligned}&\iint_Q\rho_{11}^2|\nabla y_t|^2\,dx\,dt+\sup_{0\le t\le T}\int_\Omega\rho_{11}^2|\Delta y|^2\,dx\\&\quad\le C\left(\iint_{\omega\times(0,T)}\rho_3^2|v|^2\,dx\,dt+\iint_Q\Big[\rho_7^2|h_t|^2+\rho_3^2|h|^2\Big]\,dx\,dt+\|y_0\|_{H^2}^2\right).\end{aligned}\tag{5.27}$$

Here, we have only needed $\rho_3 h\in L^2(Q)$, $\rho_7 h_t\in L^2(Q)$ and $y_0\in H^2(\Omega)\cap H_0^1(\Omega)$.

Estimates III: Now, let us multiply (5.25) by $\rho_{11}^2 y_{tt}$. Then

$$\begin{aligned}&\int_\Omega\rho_{11}^2|y_{tt}|^2\,dx+a(0)\frac{d}{dt}\int_\Omega\rho_{11}^2|\nabla y_t|^2\,dx\\&\quad\le C\Big(\int_\omega\rho_3^2|v|^2dx+\int_\Omega e^{-2s\alpha}\phi^{-1}|\varphi_t|^2\,dx+\int_\Omega[\rho_7^2|h_t|^2\,dx+\rho_3^2|h|^2]\,dx\Big)+\int_\Omega\rho_9^2|\nabla y_t|^2\,dx,\end{aligned}$$

whence we easily obtain

$$\begin{aligned}&\iint_Q\rho_{11}^2|y_{tt}|^2\,dx+\sup_{0\le t\le T}\int_\Omega\rho_{11}^2|\nabla y_t|^2\,dx\\&\quad\le C\Big(\iint_{\omega\times(0,T)}\rho_3^2|v|^2\,dx\,dt+\iint_Q\Big[\rho_7^2|h_t|^2+\rho_3^2|h|^2\Big]\,dx\,dt+\|y_0\|_{H^3}^2\Big)\end{aligned}\tag{5.28}$$

Finally, multiplying (5.25) by $-\rho_{11}^2\Delta y_t$ and integrating in space and then in time, another estimate is obtained:

$$\begin{aligned}&\sup_{0\le t\le T}\int_\Omega\rho_{11}^2|\nabla y_t|^2\,dx+\iint_Q\rho_{11}^2|\Delta y_t|^2\,dx\,dt\\&\quad\le C\Big(\iint_{\omega\times(0,T)}\rho_3^2|v|^2\,dx\,dt+\iint_Q\Big(\rho_7^2|h_t|^2+\rho_3^2|h|^2\Big)dx\,dt+\|y_0\|_{H^3}^2\Big)\end{aligned}\tag{5.29}$$

Observe again that, in order to prove (5.28) and (5.29), we have assumed that $\rho_3 h \in L^2(Q)$, $\rho_7 h_t \in L^2(Q)$, $h(\cdot\,,0) \in H_0^1(\Omega)$ and $y_0 \in H^3(\Omega)\cap H_0^1(\Omega)$. Obviously, these estimates imply (5.17).

This ends the proof. □

5.3. Proof of Theorem 5.1.1

In this section we will prove the local null controllability of the nonlinear system (5.1).

Let Y, F and Z be the following spaces of functions:

$$\begin{aligned} Y := \{(y,v) : v, v_t \in L^2(\omega\times(0,T)),\ & \iint_{\omega\times(0,T)} (\rho_7^2|v_t|^2 + \rho_3^2|v|^2)\,dx\,dt < +\infty, \\ & y, \partial_i y, \Delta y, y_t, \partial_i y_t, \Delta y_t, y_{tt} \in L^2(Q),\ \ N_1(y) + N_2(y) < +\infty, \\ & \iint_Q \rho_3^2 |y_t - a(0)\Delta y - v\tilde{1}_\omega|^2\,dx\,dt < +\infty, \\ & \iint_Q \rho_7^2 |y_{tt} - a(0)\Delta y_t - v_t\tilde{1}_\omega|^2\,dx\,dt < +\infty, \\ & v(\cdot\,,0) \in H^1(\omega),\ \ y(\cdot\,,0) \in H^3(\Omega)\cap H_0^1(\Omega), \\ & (y_t - a(0)\Delta y - v\tilde{1}_\omega)(\cdot\,,0) \in H_0^1(\Omega)\}, \end{aligned} \tag{5.30a}$$

$$F := \Big\{ g \in L^2(Q) :\ \iint_Q (\rho_3^2|g|^2 + \rho_7^2|g_t|^2)\,dx\,dt < +\infty,\ \ g(\cdot\,,0) \in H_0^1(\Omega)\Big\}, \tag{5.30b}$$

$$Z := F \times \Big(H^3(\Omega)\cap H_0^1(\Omega)\Big). \tag{5.30c}$$

We introduce the (natural) Hilbertian norms:

$$\begin{aligned} \|(y,v)\|_Y^2 :=\ & N_1(y) + N_2(y) + \iint_{\omega\times(0,T)} (\rho_7^2|v_t|^2 + \rho_3^2|v|^2)\,dx\,dt \\ & + \iint_Q \rho_3^2 |y_t - a(0)\Delta y - v\tilde{1}_\omega|^2\,dx\,dt \\ & + \iint_Q \rho_7^2 |y_{tt} - a(0)\Delta y_t - v_t\tilde{1}_\omega|^2\,dx\,dt \\ & + \|v(\cdot\,,0)\|_{H^1(\omega)}^2 + \|y(\cdot\,,0)\|_{H^3}^2, \end{aligned}$$

$$\|g\|_F^2 := \iint_Q (\rho_3^2|g|^2 + \rho_7^2|g_t|^2)\,dx\,dt + \|g(\cdot\,,0)\|_{H_0^1}^2.$$

Let us consider the mapping $\mathcal{H} : Y \mapsto Z$, with

$$\mathcal{H}(y,v) = (y_t - \nabla\cdot(a(y)\nabla y) - v\tilde{1}_\omega, y(\cdot\,,0)). \tag{5.31}$$

We will prove that there exists $\varepsilon > 0$ such that, if $(h, y_0) \in Z$ and $\|(h,y_0)\|_Z \le \varepsilon$, then the equation

$$\mathcal{H}(y,v) = (h, y_0), \quad (y,v) \in Y,$$

possesses at least one solution. In particular, this will show that (5.1) is locally null controllable and, furthermore, that the state-controls pairs that fulfill (5.2) can be found in Y.

We will apply the following version of *Liusternik's Inverse Mapping Theorem* in infinite dimensional spaces, whose proof can be found for instance in [1]. In the following statement, $B_r(0)$ and $B_\varepsilon(\xi_0)$ are open balls respectively of radii r and ε.

Theorem 5.3.1. *Let Y and Z be Banach spaces and let $\mathcal{H} : B_r(0) \subset Y \mapsto Z$ be a C^1-mapping. Let us assume that the derivative $\mathcal{H}'(0) : Y \mapsto Z$ is onto and let us set $\xi_0 = H(0)$. Then there exist $\varepsilon > 0$, a mapping $W : B_\varepsilon(\xi_0) \subset Z \mapsto Y$ and a constant $K > 0$ satisfying:*

$$\begin{cases} W(z) \in B_r(0) \text{ and } \mathcal{H}(W(z)) = z \quad \forall z \in B_\varepsilon(\xi_0), \\ \|W(z)\|_Y \leq K\|z - \mathcal{H}(0)\|_Z \quad \forall z \in B_\varepsilon(\xi_0). \end{cases}$$

Note that, in particular, W is the *inverse to the right* of $\mathcal{H}$. In order to show that Theorem 5.3.1 can be applied in this setting, we will use several lemmas:

Lemma 5.3.2. *Let $\mathcal{H} : Y \mapsto Z$ be the mapping defined by* (5.31)*. Then $\mathcal{H}$ is well defined and continuous.*

Proof: For any $(y, z) \in Y$, we have:

$$\begin{aligned}
&\|\mathcal{H}(y,v)\|_Z^2 \\
&= \iint_Q \rho_3^2[y_t - \nabla\cdot(a(y)\nabla y) - v\tilde{1}_\omega]^2\,dx\,dt \\
&\quad + \iint_Q \rho_7^2[(y_t - \nabla\cdot(a(y)\nabla y) - v\tilde{1}_\omega)_t]^2\,dx\,dt \\
&\quad + \|(y_t - \nabla\cdot(a(y)\nabla y) - v\tilde{1}_\omega)(\cdot\,,0)\|_{H_0^1}^2 + \|y(\cdot\,,0)\|_{H^3}^2 \\
&\leq 2\left(\iint_Q \rho_3^2|y_t - a(0)\Delta y - v\tilde{1}_\omega|^2\,dx\,dt + \iint_Q \rho_7^2|(y_t - a(0)\Delta y - v\tilde{1}_\omega)_t|^2\,dx\,dt\right) \\
&\quad + 2\iint_Q \rho_3^2|\nabla\cdot((a(y)-a(0))\nabla y)|^2\,dx\,dt + 2\iint_Q \rho_7^2|(\nabla\cdot(a(y)\nabla y) - a(0)\Delta y)_t|^2 dx\,dt \\
&\quad + 2\|(y_t - a(0)\Delta y - v\tilde{1}_\omega)(\cdot\,,0)\|_{H_0^1}^2 + 2\|\nabla\cdot((a(y)-a(0))\nabla y)(\cdot\,,0)\|_{H_0^1}^2 \\
&\quad + \|y(\cdot\,,0)\|_{H^3}^2 \\
&= 2I_1 + 2I_2 + 2I_3 + 2I_4 + 2I_5 + I_6.
\end{aligned}$$

From the definition of the space Y, since $N_1(y) < +\infty$ and $y(\cdot\,,0) \in H^3(\Omega)$, one easily has

$$I_1 + I_4 + I_6 \leq C\|(y,v)\|_Y^2 \tag{5.32}$$

Let us analyze I_2. Since a is C^1 and globally Lipschitz-continuous, one has:

$$\begin{aligned}
I_2 &= \iint_Q \rho_3^2|\nabla\cdot((a(y)-a(0))\nabla y)|^2\,dx\,dt \\
&\leq C\iint_Q \rho_3^2|y|^2|\Delta y|^2 dx\,dt + C\iint_Q \rho_3^2|\nabla y|^4 dx\,dt \\
&= CS_1 + CS_2
\end{aligned} \tag{5.33}$$

Since $N \leq 3$, we have $H^2(\Omega) \hookrightarrow L^\infty(\Omega)$ and $H^2(\Omega) \hookrightarrow W^{1,4}(\Omega)$ with continuous (and compact) embeddings. Therefore, from (5.8) and (5.17), one has

$$\begin{aligned} S_1 &\leq \iint_Q e^{2s\alpha_2/m} m^3 |y|^2 |\Delta y|^2 \, dx \, dt \\ &\leq \int_0^T e^{-2s\alpha_2/m} m^{-11} \left(\rho_7^2 \sup_\Omega |y|^2 \right) \left(\rho_7^2 \int_\Omega |\Delta y|^2 \, dx \right) \, dt \\ &\leq C \left(\sup_{0\leq t\leq T} \rho_7^2 \, \|\Delta y\|^2_{L^2(\Omega)} \right)^2 \\ &\leq C \|(y,v)\|_Y^4. \end{aligned} \tag{5.34}$$

and

$$\begin{aligned} S_2 &\leq \int_0^T e^{2s\alpha_2/m} m^3 \int_\Omega |\nabla y|^4 dx \, dt \\ &\leq \int_0^T e^{-2s\alpha_2/m} m^{-11} \rho_7^4 \|\Delta y\|^4_{L^2(\Omega)} \, dt \\ &\leq C \left(\sup_{0\leq t\leq T} \rho_7^2 \|\Delta y\|^2_{L^2(\Omega)} \right)^2 \\ &\leq C \|(y,v)\|_Y^4. \end{aligned} \tag{5.35}$$

From (5.34) and (5.35), we find that

$$I_2 \leq C \|(y,v)\|_Y^4. \tag{5.36}$$

On the other hand,

$$\begin{aligned} I_3 &= \iint_Q \rho_7^2 |(\nabla \cdot (a(y)\nabla y) - a(0)\Delta y)_t|^2 \, dx \, dt \\ &\leq 4 \iint_Q \rho_7^2 |a'(y) y_t \Delta y|^2 \, dx \, dt + 4 \iint_Q \rho_7^2 |(a(y) - a(0)) \Delta y_t|^2 \, dx \, dt \\ &\quad + 4 \iint_Q \rho_7^2 |a''(y) y_t|^2 |\nabla y|^4 \, dx \, dt + 16 \iint_Q \rho_7^2 |2a'(y) \nabla y \nabla y_t|^2 \, dx \, dt \\ &= 4L_1 + 4L_2 + 4L_3 + 16L_4. \end{aligned} \tag{5.37}$$

For instance,

$$\begin{aligned} L_1 &\leq C \iint_Q \rho_7^2 |y_t|^2 |\Delta y|^2 \, dx \, dt \leq C \int_0^T e^{2s\alpha_2/m} \, m^9 \, \|\Delta y_t\|^2_{L^2} \, \|\Delta y\|^2_{L^2} \, dt \\ &\leq C \left(\iint_Q \rho_{11}^2 |\Delta y_t|^2 \, dx \, dt \right) \left(\sup_{0\leq t\leq T} \rho_9^2 \|\Delta y\|^2_{L^2} \right) \\ &\leq C \|(y,v)\|_Y^4. \end{aligned} \tag{5.38}$$

We also have $L_2 + L_4 \leq C\|(y,v)\|_Y^4$ and $L_3 \leq C\|(y,v)\|_Y^6$. Accordingly,

$$I_3 \leq C(\|(y,v)\|_Y^4 + \|(y,v)\|_Y^6). \tag{5.39}$$

Finally,

$$\begin{aligned} I_5 &\leq C\||\nabla y(\cdot\,,0)|^2\|^2_{H^1_0} + C\|\Delta y(\cdot\,,0)\|^2_{H^1_0} \\ &\leq C\|\nabla y(\cdot\,,0)\|^2_{L^\infty}\|y(\cdot\,,0)\|^2_{H^2} + C\|y(\cdot\,,0)\|^2_{H_3} \\ &\leq C\|(y,v)\|^4_Y + C\|(y,v)\|^2_Y. \end{aligned} \tag{5.40}$$

Thus, we get from (5.32), (5.39) and (5.40) that $\mathcal{H} : Y \mapsto Z$ is well defined and

$$\|\mathcal{H}(y,v)\|^2_Z \leq C\|(y,v)\|^2_Y(1 + \|(y,v)\|^4_Y).$$

That $\mathcal{H}$ is continuous is easy to prove using similar arguments. □

Lemma 5.3.3. *The mapping $\mathcal{H} : Y \mapsto Z$ is continuously differentiable.*

Proof: The proof is very similar to the proof of Lemma 4.2 in [12]. Proofs of the same kind can also be found in [6, 10, 13]. We will only sketch the main ideas.

Let us first check that $\mathcal{H}$ is G-differentiable at any $(y,v) \in Y$ and let us compute the G-derivative $\mathcal{H}'(y,v)$. Thus, let us write $\mathcal{H}(y,v) = \mathcal{H}_1(y,v) + \mathcal{H}_2(y,v)$, with

$$\mathcal{H}_1(y,v) := (-\nabla\cdot(a(y)\nabla y), 0), \quad \mathcal{H}_2(y,v) := (y_t - v\tilde{1}_\omega, y(\cdot,0)).$$

Then, it is clear that both $\mathcal{H}_1$ and $\mathcal{H}_2$ are well defined and continuous. Furthermore, if $(y,v),(y'v') \in Y$ and $\sigma > 0$,

$$\begin{aligned} &\frac{1}{\sigma}[\mathcal{H}_1((y,v)+\sigma(y',v')) - \mathcal{H}_1(y,v)] \\ &\quad = \frac{1}{\sigma}\Big[\nabla\cdot\Big(a(y+\sigma y')\nabla(y+\sigma y')\Big) - \nabla\cdot\Big(a(y)\nabla y\Big)\Big] \\ &\quad = -a(y+\sigma y')\Delta y' - \frac{1}{\sigma}\big(a'(y+\sigma y')\nabla(y+\sigma y') - a'(y)\nabla y\big)\cdot\nabla y \\ &\qquad -\frac{1}{\sigma}\big(a(y+\sigma y') - a(y)\big)\Delta y. \end{aligned}$$

Let us introduce the linear mapping $D\mathcal{H}_1(y,v) \in \mathcal{L}(Y;Z)$, with

$$D\mathcal{H}_1(y,v)(y',v') := -2a'(y)\nabla y\cdot\nabla y' - a(y)\Delta y' - a''(y)y'|\nabla y|^2 - a'(y)y'\Delta y.$$

For any $(y',v') \in Y$, it is not difficult to see that

$$\frac{1}{\sigma}[\mathcal{H}_1((y,v)+\sigma(y',v')) - \mathcal{H}_1(y,v)] \to D\mathcal{H}_1(y,v)(y',v')$$

strongly in F, as $\sigma \to 0$. This shows that $\mathcal{H}_1$ is G-differentiable at (y,v), with $\mathcal{H}_1'(y,v) = D\mathcal{H}_1(y,v)$.

On the other hand, since $\mathcal{H}_2$ is linear continuous, it is also G-differentiable at any (y,v), with $\mathcal{H}_2'(y,v) = \mathcal{H}_2$.

Let us set $\mathcal{H}'(y,v) = D\mathcal{H}_1(y,v) + \mathcal{H}_2$ for all $(y,v) \in Y$. Then, $\mathcal{H}'(y,v)$ is the G-derivative of $\mathcal{H}$. Furthermore, since $a \in C^3(\mathbb{R})$ and possesses bounded derivatives, arguing as in [12], it can be

proved that the mapping $(y, v) \mapsto \mathcal{H}'(y, v)$ is continuous from Y to $\mathcal{L}(Y; Z)$. In other words, we can show that, $(y^n, v^n) \to (y, v)$ in Y, one has

$$\|(D\mathcal{H}(y^n, v^n) - D\mathcal{H}(y, v))(y', v')\|_Z \leq \varepsilon_n \|(y', v')\|_Y \quad \text{for some } \varepsilon_n \to 0.$$

The consequence is that $\mathcal{H}$ is not only G-differentiable, but also F-differentiable in the whole space Y and its F-derivative coincides with $\mathcal{H}'$ and is therefore continuous. □

Lemma 5.3.4. *Let $\mathcal{H}$ be the mapping defined by* (5.31)*. Then the bounded operator $\mathcal{H}'(0, 0) : Y \mapsto Z$ is onto.*

Proof: First, note that

$$\mathcal{H}'(0, 0)(y', v') = (y'_t - a(0)\Delta y' - v'\tilde{1}_\omega, y'(\cdot\,, 0)) \quad \forall (y', v') \in Y.$$

Consequently, the assertion is equivalent to prove that, for every $(h, y_0) \in Z$, there exists a state-control pair $(y, v) \in Y$ satisfying (5.3).

Nevertheless, the existence of a couple (y, v) with these properties is ensured by Theorem 5.2.3. Thus, $\mathcal{H}'(0, 0)$ is surjective and the lemma holds. □

Taking into account Lemmas 5.3.2, 5.3.3 and 5.3.4, we deduce that Theorem 5.3.1 can be applied in the context indicated at the beginning of the section, i.e. with Y, Z and $\mathcal{H}$ respectively given by (5.30a), (5.30c) and (5.31). Hence, Theorem 5.1.1 holds.

5.4. The convergence of ALG 1, approximations and experiments

As we already said in Section 5.1, arguing as in [6, 12], an elementary quasi-Newton algorithm can be introduced for the computation of a solution to the null control problem. To this purpose, we previously have to define an inverse $\mathcal{H}'(0, 0)^{-1}$ to the linear operator $\mathcal{H}'(0, 0)$. This can be done following the Fursikov-Imanuvilov method [13].

The argument is the following. For any $(h, y_0) \in Z$ a solution to (5.3) in Y can be obtained by solving the following extremal problem:

$$\begin{cases} \text{Minimize } \displaystyle\iint_Q \rho^2 |y|^2 \, dx\, dt + \iint_{\omega\times(0,T)} \rho_3^2 |v|^2 \, dx\, dt \\ \text{Subject to } v \in L^2(\omega \times (0, T)),\ (y, v) \text{ satisfies (5.3).} \end{cases} \tag{5.41}$$

It is known that (5.41) possesses exactly one solution, given by

$$y = \rho^{-2} L^* p, \quad v = -\rho_0^2 \, p|_{\omega\times(0,T)}\,, \tag{5.42}$$

where p is the unique solution to the Lax-Milgram problem

$$\begin{cases} \pi(p, p') = \displaystyle\iint_Q h p' \, dx\, dt + \int_\Omega y_0(x) p'(x, 0) \, dx \\ \forall p' \in P, \quad p \in P \end{cases} \tag{5.43}$$

(recall the notation introduced in the proof of Theorem 5.2.3). Accordingly, we set $\mathcal{H}'(0,0)^{-1}(h,y_0) = (y,v)$, with y and v respectively given by (5.42) and (5.43).

