ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

OPTIMIZACIÓN DE HIPERPARÁMETROS PARA UN PROBLEMA DE CONTROL ÓPTIMO NO CONVEXO

TRABAJO DE INTEGRACIÓN CURRICULAR PRESENTADO COMO REQUISITO PARA LA OBTENCIÓN DEL TÍTULO DE MATEMÁTICA

DANIELA DOMINIC RIERA SAMPEDRO

daniela.riera@epn.edu.ec

DIRECTOR: PEDRO MARTÍN MERINO ROSERO

pedro.merino@epn.edu.ec

DMQ, FEBRERO 2022

CERTIFICACIONES

Yo, DANIELA DOMINIC RIERA SAMPEDRO, declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

Daniela Dominic Riera Sampedro

Certifico que el presente trabajo de integración curricular fue desarrollado por Daniela Dominic Riera Sampedro, bajo mi supervisión.

Firmado electrónicamente por:
PEDRO MARTIN MERINO ROSERO

Pedro Martín Merino Rosero

DIRECTOR

DECLARACIÓN DE AUTORÍA

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el(los) producto(s) resultante(s) del mismo, es(son) público(s) y estará(n) a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

Daniela Dominic Riera Sampedro

Pedro Martín Merino Rosero

RESUMEN

Este trabajo está enfocado en la optimización de hiperparámetros para un problema de control óptimo no convexo. Los parámetros asociados a un problema de control óptimo (e.g. parámetros de Tickhonov) generalmente tienen un significado económico en el proceso de optimización. Por tanto escoger estos parámetros con un criterio de optimalidad, y no de manera arbitraria o empírica, es relevante para obtener soluciones “mejoradas". Por esta razón se plantea un problema binivel, para la optimización de los hiperparámetros de un problema de control óptimo no convexo, para el cual se estudiarán y determinarán sus condiciones de optimalidad.

Una vez planteado el problema binivel, primeramente trabajamos con el problema de *nivel inferior*, estableciendo una regularización mediante la cual se obtiene una familia de problemas de control óptimo (dependientes del parámetro de regularización) que aproxima al original. Luego, se establece la condición de optimalidad escalada de primer orden asociada al problema regularizado.

Dichas condiciones de optimalidad caracterizan las soluciones del problema de nivel inferior, las cuales son usadas para transformar al problema binivel en uno de un solo nivel. Finalmente, se determinan las condiciones SB-KKT que caracterizan el valor óptimo de los hiperparámetros de manera formal.

Palabras clave: control óptimo de EDP no convexo, problemas de optimización binivel, condiciones SB-KKT, hiperparámetros.

ABSTRACT

This work is focused on the optimization of hyperparameters for a non-convex optimal control problem. Associated Parameters to an optimal control problem (e.g. Tickhonov parameters) generally have economic meaning in the optimization process. Therefore choose these parameters with an optimality criterion, and not arbitrarily or empirically, is relevant to obtain solutions ïmproved". For this reason, we propose a bilevel problem, for the optimization of the hyperparameters of a non-convex optimal control problem, for which we will study and determine its optimality conditions.

Once the bilevel problem has been set, we first work with the lower level problem, establishing a regularization through which a family of optimal control problems is obtained (depending on the regularization parameter) that approximates the original. Then, the first-order scaled optimality condition associated with the regularized problem is established.

These optimality conditions characterize the solutions of the lower level problem, which are used to transform the bilevel problem into a single level problem. Finally, we will determine SB-KKT that characterize the optimal value of the hyperparameters in a formal way.

Keywords: optimal control of non-convex EDP, bilevel optimization problems, SB-KKT conditions, hyperparameters.

Índice general

Capítulo 1

Descripción del componente desarrollado

1.1. Objetivo general

Caracterizar las soluciones de un problema de optimización de hiperparámetros mediante las condiciones de optimalidad de tipo SB-KKT.

1.2. Objetivos específicos

1. Establecer la condición de optimalidad de primer orden escalada del problema de nivel inferior.
2. Determinar las condiciones SB-KKT asociadas al problema binivel.

1.3. Alcance

Existen varios estudios acerca de problemas de control óptimo no convexos similares al que se proponen en este documento, como por ejemplo [4] donde se propone una función de suavizamiento de tipo Huber para regularizar el problema de control óptimo no convexo. Además se prueba que la solución del problema regularizado aproxima la solución del problema original cuando el parámetro tiende al infinito. En [5] se realiza un estudio de un problema binivel que involucra penalizaciones en cuasinormas pero en dimensión finita.

Siguiendo algunas ideas de las investigaciones mencionadas, junto con la teoría desarrollada en [6] respecto al control óptimo de ecuaciones diferenciales parciales y de otras fuentes a las que se hará referencia cuando sea oportuno, estableceremos las condiciones de optimalidad binivel escaladas de Karush Kuhn Tucker (SB-KKT) asociadas al siguiente problema

$$(PB)\begin{cases} \underset{(u,y,\phi,\alpha,\beta)\in L^2\times H_0^1\times H_0^1\times\mathbb{R}\times\mathbb{R}}{\text{mín}} F(y,u) \\ \text{sujeto a} \\ \begin{cases} \underset{(y,u)}{\text{mín}} \dfrac{1}{2}\|y-y_d\|^2_{L^2(\Omega)} + \dfrac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + \beta\displaystyle\int_\Omega |u|^q dx \\ \text{sujeto a} \\ \begin{cases} -\Delta y + c_1 y &= u, \quad \text{en } \Omega, \\ y &= 0, \quad \text{sobre } \Gamma, \end{cases} \end{cases} \\ 0\leq\alpha\leq\hat{\alpha}, 0\leq\beta\leq\hat{\beta}. \end{cases}$$

El cual se trata de un problema binivel donde el problema de nivel inferior es un problema de control óptimo no convexo.

1.4. Ecuaciones en derivadas parciales elípticas

1.4.1. Problema de control óptimo no convexo

En el estudio de nuestro problema binivel, donde el problema de nivel inferior es un problema de control óptimo no convexo, se seguirán los lineamientos de [5] que consisten, a breves rasgos, en regularizar el problema de control óptimo, determinar las condiciones de optimalidad asociadas a este y reformular el problema binivel en uno de un solo nivel. Mediante este proceso finalmente se determinarán las condiciones SB-KKT de las soluciones óptimas.

1.4.2. Funciones de variación acotada

Para el estudio del problema de nivel inferior se considerará el espacio de las funciones de variación acotadas, $BV(\Omega)$, como en [2].

Definición. 1.4.1. *Sea Ω un dominio Lipschitz acotado. Una función $u \in L^1(\Omega)$ es una función de variación acotada si sus derivadas distribucionales $\partial_{x_i} u$, $1 \leq i \leq n$ pertencen al espacio de Banach de medidas regulares (reales) de Borel notadas por $\mathcal{M}(\Omega)$.*

En este contexto, el resultado más importante que utilizaremos es el siguiente:

Proposición. 1.4.1. *Toda sucesión acotada $(u_k)_{k\in\mathbb{N}}$ en $BV(\Omega)$ posee una subsucesión que converge débilmente * a algún $u \in BV(\Omega)$. La convergencia débil *, $u_k \overset{*}{\rightharpoonup} u$, implica que $u_k \to u$ fuertemente en $L^1(\Omega)$ y $\nabla u_k \overset{*}{\rightharpoonup} \nabla u$ en $\mathcal{M}(\Omega)^n$.*

1.4.3. Continuidad de los Operadores de Nemitski

De acuerdo con [1], decimos que f satisface (C) si

i) $s \to f(x,s)$ es continua en casi todo punto $x \in \Omega$;

ii) $x \to f(x,s)$ es medible para todo $s \in \mathbb{R}$.

Sea $p, q \geq 1$ y supongamos que

$$|f(x,s)| \leq a + b|s|^\alpha, \quad \alpha = \frac{p}{q}, \tag{1.1}$$

para algunas constantes $a, b > 0$.

El siguiente resultado corresponde al Teorema 2.2 de [1]:

Teorema. 1.4.1. *Sea $\Omega \subseteq \mathbb{R}^n$ acotado y supongamos que f satisface (C) y (1.1). Entonces el operador de Nemitski f es un mapa continuo de L^p en L^q.*

1.4.4. Condiciones de optimalidad escaladas

Adicionalmente, para obtener las condiciones de optimalidad escaladas, tanto para el problema de control óptimo, como para el problema de optimización de hiperparámetros se sigue la técnica desarrollada en [3], la cual consiste en tomar una sucesión acotada de elementos que satisfacen las condiciones de optimalidad de primer orden del problema de nivel inferior regularizado y mostrar que cualquier punto de acumulación de dicha sucesión verifica las condiciones de optimalidad escaladas de primer orden, la cual posteriormente reemplaza al problema de nivel inferior y nos permite reescribir nuevamente el problema binivel como uno de un solo nivel.

Capítulo 2

Metodología

2.1. Planteamiento del Problema

Sea $\Omega \subseteq \mathbb{R}^2$ un dominio acotado con frontera Γ de tipo Lipschitz. Para $q \in (0,1)$, consideramos la función

$$\begin{array}{ccc} L^2(\Omega) & \to & \mathbb{R} \\ u & \mapsto & \int_\Omega |u|^q. \end{array} \tag{2.1}$$

Nuestro problema de nivel inferior está dado por el siguiente problema de control óptimo que contiene un término de penalización de la forma (2.1):

$$(L) \begin{cases} \min\limits_{(y,u)} \dfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \dfrac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + \beta \displaystyle\int_\Omega |u|^q dx \\ \text{sujeto a} \\ \begin{array}{rcll} -\Delta y + c_1 y & = & u, & \text{en } \Omega, \\ y & = & 0, & \text{sobre } \Gamma. \end{array} \end{cases}$$

Donde $y \in H_0^1(\Omega)$, $u \in L^2(\Omega)$, $c_1 > 0$, $\alpha > 0$, $\beta > 0$. Así, el problema de optimización binivel viene dado por:

$$
(PB)\begin{cases}
\displaystyle\min_{(u,y,\phi,\alpha,\beta)\in L^2\times H_0^1\times H_0^1\times\mathbb{R}\times\mathbb{R}} F(y,u)\\
\text{sujeto a}\\
\begin{cases}
\displaystyle\min_{(y,u)} \frac{1}{2}\|y-y_d\|^2_{L^2(\Omega)}+\frac{\alpha}{2}\|u\|^2_{L^2(\Omega)}+\beta\int_\Omega |u|^q dx\\
\text{sujeto a}\\
\begin{cases}
-\Delta y + c_1 y &=& u, & \text{en } \Omega,\\
y &=& 0, & \text{sobre } \Gamma,
\end{cases}
\end{cases}\\
0\leq\alpha\leq\hat{\alpha}, 0\leq\beta\leq\hat{\beta}.
\end{cases}
$$

Nota. En los espacios donde se omite el dominio se entenderá que este se trata de Ω (e.g. $L^1(\Omega)=L^1$).

2.1.1. Subproblema de control óptimo y su regularización

Para plantear el problema de un solo nivel se procede a reemplazar el problema de nivel inferior por su sistema de condiciones de optimalidad de primer orden. Dado que el funcional no es diferenciable, se debe realizar una regularización. En este caso, para $\varepsilon>0$ consideremos $f_\varepsilon:\mathbb{R}\to\mathbb{R}$ definido como $f_\varepsilon(t)=(t^2+\varepsilon^2)^{q/2}$. Así, f_ε aproxima a la función $|\cdot|^q$.

