

ESCUELA POLITÉCNICA NACIONAL

FACULTAD DE CIENCIAS

OPTIMIZACIÓN BINIVEL DEL PARÁMETRO DE REGULARIZACIÓN CON
DEPENDENCIA ESPACIAL DEL MODELO DE VARIACIÓN TOTAL
GENERALIZADA PARA EL FILTRADO DE RUIDO EN IMÁGENES.

TRABAJO PREVIO A LA OBTENCIÓN DEL TÍTULO DE MAGÍSTER EN
OPTIMIZACIÓN MATEMÁTICA

TESIS

MARIBEL KATERYN HERRERA TERÁN

`kateryn.herrera@epn.edu.ec`

Director: DR. JUAN CARLOS DE LOS REYES BUENO

`juan.delosreyes@epn.edu.ec`

QUITO, JUNIO 2017

DECLARACIÓN

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

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

Maribel Kateryn Herrera Terán

CERTIFICACIÓN

Certifico que el presente trabajo fue desarrollado por MARIBEL KATERYN HERRERA TERÁN, bajo mi supervisión.

Dr. Juan Carlos De Los Reyes Bueno
Director del Trabajo

AGRADECIMIENTOS

A mis padres, por su esfuerzo y por creer en mí. A mis hermanos, por su alegría. Este trabajo es nuestro, fruto de las risas y lágrimas que juntos compartimos.

A mi director de tesis Juan Carlos De los Reyes, quien hizo posible mi participación en este proyecto de investigación, por sus palabras de aliento y por el tiempo dedicado a la supervisión de este trabajo.

A los colegas del Centro de Modelización Matemática por la oportunidad de ser parte del grupo de trabajo ModeMat.

Finalmente, al Instituto Instituto Nacional de Meteorología e Hidrología quien financió parcialmente este proyecto de titulación, dentro del marco del proyecto: *Sistema de Pronóstico del Clima y el Tiempo para todo el territorio ecuatoriano: Modelización numérica y estadística - fase dos*, a la Escuela Politécnica Nacional dentro del proyecto: *Identificación de vectores asociados a enfermedades infecciosas*.

DEDICATORIA

*A mis padres,
mis hermanos.*

A la pequeña Amelia.

A Edwin, por su invaluable apoyo.

Índice general

Resumen	VIII
Abstract	IX
1. Introducción	1
1.1. Métodos variacionales	2
1.1.1. Existencia de soluciones	4
1.2. Métodos de optimización	5
1.2.1. Método BFGS	6
1.2.2. Método de búsqueda lineal	6
1.3. Estudio del parámetro de regularización α	8
2. Modelo de variación total (TV)	11
2.1. Regularización de variación total	11
2.2. Planteamiento del problema	12
2.3. Análisis del problema binivel	13
2.4. Multiplicadores de Lagrange	14
2.4.1. Ecuación adjunta	18
2.4.2. Condición de optimalidad	18
2.5. Problema regularizado	20
2.5.1. Sistema de optimalidad	24
2.6. Tratamiento numérico	24
2.7. Implementación numérica	29
2.7.1. Experimento 1	36
2.7.2. Experimento 2	38
2.7.3. Experimento 3	40

3. Modelo de variación total generalizada (TGV)	42
3.1. Regularización de variación total generalizada	42
3.2. Planteamiento del problema	43
3.3. Análisis del problema binivel	44
3.3.1. Existencia de una solución óptima	45
3.4. Multiplicadores de Lagrange	47
3.4.1. Diferenciabilidad del operador solución	47
3.4.2. Ecuación adjunta	50
3.4.3. Condición de optimalidad	51
3.5. Problema regularizado	54
3.5.1. Sistema de optimalidad	57
3.6. Tratamiento numérico	58
3.7. Implementación numérica	65
3.7.1. Método de Newton Semi-Smooth	67
3.7.2. Filtrado de ruido de señales	72
3.7.3. Procesamiento de imágenes	75
4. Conclusiones	78
Bibliografía	80

Resumen

Nuestro objetivo es plantear y resolver un problema de optimización binivel no lineal en espacios funcionales, para determinar el valor óptimo de los parámetros con dependencia espacial del modelo de variación total (TV) y del modelo de variación total generalizada (TGV) para el filtrado de ruido en imágenes. El considerar la dependencia espacial de los parámetros nos permite filtrar ruido no uniforme en una imagen, lo que nos acerca más a situaciones reales donde no se conocen el tipo ni la distribución de ruido que puede cubrir información relevante.

Las dificultades que surgen de inmediato al tratar un problema binivel no lineal en espacios de dimensión infinita son la no diferenciabilidad de la restricción, que corresponde al problema de filtrado de ruido, lo que hace imposible verificar condiciones de calificación para sistemas de Karush-Kuhn-Tucker (KKT). Para tratar el problema de no diferenciabilidad empleamos una regularización de tipo Huber para los términos no diferenciables del modelo TV y TGV.

Los resultados analíticos que se presentan son la existencia de soluciones del problema de optimización de parámetros, la Fréchet diferenciabilidad del operador solución, lo que permite probar la existencia de multiplicadores de Lagrange. Además, la existencia del estado adjunto, que permite obtener una caracterización del gradiente del funcional de costo reducido. Se evidenció que los multiplicadores asociados a las restricciones de positividad son medidas, que computacionalmente hablando son muy costosas de implementar. Para superar esto se propuso introducir la *regularización de Moreau-Yosida*, donde se estableció el sistema de optimalidad asociado al problema regularizado y se mostró que las soluciones de los problemas regularizados convergen hacia la solución del problema original.

La estrategia numérica que se propone consiste en usar un método cuasi-Newton de segundo orden, específicamente el método BFGS, junto con los algoritmos de tipo Newton Semi-Smooth para la resolución del modelo de filtrado de ruido en el caso del modelo TGV y en el modelo TV el método de Newton clásico, estos algoritmos fueron globalizados utilizando una proyección que modifica la matriz hessiana y garantiza que sea definida positiva. Para los esquemas de discretización se empleó el método de aproximación por diferencias finitas, ampliamente utilizado en el procesamiento matemático de imágenes, debido a la estructura de la imagen digital.

Abstract

In this thesis, we study and solve a nonlinear bilevel optimization problem in function spaces. The goal is to determine the optimal spatially dependent regularization parameters in total variation (TV) and generalized total variation (TGV) image denoising models. Considering the spatial dependence of the parameters allows us to filter non-uniform noise in an image, which brings us closer to real situations where the type and distribution of noise are not known.

We present some analytical results like the existence of solutions of the problem of parameter optimization, the Fréchet differentiability of the solution operator, which allows to prove the existence of Lagrange multipliers. In addition, the existence of the adjoint state which allows to obtain a gradient characterization. The multipliers associated with the positivity constraints are regular Borel measures which are very difficult to compute. In order to overcome this issue, we proposed to introduce the Moreau-Yosida regularization, where the optimality system associated with the regularized problem was established and we prove that the solutions of regularized problems converge to the solution of the original one.

The proposed numerical strategy is to use a second-order quasi-Newton method, specifically the BFGS method, together with the Newton Semi-Smooth algorithms for the resolution of TGV image denoising model and in the case of total variation (TV) model we use a classical Newton method.

Capítulo 1

Introducción

En las dos últimas décadas, el procesamiento digital de imágenes ha ganado gran importancia dentro de distintas áreas de la ciencia, debido a que cada día la información visual es más importante y abundante. El procesamiento digital de imágenes es un conjunto de modelos y algoritmos que se aplican a imágenes deterioradas, con el objetivo de mejorar su calidad y facilitar la extracción de algún tipo de información, ya sea para la interpretación humana o de máquinas autónomas.

Entre los problemas más interesantes a tratar dentro del procesamiento de imágenes se encuentran: la deconvolución de imágenes borrosas, el escalado por la compresión de datos, la localización de objetos, la reconstrucción de partes deterioradas y el filtrado de ruido. Este último, se conoce como *denoising* y será el tema central de este trabajo.

El problema de *denoising* [2] consiste en quitar borrosidades fruto de la presencia de ruido, este fenómeno es inevitable y puede tener varias fuentes como: errores de transmisión, defectos del dispositivo, interferencia, entre otras. Se pueden identificar distintos tipos de ruido basados en el dispositivo que obtiene la imagen, algunos ejemplos son:

- Ruido gaussiano presente en tomografías por resonancia magnética (TRM).
- Ruido de poisson presente en tomografías por emisión de positrones (TEP).
- Ruido de impulso presente en sensores de cámaras.

La idea de una imagen digital es considerar una malla rectangular de píxeles y asignar a cada píxel un número, que representará el nivel de intensidad de gris o brillo en la imagen. Esta malla de píxeles puede ser tratada como una matriz de n filas y m columnas, que corresponderá a la representación matricial de una imagen. Para tener una idea, de las cámaras digitales se obtienen imágenes de tamaño 320×240 y pueden alcanzar hasta 3060×2036 en cámaras profesionales, por otro lado el tamaño de imágenes médicas de resonancia magnética está alrededor de 128×128 .

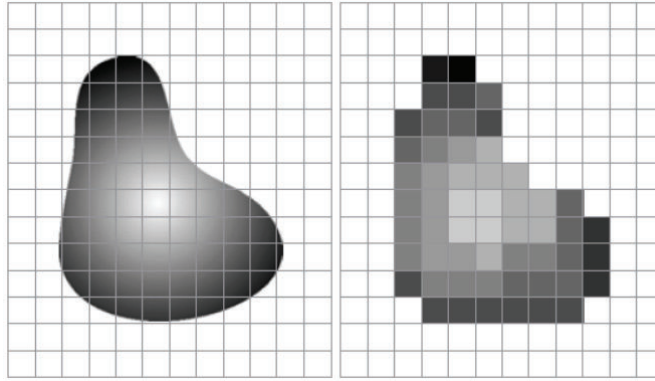


Figura 1.1. A la izquierda, la imagen continua y a la derecha la imagen digital. **Fuente:** [24]

A lo largo de los últimos años, se han desarrollado varios enfoques para abordar el procesamiento digital de imágenes. Se pueden mencionar los principales métodos basados en: la teoría del filtro, el análisis espectral, o en algunos conceptos básicos de probabilidad y estadística. Otro tipo de herramientas se basan en modelos estocásticos, en ecuaciones en derivadas parciales (EDPs), y en métodos variacionales. En este trabajo nos enfocaremos en los métodos basados en el cálculo de variaciones.

Para hacer uso de los modelos variacionales, es necesaria una representación continua de las imágenes digitales, que corresponderá a una interpolación de la imagen digital. Por lo que se considera a una imagen con dominio $\Omega \subset \mathbb{R}^2$, como una función acotada $f: \Omega \rightarrow \mathbb{R}^n$, es decir, $f \in L^\infty(\Omega)$. El dominio Ω más frecuente para una imagen f es $\Omega = (0, 1) \times (0, 1)$. Mientras que el rango dependerá del modelo de color; si la imagen considerada está en escala de grises el rango será \mathbb{R} , mientras que si es una imagen a color, considerando por ejemplo el modelo de colores RGB (Red-Green-Blue), se tiene \mathbb{R}^3 .

Varios modelos basados en métodos variacionales y en ecuaciones en derivadas parciales han sido planteados para el problema de denoising, entre los cuales se destacan: el modelo de variación total (TV) propuesto por Rudin, Osher y Fatemi [39], el modelo basado en la ecuación de calor [41], y el de, Bredies, Kunish y Pock [5] aplicando la *regularización de variación total generalizada TGV*, entre otros. En este trabajo nos concentraremos en el estudio de los modelos de variación total TV y TGV.

1.1. Métodos variacionales

Los métodos variacionales se pueden explicar mejor desde el punto de vista de los problemas inversos, donde dada una imagen f deteriorada por la presencia de ruido ϑ , se quiere hallar una imagen u sin daños, esto se describe en forma general por el siguiente esquema

$$f = u + \vartheta.$$

La presencia de la variable aleatoria ϑ que representa el ruido, da lugar a un problema *mal condicionado*. Hadamard [28] fue el primero en establecer los criterios del condicionamiento de un problema como éste.

Se dice que un problema está bien condicionado si se verifica que:

- **Existencia:** para cualquier f dado, existe una solución u .
- **Unicidad:** la solución u es única.
- **Estabilidad:** la solución u depende en forma continua del dato f .

La condición de estabilidad es constantemente quebrantada en los problemas inversos. Para superar este inconveniente se utilizan *métodos de regularización*. Existen varios enfoques, entre los principales están la regularización por filtrado, regularización a posteriori, regularización variacional y regularización iterativa. Nos enfocaremos en la regularización variacional, este método es también conocido como *método de energía* [2], el mismo que busca restaurar la estabilidad de la solución con relación a los datos. En términos generales, se puede decir que la regularización permite aproximar un problema mal condicionado por medio de una familia cercana de problemas bien condicionados. En forma general, están determinados por:

$$\min_u E[u] := \alpha R[u] + \phi(f, u). \quad (1.1)$$

- Al primer término se lo conoce como *término regularizador* o de *suavidad*. Este término cuantifica la variación dentro de una imagen, tratando que el gradiente de u varíe suavemente. Busca que las esquinas y los bordes de la imagen queden lo mejor definidas posibles, esto en el contexto de las imágenes. La elección de este término se basa en la información que se conoce previamente de la solución. Por ejemplo, si se conoce a priori que las soluciones a un cierto problema son discontinuas y constantes a trozos, conviene utilizar un término regularizador que garantice que las soluciones mantengan dichas características.
- El segundo, cuantifica el nivel de fidelidad entre la imagen solución u (imagen procesada) y la imagen original f (imagen defectuosa observada).
- El parámetro de regularización α , determina la influencia de los términos. Su valor es crítico para conseguir el equilibrio entre la estabilidad de la solución y la calidad de los datos. De hecho, si α es demasiado grande, se obtiene una solución sobre-regularizada y el suavizado en la imagen restaurada será excesivo. Por el contrario, si el parámetro α es demasiado pequeño se llegaría a la misma solución inestable por la presencia de ruido.

Para aplicar este método se parte de la hipótesis de que la imagen defectuosa f contiene suficiente información para ser restaurada, y así obtener un alto porcentaje de mejoras respecto a su estado original. Dependiendo de la elección del término de regularización $R[u]$ se han

desarrollado diversos modelos de filtrado de ruido. Nos enfocaremos en la regularización de variación total TV y la regularización de variación total generalizada TGV.

La elección de ϕ depende de la distribución de ruido, que está estrechamente relacionado con el dispositivo de adquisición o transmisión de la imagen. A continuación, se muestran algunos ejemplos de ϕ

- Ruido Gaussiano: $\phi(u, f) = (u - f)^2$,
- Ruido de Poisson : $\phi(u, f) = u - f \log(u)$,
- Ruido de Impulse: $\phi(u, f) = |u - f|$.

1.1.1. Existencia de soluciones

Nuestro primer objetivo es resolver el problema (1.1), para lo cual es necesario garantizar la existencia de una solución y posteriormente, especificar las condiciones de optimalidad asociadas. En general, para garantizar la existencia se utiliza el método directo del cálculo de variaciones [2], que consiste en: dado U un espacio de Banach y $E: U \rightarrow \mathbb{R}$ un funcional de energía, se define el problema de minimización:

$$\inf_{u \in U} E(u).$$

1. Se construye una *sucesión minimizante* $u_n \in U$, es decir, una sucesión que verifica

$$\lim_{n \rightarrow +\infty} E(u_n) = \inf_{u \in U} E(u).$$

2. Si E es *coerciva*, es decir,

$$\lim_{|u| \rightarrow +\infty} E(u) = +\infty,$$

se puede acotar uniformemente la sucesión, es decir, existe $C > 0$ tal que $|u_n|_U \leq C$. Se tienen dos casos:

- (i) Si U es reflexivo, entonces existe $u_0 \in U$ y una subsucesión u_{n_j} de u_n tal que $u_{n_j} \rightarrow u_0$, cuando $n \rightarrow \infty$.
- (ii) Si U es separable, entonces existe $u_0 \in U'$ y una subsucesión u_{n_j} de $u_n \in U'$ tal que $u_{n_j} \xrightarrow{*} u_0$, cuando $n \rightarrow \infty$.

3. Para probar que u_0 es un mínimo de E es suficiente garantizar la desigualdad

$$E(u_0) \leq \liminf_{u_{n_j} \rightarrow u_0} E(u_{n_j}),$$

la cual implica que $E(u_0) = \min_{u \in U} E(u)$.

Este esquema del cálculo de variaciones se empleará en los capítulos siguientes.

Una vez establecida la existencia de un mínimo, definimos a continuación las condiciones de optimalidad de primer orden. Si E es Gâteaux diferenciable y si el problema de optimización sin restricciones $\inf_{u \in U} E(u)$ tiene una solución u , entonces se verifica

$$E'(u) = 0.$$

Al contrario, si E es convexa, entonces una solución u de $E'(u) = 0$ es una solución del problema de minimización. Ya que uno de nuestros objetivos es encontrar el mínimo de E , se consideran dos alternativas: usar los métodos de optimización numérica o resolver el sistema $E'(u) = 0$. En la siguiente sección se introducen, de forma general, los métodos de optimización que se aplicarán tanto para encontrar la imagen filtrada de ruido, como para nuestro objetivo final que es determinar el peso óptimo del parámetro de regularización.