Obviously, (5.43) is the weak formulation of the following boundary-value problem, that is second-order in time and fourth-order in space:

$$\begin{cases} L(\rho^{-2}L^*p) + \rho_3^{-2}p1_\omega = h(x,t) & (x,t) \in Q, \\ p = 0, \quad \rho^{-2}L^*p = 0 & (x,t) \in \Sigma \\ (\rho^{-2}L^*p)|_{t=0} = y_0(x), \quad (\rho^{-2}L^*p)|_{t=T} = 0 & x \in \Omega. \end{cases}$$

Let us recall the algorithm proposed in Section 5.1:

ALG 1:

1. Choose $(y^0, v^0) \in Y$.
2. Then, for given $n \geq 0$ and $(y^n, v^n) \in Y$, compute

$$(y^{n+1}, v^{n+1}) = (y^n, v^n) - \mathcal{H}'(0,0)^{-1}(\mathcal{H}(y^n, v^n) - (0, y_0)). \tag{5.44}$$

For each n, the task reduces to the solution of a problem of the kind (5.43) with

$$h = -\nabla \cdot ((a(0) - a(y^n))\nabla y^n)). \tag{5.45}$$

The convergence of **ALG 1** is established in Theorem 5.1.2. Let us give the proof.

Proof of Theorem 5.1.2: First, note that for any initial $(y^0, v^0) \in Y$, the iterates in **ALG 1** are well defined. We will use a standard argument, appropriate for methods of this kind, that rely on the C^1 regularity of $\mathcal{H}$; see for instance [2].

Thus, let us assume that $\|y_0\|_{H^3} \leq \varepsilon$ (ε is furnished by Theorem 5.1.1) and $(y,v) \in Y$ satisfies $\mathcal{H}(y,v) = (0,y_0)$; let us set $C_H := \|\mathcal{H}'(0,0)^{-1}\|_{\mathcal{L}(Z;Y)}$ and let us assume that $0 < \overline{\theta} < 1/(2C_H)$. Since $\mathcal{H}$ is continuously differentiable, there exists $\delta > 0$ such that

$$(\tilde{y}, \tilde{v}) \in Y, \quad \|(\tilde{y}, \tilde{v})\|_Y \leq \delta \Rightarrow \|\mathcal{H}'(\tilde{y}, \tilde{v}) - \mathcal{H}'(y,v)\|_{\mathcal{L}(Y;Z)} \leq \overline{\theta}.$$

We will assume that $\|(y,v)\|_Y \leq \delta$ and we will prove that there exists $\kappa > 0$ such that, if $(y^0, v^0) \in Y$ and

$$\|(y^0, v^0) - (y,v)\|_Y \leq \kappa, \tag{5.46}$$

then the (y^n, v^n) satisfy (5.7).

Let κ be such that, if

$$(\tilde{y}, \tilde{v}) \in Y, \quad \|(\tilde{y}, \tilde{v}) - (y,v)\|_Y \leq \kappa,$$

then

$$\|\mathcal{H}(\tilde{y}, \tilde{v}) - \mathcal{H}(y,v) - \mathcal{H}'(y,v)((\tilde{y}, \tilde{v}) - (y,v))\|_Z \leq \overline{\theta}\|(\tilde{y}, \tilde{v}) - (y,v)\|_Y$$

and let us choose $(y^0, v^0) \in Y$ satisfying (5.46). Then, if we introduce $e^n := (y^n, v^n) - (y,v)$, the following holds:

$$\begin{aligned} e^{n+1} = e^n &- \mathcal{H}'(0,0)^{-1}(\mathcal{H}(y^n, v^n) - \mathcal{H}(y,v)) \\ &- \mathcal{H}'(0,0)^{-1}(\mathcal{H}(y^n, v^n) - \mathcal{H}(y,v) - \mathcal{H}'(y,v)e^n) \\ &- \mathcal{H}'(0,0)^{-1}(\mathcal{H}'(y,v) - \mathcal{H}'(0,0))e^n. \end{aligned}$$

Therefore,

$$\begin{aligned}\|e^{n+1}\|_Y \le C_H\|\mathcal{H}(y^n,v^n)-\mathcal{H}(y,v)-\mathcal{H}'(y,v)e^n\|_Z\\ + C_H\|\mathcal{H}'(y,v)-\mathcal{H}'(0,0)\|_{\mathcal{L}(Y;Z)}\|e^n\|_Y.\end{aligned}$$

Since $\|e^0\|_Y \le \kappa$, this inequality for $n=0$ yields

$$\|e^1\|_Y \le 2C_H\overline{\theta}\|e^0\|_Y$$

and, in particular, we also have $\|e^1\|_Y \le \kappa$. By induction, we then see that

$$\|e^n\|_Y \le 2C_H\overline{\theta}\|e^{n-1}\|_Y \le \cdots \le (2C_H\overline{\theta})^n\|e^0\|_Y$$

for all $n \ge 1$. This proves that $e^n \to 0$ in Y and (5.7) holds with $\theta = 2C_H\overline{\theta}$. □

In order to solve numerically the problems (5.43), it suffices in principle to construct explicity finite dimensional spaces $P_h \subset P$. Note however that this is possible but needs a considerable work. This is because the functions in P must satisfy $L^*p \in L^2_{loc}(Q)$. Thus, an approximation based on a standard triangulation of Q requires spaces P_h of functions that must be globally C^0 in all the variables and globally C^1 in space. This construction can be complex and too expensive and it is convenient to use instead a mixed formulation, as in [11, 12].

Let us briefly indicate how this can be done. Let us introduce the new variables $z = \rho^{-1}L^*p$ and $m = \rho_3^{-1}p$, the spaces $Z := \{(z,m) \in L^2(Q)\times L^2(Q) : (\rho_3 m)_t \in L^2(Q),\ \nabla(\rho_3 m) \in L^2(Q)^N\}$ and $\Lambda := \{\lambda : \rho\lambda \in L^2(Q),\ \nabla(\rho\lambda) \in L^2(Q)\}$, the bilinear forms

$$\alpha((z,m),(z',m')) := \iint_Q zz'\,dx\,dt + \iint_{\omega\times(0,T)} mm'\,dx\,dt,$$

$$\beta((z,m),\lambda) := \iint_Q \Big[\lambda\Big(z+\rho^{-1}((\rho_3 m)_t)\Big) - \nabla(\rho^{-1}\lambda)\cdot\nabla(\rho_3 m)\Big]\,dx\,dt$$

and the linear form

$$\langle \ell,(z,m)\rangle := \iint_Q \rho hm\,dx\,dt + \int_\Omega \rho_3(x,0)y_0(x)m(x,0)\,dx.$$

It is not difficult to check that $\alpha(\cdot\,,\cdot)$, $\beta(\cdot\,,\cdot)$ and ℓ are well defined and continuous, respectively, on $Z\times Z$, $Z\times\Lambda$ and Z.

Then, an appropriate mixed formulation of (5.43) is the following:

Find $(z,m) \in Z$ *and* $\lambda \in \Lambda$ *such that*

$$\begin{cases} \alpha((z,m),(z',m')) + \beta((z',m'),\lambda) = \langle \ell,(z',m')\rangle \quad \forall (z',m') \in Z,\\ \beta((z,m),\lambda') = 0 \quad \forall \lambda' \in \Lambda. \end{cases} \tag{5.47}$$

What we have to do is, consequently, to solve numerically (5.47) and then take

$$y = \rho^{-1}z, \quad v = -\rho_3^{-1}\,m|_{\omega\times(0,T)}\,.$$

Contrarily to P, it is not difficult to construct finite dimensional subspaces of Z and Λ. This leads to "natural" mixed approximations and allows efficient and computationally reasonable computations.

Thus, let $\mathcal{T}_h$ be a triangulation of the cylinder $Q = \Omega \times (0,T)$, and let us set

$$\begin{aligned} Z_h := \{(z_h, m_h) \in C^0(\overline{Q}) \times C^0(\overline{Q}) :\ & z_h|_K, m_h|_K \in \mathbb{P}_1(K)\ \ \forall K \in \mathcal{T}_h, \\ & z_h|_\Sigma = 0,\ m_h|_\Sigma = 0,\ \ z_h(\cdot,t) = m_h(\cdot\,,t) = 0 \text{ for } t \in [T-h,T]\} \end{aligned}$$

and

$$\Lambda_h := \{\lambda_h \in C^0(\overline{Q}) : \lambda_h|_K \in \mathbb{P}_1(K)\ \ \forall K \in \mathcal{T}_h,\ \lambda_h(\cdot\,,t) = 0 \text{ for } t \in [T-h,T]\}.$$

Then, we can approximate (5.47) as follows:

Find $(z_h, m_h) \in Z_h$ *and* $\lambda_h \in \Lambda_h$ *such that*

$$\begin{cases} \alpha((z_h,m_h),(z_h',m_h')) + \beta((z_h',m_h'),\lambda_h) = \langle \ell, (z_h',m_h')\rangle \quad \forall (z_h',m_h') \in Z_h, \\ \beta((z_h,m_h),\lambda_h') = 0 \quad \forall \lambda_h' \in \Lambda_h. \end{cases} \tag{5.48}$$

In the following sections, we present the results of some experiments.

5.4.1. A first test (Test 1)

The quasi-Newton method has been applied to the solution to the null controllability problem for (5.1) with the following data:

- $N = 2, \Omega = (0,1)\times(0,1), \omega = (0.2,0.8)\times(0.2,0.8), T = 0.5.$
- $y_0(x_1,x_2) = \sin(\pi x_1)\sin(\pi x_2).$
- $a(s) = \exp(-2\exp(-0.3s)).$

Note that, this choice of a has a sense. The function $s \mapsto \exp(-2\exp(-0.3s))$ can be viewed as the density profile of an isolated population under some particular circumstances. Thus, we can motivate our tests by the control of the spread of a desease in a related habitat, with diffusion depending on the number of individuals.

At each step of **ALG 1**, the problem has been rewritten in the form (5.47). Then, the finite element approximation (5.48) has been introduced. The computations have been performed with the FreeFem++ package; for a detailed description, see `http://www.freefem.org//ff++` . In this and the other tests, the stopping criterion has been

$$\frac{\|y^{n+1} - y^n\|_{L^2}}{\|y^{n+1}\|_{L^2}} \leq \kappa$$

where $y^n = \rho^{-1} z^n$, z^n is (together with some m^n and λ^n) the solution to (5.48) and $\kappa = 10^{-5}$.

The mesh is displayed in Fig. 5.1.

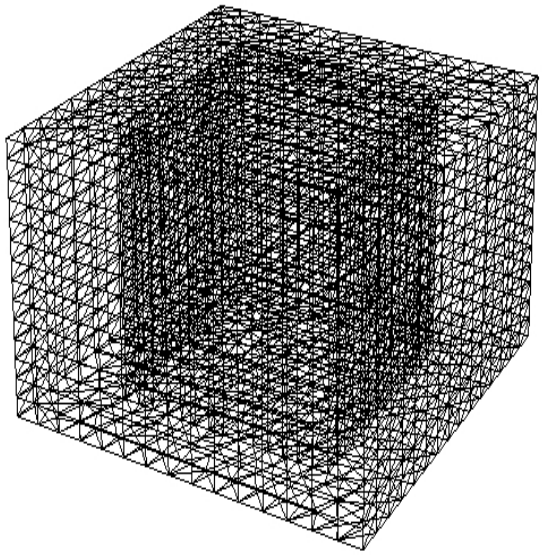


Figure 5.1: The mesh. Number of vertices: 7425. Number of tetrahedrons: 38976.

Starting from $(y^0, v^0) = \mathcal{H}'(0,0)^{-1}(0, y_0)$, convergence was reached after 11 iterates, see Fig. 5.2. The initial state can be viewed in Fig. 5.3. The computed control and the associated state are depicted in Figs. 5.4–5.7.

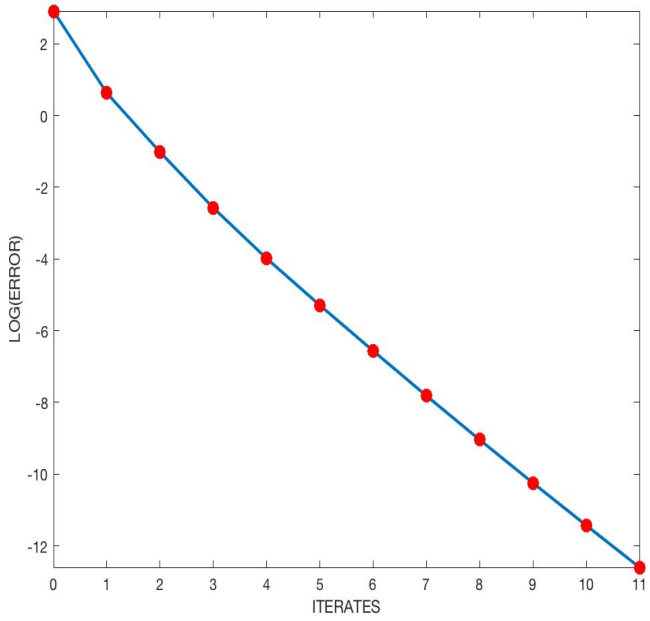


Figure 5.2: Evolution of the error at logarithmic scale for Test 1.

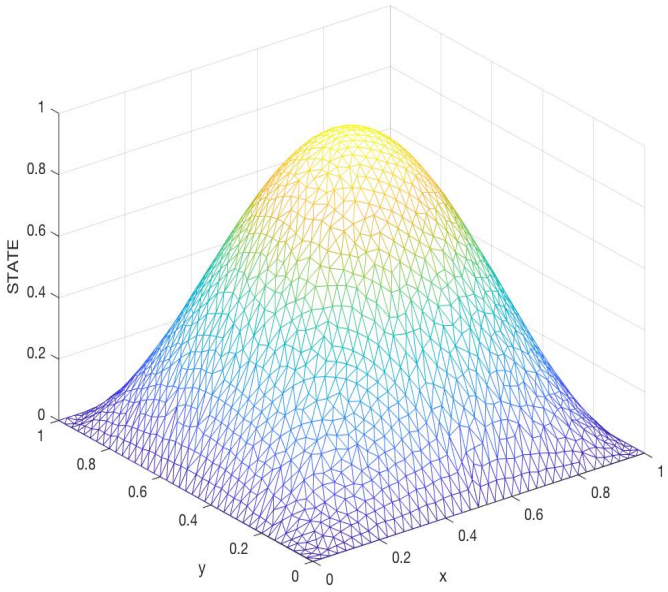
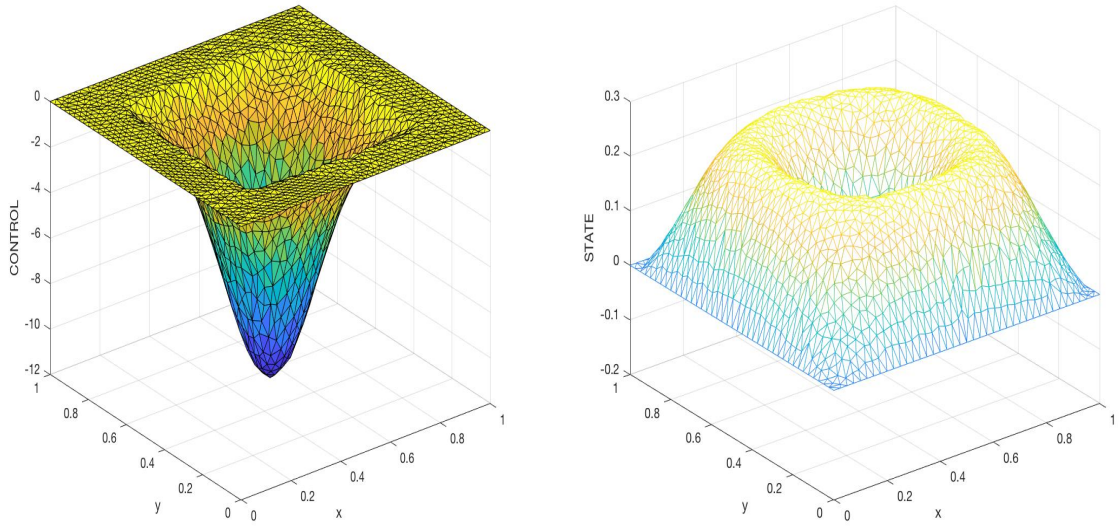
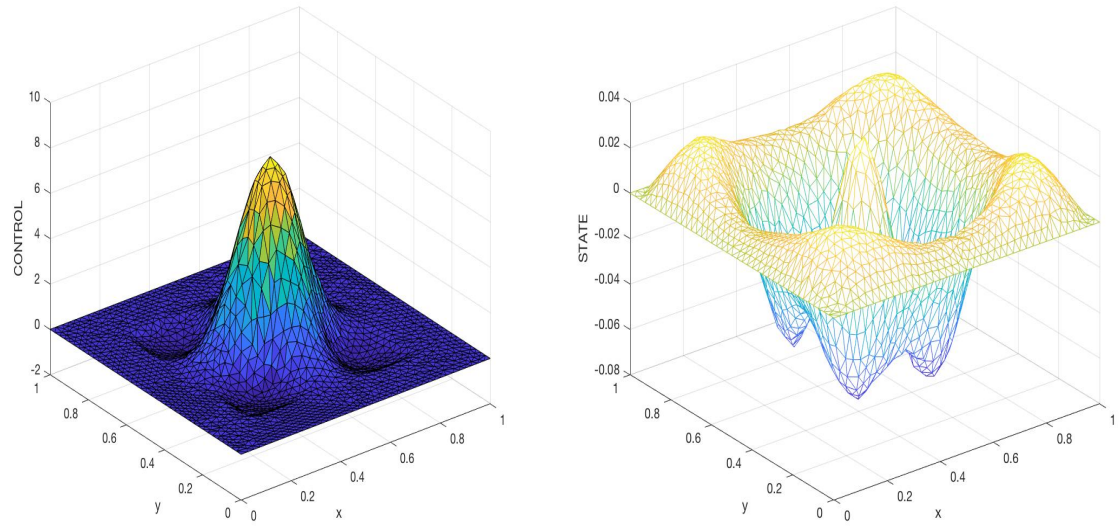


Figure 5.3: The initial state $y_0 = \sin(\pi x_1) \cdot \sin(\pi x_2)$.



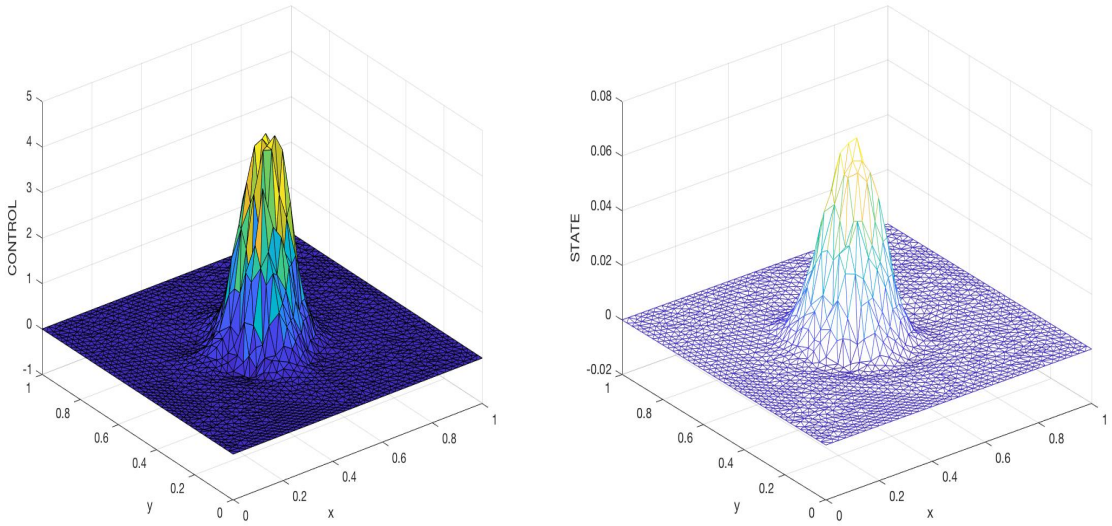
(a) Computed control (b) Associated state

Figure 5.4: The computed control and the associated state of Test 1 at $t = 0.1$. Newton method.



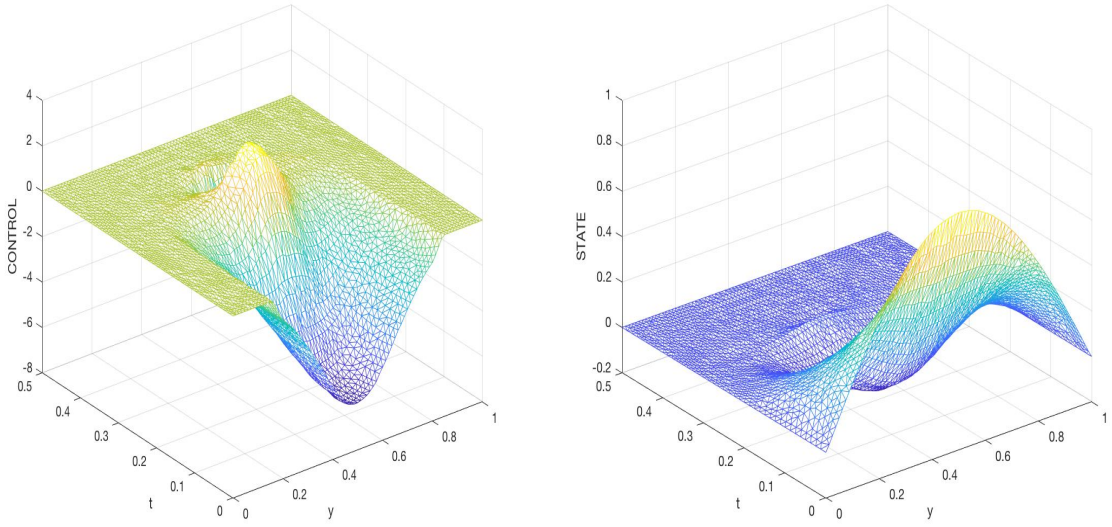
(a) Computed control (b) Associated state

Figure 5.5: The computed control and the associated state of Test 1 at $t = 0.25$.