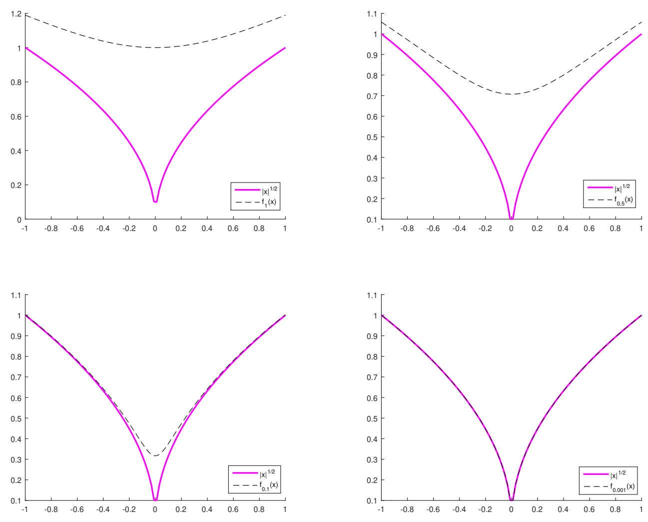

Figura 2.1: Tenemos las funciones $|\cdot|^q$ y f_ε con $q=0{,}5$ y variamos el parámetro ε. En la primera fila tomamos los valores 1 y $0{,}5$. En la segunda fila ε toma los valores $0{,}1$ y $0{,}001$, para este último vemos que la aproximación es mejor.

De este modo, introducimos la regularización de tipo Berkovier-Engelman, a la cual denotamos por Υ_ε definida para cada $\varepsilon > 0$ por:

$$u \mapsto \Upsilon_\varepsilon(u) = \int_\Omega (u(x)^2 + \varepsilon^2)^{q/2} dx.$$

Por tanto el problema de control óptimo se reformula de la siguiente forma:

$$(P_\varepsilon) \begin{cases} \min\limits_{(y,u)} \dfrac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \dfrac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + \beta\Upsilon_\varepsilon(u) \\ \text{sujeto a:} \\ \begin{aligned} -\Delta y + c_1 y &= u, \quad \text{en } \Omega, \\ y &= 0, \quad \text{sobre } \Gamma. \end{aligned} \end{cases}$$

Estudio de la ecuación de estado

Gracias al Teorema de Lax-Milgram, se tiene que para todo $w \in L^2(\Omega)$, existe un único $(y, \phi) \in H_0^1(\Omega) \times H_0^1(\Omega)$ tal que resuelve la ecuación de estado, es decir, $S_L u = (y, \phi)$. Con esto, se puede definir el operador control-estado S_L de la siguiente forma:

$$\begin{aligned} S_L : \quad L^2(\Omega) &\rightarrow H_0^1(\Omega) \times H_0^1(\Omega) \\ w &\mapsto S_L w = \begin{pmatrix} y \\ \phi \end{pmatrix}, \end{aligned} \tag{2.2}$$

tal que y satisface

$$\begin{aligned} -\Delta y + c_1 y &= u, \quad \text{en } \Omega, \\ y &= 0, \quad \text{sobre } \Gamma. \end{aligned}$$

Observemos que S_L es lineal y continuo.

Condiciones necesarias de primer orden

Lema. 2.1.1. *La función de regularización Υ_ε es Fréchet diferenciable y además, su derivada está dada por*

$$\Upsilon'_\varepsilon(u)[h] = (q(u^2 + \varepsilon^2)^{\frac{q-2}{2}} u, h)_{L^2}, \quad \forall h \in L^2(\Omega).$$

Demostración. En efecto, tenemos que:

$$\begin{aligned}
&\lim_{t\to 0}\frac{\Upsilon_\varepsilon(u+th)-\Upsilon_\varepsilon(u)}{t}\\
&=\lim_{t\to 0}\frac{1}{t}\left(\int_\Omega ((u(x)+th(x))^2+\varepsilon^2)^{\frac{q}{2}}dx-\int_\Omega (u(x)^2+\varepsilon^2)^{\frac{q}{2}}dx\right)\\
&=\lim_{t\to 0}\frac{1}{t}\int_\Omega \{((u(x)+th(x))^2+\varepsilon^2)^{\frac{q}{2}}-(u(x)^2+\varepsilon^2)^{\frac{q}{2}}\}dx.
\end{aligned}$$

Dado que la función

$$\begin{array}{rccc}
f_\varepsilon: & \mathbb{R} & \to & \mathbb{R}\\
 & z & \mapsto & (z^2+\varepsilon^2)^{\frac{q}{2}},
\end{array}$$

es de clase C^2, usando un desarrollo de Taylor alrededor de z, se tiene que

$$f_\varepsilon(z+th)-f_\varepsilon(z)=f'_\varepsilon(z)(z+th-z)+o(t^2)=q(z^2+\varepsilon^2)^{q/2-1}zth+o(t^2).$$

De lo anterior, para $h\in L^2(\Omega)$, se sigue que:

$$\begin{aligned}
\lim_{t\to 0}\frac{\Upsilon_\varepsilon(u+th)-\Upsilon_\varepsilon(u)}{t}&=\lim_{t\to 0}\frac{1}{t}\int_\Omega\left(q(u(x)^2+\varepsilon^2)^{\frac{q}{2}-1}u(x)th(x)+o(t^2)\right)dx\\
&=\lim_{t\to 0}\int_\Omega\left(q(u(x)^2+\varepsilon^2)^{\frac{q}{2}-1}u(x)h(x)+\frac{o(t^2)}{t}\right)dx\\
&=\int_\Omega q(u(x)^2+\varepsilon^2)^{\frac{q}{2}-1}u(x)h(x)dx\\
&=(q(u^2+\varepsilon^2)^{\frac{q-2}{2}}u,h)_{L^2(\Omega)}.
\end{aligned}$$

Por tanto, $\Upsilon'_\varepsilon(u)[h]$ es lineal y continua respecto de h. Notemos que:

$$\begin{aligned}
q(u^2+\varepsilon^2)^{q/2-1}u&=\frac{qu}{(u^2+\varepsilon^2)^{1-q/2}}\\
&\leq\frac{q|u|}{(\varepsilon^2)^{1-q/2}}.
\end{aligned}$$

Con lo cual, (1.1) se verifica y por el Teorema 1.4.1 concluimos que $u\mapsto q(u^2+\varepsilon^2)^{q/2-1}u$ es continua de L^2 en L^2. Esto muestra que Υ_ε es Gateux diferenciable en $u\in L^2(\Omega)$.

Ahora verifiquemos que Υ_ε es diferenciable en el sentido de Fréchet.

En efecto, tenemos que:

$$\begin{aligned}
&|\Upsilon_\varepsilon(u+h) - \Upsilon_\varepsilon(u) - (q(u^2+\varepsilon^2)^{\frac{q-2}{2}}u, h)_{L^2}| \\
=&\left|\int_\Omega [f(u(x)+h(x)) - f(u(x)) - q(u(x)^2+\varepsilon^2)^{\frac{q}{2}-1}u(x)h(x)]dx\right| \\
\leq&\int_\Omega \left| q(u(x)^2+\varepsilon^2)^{\frac{q}{2}-1}u(x)h(x) + \frac{1}{2}q\left(\frac{q}{2}-1\right)(u(x)^2+\varepsilon^2)^{\frac{q}{2}-2}2u(x)^2h(x)^2 \right. \\
&\left. +\frac{1}{2}q(u(x)^2+\varepsilon^2)^{\frac{q}{2}-1}h(x)^2 - q(u(x)^2+\varepsilon^2)^{\frac{q}{2}-1}u(x)h(x)\right| dx \\
\leq&\int_\Omega \left| q\left(\frac{q}{2}-1\right)(u(x)^2+\varepsilon^2)^{\frac{q}{2}-2}u(x)^2h(x)^2 + \frac{1}{2}q(u(x)^2+\varepsilon^2)^{\frac{q}{2}-1}h(x)^2\right| dx \\
=&\int_\Omega \left| q\left(\frac{q}{2}-1\right)\frac{u(x)^2h(x)^2}{(u(x)^2+\varepsilon^2)^{2-\frac{q}{2}}} + \frac{1}{2}q\frac{h(x)^2}{(u(x)^2+\varepsilon^2)^{1-\frac{q}{2}}}\right| dx \\
\leq&\int_\Omega \left| \left(q\left(\frac{q}{2}-1\right)\frac{u(x)^2+\varepsilon^2}{(u(x)^2+\varepsilon^2)^{2-\frac{q}{2}}} + \frac{q}{2(u(x)^2+\varepsilon^2)^{1-\frac{q}{2}}}\right) h(x)^2\right| dx \\
=&\int_\Omega \left| \left(q\left(\frac{q}{2}-1\right)\frac{1}{(u(x)^2+\varepsilon^2)^{1-\frac{q}{2}}} + \frac{q}{2(u(x)^2+\varepsilon^2)^{1-\frac{q}{2}}}\right) h(x)^2\right| dx \\
=&\int_\Omega \left| \left(q\left(\frac{q}{2}-1\right) + \frac{q}{2}\right)\frac{1}{(u(x)^2+\varepsilon^2)^{1-\frac{q}{2}}} h(x)^2\right| dx.
\end{aligned}$$

De donde se sigue que:

$$\begin{aligned}
&|\Upsilon_\varepsilon(u+h) - \Upsilon_\varepsilon(u) - (q(u^2+\varepsilon^2)^{\frac{q-2}{2}}u, h)_{L^2}| \\
\leq&\int_\Omega \frac{q^2-q}{2\varepsilon^{2\left(1-\frac{q}{2}\right)}}h(x)^2dx \\
=&C\|h\|_{L^2}^2,
\end{aligned}$$

donde $C = \dfrac{q^2-q}{2\varepsilon^{2\left(1-\frac{q}{2}\right)}}$. En resumen, hemos obtenido que

$$0 \leq |\Upsilon_\varepsilon(u+h) - \Upsilon_\varepsilon(u) - (q(u^2+\varepsilon^2)^{\frac{q-2}{2}}u, h)_{L^2}| \leq C\|h\|_{L^2}^2.$$

Por tanto,

$$0 \leq \frac{|\Upsilon_\varepsilon(u+h) - \Upsilon_\varepsilon(u) - (q(u^2+\varepsilon^2)^{\frac{q-2}{2}}u, h)_{L^2}|}{\|h\|_{L^2}} \leq C\|h\|_{L^2}.$$

Tomando el límite cuando $\|h\|_{L^2} \to 0$ en la desigualdad anterior, se sigue que

$$\lim_{\|h\|\to 0} \frac{|\Upsilon_\varepsilon(u+h) - \Upsilon_\varepsilon(u) - (q(u^2+\varepsilon^2)^{\frac{q-2}{2}}u, h)_{L^2}|}{\|h\|_{L^2}} = 0.$$

Lo cual prueba que Υ_ε es Fréchet diferenciable y además su derivada está dada por

$$\Upsilon'_\varepsilon(u)[h] = (q(u^2+\varepsilon^2)^{\frac{q-2}{2}}u, h)_{L^2}, \quad \forall h \in L^2(\Omega).$$

$\square$

Teorema. 2.1.1. *Sea $(\bar{y},\bar{u})$ un mínimo local del problema (P_ε). Entonces, existe $\phi \in H_0^1(\Omega)$ tal que, junto con $(\bar{y},\bar{u})$ satisfacen el siguiente sistema de optimalidad:*

$$\begin{aligned} -\Delta\bar{y} + c_1\bar{y} &= \bar{u}, \quad \textit{en}\,\Omega, \\ \bar{y} &= 0, \quad \textit{sobre}\,\Gamma, \end{aligned} \tag{2.3}$$

$$\begin{aligned} -\Delta\phi + c_1\phi &= \bar{y} - \bar{y}_d, \quad \textit{en}\,\Omega, \\ \phi &= 0, \qquad\quad \textit{sobre}\,\Gamma, \end{aligned} \tag{2.4}$$

$$\alpha\bar{u} + \phi + \beta q\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}} = 0. \tag{2.5}$$

Demostración. Sea $(\bar{y},\bar{u})$ un mínimo local de (P_ε). Introduciendo el estado adjunto $\phi \in H_0^1(\Omega)$ como la solución de la ecuación de estado adjunta:

$$\begin{aligned} -\Delta\phi + c_1\phi &= \bar{y} - y_d, \quad \text{en } \Omega, \\ \phi &= 0, \qquad\quad \text{sobre } \Gamma. \end{aligned}$$

La cual se trata de un problema lineal elíptico y por el Teorema de Lax-Milgran garantizamos que su solución existe. Así, podemos escribir $\phi = S^*(\bar{y} - y_d)$, donde $S^* = S$.