1.2. Métodos de optimización

Los métodos de optimización de tipo descenso usan la información de primer y segundo orden del funcional E para encontrar una solución u del problema (1.1). Estos métodos tienen asociados esquemas iterativos de la forma:

$$u_{k+1} = u_k + \lambda_k p_k, \quad k = 1, 2, \dots \quad (1.2)$$

donde λ_k es un escalar positivo que se denomina *tamaño de paso* y p_k se conoce como *dirección de descenso*, es decir, $p_k^\top \nabla E(u_k) < 0$. Esta propiedad garantiza que $E(u)$ disminuya en esta dirección. Más aún, si notamos $\nabla E(u_k)$ por ∇E_k , se sigue que p_k es de la forma

$$p_k = -H_k^{-1} \nabla E_k, \quad (1.3)$$

donde H_k es una matriz simétrica, no singular y definida positiva. Dependiendo del valor que se le asigne a H_k se determinará el nombre y orden de convergencia del método de descenso. Por ejemplo, en el método del *descenso más profundo*, H_k es la matriz identidad I muy utilizado ya que es fácil de implementar y aplicable a problemas no convexos. Sin embargo, su convergencia es lineal (lenta). Mientras que, los métodos de *Newton* y *quasi-Newton* aprovechan la información de segundo orden y por lo tanto convergen más rápido que los métodos de primer orden, sin embargo, implican un mayor costo computacional.

En el método de Newton, H_k es la hessiana de E , y en los métodos quasi-Newton, H_k es una aproximación de la hessiana. Dicha aproximación es actualizada en cada iteración con el objetivo de incorporar la información obtenida en las iteraciones anteriores. Dentro de estos métodos se destaca el método BFGS (Broyden, Fletcher, Goldfarb, Shanno) [15].

1.2.1. Método BFGS

En los métodos quasi-Newton la dirección de descenso se obtiene mediante la siguiente expresión

$$p_k = -H_k^{-1} \nabla E_k,$$

donde H_k corresponde a una aproximación de la hessiana de E . La idea general de este método es construir aproximaciones de la hessiana mediante las ecuaciones del tipo secante, que en forma general se escriben como

$$H_{k+1} \underbrace{(u_{k+1} - u_k)}_{s_k} = \underbrace{\nabla E(u_{k+1}) - \nabla E(u_k)}_{z_k}.$$

Ya que matriz H_{k+1} no está unívocamente determinada por esta ecuación, se exigen criterios adicionales para la construcción de dichas matrices; uno de estos, consiste en actualizar H_{k+1} a partir de una modificación sencilla de H_k , de manera que se preserve la simetría. La alternativa que se emplea en este trabajo es, escoger H_{k+1} como la solución del problema:

$$\begin{cases} \min_B \|W(H^{-1} - H_k^{-1})W\|_F^2 \\ \text{subjeto a:} \\ H = H^T, \\ Hs_k = z_k, \end{cases} \quad (1.4)$$

donde $\|\cdot\|_F$ representa la norma de Frobenius y W es una matriz simétrica definida positiva tal que $W^2 s_k = z_k$. La solución del problema (1.4) es

$$H_{k+1} = H_k - \frac{H_k s_k s_k^T H_k}{s_k^T H_k s_k} + \frac{z_k z_k^T}{z_k^T s_k},$$

que corresponde a la actualización del método BFGS (Broyden, Fletcher, Goldfarb, Shanno). Si se verifica la condición de la curvatura, $z_k^T s_k > 0$ y usando las formulas de Sherman-Morrison-Woogbury se obtiene la actualización de la inversa de H_k notada por B_k ,

$$B_{k+1} = B_k + \frac{(s_k - B_k z_k) s_k^T + s_k (s_k - B_k z_k)^T}{s_k^T z_k} - \frac{(s_k - B_k z_k)^T z_k}{(s_k^T z_k)^2} s_k s_k^T.$$

Bajo ciertas hipótesis se puede demostrar que el algoritmo del BFGS está bien definido y genera una sucesión $\{u_k\}_{k \in \mathbb{N}}$ que converge superlinealmente hacia \bar{u} un mínimo local de (1.1).

1.2.2. Método de búsqueda lineal

La idea general en los algoritmos de optimización, consiste en elegir una dirección de descenso p_k y buscar a lo largo de ella, partiendo de la iteración actual u_k , el nuevo valor u_{k+1} que haga más pequeño el valor de la función objetivo E , una vez elegida la dirección

de descenso hace falta decidir cuanto moverse a lo largo de la misma.

El tamaño de paso λ que es la distancia a moverse a lo largo de p_k se puede encontrar dando solución al subproblema:

$$\min_{\lambda > 0} E(u_k + \lambda p_k),$$

la solución de este problema puede ser excesivamente costosa. En su lugar, se escogen los pasos de acuerdo a estrategias de *búsqueda lineal*. Una búsqueda popular es la regla de Armijo, donde dada una dirección de descenso p_k en u_k , escoger λ_k tal que verifique

$$E(u_k + \lambda_k p_k) \leq E(u_k) + \epsilon \lambda_k (\nabla E(u_k), p_k) \quad (1.5)$$

donde $\epsilon \in (0, 1)$. En este trabajo se propone utilizar un modelo polinomial [29] para determinar el tamaño del paso con la regla de Armijo.

La estrategia para elegir λ_k consiste en empezar con $\lambda_k = 1$ y si $u_k + p_k$ no verifica la condición de Armijo (1.5) se tiene que reducir λ_k hasta obtener un aceptable $u_k + \lambda_k p_k$, donde $\epsilon > 0$ es una constante.

Se define para cada iteración del método de optimización $\phi(\lambda) = E(u_k + \lambda p_k)$, que es la restricción unidimensional de E a la línea que pasa por u_k en la dirección p_k . La idea principal del método es aprovechar la información disponible sobre ϕ . En cada iteración se conoce que:

$$\phi(0) = E(u_k), \quad \phi'(0) = (\nabla E(u_k), p_k) < 0,$$

antes de calcular $E(u_k + p_k)$, se tiene también que

$$\phi(1) = E(u_k + p_k).$$

Si $E(u_k + p_k)$ satisface

$$\phi(1) > \phi(0) + \epsilon \phi'(0), \quad (1.6)$$

se plantea el modelo cuadrático de $\phi(\lambda)$, como

$$q(\lambda) = a + b\lambda + c\lambda^2, \quad a, b, c \in \mathbb{R},$$

con las condiciones

$$q(0) = \phi(0), \quad q(1) = \phi(1), \quad q'(0) = \phi'(0).$$

Se puede probar que el modelo cuadrático que satisface estas condiciones está dado por la siguiente expresión

$$q(\lambda) = \phi(0) + \phi'(0)\lambda + (\phi(1) - \phi(0) - \phi'(0))\lambda^2. \quad (1.7)$$

Así, el mínimo global de $q(\lambda)$ es:

$$\lambda^{\min} = \frac{-\phi'(0)}{2(\phi(1) - \phi(0) - \phi'(0))} > 0.$$

En efecto,

$$q''(\lambda) = 2(\phi(1) - \phi(0) - \phi'(0)) > 0,$$

de (1.6)

$$\phi(1) > \phi(0) + \epsilon\phi'(0) > \phi(0) + \phi'(0),$$

pues $\phi'(0) < 0$ y $\epsilon < 1$. Por lo tanto, se toma $\lambda^{(l)} = \lambda^{\min}$. Si $\phi(\lambda^{(l)})$ no satisface la regla de Armijo, se repite el proceso y se elige el siguiente tamaño del paso λ_{l+1} de acuerdo con

$$\lambda_{l+1} = \begin{cases} \underline{k}\lambda^l & \text{if } \lambda^{\min} < \underline{k}\lambda^l, \\ \lambda^{\min} & \text{if } \underline{k}\lambda^l \leq \lambda^{\min} \leq \bar{k}\lambda^l, \\ \bar{k}\lambda^l & \text{if } \lambda^{\min} > \bar{k}\lambda^l, \end{cases}$$

de donde, $\lambda_{l+1} \in [\underline{k}\lambda^l, \bar{k}\lambda^l]$.

1.3. Estudio del parámetro de regularización α

Como se analizó antes, el parámetro de regularización α determina la influencia de los términos en el modelo de filtrado de ruido, por lo que su elección es crucial para conseguir el equilibrio entre la estabilidad de la solución y la calidad de los datos. Para evidenciar la influencia del parámetro de regularización en la Figura 1.2 se muestra la solución del problema del filtrado de ruido considerando diferentes valores α .

Se puede observar que si α es demasiado pequeño, se obtiene una solución con un suavizado excesivo en la imagen restaurada. Por el contrario, si el parámetro α es demasiado grande se obtiene una imagen con presencia de ruido y por lo tanto no se ha resuelto el problema de filtrado de ruido. La pregunta que surge inmediatamente es: ¿ existe un método sistemático para elegir el parámetro de regularización α ? La respuesta es afirmativa y una de las alternativas consiste en plantear la estimación óptima del parámetro de regularización mediante un problema de optimización en dos niveles. Esta técnica proviene del aprendizaje automático de máquinas conocido como *Machine Learning* que consiste en aprender información de una muestra de datos que se aplicará luego a los términos desconocidos.

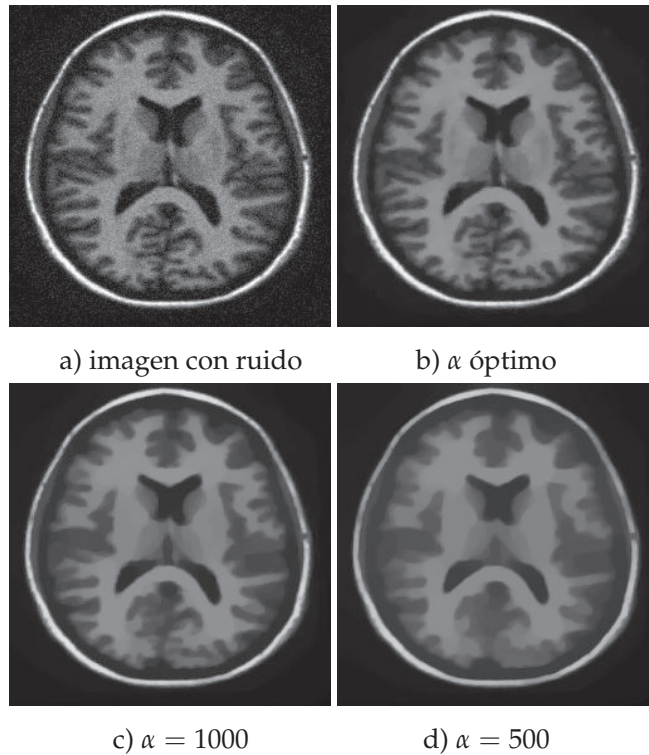


Figura 1.2. Filtrado de ruido considerando distintos valores del parámetro de regularización.

Los parámetros de regularización se consideran usualmente como valores reales. En este trabajo se propone considerar la dependencia espacial de los parámetros, con el objetivo de filtrar ruido no uniforme en una imagen, lo que nos permite acercarnos más a situaciones reales donde no se conocen el tipo ni la distribución de ruido que puede cubrir información relevante en una imagen. Así, consideramos los pesos en (1.1) como funciones continuas $P(\alpha)$ que tienen como dominio un espacio vectorial de dimensión finita \mathbb{R}^n . La idea de considerar espacios de dimensión finita surge de los dispositivos médicos, donde el número de valores que se pueden calibrar para adquirir una imagen es finito.

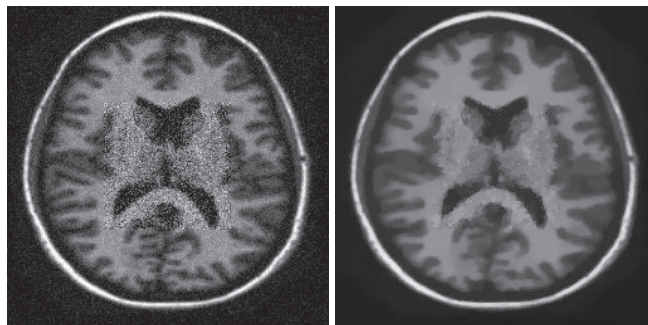


Figura 1.3. A la izquierda, la imagen con ruido no uniforme y a la derecha la imagen filtrada.

Para evidenciar que es importante conocer la distribución y localización del ruido, a continuación en la Figura 1.3 se presenta una imagen con ruido no uniforme, donde se aprecia

dos intensidades de ruido. El ruido más fuerte está localizado en el centro de la imagen y uno leve en el resto de la imagen. Si se considera filtrar la imagen utilizando un parámetro real el ruido en la parte central persiste.

A continuación, se citan algunos de los trabajos relacionados con el tema central de esta investigación,

- Dong, Hintermüller y Rincon ([22], 2011): consideran el parámetro α con dependencia espacial y plantean el algoritmo (SA-TV), usando intervalos de confianza y estimadores de varianza local, para lo cual se debe conocer información a priori como el tipo y la desviación estándar del ruido.
- De los Reyes y Schoenlieb ([19], 2013): proponen analizar y resolver un problema de optimización binivel no lineal en espacios funcionales para determinar el peso óptimo del regularizador escalar y el tipo de ruido, usando el modelo de variación total.
- Hintermüller, Papafitsoros, y Rautenberg ([30], 2013): reportan resultados analíticos del problema en una dimensión usando la regularización de variación total.
- De los Reyes, Schoenlieb y Valkonen([14], 2015): proponen analizar y resolver un problema de optimización binivel no lineal en espacios funcionales para determinar el peso del regularizador escalar de modelos de variación total generalizada.
- Cao, De los Reyes y Schoenlieb ([9], 2016): consideran la dependencia espacial del parámetro de regularización en espacios de Sobolev, $\alpha \in H^1(\Omega)$, en el modelo de variación total

Como se mostró, la idea de plantear un problema de optimización binivel no lineal, para determinar el valor óptimo de los parámetros reales del modelo TV y TGV para el filtrado de ruido considerando un conjunto de entrenamiento, fue propuesta por De los Reyes y Schönlieb en [19] y [14]. Siguiendo con esta idea, el objetivo del presente trabajo es plantear y resolver un problema de optimización binivel no lineal en espacios funcionales, para determinar el peso del valor óptimo de los parámetros con dependencia espacial que permita filtrar ruido no uniforme.

Capítulo 2

Modelo de variación total (TV)

En este capítulo, nos enfocaremos en el análisis del modelo de variación total TV, donde se abordarán temas de regularización debido a la no diferenciabilidad del modelo de filtrado de ruido. Esto permite obtener una condición necesaria y suficiente del modelo TV para el filtrado de ruido. Se exponen algunos resultados teóricos como la existencia de soluciones del problema de optimización de parámetros y la existencia de multiplicadores de Lagrange del problema de optimización no lineal con restricciones de igualdad y desigualdad. Se evidencia que el multiplicador asociado a la restricción de positividad del parámetro de optimización es una medida, que computacionalmente hablando es muy costosa de implementar. Para superar esto se propone introducir la *Regularización de Moreau-Yosida* [32]. Finalmente, se establece el sistema de optimalidad asociado al problema regularizado y se muestra que las soluciones de los problemas regularizados convergen hacia la solución del problema original.

2.1. Regularización de variación total

La técnica de *regularización de variación total* fue desarrollada por Rudin, Osher, y Fatemi [39], para dar solución al filtrado de ruido gaussiano. Sean Ω un subconjunto acotado de \mathbb{R}^2 , $f : \Omega \rightarrow \mathbb{R}$ la imagen observada y $BV(\Omega)$ el conjunto de las funciones de variación acotada. El modelo TV para el filtrado de ruido consiste en resolver el problema de minimización

$$\min_{u \in BV(\Omega)} \mathcal{J}(u, \alpha) := |Du|(\Omega) + \frac{\alpha}{2} \|u - f\|_{L^2(\Omega)}^2, \quad (2.1)$$

donde la solución u es la aproximación de la imagen sin ruido, $\alpha > 0$ que se conoce como parámetro de regularización y el conjunto de funciones de variación acotada $BV(\Omega)$ es un subconjunto de $L^1(\Omega)$, tal que $|Du|(\Omega) < \infty$, donde $|Du|(\Omega)$ es la variación total de u y se define como sigue

$$|Du|(\Omega) = \sup \left\{ \int_{\Omega} u \nabla \cdot \phi \, dx : \phi \in C_0^\infty(\Omega, \mathbb{R}^2), \|\phi\|_\infty \leq 1 \right\},$$

$C_0^\infty(\Omega, \mathbb{R}^2)$ denota el espacio de las funciones continuamente diferenciables con soporte compacto, equipado con la norma

$$\|\phi\|_\infty = \left(\sum_{i=1}^2 \sup_{x \in \Omega} |\phi_i(x)|^2 \right)^{1/2}.$$

Si $u \in W^{1,1}$ entonces

$$|Du|(\Omega) = \int_{\Omega} |Du| dx.$$

Como se explico en el primer capítulo, en la sección de métodos variacionales, el primer término de (2.1) cuantifica la variación dentro de una imagen y tiene como objetivo que las esquinas y bordes queden lo mejor definidos, mientras que el segundo, garantiza que la imagen solución u preserve la características de la imagen observada f . El parámetro de regularización α determina la influencia de los términos en (2.1), por lo que su valor es crítico, pues si este es pequeño se obtiene una solución sobrerregularizada y su suavizado será excesivo. Por el contrario, si su valor es grande no se resolverá el problema de ruido. Es así que nuestro objetivo es determinar su valor óptimo en el caso de imágenes que presentan ruido no uniforme, en la siguiente sección se presenta el planteamiento del problema.

2.2. Planteamiento del problema

El parámetro α se considera usualmente como escalar. En este trabajo se propone considerar la dependencia espacial del parámetro, con el objetivo de filtrar ruido no uniforme en una imagen, lo que nos permite acercarnos más a situaciones reales donde no se conocen el tipo ni la distribución de ruido que puede cubrir información relevante. Así, consideramos el peso de (2.1) como una función continua $P(\alpha)$ que tiene como dominio un espacio vectorial de dimensión finita \mathbb{R}^N . La idea de considerar espacios de dimensión finita surge de los dispositivos médicos, donde el número de valores que se pueden calibrar para adquirir una imagen es finito.