(a) Computed control (b) Associated state

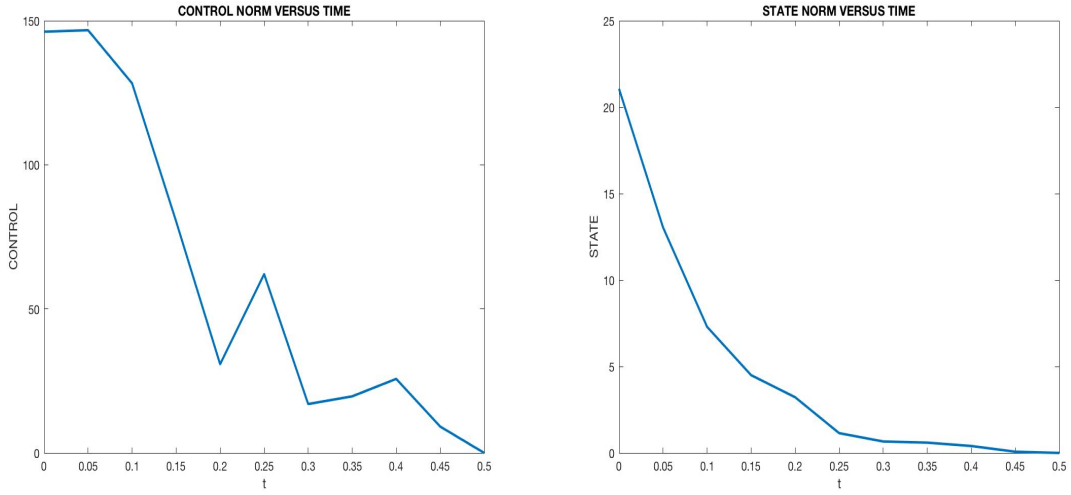
Figure 5.6: The computed control and the associated state of Test 1 at $t = 0.4$.



(a) Computed control (b) Associated state

Figure 5.7: The computed control and the associated state of Test 1 at $x_1 = 0.68$.

The L^2 norms of the computed control and the state as functions of t are displayed in Fig. 5.8.



(a) Computed control (b) Associated state

Figure 5.8: Evolution in time of the L^2 norms of the control and the state for Test 1.

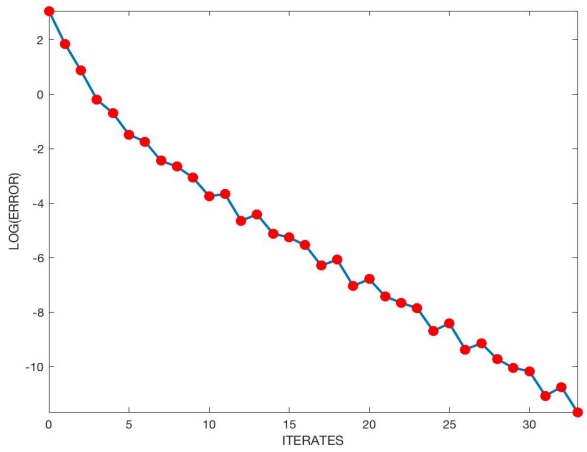


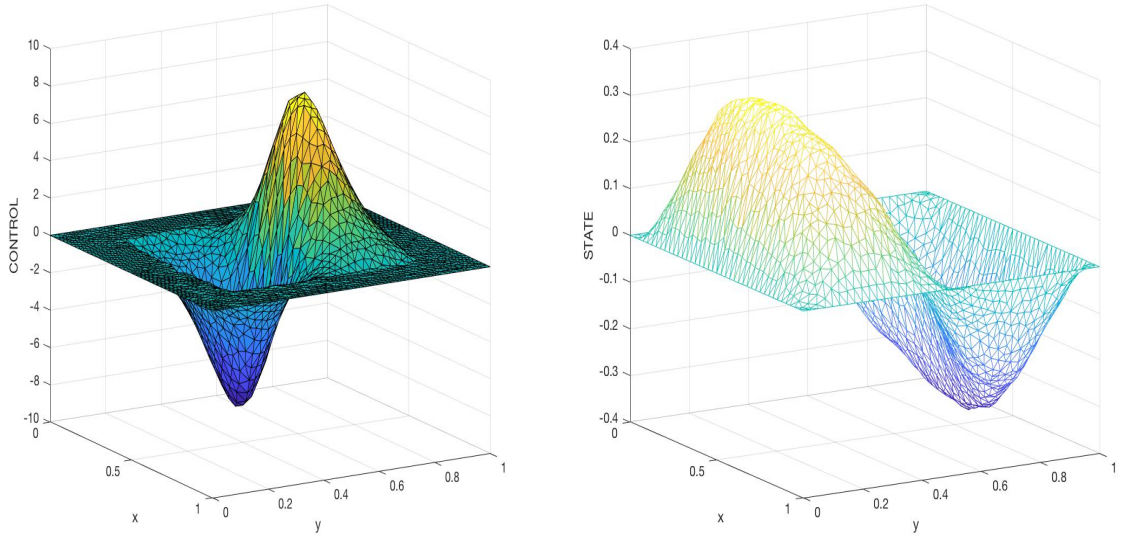
Figure 5.9: Evolution of the error at logarithmic scale for Test 2.

5.4.2. A second test (Test 2)

In a second experiment, we have taken the same Ω, ω and T and we have fixed

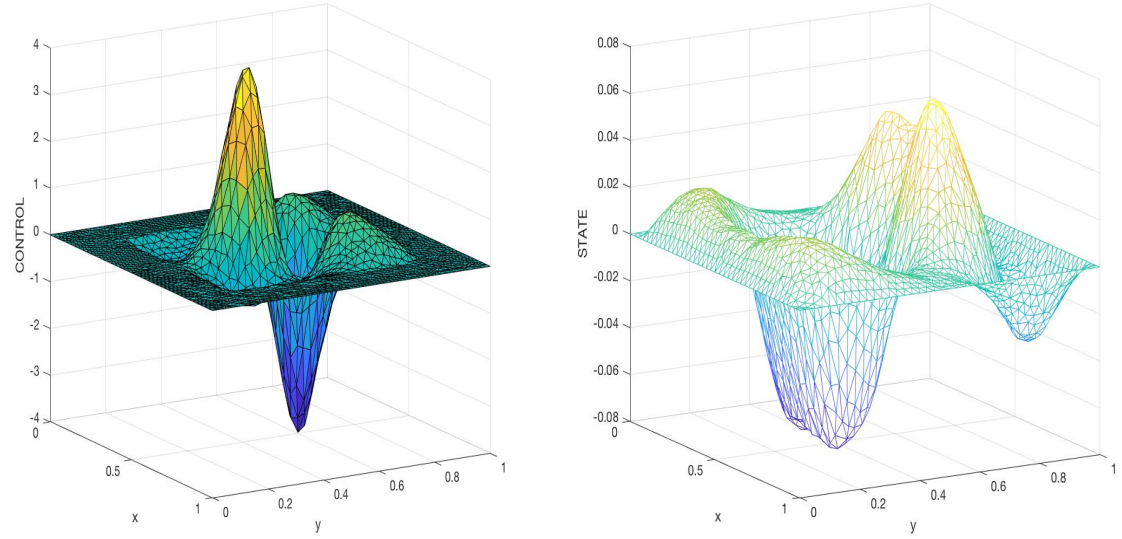
- $y_0(x, y) = \sin(\pi x_1)\sin(2\pi x_2)$.
- $a(s) = a_0(1 + 5 \cdot \sin(50s))$, with $a_0 = e^{-2}$.

This time, starting from $(y^0, v^0) = \mathcal{H}'(0,0)^{-1}(0, y_0)$, convergence was reached after 33 iterates (see Fig. 5.9). The initial state and the computed control and associated state are depicted in Figs. 5.10–5.13.



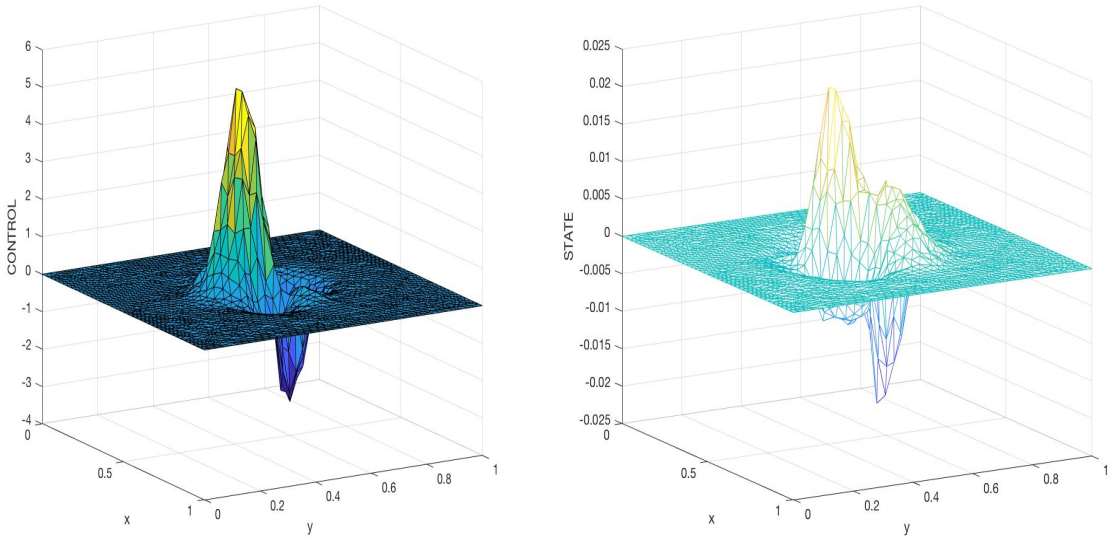
(a) Computed control (b) Associated state

Figure 5.10: The computed control and the associated state of Test 2 at $t = 0.1$.



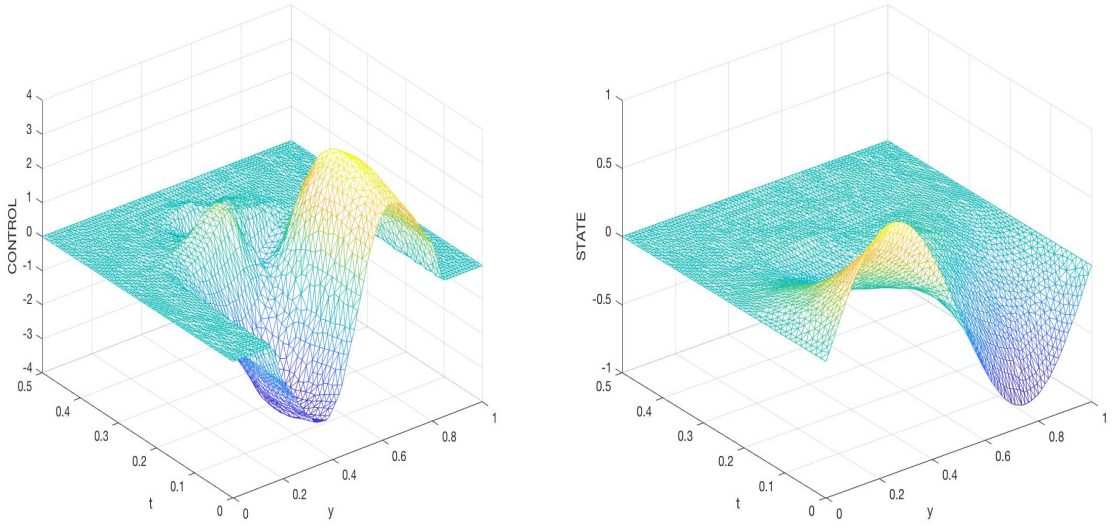
(a) Computed control (b) Associated state

Figure 5.11: The computed control and the associated state of Test 2 at $t = 0.25$.



(a) Computed control (b) Associated state

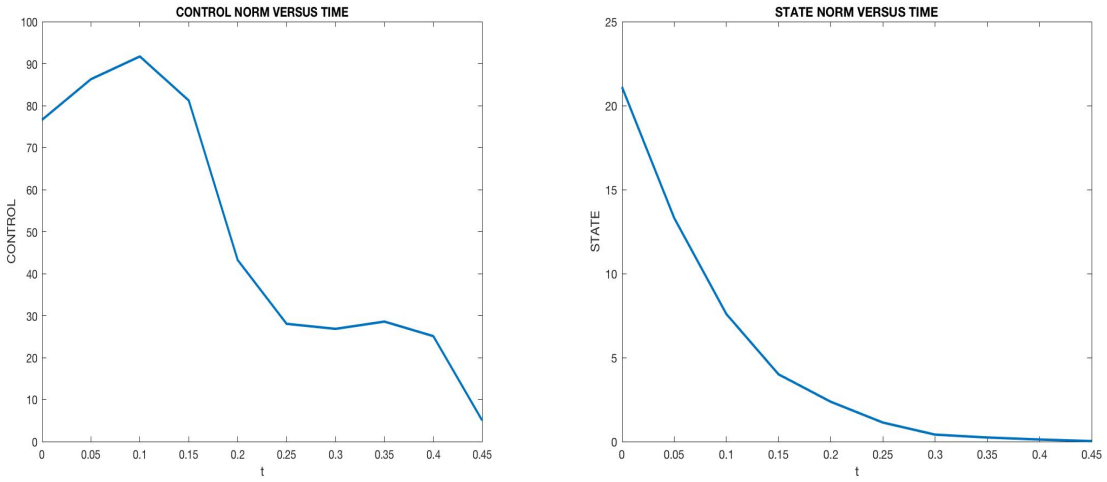
Figure 5.12: The computed control and the associated state of Test 2 at $t = 0.4$.



(a) Computed control (b) Associated state

Figure 5.13: The computed control and the associated state of Test 2 at $x_1 = 0.68$.

Finally, the evolution in time of the L^2 norms of the control and the state is depicted in Fig. 5.14.



(a) Computed control

(b) Associated state

Figure 5.14: Evolution in time of the L^2 norms of the control and the state for Test 2.

The results of these numerical experiments corroborate the analysis performed at the beginning of Section 5.4. Thus, we conclude that the local null controllability problem for (5.1) can be solved theoretically and numerically. An interesting related open question is whether global controllability can be established under appropriate assumptions on a.

Bibliography

[1] Alekseev V.M., Tikhomorov V.M., Formin S.V., *Optimal Control,* Consultants Bureau, New York, 1987.

[2] Argyros I.K., *Convergence and applications of Newton-type iterations,* Springer, New York, 2008.

[3] Boyer F., Hubert F., Le Rousseau J.,*Discrete Carleman estimates for elliptic operators in arbitrary dimension and applications,* SIAM J. Control Optim., 48, 5357–5397, 2010.

[4] Boyer F.,*On the penalised HUM approach and its applications to the numerical approximation of null controls for parabolic problems,* CANUM 2012, Super-Besse, ESAIM Proceedings, EDP Sciences, Les Ulis, 2013.

[5] Carthel C., Glowinski R., Lions J.-L.,*On exact and approximate boundary controllability for the heat equation: a numerical approach,* J. Optim. Theory Appl., 82(3), 429–484, 1994.

[6] Clark H.R., Fernández-Cara E., Límaco J., Medeiros L.A., *Theoretical and numerical local null controllability for a parabolic system with local and nonlocal nonlinearities,* Applied Mathematics and Computation, 223, 483–505, 2013.

[7] Fabre C., Puel J.-P., Zuazua E., *Approximate controllability of the semilinear heat equation,* Proc. R. Soc. Edinburgh 125A, 1995.

[8] Fernández-Cara E., Guerrero S., *Global Carleman inequalities for parabolic systems and applications to controllability,* SIAM J. Control Optim., 45(4), 1395–1446, 2006.

[9] Fernández-Cara E., Límaco J., Marín-Gayte I., *Null controllability of a non-linear parabolic equation*, submited.

[10] Fernández-Cara E., Lu Q., Zuazua E.,*Null controllability of linear heat and wave equations with nonlocal spatial terms,* SIAM J. Control Optim., 54(4), 2009–2019, 2016.

[11] Fernández-Cara E., Münch A., Souza D.A., *On the numerical Controllability of the two-dimensional heat, Stokes and Navier-Stokes equations,* J. Sci Comput., 70, 78–85, 2017.

[12] Fernández-Cara E., Nina-Huamán D., Núñez-Chávez M.R., Vieira F.B., *On the theoretical and numerical control of a 1D nonlinear parabolic PDE,* J. Optim. Theory Appl., 175(3), 652–682, 2017.

[13] Fursikov A.V., Imanuvilov O.Yu., *Controllability of Evolution Equations,* Lecture Notes Series, Seoul National University, Research Institute of Mathematics, Global Analysis Research Center, Seoul, 34,1996.

[14] Imanuvilov O.Yu.,*Remarks on exact controllability for the Navier-Stokes equations,* ESAIM Control Optim. Calc. Var., 6, 39–72, 2001.

[15] Labbé S., Trélat E., *Uniform controllability of semi-discrete approximations of parabolic control systems,* Syst. Control Lett., 55, 597–609, 2006.

[16] Zuazua E.,*Controllability and observability of partial differential equations: some results and open problems. Handbook of differential equations: evolutionary equations,* Hand. Differ. Equ./North-Holland, Amsterdam, 3, 527–621, 2007.

Capítulo 6

Approximation of null controls for semilinear heat equations using a least-squares approach

The null distributed controllability of the semilinear heat equation $y_t - \Delta y + g(y) = f\,1_\omega$, assuming that g satisfies the growth condition $g(s)/(|s|\log^{3/2}(1+|s|)) \to 0$ as $|s| \to \infty$ and that $g' \in L^\infty_{loc}(\mathbb{R})$ has been obtained by Fernández-Cara and Zuazua in 2000. The proof based on a fixed point argument makes use of precise estimates of the observability constant for a linearized heat equation. It does not provide however an explicit construction of a null control. Assuming that $g' \in W^{s,\infty}(\mathbb{R})$ for one $s \in (0,1]$, we construct an explicit sequence converging strongly to a null control for the solution of the semilinear equation. The method, based on a least-squares approach, generalizes Newton type methods and guarantees the convergence whatever be the initial element of the sequence. In particular, after a finite number of iterations, the convergence is super linear with a rate equal to $1+s$. Numerical experiments in the one dimensional setting illustrate our analysis. This chapter is based on the paper [17], in collaboration with A. Münch and J. Lémoine.

6.1. Introduction

Let $\Omega \subset \mathbb{R}^d$, $1 \leq d \leq 3$, be a bounded connected open set whose boundary $\partial\Omega$ is Lipschitz. Let ω be any non-empty open set of Ω and let $T > 0$. We note $Q_T = \Omega \times (0,T)$, $q_T = \omega \times (0,T)$ and $\Sigma_T = \partial\Omega \times (0,T)$. We are concerned with the null controllability problem for the following semilinear heat equation

$$\begin{cases} y_t - \Delta y + g(y) = f1_\omega \quad \text{in} \quad Q_T, \\ y = 0 \text{ on } \Sigma_T, \quad y(\cdot,0) = u_0 \text{ in } \Omega, \end{cases} \tag{6.1}$$

where $u_0 \in L^2(\Omega)$ is the initial state of y and $f \in L^2(q_T)$ is a *control* function. We assume moreover that the nonlinear function $g : \mathbb{R} \mapsto \mathbb{R}$ is, at least, locally Lipschitz-continuous. Following [13], we will also assume for simplicity that g satisfies

$$|g'(s)| \leq C(1+|s|^m) \quad \text{a.e., with } 1 \leq m \leq 1 + 4/d. \tag{6.2}$$

Under this condition, (6.1) possesses exactly one local in time solution. Moreover, under the growth condition

$$|g(s)| \leq C(1 + |s| \log(1 + |s|)) \quad \forall s \in \mathbb{R}, \tag{6.3}$$

the solutions to (6.1) are globally defined in $[0, T]$ and one has

$$y \in C^0([0, T]; L^2(\Omega)) \cap L^2(0, T; H_0^1(\Omega)), \tag{6.4}$$

see [3]. Recall that, without a growth condition of the kind (6.3), the solutions to (6.1) can blow up before $t = T$; in general, the blow-up time depends on g and the size of $\|u_0\|_{L^2(\Omega)}$.

The system (6.1) is said to be *controllable* at time T if, for any $u_0 \in L^2(\Omega)$ and any globally defined bounded trajectory $y^\star \in C^0([0, T]; L^2(\Omega))$ (corresponding to the data $u_0^\star \in L^2(\Omega)$ and $f^\star \in L^2(q_T)$), there exist controls $f \in L^2(q_T)$ and associated states y that are again globally defined in $[0, T]$ and satisfy (6.4) and

$$y(x, T) = y^\star(x, T), \quad x \in \Omega. \tag{6.5}$$

We refer to [5] for an overview of control problems in nonlinear situations. The uniform controllability strongly depends on the nonlinearity g. Fernández-Cara and Zuazua proved in [13] that if g is too "super-linear" at infinity, then, for some initial data, the control cannot compensate the blow-up phenomenon occurring in $\Omega \backslash \overline{\omega}$:

Theorem 6.1.1 ([13]). *There exist locally Lipschitz-continuous functions g with $g(0) = 0$ and*

$$|g(s)| \sim |s| \log^p(1 + |s|) \quad \textit{as} \quad |s| \to \infty, \quad p > 2,$$

such that (6.1) *fails to be controllable for all $T > 0$.*

On the other hand, Fernández-Cara and Zuazua also proved that if p is small enough, then the controllability holds true uniformly.