A continuación, tenemos que el Lagrangiano está dado por:

$$\begin{aligned} \mathscr{L}: \quad H_0^1(\Omega)\times L^2(\Omega)\times H_0^1(\Omega) &\to \mathbb{R} \\ (\bar{y},\bar{u},\phi) &\mapsto \mathscr{L}(\bar{y},\bar{u},\phi) = J(\bar{u},\bar{y}) - \textstyle\int_\Omega(\nabla\bar{y}\cdot\nabla\phi + c_1\bar{y}\phi - \bar{u}\phi)dx, \end{aligned}$$

donde

$$J(\bar{u},\bar{y}) = \frac{1}{2}\|\bar{y}-y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|\bar{u}\|^2_{L^2(\Omega)} + \beta\Upsilon_\varepsilon(\bar{u}).$$

Tomando la derivada parcial del Lagrangiano con respecto del estado

$\bar{y}$ tenemos que:

$$\begin{aligned}\mathscr{L}_{\bar{y}}(\bar{y},\bar{u},\phi)[w] &= (\bar{y}-y_d,w) - \int_\Omega (\nabla w\cdot\nabla\phi + c_1 w\phi)dx\\ &= \int_\Omega (\bar{y}-y_d)wdx - \int_\Omega (\nabla\phi\cdot\nabla w + c_1\phi w)dx\\ &= 0, \quad \forall w\in H_0^1(\Omega).\end{aligned}$$

Lo cual implica que

$$\int_\Omega (\nabla\phi\cdot\nabla w + c_1\phi w)dx = \int_\Omega (\bar{y}-y_d)wdx, \quad \forall w\in H_0^1(\Omega).$$

Por otro lado, tomando la derivada parcial con respecto de $\bar{u}$ obtenemos que

$$\mathscr{L}_{\bar{u}}(\bar{y},\bar{u},\phi)[h] = \alpha(\bar{u},h)_{L^2} + \int_\Omega h\phi dx + \beta\Upsilon'_\varepsilon(\bar{u})[h]dx = 0, \quad \forall h\in L^2(\Omega),$$

de donde, gracias al Lema 2.1.1, se sigue que

$$\alpha\bar{u} + \phi + \beta q(\bar{u}^2+\varepsilon^2)^{q/2-1}\bar{u} = 0.$$

Ahora, tomando la derivada con respecto a ϕ se tiene que

$$\mathscr{L}_\phi(\bar{y},\bar{u},\phi)[\theta] = -\int_\Omega (\nabla\bar{y}\cdot\nabla\theta + c_1\bar{y}\theta - \bar{u}\theta)dx = 0, \quad \forall\theta\in H_0^1.$$

Lo cual nos indica que

$$\int_\Omega (\nabla\bar{y}\cdot\nabla\theta + c_1\bar{y}\theta)dx = \int_\Omega \bar{u}\theta dx, \quad \forall\theta\in H_0^1(\Omega).$$

En resumen, tenemos que el sistema de optimalidad de (P_ε) es el siguiente:

$$\int_\Omega (\nabla\bar{y}\cdot\nabla\theta + c_1\bar{y}\theta)dx = \int_\Omega u\theta dx, \quad \forall\theta\in H_0^1(\Omega), \tag{2.6}$$

$$\int_\Omega (\nabla\phi\cdot\nabla w + c_1\phi w)dx = \int_\Omega (\bar{y}-y_d)wdx, \quad \forall w\in H_0^1(\Omega), \tag{2.7}$$

$$\alpha\bar{u} + \phi + \beta q\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}} = 0.$$

Donde (2.6) es la formulación débil de:

$$\begin{aligned} -\Delta \bar{y} + c_1 \bar{y} &= \bar{u}, \quad \text{en } \Omega, \\ \bar{y} &= 0, \quad \text{sobre } \Gamma, \end{aligned}$$

y (2.7) es la formulación débil de:

$$\begin{aligned} -\Delta \phi + c_1 \phi &= \bar{y} - y_d, \quad \text{en } \Omega, \\ \phi &= 0, \qquad\quad \text{sobre } \Gamma. \end{aligned}$$

□

2.2. Análisis del problema de optimización de hiperparámetros

En esta sección estudiaremos la existencia de soluciones del problema de optimización de hiperparámetros y determinaremos las condiciones de optimalidad del problema de nivel inferior.

Usando las ideas en [5], reemplazamos la restricción del problema binivel (PB) por sus condiciones de optimalidad, obtenidas en el Teorema 2.1.1, así tenemos el siguiente problema de un solo nivel:

$$(P) \begin{cases} \min\limits_{(u,y,\phi,\alpha,\beta)\in L^2\times H_0^1\times H_0^1\times\mathbb{R}\times\mathbb{R}} F(y,u) \\ \text{sujeto a:} \\ \begin{aligned} -\Delta y + c_1 y &= u, \quad \text{en } \Omega, \\ y &= 0, \quad \text{sobre } \Gamma, \end{aligned} \\ \begin{aligned} -\Delta \phi + c_1 \phi &= y - y_d, \quad \text{en } \Omega, \\ \phi &= 0, \qquad\quad \text{sobre } \Gamma, \end{aligned} \\ \alpha u + \phi + \beta q \frac{u}{(u^2+\varepsilon^2)^{1-q/2}} = 0; \\ 0 \le \alpha \le \hat{\alpha}, 0 \le \beta \le \hat{\beta}. \end{cases}$$

2.2.1. Existencia de soluciones

Notemos que, de (2.5) se tiene que

$$\bar{u} = -\frac{1}{\alpha}\left(\phi + \beta q \frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right). \tag{2.8}$$

Reemplazando (2.8) en (2.3) obtenemos que

$$-\Delta\bar{y} + c_1 y = -\frac{1}{\alpha}\left(\phi + \beta q \frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right). \tag{2.9}$$

Así, el sistema de optimalidad de (P_ε) se reduce a lo siguiente:

$$\begin{aligned}
&-\Delta\bar{y} + c_1\bar{y} = -\frac{1}{\alpha}\left(\phi + \beta q \frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right),\\
&-\Delta\phi + c_1\phi = \bar{y} - y_d.
\end{aligned}$$

De donde se sigue que:

$$-\Delta\bar{y} + c_1\bar{y} + \frac{1}{\alpha}\phi = \left(\beta q \frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right), \tag{2.10}$$

$$-\Delta\phi + c_1\phi - \bar{y} = -y_d. \tag{2.11}$$

Lo cual podemos escribir de la siguiente forma

$$\begin{pmatrix} -\Delta + c_1 I & \frac{1}{\alpha} I \\ -I & -\Delta + c_1 I \end{pmatrix}\begin{pmatrix} \bar{y} \\ \phi \end{pmatrix} = \begin{pmatrix} -\frac{\beta q}{\alpha}\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}} \\ -y_d \end{pmatrix}. \tag{2.12}$$

De este modo, definimos

$$a: \quad H_0^1(\Omega)\times H_0^1(\Omega) \quad \to \quad \mathbb{R}, \tag{2.13}$$

tal que

$$\begin{aligned}
a((y,\phi),(v_1,v_2)) &= (\nabla y, \nabla v_1) + (c_1 y, v_1) + \frac{1}{\alpha}(\phi, v_1)\\
&\quad + (\nabla\phi, \nabla v_2) + (c_1\phi, v_2) - (y, v_2), \quad \forall (v_1,v_2)\in H_0^1\times H_0^1.
\end{aligned}$$

Tenemos que

$$a((y,\phi),(v_1,v_2)) = \ell(v_1,v_2), \quad \forall (v_1,v_2)\in H_0^1\times H_0^1,$$

donde
$$\ell(v_1, v_2) = \left(-\frac{\beta q}{\alpha}\frac{u}{(u^2+\varepsilon^2)^{1-q/2}}, v_1\right) + (-y_d, v_2).$$
Notemos que a es continua y coerciva. En efecto, mostremos que existe $c > 0$ tal que

$$a((y,\phi),(y,\phi)) \geq c\|(y,\phi)\|^2 = c(\|y\|^2_{H_0^1} + \|\phi\|^2_{H_0^1}), \quad \forall (y,\phi) \in H_0^1(\Omega) \times H_0^1(\Omega).$$

Tenemos que

$$\begin{aligned} a((y,\phi),(y,\phi)) = \|\nabla y\|^2_{L^2} + c_1\|y\|^2_{L^2} - \left(1 - \frac{1}{\alpha}\right)(y,\phi) \\ + \|\nabla \phi\|^2_{L^2} + c_1\|\phi\|^2_{L^2}. \end{aligned}$$

Por otro lado, se tiene que

$$\begin{aligned} \left(1 - \frac{1}{\alpha}\right)(y,\phi) &\leq \left|\frac{1}{\alpha} - 1\right| |(y,\phi)| \\ &\leq \left|\frac{1}{\alpha} - 1\right| \|y\|\|\phi\|, \end{aligned}$$

usando la desigualdad de Young se sigue que

$$\left(1 - \frac{1}{\alpha}\right)(y,\phi) \leq \left|\frac{1}{\alpha} - 1\right| (\|y\|^2_{L^2} + \|\phi\|^2_{L^2}).$$

Por tanto
$$-\left(1 - \frac{1}{\alpha}\right)(y,\phi) \geq -\left|\frac{1}{\alpha} - 1\right| (\|y\|^2_{L^2} + \|\phi\|^2_{L^2}).$$
Con lo cual obtenemos que

$$a((y,\phi),(y,\phi)) \geq \|\nabla y\|^2_{L^2} + \|\nabla \phi\|^2_{L^2} + c_1(\|y\|^2_{L^2} + \|\phi\|^2_{L^2}) - \frac{|1/\alpha - 1|}{2}(\|y\|^2_{L^2} + \|\phi\|^2_{L^2}).$$

Si $c_1 > \dfrac{\left|\frac{1}{\alpha} - 1\right|}{2}$, entonces

$$a((y,\phi),(y,\phi)) \geq \|\nabla y\|^2_{L^2} + \|\nabla \phi\|^2_{L^2}.$$

De donde, por la desigualdad de Poincaré, sabemos existe una constante c tal que
$$a((y,\phi),(y,\phi)) \geq c(\|y\|^2_{H_0^1} + \|\phi\|^2_{H_0^1}). \tag{2.14}$$

De esta manera, usando (2.10) y (2.11), (P) puede reescribirse del siguiente modo

$$
(PB2) \begin{cases} \displaystyle\min_{(u,y,\phi,\alpha,\beta)\in L^2\times H_0^1\times H_0^1\times\mathbb{R}\times\mathbb{R}} F(y,u) \\ \text{sujeto a:} \\ \begin{aligned} -\Delta y + c_1 y + \tfrac{1}{\alpha}\phi &= -\tfrac{\beta q}{\alpha}\left(\tfrac{u}{(u^2+\varepsilon^2)^{1-q/2}}\right), \\ -\Delta\phi + c_1\phi - y &= -y_d, \\ 0\le\alpha\le\hat{\alpha}, &\quad 0\le\beta\le\hat{\beta}. \end{aligned} \end{cases}
$$

Hipótesis. 2.2.1. *Supongamos que:*

1) F *es acotado inferiormente;*

2) $F(\cdot,\cdot) : BV(\Omega)\times L^1(\Omega)\to\mathbb{R}$ *es continua;*

3) $F(y,u)\to+\infty$ *si* $\|u\|_{BV}\to\infty$, $\forall y$.