La idea de plantear un problema de optimización, para determinar el valor óptimo de los parámetros reales del modelo TV para el filtrado de ruido considerando conjuntos de entrenamiento, fue propuesta en [19]. Siguiendo con esta idea, el problema de optimización binivel para determinar $P(\alpha)$, tiene como restricción el modelo del filtrado de ruido TV y será:

$$\min_{\alpha \in \mathbb{R}^N} \frac{1}{2} \|f_0 - u_\alpha\|_{L^2(\Omega)}^2 + \xi \|\alpha\|_{\mathbb{R}^N}^2 := F(u_\alpha) \quad (2.2a)$$

sujeto a

$$u_\alpha \in \arg \min_{u \in BV(\Omega)} \mathcal{J}(u, \alpha) := |Du|(\Omega) + \frac{1}{2} \int_{\Omega} P(\alpha) (u - f)^2 dx, \quad (2.2b)$$

$$P(\alpha) \geq 0, \text{ en todo punto de } \Omega, \quad (2.2c)$$

donde f_0 es una imagen tomada con la máxima precisión posible, esta corresponde al con-

junto de entrenamiento que en este caso simplificado se considera un conjunto de un elemento. La idea de trabajar con conjuntos de entrenamiento proviene del *Machine Learning*.

2.3. Análisis del problema binivel

Las dificultades que surgen de inmediato al tratar un problema binivel no lineal en espacios de dimensión infinita son la no diferenciabilidad de la restricción (2.2b), lo que hace imposible verificar condiciones de calificación para sistemas de Karush-Kuhn-Tucker (KKT). En este trabajo se propone utilizar técnicas de regularización para demostrar la existencia de los multiplicadores de Lagrange y determinar así el sistema de optimalidad asociado al problema binivel que permita estimar el valor óptimo de los parámetros.

Empecemos por reemplazar el término TV de la restricción (2.2b) por su versión regularizada $TV^{\gamma,\mu}$, con parámetro $\gamma > 0$ asociado a la regularización de Huber y parámetro $\mu > 0$ asociado al término elíptico,

$$TV^{\gamma,\mu}_\alpha(u) = \int_{\Omega} |Du|_{\gamma} + \frac{\mu}{2} \|u\|_{H^1(\Omega)}^2, \quad (2.3)$$

donde $|\cdot|_{\gamma}: \mathbb{R}^n \rightarrow \mathbb{R}$ es la regularización de Huber C^1 y esta definida por,

$$|z|_{\gamma} = \begin{cases} \|z\| + \frac{\gamma}{2} t_1^2 - t_2 + A & \text{si } \gamma \|z\| \geq 1 + \frac{1}{2\gamma}, \\ B \|z\| + C \|z\|^2 - \frac{\gamma^3}{6} \|z\|^3 + D & \text{si } 1 - \frac{1}{2\gamma} \leq \gamma \|z\| \leq 1 + \frac{1}{2\gamma}, \\ \frac{\gamma}{2} \|z\|^2 & \text{si } \gamma \|z\| \leq 1 - \frac{1}{2\gamma}, \end{cases}$$

$\|\cdot\|$ corresponde a la norma euclideana en \mathbb{R}^n , $t_1 = \frac{1}{\gamma}(1 - \frac{1}{2\gamma})$, $t_2 = \frac{1}{\gamma}(1 + \frac{1}{2\gamma})$,

$$A = B(t_1 - t_2) + \frac{C}{2}(t_1^2 - t_2^2) - \frac{\gamma^3}{6}(t_1^3 - t_2^3), \quad B = 1 - \frac{(2\gamma + 1)^2}{8\gamma},$$

$$C = \frac{\gamma}{2}(2\gamma + 1), \quad D = \left(\frac{\gamma}{2} - \frac{B}{2}\right) t_1^2 - B \|t_1\| + \frac{\gamma^3}{6} \|t_1\|^3.$$

Considerar $TV^{\gamma,\mu}$ permite que el problema de optimización inferior (2.2b) se pueda reemplazar por su condición necesaria y suficiente de optimalidad. Se obtiene así el siguiente problema de optimización con restricciones,

$$\min_{\alpha \in \mathbb{R}^N} F(u_\alpha) \quad (2.4a)$$

sujeto a

$$\mu(u, \phi)_{H^1(\Omega)} + (h_\gamma(Du), D\phi)_{L^2(\Omega)} + \int_{\Omega} P(\alpha)(u - f)\phi \, dx = 0, \quad \forall \phi \in H^1(\Omega), \quad (2.4b)$$

$$P(\alpha) \geq 0, \text{ en todo punto de } \Omega \quad (2.4c)$$

donde h_γ es la derivada de $|\cdot|_\gamma$ determinada por,

$$h_\gamma(z) = \begin{cases} \frac{z}{\|z\|} & \text{si } \gamma\|z\| \geq 1 + \frac{1}{2\gamma}, \\ \frac{z}{\|z\|} \left(1 - \frac{\gamma}{2}(1 - \gamma\|z\| + \frac{1}{2\gamma})^2\right) & \text{si } 1 - \frac{1}{2\gamma} \leq \gamma\|z\| \leq 1 + \frac{1}{2\gamma}, \\ \gamma z & \text{si } \gamma\|z\| \leq 1 - \frac{1}{2\gamma}. \end{cases} \quad (2.5)$$

Así, $h_\gamma(Du)$ corresponde a una regularización de conjuntos activos e inactivos del subdiferencial de $|Du|$, es decir, h_γ coincide con todos los elementos del subdiferencial fuera de una vecindad de 0.

2.4. Multiplicadores de Lagrange

En esta sección, se prueba la existencia de multiplicadores de Lagrange para el problema de optimización de parámetros (2.4) y se establece el sistema de optimalidad asociado. Además, se obtiene una caracterización del gradiente para el funcional de costo reducido, la cual juega un papel importante en el desarrollo de algoritmos de solución rápida para los problemas de aprendizaje. Para esto se prueba en primer lugar que el operador solución es diferenciable.

En el trabajo [19] se analizó y planteó el problema de optimización de parámetros considerando un problema de optimización binivel no lineal en espacios funcionales, para el caso de parámetros escalares y con dependencia espacial del modelo TV. Por esta razón varios resultados teóricos se utilizan sin demostración, por ejemplo se presenta el resultado de existencia de una solución del problema de filtrado de ruido (2.4b) para $\gamma > 0$ fijo. Por otro lado, la implementación numérica se realizó solo para el caso del parámetro escalar y es ahí nuestra contribución, pues en este trabajo se pretende aproximar numéricamente el parámetro con dependencia espacial del modelo TV.

El operador solución asociado al problema de filtrado de ruido TV con parámetro con dependencia espacial es $S : X \mapsto H^1(\Omega)$, tal que

$$S(\alpha) = u.$$

En [19] se muestra que el operador solución es *Gâteaux* diferenciable, donde X es un espacio Hilbert tal que $X \hookrightarrow L^2(\Omega)$ (\hookrightarrow denota la inmersión continua). Sin embargo, este resultado fue mejorado en [9] donde los autores prueban para el caso $\alpha \in H^1(\Omega)$, que el operador solución es en realidad Fréchet diferenciable.

En nuestro caso, el operador solución asigna a cada $P(\alpha)$ la solución u de (2.4b), note que S puede verse como la composición de funciones *Gâteaux* diferenciables

$$S = G \circ P,$$

donde $P : \mathbb{R}^N \rightarrow C(\Omega)$ y $G : C(\Omega) \rightarrow H^1(\Omega)$.

Así, la derivada del operador S en α , en la dirección $\delta_\alpha \in \mathbb{R}^N$, esta caracterizada por la solución única $z \in H^1(\Omega)$ de la ecuación linealizada:

$$\begin{aligned} \mu(z, \phi)_{H^1(\Omega)} + (h'_\gamma(Du)Dz, D\phi)_{L^2(\Omega)} \\ + \int_{\Omega} P(\alpha) z \phi \, dx + \int_{\Omega} P'(\alpha) \delta_\alpha (u - f) \phi \, dx = 0, \text{ para todo } \phi \in H^1(\Omega). \end{aligned} \quad (2.6)$$

La existencia y unicidad de soluciones de (2.6) se sigue del teorema de Lax-Milgram usando las propiedades de monotonía de h_γ . En efecto, se define la forma bilineal $a(\cdot, \cdot)$ como

$$a(w, \phi) = \mu(w, \phi)_{H^1(\Omega)} + (h'_\gamma(Du)Dw, D\phi)_{L^2(\Omega)} + \int_{\Omega} P(\alpha) u w \phi \, dx.$$

Se debe probar que $a(\cdot, \cdot)$ es continua y coerciva en $H^1(\Omega)$, es decir, que para todo $w, \phi \in H^1(\Omega)$ existen k, K tales que se verifica

$$|a(w, \phi)| \leq K \|\phi\|_{H^1(\Omega)} \|w\|_{H^1(\Omega)} \quad \text{y} \quad a(w, w) \geq k \|w\|_{H^1(\Omega)}^2.$$

Primero, se debe probar que $h'_\gamma(Du)Dw \in L^\infty(\Omega)$ para demostrar la continuidad de la forma bilineal. En efecto,

$$h'_\gamma(Du)D\phi = \begin{cases} \frac{D\phi}{|Du|} - \frac{\langle Du, D\phi \rangle Du}{|Du|^3} & \text{si } \gamma|Du| \geq 1 + \frac{1}{2\gamma}, \\ \left(1 - \frac{\gamma}{2}\theta^2(Du)\right) \left(\frac{D\phi}{|Du|} - \frac{Du\langle D\phi, Du \rangle}{|Du|^3}\right) \\ \quad + \gamma^2\theta(Du) \frac{\langle Du, D\phi \rangle Du}{|Du|^2} & \text{si } 1 - \frac{1}{2\gamma} \leq \gamma|Du| \leq 1 + \frac{1}{2\gamma}, \\ \gamma D\phi & \text{si } \gamma|Du| \leq 1 - \frac{1}{2\gamma}. \end{cases} \quad (2.7)$$

donde $\theta(Du) = 1 - \gamma|Du| + \frac{1}{2\gamma}$. Se definen los conjuntos activo, semi-activo e inactivo como sigue,

$$\begin{aligned} \mathcal{A}^\gamma &:= \left\{ u \in \Omega : \gamma|u| \geq 1 + \frac{1}{2\gamma} \right\}, \\ \mathcal{S}^\gamma &:= \left\{ u \in \Omega : 1 - \frac{1}{2\gamma} \leq \gamma|u| \leq 1 + \frac{1}{2\gamma} \right\}, \\ \mathcal{I}^\gamma &:= \left\{ u \in \Omega : \gamma|u| \leq 1 - \frac{1}{2\gamma} \right\}. \end{aligned}$$

A continuación, se analizan estos conjuntos por separado:

- en el conjunto activo \mathcal{A}^γ se verifica que:

$$\frac{1}{|Du|} \leq \frac{2\gamma^2}{1 + 2\gamma}'$$

de donde,

$$\begin{aligned}
|h'_\gamma(Du)D\phi| &= \left| \frac{D\phi}{|Du|} - \frac{Du\langle D\phi, Du \rangle}{|Du|^3} \right| \leq \left| \frac{D\phi}{|Du|} \right| + \left| \frac{Du\langle D\phi, Du \rangle}{|Du|^3} \right|, \\
&\leq \frac{|D\phi|}{|Du|} + \frac{|Du||\langle D\phi, Du \rangle|}{|Du|^3} \leq \frac{|D\phi|}{|Du|} + \frac{|Du|^2|D\phi|}{|Du|^3}, \\
&= 2\frac{|D\phi|}{|Du|} \leq 4\frac{\gamma^2}{1+2\gamma}|D\phi|. \tag{2.8}
\end{aligned}$$

- En el conjunto semi-activo \mathcal{S}^γ se verifica que,

$$\frac{2\gamma^2}{1+2\gamma} \leq \frac{1}{|Du|} \leq \frac{2\gamma^2}{2\gamma-1},$$

de donde $\theta(Du) = (1 - \gamma|Du| + \frac{1}{2\gamma})$ verifica que $0 \leq \theta \leq \frac{1}{\gamma}$. Además,

$$0 \leq \theta^2 \leq \frac{1}{\gamma^2}$$

y

$$0 \leq \frac{2\gamma-1}{2\gamma} \leq (1 - \frac{\gamma}{2}\theta^2(Du)) \leq 1.$$

Por otro lado,

$$\begin{aligned}
|h'_\gamma(Du)D\phi| &= \left| \left(1 - \frac{\gamma}{2}\theta^2(Du)\right) \left(\frac{D\phi}{|Du|} - \frac{Du\langle D\phi, Du \rangle}{|Du|^3} \right) + \gamma^2\theta(Du) \frac{\langle Du, D\phi \rangle Du}{|Du|^2} \right|, \\
&\leq \left|1 - \frac{\gamma}{2}\theta^2(Du)\right| \left| \frac{D\phi}{|Du|} - \frac{Du\langle D\phi, Du \rangle}{|Du|^3} \right| + |\gamma^2\theta(Du)| \left| \frac{\langle Du, D\phi \rangle Du}{|Du|^2} \right|, \\
&\leq \left| \frac{D\phi}{|Du|} - \frac{Du\langle D\phi, Du \rangle}{|Du|^3} \right| + \gamma \frac{|Du|^2|D\phi|}{|Du|^2}.
\end{aligned}$$

Usando la estimación (2.8), es decir,

$$\left| \frac{D\phi}{|Du|} - \frac{Du\langle D\phi, Du \rangle}{|Du|^3} \right| \leq 2\frac{|D\phi|}{|Du|},$$

se sigue que

$$\begin{aligned}
|h'_\gamma(Du)D\phi| &\leq 2\frac{|D\phi|}{|Du|} + \gamma|D\phi|, \\
&\leq \frac{4\gamma^2}{2\gamma-1}|D\phi| + \gamma|D\phi| = \left(\frac{4\gamma^2}{2\gamma-1} + \gamma \right) |D\phi|. \tag{2.9}
\end{aligned}$$

- Finalmente, en el conjunto inactivo \mathcal{I}^γ ,

$$|h'_\gamma(Du)D\phi| = \gamma|D\phi|. \tag{2.10}$$

A partir de (2.8), (2.9) y (2.10), se puede concluir que

$$|h'_\gamma(Du)D\phi| \leq C|D\phi|,$$

donde $C \geq 0$ y se define como

$$C := \max \left\{ \frac{4\gamma^2}{2\gamma+1}, \frac{4\gamma^2}{2\gamma-1} + \gamma, \gamma \right\},$$

de donde $h'_\gamma(Du)Dw \in L^\infty(\Omega)$. Considerando este resultado, se prueba que $a(\cdot, \cdot)$ es bicon-
tinua

$$\begin{aligned} a(w, \phi) &= \mu(w, \phi)_{H^1(\Omega)} + (h'_\gamma(Du)Dw, D\phi)_{L^2(\Omega)} + (P(\lambda), w\phi)_{L^2(\Omega)}, \\ &\leq \mu|(w, \phi)_{H^1(\Omega)}| + |(h'_\gamma(Du)Dw, D\phi)_{L^2(\Omega)}| + |(P(\lambda), w\phi)_{L^2(\Omega)}|, \\ &\leq \mu\|w\|_{H^1(\Omega)}\|\phi\|_{H^1(\Omega)} + \|h'_\gamma(Du)Dw\|_{L^\infty(\Omega)}\|D\phi\|_{L^2(\Omega)} \\ &\quad + \|P(\lambda)\|_{L^\infty(\Omega)}\|w\phi\|_{L^2(\Omega)}. \end{aligned}$$

Usando la inmersión continua de H^1 en L^p , para $p > 1$ pero distinto de infinito, se sigue que
existe $C_1, C_2 > 0$ tales que

$$\begin{aligned} \|P(\lambda)\|_{L^\infty(\Omega)}\|w\phi\|_{L^2(\Omega)} &\leq \|P(\lambda)\|_{L^\infty(\Omega)}\|w\|_{L^4(\Omega)}\|\phi\|_{L^4(\Omega)} \\ &\leq \|P(\lambda)\|_{L^\infty(\Omega)}\|C_1\|w\|_{H^1(\Omega)}\|C_2\|\phi\|_{H^1(\Omega)} \\ &\leq C_3\|w\|_{H^1(\Omega)}\|\phi\|_{H^1(\Omega)}, \end{aligned}$$

pues $P(\lambda)$ es constante respecto a w y a ϕ . Reemplazando, esta estimación se obtiene que

$$\begin{aligned} a(w, \phi) &\leq \mu\|w\|_{H^1(\Omega)}\|\phi\|_{H^1(\Omega)} + C_4\|Dw\|_{L^2(\Omega)}\|D\phi\|_{L^2(\Omega)} + C_3\|w\|_{H^1(\Omega)}\|\phi\|_{H^1(\Omega)} \\ &\leq (\mu + C_4 + C_2)\|w\|_{H^1(\Omega)}\|\phi\|_{H^1(\Omega)} \\ &\leq K\|w\|_{H^1(\Omega)}\|\phi\|_{H^1(\Omega)}, \end{aligned}$$

con lo cual queda demostrada la continuidad de la forma bilineal. Resta probar que $a(\cdot, \cdot)$ es
coerciva. En primero lugar se debe probar que $(h'_\gamma(Du)D\phi, D\phi)_{L^2(\Omega)} \geq 0$.