Theorem 6.1.2 ([13]). *Let $T > 0$ be given. Assume that* (6.1) *admits at least one solution $y^\star$, globally defined in $[0, T]$ and bounded in Q_T. Assume that $g : \mathbb{R} \mapsto \mathbb{R}$ is locally Lipschitz-continuous and satisfies* (6.2) *and*

$$\frac{g(s)}{|s| \log^{3/2}(1 + |s|)} \to 0 \quad \textit{as} \quad |s| \to \infty. \tag{6.6}$$

Then (6.1) *is controllable at time T.*

Therefore, if $|g(s)|$ does not grow at infinity faster than $|s| \log^p(1 + |s|)$ for any $p < 3/2$, then (6.1) is controllable. This result extends [9] obtaining the uniform controllability for any $p < 1$. We also mention [1] which gives the same result assuming additional sign condition on g, namely $g(s)s \geq -C(1 + s^2)$ for all $s \in \mathbb{R}$ and some $C > 0$. The problem remains open when g behaves at infinity like $|s| \log^p(1 + |s|)$ with $3/2 \leq p \leq 2$. We mention however the recent work of Le Balc'h [16] where uniform controllability results are obtained for $p \leq 2$ assuming additional sign conditions on g, notably that $g(s) > 0$ for $s > 0$ or $g(s) < 0$ for $s < 0$. This condition is not satisfied for $g(s) = -s \log^p(1 + |s|)$. Let us also mention [6] in the context of Theorem 6.1.1

where a positive boundary controllability result is proved for a specific class of initial and final data and T large enough.

In the sequel, for simplicity, we shall assume that $g(0) = 0$ and that $f^\star \equiv 0, u_0^\star \equiv 0$ so that $y^\star$ is the null trajectory. The proof given in [13] is based on a fixed point method. Precisely, it is shown that the operator $\Lambda : L^\infty(Q_T) \to L^\infty(Q_T)$, where $y_z := \Lambda z$ is a null controlled solution of the linear boundary value problem

$$\begin{cases} y_{z,t} - \Delta y_z + y_z\, \tilde{g}(z) = f_z 1_\omega \quad \text{in} \quad Q_T \\ y_z = 0 \text{ on } \Sigma_T, \quad y_z(\cdot,0) = u_0 \quad \text{in} \quad \Omega \end{cases}, \qquad \tilde{g}(s) := \begin{cases} g(s)/s & s \neq 0, \\ g'(0) & s = 0, \end{cases} \tag{6.7}$$

maps a closed ball $B(0,M) \subset L^\infty(Q_T)$ into itself, for some $M > 0$. The Kakutani's theorem then provides the existence of at least one fixed point for the operator Λ, which is also a controlled solution for (6.1).

The main goal of this work is to determine an approximation of the controllability problem associated to (6.1), that is to construct an explicit sequence $(f_k)_{k\in\mathbb{N}}$ converging strongly toward a null control for (6.1). A natural strategy is to take advantage of the method used in [13, 16] and consider the Picard iterates associated with the operator Λ: $y_{k+1} = \Lambda(y_k)$, $k \geq 0$ initialized with any element $y_0 \in B(0,M)$. The sequence of controls is then $(f_k)_{k\in\mathbb{N}}$ so that $f_k \in L^2(q_T)$ is a null control for y_k solution of

$$\begin{cases} y_{k,t} - \Delta y_k + y_k\, \tilde{g}(y_{k-1}) = f_k 1_\omega \quad \text{in} \quad Q_T, \\ y_k = 0 \text{ on } \Sigma_T, \quad y_k(\cdot,0) = u_0 \text{ in } \Omega. \end{cases} \tag{6.8}$$

Numerical experiments for $d = 1$ reported in [11] exhibit the non convergence of the sequences $(y_k)_{k\in\mathbb{N}}$ and $(f_k)_{k\in\mathbb{N}}$ for some initial conditions large enough. This phenomenon is related to the fact that the operator Λ is *a priori* not contractant. We also refer to [2] where this strategy is implemented. Still in the one dimensional case, a least-squares type approach, based on the minimization over $L^2(Q_T)$ of the functional $R : L^2(Q_T) \to \mathbb{R}^+$ defined by $R(z) := \|z - \Lambda(z)\|_{L^2(Q_T)}$ is introduced and analyzed in [11]. Assuming that $\tilde{g} \in C^1(\mathbb{R})$ and $g' \in L^\infty(\mathbb{R})$, it is proved first that $R \in C^1(L^2(Q_T);\mathbb{R}^+)$ and secondly that, if $\|u_0\|_{L^\infty(\Omega)}$ is small enough, then any critical point for R is a fixed point for Λ. Under this smallness assumption on the data, numerical experiments reported in [11] display the convergence of minimizing sequences for R (based on a gradient method) and a better behavior than the Picard iterates. The analysis of convergence is however not performed. As is usual for nonlinear problems and considered in [11], we may also employ a Newton type method to find a zero of the mapping $\widetilde{F} : Y \mapsto W$ defined by

$$\widetilde{F}(y,f) = (y_t - \Delta y + g(y) - f1_\omega, y(\cdot\,,0) - u_0, y(\cdot,T)) \quad \forall (y,f) \in Y \tag{6.9}$$

for some appropriate Hilbert spaces Y and W (see below). It is shown for $d = 1$ in [11] that, if $g \in C^1(\mathbb{R})$ and $g' \in L^\infty(\mathbb{R})$, then $\widetilde{F} \in C^1(Y;W)$ allowing to derive the Newton iterative sequence: given (y_0, f_0) in Y, define the sequence $(y_k, f_k)_{k\in\mathbb{N}}$ iteratively as follows $(y_{k+1}, f_{k+1}) = (y_k, f_k) - (Y_k, F_k)$ where F_k is a control for Y_k solution of

$$\begin{cases} Y_{k,t} - \Delta Y_k + g'(y_k)\, Y_k = F_k\, 1_\omega + y_{k,t} - \Delta y_k + g(y_k) - f_k 1_\omega, & \text{in} \quad Q_T, \\ Y_k = 0, & \text{on} \quad \Sigma_T, \\ Y_k(\cdot,0) = u_0 - y_k(\cdot,0), & \text{in} \quad \Omega \end{cases} \tag{6.10}$$

such that $Y_k(\cdot,T) = -y_k(\cdot,T)$ in Ω. Once again, numerical experiments for $d = 1$ in [11] exhibits the lack of convergence of the Newton method for large enough initial condition, for which the solution y is not close enough to the zero trajectory. As far as we know, the construction of a convergent approximation $(f_k)_{k\in\mathbb{N}}$ in the general case where the initial data to be controlled is arbitrary in $L^2(\Omega)$ remains an open issue. Still assuming that $g' \in L^\infty(\mathbb{R})$ and in addition that there exists one s in $(0,1]$ such that $\sup_{a,b\in\mathbb{R},a\neq b} \frac{|g'(a)-g'(b)|}{|a-b|^s} < \infty$, we construct, for any initial data $u_0 \in L^2(\Omega)$, a strongly convergent sequence $(f_k)_{k\in\mathbb{N}}$ toward a control for (6.1). Moreover, after a finite number of iterates related to the norm $\|g'\|_{L^\infty(\mathbb{R})}$, the convergence is super linear with a rate equal to $1+s$. This is done (following and improving [21] devoted to a linear case) by introducing a quadratic functional which measures how a pair $(y,f) \in Y$ is close to a controlled solution for (6.1) and then by determining a particular minimizing sequence enjoying the announced property. A natural example of so-called error (or least-squares) functional is given by $\widetilde{E}(y,f) := \frac{1}{2}\|\widetilde{F}(y,f)\|_W^2$ to be minimized over Y. In view of controllability results for (6.1), the non-negative functional $\widetilde{E}$ achieves its global minimum equal to zero for any control pair $(y,f) \in Y$ of (6.1).

The paper is organized as follows. In Section 6.2, we first derive a controllability result for a linearized heat equation with potential in $L^\infty(Q_T)$ and source term in $L^2(0,T;H^{-1}(\Omega))$. Then, in Section 6.3, we define the least-squares functional E and the corresponding optimization problem (6.26) over the Hilbert space $\mathcal{A}$. We show that E is Gateaux-differentiable over $\mathcal{A}$ and that any critical point (y,f) for E for which $g'(y)$ belongs to $L^\infty(Q_T)$ is also a zero of E (see Proposition 6.3.2). This is done by introducing a descent direction (Y^1,F^1) for $E(y,f)$ for which $E'(y,f)\cdot(Y^1,F^1)$ is proportional to $E(y,f)$. Then, assuming that the nonlinear function g is such that g' belongs to $W^{s,\infty}(\mathbb{R})$ for one s in $(0,1]$, we determine a minimizing sequence based on (Y^1,F^1) which converges strongly to a controlled pair for the semilinear heat equation (6.1). Moreover, we prove that after a finite number of iterates, the convergence enjoys a rate equal to $1+s$ (see Theorem 6.3.3 for $s=1$ and Theorem 6.3.4 for $s \in (0,1)$). We also emphasize that this least-squares approach coincides with the damped Newton method one may use to find a zero of a mapping similar to $\widetilde{F}$ mentioned above; we refer to Remark 8. This explains the convergence of our approach with a super-linear rate. Section 6.4 gives some numerical illustrations of our result in the one dimensional case and a nonlinear function g for which $g' \in W^{1,\infty}(\mathbb{R})$. We conclude in Section 6.5 with some perspectives. As far as we know, the analysis of convergence presented in this work, though some restrictive hypotheses on the nonlinear function g, is the first one in the context of controllability for partial differential equations.

Along the text, we shall denote by $\|\cdot\|_\infty$ the usual norm in $L^\infty(\mathbb{R})$, $(\cdot,\cdot)_X$ the scalar product of X (if X is a Hilbert space) and by $\langle\cdot,\cdot\rangle_{X,Y}$ the duality product between the spaces X and Y.

6.2. A controllability result for a linearized heat equation with $L^2(H^{-1})$ right hand side

We give in this section a controllability result for a linear heat equation with potential in $L^\infty(Q_T)$ and right hand side in $L^2(0,T;H^{-1}(\Omega))$. As this work concerns the null controllability of parabolic equation, we shall make use of Carleman type weights introduced in this context notably in [14] (we also refer to [10] for a review). Here, we assume that such weights ρ, ρ_0, ρ_1

and ρ_2 blow up as $t \to T^-$ and satisfy:

$$\begin{cases} \rho = \rho(x,t),\, \rho_0 = \rho_0(x,t),\, \rho_1 = \rho_1(x,t) \text{ and } \rho_2 = \rho_2(x,t) \text{ are continuous and } \geq \rho_* > 0 \text{ in } Q_T \\ \rho, \rho_0, \rho_1, \rho_2 \in L^\infty(Q_{T-\delta}) \quad \forall \delta > 0. \end{cases} \tag{6.11}$$

Precisely, we will take $\rho_0 = (T-t)^{3/2}\rho$, $\rho_1 = (T-t)\rho$ and $\rho_2 = (T-t)^{1/2}\rho$ where ρ is defined as follow

$$\rho(x,t) = \exp\Big(\frac{s\beta(x)}{\ell(t)}\Big), \quad s \geq C(\Omega, \omega, T, \|g'\|_\infty) \tag{6.12}$$

with $\ell(t) = \begin{cases} t(T-t) & \text{si } t \geq T/4 \\ 3T^2/16 & \text{si } 0 \leq t < T/4 \end{cases}$. Here $\beta(x) = \exp(2\lambda m\|\eta^0\|_\infty) - \exp(\lambda(m\|\eta^0\|_\infty + \eta^0(x)))$, $m > 1$, $\eta^0 \in \mathcal{C}(\overline{\Omega})$ satisfies $\eta^0 > 0$ in Ω, $\eta^0 = 0$ on $\partial\Omega$ and $|\nabla\eta^0| > 0$ in $\overline{\Omega\backslash\omega}$ (see [10], Lemma 1.2, p.1401).

In the next section, we shall make use the following controllability result.

Proposition 6.2.1. *Assume $A \in L^\infty(Q_T)$, $\rho_2 B \in L^2(0,T;H^{-1}(\Omega))$ and $z_0 \in L^2(\Omega)$. Then there exists a control $v \in L^2(\rho_0, q_T)$ such that the weak solution z of*

$$\begin{cases} z_t - \Delta z + Az = v1_\omega + B \quad \textit{in} \quad Q_T, \\ z = 0 \textit{ on } \Sigma_T, \quad z(\cdot,0) = z_0 \textit{ in } \Omega \end{cases} \tag{6.13}$$

satisfies

$$z(\cdot,T) = 0 \textit{ in } \Omega. \tag{6.14}$$

Moreover, the unique control u which minimizes together with the corresponding solution z the functional $J : L^2(\rho, Q_T) \times L^2(\rho_0, q_T) \to \mathbb{R}^+$ defined by $J(z,v) := \frac{1}{2}\|\rho\, z\|^2_{L^2(Q_T)} + \frac{1}{2}\|\rho_0\, v\|^2_{L^2(q_T)}$ satisfies the following estimate

$$\|\rho\, z\|_{L^2(Q_T)} + \|\rho_0\, v\|_{L^2(q_T)} \leq C\bigg(\|\rho_2 B\|_{L^2(0,T;H^{-1}(\Omega))} + \|z_0\|_{L^2(\Omega)}\bigg) \tag{6.15}$$

for some constant $C = C(\Omega, \omega, T, \|A\|_\infty)$.

The controlled solution also satisfies, for some constant $C = C(\Omega, \omega, T, \|A\|_\infty)$, the estimate

$$\|\rho_1 z\|_{L^\infty(0,T;L^2(\Omega))} + \|\rho_1 \nabla z\|_{L^2(Q_T)^d} \leq C\bigg(\|\rho_2\, B\|_{L^2(0,T;H^{-1}(\Omega))} + \|z_0\|_{L^2(\Omega)}\bigg). \tag{6.16}$$

Proof:

Let us first set

$$P_0 = \{q \in C^2(\overline{Q_T}) \ : \ q = 0 \text{ on } \Sigma_T\}.$$

The bilinear form

$$(p,q)_P := \iint_{Q_T} \rho^{-2} L_A^\star p\, L_A^\star q + \iint_{q_T} \rho_0^{-2} p\, q$$

where $L_A^\star q := -q_t - \Delta q + Aq$, is a scalar product on P_0 (see [12]). The completion P of P_0 for the norm $\|\cdot\|_P$ associated to this scalar product is a Hilbert space and the following result proved in [14] holds.

Lemma 6.2.1. *There exists $C = C(\Omega, \omega, T, \|A\|_\infty) > 0$ such that one has the following Carleman estimate, for all $p \in P$:*

$$\iint_{Q_T} \left(\rho_1^{-2} |\nabla p|^2 + \rho_0^{-2} |p|^2 \right) \leq C \|p\|_P^2. \tag{6.17}$$

Remark 2. We denote by P (instead of P_A) the completion of P_0 for the norm $\| \cdot \|_P$ since P does not depend on A (see [11]).

Lemma 6.2.2. *There exists $C = C(\Omega, \omega, T, \|A\|_\infty) > 0$ such that one has the following observability inequality, for all $p \in P$:*

$$\|p(\cdot, 0)\|_{L^2(\Omega)} \leq C \|p\|_P. \tag{6.18}$$

Proof:
From the definition of ρ_0, ρ_1 and ρ_2, $P \hookrightarrow H^1(0, \frac{T}{2}; L^2(\Omega)) \hookrightarrow C([0, \frac{T}{2}]; L^2(\Omega))$ where each imbedding is continuous. The result follows from Lemma 6.2.1. □

Lemma 6.2.3. *There exists $p \in P$ unique solution of*

$$(p, q)_P = \int_\Omega z_0 q(0) + \int_0^T \langle \rho_2 B, \rho_2^{-1} q \rangle_{H^{-1}(\Omega) \times H_0^1(\Omega)}, \quad \forall q \in P. \tag{6.19}$$

This solution satisfies the following estimate :

$$\|p\|_P \leq C \left(\|\rho_2 \, B\|_{L^2(0,T;H^{-1}(\Omega))} + \|z_0\|_{L^2(\Omega)} \right)$$

where $C = C(\Omega, \omega, T, \|A\|_\infty) > 0$.

Proof:
The linear map $L_1 : P \to \mathbb{R}$, $q \mapsto \int_0^T \langle \rho_2 B, \rho_2^{-1} q \rangle_{H^{-1}(\Omega) \times H_0^1(\Omega)}$ is continuous. Indeed, for all $q \in P$

$$\Big| \int_0^T \langle \rho_2 B, \rho_2^{-1} q \rangle_{H^{-1}(\Omega) \times H_0^1(\Omega)} \Big| \leq \Big(\int_0^T \|\rho_2 B\|_{H^{-1}(\Omega)}^2 \Big)^{1/2} \Big(\int_0^T \|\rho_2^{-1} q\|_{H_0^1(\Omega)}^2 \Big)^{1/2}$$

and a.e. in $(0,T)$ $\|\rho_2^{-1} q\|_{H_0^1(\Omega)}^2 = \|\rho_2^{-1} q\|_{L^2(\Omega)}^2 + \|\nabla(\rho_2^{-1} q)\|_{L^2(\Omega)^d}^2$. But since $\rho_0 \leq T \rho_2$ a.e. t in $(0,T)$

$$\|\rho_2^{-1} q\|_{L^2(\Omega)}^2 \leq \frac{1}{T^2} \|\rho_0^{-1} q\|_{L^2(\Omega)}^2, \ a.e.\ t \in (0,T).$$

Moreover

$$\nabla(\rho_2^{-1} q) = \nabla(\rho_2^{-1}) q + \rho_2^{-1} \nabla q = -\frac{s \nabla \beta(x)}{\ell(t)(T-t)^{1/2}} \rho^{-1} + \rho_2^{-1} \nabla q$$

and thus, since $\rho_1 \leq T^{1/2} \rho_2$ a.e. t in $(0,T)$:

$$\begin{aligned} \|\nabla(\rho_2^{-1} q)\|_{L^2(\Omega)^d}^2 &\leq \left\| \frac{s \nabla \beta(x)}{\ell(t)(T-t)^{1/2}} \rho^{-1} q \right\|_{L^2(\Omega)^d}^2 + \|\rho_2^{-1} \nabla q\|_{L^2(\Omega)^d}^2 \\ &\leq C(\Omega, \omega, T, \|A\|_\infty) \big(\|\rho_0^{-1} q\|_{L^2(\Omega)}^2 + \|\rho_1^{-1} \nabla q\|_{L^2(\Omega)}^2 \big). \end{aligned}$$

We then deduce that, a.e. in $(0,T)$

$$\|\rho_2^{-1}q\|^2_{H_0^1(\Omega)} \leq C(\Omega,\omega,T,\|A\|_\infty)\big(\|\rho_0^{-1}q\|^2_{L^2(\Omega)} + \|\rho_1^{-1}\nabla q\|^2_{L^2(\Omega)^d}\big)$$

and from the Carleman estimate (6.17) that

$$\Big(\int_0^T \|\rho_2^{-1}q\|^2_{H_0^1(\Omega)}\Big)^{1/2} \leq C(\Omega,\omega,T,\|A\|_\infty)\|q\|_P$$

and therefore

$$\Big|\int_0^T \langle \rho_2 B, \rho_2^{-1}q\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)}\Big| \leq C(\Omega,\omega,T,\|A\|_\infty)\Big(\int_0^T \|\rho_2 B\|^2_{H^{-1}(\Omega)}\Big)^{1/2}\|q\|_P.$$

Thus L_1 is continuous.

From (6.18) we easily deduce that the linear map $L_2 : P \to \mathbb{R}$, $q \mapsto \int_\Omega z_0 q(0)$ is continuous. Using Riesz's theorem, we conclude that there exists exactly one solution $p \in P$ of (6.19). □

Let us now introduce the convex set

$$C(z_0,T) = \Big\{(z,v) : \rho z \in L^2(Q_T),\ \rho_0 v \in L^2(q_T),\ (z,v) \text{ solves } (6.13)-(6.14) \text{ in the transposition sense}\Big\}$$

that is (z,v) is solution of

$$\iint_{Q_T} z L_A^\star q = \iint_{q_T} vq + \int_\Omega z_0 q(0) + \int_0^T \langle B, q\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)}, \quad \forall q \in P.$$

Let us remark that if $(z,v) \in C(z_0,T)$, then since $z_0 \in L^2(\Omega)$, $v \in L^2(q_T)$ and $B \in L^2(0,T;H^{-1}(\Omega))$, z must coincide with the unique weak solution of (6.13) associated to v.

We can now claim that $C(z_0,T)$ is a non empty. Indeed we have :

Lemma 6.2.4. *Let $p \in P$ defined in Lemma 6.2.3 and (z,v) defined by*

$$z = \rho^{-2}L_A^\star p \quad \textit{and} \quad v = -\rho_0^{-2}p|_{q_T}. \tag{6.20}$$

Then $(z,v) \in C(z_0,T)$ and satisfies the following estimate

$$\|\rho\, z\|_{L^2(Q_T)} + \|\rho_0\, v\|_{L^2(q_T)} \leq C\bigg(\|\rho_2\, B\|_{L^2(0,T;H^{-1}(\Omega))} + \|z_0\|_{L^2(\Omega)}\bigg) \tag{6.21}$$

where $C = C(\Omega,\omega,T,\|A\|_\infty) > 0$.