Bajo estas hipótesis, consideremos una sucesión minimizante fáctible $(u_k, y_k, \phi_k, \alpha_k, \beta_k)_{k\in\mathbb{N}}\subseteq BV(\Omega)\times H_0^1\times H_0^1\times\mathbb{R}^2$. Entonces

$$
F(y_k,u_k)\to\inf\{F(y,u) : S_L u-(y,\phi)=0, 0\le\alpha\le\hat{\alpha}, 0\le\beta\le\hat{\beta}\}, \tag{2.15}
$$

cuando $k\to\infty$.

De la Hipótesis 3) se sigue que $(u_k)_{k\in\mathbb{N}}$ es acotada en BV, puesto que si no lo fuera entoces $\|u_k\|\to\infty$, así

$$
F(y_k,u_k)\to\infty,
$$

lo cual contradice (2.15).

Por otra parte notemos que

$$
-\frac{\beta q}{\alpha}\frac{u}{(u^2+\varepsilon^2)}\le\frac{\beta q}{\alpha(\varepsilon^2)^{1-q/2}}|u|,
$$

con lo cual (1.1) se verifica, y por el Teorema 1.4.1 concluimos que $u\mapsto -\dfrac{\beta q}{\alpha}\dfrac{u}{(u^2+\varepsilon^2)^{1-q/2}}$ es continuo de L^1 en L^1.

Además, de (2.14), para cada $k \in \mathbb{N}$ tenemos que:

$$\begin{aligned}
\|y_k\|_{H_0^1}^2 + \|\phi_k\|_{H_0^1}^2 &\leq a((y_k, \phi_k), (y_k, \phi_k)) \\
&= \left(-\frac{\beta q}{\alpha} \frac{u_k}{(u_k^2 + \varepsilon^2)^{1-q/2}}, y_k \right) + (y_d, \phi_k) \\
&\leq \tilde{c} \left(\frac{\beta q}{\alpha} \left\| \frac{u_k}{(u_k^2 + \varepsilon^2)^{1-q/2}} \right\|_{L^2} \right) \|y_k\| + \|y_d\| \|\phi_k\| \\
&\leq \frac{\tilde{c}}{\varepsilon^{2-q}} \frac{\beta q}{\alpha} \|u_k\|_{L^2} (\|y_k\| + \|\phi_k\|),
\end{aligned}$$

con $c_\varepsilon = \frac{\tilde{c}}{\varepsilon^{2-q}}$. Lo cual implica que $(y_k)_{k \in \mathbb{N}}, (\phi_k)_{k \in \mathbb{N}}$ son acotadas. En consecuencia, la sucesión $(u_k, y_k, \phi_k, \alpha_k, \beta_k)_{k \in \mathbb{N}}$ es acotada en $BV \times H_0^1 \times H_0^1 \times \mathbb{R}^2$.

Por tanto, existen subsucesiones tales que:

$$\begin{aligned}
u_k &\overset{*}{\rightharpoonup} u, && \text{en } BV, \\
y_k &\rightharpoonup y, && \text{en } H_0^1, \\
\phi_k &\rightharpoonup \phi, && \text{en } H_0^1, \\
\alpha_k &\to \alpha, && \text{en } \mathbb{R}, \\
\beta_k &\to \beta, && \text{en } \mathbb{R},
\end{aligned}$$

donde $(u, y, \phi, \alpha, \beta) \in BV \times H_0^1 \times H_0^1 \times \mathbb{R}^2$ es factible. En efecto, vemos que

$$\lim_{k \to \infty} a((y_k, \phi_k), (v_1, v_2)) = a((y, \phi), (v_1, v_2))$$

y

$$\lim_{k \to \infty} \left(-\frac{\beta q}{\alpha} \frac{u_k}{(u_k^2 + \varepsilon^2)^{1-q/2}}, v_1 \right) = \left(-\frac{\beta q}{\alpha} \frac{u}{(u^2 + \varepsilon^2)^{1-q/2}}, v_1 \right),$$

por tanto

$$a((y, \phi), (v_1, v_2)) = \left(-\frac{\beta q}{\alpha} \frac{u}{(u^2 + \varepsilon^2)^{1-q/2}}, v_1 \right) + (-y_d, v_2).$$

Además, para cada $k \in \mathbb{N}$, se tiene que:

$$\begin{aligned}
0 \leq \alpha_k \leq \hat{\alpha}, \\
0 \leq \beta_k \leq \hat{\beta},
\end{aligned}$$

tomando el límite cuando $k \to \infty$ en las desigualdades anteriores, se sigue que:

$$0 \leq \alpha \leq \hat{\alpha},$$
$$0 \leq \beta \leq \hat{\beta}.$$

Como $y_k \rightharpoonup y$, en H_0^1, entonces $y_k \to y$ en L^2. Con lo cual, finalmente tenemos que:

$$\begin{aligned}\text{ínf}\{F(y,u) : S_L u - (y,\phi) = 0, 0 \leq \alpha \leq \hat{\alpha}, 0 \leq \beta \leq \hat{\beta}\} &= \lim_{k\to\infty} F(y_k, u_k) \\ &= F(u,y).\end{aligned}$$

Por lo que concluimos que existe solución en $BV \times H_0^1 \times H_0^1 \times H_0^1 \times \mathbb{R}^2$.

2.2.2. Condiciones de optimalidad

Teorema. 2.2.1. *Sea $(\bar{u}, \bar{y}, \bar{\phi}, \bar{\alpha}, \bar{\beta})$ un mínimo local de $(PB2)$. Entonces existen $\eta_1, \eta_2 \in H_0^1(\Omega)$, $\lambda, \mu \in \mathbb{R}^2$ tales que se satisface el siguiente sistema de optimalidad:*

$$\begin{aligned}
&\partial_{\bar{u}} F(\bar{y}, \bar{u}) = \int_\Omega \frac{\bar{\beta} q}{\bar{\alpha}} \left(\frac{(\bar{u}^2 + \varepsilon^2)^{1-q/2} - (2-q)\bar{u}^2(\bar{u}^2+\varepsilon^2)^{-q/2}}{(\bar{u}^2+\varepsilon^2)^{2-q}} \right) \eta_1, \\
&-\Delta \eta_1 + c_1 \eta_1 - \eta_2 = \partial_{\bar{y}} F(\bar{y}, \bar{u}), \\
&-\Delta \eta_2 + c_1 \eta_2 + \frac{1}{\bar{\alpha}} \eta_1 = 0, \\
&\lambda^T((\bar{\alpha}, \bar{\beta}) - (\hat{\alpha}, \hat{\beta})) = 0, \\
&\mu^T(\bar{\alpha}, \bar{\beta}) = 0, \\
&-\Delta \bar{y} + c_1 \bar{y} + \frac{1}{\bar{\alpha}} \bar{\phi} = -\frac{\bar{\beta} q}{\bar{\alpha}} \left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}} \right), \\
&-\Delta \bar{\phi} + c_1 \bar{\phi} - \bar{y} = -y_d, \\
&0 \leq \bar{\alpha} \leq \hat{\alpha}, 0 \leq \bar{\beta} \leq \hat{\beta}, \\
&\lambda \geq 0, \mu \geq 0.
\end{aligned} \tag{2.16}$$

Demostración. Sea $(\bar{u}, \bar{y}, \bar{\phi}, \bar{\alpha}, \bar{\beta})$ un mínimo local de $(PB2)$. Introducimos la ecuación adjunta definida por:

$$\begin{pmatrix} -\Delta + c_1 I & -I \\ \frac{1}{\alpha} I & -\Delta + c_1 I \end{pmatrix} \begin{pmatrix} \eta_1 \\ \eta_2 \end{pmatrix} = \begin{pmatrix} \partial_{\bar{y}} F \\ 0 \end{pmatrix} \tag{2.17}$$

De los mismos argumentos usados para (2.12), por el Teorema de Lax-Milgram existen $(\eta_1, \eta_2) \in H_0^1(\Omega) \times H_0^1(\Omega)$. De este modo el Lagrangiano para $(PB2)$ está definido por:

$$\begin{aligned} \mathscr{L}((\bar{u}, \bar{y}, \bar{\phi}, \bar{\alpha}, \bar{\beta}), \eta_1, \eta_2, \lambda, \mu) = {} & F(\bar{y}, \bar{u}) \\ & + \int_\Omega \left(\nabla \bar{y} \nabla \eta_1 + c_1 \bar{y} \eta_1 + \frac{1}{\bar{\alpha}} \bar{\phi} \eta_1 + \frac{\bar{\beta} q}{\bar{\alpha}} \left(\frac{\bar{u}}{(\bar{u}^2 + \varepsilon^2)^{1-q/2}} \right) \eta_1 \right) dx \\ & + \int_\Omega \left(\nabla \bar{\phi} \nabla \eta_2 + c_1 \bar{\phi} \eta_2 - \bar{y} \eta_2 + y_d \eta_2 \right) dx \\ & + \lambda^T ((\bar{\alpha}, \bar{\beta}) - (\hat{\alpha}, \hat{\beta})) - \mu^T (\bar{\alpha}, \bar{\beta}). \end{aligned}$$

Primero, para la derivada parcial con respecto a $\bar{u}$, obtenemos que:

$$\begin{aligned} & \mathscr{L}_{\bar{u}}((\bar{u}, \bar{y}, \bar{\phi}, \bar{\alpha}, \bar{\beta}), \eta_1, \eta_2, \lambda, \mu) = \langle \partial_{\bar{u}} F(\bar{y}, \bar{u}), h \rangle \\ & + \int_\Omega \frac{\bar{\beta} q}{\bar{\alpha}} \left(\frac{h(\bar{u}^2 + \varepsilon^2)^{1-q/2} - (1 - q/2) \bar{u}^2 (\bar{u}^2 + \varepsilon^2)^{-q/2} 2h}{(\bar{u}^2 + \varepsilon^2)^{2-q}} \right) \eta_1 dx = 0, \quad \forall h \in L^2(\Omega). \end{aligned}$$

Por tanto

$$\langle \partial_{\bar{u}} F(\bar{y}, \bar{u}), h \rangle = \int_\Omega \frac{\bar{\beta} q}{\bar{\alpha}} \left(\frac{h(\bar{u}^2 + \varepsilon^2)^{1-q/2} - (1 - q/2) \bar{u}^2 (\bar{u}^2 + \varepsilon^2)^{-q/2} 2h}{(\bar{u}^2 + \varepsilon^2)^{2-q}} \right) \eta_1, \quad \forall h \in L^2(\Omega).$$