- En el conjunto activo \mathcal{A}^γ , se verifica por la desigualdad de Cauchy-Schwarz que

$$\begin{aligned} (h'_\gamma(Du)D\phi, D\phi)_{L^2(\Omega)} &= \frac{\langle D\phi, D\phi \rangle}{|Du|} - \frac{\langle D\phi, Du \rangle^2}{|Du|^3}, \\ &= \frac{|D\phi|^2}{|Du|} - \frac{\langle D\phi, Du \rangle^2}{|Du|^3}, \\ &= \frac{|D\phi|^2|Du|^2 - \langle D\phi, Du \rangle^2}{|Du|^3} \geq 0. \end{aligned}$$

- En el conjunto semi-activo \mathcal{S}^γ , se verifica que

$$(1 - \frac{\gamma}{2}\theta^2(Du)) \geq 0 \quad \text{y} \quad \gamma^2\theta(Du) \geq 0,$$

así,

$$\begin{aligned} (h'_\gamma(Du)D\phi, D\phi) &= (1 - \frac{\gamma}{2}\theta^2(Du)) \left(\frac{\langle D\phi, D\phi \rangle}{|Du|} - \frac{\langle D\phi, Du \rangle^2}{|Du|^3} \right) + \gamma^2\theta(Du) \frac{\langle Du, D\phi \rangle^2}{|Du|^2} \\ &\geq \frac{\langle Du, D\phi \rangle^2}{|Du|^2} \geq 0. \end{aligned}$$

- En el caso del conjunto inactivo \mathcal{I}^γ , se tiene de forma directa que

$$(h'(Du)D\phi, D\phi)_{L^2(\Omega)} = \gamma \langle D\phi, D\phi \rangle \geq 0.$$

Finalmente, se verifica que $P(\lambda) \geq 0$, de donde

$$\begin{aligned} a(w, w) &= \mu(w, w)_{H^1(\Omega)} + (h'_\gamma(Du)Dw, Dw)_{L^2(\Omega)} + (P(\lambda)w, w)_{L^2(\Omega)}, \\ &\geq \mu \|w\|_{H^1(\Omega)}^2 \end{aligned}$$

2.4.1. Ecuación adjunta

El objetivo de esta sección es caracterizar la ecuación adjunta del problema de optimización de parámetros, usando el método del Lagrangiano. La existencia de una solución única de la ecuación adjunta se sigue ya que la ecuación linealizada está bien planteada. Se define el Lagrangiano asociado al problema (2.4) como :

$$\mathcal{L}(u, \alpha, p) = F(u) + \mu(u, p)_{H^1(\Omega)} + \int_{\Omega} h_\gamma(Du)Dp + \int_{\Omega} P(\alpha)(u - f) p dx.$$

Tomando la derivada con respecto a la variable de estado u en la dirección δ_u , se obtiene la condición necesaria de optimalidad

$$\mathcal{L}'_u(u, \alpha, p)[\delta_u] = \nabla_u F(u)\delta_u + \mu(p, \delta_u)_{H^1(\Omega)} + \int_{\Omega} h'_\gamma(Du)\delta_u Dp + \int_{\Omega} P(\alpha) \delta_u p = 0.$$

2.4.2. Condición de optimalidad

Utilizando la diferenciabilidad del operador solución y que la ecuación adjunta está bien planteada, a continuación se establece el sistema de optimalidad asociado al problema de optimización:

$$\min_{\alpha \in \mathbb{R}^N} F(u_\alpha) = \frac{1}{2} \|f_0 - u\|_{L^2(\Omega)}^2 + \xi \|\alpha\|_{\mathbb{R}^N}^2 \quad (2.11a)$$

sujeto a

$$\begin{aligned} \mu(u, \phi)_{H^1(\Omega)} + (h_\gamma(Du), D\phi)_{L^2(\Omega)} + \int_{\Omega} P(\alpha)(u - f)\phi \, dx = 0, \text{ para todo } \phi \in H^1(\Omega), \\ P(\alpha) \geq 0 \text{ en todo punto de } \Omega. \end{aligned} \quad (2.11b)$$

Además, se presenta la caracterización del gradiente del funcional de costo reducido, la cual juega un papel importante en el desarrollo de algoritmos de solución rápida para los problemas de aprendizaje.

TEOREMA 2.1. *Sea $\bar{\alpha} \in \mathbb{R}^N$ una solución óptima local del problema (2.11). Si existen $\alpha_0 \in \mathbb{R}^N$ tal que*

$$P(\bar{\alpha}) + \nabla_{\alpha} P(\bar{\alpha})^T (\alpha_0 - \bar{\alpha}) > 0, \quad (2.12)$$

entonces existen multiplicadores de Lagrange $p \in H^1(\Omega)$ y $\lambda \in \mathcal{M}(\Omega)$, tales que el siguiente sistema se verifica:

$$\mu(u, \phi)_{H^1(\Omega)} + (h_\gamma(Du), D\phi)_{L^2(\Omega)} + \int_{\Omega} P(\alpha)(u - f)\phi \, dx = 0, \text{ para todo } \phi \in H^1(\Omega), \quad (2.13a)$$

$$\begin{aligned} \mu(p, \phi)_{H^1(\Omega)} + (h'_\gamma(Du)^* Dp, D\phi)_{L^2(\Omega)} \\ + \int_{\Omega} P(\alpha) p \phi \, dx = -\nabla_u F(u), \text{ para todo } \phi \in H^1(\Omega), \end{aligned} \quad (2.13b)$$

$$2\zeta(\alpha, \delta_\alpha)_{\mathbb{R}^N} + \int_{\Omega} P'(\alpha)\delta_\alpha(u - f) p \, dx \quad (2.13c)$$

$$- (P'(\alpha)^* \lambda, \delta_\alpha)_{\mathbb{R}^N} = 0, \text{ para todo } \delta_\alpha \in \mathbb{R}^N, \quad (2.13d)$$

$$\int_{\Omega} (z - P(\alpha)) d\lambda \geq 0, \text{ para todo } z \in C(\Omega) : z \geq 0. \quad (2.13e)$$

DEMOSTRACIÓN 2.1. *Considere el funcional de costo reducido*

$$\mathcal{F}(\alpha) = F(u(\alpha), \alpha) = \|f_0 - S(\alpha)\|_{L^2(\Omega)}^2 + \zeta \|\alpha\|_{\mathbb{R}^N}^2.$$

El problema de optimización binivel se puede formular como sigue

$$\min_{\alpha \in \mathbb{R}^N} \mathcal{F}(\alpha)$$

sujeto a:

$$P(\alpha) \geq 0, \text{ en todo punto de } \Omega$$

donde $\mathcal{F} : \mathbb{R}^N \rightarrow \mathbb{R}$ y $P : \mathbb{R}^N \rightarrow C(\Omega)$. Se define el cono

$$K := \{t \in C(\Omega) : t(x) \geq 0\}.$$

Las restricciones se pueden escribir como $G(\alpha) := P(\alpha) \in K$. La existencia de multiplicadores

de Lagrange se sigue de [44], si la condición de Slater se verifica, es decir, si

$$\exists \alpha_0 \text{ tal que } G(\alpha) + G'(\alpha)(\alpha_0 - \alpha) \in \text{int } K. \quad (2.14)$$

Ya que K tiene interior no vacío y por hipótesis se cumple (2.12), entonces la condición de Slater se verifica. Se sigue así que existe un multiplicador

$$\lambda \in K^+ := \{\vartheta \in \mathcal{M}(\Omega) : \langle \vartheta, z \rangle_{\mathcal{M}(\Omega), \mathcal{C}(\Omega)} \geq 0, \forall z \in K\},$$

tal que

$$\mathcal{F}'(\bar{\alpha}) - G'(\bar{\alpha})^* \lambda = 0.$$

Tomando la derivada con respecto a α y denotando a z la solución de la ecuación linealizada (2.6) se tiene, junto con la ecuación adjunta (2.13b), que

$$\begin{aligned} \mathcal{F}'(\alpha)[\theta] &= \nabla_u F(u)z + 2\xi(\alpha, \theta), \\ &= 2\xi(\alpha, \theta) - \mu(p, z) - \int_{\Omega} h'_{\gamma}(Du)(Dp)Dz - \int_{\Omega} P(\alpha)pz \, dx, \\ &= 2\xi(\alpha, \theta) - \mu(z, p) - \int_{\Omega} h'_{\gamma}(Du)Dz(Dp) - \int_{\Omega} P(\alpha)z p \, dx \end{aligned}$$

la cual, teniendo en cuenta la ecuación linealizada, se convierte en

$$\mathcal{F}'(\alpha)[\theta] = 2\xi(\alpha, \theta) + \int_{\Omega} P'(\alpha)\theta(u - f)p. \quad (2.15)$$

Por otro lado, de la estructura de G se obtiene que

$$\begin{aligned} \langle G'(\bar{\alpha})^* \lambda, \theta \rangle &= \langle \lambda, G'(\bar{\alpha})\theta \rangle_{\mathcal{M}(\Omega), \mathcal{C}(\Omega)}, \\ &= \langle \lambda, P'(\bar{\alpha})\theta \rangle_{\mathcal{M}(\Omega), \mathcal{C}(\Omega)}, \end{aligned}$$

y, por lo tanto

$$2\xi\alpha - P'(\bar{\alpha})^* \lambda + \int_{\Omega} P'(\alpha)(u - f)p = 0.$$

OBSERVACIÓN. Si P es lineal respecto al parámetro, la condición de Slater (2.14) se reduce a la existencia de α_0 , tal que $P(\alpha_0) > 0$.

OBSERVACIÓN. Se puede considerar $P(\alpha)(x)$ como una función continua a trozos. En este caso el multiplicador λ pertenecen al dual del espacio $L^{\infty}(\Omega)$, cuyos elementos no son necesariamente medidas.

2.5. Problema regularizado

En la sección anterior se evidenció que el multiplicador de Lagrange asociado a la restricción de desigualdad es una medida, cuya implementación numérica es muy complicada. Como alternativa a esto se introduce la *Regularización de Moreau-Yosida* [32], que consiste en

penalizar el funcional de costo con una función C^1 cuando la restricción de positividad no se verifica [15]. Esta técnica ha sido ampliamente utilizada en trabajos como [17], [18], [16] donde se consideran problemas de control óptimo con restricciones de desigualdad. Se define la familia de problemas η -penalizados como:

$$\min_{\alpha \in \mathbb{R}^N} F_\eta(u_\alpha) := F(u_\alpha) + \frac{1}{2\eta} \|\min(0, \eta P(\alpha))\|_{L^2(\Omega)}^2 \quad (2.16a)$$

sujeto a

$$\mu(u, \phi)_{H^1(\Omega)} + (h_\gamma(Du), D\phi)_{L^2(\Omega)} + \int_{\Omega} P(\alpha)(u - f)\phi \, dx = 0, \text{ para todo } \phi \in H^1(\Omega). \quad (2.16b)$$

En el siguiente teorema, se estudia la convergencia de las soluciones del problema η -regularizado a las soluciones del problema original.

TEOREMA 2.2. *La sucesión $\{(u_{\alpha_\eta}, \alpha_\eta)\}_{\eta>0}$ de soluciones del problema η -regularizado (2.16) contiene una subsucesión que converge fuertemente en $H^1(\Omega) \times \mathbb{R}^N$ a una solución óptima $(\bar{u}, \bar{\alpha})$ del problema original (2.11).*

DEMOSTRACIÓN 2.2. *Sea $(\bar{u}, \bar{\alpha}) \in H^1(\Omega) \times \mathbb{R}^N$ una solución global del problema (2.11). De las propiedades del funcional de costo regularizado se verifica que,*

$$F(u_{\alpha_\eta}) \leq F_\eta(u_{\alpha_\eta}) = F(u_{\alpha_\eta}) + \underbrace{\frac{1}{2\eta} \|\min(0, \eta P(u_{\alpha_\eta}))\|_{L^2(\Omega)}^2}_{\geq 0}$$

además,

$$F_\eta(u_{\alpha_\eta}) \leq F_\eta(\bar{u}) = F(\bar{u}). \quad (2.17)$$

Consecuentemente, puesto que $\xi > 0$, se sigue que la sucesión $\{\alpha_\eta\}_{\eta>0}$ es uniformemente acotada en \mathbb{R}^N , de donde existe una subsucesión en \mathbb{R}^N notada con el mismo nombre, tal que

$$\alpha_\eta \rightarrow \hat{\alpha}.$$

Por otro lado, notando $u_\eta := u_{\alpha_\eta}$ se prueba que $\{u_\eta\}_{\eta>0}$ es uniformemente acotado en $H^1(\Omega)$. En efecto, usando la ecuación variacional (2.16b), se puede estimar $\|u_\eta\|_{H^1(\Omega)}$, donde u_η corresponde a la solución de (2.16b) respecto a α_η

$$\mu(u_\eta, \phi)_{H^1(\Omega)} + (h_\gamma(Du_\eta), D\phi)_{L^2(\Omega)} + \int_{\Omega} P(\alpha_\eta)(u_\eta - f)\phi \, dx = 0, \text{ para todo } \phi \in H^1(\Omega).$$

Tomando $\phi = u_\eta$ se sigue que,

$$\mu\|u_\eta\|_{H^1(\Omega)}^2 + (h_\gamma(Du_\eta), Du_\eta)_{L^2(\Omega)} + \int_{\Omega} P(\alpha_\eta)u_\eta^2 = \int_{\Omega} P(\alpha_\eta)f u_\eta \, dx,$$

y dado que $(h_\gamma(Du_\eta), Du_\eta)_{L^2(\Omega)} \geq 0$ y $P(\alpha_\eta)u_\eta^2 \geq 0$, pues $P(\alpha_\eta) \geq 0$, se obtiene que

$$\mu \|u_\eta\|_{H^1(\Omega)}^2 \leq \left| \int_{\Omega} P(\alpha_\eta) f u_\eta \, dx \right| \leq \|P(\alpha_\eta)\|_{L^\infty(\Omega)} \|f\|_{L^2(\Omega)} \|u_\eta\|_{L^2(\Omega)},$$

donde $\|P(\alpha_\eta)\|_{L^\infty(\Omega)}$ tiene sentido ya que $P(\alpha_\eta) \in C(\Omega)$ y por extensión continua $P(\alpha_\eta) \in C(\bar{\Omega})$, es decir, que $P(\alpha_\eta)$ es una función continua sobre un compacto, así $P(\alpha_\eta) \in L^\infty(\Omega)$. Usando $\|u_\eta\|_{L^2(\Omega)} \leq \|u_\eta\|_{H^1(\Omega)}$ se sigue que

$$\|u_\eta\|_{H^1(\Omega)} \leq \frac{C_1}{\mu} \|P(\alpha_\eta)\|_{L^\infty(\Omega)}.$$

Por otro lado, se probó que α_η es uniformemente acotado, es decir, $|\alpha_\eta| \leq C$ de donde $P(\alpha_\eta)$ es acotado en $C(\Omega)$ por estar definido sobre un conjunto cerrado y acotado $S := \{\alpha_\eta : |\alpha_\eta| \leq C\}$ (ver [1, Teorema 1.2]). Así se obtiene que

$$\|u_\eta\|_{H^1(\Omega)} \leq \frac{C_1}{\mu} \|P(\alpha_\eta)\|_{L^\infty(\Omega)} \leq C_2, \quad (2.18)$$

es decir, que u_η esta uniformemente acotada $H^1(\Omega)$, de donde existe una subsucesión notada con el mismo nombre $u_\eta \rightharpoonup \hat{u}$ débilmente en $H^1(\Omega)$. Ahora, se debe probar que el punto límite \hat{u} es factible, es decir, satisface la restricción de positividad del problema original (2.11). De (2.17) el siguiente término

$$\frac{1}{2\eta} \|\min(0, \eta P(\alpha_\eta))\|_{L^2(\Omega)}^2 \quad (2.19)$$

es uniformemente acotado con respecto a η , pues por (2.17) se tiene que

$$\begin{aligned} 0 &\leq \frac{1}{2\eta} \|\min(0, \eta P(\alpha_\eta))\|_{L^2(\Omega)}^2 \leq F(\bar{u}) := C \\ 0 &\leq \eta \frac{1}{2} \|\min(0, \eta P(\alpha_\eta))\|_{L^2(\Omega)}^2 \leq C \\ 0 &\leq \|\min(0, \eta P(\alpha_\eta))\|_{L^2(\Omega)}^2 \leq \frac{2C}{\eta}. \end{aligned}$$

Así,

$$\lim_{\eta \rightarrow \infty} \|\min(0, P(\alpha_\eta))\|_{L^2(\Omega)}^2 = 0.$$

Aplicando el lema de Fatou, se obtiene que

$$\int_{\Omega} \liminf_{\eta \rightarrow \infty} \min(0, P(\alpha_\eta))^2 \leq \liminf_{\eta \rightarrow \infty} \int_{\Omega} \min(0, P(\alpha_\eta))^2 = 0,$$

entonces

$$\int_{\Omega} \min(0, P(\hat{u}))^2 \leq 0$$

lo que implica que $\min(0, P(\hat{u})) = 0$. Entonces, se verifica que $P(\hat{u}) \geq 0$ en Ω . Considerando

además que el funcional de costo es débilmente semicontinuo inferior, es decir,

$$F(\hat{u}) \leq \liminf_{\eta \rightarrow \infty} F(u_{\alpha_\eta}),$$

donde \hat{u} denota la solución de (2.16b) asociada a $\hat{\alpha}$. Además, se verifica que

$$\liminf_{\eta \rightarrow \infty} F(u_{\alpha_\eta}) \leq \liminf_{\eta \rightarrow \infty} F_\eta(u_{\alpha_\eta}) \leq F(\bar{u}),$$

de donde, se sigue que $(\hat{u}, \hat{\alpha})$ es solución de (2.11). Se denota al óptimo $(\hat{u}, \hat{\alpha})$ por $(\bar{u}, \bar{\alpha})$.