Proof:

Let us prove that (z,v) belongs to $C(z_0,T)$. From the definition of P, $\rho z \in L^2(Q_T)$ and $\rho_0 v \in L^2(q_T)$ and from the definition of ρ, ρ_0, ρ_2, $z \in L^2(Q_T)$ and $v \in L^2(q_T)$. In view of (6.19), (z,v) is solution of

$$\iint_{Q_T} z L_A^\star q = \iint_{q_T} vq + \int_\Omega z_0 q(0) + \int_0^T \langle \rho_2 B, \rho_2^{-1}q\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)}, \quad \forall q \in P \tag{6.22}$$

that is, since from the definition of ρ_2, $B \in L^2(0,T;H^{-1}(\Omega))$ and $\int_0^T \langle \rho_2 B, \rho_2^{-1} q\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)} = \int_0^T \langle B, q\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)}$, z is the solution of (6.13) associated to v in the transposition sense. Thus $C(z_0,T) \neq \emptyset$. □

Let us now consider the following extremal problem, introduced by Fursikov and Imanuvilov [14]

$$\begin{cases} \text{Minimize } J(z,v) = \dfrac{1}{2}\|(z,v)\|^2_{L^2(\rho^2;Q_T)\times L^2(\rho_0^2;q_T)} = \dfrac{1}{2}\displaystyle\iint_{Q_T} \rho^2|z|^2 + \dfrac{1}{2}\iint_{q_T} \rho_0^2|v|^2 \\ \text{Subject to } (z,v) \in C(z_0,T). \end{cases} \tag{6.23}$$

Then $(z,v) \mapsto J(z,v)$ is clearly strictly convex and continuous on $L^2(\rho^2;Q_T)\times L^2(\rho_0^2;q_T)$. Therefore (6.23) possesses at most a unique solution in $C(z_0,T)$. More precisely we have :

Proposition 6.2.2. *$(z,v) \in C(z_0,T)$ defined in Lemma 6.2.4 is the unique solution of (6.23).*

Proof:

Let $(y,w) \in C(z_0,T)$. Since J is convex and differentiable on $L^2(\rho^2;Q_T)\times L^2(\rho_0^2;q_T)$ we have :

$$\begin{aligned} J(y,w) &\geq J(z,v) + \iint_{Q_T} \rho^2 z(y-z) + \iint_{q_T} \rho_0^2 v(w-v) \\ &= J(z,v) + \iint_{Q_T} L^\star p(y-z) - \iint_{q_T} p(w-v) \\ &= J(z,v) \end{aligned}$$

y being the solution of (6.13) associated to w in the transposition sense. Hence (z,v) solves (6.23).

To finish the proof of Proposition 6.2.1, it suffices to prove that (z,v) satisfies the estimate (6.16). Since z is a weak solution of (6.13) associated to v, $z \in L^2(0,T;H_0^1(\Omega))$ and $z_t \in L^2(0,T;H^{-1}(\Omega))$. Multiplying (6.13) by $\rho_1^2 z$ and integrating by part we obtain, a.e. t in $(0,T)$

$$\begin{aligned} \frac{1}{2}\partial_t \int_\Omega |z|^2\rho_1^2 - \int_\Omega |z|^2\rho_1\partial_t\rho_1 + \int_\Omega \rho_1^2|\nabla z|^2 + 2\int_\Omega \rho_1 z\nabla\rho_1\cdot\nabla z + \int_\Omega \rho_1^2 Azz \\ = \int_\omega v\rho_1^2 z + \langle B, \rho_1^2 z\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)}. \end{aligned}$$

But $\partial_t\rho_1 = -\rho - (T-t)\frac{s\beta\ell'(t)}{\ell(t)^2}\rho$, so that

$$\Big|\int_\Omega |z|^2\rho_1\partial_t\rho_1\Big| \leq C(\Omega,\omega,T,\|A\|_\infty)\int_\Omega |\rho z|^2.$$

Since $\nabla\rho_1 = (T-t)\nabla\rho = (T-t)\frac{s\nabla\beta}{\ell(t)}\rho$ we have

$$\Big|\int_\Omega \rho_1 z\nabla\rho_1\cdot\nabla z\Big| \leq C(\Omega,\omega,T,\|A\|_\infty)\Big(\int_\Omega |\rho_1\nabla z|^2\Big)^{1/2}\Big(\int_\Omega |\rho z|^2\Big)^{1/2}.$$

The following estimates also hold

$$\Big|\int_\Omega \rho_1^2 Azz\Big| \leq C(T,\|A\|_\infty)\int_\Omega |\rho z|^2,$$

$$\Big|\int_\omega v\rho_1^2 z\Big| \le T^{1/2}\Big|\int_\omega \rho_0 v\rho z\Big| \le T^{1/2}\Big(\int_\omega |\rho_0 v|^2\Big)^{1/.2}\Big(\int_\Omega |\rho z|^2\Big)^{1/2}$$

and

$$\begin{aligned}|\langle B,\rho_1^2 z\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)}| &= |\langle \rho_1 B,\rho_1 z\rangle_{H^{-1}(\Omega)\times H_0^1(\Omega)}| \le \|\rho_1 B\|_{H^{-1}(\Omega)}\|\rho_1 z\|_{H_0^1(\Omega)}\\ &\le C(\Omega,\omega,T,\|A\|_\infty)\|\rho_2 B\|_{H^{-1}(\Omega)}\big(\|\rho z\|_{L^2(\Omega)}+\|\rho_1\nabla z\|_{L^2(\Omega)^d}\big).\end{aligned}$$

Thus we easily obtain that

$$\partial_t\int_\Omega \rho_1^2|z|^2+\int_\Omega \rho_1^2|\nabla z|^2 \le C(\Omega,\omega,T,\|A\|_\infty)\left(\|\rho_2 B\|^2_{H^{-1}(\Omega)}+\int_\Omega \rho^2|z|^2+\int_\omega |\rho_0 v|^2\right)$$

and therefore, using (6.21), for all $t\in[0,T]$:

$$\Big(\int_\Omega \rho_1^2|z|^2\Big)(t)+\iint_{Q_t}\rho_1^2|\nabla z|^2 \le C(\Omega,\omega,T,\|A\|_\infty)\left(\|\rho_2\, B\|^2_{L^2(0,T;H^{-1}(\Omega))}+\|z_0\|^2_{L^2(\Omega)}\right)$$

which gives (6.16) and concludes the proof of Proposition 6.2.1.

6.3. The least-squares method and its analysis

For any $s\in[0,1]$, we define the space

$$W_s=\left\{g\in\mathcal{C}(\mathbb{R}),\ g(0)=0,\ g'\in L^\infty(\mathbb{R}),\ \sup_{a,b\in\mathbb{R},a\neq b}\frac{|g'(a)-g'(b)|}{|a-b|^s}<\infty\right\}.$$

The case $s=0$ reduces to $W_0=\{g\in\mathcal{C}(\mathbb{R}),\ g(0)=0,\ g'\in L^\infty(\mathbb{R})\}$ while the case $s=1$ corresponds to $W_1=\{g\in\mathcal{C}(\mathbb{R}),\ g(0)=0,\ g'\in L^\infty(\mathbb{R}), g''\in L^\infty(\mathbb{R})\}$.

In the sequel, we shall assume that there exists one $s\in(0,1]$ for which the nonlinear function g belongs to W_s. Remark that $g\in W_s$ for some $s\in[0,1]$ satisfies hypotheses (6.2) and (6.6). We shall also assume that $u_0\in L^2(\Omega)$.

6.3.1. The least-squares method

We introduce the vectorial space $\mathcal{A}_0$ as follows

$$\begin{aligned}\mathcal{A}_0=\Big\{&(y,f): \rho\, y\in L^2(Q_T),\ \rho_1\,\nabla y\in L^2(Q_T)^d,\ \rho_0 f\in L^2(q_T),\\ &\rho_2(y_t-\Delta y-f\,1_\omega)\in L^2(0,T;H^{-1}(\Omega)),\ y(\cdot,0)=0 \text{ in }\Omega,\ y=0 \text{ on }\Sigma_T\Big\}\end{aligned} \tag{6.24}$$

where ρ, ρ_2, ρ_1 and ρ_0 are defined in (6.12). Since $L^2(0,T;H^{-1}(\Omega))$ is also a Hilbert space, $\mathcal{A}_0$ endowed with the following scalar product

$$\begin{aligned}\big((y,f),(\overline{y},\overline{f})\big)_{\mathcal{A}_0} &= \big(\rho y,\rho\overline{y}\big)_2+\big(\rho_1\nabla y,\rho_1\nabla\overline{y}\big)_2+\big(\rho_0 f,\rho_0\overline{f}\big)_2\\ &\quad+\big(\rho_2(y_t-\Delta y-f\,1_\omega),\rho_2(\overline{y}_t-\Delta\overline{y}-\overline{f}\,1_\omega)\big)_{L^2(0,T;H^{-1}(\Omega))}\end{aligned}$$

is a Hilbert space. The corresponding norm is $\|(y,f)\|_{\mathcal{A}_0} = \sqrt{((y,f),(y,f))_{\mathcal{A}_0}}$. We also consider the convex space

$$\mathcal{A} = \Big\{ (y,f) : \rho\, y \in L^2(Q_T),\ \rho_1\, \nabla y \in L^2(Q_T)^d,\ \rho_0 f \in L^2(q_T), \\ \rho_2(y_t - \Delta y - f\, 1_\omega) \in L^2(0,T;H^{-1}(\Omega)),\ y(\cdot,0) = u_0 \text{ in } \Omega,\ y = 0 \text{ on } \Sigma_T \Big\} \tag{6.25}$$

so that we can write $\mathcal{A} = (\overline{y},\overline{f}) + \mathcal{A}_0$ for any element $(\overline{y},\overline{f}) \in \mathcal{A}$. We endow $\mathcal{A}$ with the same norm. Clearly, if $(y,f) \in \mathcal{A}$, then $y \in C([0,T];L^2(\Omega))$ and since $\rho\, y \in L^2(Q_T)$, then $y(\cdot,T) = 0$. The null controllability requirement is therefore incorporated in the spaces $\mathcal{A}_0$ and $\mathcal{A}$.

For any fixed $(\overline{y},\overline{f}) \in \mathcal{A}$, we can now consider the following extremal problem :

$$\min_{(y,f)\in\mathcal{A}_0} E(\overline{y}+y, f+\overline{f}) \tag{6.26}$$

where $E : \mathcal{A} \to \mathbb{R}$ is defined as follows

$$E(y,f) := \frac{1}{2}\left\| \rho_2\Big(y_t - \Delta y + g(y) - f\, 1_\omega \Big) \right\|^2_{L^2(0,T;H^{-1}(\Omega))} \tag{6.27}$$

justifying the least-squares terminology we have used.

Let us remark that, if $g \in W_s$ for one $s \geq 0$, then g is Lipschitz and thus, since $g(0) = 0$, there exists $K > 0$ such that $|g(\xi)| \leq K|\xi|$ for all $\xi \in \mathbb{R}$. Consequently, $\rho_2 g(y) \in L^2(Q_T)$ (and then $\rho_2 g(y) \in L^2(0,T;H^{-1}(\Omega))$) since

$$\|\rho_2 g(y)\|_{L^2(Q_T)} = \|(\rho_2\rho^{-1})\rho g(y)\|_{L^2(Q_T)} = \|(T-t)^{1/2}\rho g(y)\|_{L^2(Q_T)} \leq T^{1/2}K\|\rho y\|_{L^2(Q_T)}.$$

Since any $g \in W_s$ satisfies hypotheses (6.2) and (6.6), the controllability result of Theorem 6.1.2 given in [13] implies the existence of at least one pair $(y,f) \in \mathcal{A}$ such that $E(y,f) = 0$. The extremal problem (6.26) admits therefore solutions. Conversely, any pair $(y,f) \in \mathcal{A}$ for which $E(y,f)$ vanishes is a controlled pair of (6.1). In this sense, the functional E is a so-called error functional which measures the deviation of (y,f) from being a solution of the underlying nonlinear equation. We emphasize that the $L^2(0,T;H^{-1}(\Omega))$ norm in E indicates that we are looking for weak solutions of the parabolic equation (6.1). We refer to [19] where a similar so-called weak least-squares method is employed to approximate the solutions of the unsteady Navier-Stokes equation.

A practical way of taking a functional to its minimum is through some clever use of descent directions, i.e the use of its derivative. In doing so, the presence of local minima is always something that may dramatically spoil the whole scheme. The unique structural property that discards this possibility is the strict convexity of the functional E. However, for nonlinear equation like (6.1), one cannot expect this property to hold for the functional E. Nevertheless, we insist in that one may construct a particular minimizing sequence which cannot converge except to a global minimizer leading E down to zero.

In order to construct such minimizing sequence, we look, for any $(y,f) \in \mathcal{A}$, for a pair $(Y^1,F^1) \in \mathcal{A}_0$ solution of the following formulation

$$\begin{cases} Y^1_t - \Delta Y^1 + g'(y)Y^1 = F^1 1_\omega + \big(y_t - \Delta y + g(y) - f\, 1_\omega\big) \quad \text{in} \quad Q_T, \\ Y^1 = 0 \text{ on } \Sigma_T, \quad Y^1(\cdot,0) = 0 \text{ in } \Omega. \end{cases} \tag{6.28}$$

Since $(Y^1, F^1) \in \mathcal{A}_0$, F^1 is a null control for Y^1. We have the following property.

Proposition 6.3.1. *Let any $(y, f) \in \mathcal{A}$. There exists a pair $(Y^1, F^1) \in \mathcal{A}_0$ solution of (6.28) which satisfies the following estimate:*

$$\|(Y^1, F^1)\|_{\mathcal{A}_0} \leq C\sqrt{E(y, f)} \tag{6.29}$$

for some $C = C(\Omega, \omega, T, \|g'\|_\infty) > 0$.

Proof:

For all $(y, f) \in \mathcal{A}$ we have $\rho_2(y_t - \Delta y + g(y) - f1_\omega) \in L^2(0, T; H^{-1}(\Omega))$. The existence of a null control F^1 is therefore given by Proposition 6.2.1. Choosing the control F^1 which minimizes together with the corresponding solution Y^1 the functional J defined in Proposition 6.2.1, we get the following estimate (since $Y^1(\cdot, 0) = 0$)

$$\begin{aligned} \|\rho\, Y^1\|_{L^2(Q_T)} + \|\rho_0 F^1\|_{L^2(q_T)} &\leq C\|\rho_2(y_t - \Delta y + g(y) - f1_\omega)\|_{L^2(0,T;H^{-1}(\Omega))} \\ &\leq C\sqrt{E(y, f)} \end{aligned} \tag{6.30}$$

and

$$\begin{aligned} \|\rho_1\, Y^1\|_{L^\infty(0,T;L^2(\Omega))} + \|\rho_1 \nabla Y^1\|_{L^2(Q_T)^d} &\leq C\|\rho_2(y_t - \Delta y + g(y) - f1_\omega)\|_{L^2(0,T;H^{-1}(\Omega))} \\ &\leq C\sqrt{E(y, f)} \end{aligned} \tag{6.31}$$

for some $C = C(\Omega, \omega, T, \|g\|_\infty)$ independent of Y^1, F^1 and y. Eventually, from the equation solved by Y^1,

$$\begin{aligned} &\|\rho_2(Y^1_t - \Delta Y^1 - F^1\, 1_\omega)\|_{L^2(0,T;H^{-1}(\Omega))} \\ &\qquad \leq \|\rho_2 g'(y) Y^1\|_{L^2(Q_T)} + \|\rho_2(y_t - \Delta y + g(y) - f\, 1_\omega)\|_{L^2(0,T;H^{-1}(\Omega))} \\ &\qquad \leq \|(T-t)^{1/2} g'(y)\|_\infty \|\rho Y^1\|_{L^2(Q_T)} + \sqrt{2E(y, f)} \\ &\qquad \leq \max\left(1, \|(T-t)^{1/2} g'\|_\infty\right) C\sqrt{E(y, f)} \end{aligned} \tag{6.32}$$

which proves that (Y^1, F^1) belongs to $\mathcal{A}_0$. □

Remark 3. From (6.28), $z = y - Y^1$ is a null controlled solution satisfying

$$\begin{cases} z_t - \Delta z + g'(y) z = (f - F^1)1_\omega - g(y) + g'(y) y \quad \text{in} \quad Q_T, \\ z = 0 \text{ on } \Sigma_T, \quad z(\cdot, 0) = u_0 \text{ in } \Omega \end{cases} \tag{6.33}$$

by the control $(f - F^1) \in L^2(\rho_0, q_T) := \{f; \rho_0 f \in L^2(q_T)\}$.

Remark 4. We emphasize that the presence of a right hand side in (6.28), namely $y_t - \Delta y + g(y) - f\, 1_\omega$, forces us to introduce from the beginning the weights ρ_0, ρ_1, ρ_2 and ρ in the spaces $\mathcal{A}_0$ and $\mathcal{A}$. This can be seen from the equality (6.19): since $\rho_2^{-1} q$ belongs to $L^2(0, T; H^1(\Omega))$ for all $q \in P$, we need to impose that $\rho_2 B \in L^2(0, T; H^{-1}(\Omega))$ with here $B = y_t - \Delta y + g(y) - f\, 1_\omega$. Working with the linearized equation (6.7) (introduced in [13]) which does not make appear an additional right hand side, we may avoid the introduction of Carleman type weights. Actually, the authors in (6.7) consider controls of minimal $L^\infty(q_T)$ norm. Introduction of weights allows however the characterization (6.19), which is very convenient at the practical level. We refer to [12] where this is discussed at length.

The interest of the pair $(Y^1, F^1) \in \mathcal{A}_0$ lies in the following result.

Proposition 6.3.2. *Let $(y, f) \in \mathcal{A}$ and let $(Y^1, F^1) \in \mathcal{A}_0$ be a solution of (6.28). Then the derivative of E at the point $(y, f) \in \mathcal{A}$ along the direction (Y^1, F^1) given by $E'(y, f) \cdot (Y^1, F^1) := \lim_{\eta \to 0, \eta \neq 0} \frac{E((y,f)+\eta(Y^1,F^1)) - E(y,f)}{\eta}$ satisfies*

$$E'(y, f) \cdot (Y^1, F^1) = 2E(y, f). \tag{6.34}$$

Proof:

We preliminary check that for all $(Y, F) \in \mathcal{A}_0$, E is differentiable at the point $(y, f) \in \mathcal{A}$ along the direction $(Y, F) \in \mathcal{A}_0$. For all $\lambda \in \mathbb{R}$, simple computations lead to the equality

$$E(y + \lambda Y, f + \lambda F) = E(y, f) + \lambda E'(y, f) \cdot (Y, F) + h((y, f), \lambda(Y, F))$$

with

$$E'(y, f)\cdot(Y, F) := \Big(\rho_2(y_t - \Delta y + g(y) - f\,1_\omega), \rho_2(Y_t - \Delta Y + g'(y)Y - F\,1_\omega)\Big)_{L^2(0,T;H^{-1}(\Omega))} \tag{6.35}$$

and

$$\begin{aligned} h((y, f), \lambda(Y, F)) =& \lambda\Big(\rho_2(Y_t - \Delta Y + g'(y)Y - F\,1_\omega), \rho_2 l(y, \lambda Y)\Big)_{L^2(0,T;H^{-1}(\Omega))} \\ &+ \frac{\lambda^2}{2}\|\rho_2(Y_t - \Delta Y + g'(y)Y - F\,1_\omega)\|^2_{L^2(0,T;H^{-1}(\Omega))} \\ &+ \Big(\rho_2(y_t - \Delta y + g(y) - f\,1_\omega), \rho_2 l(y, \lambda Y)\Big)_{L^2(0,T;H^{-1}(\Omega))} \\ &+ \frac{1}{2}\|\rho_2 l(y, \lambda Y)\|^2_{L^2(0,T;H^{-1}(\Omega))} \end{aligned}$$

where $l(y, \lambda Y) = g(y + \lambda Y) - g(y) - \lambda g'(y)Y$.