Tomando la derivada del Lagrangiano respecto a $\bar{y}$ tenemos que:

$$\mathscr{L}_{\bar{y}}((\bar{u}, \bar{y}, \bar{\phi}, \bar{\alpha}, \bar{\beta}), \eta_1, \eta_2, \lambda, \mu) = \langle \partial_{\bar{y}} F(\bar{y}, \bar{u}), w \rangle + \int_\Omega (\nabla w \nabla \eta_1 + c_1 w \eta_1 - w \eta_2) dx = 0, \quad \forall w \in H_0^1.$$

Lo cual implica que

$$\int_\Omega \nabla \eta_1 \nabla w + c_1 \eta_1 w - \eta_2 w = \langle \partial_{\bar{y}} F(\bar{y}, \bar{u}), w \rangle, \quad \forall w \in H_0^1. \tag{2.18}$$

Por otro lado, tomando la derivada parcial respecto a ϕ obtenemos que:

$$\mathscr{L}_{\bar{\phi}}((\bar{u}, \bar{y}, \bar{\phi}, \bar{\alpha}, \bar{\beta}), \eta_1, \eta_2, \lambda, \mu) = \int_\Omega \left(\frac{1}{\bar{\alpha}} \theta \eta_1 + \nabla \theta \nabla \eta_2 + c_1 \theta \eta_2 \right) dx = 0, \quad \forall \theta \in H_0^1(\Omega).$$

De donde se sigue que

$$\int_\Omega \nabla\eta_2\nabla\theta + c_1\eta_2\theta + \frac{1}{\alpha}\eta_1\theta = 0. \tag{2.19}$$

Ahora, derivando con respecto a $(\bar{\alpha},\bar{\beta})$ tenemos que:

$$\mathscr{L}_{(\bar{\alpha},\bar{\beta})}((\bar{u},\bar{y},\bar{\phi},\bar{\alpha},\bar{\beta}),\eta_1,\eta_2,\lambda,\mu) = \lambda - \mu$$
$$+\left(-\int_\Omega\left(\frac{1}{\bar{\alpha}^2}\bar{\phi}\eta_1 + \frac{\bar{\beta}q}{\bar{\alpha}^2}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1\right)dx, \int_\Omega \frac{q}{\bar{\alpha}}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1 dx\right) = 0.$$

Lo cual nos indica que

$$\lambda-\mu = \left(-\int_\Omega\left(\frac{1}{\bar{\alpha}^2}\bar{\phi}\eta_1 + \frac{\bar{\beta}q}{\bar{\alpha}^2}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1\right)dx, \int_\Omega \frac{q}{\bar{\alpha}}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1 dx\right)$$

Como $\lambda \geq 0, \mu \geq 0$, tomamos

$$\lambda = \left(-\int_\Omega\left(\frac{1}{\bar{\alpha}^2}\bar{\phi}\eta_1 + \frac{\bar{\beta}q}{\bar{\alpha}^2}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1\right)dx, \int_\Omega \frac{q}{\bar{\alpha}}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1 dx\right)_+ \in \mathbb{R}^2$$

y

$$\mu = -\left(-\int_\Omega\left(\frac{1}{\bar{\alpha}^2}\bar{\phi}\eta_1 + \frac{\bar{\beta}q}{\bar{\alpha}^2}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1\right)dx, \int_\Omega \frac{q}{\bar{\alpha}}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1 dx\right)_- \in \mathbb{R}^2.$$

Notemos que (2.18) es la formulación débil de

$$-\Delta\eta_1 + c_1\eta_1 - \eta_2 = \partial_{\bar{y}}F(\bar{y},\bar{u}) \tag{2.20}$$

y (2.19) es la formulación débil de

$$-\Delta\eta_2 + c_1\eta_2 + \frac{1}{\alpha}\eta_1 = 0. \tag{2.21}$$

Además, derivando con respecto a los multiplicadores recuperamos las restricciones de $(PB2)$.

En resumen, tomando:

$$\lambda = \left(-\int_\Omega\left(\frac{1}{\bar{\alpha}^2}\bar{\phi}\eta_1 + \frac{\bar{\beta}q}{\bar{\alpha}^2}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1\right)dx, \int_\Omega \frac{q}{\bar{\alpha}}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1 dx\right)_+ \in \mathbb{R}^2,$$
$$\mu = -\left(-\int_\Omega\left(\frac{1}{\bar{\alpha}^2}\bar{\phi}\eta_1 + \frac{\bar{\beta}q}{\bar{\alpha}^2}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1\right)dx, \int_\Omega \frac{q}{\bar{\alpha}}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right)\eta_1 dx\right)_- \in \mathbb{R}^2,$$

se verfica el sistema de optimalidad (2.16).

$\square$

2.3. Condiciones SB-KKT

En esta sección presentamos las condiciones de optimalidad binivel escaladas KKT (SB-KKT) para nuestro problema de optimización de hiperparámetros. Estas condiciones pueden ser vistas como una extensión de las condiciones de optimalidad escaladas de primer orden introducidas en [3].

Además verificaremos que dichas condiciones no son más que condiciones de optimalidad necesarias para un problema de optimización de un solo nivel obtenido al reemplazar el problema de nivel inferior con sus condiciones de optimalidad escaladas de primer orden.

Empezamos considerando la regularización propuesta en [3], para (L), de la siguiente forma:

Para cada $\gamma > 0$, definimos $s_\gamma : \mathbb{R} \to \mathbb{R}$ de la siguiente manera

$$s_\gamma(t) = \begin{cases} |t|, & |t| > \gamma, \\ \frac{t^2}{2\gamma} + \frac{\gamma}{2}, & |t| \leq \gamma. \end{cases} \tag{2.22}$$

La función s_γ, así definida es continuamente diferenciable. De esta manera s_γ^q aproxima a la función $|\cdot|^q$, y además la derivada de s_γ^q está dada por

$$((s_\gamma(t))^q)' = \begin{cases} q|t|^{q-1} sgn(t), & |t| > \gamma, \\ q\left(\frac{t^2}{2\gamma} + \frac{\gamma}{2}\right)^{q-1} \frac{t}{\gamma}, & |t| \leq \gamma. \end{cases}$$

Sea $\psi_\gamma(t) = ((s_\gamma(t))^q)'$, para cada $t \in \mathbb{R}$.

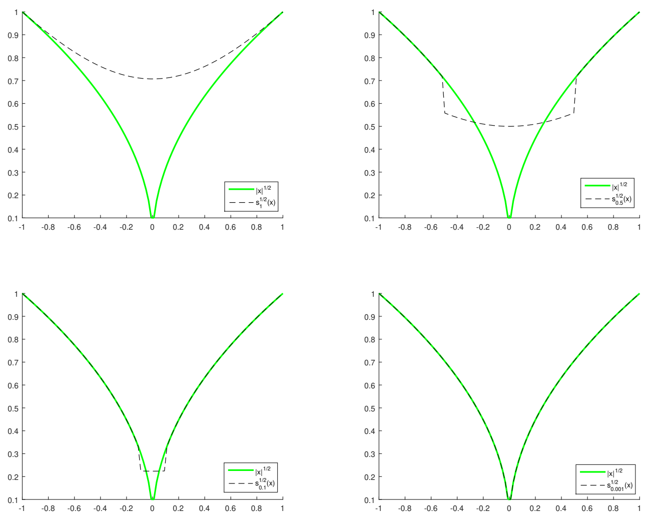

Figura 2.2: Tenemos las funciones $|\cdot|^q$ y s_γ^q con $q = 0{,}5$ y variamos el parámetro γ. En la primera fila tomamos los valores 1 y $0{,}5$. En la segunda fila γ toma los valores $0{,}1$ y $0{,}001$, para este último vemos que la aproximación es mejor.

Por tanto, tomamos Υ_γ definida para cada $\gamma > 0$ por

$$u \mapsto \Upsilon_\gamma(u) = \int_\Omega s_\gamma(u(x))^q dx.$$

Con lo cual definimos la aproximación suave de (L) como sigue

$$(L_\gamma)\begin{cases} \text{mín } f_\gamma(u) = \frac{1}{2}\|y - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + \beta\Upsilon_\gamma(u) \\ \text{sujeto a} \\ \begin{array}{rcll} -\Delta y & = & u, & \text{en } \Omega, \\ y & = & 0, & \text{sobre } \Gamma. \end{array} \end{cases}$$

Usando el operador control-estado $S_L : L^2(\Omega) \to H_0^1(\Omega) \times H_0^1(\Omega)$, definido en (2.2), obtenemos el problema reducido

$$(L_\gamma R)\left\{ \min_u f_\gamma(u) = \frac{1}{2}\|Su - y_d\|^2_{L^2(\Omega)} + \frac{\alpha}{2}\|u\|^2_{L^2(\Omega)} + \beta\Upsilon_\gamma(u). \right.$$

Lema. 2.3.1. *Sean* $\psi_\gamma(t) = \begin{cases} q|t|^{q-1}sgn(t), & |t| > \gamma, \\ q\left(\frac{t^2}{2\gamma} + \frac{\gamma}{2}\right)^{q-1}\frac{t}{\gamma}, & |t| \le \gamma, \end{cases}$ *y* $\eta_\gamma(t) = t\psi_\gamma(t)$. *En-*

tonces existen $a, b, \tilde{a}, \tilde{b} \in \mathbb{R}$ *tales que*

$$|\psi_\gamma(t)| \leq a + b|t|,$$

y

$$|\eta_\gamma(t)| \leq \tilde{a} + \tilde{b}|t|.$$

Demostración. Si $|t| > \gamma$, tenemos que

$$|\psi_\gamma(t)| = q|t|^{q-1}.$$

Por otro lado, notemos que

$$|\psi(t)| = \frac{q|t|^q}{|t|} \leq q\frac{1}{\gamma}|t|^q,$$

de donde, se sigue que

$$\ln(|\psi_\gamma(t)|) \leq \ln\left(q\frac{1}{\gamma}\right) + q\ln(|t|) \leq \ln\left(q\frac{1}{\gamma}\right) + \ln(|t|).$$

Así

$$|\psi_\gamma(t)| \leq e^{\ln(q/\gamma)+\ln(|t|)} \leq c_1 + \frac{q}{\gamma}|t|, \tag{2.23}$$

con $c_1 > 0$ cualquiera. En este caso también tenemos que

$$|\eta_\gamma(t)| = q|t|^q.$$

Luego,

$$\ln(|\eta_\gamma(t)|) = \ln(q) + q\ln(|t|) \leq \ln(q) + \ln(|t|),$$

de donde obtenemos que

$$|\eta_\gamma(t)| \leq e^{\ln(q)+\ln(|t|)} = q|t|.$$

Por tanto

$$|\eta_\gamma(t)| \leq c_2 + q|t|, \tag{2.24}$$

con $c_2 > 0$, cualquiera.

Si $|t| \leq \gamma$, entonces

$$|f(t)| = \left| q\left(\frac{t^2}{2\gamma} + \frac{\gamma}{2}\right)^{q-1} \frac{t}{\gamma} \right| = q \left| \left(\frac{t^2}{2\gamma} + \frac{\gamma}{2}\right)^{q-1} \right| \left| \frac{t}{\gamma} \right| \leq q \left| \left(\frac{t^2}{2\gamma} + \frac{\gamma}{2}\right)^{q-1} \right|.$$

Además notemos que

$$\frac{\gamma}{2} \leq \frac{t^2}{2\gamma} + \frac{\gamma}{2},$$

así

$$\left(\frac{t^2}{2\gamma} + \frac{\gamma}{2}\right)^{q-1} \leq \left(\frac{\gamma}{2}\right)^{q-1}.$$

Lo cual nos indica que

$$|\psi_\gamma(t)| \leq q \left| \left(\frac{\gamma}{2}\right)^{q-1} \right| \leq q \left| \left(\frac{\gamma}{2}\right)^{q-1} \right| + c_3|t|, \tag{2.25}$$

con $c_3 > 0$ cualquiera.