Finalmente, para verificar la convergencia fuerte se emplea la diferencia de la ecuación variacional (2.16b) para $u_\eta := u_{\alpha_\eta}$ y \bar{u} y se sigue que,

$$\begin{aligned} \mu(u_\eta - \bar{u}, \phi)_{H^1(\Omega)} + (h_\gamma(Du_\eta) - h_\gamma(D\bar{u}), D\phi)_{L^2(\Omega)} + \\ \int_{\Omega} P(\alpha_\eta) (u_\eta - f)\phi - P(\bar{\alpha})(\bar{u} - f)\phi \, dx = 0, \quad \text{para todo } \phi \in H^1(\Omega). \end{aligned}$$

Tomando $\phi = u_\eta - \bar{u}$ y sumando y restando el término

$$\int_{\Omega} P(\bar{\alpha}) u_\eta \, dx,$$

se obtiene que

$$\begin{aligned} \mu \|u_\eta - \bar{u}\|_{H^1(\Omega)}^2 + (h_\gamma(Du_\eta) - h_\gamma(D\bar{u}), D(u_\eta - \bar{u}))_{L^2(\Omega)} + \int_{\Omega} P(\bar{\alpha})(u_\eta - \bar{u})^2 \\ = - \int_{\Omega} [(P(\alpha_\eta) + P(\bar{\alpha})) u_\eta + (P(\alpha_\eta) - P(\bar{\alpha})) f] (u_\eta - \bar{u}) \, dx. \end{aligned}$$

Por la monotonía de $h_\gamma(\cdot)$ y ya que $P(\bar{\alpha})(u_\eta - \bar{u})^2 \geq 0$, se verifica que

$$\begin{aligned} \mu \|u_\eta - \bar{u}\|_{H^1(\Omega)}^2 &\leq \left| \int_{\Omega} [(P(\alpha_\eta) - P(\bar{\alpha})) u_\eta + (P(\alpha_\eta) - P(\bar{\alpha})) f] (u_\eta - \bar{u}) \, dx \right| \\ &\leq \| (P(\alpha_\eta) - P(\bar{\alpha})) u_\eta + (P(\alpha_\eta) - P(\bar{\alpha})) f \|_{L^2(\Omega)} \|u_\eta - \bar{u}\|_{L^2(\Omega)}, \\ &\leq (\| (P(\alpha_\eta) - P(\bar{\alpha})) u_\eta \|_{L^2(\Omega)} \\ &\quad + \|P(\alpha_\eta) - P(\bar{\alpha})\|_{L^2(\Omega)} \|f\|_{L^\infty(\Omega)}) \|u_\eta - \bar{u}\|_{L^2(\Omega)}, \\ &\leq (\| (P(\alpha_\eta) - P(\bar{\alpha})) \|_{L^\infty(\Omega)} \|u_\eta\|_{L^2(\Omega)} \\ &\quad + \|P(\alpha_\eta) - P(\bar{\alpha})\|_{L^2(\Omega)} \|f\|_{L^\infty(\Omega)}) \|u_\eta - \bar{u}\|_{L^2(\Omega)}. \end{aligned}$$

Ya que P es $C^1(\mathbb{R}^N)$, es Lipschitz continua en una vecindad de $\bar{\alpha}$ se verifica que

$$\begin{aligned} \mu \|u_\eta - \bar{u}\|_{H^1(\Omega)}^2 &\leq (L_1(\bar{\alpha}) \|\alpha_\eta - \bar{\alpha}\|_{\mathbb{R}^N} \|u_\eta\|_{L^2(\Omega)} \\ &\quad + L_2(\bar{\alpha}) \|\alpha_\eta - \bar{\alpha}\|_{\mathbb{R}^N} \|f\|_{L^\infty(\Omega)}) \|u_\eta - \bar{u}\|_{L^2(\Omega)}. \end{aligned}$$

Así, usando (2.18) y $\|u_\eta - \bar{u}\|_{L^2(\Omega)} \leq \|u_\eta - \bar{u}\|_{H^1(\Omega)}$ se obtiene que

$$\begin{aligned} \mu \|u_\eta - \bar{u}\|_{H^1(\Omega)}^2 &\leq \left(L_1(\bar{\alpha}) \|\alpha_\eta - \bar{\alpha}\|_{\mathbb{R}^N} \|u_\eta\|_{H^1(\Omega)} + L_2(\bar{\alpha}) \|\alpha_\eta - \bar{\alpha}\|_{\mathbb{R}^N} \|f\|_{L^\infty(\Omega)} \right) \|u_\eta - \bar{u}\|_{H^1(\Omega)}, \\ &\leq (L_1(\bar{\alpha}) C_2 \|\alpha_\eta - \bar{\alpha}\|_{\mathbb{R}^N} + C_1 L_2(\bar{\alpha}) \|\alpha_\eta - \bar{\alpha}\|_{\mathbb{R}^N}) \|u_\eta - \bar{u}\|_{H^1(\Omega)}, \\ \|u_\eta - \bar{u}\|_{H^1(\Omega)} &\leq L_1(\bar{\alpha}) C_2 \|\alpha_\eta - \bar{\alpha}\|_{\mathbb{R}^N} + C_1 L_2(\bar{\alpha}) \|\alpha_\eta - \bar{\alpha}\|_{\mathbb{R}^N}, \\ &\leq L_1(\bar{\alpha}) C_2 \|\alpha_\eta - \bar{\alpha}\|_{\mathbb{R}^N} + C_1 L_2(\bar{\alpha}) \|\alpha_\eta - \bar{\alpha}\|_{\mathbb{R}^N}, \end{aligned}$$

por lo tanto, si $\alpha_\eta \rightarrow \bar{\alpha}$ cuando $\eta \rightarrow \infty$ entonces $u_\eta \rightarrow \bar{u}$.

2.5.1. Sistema de optimalidad

El Lagrangiano asociado al problema regularizado (2.16) se define a continuación:

$$\mathcal{L}_\eta(u, \alpha, p) = F_\eta(u) + \mu(u, p)_{H^1(\Omega)} + (h_\gamma(Du), Dp)_{L^2(\Omega)} + \int_\Omega P(\alpha)(u - f)p.$$

Tomando la derivada con respecto a la variable de estado u en la dirección δ_u , se obtiene la condición necesaria de optimalidad

$$\nabla_u \mathcal{L}_\eta(u, \alpha, p)[\delta_u] = \nabla_u F_\eta(u) \delta_u + \mu(p, \delta_u)_{H^1(\Omega)} + (h'_\gamma(Du)^* Dp, D\delta_u) + \int_\Omega P(\alpha) p \delta_u = 0, \quad (2.20)$$

que se conoce como ecuación adjunta. Por otro lado, tomando la derivada parcial con respecto a la variable de control α en la dirección δ_α se obtiene:

$$\nabla_\alpha \mathcal{L}_\eta(u, \alpha, p)[\delta_\alpha] = \nabla_\alpha F_\eta(u) \delta_\alpha + \int_\Omega P'(\alpha) \delta_\alpha (u - f) p = 0,$$

que se conoce como ecuación del gradiente.

OBSERVACIÓN. Como resultado inmediato se obtiene la caracterización del gradiente del funcional de costo reducido:

$$\nabla \mathcal{F}_\eta(\alpha) := \nabla_\alpha F(u) + \int_\Omega \min(0, \eta P(\alpha)) P'(\alpha) dx + \int_\Omega P'(\alpha) (u - f) p. \quad (2.21)$$

Este resultado desempeña un rol importante en el diseño de algoritmos de optimización para determinar de forma eficiente la solución numérica del problema (2.16).

2.6. Tratamiento numérico

La solución numérica del problema de optimización de parámetros se puede abordar desde dos puntos de vista, el primero consiste en utilizar un método de Newton para resolver el sistema de optimalidad, y el segundo, usar un algoritmo iterativo para la solución del problema de optimización, con la caracterización del gradiente del funcional de costo reducido. En este trabajo nos enfocaremos en los algoritmos iterativos para la resolución

de problemas de optimización binivel no lineales, específicamente el método quasi-Newton BFGS.

Si $\nabla \mathcal{F}_\eta$ es localmente Lipschitz continua, el algoritmo BFGS está bien definido y genera una sucesión $\{\alpha_k\}_{k \in \mathbb{N}}$ que converge a un punto estacionario $\hat{\alpha}$. Esto gracias a que la Lipschitz continuidad local implica la continuidad uniforme en los conjunto de nivel. Para probar que el gradiente del funcional de costo reducido es localmente Lipschitz continuo se emplean estimaciones de la variable de estado y del estado adjunto que se muestran en los siguientes resultados.

LEMA 2.3. Si $f \in L^\infty(\Omega)$ y $u_\alpha \in C^{1,\rho}(\Omega)$, para algún $\rho \in (0,1)$ entonces para $\alpha^* \in \mathbb{R}^N$ arbitrario pero fijo y $\alpha, \hat{\alpha} \in V_\delta(\alpha^*)$ se verifica:

$$\begin{aligned} i) \|p\|_{H^1(\Omega)} &\leq K_1, \\ ii) \|p - \hat{p}\|_{H^1(\Omega)} &\leq K_2(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}. \end{aligned}$$

DEMOSTRACIÓN 2.3. Se notará $u := u_\alpha$ como la solución de la ecuación de estado (2.16b) correspondiente a α y $p := p_\alpha$ como la solución de la ecuación adjunta (2.20) correspondiente a α . Empecemos estimando $\|p\|_{H^1(\Omega)}$, para lo se considera la ecuación adjunta (2.20), es decir,

$$\nabla_u F(u)\phi + \mu(p, \phi)_{H^1(\Omega)} + \int_\Omega h'_\gamma(Du)DpD\phi + \int_\Omega P(\alpha)p\psi = 0, \quad \forall \phi \in H^1(\Omega).$$

Tomando $\phi = p$, se sigue que

$$\mu \|p\|_{H^1(\Omega)}^2 + \int_\Omega h'_\gamma(Du)(Dp)^2 + \int_\Omega P(\alpha)p^2 = - \int_\Omega (u - f_0)p.$$

Gracias a la positividad de $h'_\gamma(\cdot)$ y de $P(\alpha)$, se sigue que

$$\begin{aligned} \mu \|p\|_{H^1(\Omega)}^2 &\leq \|u - f_0\|_{L^2(\Omega)} \|p\|_{L^2(\Omega)}, \\ &\leq (\|u\|_{L^2(\Omega)} + \|f_0\|_{L^2(\Omega)}) \|p\|_{H^1(\Omega)}, \\ \|p\|_{H^1(\Omega)} &\leq \frac{K_1}{\mu} (\|f_0\|_{L^2(\Omega)} + K_2 \|u\|_{H^1(\Omega)}), \end{aligned}$$

y usando la estimación (2.18) de $\|u\|_{H^1(\Omega)}$ que se probó previamente, se tiene que

$$\|p\|_{H^1(\Omega)} \leq K_3.$$

Resta hallar la estimación de $\|p - \hat{p}\|_{H^1(\Omega)}$, para la cual se usa exactamente el proceso anterior

$$\begin{aligned} (\nabla_u F(u) - \nabla_u F(\hat{u}))\phi + \mu(p - \hat{p}, \phi)_{H^1(\Omega)} + \int_\Omega (h'_\gamma(Du)(Dp) - h'_\gamma(D\hat{u})(D\hat{p}))D\phi \\ + \int_\Omega (P(\alpha)p - P(\hat{\alpha})\hat{p})\phi = 0. \end{aligned}$$

Sumando y restando los términos

$$\int_{\Omega} h'_{\gamma}(Du) D\hat{p} D\phi, \quad \int_{\Omega} P(\alpha) \hat{p} D\phi,$$

y tomando $\phi = (p - \hat{p})$, se obtiene

$$\begin{aligned} \mu \|p - \hat{p}\|_{H^1(\Omega)}^2 + \int_{\Omega} h'_{\gamma}(Du) D(p - \hat{p}) D(p - \hat{p}) + \int_{\Omega} P(\alpha) (p - \hat{p})^2 \\ = - \int_{\Omega} (P(\alpha) - P(\hat{\alpha})) \hat{p} (p - \hat{p}) - \int_{\Omega} (u - \hat{u}) (p - \hat{p}) \\ - \int_{\Omega} (h'_{\gamma}(Du) - h'_{\gamma}(D\hat{u})) D\hat{p} D(p - \hat{p}). \end{aligned}$$

Gracias a la positividad de $h'_{\gamma}(\cdot)$ y ya que $P(\alpha) \geq 0$ se verifica que,

$$\begin{aligned} \mu \|p - \hat{p}\|_{H^1(\Omega)}^2 &\leq \left(\|u - \hat{u}\|_{L^2(\Omega)} + \|P(\alpha) - P(\hat{\alpha})\|_{L^{\infty}(\Omega)} \|\hat{p}\|_{L^2(\Omega)} \right) \|p - \hat{p}\|_{L^2(\Omega)} \\ &\quad + \| (h'_{\gamma}(Du) - h'_{\gamma}(D\hat{u})) D\hat{p} \|_{L^2(\Omega)} \|p - \hat{p}\|_{L^2}, \\ &\leq C \left\{ \|u - \hat{u}\|_{L^2(\Omega)} + \|P(\alpha) - P(\hat{\alpha})\|_{L^{\infty}(\Omega)} \|\hat{p}\|_{L^2(\Omega)} \right. \\ &\quad \left. + \| (h'_{\gamma}(Du) - h'_{\gamma}(D\hat{u})) D\hat{p} \|_{L^2(\Omega)} \right\} \|p - \hat{p}\|_{H^1(\Omega)}. \end{aligned}$$

Por otro lado, se debe estimar

$$\| (h'_{\gamma}(Du) - h'_{\gamma}(D\hat{u})) D\hat{p} \|_{L^2(\Omega)},$$

para lo cual se requiere regularidad extra tanto del estado como del estado adjunto. A partir de la hipótesis $f \in L^{\infty}(\Omega)$ y $u_{\alpha} \in C^{1,\rho}(\Omega)$, para algún $\rho \in (0, 1)$ entonces

$$\|u_{\alpha} - u_{\hat{\alpha}}\|_{W^{1,r}} \leq C \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}, \quad (2.22)$$

para todo $\alpha \in V(\alpha^*)$, y para algún $r > 2$. Este resultado de regularidad para sistemas de segundo orden elípticos semilineales puede ser probada bajo ciertos supuestos sobre el dominio (ver [25, Teorema 1, Rem. 14]).

En el caso del estado adjunto, bajo la misma hipótesis y usando el resultado para sistemas elípticos lineales [42], se verifica para cualquier $s > 2$ que

$$\|p_{\alpha}\|_{W^{1,s}} \leq \hat{C}. \quad (2.23)$$

Observe que,

$$\| (h'_{\gamma}(Du) - h'_{\gamma}(D\hat{u})) (D\hat{p}) \|_{L^2(\Omega)} \leq \|h'_{\gamma}(Du) - h'_{\gamma}(D\hat{u})\|_{L^p(\Omega)} \|D\hat{p}\|_{L^q(\Omega)}$$

donde $1/p + 1/q = 1/2$. Ya que $h'_{\gamma}(\cdot)$ es Lipschitz continuo de $L^p(\Omega)$ a $L^p(\Omega)$ (cf. [34]), y por los resultados (2.22) y (2.23) con $r = p, s = q = 2p/(p - 2)$ respectivamente, se obtiene

que

$$\begin{aligned} \|(h'_\gamma(Du) - h'_\gamma(D\hat{u}))(D\hat{p})\|_{L^2(\Omega)} &\leq M_\gamma \|D(u - \hat{u})\|_{L^p(\Omega)} \|p\|_{W^{1,q}(\Omega)}, \\ &\leq M_\gamma M_1 \|u - \hat{u}\|_{W^{1,r}(\Omega)} \|p\|_{W^{1,q}(\Omega)}, \\ &\leq M_2 \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}. \end{aligned}$$

Así, utilizando estos resultados se sigue que

$$\mu \|p - \hat{p}\|_{H^1(\Omega)}^2 \leq C \left\{ \|u - \hat{u}\|_{L^2(\Omega)} + \|P(\alpha) - P(\hat{\alpha})\|_{L^\infty(\Omega)} \|\hat{p}\|_{L^2(\Omega)} + M_2 \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} \right\} \|p - \hat{p}\|_{H^1(\Omega)}.$$

Por otro lado, P es continuamente diferenciable en \mathbb{R}^N , de donde se sigue que P es localmente Lipschitz continua. Esto se emplea para acotar $\|P(\alpha) - P(\hat{\alpha})\|_{L^\infty(\Omega)}$, como se muestra a continuación

$$\|P(\alpha) - P(\alpha^*)\|_{L^\infty(\Omega)} \leq L_1(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}.$$

Así se obtiene que

$$\begin{aligned} \mu \|p - \hat{p}\|_{H^1(\Omega)}^2 &\leq C_1 \{ \|u - \hat{u}\|_{H^1(\Omega)} + L_1(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} \|\hat{p}\|_{H^1(\Omega)} + M_2 \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} \} \|p - \hat{p}\|_{H^1(\Omega)}, \\ \mu \|p - \hat{p}\|_{H^1(\Omega)} &\leq C_2(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} + C_3 \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}, \\ \|p - \hat{p}\|_{H^1(\Omega)} &\leq C_4(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}, \end{aligned}$$

que corresponde al resultado deseado.

En el siguiente resultado, se prueba para η fijo la Lipschitz continuidad local en el caso del modelo TV. Se emplea la definición de Lipschitz continuidad de [23], dada por:

DEFINICIÓN 2.1. F se dice localmente Lipschitz continua en un conjunto X si para cada $x \in X$ existe una vecindad $V(x)$ y una constante $L = L(x)$ tales que

$$\|F(x_1) - F(x_2)\| \leq L(x) \|x_1 - x_2\|, \quad \forall x_1, x_2 \in V(x) \cap X.$$

TEOREMA 2.4. Si $f \in L^\infty(\Omega)$ y $u_\alpha \in C^{1,\rho}(\Omega)$, para algún $\rho \in (0,1)$ entonces $\nabla \mathcal{F}_\eta$ es localmente Lipschitz continua.