The application $(Y, F) \to E'(y, f) \cdot (Y, F)$ is linear and continuous from $\mathcal{A}_0$ to $\mathbb{R}$ as it satisfies

$$\begin{aligned} &|E'(y, f) \cdot (Y, F)| \\ &\quad \le \|\rho_2(y_t - \Delta y + g(y) - f\,1_\omega)\|_{L^2(0,T;H^{-1}(\Omega))}\|\rho_2(Y_t - \Delta Y + g'(y)Y - F\,1_\omega)\|_{L^2(0,T;H^{-1}(\Omega))} \\ &\quad \le \sqrt{2E(y, f)}\Big(\|\rho_2(Y_t - \Delta Y - F\,1_\omega)\|_{L^2(0,T;H^{-1}(\Omega))} + \|\rho_2 g'(y)Y\|_{L^2(Q_T)}\Big) \\ &\quad \le \sqrt{2E(y, f)}\Big(\|\rho_2(Y_t - \Delta Y - F\,1_\omega)\|_{L^2(0,T;H^{-1}(\Omega))} + \|(T - t)^{1/2} g'(y)\|_{L^\infty(Q_T)}\|\rho Y\|_{L^2(Q_T)}\Big) \\ &\quad \le \sqrt{2E(y, f)}\max\Big(1, \|(T - t)^{1/2} g'\|_\infty\Big)\|(Y, F)\|_{\mathcal{A}_0}. \end{aligned}$$

Similarly, for all $\lambda \in \mathbb{R}^\star$

$$\begin{aligned} |\frac{1}{\lambda}h((y, f), \lambda(Y, F))| \le& \Big(\lambda\|\rho_2(Y_t - \Delta Y + g'(y)Y - F\,1_\omega)\|_{L^2(0,T;H^{-1}(\Omega))} + \sqrt{2E(y, f)} \\ &+ \frac{1}{2}\|\rho_2 l(y, \lambda Y)\|_{L^2(0,T;H^{-1}(\Omega))}\Big)\frac{1}{\lambda}\|\rho_2 l(y, \lambda Y)\|_{L^2(0,T;H^{-1}(\Omega))} \\ &+ \frac{\lambda}{2}\|\rho_2(Y_t - \Delta Y + g'(y)Y - F\,1_\omega)\|^2_{L^2(0,T;H^{-1}(\Omega))}. \end{aligned}$$

Since $g' \in L^\infty(\mathbb{R})$ we have for a.e. $(x,t) \in Q_T$:

$$\rho_2|\frac{1}{\lambda}l(y,\lambda Y)| = \rho_2\Big|\frac{g(y+\lambda Y)-g(y)}{\lambda} - g'(y)Y\Big| \leq 2\|g'\|_\infty|\rho_2 Y|$$

and $\rho_2 Y \in L^2(Q_T)$. Moreover, for a.e. $(x,t) \in Q_T$, $\rho_2|\frac{1}{\lambda}l(y,\lambda Y)| = \rho_2|\frac{g(y+\lambda Y)-g(y)}{\lambda} - g'(y)Y| \to 0$ as $\lambda \to 0$; it follows from the Lebesgue's Theorem that

$$\frac{1}{\lambda}\|\rho_2 l(y,\lambda Y)\|_{L^2(Q_T)} \to 0 \text{ as } \lambda \to 0.$$

It is now easy to see that

$$h((y,f),\lambda(Y,F)) = o(\lambda)$$

and that the functional E is differentiable at the point $(y,f) \in \mathcal{A}$ along the direction $(Y,F) \in \mathcal{A}_0$. Eventually, the equality (6.34) follows from the definition of the pair (Y^1,F^1) given in (6.28). □

Remark that from the equality (6.35), the derivative $E'(y,f)$ is independent of (Y,F). We can then define the norm $\|E'(y,f)\|_{(\mathcal{A}_0)'} := \sup_{(Y,F)\in\mathcal{A}_0,(Y,F)\neq(0,0)} \frac{E'(y,f)\cdot(Y,F)}{\|(Y,F)\|_{\mathcal{A}_0}}$ associated to $(\mathcal{A}_0)'$, the set of the linear and continuous applications from $\mathcal{A}_0$ to $\mathbb{R}$.

Combining the equality (6.34) and the inequality (6.29), we deduce the following estimates of $E(y,f)$ in term of the norm of $E'(y,f)$.

Proposition 6.3.3. *For any $(y,f) \in \mathcal{A}$, the inequalities holds true*

$$C_1(\Omega,\omega,T,\|g'\|_\infty)\|E'(y,f)\|_{\mathcal{A}_0'} \leq \sqrt{E(y,f)} \leq C_2(\Omega,\omega,T,\|g'\|_\infty)\|E'(y,f)\|_{\mathcal{A}_0'}$$

for some constants $C_1, C_2 > 0$.

Proof:
(6.34) rewrites $E(y,f) = \frac{1}{2}E'(y,f)\cdot(Y^1,F^1)$ where $(Y^1,F^1) \in \mathcal{A}_0$ is solution of (6.28) and therefore, with (6.29)

$$E(y,f) \leq \frac{1}{2}\|E'(y,f)\|_{\mathcal{A}_0'}\|(Y^1,F^1)\|_{\mathcal{A}_0} \leq C(\Omega,\omega,T,\|g'\|_\infty)\|E'(y,f)\|_{\mathcal{A}_0'}\sqrt{E(y,f)}.$$

On the other hand, for all $(Y,F) \in \mathcal{A}_0$ (see the proof of Proposition 6.3.2) :

$$|E'(y,f)\cdot(Y,F)| \leq \sqrt{2E(y,f)}\max\Big(1,\|(T-t)^{1/2}g'\|_\infty\Big)\|(Y,F)\|_{\mathcal{A}_0}$$

and thus

$$C_1(\Omega,\omega,T,\|g'\|_\infty)\|E'(y,f)\|_{\mathcal{A}_0'} \leq \sqrt{E(y,f)}.$$

□

In particular, any *critical* point $(y,f) \in \mathcal{A}$ for E (i.e. for which $E'(y,f)$ vanishes) is a zero for E, a pair solution of the controllability problem. In other words, any sequence $(y_k,f_k)_{k>0}$ satisfying $\|E'(y_k,f_k)\|_{(\mathcal{A}_0)'} \to 0$ as $k \to \infty$ is such that $E(y_k,f_k) \to 0$ as $k \to \infty$. We insist that this property does not imply the convexity of the functional E (and *a fortiori* the strict convexity of

E, which actually does not hold here in view of the multiple zeros for E) but show that a minimizing sequence for E can not be stuck in a local minimum. Far from the zeros of E, in particular, when $\|(y,f)\|_{\mathcal{A}} \to \infty$, the right hand side inequality indicates that E tends to be convex. On the other side, the left inequality indicates the functional E is flat around its zero set. As a consequence, gradient based minimizing sequences may achieve a very low rate of convergence (we refer to [21] and also [18] devoted to the Navier-Stokes equation where this phenomenon is observed).

6.3.2. A strongly converging minimizing sequence for E

We now examine the convergence of an appropriate sequence $(y_k, f_k) \in \mathcal{A}$. In this respect, we observe that the equality (6.34) shows that $-(Y^1, F^1)$ given by the solution of (6.28) is a descent direction for the functional E. Therefore, we can define at least formally, for any $m \geq 1$, a minimizing sequence $(y_k, f_k)_{k\in\mathbb{N}}$ as follows:

$$\begin{cases} (y_0, f_0) \in \mathcal{A}, \\ (y_{k+1}, f_{k+1}) = (y_k, f_k) - \lambda_k (Y_k^1, F_k^1), \quad k \geq 0, \\ \lambda_k = argmin_{\lambda\in[0,m]} E\big((y_k, f_k) - \lambda (Y_k^1, F_k^1)\big) \end{cases} \tag{6.36}$$

where $(Y_k^1, F_k^1) \in \mathcal{A}_0$ is such that F_k^1 is a null control for Y_k^1, solution of

$$\begin{cases} Y_{k,t}^1 - \Delta Y_k^1 + g'(y_k) Y_k^1 = F_k^1 1_\omega + (y_{k,t} - \Delta y_k + g(y_k) - f_k 1_\omega) \quad \text{in} \quad Q_T, \\ Y_k^1 = 0 \text{ on } \Sigma_T, \quad Y_k^1(\cdot, 0) = 0 \text{ in } \Omega \end{cases} \tag{6.37}$$

and minimizes the functional J defined in Proposition 6.2.1. The direction Y_k^1 vanishes when E vanishes.

We first perform the analysis assuming the non linear function g in W_1, notably that $g'' \in L^\infty(\mathbb{R})$ (the derivatives here are in the sense of distribution). We first prove the following lemma.

Lemma 6.3.1. *Assume $g \in W_1$. Let $(y,f) \in \mathcal{A}$ and $(Y^1, F^1) \in \mathcal{A}_0$ defined by (6.28). For any $\lambda \in \mathbb{R}$ and $k \in \mathbb{N}$, the following estimate holds*

$$E\big((y,f) - \lambda(Y^1, F^1)\big) \leq E(y,f) \left(|1-\lambda| + \lambda^2 C(\Omega, \omega, T, \|g'\|_\infty) \|g''\|_\infty \sqrt{E(y,f)} \right)^2. \tag{6.38}$$

Proof:
With $g \in W_1$, we write that

$$|l(y, -\lambda Y^1)| = |g(y - \lambda Y^1) - g(y) + \lambda g'(y) Y^1| \leq \frac{\lambda^2}{2} \|g''\|_\infty (Y^1)^2 \tag{6.39}$$

and obtain that

$$
\begin{aligned}
&2E\big((y,f)-\lambda(Y^1,F^1)\big)\\
&=\Big\|\rho_2\big(y_t-\Delta y_k+g(y)-f\,1_\omega\big)-\\
&\qquad\qquad\lambda\rho_2\big(Y^1_t-\Delta Y^1+g'(y)Y^1-F\,1_\omega\big)+\rho_2 l(y,-\lambda Y^1)\Big\|^2_{L^2(0,T;H^{-1}(\Omega))}\\
&=\Big\|\rho_2(1-\lambda)\big(y_t-\Delta y+g(y)-f\,1_\omega\big)+\rho_2 l(y,-\lambda Y^1)\Big\|^2_{L^2(0,T;H^{-1}(\Omega))}\\
&\le\Big(\big\|\rho_2(1-\lambda)\big(y_t-\Delta y+g(y)-f\,1_\omega\big)\big\|_{L^2(0,T;H^{-1}(\Omega))}+\big\|\rho_2 l(y,-\lambda Y^1)\big\|_{L^2(0,T;H^{-1}(\Omega))}\Big)^2\\
&\le 2\bigg(|1-\lambda|\sqrt{E(y,f)}+\frac{\lambda^2}{2\sqrt{2}}\|g''\|_\infty\|\rho_2(Y^1)^2\|_{L^2(0,T;H^{-1}(\Omega))}\bigg)^2.
\end{aligned}
\tag{6.40}
$$

For $d=3$ (similar estimates hold for $d=1$ and $d=2$), using the continuous embedding of $L^{6/5}(\Omega)$ into $H^{-1}(\Omega)$, we have:

$$
\begin{aligned}
\|\rho_2(Y^1)^2\|^2_{L^2(0,T;H^{-1}(\Omega))}&\le C(\Omega)\|\rho_2(Y^1)^2\|^2_{L^2(0,T;L^{6/5}(\Omega))}\\
&\le C(\Omega)\int_0^T\|\rho_2 Y^1\|^2_{L^3(\Omega)}\|Y^1\|^2_{L^2(\Omega)}\\
&\le C(\Omega)\int_0^T\|\rho Y^1\|_{L^2(\Omega)}\|\rho_1 Y^1\|_{L^6(\Omega)}\|Y^1\|^2_{L^2(\Omega)}\\
&\le C(\Omega)\int_0^T\|\rho Y^1\|_{L^2(\Omega)}\|\nabla(\rho_1 Y^1)\|_{L^2(\Omega)^d}\|Y^1\|^2_{L^2(\Omega)}.
\end{aligned}
$$

From the definition of ρ and ρ_1 we have $\nabla\rho_1=\frac{s\nabla\beta}{\ell(t)(T-t)}\rho_1=\frac{s\nabla\beta}{\ell(t)}\rho$ and therefore a.e. t in $(0,T)$

$$
\begin{aligned}
\|\nabla(\rho_1 Y^1)\|_{L^2(\Omega)^d}&\le\|\nabla(\rho_1)Y^1\|_{L^2(\Omega)^d}+\|\rho_1\nabla Y^1\|_{L^2(\Omega)^d}\\
&\le C(\Omega,\omega,T,\|g'\|_\infty)\|\rho Y^1\|_{L^2(\Omega)}+\|\rho_1\nabla Y^1\|_{L^2(\Omega)^d}
\end{aligned}
$$

and thus

$$
\begin{aligned}
\|\rho_2(Y^1)^2\|^2_{L^2(0,T;H^{-1}(\Omega))}\le C(\Omega,\omega,T,\|g'\|_\infty)&\|\rho_1 Y^1\|^2_{L^\infty(0,T;L^2(\Omega))}\|\rho Y^1\|_{L^2(Q_T)}\\
&\times\big(\|\rho Y^1\|_{L^2(Q_T)}+\|\rho_1\nabla Y^1\|_{L^2(Q_T)^d}\big).
\end{aligned}
$$

Using (6.30) and (6.31), we obtain

$$
\|\rho_0(Y^1)^2\|^2_{L^2(0,T;H^{-1}(\Omega))}\le C(\Omega,\omega,T,\|g'\|_\infty)E(y,f)^2,
\tag{6.41}
$$

from which we get (6.38).

Proceeding as in [20], we are now in position to prove the following convergence result for the sequence $(E(y_k,f_k))_{(k\ge0)}$.

Proposition 6.3.4. *Assume $g \in W_1$. Let $(y_k, f_k)_{k\in\mathbb{N}}$ be the sequence defined by (6.36). Then $E(y_k, f_k) \to 0$ as $k \to \infty$. Moreover, there exists $k_0 \in \mathbb{N}$ such that the sequence $(E(y_k, f_k))_{k\geq k_0}$ decays quadratically.*

Proof:
We define the real function p_k as follows

$$p_k(\lambda) = |1-\lambda| + \lambda^2 c_1 \sqrt{E(y_k, f_k)} \quad \text{where} \quad c_1 := C(\Omega, \omega, T, \|g'\|_\infty)\|g''\|_\infty.$$

Lemma 6.3.1 with $(y, f) = (y_k, f_k)$ allows to write that

$$c_1\sqrt{E(y_{k+1}, f_{k+1})} \leq c_1\sqrt{E(y_k, f_k)}p_k(\widetilde{\lambda}_k), \quad \forall k \geq 0 \tag{6.42}$$

with $p_k(\widetilde{\lambda}_k) := \min_{\lambda\in[0,m]} p_k(\lambda)$.

If $c_1\sqrt{E(y_0, f_0)} < 1$ (and thus $c_1\sqrt{E(y_k, f_k)} < 1$ for all $k \in \mathbb{N}$) then

$$p_k(\widetilde{\lambda}_k) = \min_{\lambda\in[0,m]} p_k(\lambda) \leq p_k(1) = c_1\sqrt{E(y_k, f_k)}$$

and thus

$$c_1\sqrt{E(y_{k+1}, f_{k+1})} \leq \big(c_1\sqrt{E(y_k, f_k)}\big)^2 \tag{6.43}$$

implying that $c_1\sqrt{E(y_k, f_k)} \to 0$ as $k \to \infty$ with a quadratic rate.

If now $c_1\sqrt{E(y_0, f_0)} \geq 1$, we check that $I := \{k \in \mathbb{N},\ c_1\sqrt{E(y_k, f_k)} \geq 1\}$ is a finite subset of $\mathbb{N}$. For all $k \in I$, since $c_1\sqrt{E(y_k, f_k)} \geq 1$,

$$\min_{\lambda\in[0,m]} p_k(\lambda) = \min_{\lambda\in[0,1]} p_k(\lambda) = p_k\Big(\frac{1}{2c_1\sqrt{E(y_k, f_k)}}\Big) = 1 - \frac{1}{4c_1\sqrt{E(y_k, f_k)}}$$

and thus, for all $k \in I$,

$$c_1\sqrt{E(y_{k+1}, f_{k+1})} \leq \Big(1 - \frac{1}{4c_1\sqrt{E(y_k, f_k)}}\Big)c_1\sqrt{E(y_k, f_k)} = c_1\sqrt{E(y_k, f_k)} - \frac{1}{4}. \tag{6.44}$$

This inequality implies that the sequence $(c_1\sqrt{E(y_k, f_k)})_{k\in\mathbb{N}}$ strictly decreases and then that the sequence $(p_k(\widetilde{\lambda}_k)_{k\in\mathbb{N}}$ decreases as well. Thus the sequence $(c_1\sqrt{E(y_k, f_k)})_{k\in\mathbb{N}}$ decreases to 0 at least linearly and there exists $k_0 \in \mathbb{N}$ such that for all $k \geq k_0$, $c_1\sqrt{E(y_k, f_k)} < 1$, that is I is a finite subset of $\mathbb{N}$. Arguing as in the first case, it follows that $c_1\sqrt{E(y_k, f_k)} \to 0$ as $k \to \infty$. In both cases, remark that $p_k(\widetilde{\lambda}_k)$ decreases with respect to k. □

Remark 5. Writing from (6.44) that $c_1\sqrt{E(y_k, f_k)} \leq c_1\sqrt{E(y_0, f_0)} - \frac{k}{4}$ for all k such that $c_1\sqrt{E(y_k, f_k)} \geq 1$, we obtain that

$$k_0 \leq \left\lfloor 4(c_1\sqrt{E(y_0, f_0)} - 1) + 1 \right\rfloor$$

where $\lfloor x \rfloor$ denotes the integer part of $x \in \mathbb{R}^+$.

We also have the following convergence of the optimal sequence $\{\lambda_k\}_{k>0}$.

Lemma 6.3.2. *The sequence $\{\lambda_k\}_{k>0}$ defined in (6.36) converges to 1 as $k \to \infty$.*

Proof:
In view of (6.40), we have, as long as $E(y_k, f_k) > 0$, since $\lambda_k \in [0, m]$

$$\begin{aligned}(1-\lambda_k)^2 &= \frac{E(y_{k+1}, f_{k+1})}{E(y_k, f_k)} - 2(1-\lambda_k)\frac{\langle \rho_2\big(y_{k,t} + \Delta y_k + g(y_k) - f_k\, 1_\omega\big), \rho_2 l(y_k, \lambda_k Y_k^1)\rangle_{L^2(0,T;H^{-1}(\Omega))}}{E(y_k, f_k)} \\ &\qquad - \frac{\big\|\rho_2 l(y_k, \lambda_k Y_k^1)\big\|^2_{L^2(0,T;H^{-1}(\Omega))}}{2E(y_k)} \\ &\le \frac{E(y_{k+1}, f_{k+1})}{E(y_k, f_k)} - 2(1-\lambda_k)\frac{\langle \rho_2\big(y_{k,t} + \Delta y_k + g(y_k) - f_k\, 1_\omega\big), \rho_2 l(y_k, \lambda_k Y_k^1)\rangle_{L^2(0,T;H^{-1}(\Omega))}}{E(y_k, f_k)} \\ &\le \frac{E(y_{k+1}, f_{k+1})}{E(y_k, f_k)} + 2\sqrt{2}m\frac{\sqrt{E(y_k, f_k)}\|\rho_2 l(y_k, \lambda_k Y_k^1)\|_{L^2(0,T;H^{-1}(\Omega))}}{E(y_k, f_k)} \\ &\le \frac{E(y_{k+1}, f_{k+1})}{E(y_k, f_k)} + 2\sqrt{2}m\frac{\|\rho_2 l(y_k, \lambda_k Y_k^1)\|_{L^2(0,T;H^{-1}(\Omega))}}{\sqrt{E(y_k, f_k)}}\end{aligned}$$

But, from (6.39) and (6.41)

$$\begin{aligned}\|\rho_2 l(y_k, \lambda_k Y_k^1)\|_{L^2(0,T;H^{-1}(\Omega))} &\le \frac{\lambda_k^2}{2\sqrt{2}}\|g''\|_\infty \|\rho_2 (Y_k^1)^2\|_{L^2(0,T;H^{-1}(\Omega))} \\ &\le m^2\|g''\|_\infty C(T, \Omega, \omega, \|g'\|_\infty) E(y_k, f_k)\end{aligned}$$

and thus

$$(1-\lambda_k)^2 \le \frac{E(y_{k+1}, f_{k+1})}{E(y_k, f_k)} + m^2\|g''\|_\infty C(\Omega, \omega, T, \|g'\|_\infty)\sqrt{E(y_k, f_k)}.$$

Consequently, since $E(y_k, f_k) \to 0$ and $\frac{E(y_{k+1}, f_{k+1})}{E(y_k, f_k)} \to 0$, we deduce that $(1-\lambda_k)^2 \to 0$. □

We are now in position to prove the following convergence result.

Theorem 6.3.3. *Assume $g \in W_1$. Let $(y_k, f_k)_{k\in\mathbb{N}}$ be the sequence defined by (6.36). Then, $(y_k, f_k)_{k\in\mathbb{N}} \to (y, f)$ in $\mathcal{A}$ where f is a null control for y solution of* (6.1)*. Moreover, the convergence is quadratic after a finite number of iterates.*

Proof:
For all $k \in \mathbb{N}$, let $F_k = -\sum_{n=0}^k \lambda_n F_n^1$ and $Y_k = \sum_{n=0}^k \lambda_n Y_n^1$. Let us prove that $\big((Y_k, F_k)\big)_{k\in\mathbb{N}}$ converge in $\mathcal{A}_0$, i.e. that the series $\sum \lambda_n (F_n^1, Y_n^1)$ converges in $\mathcal{A}_0$. Using that $\|(Y_k^1, F_k^1)\|_{\mathcal{A}_0} \le C\sqrt{E(y_k, f_k)}$ for all $k \in \mathbb{N}$ (see (6.29)), we write

$$\sum_{n=0}^k \lambda_n \|(Y_n^1, F_n^1)\|_{\mathcal{A}_0} \le m\sum_{n=0}^k \|(Y_n^1, F_n^1)\|_{\mathcal{A}_0} \le C\sum_{n=0}^k \sqrt{E(y_n, f_n)}.$$

But $\big(\sqrt{E(y_n, f_n)}\big)_{k\in\mathbb{N}}$ and $\big(p_k(\widetilde{\lambda_k})\big)_{k\in\mathbb{N}}$ are decreasing sequences so that

$$\sqrt{E(y_n, f_n)} \le p_n(\widetilde{\lambda_n})\sqrt{E(y_{n-1}, f_{n-1})} \le p_0(\widetilde{\lambda_0})\sqrt{E(y_{n-1}, f_{n-1})} \le p_0(\widetilde{\lambda_0})^n\sqrt{E(y_0, f_0)}$$

so that, since $p_0(\widetilde{\lambda}_0) < 1$:

$$\sum_{n=0}^{k} \sqrt{E(y_n, f_n)} \leq \sqrt{E(y_0, f_0)} \frac{1 - p(\widetilde{\lambda}_0)^{k+1}}{1 - p(\widetilde{\lambda}_0)} \leq \sqrt{E(y_0, f_0)} \frac{1}{1 - p(\widetilde{\lambda}_0)}.$$

We deduce that the series $\sum_n \lambda_n(Y_n^1, F_n^1)$ is normally convergent and so convergent. Consequently, there exists $(Y, F) \in \mathcal{A}_0$ such that $(Y_k, F_k)_{k\in\mathbb{N}}$ converges to (Y, F) in $\mathcal{A}_0$.