Notemos también que

$$|\eta_\gamma(t)| = \left| q\left(\frac{t^2}{2\gamma} + \frac{\gamma}{2}\right)^{q-1} \frac{t^2}{\gamma} \right|.$$

Dado que $|t| \leq \gamma$ se tiene que

$$\frac{t^2}{2\gamma} \leq \frac{\gamma}{2},$$

luego,

$$\frac{t^2}{2\gamma} + \frac{t^2}{2\gamma} \leq \frac{t^2}{2\gamma} + \frac{\gamma}{2}.$$

De aquí que

$$\left(\frac{t^2}{2\gamma} + \frac{\gamma}{2}\right)^{q-1} \leq \left(\frac{t^2}{2\gamma} + \frac{t^2}{2\gamma}\right)^{q-1} = \left(\frac{t^2}{\gamma}\right)^{q-1}.$$

En consecuencia

$$|\eta_\gamma(t)| \leq q \left| \left(\frac{t^2}{\gamma}\right)^{q-1} \frac{t}{\gamma} \right| = q \left| \left(\frac{t^2}{\gamma}\right)^{q} \right|.$$

Además
$$t^2 \leq \gamma |t|,$$
lo cual implica que
$$\frac{t^2}{\gamma} \leq |t|,$$
así
$$\left(\frac{t^2}{\gamma}\right)^q \leq |t|^q,$$
con lo cual
$$|\eta_\gamma(t)| \leq q|t|^q.$$
De donde tenemos que
$$\ln|\eta_\gamma(t)| = \ln q + q\ln(|t|) \leq \ln q + \ln t,$$
así
$$|\eta_\gamma(t)| \leq e^{\ln(q)+\ln(|t|)} = q|t|.$$
De este modo se obtiene que
$$|\eta_\gamma(t)| \leq c_4 + q|t|. \tag{2.26}$$

De (2.23) y (2.25), tomando $a := q\left|\left(\frac{\gamma}{2}\right)^{q-1}\right|$ y $b := \frac{q}{\gamma}$ se sigue que
$$|\psi_\gamma(t)| \leq a + b|t|.$$
De (2.24) y (2.26), tomando $\tilde{a} > 0$ cualquiera y $\tilde{b} := q$ obtenemos que
$$|\eta_\gamma(t)| \leq \tilde{a} + \tilde{b}|t|.$$
□

Lema. 2.3.2. *La función de regularización Υ_γ es Fréchet diferenciable y su derivada está dada por*
$$\Upsilon'_\gamma(u)[h] = (\psi_\gamma(u), h)_{L^2}, \quad \forall h \in L^2(\Omega).$$

Demostración. Sean $h \in L^2(\Omega)$, cualquiera. Notemos que:

$$\begin{aligned}
\frac{\Upsilon_\gamma(u+th)-\Upsilon_\gamma(u)}{t} &= \frac{1}{t}\int_\Omega [s_\gamma(u(x)+th(x))^q - s_\gamma(u(x))^q]dx \\
&= \frac{1}{t}\int_\Omega [\psi_\gamma(u(x))th(x) + o(t^2)]dx \\
&= \int_\Omega \left(\psi_\gamma(u(x))h(x) + \frac{o(t^2)}{t}\right) dx.
\end{aligned}$$

Tomando el límite cuando $t \to 0$ en esta última igualdad, obtenemos que

$$\lim_{t\to 0} \frac{\Upsilon_\gamma(u+th)-\Upsilon_\gamma(u)}{t} = (\psi_\gamma(u), h)_{L^2}, \quad \forall h \in L^2(\Omega).$$

Adicionalmente, del Lema 2.3.1 sabemos existen $a, b > 0$ tales que

$$|\psi_\gamma(t)| \le a + b|t|,$$

lo cual muestra que (1.1) se verifica, por tanto se cumple el Teorema 1.4.1, garantizando que ψ_γ es continua de L^2 en L^2. Por lo tanto, Υ_γ es Gateux diferenciable en $u \in L^2(\Omega)$. Ahora probemos que Υ_γ es diferenciable en el sentido de Fréchet. Para cada $h \in L^2(\Omega)$, tenemos que

$$\begin{aligned}
0 \le &|\Upsilon_\gamma(u+h) - \Upsilon_\gamma(u) - \Upsilon'_\gamma(u)[h]| \\
\le &\int_\Omega |\psi_\gamma(u(x))h(x) + O(h^2) - \psi_\gamma(u(x))h(x)|dx,
\end{aligned}$$

tomando el límite cuando $\|h\| \to 0$ en la desigualdad anterior, se sigue que

$$\lim_{\|h\|\to 0} \frac{|\Upsilon_\gamma(u+h) - \Upsilon_\gamma(u) - \Upsilon'_\gamma(u)[h]|}{\|h\|_{L^2}} = 0,$$

Lo cual prueba que Υ_γ es Fréchet diferenciable y su derivada es

$$\Upsilon_\gamma(u)[h] = (\psi_\gamma(u), h)_{L^2}, \quad \forall h \in L^2(\Omega).$$

□

Dado que $BV(\Omega) \hookrightarrow L^2(\Omega)$, con una inyección continua (en dos dimensiones [2]), entonces se obtiene el siguiente resultado:

Teorema. 2.3.1. *Sea $\bar{u} \in BV(\Omega)$ un mínimo local de f. Decimos que $\bar{u}$*

satisface la condición necesaria de primer orden de $(L_\gamma R)$ si

$$\phi + \alpha\bar{u} + \beta\psi_\gamma(\bar{u}) = 0. \tag{2.27}$$

Demostración. Sea $\bar{u} \in BV$ un mínimo local de f. Tenemos que las condiciones necesarias de primer orden para un mínimo local de f, están dadas por $f'_\gamma(\bar{u}) = 0$. Es decir,

$$f'_\gamma(\bar{u})[h] = (S\bar{u} - y_\alpha, Sh)_{L^2} + \alpha(\bar{u}, h)_{L^2} + \beta(\psi_\gamma(\bar{u}), h)_{L^2}, \quad \forall h \in L^2(\Omega).$$

Como S es autoadjunto se tiene que:

$$\begin{aligned} f'_\gamma(\bar{u})[h] &= (S(S\bar{u} - y_\alpha) + \alpha\bar{u} + \beta\psi_\gamma(\bar{u}), h)_{L^2} \\ &= (\phi + \alpha\bar{u} + \beta\psi_\gamma(\bar{u}), h)_{L^2} \\ &= 0, \quad \forall h \in L^2(\Omega). \end{aligned}$$

Lo cual indica que $\bar{u}$ satisface la condición necesaria de primer orden (2.27).

□

Consideremos la siguiente condición necesaria de primer orden dada en [3], donde el escalamiento de la condición necesaria de primer orden usual está dado por la multiplicación del control por dicha condición.

Definición. 2.3.1. *Sea $\bar{u} \in BV(\Omega)$. Decimos que $\bar{u}$ satisface la condición escalada necesaria de primer orden de (L) si*

$$\bar{u}\phi + \alpha\bar{u}^2 + \beta q|\bar{u}|^q = 0. \tag{2.28}$$

Teorema. 2.3.2. *Sea $(\bar{u}_{\gamma_k})_{k\in\mathbb{N}} \subseteq BV(\Omega)$, con $\gamma_k > 0$, para cada $k \in \mathbb{N}$ tal que $\gamma_k \to 0$ cuando $k \to \infty$. Si $(\bar{u}_{\gamma_k})_{k\in\mathbb{N}}$ es acotada y además para cada $k \in \mathbb{N}$, $\bar{u}_{\gamma_k}$ satisface la condición necesaria de primer orden (2.27). Entonces cualquier punto de acumulación de $(\bar{u}_{\gamma_k})_{k\in\mathbb{N}}$ satisface la condición necesaria de primer orden (2.28).*

Demostración. Dado que $(\bar{u}_{\gamma_k})_{k\in\mathbb{N}} \subseteq BV(\Omega)$ es una sucesión acotada, esta admite una subsucesión $(\bar{u}_{k_l})_{l\in\mathbb{N}}$ tal que

$$\bar{u}_{k_l} \overset{*}{\rightharpoonup} \bar{u},$$

con $\bar{u} \in BV(\Omega)$. Gracias a la Proposición 1.4.1, se sigue que

$$\bar{u}_{\gamma_{k_l}} \to \bar{u}, \quad \text{en } L^1.$$

Nombrando $(\bar{u}_{\gamma_k})_{k\in\mathbb{N}}$ a $(\bar{u}_{\gamma_{k_l}})_{l\in\mathbb{N}}$, obtenemos que

$$\bar{u}_{\gamma_k} \to \bar{u}, \quad \text{en } L^1 \quad \text{y} \quad \bar{u}_{\gamma_k} \rightharpoonup \bar{u}, \quad \text{en } L^2.$$

Notemos por $(y_{\gamma_k})_{k\in\mathbb{N}}$ y $(\phi_{\gamma_k})_{k\in\mathbb{N}}$ las sucesiones de los estados y estados adjuntos asociados que satisfacen:

$$\begin{aligned}
-\Delta y_{\gamma_k} + c_1 y_{\gamma_k} &= -\frac{1}{\alpha}\left(\phi_k + \beta q \frac{\bar{u}_{\gamma_k}}{(\bar{u}_{\gamma_k}^2 + \varepsilon^2)^{1-q/2}}\right), \\
-\Delta \phi_{\gamma_k} + c_1 \phi_{\gamma_k} &= y_{\gamma_k} - y_d,
\end{aligned}$$

para cada $k \in \mathbb{N}$.

Además,

$$y_{\gamma_k} \rightharpoonup y, \quad \text{en } H_0^1(\Omega)$$

y

$$\phi_{\gamma_k} \rightharpoonup \phi, \quad \text{en } H_0^1(\Omega).$$

Como $H_0^1(\Omega) \hookrightarrow L^2(\Omega)$ con una inyección compacta, se tiene que

$$y_{\gamma_k} \to y, \quad \text{en } L^2(\Omega)$$

y

$$\phi_{\gamma_k} \to \phi, \quad \text{en } L^2(\Omega),$$

respectivamente.

Por otro lado, de (2.27) se tiene que

$$\bar{u}_{\gamma_k}(x)\phi_{\gamma_k}(x) + \alpha\bar{u}_{\gamma_k}(x)^2 + \beta\bar{u}_{\gamma_k}(x)\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x)) = 0, \quad \text{a.e } x \in \Omega, \forall k \in \mathbb{N}.$$

Por lo cual tenemos que

$$\int_{\Omega} |\bar{u}_{\gamma_k}(x)\phi_{\gamma_k}(x) + \alpha\bar{u}_{\gamma_k}(x)^2 + \beta\bar{u}_{\gamma_k}(x)\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x))|dx = 0 < \infty, \quad \forall k \in \mathbb{N}. \quad (2.29)$$

Es decir,

$$\bar{u}_{\gamma_k}\phi_{\gamma_k} + \alpha\bar{u}_{\gamma_k}^2 + \beta\bar{u}_{\gamma_k}\psi_{\gamma_k}(\bar{u}_{\gamma_k}) \in L^1(\Omega),$$

para cada $k \in \mathbb{N}$.