DEMOSTRACIÓN 2.4. Sea $\alpha^* \in \mathbb{R}^N$ arbitrario pero fijo y $\alpha, \hat{\alpha} \in V_\delta(\alpha^*)$. Se debe hallar una constante $L(\alpha^*) > 0$ tal que,

$$\|\nabla \mathcal{F}_\eta(\alpha) - \nabla \mathcal{F}_\eta(\hat{\alpha})\| \leq L(\alpha^*) \|\alpha - \hat{\alpha}\|.$$

Se consideran $u := u_\alpha, \hat{u} := u_{\hat{\alpha}}, p := p_\alpha$ y $\hat{p} := p_{\hat{\alpha}}$, donde u_α denota la solución de (2.16b)

asociada con α y p el estado adjunto solución de (2.20) para u_α .

$$\begin{aligned} \|\nabla \mathcal{F}_\eta(\alpha) - \nabla \mathcal{F}_\eta(\hat{\alpha})\| &\leq 2\xi \|\alpha - \hat{\alpha}\| + \left\| \int_{\Omega} \min(0, \eta P(\alpha)) P'(\alpha) - \int_{\Omega} \min(0, \eta P(\hat{\alpha})) P'(\hat{\alpha}) \right\| \\ &\quad + \left\| \int_{\Omega} P'(\alpha) (u - f) p - \int_{\Omega} P'(\hat{\alpha}) (\hat{u} - f) \hat{p} \right\|. \end{aligned}$$

En el caso del segundo término de $\|\nabla \mathcal{F}_\eta(\alpha) - \nabla \mathcal{F}_\eta(\hat{\alpha})\|$ se suma y resta el término

$$\min(0, \eta P(\hat{\alpha})) P'(\alpha)$$

para usar la Lipschitz continuidad del mín. Se obtiene que

$$\begin{aligned} &\left\| \int_{\Omega} \min(0, \eta P(\alpha)) P'(\alpha) - \int_{\Omega} \min(0, \eta P(\hat{\alpha})) P'(\hat{\alpha}) \right\| \\ &\leq \left\| \min(0, \eta P(\alpha)) - \min(0, \eta P(\hat{\alpha})) \right\|_{L^2(\Omega)} \|P'(\alpha)\|_{L^2(\Omega)} \\ &\quad + \left\| \min(0, \eta P(\hat{\alpha})) \right\|_{L^2(\Omega)} \|P'(\alpha) - P'(\hat{\alpha})\|_{L^2(\Omega)}, \\ &\leq \eta \|P(\alpha) - P(\hat{\alpha})\|_{L^2(\Omega)} \|P'(\alpha)\|_{L^2(\Omega)} + M_1 \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}, \\ &\leq L_2(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} \|P'(\alpha)\|_{L^2(\Omega)} + M_1 \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}, \end{aligned}$$

ya que P tiene derivada Lipschitz continua local, se verifica que

$$\begin{aligned} \|P'(\alpha)\|_{L^2(\Omega)} &= \|P'(\alpha) - P'(\alpha^*) + P'(\alpha^*)\|_{L^2(\Omega)}, \\ &\leq \|P'(\alpha) - P'(\alpha^*)\|_{L^2(\Omega)} + \|P'(\alpha^*)\|_{L^2(\Omega)}, \\ &\leq L_2(\alpha^*) \|\alpha - \alpha^*\|_{\mathbb{R}^N} + M_4, \\ &\leq L_2(\alpha^*) \delta + M_4, \end{aligned}$$

lo que nos permite obtener

$$\begin{aligned} &\left\| \int_{\Omega} \min(0, \eta P(\alpha)) P'(\alpha) - \int_{\Omega} \min(0, \eta P(\hat{\alpha})) P'(\hat{\alpha}) \right\| \\ &\leq L_2(\alpha^*) \|\alpha - \hat{\alpha}\| (L_2(\alpha^*) \delta + M_4) + M_1(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}, \\ &\leq M_5(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}. \end{aligned}$$

Ahora, acotemos los dos últimos términos de $\|\nabla \mathcal{F}_\eta(\alpha) - \nabla \mathcal{F}_\eta(\hat{\alpha})\|$,

$$\begin{aligned} \left\| \int_{\Omega} P'(\alpha) (u - f) p - \int_{\Omega} P'(\hat{\alpha}) (\hat{u} - f) \hat{p} \right\| &\leq \underbrace{\int_{\Omega} \|P'(\alpha) u p - P'(\hat{\alpha}) \hat{u} \hat{p}\|}_a \\ &\quad + \underbrace{\int_{\Omega} \|P'(\alpha) p - P'(\hat{\alpha}) \hat{p} f\|}_b. \end{aligned}$$

Analizamos estos términos por separado. En el caso del término a se requiere sumar y restar los términos

$$P'(\alpha) \hat{u} p, \quad P'(\hat{\alpha}) \hat{u} p,$$

para obtener

$$\begin{aligned}
\int_{\Omega} \|P'(\alpha)u p - P'(\hat{\alpha})\hat{u}\hat{p}\| &\leq \|P'(\alpha)\|_{L^\infty(\Omega)} \|u - \hat{u}\|_{L^2(\Omega)} \|p\|_{L^2(\Omega)} \\
&\quad + (\|P'(\alpha) - P'(\hat{\alpha})\|_{L^\infty(\Omega)} \|p\|_{L^2(\Omega)} + \|P'(\hat{\alpha})\|_{L^\infty(\Omega)} \|p - \hat{p}\|_{L^2(\Omega)}) \|\hat{u}\|_{L^2(\Omega)}, \\
&\leq (L_2(\alpha^*)\delta + M_4) \|u - \hat{u}\|_{H^1(\Omega)} \|p\|_{H^1(\Omega)} \\
&\quad + (L_3(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} \|p\|_{H^1(\Omega)} + (L_2(\alpha^*)\delta + M_4) \|p - \hat{p}\|_{H^1(\Omega)}) \|\hat{u}\|_{H^1(\Omega)},
\end{aligned}$$

$$\begin{aligned}
\int_{\Omega} \|P'(\alpha)u p - P'(\hat{\alpha})\hat{u}\hat{p}\| &\leq (L_2(\alpha^*)\delta + M_4) M_1(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} K_1(\alpha^*) \\
&\quad + (L_3(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} K_1(\alpha^*) + (L_2(\alpha^*)\delta + M_4) K_2(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}) C_1(\alpha^*) \\
&\leq M_5(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}.
\end{aligned}$$

Considere ahora el término b , si se suma y restar el término, $P'(\hat{\alpha}) p$ se obtiene

$$\begin{aligned}
\int_{\Omega} \|(P'(\alpha)p - P'(\hat{\alpha})\hat{p})f\| &\leq \|P'(\alpha)p - P'(\hat{\alpha})\hat{p}\|_{L^2(\Omega)} \|f\|_{L^2(\Omega)}, \\
&\leq (\|P'(\alpha) - P'(\hat{\alpha})\|_{L^\infty(\Omega)} \|p\|_{L^2(\Omega)} + \|P'(\hat{\alpha})\|_{L^\infty(\Omega)} \|p - \hat{p}\|_{L^2(\Omega)}) \|f\|_{L^2(\Omega)}, \\
&\leq (L_1(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} \|p\|_{H^1(\Omega)} + (L_2(\alpha^*)\delta + M_4) \|p - \hat{p}\|_{H^1(\Omega)}) C_1, \\
&\leq (L_1(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} K_1(\alpha^*) + (L_2(\alpha^*)\delta + M_4) K_2(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}) C_1, \\
&\leq K_3(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}.
\end{aligned}$$

Por tanto, se sigue que

$$\begin{aligned}
\left\| \int_{\Omega} P'(\alpha)(u - f)p - \int_{\Omega} P'(\hat{\alpha})(\hat{u} - f)\hat{p} \right\| &\leq M_5(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} + K_3(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}, \\
&\leq M_6(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}.
\end{aligned}$$

Finalmente,

$$\begin{aligned}
\|\nabla \mathcal{F}_\eta(\alpha) - \nabla \mathcal{F}_\eta(\hat{\alpha})\| &\leq 2\zeta \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} + M_5(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} + M_6(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}, \\
&\leq L(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N},
\end{aligned}$$

que corresponde al resultado.

2.7. Implementación numérica

En esta sección se abordará la implementación numérica del problema de optimización de parámetros, para lo cual empezaremos por presentar de forma resumida el trabajo desarrollado hasta este punto. A continuación se presentan los esquemas de discretización de las ecuaciones de estado, adjunta y del gradiente. Para finalizar se discuten los resultados y conclusiones obtenidos en los experimentos numéricos, a partir de la implementación de

los algoritmos propuestos para resolver el problema de optimización del parámetro con dependencia espacial presente en el modelo de variación total para el filtrado de ruido.

Los resultados expuestos en las secciones 2.3 y 2.5 nos permiten abordar el problema de optimización binivel (2.2) como un problema de optimización con restricciones de igualdad, el mismo que está determinado por:

$$\min_{\alpha \in \mathbb{R}^N} F(u_\alpha) + \frac{1}{2\eta} \|\min(0, \eta P(\alpha))\|_{L^2(\Omega)}^2$$

sujeto a

$$\mu(u, \phi)_{H^1(\Omega)} + (h_\gamma(Du), D\phi)_{L^2(\Omega)} + \int_{\Omega} P(\alpha)(u - f)\phi \, dx = 0, \text{ para todo } \phi \in H^1(\Omega),$$

donde

$$F(u_\alpha) = \frac{1}{2} \|f_0 - u\|_{L^2(\Omega)}^2 + \xi \|\alpha\|_{\mathbb{R}^N}^2.$$

Como se explico previamente la solución numérica de este problema de optimización se puede abordar desde dos puntos de vista, en el primero se puede utilizar un método de Newton para resolver el sistema de optimalidad asociado, y en el segundo, consiste en usar un algoritmo iterativo para la solución del problema de optimización en si. Usaremos la segunda opción empleando el método quasi-Newton BFGS, esto gracias a la caracterización que se encontró del gradiente del funcional de costo reducido (2.21). El esquema del método de BFGS requiere resolver en cada iteración la ecuación de estado y la ecuación adjunta hasta que se verifique el criterio de parada, por lo cual empezaremos por presentar los algoritmos empleados para resolver estas ecuaciones.

La ecuación de estado (2.16b) corresponde a la condición de optimalidad de primer orden del problema de optimización para el filtrado de ruido (2.3). Para encontrar la solución numérica a esta ecuación no lineal se emplea el método de Newton. Sea q una variable auxiliar, se define el operador H como

$$H(u, q) = \begin{pmatrix} \mu(u, \phi)_{H^1(\Omega)} + (q, D\phi)_{L^2(\Omega)} + \int_{\Omega} P(\alpha)(u - f)\phi \, dx \\ q - h_\gamma(Du) \end{pmatrix},$$

así la ecuación de estado será $H(u, q) = 0$. El esquema del método de Newton para resolver esta ecuación no lineal es

$$H'(u, q)(\delta u, \delta q) = -H(u, q),$$

donde $(\delta u, \delta q)$ corresponde a la dirección en la cual se deriva el operador H ,

$$H'(u, q)(\delta u, \delta q) = \begin{pmatrix} \mu(\delta u, \phi)_{H^1(\Omega)} + \int_{\Omega} P(\alpha)\delta u\phi \, dx + (\delta q, D\phi)_{L^2(\Omega)} \\ -h'_\gamma(Du)D\delta u + \delta q \end{pmatrix},$$

$h_\gamma(\cdot)$ denota la derivada de la regularización de Huber $|\cdot|_\gamma$ determinado por (2.5) y $h'_\gamma(\cdot)$

su derivada descrita en (2.5) como sigue

$$h'_\gamma(Du)D\phi = \begin{cases} \frac{D\phi}{|Du|} - \frac{\langle Du, D\phi \rangle Du}{|Du|^3} & A\gamma, \\ (1 - \frac{\gamma}{2}\theta^2(Du)) \left(\frac{D\phi}{|Du|} - \frac{Du\langle D\phi, Du \rangle}{|Du|^3} \right) + \gamma^2\theta(Du) \frac{\langle Du, D\phi \rangle Du}{|Du|^2} & S\gamma, \\ \gamma D\phi & I\gamma. \end{cases}$$

Esta matriz deberá ser modificada con una proyección para garantizar que la matriz $H'(u, q)$ sea definida positiva. Como se muestra en el trabajo [7] la proyección consiste en reemplazar

$$\frac{\langle Du, D\phi \rangle Du}{|Du|^3} = \frac{Du \otimes Du}{|Du|^3} D\phi,$$

por

$$\frac{q}{\max(1, |q|)} \otimes \frac{Du}{|Du|^2} D\phi.$$

Para la implementación numérica se empleó esquemas de discretización con diferencias finitas, ampliamente utilizado en el procesamiento matemático de imágenes debido a la estructura de la imagen digital. Se consideran $\Omega = (0, 1) \times (0, 1)$ el dominio de las imágenes en escala de grises, $n \times n$ la dimensión de las mismas, $h = 1/(n + 1)$ el paso de discretización de la malla, \mathbf{G} matriz de discretización del gradiente, \mathbf{A} la matriz de discretización del laplaciano, div matriz de discretización de la divergencia y \mathbf{I} la matriz identidad. El algoritmo para resolver la ecuación de estado para α fijo se presenta a continuación:

Algorithm 1 Método de Newton–filtrado de ruido (ecuación de estado)

- 1: Dados f, α, μ, γ
- 2: Evaluar $P(\alpha)$
- 3: **Inicializar** $k = 0, u_0, q_0$
- 4: **Repetir**
- 5: $H'(u_k, q_k)(\Delta u_k, \Delta q_k) = -H(u_k, q_k)$, con

$$H(u, q) = \begin{pmatrix} -\mu \mathbf{A} u + \mu u - \mathit{div} q + P(\alpha)(u - f) \\ q - h_\gamma(\mathbf{G} u) \end{pmatrix},$$

$$H'(u, q) = \begin{pmatrix} -\mu \mathbf{A} + \mu \mathbf{I} + P(\alpha) \mathbf{I} & -\mathit{div} \mathbf{I} \\ -h'_\gamma(\mathbf{G} u) \mathbf{G} & \mathbf{I} \end{pmatrix}.$$

- 6: **Actualizar**
 - 7: $(u_{k+1}, q_{k+1}) = (u_k, q_k) + (\Delta u_k, \Delta q_k)$
 - 8: $k = k + 1$
 - 9: **Hasta** verificar criterio de parada $\|(\Delta u_k, \Delta q_k)\| < tol$
 - 10: **Retornar** u_k .
-

La ecuación adjunta (2.20), corresponde a una ecuación lineal respecto a la variable p . Esta ecuación se puede describir utilizando una variable auxiliar r tal que

$$\begin{aligned} \mu(p, \phi)_{H^1(\Omega)} + (r, D\phi) + \int_{\Omega} P(\alpha) p \phi &= -(u - f_o, \phi), \\ r &= h'_{\gamma}(Du)Dp, \end{aligned}$$

donde f_o es la imagen original o una aproximación de la misma y $h'_{\gamma}(\cdot)$ está determinado por (2.5). Usando esquemas de discretización en diferencias finitas se obtiene el siguiente algoritmo para determinar el estado adjunto p , para α fijo

Algorithm 2 Ecuación adjunta

- 1: Dados $f_o, u, \alpha, \mu, \gamma$
- 2: Evaluar $P(\alpha)$
- 3: **Resolver** el sistema de ecuaciones

$$\begin{pmatrix} -\mu\mathbb{A} + \mu\mathbb{I} + P(\alpha)\mathbb{I} & -div \\ -h'_{\gamma}(Gu)G & \mathbb{I} \end{pmatrix} \begin{pmatrix} p \\ r \end{pmatrix} = \begin{pmatrix} -u + f_o \\ \mathbf{0} \end{pmatrix}$$

- 4: **Retornar** p .
-

Gracias a la caracterización del gradiente del funcional de costo reducido obtenida en (2.21), es decir,

$$\nabla \mathcal{F}_{\eta}(\alpha) = 2\xi\alpha + \int_{\Omega} \text{mín}(0, \eta P(\alpha)) P'(\alpha) dx + \int_{\Omega} P'(\alpha) (u - f) p dx$$

con u y p soluciones de la ecuación de estado y adjunta para α , respectivamente, se propone utilizar el método BFGS para resolver el problema de optimización de parámetros con dependencia espacial. Las integrales dobles se aproximaron usando la regla de punto medio (RPM) determinada por:

$$\int_a^b \int_c^d f(x, y) dy dx \approx RPM(f, h)$$

donde,

$$\begin{aligned} RPM(f, h) &= \frac{1}{4} h^2 \{ f(a, c) + f(b, c) + f(a, d) + f(b, d) + \\ &2 \sum_{i=1}^{n-2} f(x_i, c) + 2 \sum_{i=1}^{n-2} f(x_i, d) + 2 \sum_{i=1}^{n-2} f(a, y_i) + \\ &2 \sum_{i=1}^{n-2} f(b, y_i) + 4 \sum_{i=1}^{n-2} \sum_{j=1}^{n-2} f(x_i, y_j) \}. \end{aligned}$$

Así se obtiene la aproximación del gradiente,

$$\nabla \mathcal{F}_{\eta}(\alpha) \approx 2\xi\alpha + RPM(\text{mín}(0, \eta P(\alpha)) P'(\alpha)) + RPM(P'(\alpha) (u - f) p),$$

donde $P'(\cdot)$ corresponde a la derivada de P respecto al parámetro. Siguiendo las ideas presentadas en el trabajo [14] se eligió como criterio de parada

$$\frac{\|\alpha_{k+1} - \alpha_k\|_{\mathbb{R}^N}}{\|\alpha_{k+1}\|_{\mathbb{R}^N}} < tol$$

y se fija la tolerancia como $tol = 10^{-6}$.