Denoting $y = y_0 + Y$ and $f = f_0 + F$, we then have that $(y_k, f_k)_{k\in\mathbb{N}} = (y_0 + Y_k, f_0 + F_k)_{k\in\mathbb{N}}$ converges to (y, f) in $\mathcal{A}$.

It suffices now to verify that the limit (y, f) satisfies $E(y, f) = 0$. We write that $(Y_k^1, F_k^1) \in \mathcal{A}_0$ and $(y_k, f_k) \in \mathcal{A}$ solve the

$$\begin{cases} Y_{k,t}^1 - \Delta Y_k^1 + g'(y_k) \cdot Y_k^1 = F_k^1 1_\omega - (y_{k,t} - \Delta y_k + g(y_k) - f_k 1_\omega) \quad \text{in} \quad Q_T, \\ Y_k^1 = 0 \text{ on } \Sigma_T, \quad Y_k^1(\cdot, 0) = 0 \text{ in } \Omega. \end{cases} \tag{6.45}$$

Using that (Y_k^1, F_k^1) goes to zero in $\mathcal{A}_0$ as $k \to \infty$, we pass to the limit in (6.45) and get, since $g \in W_1$, that $(y, f) \in \mathcal{A}$ solves (6.1), that is $E(y, f) = 0$. □

In particular, along the sequence $(y_k, f_k)_k$ defined by (6.36), we have the following coercivity property for E, which confirms the strong convergence of the sequence $(y_k, f_k)_{k>0}$. In view of the non uniqueness of the zeros of E, remark that this property is not true in general for all (y, f) in $\mathcal{A}$.

Proposition 6.3.5. *Let $(y_k, f_k)_{k>0}$ defined by (6.36) and $(\overline{y}, \overline{f})$ its limit. Then, there exists a positive constant C such that*

$$\|(\overline{y}, \overline{f}) - (y_k, f_k)\|_{\mathcal{A}_0} \leq C\sqrt{E(y_k, f_k)}, \quad \forall k > 0. \tag{6.46}$$

Proof:
We write that

$$\begin{aligned} \|(\overline{y}, \overline{f}) - (y_k, f_k)\|_{\mathcal{A}_0} = \|\sum_{p=k+1}^{\infty} \lambda_p(Y_p^1, F_p^1)\|_{\mathcal{A}} &\leq m \sum_{p=k+1}^{\infty} \|(Y_p^1, F_p^1)\|_{\mathcal{A}_0} \\ &\leq m\, C \sum_{p=k+1}^{\infty} \sqrt{E(y_p, f_p)} \\ &\leq m\, C \sum_{p=k+1}^{\infty} p_0(\widetilde{\lambda}_0)^{p-k} \sqrt{E(y_k, f_k)} \\ &\leq m\, C \frac{p_0(\widetilde{\lambda}_0)}{1 - p_0(\widetilde{\lambda}_0)} \sqrt{E(y_k, f_k)}. \end{aligned}$$

□

We emphasize, in view of the non uniqueness of the zeros of E, that an estimate (similar to (6.46)) of the form $\|(\overline{y}, \overline{f}) - (y, f)\|_{\mathcal{A}_0} \leq C\sqrt{E(y, f)}$ does not hold for all $(y, f) \in \mathcal{A}$. We also mention the fact that the sequence $(y_k, f_k)_{k>0}$ and its limits $(\overline{y}, \overline{f})$ are uniquely determined from the initial guess (y_0, f_0) and from our criterion of selection of the control F^1. In other words, the solution $(\overline{y}, \overline{f})$ is unique up to the element (y_0, f_0) and the functional J.

6.3.3. The case $g \in W_s$, $0 \leq s < 1$ and additional remarks

The results of the previous subsection devoted to the case $s = 1$ still hold if we assume only that $g \in W_s$ for one $s \in (0,1)$. For any $g \in W_s$, we introduce the notation $\|g'\|_{\widetilde{W}^{s,\infty}(\mathbb{R})} := \sup_{a,b\in\mathbb{R},a\neq b} \frac{|g'(a)-g'(b)|}{|a-b|^s}$. We have the following result.

Theorem 6.3.4. *Assume that there exists $s \in (0,1)$ such that $g \in W_s$. Let $(y_k, f_k)_{k\in\mathbb{N}}$ be the sequence defined by (6.36). Then, $(y_k, f_k)_{k\in\mathbb{N}} \to (y, f)$ in $\mathcal{A}$ where f is a null control for y solution of* (6.1)*. Moreover, after a finite number of iterates, the rate of convergence is equal to $1 + s$.*

Proof:

We briefly sketch the proof, close to the proof of Theorem 6.3.3 for the case $s = 1$.

-We first prove for any $(y, f) \in \mathcal{A}$ and $\lambda \in \mathbb{R}$ the following inequality (similar to the inequality (6.38))

$$E\big((y,f) - \lambda(Y^1, F^1)\big) \leq E(y,f)\bigg(|1-\lambda| + \lambda^{1+s} c_1 E(y,f)^{s/2}\bigg)^2 \tag{6.47}$$

with $c_1 = C(T, \Omega, \omega, \|g'\|_\infty)\|g'\|_{\widetilde{W}^{s,\infty}(\mathbb{R})}$ and $(Y^1, F^1) \in \mathcal{A}_0$ the solution of (6.37) which minimizes J. For any $(x, y) \in \mathbb{R}^2$ and $\lambda \in \mathbb{R}$, we write $g(x + \lambda y) - g(x) = \int_0^\lambda y g'(x + \xi y) d\xi$ leading to

$$\begin{aligned}
|g(x+\lambda y) - g(x) - \lambda g'(x) y| &\leq \int_0^\lambda |y||g'(x+\xi y) - g'(x)| d\xi \\
&\leq \int_0^\lambda |y|^{1+s} |\xi|^s \frac{|g'(x+\xi y) - g'(x)|}{|\xi y|^s} d\xi \\
&\leq \|g'\|_{\widetilde{W}^{s,\infty}(\mathbb{R})} |y|^{1+s} \frac{\lambda^{1+s}}{1+s}.
\end{aligned}$$

It follows that

$$|l(y, -\lambda Y^1)| = |g(y - \lambda Y^1) - g(y) + \lambda g'(y) Y^1| \leq \|g'\|_{\widetilde{W}^{s,\infty}(\mathbb{R})} \frac{\lambda^{1+s}}{1+s} |Y^1|^{1+s}$$

and

$$\begin{aligned}
\big\|\rho_2 l(y, \lambda Y^1)\big\|_{L^2(0,T;H^{-1}(\Omega))} &\leq \big\|\rho_2 l(y, \lambda Y^1)\big\|_{L^2(0,T;L^{6/5}(\Omega))} \\
&\leq \|g'\|_{\widetilde{W}^{s,\infty}(\mathbb{R})} \frac{\lambda^{1+s}}{1+s} \big\|\rho_2 |Y^1|^{1+s}\big\|_{L^2(0,T;L^{6/5}(\Omega))}.
\end{aligned}$$

But

$$\begin{aligned}
\big\|\rho_2 |Y^1|^{1+s}\big\|^2_{L^2(0,T;L^{6/5}(\Omega))} &= \int_0^T \big\|\rho_2 |Y^1|^{1+s}\big\|^2_{L^{6/5}(\Omega)} \leq \int_0^T \big\|\rho_2 Y^1\big\|^2_{L^3(\Omega)} \big\||Y^1|^s\big\|^2_{L^2(\Omega)} \\
&\leq \int_0^T \big\|\rho Y^1\big\|_{L^2(\Omega)} \big\|\rho_1 Y^1\big\|_{L^6(\Omega)} \big\|Y^1\big\|^{2s}_{L^{2s}(\Omega)} \\
&\leq C(\Omega) \int_0^T \big\|\rho Y^1\big\|_{L^2(\Omega)} \big\|\nabla(\rho_1 Y^1)\big\|_{L^2(\Omega)^d} \big\|Y^1\big\|^{2s}_{L^{2s}(\Omega)} \\
&\leq C(\Omega) \big\|\rho Y^1\big\|_{L^2(Q_T)} \big\|\nabla(\rho_1 Y^1)\big\|_{L^2(Q_T)^d} \big\|Y^1\big\|^{2s}_{L^\infty(0,T;L^{2s}(\Omega))} \\
&\leq C(\Omega) \big\|\rho Y^1\big\|_{L^2(Q_T)} \big\|\nabla(\rho_1 Y^1)\big\|_{L^2(Q_T)^d} \big\|Y^1\big\|^{2s}_{L^\infty(0,T;L^2(\Omega))}.
\end{aligned}$$

Since $\|\nabla(\rho_1 Y^1)\|_{L^2(\Omega)^d} \le C(\Omega,\omega,T,\|g'\|_\infty)\|\rho Y^1\|_{L^2(\Omega)} + \|\rho_1\nabla Y^1\|_{L^2(\Omega)^d}$, we finally get

$$\big\|\rho_2|Y^1|^{1+s}\big\|^2_{L^2(0,T;L^{6/5}(\Omega))} \le C(\Omega,\omega,T,\|g'\|_\infty)\|\rho Y^1\|_{L^2(Q_T)} \times \big(\|\rho Y^1\|_{L^2(Q_T)} + \|\rho_1\nabla Y^1\|_{L^2(Q_T)^d}\big)\|\rho_1 Y\|^{2s}_{L^\infty(0,T;L^2(\Omega))}.$$

The first inequality of (6.40) then leads to (6.47).

- We then check that the sequence $(E(y_k,f_k))_{k\in\mathbb{N}}$ goes to zero as $k\to\infty$. We define p_k as follows

$$p_k(\lambda) = |1-\lambda| + \lambda^{1+s}c_1 E(y_k,f_k)^{s/2}$$

so that

$$\sqrt{E(y_{k+1},f_{k+1})} \le \sqrt{E(y_k,f_k)}p_k(\widetilde{\lambda}_k), \quad \forall k\ge 0$$

with $p_k(\widetilde{\lambda}_k) = \min_{\lambda\in[0,m]} p_k(\lambda)$. We have $p_k(\widetilde{\lambda}_k) := \min_{\lambda\in[0,m]} p_k(\lambda) \le p_k(1) = c_1 E(y_k,f_k)^{s/2}$ and thus

$$c_2\sqrt{E(y_{k+1},f_{k+1})} \le \big(c_2\sqrt{E(y_k,f_k)}\big)^{1+s}, \qquad c_2 := c_1^{1/s}.$$

If $c_2\sqrt{E(y_0,f_0)} < 1$ (and thus $c_2\sqrt{E(y_k,f_k)} < 1$ for all $k\in\mathbb{N}$) then the above inequality implies that $c_2\sqrt{E(y_k,f_k)} \to 0$ as $k\to\infty$. If $c_2\sqrt{E(y_0,f_0)} \ge 1$ then let $I = \{k\in\mathbb{N},\ c_2\sqrt{E(y_k,f_k)} \ge 1\}$. I is a finite subset of $\mathbb{N}$; for all $k\in I$, since $c_2\sqrt{E(y_k,f_k)} \ge 1$

$$\min_{\lambda\in[0,m]} p_k(\lambda) = \min_{\lambda\in[0,1]} p_k(\lambda) = p_k\Big(\frac{1}{(1+s)^{1/s}c_2\sqrt{E(y_k,f_k)}}\Big) = 1 - \frac{s}{(1+s)^{\frac{1}{s}+1}}\frac{1}{c_2\sqrt{E(y_k,f_k)}}$$

and thus, for all $k\in I$,

$$c_2\sqrt{E(y_{k+1},f_{k+1})} \le \Big(1-\frac{s}{(1+s)^{\frac{1}{s}+1}}\frac{1}{c_2\sqrt{E(y_k,f_k)}}\Big)c_2\sqrt{E(y_k,f_k)} = c_2\sqrt{E(y_k,f_k)} - \frac{s}{(1+s)^{\frac{1}{s}+1}}.$$

This inequality implies that the sequence $(c_2\sqrt{E(y_k,f_k)})_{k\in\mathbb{N}}$ strictly decreases and then that the sequence $(p_k(\widetilde{\lambda}_k))_{k\in\mathbb{N}}$ decreases as well. Thus the sequence $(c_2\sqrt{E(y_k,f_k)})_{k\in\mathbb{N}}$ decreases to 0 at least linearly and there exists $k_0\in\mathbb{N}$ such that for all $k\ge k_0$, $c_2\sqrt{E(y_k,f_k)} < 1$, that is I is a finite subset of $\mathbb{N}$. Similarly, the optimal parameter λ_k goes to one as $k\to\infty$.

- Using that the sequence $(E(y_k,f_k))_{k\in\mathbb{N}}$ goes to zero, we conclude exactly as in the proof of Theorem 6.3.3. □

On the other hand, if we assume only that g belongs to W_0, then we can not expect the convergence of the sequence $(y_k,f_k)_{k>0}$ if $\|g'\|_\infty$ is too large.

Remark 6. Assume that $g\in W_0$. Let any $(y,f)\in\mathcal{A}$ and (Y^1,F^1) the solution of (6.28) which minimizes J. The following inequality holds :

$$E\big((y,f)-\lambda(Y^1,F^1)\big) \le E(y,f)\bigg(|1-\lambda| + \lambda C(\Omega,\omega,T,\|g'\|_\infty)\|g'\|_\infty\bigg)^2$$

for all $\lambda \in \mathbb{R}$ where $C(\Omega, \omega, T, \|g'\|_\infty) \geq 0$ increases with $\|g'\|_\infty$. Indeed, this is a consequence of the following inequality, for all $(y, f) \in \mathcal{A}$, $(Y, F) \in \mathcal{A}_0$:

$$\begin{aligned} 2E\big((y,f) - \lambda(Y^1, F^1)\big) &\leq \Big(\big\| \rho_2 (1-\lambda)\big(y_t - \Delta y + g(y) - f\, 1_\omega\big)\big\|_{L^2(0,T;H^{-1}(\Omega))} \\ &\qquad + \big\| \rho_2 l(y, \lambda Y^1)\big\|_{L^2(0,T;H^{-1}(\Omega))} \Big)^2 \\ &\leq \Big(|1-\lambda| \sqrt{2E(y,f)} + 2\lambda \|(T-t)^{1/2} g'(y)\|_{L^\infty(Q_T)} \|\rho Y\|_{L^2(Q_T)} \Big)^2. \end{aligned}$$

As a consequence, we get that the sequence $(E(y_k, f_k))_{k\geq 0}$ decreases to 0 if g satisfies

$$C(\Omega, \omega, T, \|g'\|_\infty)\|g'\|_\infty < 1.$$

□

Remark 7. The estimate (6.29) is a key point in the convergence analysis and is independent of the choice of the functional J defined by $J(Y^1, F^1) = \frac{1}{2}\|\rho_0 F^1\|^2_{L^2(q_T)} + \frac{1}{2}\|\rho Y\|^2_{L^2(Q_T)}$ (see Proposition 6.2.1) in order to select a pair (Y^1, F^1) in $\mathcal{A}_0$. Thus, we may consider other weighted functionals, for instance $J(Y^1, F^1) = \frac{1}{2}\|\rho_0 F^1\|^2_{L^2(q_T)}$ as discussed in [22].

Remark 8. If we introduce $F : \mathcal{A} \to L^2(0,T; H^{-1}(\Omega))$ by $F(y,f) := \rho^{-2}(y_t - \Delta y + g(y) - f\, 1_\omega)$, we get that $E(y,f) = \frac{1}{2}\|F(y,f)\|^2_{L^2(0,T;H^{-1}(\Omega))}$ and observe that, for $\lambda_k = 1$, the algorithm (6.36) coincides with the Newton algorithm associated to the mapping F. This explains notably the quadratic convergence of Theorem 6.3.3 in the case $g \in W_1$ for which we have a control of g'' in $L^\infty(Q_T)$. The optimization of the parameter λ_k allows to get a global convergence of the algorithm and leads to the so-called damped Newton method (for F). Under general hypothesis, global convergence for this kind of method is achieved, with a linear rate (for instance; we refer to [7, Theorem 8.7]). As far as we know, the analysis of damped type Newton methods for partial differential equations has deserved very few attention in the literature. We mention [19, 23] in the context of fluids mechanics.

Remark 9. Suppose to simplify that λ_k equals one (corresponding to the standard Newton method). Then, for each k, the optimal pair $(Y^1_k, F^1_k) \in \mathcal{A}_0$ is such that the element (y_{k+1}, f_{k+1}) minimizes over $\mathcal{A}$ the functional $(z, v) \to J(z - y_k, v - f_k)$. Instead, we may also select the pair (Y^1_k, F^1_k) such that the element (y_{k+1}, f_{k+1}) minimizes the functional $(z, v) \to J(z, v)$. This leads to the following sequence $\{y_k, f_k\}_k$ defined by

$$\begin{cases} y_{k+1,t} - \Delta y_{k+1} + g'(y_k) y_{k+1} = f_{k+1} 1_\omega + g'(y_k) y_k - g(y_k), & \text{in } Q_T, \\ y_k = 0, & \text{on } \Sigma_T, \\ (y_{k+1}(\cdot,0), y_{k+1,t}(\cdot,0)) = (u_0, u_1), & \text{in } \Omega. \end{cases} \tag{6.48}$$

This is actually the formulation used in [11]. This formulation is different and the analysis of convergence (at least in the framework of our least-squares setting) is less direct because it is necessary to have a control of the right hand side term $g'(y_k)y_k - g(y_k)$.

Remark 10. We emphasize that the explicit construction used here allows to recover the null controllability property of (6.1) for nonlinearities g in W_s for one $s \in (0,1]$. We do not use a fixed point argument as in [13]. On the other hand, the conditions we make on g are more restrictives that in [13]. Eventually, it is also important to remark these additional conditions on g does not imply a priori a contraction property of the operator Λ introduced in [13] and mentioned in the introduction. Assume $g \in W_1$ and let z^i in $L^\infty(Q_T)$, $i = 1, 2$. If (y_{z^i}, f_{z^i}), $i = 1, 2$ are a controlled pair for the system (6.7) minimizing the functional J, then the following inequality holds :

$$\|\rho_0(f_{z^1} - f_{z^2})\|_{L^2(q_T)} + \|\rho(y_{z^1} - y_{z^2})\|_{L^2(Q_T)} \leq C(\Omega, \omega, T, \|\tilde{g}\|_\infty)\|g''\|_\infty \|u_0\|_{L^2(\Omega)} \|z^1 - z^2\|_{L^\infty(Q_T)} \tag{6.49}$$

where $C(\Omega, \omega, T, \|\tilde{g}\|_\infty)$ is the constant appearing in (6.15). In order to ensure a contraction property, we need a priori to add a smallness assumption on the data g and u_0.

6.4. Numerical illustrations

We illustrate in this section our results of convergence. We first provide some practical details of the algorithm (6.36) then discussed some experiments in the one dimensional case.