Ahora, consideremos los siguientes casos:

i. Si $|\bar{u}_{\gamma_k}(x)| > \gamma$, entonces

$$\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x)) = q|\bar{u}_{\gamma_k}(x)|^{q-1} sgn(\bar{u}_{\gamma_k}(x)),$$

multiplicando esta última igualdad por $\bar{u}_{\gamma_k}$, se sigue que

$$\bar{u}_{\gamma_k}(x)\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x)) = q|\bar{u}_{\gamma_k}(x)|^{q-1} sgn(\bar{u}_{\gamma_k}(x))\bar{u}_{\gamma_k}(x) = q|\bar{u}_{\gamma_k}(x)|^{q-1}|\bar{u}_{\gamma_k}(x)|,$$

así

$$u_{\gamma_k}(x)\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x)) = q|\bar{u}_{\gamma_k}(x)|^q.$$

ii. Si $|\bar{u}_{\gamma_k}(x)| \leq \gamma_k$, entonces

$$0 \leq |\bar{u}_{\gamma_k}(x)| \leq \gamma_k. \quad (2.30)$$

Tomando el límite cuando $k \to \infty$ en (2.30), se tiene que

$$\lim_{k\to\infty} \bar{u}_{\gamma_k}(x) = 0. \quad (2.31)$$

En este caso, también tenemos que

$$\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x)) = q\left(\frac{\bar{u}_{\gamma_k}(x)^2}{2\gamma_k} + \frac{\gamma_k}{2}\right)^{q-1} \frac{\bar{u}_{\gamma_k}(x)}{\gamma_k}.$$

De donde, al multiplicar por u_{γ_k}, se sigue que

$$\bar{u}_{\gamma_k}(x)\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x)) = q\left(\frac{\bar{u}_{\gamma_k}(x)^2}{2\gamma_k} + \frac{\gamma_k}{2}\right)^{q-1} \frac{\bar{u}_{\gamma_k}(x)^2}{\gamma_k}. \quad (2.32)$$

Adicionalmente, como $|\bar{u}_{\gamma_k}(x)| \leq \gamma_k$, entonces

$$\frac{\bar{u}_{\gamma_k}(x)^2}{2\gamma_k} \leq \frac{\gamma_k}{2},$$

sumando $\dfrac{\bar{u}_{\gamma_k}(x)^2}{2\gamma_k}$ en ambos lados de la desigualdad anterior se obtiene que

$$\frac{\bar{u}_{\gamma_k}(x)^2}{2\gamma_k} + \frac{\bar{u}_{\gamma_k}(x)^2}{2\gamma_k} \leq \frac{\gamma_k}{2} + \frac{\bar{u}_{\gamma_k}(x)^2}{2\gamma_k}.$$

Por tanto,

$$\left(\frac{\gamma_k}{2} + \frac{\bar{u}_{\gamma_k}(x)^2}{2\gamma_k}\right)^{q-1} \leq \left(\frac{\bar{u}_{\gamma_k}(x)^2}{2\gamma_k} + \frac{\bar{u}_{\gamma_k}(x)^2}{2\gamma_k}\right)^{q-1} = \left(\frac{\bar{u}_{\gamma_k}(x)^2}{\gamma_k}\right)^{q-1}. \tag{2.33}$$

Así, junto con (2.32), se sigue que:

$$\begin{aligned} \bar{u}_{\gamma_k}(x)\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x)) &\leq q\left(\frac{\bar{u}_{\gamma_k}(x)^2}{\gamma_k}\right)^{q-1} \frac{\bar{u}_{\gamma_k}(x)^2}{\gamma_k} \\ &= q\left(\frac{\bar{u}_{\gamma_k}(x)^2}{\gamma_k}\right)^{q}. \end{aligned} \tag{2.34}$$

Nuevamente, usando el hecho de que $|u_{\gamma_k}(x)| \leq \gamma_k$, se obtiene

$$\frac{|\bar{u}_{\gamma_k}(x)|}{\gamma_k} \leq 1.$$

Multiplicando por $|\bar{u}_{\gamma_k}(x)|$ la desigualdad anterior se sigue que

$$\frac{|\bar{u}_{\gamma_k}(x)|^2}{\gamma_k} \leq |\bar{u}_{\gamma_k}(x)|,$$

luego

$$\left(\frac{\bar{u}_{\gamma_k}(x)^2}{\gamma_k}\right)^{q} \leq |\bar{u}_{\gamma_k}(x)|^q,$$

con lo cual, de (2.34) obtenemos

$$\bar{u}_{\gamma_k}(x)\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x)) \leq q|\bar{u}_{\gamma_k}(x)|^q.$$

De esta última desigualdad y de (2.31), se sigue que

$$\lim_{k\to\infty} \bar{u}_{\gamma_k}(x)\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x)) = 0, \quad \text{a.e } x \in \Omega.$$

Dado que $\bar{u}_{\gamma_k} \to \bar{u}$ en L^1, existe una subsucesión $(\bar{u}_{\gamma_k}(x))_{k\in\mathbb{N}}$ convergente a $\bar{u}(x)$ a.e $x \in \Omega$. De este modo tenemos que

$$\begin{aligned}
&\lim_{k\to\infty} (\bar{u}_{\gamma_k}(x)\phi_{\gamma_k}(x) + \alpha\bar{u}_{\gamma_k}(x)^2 + \beta\bar{u}_{\gamma_k}(x)\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x)) \\
&= \lim_{k\to\infty} \bar{u}_{\gamma_k}(x)\phi_{\gamma_k}(x) + \alpha \lim_{k\to\infty} \bar{u}_{\gamma_k}(x)^2 + \beta \lim_{k\to\infty} \bar{u}_{\gamma_k}(x)\psi_{\gamma_k}(\bar{u}_{\gamma_k}(x)) \\
&= \bar{u}(x)\phi(x) + \alpha\bar{u}(x)^2 + \beta q|\bar{u}(x)|^q \\
&= 0, \quad \text{a.e } x \in \Omega.
\end{aligned}$$

Por lo anterior y de (2.29), gracias al Teorema de Convergencia Dominada de Lebesgue, concluimos que

$$\lim_{k\to\infty} (\bar{u}_{\gamma_k}\phi_{\gamma_k} + \alpha\bar{u}_{\gamma_k}^2 + \beta\bar{u}_{\gamma_k}\psi_{\gamma_k}(\bar{u}_{\gamma_k})) = \bar{u}\phi + \alpha\bar{u}^2 + \beta q|\bar{u}|^q, \quad \text{en } L^1.$$

Lo cual muestra que $\bar{u}$ satisface (2.28).

$\square$

Como antes, obtenemos un problema de un solo nivel reemplazando el problema de nivel inferior en (PB) con la condición escalada de primer orden (2.28):

$$(SP)\begin{cases}
\min\limits_{(u,y,\phi,\alpha,\beta)\in BV(\Omega)\times H_0^1\times H_0^1\times\mathbb{R}^2} F(y,u) \\
\text{sujeto a} \\
\quad u\phi + \alpha u^2 + \beta q|u|^q = 0, \quad \text{en } L^1, \\
\quad (\nabla y, \nabla v_1) + c_1(y,v_1) = (u,v_1), \quad \forall v_1 \in H_0^1, \\
\quad (\nabla \phi, \nabla v_2) + c_1(y,v_2) = (y - y_d, v_2), \quad \forall v_2 \in H_0^1, \\
\quad 0 \le \alpha \le \hat{\alpha}, 0 \le \beta \le \hat{\beta}.
\end{cases}$$

Reformulamos (SP) de la siguiente forma

$$(SP2)\begin{cases}
\min\limits_{(u,(\alpha,\beta))\in BV(\Omega)\times H_0^1\times\mathbb{R}^2} F(Su,u) \\
\text{sujeto a} \\
\quad uS(Su - y_d) + \alpha u^2 + \beta q|u|^q = 0, \quad \text{en } L^1, \\
\quad 0 \le \alpha \le \hat{\alpha}, \\
\quad 0 \le \beta \le \hat{\beta}.
\end{cases}$$

Dado que el término $|\cdot|^q$ no es diferenciable, introducimos la regularización Υ_γ presentada anteriormente.

$$(SPR)\begin{cases} \min\limits_{(u,(\alpha,\beta))\in BV(\Omega)\times\mathbb{R}^2} F(Su,u) \\ \text{sujeto a} \\ \quad uS(Su-y_d)+\alpha u^2+\beta q\Upsilon_\gamma(u)=0, \quad \text{en } L^1, \\ \quad \alpha-\hat{\alpha}\le 0, \\ \quad \beta-\hat{\beta}\le 0, \\ \quad -\alpha\le 0, \\ \quad -\beta\le 0. \end{cases}$$

Teorema. 2.3.3. *Sea* $(\bar{u},(\bar{\alpha},\bar{\beta}))\in BV(\Omega)\times\mathbb{R}^2$ *un mínimo local de* (SPR)*, entonces existen* $\theta\in L^\infty(\Omega),\mu,\lambda\in\mathbb{R}^2$ *tales que se satisfacen las condiciones SB-KKT:*

$$\begin{aligned} &\partial_{\bar{u}}F(S\bar{u},\bar{u})=-\int_\Omega[2\bar{\alpha}\bar{u}+SS\bar{u}+\bar{u}SS-Sy_d+\bar{\beta}q\psi_\gamma(\bar{u})]\theta, \\ &\bar{u}S(S\bar{u}-y_d)+\bar{\alpha}\bar{u}^2+\bar{\beta}q\Upsilon_\gamma(\bar{u})=0, \\ &\mu^T((\bar{\alpha},\bar{\beta})-(\hat{\alpha},\hat{\beta}))=0, \\ &\lambda^T(\bar{\alpha},\bar{\beta})=0, \\ &0\le\bar{\alpha}\le\hat{\alpha},0\le\bar{\beta}\le\hat{\beta}, \\ &\lambda\ge 0,\mu\ge 0. \end{aligned} \tag{2.35}$$

Demostración. Sea $(\bar{u},(\bar{\alpha},\bar{\beta}))$ un mínimo local de (SPR). Formulamos el Lagrangiano para (SPR), con $\theta\in L^\infty(\Omega)$, $\mu,\lambda\in\mathbb{R}^2$.

$$\begin{aligned} \mathscr{L}(\bar{u},(\bar{\alpha},\bar{\beta}),\theta,\mu,\lambda)=F(S\bar{u},\bar{u})+ & \\ \int_\Omega(\alpha\bar{u}(x)^2\theta+\bar{u}(x)(SS\bar{u}(x))\theta-\theta\bar{u}(x)Sy_d(x)+\beta qs_\gamma(\bar{u}(x))^q\theta)dx & \\ +\mu^T((\bar{\alpha},\bar{\beta})-(\hat{\alpha},\hat{\beta}))-\lambda^T(\bar{\alpha},\bar{\beta}). & \end{aligned}$$

Por facilidad en la notación omitimos el punto de derivación. La derivada de $\mathscr{L}$ con respecto de $\bar{u}$ es:

$$\begin{aligned} \mathscr{L}_{\bar{u}}&=\langle\partial_{\bar{u}}F(S\bar{u},\bar{u}),h\rangle \\ &+\int_\Omega(2\bar{\alpha}\bar{u}(x)h(x)\theta+h(x)SS\bar{u}(x)\theta+\bar{u}(x)SSh(x)\theta-\theta h(x)Sy_d(x)+\bar{\beta}q\psi_\gamma(\bar{u}(x))h(x)\theta)dx \\ &=0,\forall h\in L^2(\Omega). \end{aligned}$$

De donde se sigue que

$$\langle \partial_{\bar{u}} F(S\bar{u}, \bar{u}), h \rangle = -\int_\Omega [2\bar{\alpha}\bar{u}h + hSS\bar{u} + \bar{u}SSh - hSy_d + \bar{\beta}q\psi_\gamma(\bar{u})h]\theta, \quad \forall h \in L^2(\Omega).$$

Ahora, tomando la derivada de $\mathscr{L}$ con respecto de $(\bar{\alpha}, \bar{\beta})$ obtenemos que:

$$\mathscr{L}_{(\bar{\alpha},\bar{\beta})} = \left(\int_\Omega \theta \bar{u}(x)^2 dx, \int_\Omega q s_\gamma(\bar{u}(x))^q \theta dx \right) + \mu - \lambda = 0.$$

Lo cual implica que

$$\lambda - \mu = \left(\int_\Omega \theta \bar{u}(x)^2 dx, \int_\Omega q s_\gamma(\bar{u}(x))^q \theta dx \right).$$

Como $\lambda \geq 0$, $\mu \geq 0$, de lo anterior tenemos que

$$\lambda = \left(\int_\Omega \theta \bar{u}(x)^2 dx, \int_\Omega q s_\gamma(\bar{u}(x))^q \theta dx \right)_+$$

y

$$\mu = -\left(\int_\Omega \theta \bar{u}(x)^2 dx, \int_\Omega q s_\gamma(\bar{u}(x))^q \theta dx \right)_-.$$

Por otra parte, derivando con respecto a los multiplicadores recuperamos las restricciones de (SPR).