A continuación se presenta el algoritmo

Algorithm 3 BFGS optimización de parámetros

- 1: Dados $f, f_o, \mu, \gamma, \eta, \xi$
- 2: **Inicializar** $k = 0, B_0, \alpha_0$
- 3: **Repetir**
- 4: Calcular $P(\alpha_k)$ y $P'(\alpha_k)$
- 5: Resolver la ecuación de estado usando el algoritmo 1, u_k
- 6: Resolver la ecuación adjunta usando el algoritmo 2, p_k
- 7: Calcular $\nabla \mathcal{F}_\eta(\alpha_k)$,
- 8: $s_k = \alpha_{k+1} - \alpha_k$,
- 9: $z_k = \nabla \mathcal{F}_\eta(\alpha_{k+1}) - \nabla \mathcal{F}_\eta(\alpha_k)$
- 10: **If** $s_k^T z_k > 0$ **then**

$$B_{k+1} = B_k + \frac{(s_k - B_k z_k) s_k^T + s_k (s_k - B_k z_k)^T}{s_k^T z_k} - \frac{(s_k - B_k z_k)^T z_k}{(s_k^T z_k)^2} s_k s_k^T.$$

- 11: Calcular la dirección de descenso como
 - 12: $d_k = -B_{k+1} \nabla \mathcal{F}_\eta(\alpha_k)$
 - 13: **else**
 - 14: Calcular la dirección de descenso como
 - 15: $d_k = -B_k \nabla \mathcal{F}_\eta(\alpha_k)$
 - 16: **End**
 - 17: Determinar el paso de descenso λ usando una de búsqueda lineal
 - 18: **Actualizar**
 - 19: $\alpha_{k+1} = \alpha_k + \lambda d_k$
 - 20: $k = k + 1$
 - 21: **Hasta** verificar criterio de parada.
 - 22: **Retornar** λ_k .
-

OBSERVACIÓN. Las preguntas que debemos responder ahora son: ¿se puede considerar cualquier función $P: \mathbb{R}^n \rightarrow C(\Omega)$ como parámetro de regularización?, ¿la dificultad de hallar el parámetro α depende de la función P que se eligió?. Para responder a estas preguntas se presenta a continuación el análisis del problema de optimización en una dimensión: sea $0 \leq x \leq 1, \nu \in \mathbb{R}$, resolver

$$\min_{\alpha \in \mathbb{R}^N} \|P(\alpha)(x) - \nu\|^2.$$

Se consideró para el análisis las siguientes funciones:

- $P(\alpha) = \alpha_1 x + \alpha_2$,
- $P(\alpha) = \alpha_1 x^2 + \alpha_2 x$,
- $P(\alpha) = \alpha_1 \sin(x) + \alpha_2 \cos(x)$,
- $P(\alpha) = \alpha_1 \sin(\alpha_2 x)$.

Las funciones lineales respecto a α no generan inconveniente, pues el funcional objetivo

$$\|P(\alpha)(x) - v\|^2$$

es convexo como se observa en la Figura 2.1, lo que implica que cualquier mínimo local es un mínimo global.

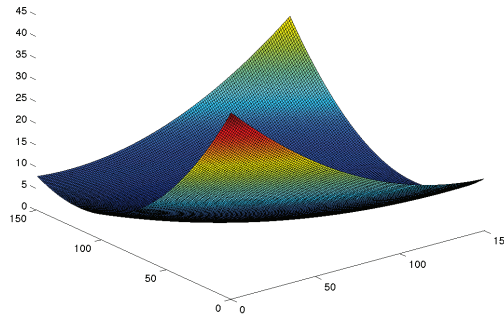


Figura 2.1. Funcional objetivo considerando $P(\alpha) = \alpha_1 \sin(x) + \alpha_2 \cos(x)$.

Por otro lado la última función, $P(\alpha) = \alpha_1 \sin(\alpha_2 x)$, hace que el funcional objetivo presente oscilaciones que tienen un efecto devastador para el problema de optimización, pues dificulta la estimación de los parámetros, ya que nos encontramos con diferentes mínimos del problema para puntos iniciales diferentes, numéricamente hablando. Para resolver este problema se puede convexificar la función usando una regularización. A manera de ejemplo se muestra la regularización de Tikhonov,

$$\min_{\alpha \in \mathbb{R}^2} \|\alpha_1 \sin(\alpha_2 x) - v\|^2 + \underbrace{\zeta_1 \|\alpha_1\|^2 + \zeta_2 \|\alpha_2\|^2}_{\text{R. Tikhonov}}.$$

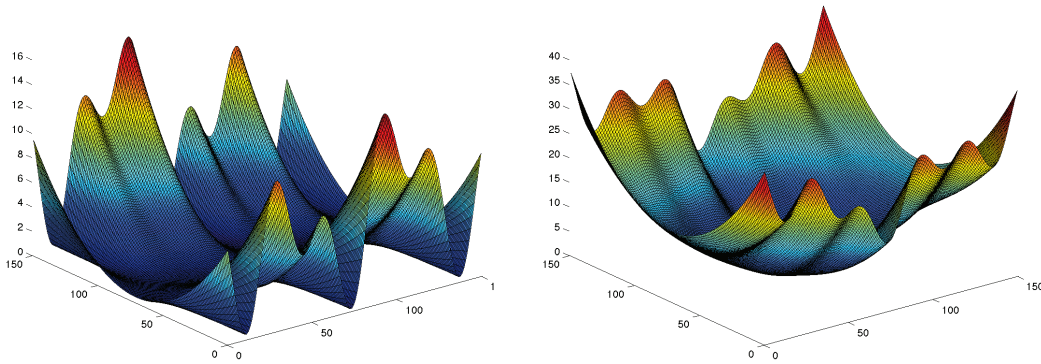


Figura 2.2. A la izquierda, el funcional original y a la derecha el funcional regularizado.

En la Figura 2.2 se puede observar este efecto. Para no incluir términos de regularización que implican la presencia de más parámetros, nos enfocaremos en funciones lineales respecto al parámetro α .

Los parámetros empleados en los experimentos son: número de puntos de la malla $n = 128$, tamaño de la malla $h = 1/(n + 1)$, término de Tikhonov $\zeta = 1e - 10$, regularización Moreau Yosida $\eta = 1e4$, regularización de Huber $\gamma = 1e4$ y la regularización elíptica $\mu = 1e - 12$.

Para la comparación de las imágenes resultantes se usaron medidas de calidad como: la relación pico de señal a ruido PSNR (Peak signal-to-noise ratio), la relación señal ruido SNR (Signal to noise ratio) y la similitud estructural SSIM.

PSNR es un término utilizado como medida de la calidad de la reconstrucción en el área de la compresión de imágenes, se define como:

$$PSNR(I, K) = 10 \cdot \log_{10} \left(\frac{MAX_I^2}{MSE} \right) = 20 \cdot \log_{10} \left(\frac{MAX_I}{\sqrt{MSE}} \right)$$

donde MSE es el error cuadrático medio, entre dos imágenes I y K de tamaño $M \times N$ definido por

$$MSE(I, K) = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} ||I(i, j) - K(i, j)||^2,$$

MAXI es el máximo valor que puede tomar un píxel en la imagen. Para una imagen en formato RGB, la definición del PSNR es la misma, pero el MSE se calcula como la media aritmética de los MSEs de los tres colores (R, G y B). Los valores típicos que adopta este parámetro están entre 30 y 50, siendo mayor cuanto mejor es la codificación.

SSIM es una medida que busca cuantificar de forma numérica la calidad visual de una

imagen para un observador humano. Está dado por la expresión:

$$SSIM(X, Y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)},$$

con μ_x la media de X , μ_y la media de Y , σ_x^2 la varianza de X , σ_y^2 la varianza de Y , σ_{xy} la covarianza de X, Y , $c_1 = (k_1L)^2$, $c_2 = (k_2L)^2$ variables para estabilizar la división, L rango del valor de los pixeles y $k_1 = 0,01, k_2 = 0,03$ constantes. El valor SSIM está entre 0 y 1. A mayor valor, más similares son las imágenes.

SNR es una medida usada en ciencia e ingeniería para medir el nivel de una señal deseada con el nivel de ruido,

$$SNR(f) = 20 \times \log_{10} \left(\frac{\|f_o\|_{L^2(\Omega)}}{\|f - f_o\|_{L^2(\Omega)}} \right).$$

Las imágenes que se emplean en los experimentos presentan ruido no uniforme, estas se obtuvieron usando simulación de ruido, que consiste en: dada una imagen original, se perturba sumando dos variables aleatorias con distribución Gaussiana de media 0 y desviación estándar σ_1 y σ_2 en distintas localizaciones, para obtener ruido no uniforme. Para que sea un reto mayor el filtrado de ruido, se consideran regiones de tamaño a , y el esquema se resume en la Figura 3.4.

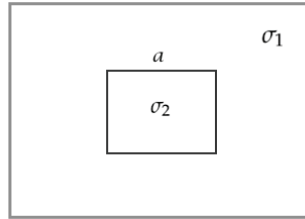


Figura 2.3. Simulación de ruido no uniforme.

Se utilizaron imágenes f_o con distintas características como: brillo, bordes, contenido y textura. Lo que permitió analizar el comportamiento de los algoritmos propuestos en este trabajo. La implementación numérica se realizó haciendo uso del software de programación Matlab, en un equipo con procesador de Intel(R) Core(TM) i5 de 2.5 GHz con 4GB de memoria RAM.

2.7.1. Experimento 1

En este experimento se mostrará la importancia de emplear parámetros con dependencia espacial en lugar de parámetros escalares para filtrar imágenes con ruido no uniforme. Para esto se resolvió el problema de optimización de parámetros escalares, es decir, $P(\alpha) = \alpha$ y para el problema con dependencia espacial $P(\alpha) = \alpha_1x + \alpha_2y + \alpha_3$.

Se considera una imagen con detalles para apreciar visualmente el efecto. La imagen original fue perturbada sumando dos variables aleatorias con distribución Gaussiana de media 0 y desviación estándar $\sigma_1 = 0.002$, $\sigma_2 = 0.02$, en distintas localizaciones para obtener ruido no uniforme, como se observa en la Figura 2.4,

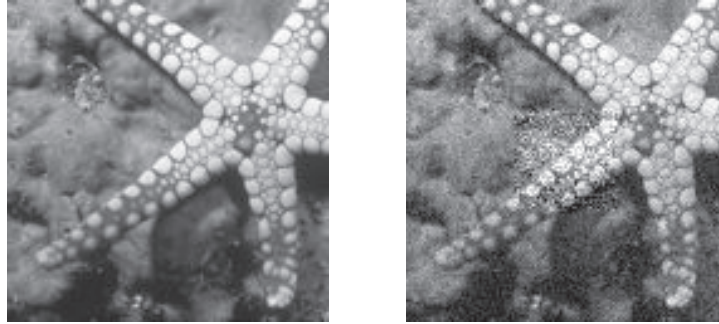


Figura 2.4. Imagen original e imagen con ruido no uniforme.

En la Tabla 2.1 se muestran los resultados obtenidos usando el modelo de variación total para el filtrado del ruido con parámetros escalares y con dependencia espacial. Se fija como punto inicial $\alpha_0 = 1500$ en el caso escalar y en el caso espacial $\alpha_0 = 1500(1, 1, 1)$. Como se esperaba los resultados obtenidos considerando el parámetro con dependencia espacial son mejores.

Parámetro $P(\alpha)$	SNR	PSNR	SSIM	Parámetro óptimo α^*	F. objetivo
α	39.71	25.987	0.803	2054.0	1.618e-3
$\alpha_1x + \alpha_2y + \alpha_3$	40.21	26.234	0.814	(886.59,734.35,1554.0)	1.487e-3

Tabla 2.1. Comparación del parámetro escalar y con dependencia espacial.

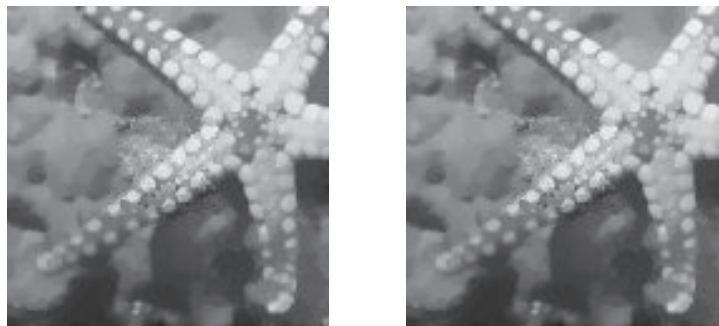


Figura 2.5. A la izquierda imagen solución con α escalar y a la derecha la solución considerando dependencia espacial.

Consideremos ahora una perturbación de mayor tamaño, $a = 40$, para la cual se considero variables aleatorias con distribución Gaussiana de media 0 y desviación estándar $\sigma_1 = 0.001$, $\sigma_2 = 0.01$, en distintas localizaciones para obtener ruido no uniforme como se observa en la Figura 2.6,

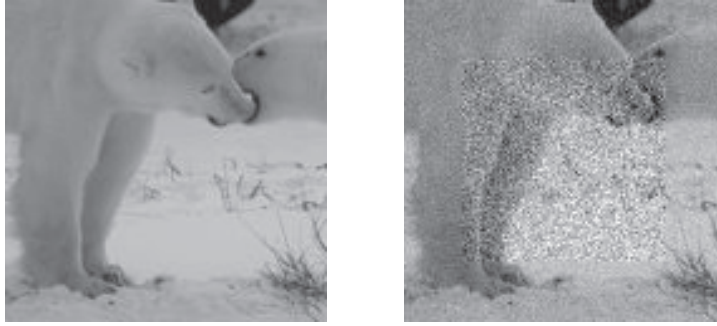


Figura 2.6. Imagen original e imagen con ruido no uniforme.

En la Tabla 2.2 se muestran los resultados obtenidos usando el modelo de variación total para el filtrado del ruido con parámetros escalares y con dependencia espacial. Se fija como punto inicial $\alpha_0 = 1000$ en el caso escalar y en el caso espacial $\alpha_0 = 1000(1, 1, 1)$. Como se esperaba los resultados obtenidos considerando el parámetro con dependencia espacial son mejores, respecto del valor de SSIM.

Parámetro $P(\alpha)$	SNR	PSNR	SSIM	Parámetro óptimo α^*	F. objetivo
α	49.2651	27.432	0.8088	13618.0	6.979e-3
$\alpha_1x + \alpha_2y + \alpha_3$	50.0049	27.802	0.8151	(670.65, 649.36, 797.19)	6.213e-4

Tabla 2.2. Comparación del parámetro escalar y con dependencia espacial.

Visualmente los resultados nos muestran que los bordes no están definidos, en el siguiente capítulo analizaremos el modelo de variación total generalizada que nos permitirá mejorar estos resultados.

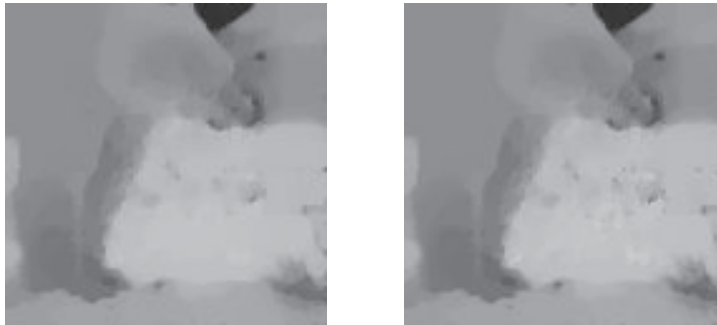


Figura 2.7. A la izquierda imagen solución con α escalar y a la derecha la solución considerando dependencia espacial.

2.7.2. Experimento 2

Este experimento se realizó para diferentes funciones $P(\cdot)$ y tiene como objetivo analizar la convergencia del método propuesto y ver la influencia de estas en el filtrado de ruido. La imagen que se considero para este experimento contiene tres regiones de dimensión $a = 10$

con distintas localizaciones con el fin de evaluar la capacidad del algoritmo para recuperar los detalles. Las variables aleatorias que se usaron para representar ruido no uniforme tienen distribución Gaussiana de media 0 y desviación estándar $\sigma_1 = 0.0005$, $\sigma_2 = 0.005$, como se observa en la Figura 2.8,



Figura 2.8. Imagen original e imagen con ruido no uniforme.

En la Tabla 2.3 se muestran los resultados obtenidos para diferentes funciones $P(\cdot)$, los mismos que se comparan respecto de la medida SSIM. La condición inicial en el caso del parámetro escalar fue $\alpha_0 = 500$ y el caso con dependencia espacial $\alpha_0 = 500(1, 1, 1)$. El mejor resultado se obtuvo utilizando la función $P(\alpha) = \alpha_1 \sin(x) + \alpha_2 \cos(y) + \alpha_3$ que alcanzó 0.8196 en el índice SSIM. En el caso de la función trigonométrica al algoritmo le toma más iteraciones para determinar el parámetro óptimo.