6.4.1. Approximation - Algorithm

Each iterate of the algorithm (6.36) requires the determination of the null control of F_k^1 for Y_k^1 solution of

$$\begin{cases} Y_{k,t}^1 - \Delta Y_k^1 + g'(y_k) Y_k^1 = F_k^1 1_\omega + B_k, & \text{in } Q_T, \\ Y_k^1 = 0, & \text{on } \Sigma_T, \\ Y_k^1(\cdot, 0) = 0, & \text{in } \Omega \end{cases} \tag{6.50}$$

with $B_k := y_{k,t} - \Delta y_k + g(y_k) - f_k 1_\omega$. From Lemma 6.2.4, the pair (F_k^1, Y_k^1) which minimizes the functional J is given by

$$Y_k^1 = \rho^{-2} L^\star_{g'(y_k)} p_k, \quad F_k^1 = -\rho_0^{-2} p_k \, 1_{q_T}$$

where $p_k \in P$ solves the formulation

$$\iint_{Q_T} \rho^{-1} L^\star_{g'(y_k)} p_k \, \rho^{-1} L^\star_{g'(y_k)} \overline{p} + \iint_{q_T} \rho_0^{-1} p_k \, \rho_0^{-1} \overline{p} = \int_0^T < \rho_2 B_k, \rho_2^{-1} \overline{p} >_{H^{-1}(\Omega), H_0^1(\Omega)} \, dt \quad \forall \overline{p} \in P. \tag{6.51}$$

The numerical approximation of this variational formulation (of second order in time and fourth order in space) has been discussed at length in [12]. In order, first to avoid numerical instabilities (due to the presence of exponential functions in the formulation), and second to make appear explicitly the controlled solution, we introduce the new variables

$$m_k = \rho_0^{-1} p, \qquad z_k = \rho^{-1} L^\star_{g'(y_k)} p_k.$$

Since $\rho_2^{-1} p \in L^2(0, T; H_0^1(\Omega))$, we obtain notably that $\rho_2^{-1} p = \rho_2^{-1} \rho_0 m = (T - t) m \in L^2(0, T; H_0^1(\Omega))$. From (6.51), the pair $(m_k, z_k) \in \mathcal{M} \times L^2(Q_T)$ with $\mathcal{M} := \rho_0^{-1} P$ solves

$$\iint_{Q_T} z_k \, \overline{z} + \iint_{q_T} m_k \, \overline{m} = \int_0^T < \rho_2 B_k, (T - t)\overline{m} >_{H^{-1}(\Omega), H_0^1(\Omega)} \, dt \quad \forall (\overline{m}, \overline{z}) \in \mathcal{M} \times L^2(Q_T) \tag{6.52}$$

subject to the constraint $z_k = \rho^{-1} L^\star_{g'(y_k)}(\rho_0 m_k)$. This constraint leads to the following well-posed mixed formulation : find $(m_k, z_k, \lambda_k) \in \mathcal{M} \times L^2(Q_T) \times L^2(Q_T)$ solution of

$$\left\{ \begin{array}{l} \displaystyle\iint_{Q_T} z_k\,\overline{z} + \iint_{q_T} m_k\,\overline{m} + \iint_{Q_T} \lambda_k \Big(\overline{z} - \rho^{-1} L^\star_{g'(y)}(\rho_0\,\overline{m})\Big) \\ \qquad\qquad = \displaystyle\int_0^T < \rho_2 B_k, (T-t)\overline{m} >_{H^{-1}(\Omega), H^1_0(\Omega)}\, dt, \quad \forall (\overline{m}, \overline{z}) \in \mathcal{M} \times L^2(Q_T), \\ \displaystyle\iint_{Q_T} \overline{\lambda}\Big(z_k - \rho^{-1} L^\star_{g'(y_k)}(\rho_0\, m)\Big) = 0, \quad \forall \overline{\lambda} \in L^2(Q_T). \end{array} \right. \tag{6.53}$$

The variable $\lambda_k \in L^2(Q_T)$ is a Lagrange multiplier. Moreover, from the unique solution (m_k, z_k), we get the explicit form of the controlled pair (Y^1_k, F^1_k) as follows:

$$Y^1_k = \rho^{-1} z_k, \qquad F^1_k = -\rho_0^{-1} m_k\, 1_{q_T}.$$

The algorithm associated to the sequence $(y_k, f_k)_{k>0}$ (see (6.36)) may be developed as follows: given $\epsilon > 0$ and $m \geq 1$,

I. We determine the controlled pair (y_0, f_0) which minimizes the functional J associated to the linear case (for which $g \equiv 0$ in (6.1)). (y_0, f_0) is given by

$$(y_0, f_0) = (\rho^{-1} z_0, -\rho_0^{-1} m_0\, 1_{q_T})$$

where (z_0, m_0) solves the formulation :

$$\left\{ \begin{array}{l} \displaystyle\iint_{Q_T} z\,\overline{z} + \iint_{q_T} m\,\overline{m} + \iint_{Q_T} \lambda \Big(\overline{z} - \rho^{-1} L^\star_0(\rho_0\,\overline{m})\Big) = \iint_\Omega \rho_0(\cdot,0)\, u_0\, \overline{m}(\cdot,0), \\ \qquad\qquad \forall (\overline{m}, \overline{z}) \in \mathcal{M} \times L^2(Q_T), \\ \displaystyle\iint_{Q_T} \overline{\lambda}\Big(z - \rho^{-1} L^\star_0(\rho_0\, m)\Big) = 0, \quad \forall \overline{\lambda} \in L^2(Q_T). \end{array} \right. \tag{6.54}$$

In view of Proposition 6.2.1, we check that (y_0, f_0) belongs to $\mathcal{A}$.

II. Assume now that (λ_k, f_k) is computed for some $k \geq 0$. We then compute $c_k \in L^2(0,T; H^1_0(\Omega))$, unique solution of

$$\int_{Q_T} \nabla c_k \cdot \nabla \overline{c} = \int_0^T < \rho_2(y_{k,t} - \Delta y_k + g(y_k) - f_k\, 1_\omega), \overline{c} >_{H^{-1}(\Omega), H^1_0(\Omega)}, \tag{6.55}$$

for all $\overline{c} \in L^2(0,T; H^1_0(\Omega))$ and then
$E(y_k, f_k) = \frac{1}{2} \|\rho_2(y_{k,t} - \Delta y_k + g(y_k) - f_k\, 1_\omega)\|^2_{L^2(0,T;H^{-1}(\Omega))} = \frac{1}{2} \|\nabla c_k\|^2_{L^2(Q_T)}$.

III. If $E(y_k, f_k) < \epsilon$, the approximate controlled pair is given by $(\overline{y}, \overline{f}) = (y_k, f_k)$ and the algorithm stops. Otherwise, we determine the solution $(Y^1_k, F^1_k) = (\rho^{-1} z_k, -\rho_0^{-1} m_k\, 1_{q_T})$ where (z_k, m_k) solves (6.53).

IV. Set $(y_{k+1}, f_{k+1}) = (y_k, f_k) - \lambda_k(Y_k^1, F_k^1)$ where λ_k minimizes over $[0, m]$ the scalar functional $\lambda \to E((y_k, f_k) - \lambda(Y_k^1, F_k^1))$ defined by (see (6.40))

$$2E\big((y_k, f_k) - \lambda(Y_k^1, F_k^1)\big) = \Big\| \rho_2(1-\lambda)\big(y_{k,t} - \Delta y_k + g(y_k) - f_k\, 1_\omega\big) + \rho_2 l(y_k, -\lambda Y_k^1) \Big\|^2_{L^2(0,T;H^{-1}(\Omega))} \tag{6.56}$$

with $l(y_k, -\lambda Y_k^1) = g(y_k - \lambda Y_k^1) - g(y_k) + \lambda g'(y_k) Y_k^1$. The minimization is performed using a line search method. Return to step 2.

We use the conformal space-time finite element method described in [12]. We consider a regular family $\mathcal{T} = \{\mathcal{T}_h; h > 0\}$ of triangulation of Q_T such that $\overline{Q_T} = \cup_{K \in \mathcal{T}_h} K$. The family $\mathcal{T}$ is indexed by $h = max_{K \in \mathcal{T}_h} diam(K)$. The variable z_k and λ_k are approximated with the space $P_h = \{p_h \in C(\overline{Q_T}); p_h|_K \in \mathbb{P}_1(K), \forall K \in \mathcal{T}_h\} \subset L^2(Q_T)$ where $\mathbb{P}_1(K)$ denotes the space of affine functions both in x and t. The variable m_k is approximated with the space $V_h = \{v_h \in C^1(\overline{Q_T}); v_h|_K \in \mathbb{P}(K), \forall K \in \mathcal{T}_h\} \subset \mathcal{M}$ where $\mathbb{P}(K)$ denotes the Hsieh-Clough-Tocher C^1 element (we refer to [4] page 356). These conformal approximation leads to a strong convergent approximation of the control and the controlled solution with respect to the parameter h.

6.4.2. Experiments

We present some numerical experiments in the one dimensional setting with $\Omega = (0, 1)$. The control is located on $\omega = (0.1, 0.3)$. We consider $T = 1/2$; moreover, in order to reduce the dissipation of the solution of (6.1) when $g \equiv 0$, we replace the term $-\Delta y$ in (6.1) by $-\nu\Delta y$ with $\nu > 0$ small, here $\nu = 10^{-1}$. We consider the nonlinear even function g as follows

$$g(s) = \begin{cases} l(s), & s \in [-a, a], \\ -|s|^\alpha \log^{3/2}(1 + |s|), & |s| \geq a \end{cases} \tag{6.57}$$

with $a, \alpha \in (0, 1)$. l denotes the (even) polynomial of order two such that $l(0) = 0$, $l(a) = -|a|^\alpha \log^{3/2}(1 + |a|)$ and $(-|s|^\alpha \log^{3/2}(1 + |s|))'(s = a) = l'(a)$. We use in the sequel the values $a = 10^{-1}$ and $\alpha = 0.95$. We check that g belongs to W_1, in particular $g'' \in L^\infty(\mathbb{R})$ in the sense of distribution. Remark as well that g is sublinear.

As for the initial condition to be controlled, we consider simply $u_0(x) = \beta \sin(\pi x)$ parametrized by $\beta > 0$.

The experiments are performed with the Freefem++ package developed at the Sorbonne university (see [15]), very well-adapted to the space-time formulation we employ. The algorithm is stopped when the value $E(y_k, f_k)$ is less than $\epsilon = 10^{-6}$. The optimal steps λ_k are searched in the interval $[0, 1]$.

Table 6.1, 6.2 and 6.3 collect some norms from the sequence $(y_k, f_k)_{k \geq 0}$ defined by the algorithm (6.36), initialized with the linear controlled solution, for $\beta = 10.$, $\beta = 10^2$ and $\beta = 10^3$ respectively. We use a structured mesh composed of $20\,000$ triangles, $10\,201$ vertices and for which $h \approx 1.11 \times 10^{-2}$. For $\beta = 10$, we observe the convergence after 4 iterates. The optimal steps λ_k are very close to one since $\max_k |\lambda_k - 1| < 0.05$; consequently, the algorithm (6.36) provides similar

results than the Newton algorithm (for which $\lambda_k = 1$ for all k). For $\beta = 10^2$, the convergence remains fast and is reached after 8 iterates. We can observe that some optimal steps differ from one since $\max_k |\lambda_k - 1| > 0.4$. Nevertheless, the Newton algorithm still converge after 17 iterates. More interestingly, the value $\beta = 10^3$ illustrates the features and robustness of the algorithm: the convergence is achieved after 19 iterates. Far away from a zero of E, the variations of the error functional $E(y_k, f_k)$ are first quite slow, then increase to become very fast after 16 iterates, when λ_k is close to one. In contrast, for $\beta = 10^3$., the Newton algorithm, still initialized with the linear solution diverges (see Table 6.4). As discussed in [19], in that case, a continuation method with respect to the parameter β may be combined with the Newton algorithm.

On the contrary, we mention that with these data, the sequences obtained from the algorithm (6.8) based on the linearization introduced in [13], remain bounded but do not converge, including for the value $\beta = 10$. The convergence is observed for instance with a larger size of the domain ω, for instance $\omega = (0.2, 0.8)$ (see [11, section 4.2]).

♯iterate k	$\frac{\|y_k - y_{k-1}\|_{L^2(Q_T)}}{\|y_{k-1}\|_{L^2(Q_T)}}$	$\frac{\|f_k - f_{k-1}\|_{L^2(q_T)}}{\|f_{k-1}\|_{L^2(q_T)}}$	$\|y_k\|_2$	$\|f_k\|_{2,q_T}$	$\sqrt{2E(y_k)}$	λ_k
0	–	–	4.528	4.391	5.58×10^{-1}	0.961
1	1.83×10^{-2}	1.28×10^{-3}	4.651	4.402	1.81×10^{-3}	0.996
2	4.45×10^{-4}	9.07×10^{-5}	4.661	4.403	2.72×10^{-6}	1.
3	1.12×10^{-6}	3.74×10^{-7}	4.662	4.404	4.88×10^{-8}	1.

Table 6.1: $\beta = 10.$; Results for the algorithm (6.36).

♯iterate k	$\frac{\|y_k - y_{k-1}\|_{L^2(Q_T)}}{\|y_{k-1}\|_{L^2(Q_T)}}$	$\frac{\|f_k - f_{k-1}\|_{L^2(q_T)}}{\|f_{k-1}\|_{L^2(q_T)}}$	$\|y_k\|_2$	$\|f_k\|_{2,q_T}$	$\sqrt{2E(y_k)}$	λ_k
0	–	–	45.28	43.91	9.31×10^{-1}	0.534
1	8.41×10^{-1}	1.23×10^{-2}	35.8908	38.76	1.12×10^{-1}	0.591
2	1.93×10^{-1}	2.91×10^{-3}	36.7302	38.92	3.40×10^{-2}	0.701
3	3.65×10^{-2}	1.01×10^{-3}	37.0919	39.12	6.12×10^{-3}	0.812
4	1.12×10^{-2}	2.69×10^{-4}	37.2124	40.01	1.12×10^{-3}	0.881
5	3.23×10^{-4}	4.23×10^{-5}	37.2426	40.04	2.13×10^{-4}	0.912
6	1.27×10^{-5}	6.23×10^{-6}	37.2518	40.05	3.05×10^{-5}	0.999
7	5.09×10^{-6}	8.12×10^{-7}	37.2520	40.05	2.10×10^{-6}	0.999
8	7.40×10^{-8}	8.21×10^{-9}	37.2520	40.05	5.10×10^{-9}	1.

Table 6.2: $\beta = 10^2$; Results for the algorithm (6.36).

♯iterate k	$\frac{\|y_k-y_{k-1}\|_{L^2(Q_T)}}{\|y_{k-1}\|_{L^2(Q_T)}}$	$\frac{\|f_k-f_{k-1}\|_{L^2(q_T)}}{\|f_{k-1}\|_{L^2(q_T)}}$	$\|y_k\|_2$	$\|f_k\|_{2,q_T}$	$\sqrt{2E(y_k)}$	λ_k
0	–	–	452.80	439.18	9.809×10^{-1}	0.4215
1	8.21×10^{-1}	6.00×10^{-1}	320.12	330.15	8.536×10^{-1}	0.3919
2	6.19×10^{-1}	3.29×10^{-2}	324.02	334.12	8.012×10^{-1}	0.1566
3	4.18×10^{-1}	1.37×10^{-2}	325.65	338.21	7.953×10^{-1}	0.1767
4	3.11×10^{-2}	1.34×10^{-2}	326.11	340.12	7.851×10^{-1}	0.0937
5	2.98×10^{-2}	5.85×10^{-3}	326.35	342.24	7.688×10^{-2}	0.0491
6	3.37×10^{-2}	7.00×10^{-3}	326.91	344.65	7.417×10^{-2}	0.1296
7	4.17×10^{-2}	9.69×10^{-3}	327.23	346.12	6.864×10^{-2}	0.1077
8	2.89×10^{-2}	8.09×10^{-3}	327.42	347.19	6.465×10^{-2}	0.0859
9	1.09×10^{-2}	6.40×10^{-3}	327.49	347.29	6.182×10^{-2}	0.0968
10	1.02×10^{-2}	6.72×10^{-3}	327.92	347.38	5.805×10^{-2}	0.1184
11	6.32×10^{-3}	6.91×10^{-3}	328.13	347.41	5.371×10^{-2}	0.1730
12	5.53×10^{-3}	7.41×10^{-3}	328.16	347.43	4.825×10^{-2}	0.2579
13	4.32×10^{-3}	8.22×10^{-3}	328.19	347.45	4.083×10^{-2}	0.3817
14	2.13×10^{-3}	8.14×10^{-3}	328.21	347.48	3.164×10^{-2}	0.4946
15	3.57×10^{-3}	7.34×10^{-3}	328.22	347.50	2.207×10^{-2}	0.8294
16	1.01×10^{-3}	6.68×10^{-3}	328.25	347.51	1.174×10^{-2}	0.9845
17	5.68×10^{-4}	3.84×10^{-4}	328.26	347.51	2.191×10^{-3}	0.9999
18	2.14×10^{-4}	5.85×10^{-5}	328.26	347.52	4.674×10^{-5}	1.
19	3.21×10^{-6}	1.57×10^{-7}	328.27	347.52	5.843×10^{-7}	–

Table 6.3: $\beta = 10^3$; Results for the algorithm (6.36).

♯iterate k	$\frac{\|y_k-y_{k-1}\|_{L^2(Q_T)}}{\|y_{k-1}\|_{L^2(Q_T)}}$	$\frac{\|f_k-f_{k-1}\|_{L^2(q_T)}}{\|f_{k-1}\|_{L^2(q_T)}}$	$\|y_k\|_2$	$\|f_k\|_{2,q_T}$	$\sqrt{2E(y_k)}$
0	–	–	452.80	439.18	9.809×10^{-1}
1	9.76×10^{-1}	1.05	330.21	334.15	9.812×10^{-1}
2	1.02	1.11	344.37	336.12	1.356
3	1.27	1.13	366.92	338.23	4.319
4	1.18	1.25	406.06	343.12	4.799
5	1.01	1.14	481.53	405.03	13.131

Table 6.4: $\beta = 10^3$; Results for the algorithm (6.36) with $\lambda_k = 1$ for all k.

6.5. Conclusions and perspectives

We have constructed an explicit sequence of functions $(f_k)_k$ converging strongly in the $L^2(q_T)$ norm toward a null control for the semilinear heat equation $y_t - \Delta y + g(y) = f\,1_\omega$. The construction of the sequence is based on the minimization of a $L^2(0,T;H^{-1}(\Omega))$ least-squares functional. The use of a specific descent direction allows to achieve a global convergence (uniform with respect to the data and to the initial guess) with a super-linear rate related to the regularity of the nonlinear

function g. Experiment confirms the robustness of the approach. In this analysis, we have assumed in particular that the derivative g' of g is uniformly bounded in $\mathbb{R}$. This allows to get a uniform bound of the constant of the form $C(\Omega, \omega, T, \|g'(y)\|_\infty)$ appearing from the Carleman estimate (6.17). In order to remove this assumption and be able to consider super-linear function g (as in the seminal work [13] by Fernández-Cara and Zuazua, assuming that g is locally Lipschitz-continuous and the asymptotic behavior (6.6)), we need to refine the analysis and exploit the structure of the constant $C(\Omega, \omega, T, \|g'(y)\|_{L^\infty})$ (as done in [8] for the observability constant). This may allow, assuming the above hypotheses of [13], not only to recover the null controllability of (6.1) but also, to construct, within the algorithm (6.36), approximations of null controls.

We also emphasize that this least-squares approach is very general and may be used to address other PDEs. Following [19] devoted to the direct problem, one may notably study the applicability of the method to approximate controls for the Navier-Stokes system. We also mentioned the case of nonlinear wave equation studied in [24] making use of a fixed point strategy.

Bibliography

[1] V. Barbu, *Exact controllability of the superlinear heat equation*, Appl. Math. Optim., 42(1),73–89, 2000.

[2] F. Boyer, *On the penalised HUM approach and its applications to the numerical approximation of null-controls for parabolic problems*, In CANUM 2012, 41e Congrès National d'Analyse Numérique, volume 41 of ESAIM Proc., 15–58, 2013.

[3] T. Cazenave, A. Haraux, *An introduction to semilinear evolution equations*, volume 13 of Oxford Lecture Series in Mathematics and its Applications,k The Clarendon Press, Oxford University Press, New York, 1998.

[4] P. G. Ciarlet, *The finite element method for elliptic problems*, volume 40 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2002.

[5] J. M. Coron, *Control and nonlinearity*, volume 136 of Mathematical Surveys and Monographs, American Mathematical Society, Providence, RI, 2007.

[6] J. M. Coron, E. Trélat, *Global steady-state controllability of one-dimensional semilinear heat equations*, SIAM J. Control Optim., 43(2), 549–569, 2004.

[7] P. Deuflhard, *Newton methods for nonlinear problems*, volume 35 of Springer Series in Computational Mathematics, Springer, Heidelberg, 2011.

[8] T. Duyckaerts, X. Zhang, E. Zuazua, *On the optimality of the observability inequalities for parabolic and hyperbolic systems with potentials* Ann. Inst. H. Poincaré Anal. Non Linéaire, 25(1), 1–41, 2008.

[9] E. Fernández-Cara, *Null controllability of the semilinear heat equation*, ESAIM Control Optim. Calc. Var., 2, 87–103, 1997.

[10] E. Fernández-Cara, S. Guerrero, *Global Carleman inequalities for parabolic systems and applications to controllability*, SIAM J. Control Optim., 45(4), 1399–1446, 2006.

[11] E. Fernández-Cara, A. Münch, *Numerical null controllability of semi-linear 1-D heat equations: fixed point, least squares and Newton methods*, Math. Control Relat. Fields, 2(3):217–246, 2012.

[12] E. Fernández-Cara, A. Münch, *Strong convergent approximations of null controls for the 1D heat equation*, SeMA J., 61, 49–78, 2013.

[13] E. Fernández-Cara, E. Zuazua, *Null and approximate controllability for weakly blowing up semilinear heat equations* Ann. Inst. H. Poincaré Anal. Non Linéaire, 17(5), 583–616, 2000.

[14] A. V. Fursikov, O. Yu. Imanuvilov, *Controllability of evolution equations*, volume 34 of Lecture Notes Series, Seoul National University, Research Institute of Mathematics, Global Analysis Research Center, Seoul, 1996.

[15] F. Hecht., *New development in Freefem++*, J. Numer. Math., 20(3-4), 251–265, 2012.

[16] K. Le Balc'h, *Global null-controllability and nonnegative-controllability of slightly superlinear heat equations*, J. Math. Pures Appl. (9), 135, 103–139, 2020.

[17] J. Lemoine, A. Münch, I. Marín-Gayte, *Approximation of null controls for semilinear heat equations using a least-squares approach*, submited.

[18] J. Lemoine, A. Münch, P. Pedregal, *Analysis of continuous H^{-1}-least-squares approaches for the steady Navier-Stokes system*, to appear in Applied Mathematics and Optimization.

[19] J. Lemoine, A. Münch, *A fully space-time least-squares method for the unsteady Navier-Stokes system*, submited, arXiv:1909.05034.

[20] J. Lemoine, A. Münch, *Resolution of the implicit Euler scheme for the Navier-Stokes equation through a least-squares method*, to appear in Numerische Mathematik, 2020.

[21] A. Münch, P. Pedregal *Numerical null controllability of the heat equation through a least squares and variational approach*, European J. Appl. Math., 25(3), 277–306, 2014.

[22] A. Münch, D. A. Souza, *A mixed formulation for the direct approximation of L^2-weighted controls for the linear heat equation*, Adv. Comput. Math., 42(1), 85–125, 2016.

[23] P. Saramito, *A damped Newton algorithm for computing viscoplastic fluid flows*, J. Non-Newton. Fluid Mech., 238, 6–15, 2016.

[24] E. Zuazua, *Exact controllability for semilinear wave equations in one space dimension*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 10(1), 109–129, 1993.