En consecuencia $(\bar{u}, (\bar{\alpha}, \bar{\beta}))$ satisface el sistema de optimalidad SB-KKT (2.35).

$\square$

Capítulo 3

Resultados, conclusiones y recomendaciones

3.1. Resultados

A continuación se presenta un resumen de los resultados que obtuvimos en este trabajo.

Dado que la función objetivo del problema de nivel inferior de (PB) es no diferenciable, introducimos la regularización Υ_ε, para la cual se demostró el siguiente resultado:

Lema. 3.1.1. *La función de regularización Υ_ε es Fréchet diferenciable y además, su derivada está dada por*

$$\Upsilon'_\varepsilon(u)[h] = (q(u^2+\varepsilon^2)^{\frac{q-2}{2}}u, h)_{L^2}, \quad \forall h \in L^2(\Omega).$$

Luego, para obtener las condiciones de optimalidad asociadas a (P_ε) demostramos el siguiente teorema:

Teorema. 3.1.1. *Sea $(\bar{y}, \bar{u})$ un mínimo local del problema (P_ε). Entonces, existe $\phi \in H_0^1(\Omega)$ tal que, junto con $(\bar{y}, \bar{u})$ satisfacen el siguiente sistema de optimalidad:*

$$\begin{aligned} -\Delta\bar{y} + c_1\bar{y} &= \bar{u}, \quad \textit{en}\, \Omega, \\ \bar{y} &= 0, \quad \textit{sobre}\, \Gamma, \end{aligned}$$

$$\begin{aligned} -\Delta\phi + c_1\phi &= \bar{y} - \bar{y}_d, \quad \textit{en}\, \Omega, \\ \phi &= 0, \quad \textit{sobre}\, \Gamma, \end{aligned}$$

$$\alpha\bar{u}+\phi+\beta q\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}=0.$$

Luego de reescribir (P) como $(PB2)$ obtuvimos su sistema de optimalidad:

Teorema. 3.1.2. *Sea* $(\bar{u},\bar{y},\bar{\phi},\bar{\alpha},\bar{\beta})$ *un mínimo local de* $(PB2)$*. Entonces existen* $\eta_1,\eta_2\in H_0^1(\Omega)$*,* $\lambda,\mu\in\mathbb{R}^2$ *tales que se satisface el siguiente sistema de optimalidad:*

$$\begin{aligned}
&\partial_{\bar{u}}F(\bar{y},\bar{u})=\int_\Omega\frac{\beta q}{\alpha}\left(\frac{(\bar{u}^2+\varepsilon^2)^{1-q/2}-(1-q/2)\bar{u}^2(\bar{u}^2+\varepsilon^2)^{-q/2}2}{(\bar{u}^2+\varepsilon^2)^{2-q}}\right)\eta_1,\\
&-\Delta\eta_1+c_1\eta_1-\eta_2=\partial_{\bar{y}}F(\bar{y},\bar{u}),\\
&-\Delta\eta_2+c_1\eta_2+\frac{1}{\bar{\alpha}}\eta_1=0,\\
&\lambda^T((\bar{\alpha},\bar{\beta})-(\hat{\alpha},\hat{\beta}))=0,\\
&\mu^T(\bar{\alpha},\bar{\beta})=0,\\
&-\Delta\bar{y}+c_1\bar{y}+\frac{1}{\bar{\alpha}}\bar{\phi}=-\frac{\bar{\beta}q}{\bar{\alpha}}\left(\frac{\bar{u}}{(\bar{u}^2+\varepsilon^2)^{1-q/2}}\right),\\
&-\Delta\bar{\phi}+c_1\bar{\phi}-\bar{y}=-y_d,\\
&0\le\bar{\alpha}\le\hat{\alpha},0\le\bar{\beta}\le\hat{\beta},\\
&\lambda\ge 0,\mu\ge 0.
\end{aligned}$$

Para poder plantear las condiciones de optimalidad escaladas de primer orden asociadas a (SPR), nuevamente fue necesario utilizar una regularización, a la cual denotamos por Υ_γ y para la cual también demostramos su diferenciabilidad, haciendo uso de los siguientes resultados:

Lema. 3.1.2. *Sean* $\psi_\gamma(t)=\begin{cases} q|t|^{q-1}sgn(t), & |t|>\gamma,\\ q\left(\frac{t^2}{2\gamma}+\frac{\gamma}{2}\right)^{q-1}\frac{t}{\gamma}, & |t|\le\gamma,\end{cases}$ *y* $\eta_\gamma(t)=t\psi_\gamma(t)$*. Entonces existen* $a,b,\tilde{a},\tilde{b}\in\mathbb{R}$ *tales que*

$$|\psi_\gamma(t)|\le a+b|t|,$$

y

$$|\eta_\gamma(t)|\le\tilde{a}+\tilde{b}|t|.$$

Lema. 3.1.3. *La función de regularización* Υ_γ *es Fréchet diferenciable y su* *derivada está dada por*

$$\Upsilon_\gamma'(u)[h] = (\psi_\gamma(u), h)_{L^2}, \quad \forall h \in L^2(\Omega).$$

Gracias a lo anterior pudimos determinar la condición necesaria de primer orden asociada a $(L_\gamma R)$.

Teorema. 3.1.3. *Sea $\bar{u} \in BV(\Omega)$ un mínimo de f. Decimos que $\bar{u}$ satisface la condición necesaria de primer orden de $(L_\gamma R)$ si*

$$\phi + \alpha\bar{u} + \beta\psi_\gamma(\bar{u}) = 0. \tag{3.1}$$

Con lo cual se probó lo siguiente

Teorema. 3.1.4. *Sea $(\bar{u}_{\gamma_k})_{k\in\mathbb{N}} \subseteq BV(\Omega)$, con $\gamma_k > 0$, para cada $k \in \mathbb{N}$ tal que $\gamma_k \to 0$ cuando $k \to \infty$. Si $(\bar{u}_{\gamma_k})_{k\in\mathbb{N}}$ es acotada y además para cada $k \in \mathbb{N}$, $\bar{u}_{\gamma_k}$ satisface la condición necesaria de primer orden (2.27). Entonces cualquier punto de acumulación de $(\bar{u}_{\gamma_k})_{k\in\mathbb{N}}$ satisface la condición necesaria de primer orden (2.28).*

Finalmente, obtuvimos el sistema SB-KKT para (SPR).

Teorema. 3.1.5. *Sea $(\bar{u}, (\bar{\alpha}, \bar{\beta})) \in BV(\Omega) \times \mathbb{R}^2$ un mínimo local de (SPR), entonces existen $\theta \in L^\infty(\Omega), \mu, \lambda \in \mathbb{R}^2$ tales que se satisfacen las condiciones SB-KKT:*

$$\begin{aligned}
&\partial_{\bar{u}} F(S\bar{u}, \bar{u}) = -\int_\Omega [2\bar{\alpha}\bar{u} + SS\bar{u} + \bar{u}SS - Sy_d + \bar{\beta}q\psi_\gamma(\bar{u})]\theta, \\
&\bar{u}S(S\bar{u} - y_d) + \bar{\alpha}\bar{u}^2 + \bar{\beta}q\Upsilon_\gamma(\bar{u}) = 0, \\
&\mu^T((\bar{\alpha}, \bar{\beta}) - (\hat{\alpha}, \hat{\beta})) = 0, \\
&\lambda^T(\bar{\alpha}, \bar{\beta}) = 0, \\
&0 \leq \bar{\alpha} \leq \hat{\alpha}, 0 \leq \bar{\beta} \leq \hat{\beta}, \\
&\lambda \geq 0, \mu \geq 0.
\end{aligned}$$

3.2. Conclusiones y recomendaciones

3.2.1. Conclusiones

- Para determinar las condiciones SB-KKT para el problema binivel se tuvo que cambiar del espacio $L^2(\Omega)$ al espacio $BV(\Omega)$, dado que en este último, la convergencia débil implica la convergencia fuerte, propiedad que fue necesaria para obtener la condición de optimalidad de primer orden escalada asociada al problema de nivel inferior.

- Puesto que el problema binivel está definido en un espacio de dimensión infinita, donde la restricción no es diferenciable se utilizó una regularización adecuada, para el término no diferenciable del problema de control óptimo no convexo, que nos permitió determinar las condiciones SB-KKT.

- Para tratar el problema binivel, se reemplazó la restricción por sus condiciones de optimalidad. Sin embargo, al emplear esta técnica la región factible del nuevo problema puede ser más grande que la región factible del problema original. A pesar de que esto puede significar una modificación del problema original, es necesario, porque no se puede resolver el problema binivel tal cual se presenta inicialmente, además determinar el valor óptimo del hiperparámetro en un espacio más amplio puede garantizar una mejor elección.

- Como se menciona en [5] se espera que las condiciones SB-KKT pueden ser usadas para realizar análisis de convergencia.

3.2.2. Recomendaciones

- Se recomienda el estudio de la existencia del multiplicador $\theta \in L^\infty(\Omega)$ asociado a la restricción de igualdad del problema (SPR).

- Se podría realizar un estudio adicional sobre la obtención del sistema de optimalidad SB-KKT asociado al problema original a partir del sistema de optimalidad SB-KKT del problema regularizado haciendo tender el parámetro al infinito.

Referencias bibliográficas

[1] Antonio Ambrosetti and Giovanni Prodi. *A primer of nonlinear analysis*. Cambridge University Press, 1995.

[2] Eduardo Casas and Karl Kunisch. Analysis of optimal control problems of semilinear elliptic equations by bv-functions. *Set-Valued and Variational Analysis*, 27(2):355–379, 2019.

[3] Xiaojun Chen, Fengmin Xu, and Yinyu Ye. Lower bound theory of nonzero entries in solutions of ℓ_2-ℓ_p minimization. *SIAM Journal on Scientific Computing*, 32(5):2832–2852, 2010.

[4] Pedro Merino. A difference-of-convex functions approach for sparse pde optimal control problems with nonconvex costs. *Computational Optimization and Applications*, 74(1):225–258, 2019.

[5] Takayuki Okuno, Akiko Takeda, Akihiro Kawana, and Motokazu Watanabe. On ℓ_p-hyperparameter learning via bilevel nonsmooth optimization. *Journal of Machine Learning Research*, 22(245):1–47, 2021.

[6] Fredi Tröltzsch. *Optimal control of partial differential equations: theory, methods, and applications*, volume 112. American Mathematical Soc., 2010.