Parámetro $P(\alpha)$	SNR	PSNR	SSIM	Parámetro óptimo α^*	F. objetivo	Número de iteraciones
α	44.467	27.736	0.7816	1680.2	1.011e-3	9
$\alpha_1 x + \alpha_2 y + \alpha_3$	45.849	28.427	0.8091	(530.84, 796.85, 1332.5)	8.928e-4	15
$\alpha_1 x^2 + \alpha_2 y^2 + \alpha_3$	45.408	28.207	0.7992	(-112.41, 196.46, 1862.0)	1.008e-3	3
$\alpha_1 \sin(x) + \alpha_2 \cos(y) + \alpha_3$	45.953	28.479	0.8196	(482.87, 888.43, 1209.4)	8.641e-4	14

Tabla 2.3. Número de iteraciones para distintos valores del parámetro.

En la Figura 2.9 se comparan las imágenes obtenidas para las diferentes funciones $P(\cdot)$.



Figura 2.9. Resultados obtenidos para distintos $P(\cdot)$.

2.7.3. Experimento 3

Este experimento se realizó considerando diferentes puntos iniciales. Esto tiene como objetivo analizar la sensibilidad respecto a las condiciones iniciales del método propuesto. Se usó la imagen perturbada sumando dos variables aleatorias con distribución Gaussiana de media 0 y desviación estándar $\sigma_1 = 0.002$, $\sigma_2 = 0.02$, en distintas localizaciones para obtener ruido no uniforme como se observa en la Figura 2.10,

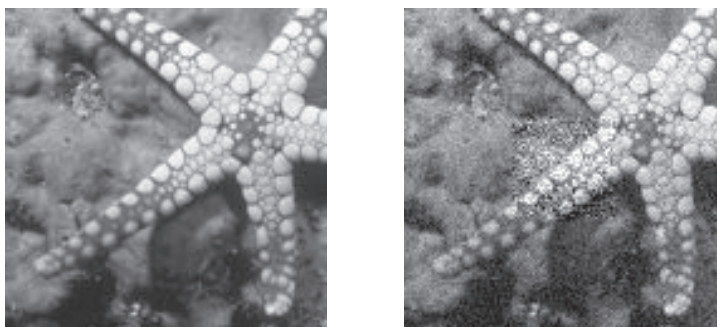


Figura 2.10. Imagen original e imagen con ruido no uniforme.

A continuación, en la Tabla 2.4, se muestran los resultados para distintos puntos iniciales en el algoritmo de optimización de parámetros para $P(\alpha) = \alpha_1 x + \alpha_2 y + \alpha_3$.

Condición inicial α_0	Número de iteraciones	PSNR	SNR	Parámetro óptimo α^*	F. objetivo
0	–	–	–	–	–
100(1, 1, 1)	13	26.22	40.20	(979.52, 568.97, 1492.73)	1.48e-3
500(1, 1, 1)	22	26.24	40.21	(847.18, 522.91, 1580.80)	1.47e-3
1000(1, 1, 1)	11	26.24	40.22	(803.40, 541.56, 1597.50)	1.47e-3
1500(1, 1, 1)	4	26.12	39.98	(872.99, 678.31, 1424.9)	1.48e-3
2000(1, 1, 1)	6	26.24	40.21	(662.87, 346.40, 1735.50)	1.54e-3
(0, 0, 2054)	–	–	–	–	–
(100, 100, 2054)	3	25.57	38.89	(979.52, 568.97, 1492.73)	1.48e-3

Tabla 2.4. Comparación de resultados para distintas condiciones iniciales.

Se evidenció la necesidad de inicializar el algoritmo en un valor distinto a cero, caso contrario no se puede garantizar la convergencia del método mucho menos resolver el problema del filtrado de ruido.

Los experimentos nos han permitido mostrar que los algoritmos planteados son potentes, pues son capaces de recuperar la imagen del ruido utilizando menos de 20 iteraciones lo que es eficiente, considerando que las matrices que se manipulan, en cada iteración, son de dimensión $3n^2 \times 3n^2$, con n la dimensión de las imágenes.

Capítulo 3

Modelo de variación total generalizada (TGV)

En este capítulo, nos enfocaremos en el análisis del modelo de variación total generalizada TGV, donde se abordarán temas de regularización debido a la no diferenciabilidad del modelo de filtrado de ruido. Esto permite obtener una condición necesaria y suficiente del modelo TGV. Se exponen algunos resultados teóricos como la existencia de soluciones del problema de optimización de parámetros y la existencia de multiplicadores de Lagrange del problema de optimización no lineal con restricciones de igualdad y desigualdad. Se evidencia que los multiplicadores asociados a las restricciones de positividad del parámetro de optimización son medidas, que computacionalmente hablando son muy costosas de implementar. Para superar esto se propone introducir la *Regularización de Moreau-Yosida* [32]. Finalmente, se establece el sistema de optimalidad asociado al problema regularizado y se muestra que las soluciones de los problemas regularizados convergen hacia la solución del problema original.

3.1. Regularización de variación total generalizada

La regularización de variación total (TV) [39] preserva los bordes y se comporta muy bien si la imagen a ser reconstruida, vista como función, es constante a trozos. Sin embargo, si se consideran funciones lineales a trozos, que es el caso de las imágenes, se produce un efecto de escalonamiento [38]. Una posibilidad para contrarrestar tal efecto es introducir derivadas de orden superior en la regularización. Así, la *regularización de variación total generalizada (TGV)* fue introducida como una variante de orden superior de la regularización de variación total en el procesamiento digital de imágenes. En este trabajo, nos concentraremos en el modelo de variación total generalizada de segundo orden TGV^2 [5].

Sean Ω un subconjunto acotado de \mathbb{R}^2 , $f : \Omega \rightarrow \mathbb{R}$ la imagen, en escala de grises, observada y $BV(\Omega)$ el conjunto de las funciones de variación acotada. El modelo TGV^2 para el filtrado de ruido consiste en encontrar $u \in BV(\Omega)$, es decir una función con variación

acotada, solución del problema de minimización,

$$\min_{u \in BV(\Omega)} TGV_{\alpha, \beta}^2(u) + \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2, \quad (3.1)$$

donde α, β son parámetros positivos y

$$TGV_{\alpha, \beta}^2(u) := \min_{w \in BD(\Omega)} \alpha \|Du - w\|_{M(\Omega; \mathbb{R}^2)} + \beta \|Ew\|_{M(\Omega; Sym(\mathbb{R}^2))}, \quad (3.2)$$

con BD el conjunto de las funciones con deformación acotada,

$$BD(\Omega) := \{w \in L^1(\Omega; \mathbb{R}^n) : \|Ew\|_{M(\Omega; Sym(\mathbb{R}^2))} < \infty\},$$

E denota el gradiente simetrizado, $Sym(\mathbb{R}^2)$ el espacio de tensores simétricos de orden 2 con argumentos en \mathbb{R}^2 , $\mathcal{M}(\Omega; Sym(\mathbb{R}^2))$ el espacio de las medidas de Borel con valores en $Sym(\mathbb{R}^2)$ y

$$\|Du\|_{M(\Omega; \mathbb{R}^2)} = \sup \left\{ \int_{\Omega} u \nabla \cdot \phi \, dx : \phi \in C_0^\infty(\Omega, \mathbb{R}^2), \|\phi\|_{C_0^\infty(\Omega, \mathbb{R}^2)} \leq 1 \right\},$$

$$\|Ew\|_{M(\Omega; Sym(\mathbb{R}^2))} = \sup \left\{ \int_{\Omega} w \nabla \cdot \phi \, dx : \phi \in C_0^\infty(\Omega, Sym(\mathbb{R}^2)), \|\phi\|_{C_0^\infty(\Omega, Sym(\mathbb{R}^2))} \leq 1 \right\}.$$

El primer término de (3.1) cuantifica la variación dentro de una imagen y tiene como objetivo que las esquinas y bordes queden lo mejor definidos, mientras que el segundo, garantiza que la imagen solución u preserve las características de la imagen observada f , este se conoce como término de fidelidad. Los parámetros de regularización α, β determinan la influencia de los términos en (3.1), por lo que su valor es crítico, pues si este es grande se obtiene una solución sobrerregularizada y su suavizado será excesivo. Por el contrario, si su valor es pequeño no se resolverá el problema de ruido.

3.2. Planteamiento del problema

Los parámetros α, β se consideran usualmente como valores reales. En este trabajo se propone considerar la dependencia espacial de los parámetros, con el objetivo de filtrar ruido no uniforme en una imagen, lo que nos permite acercarnos más a situaciones reales donde no se conocen el tipo ni la distribución de ruido que puede cubrir información relevante. Así, consideramos los pesos de (3.1) como funciones continuas $P(\alpha)$ y $Q(\beta)$ que tienen como dominio un espacio vectorial de dimensión finita \mathbb{R}^n y \mathbb{R}^m , respectivamente. La idea de considerar espacios de dimensión finita surge de los dispositivos médicos, donde el número de valores que se pueden calibrar para adquirir una imagen es finito.

La idea de plantear un problema de optimización, para determinar el valor óptimo de los parámetros reales del modelo TGV^2 para el filtrado de ruido considerando conjuntos de entrenamiento, fue propuesta en [14]. Siguiendo con esta idea, el problema de optimización

binivel para determinar $P(\alpha)$ y $Q(\beta)$, tiene como restricción el modelo del filtrado de ruido TGV^2 y será:

$$\min_{(\alpha, \beta) \in \mathbb{R}^{N+M}} \frac{1}{2} \|f_0 - u_{\alpha, \beta}\|_{L^2(\Omega)} + \xi (\|\alpha\|_{\mathbb{R}^N}^2 + \|\beta\|_{\mathbb{R}^M}^2) := F(u_{\alpha, \beta}) \quad (3.3a)$$

sujeto a

$$u_{\alpha, \beta} \in \arg \min_{u \in BV(\Omega)} \mathcal{J}(u, P(\alpha), Q(\beta)) := TGV_{P(\alpha), Q(\beta)}^2(u) + \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2 \quad (3.3b)$$

$$P(\alpha), Q(\beta) \geq 0, \text{ en todo punto de } \Omega \quad (3.3c)$$

donde f_0 es una imagen tomada con la máxima precisión posible, esta corresponde al conjunto de entrenamiento que en este caso simplificado se considera un conjunto de un elemento. La idea de trabajar con conjuntos de entrenamiento proviene del *Machine Learning*. Esta técnica consiste en aprender información de una muestra de datos que se aplicará luego a los términos desconocidos.

3.3. Análisis del problema binivel

Las dificultades que surgen de inmediato al tratar un problema binivel no lineal en espacios de dimensión infinita son la no diferenciabilidad de la restricción (3.3b), lo que hace imposible verificar condiciones de calificación para sistemas de Karush-Kuhn-Tucker (KKT). En este trabajo se propone utilizar técnicas de regularización para demostrar la existencia de los multiplicadores de Lagrange y determinar así el sistema de optimalidad asociado al problema binivel que permita estimar el valor óptimo de los parámetros.

Empecemos por reemplazar el término TGV^2 de la restricción (3.3b) por su versión regularizada $TGV^{2, \gamma, \mu}$, con parámetro $\gamma > 0$ asociado a la regularización de Huber y parámetro $\mu > 0$ asociado al término elíptico,

$$TGV_{\alpha, \beta}^{2, \gamma, \mu}(u) := \min_{w \in \mathbb{H}^1(\Omega)} \int_{\Omega} P(\alpha) |Du - w|_{\gamma} + \int_{\Omega} Q(\beta) |Ew|_{\gamma} + \frac{\mu}{2} (\|u\|_{H^1(\Omega)}^2 + \|w\|_{H^1(\Omega)}^2), \quad (3.4)$$

donde $\mathbb{H}^1(\Omega) = H^1(\Omega; \mathbb{R}^N)$ y la regularización de Huber $C^1 |\cdot|_{\gamma}: \mathbb{R}^n \rightarrow \mathbb{R}$ esta definida por,

$$|z|_{\gamma} = \begin{cases} \|z\| + \frac{\gamma}{2} t_1^2 - t_2 + A & \text{si } \gamma \|z\| \geq 1 + \frac{1}{2\gamma}, \\ B \|z\| + C \|z\|^2 - \frac{\gamma^3}{6} \|z\|^3 + D & \text{si } 1 - \frac{1}{2\gamma} \leq \gamma \|z\| \leq 1 + \frac{1}{2\gamma}, \\ \frac{\gamma}{2} \|z\|^2 & \text{si } \gamma \|z\| \leq 1 - \frac{1}{2\gamma}, \end{cases}$$

$\|\cdot\|$ corresponde a la norma euclidiana en \mathbb{R}^n , $t_1 = \frac{1}{\gamma}(1 - \frac{1}{2\gamma})$, $t_2 = \frac{1}{\gamma}(1 + \frac{1}{2\gamma})$,

$$\begin{aligned} A &= B(t_1 - t_2) + \frac{C}{2}(t_1^2 - t_2^2) - \frac{\gamma^3}{6}(t_1^3 - t_2^3), & B &= 1 - \frac{(2\gamma + 1)^2}{8\gamma}, \\ C &= \frac{\gamma}{2}(2\gamma + 1), & D &= \left(\frac{\gamma}{2} - \frac{B}{2}\right)t_1^2 - B\|t_1\| + \frac{\gamma^3}{6}\|t_1\|^3. \end{aligned}$$

Considerar $TGV^{2,\gamma,\mu}$ permite que el problema de optimización inferior (3.3b) se pueda reemplazar por su condición necesaria y suficiente de optimalidad. Se obtiene así el siguiente problema de optimización con restricciones,

$$\min_{(\alpha,\beta) \in \mathbb{R}^{N+M}} F(u_{\alpha,\beta}) \quad (3.5a)$$

sujeto a

$$\begin{aligned} a(y, \Psi) + \int_{\Omega} P(\alpha)h_{\gamma}(Du - w)(D\phi - \phi) dx + \int_{\Omega} Q(\beta)h_{\gamma}(Ew)E\phi dx \\ + \int_{\Omega} (u - f)\phi dx = 0, \text{ para todo } \Psi \in Y, \end{aligned} \quad (3.5b)$$

$$P(\alpha), Q(\beta) \geq 0, \text{ en todo punto de } \Omega \quad (3.5c)$$

donde $a(y, y) = \mu \left(\|u\|_{H^1(\Omega)}^2 + \|w\|_{\mathbb{H}^1(\Omega)}^2 \right)$, $Y = H^1(\Omega) \times \mathbb{H}^1(\Omega)$, $y = (u, w) \in Y$, $\Psi = (\phi, \varphi) \in Y$ y h_{γ} la derivada de $|\cdot|_{\gamma}$ determinada por,

$$h_{\gamma}(z) = \begin{cases} \frac{z}{\|z\|} & \text{si } \gamma\|z\| \geq 1 + \frac{1}{2\gamma}, \\ \frac{z}{\|z\|} \left(1 - \frac{\gamma}{2}(1 - \gamma\|z\| + \frac{1}{2\gamma})^2 \right) & \text{si } 1 - \frac{1}{2\gamma} \leq \gamma\|z\| \leq 1 + \frac{1}{2\gamma}, \\ \gamma z & \text{si } \gamma\|z\| \leq 1 - \frac{1}{2\gamma}. \end{cases} \quad (3.6)$$

Así, $h_{\gamma}(Du)$ corresponde a una regularización de conjuntos activos e inactivos del subdiferencial de $|Du|$, es decir, h_{γ} coincide con todos los elementos del subdiferencial fuera de una vecindad de 0.

3.3.1. Existencia de una solución óptima

La existencia de una solución óptima para el problema de optimización binivel (3.3) con $\mu = 0$, es decir sin término elíptico, se demostró en [20]. Así, nos centramos en el caso $\mu > 0$ ya que el término elíptico es necesario para probar la existencia de multiplicadores de Lagrange en la siguiente sección.

La existencia de una solución óptima para el problema optimización de parámetros (3.3) es un caso especial de los problemas binivel considerados en [14], donde la existencia de parámetros reales óptimos en \mathbb{R}^2 fue probado. Se utilizará para la demostración el resultado auxiliar sobre semicontinuidad inferior de la función de Huber, demostrado en el mismo trabajo.

LEMA 3.1. Sean $u, v \in L^p(\Omega)$, $1 \leq p < \infty$. Entonces, el funcional $u \mapsto \int_{\Omega} |u - v|_{\gamma} dx$, donde $|\cdot|_{\gamma}$ es la regularización de Huber, es semicontinuo inferior con respecto a la convergencia débil* en $\mathcal{M}(\Omega; \mathbb{R}^d)$.

DEMOSTRACIÓN 3.1. La demostración de este resultado se presenta en [14, Lema 2.1]

TEOREMA 3.2. Existe una solución óptima $(\hat{\alpha}, \hat{\beta}) \in \mathbb{R}^{N+M}$ del problema de optimización binivel (3.3).

DEMOSTRACIÓN 3.2. Sea $(\alpha_n, \beta_n) \subset \mathbb{R}^{N+M}$ una sucesión minimizante. Gracias a la presencia del término de Tikhonov en el funcional de costo (3.3a), se obtiene que la sucesión (α_n, β_n) esta acotada en \mathbb{R}^{N+M} , es decir, que existe una subsucesión $\{(\alpha_n, \beta_n)\}$ que converge a $(\hat{\alpha}, \hat{\beta})$ en \mathbb{R}^{N+M} . Por otro lado, se verifica que

$$J^{\gamma, \mu}(u_n; P(\alpha_n), Q(\beta_n)) \leq J^{\gamma, \mu}(u; P(\alpha_n), Q(\beta_n)), \quad \text{para todo } u \in H^1(\Omega),$$

en particular esto se verifica para $u = 0$. Por lo tanto,

$$\frac{1}{2} \|u_n - f\|_{L^2(\Omega)}^2 + \text{TGV}_{P(\alpha_n), Q(\beta_n)}^{2, \gamma, \mu} \leq \frac{1}{2} \|f\|_{L^2(\Omega)}^2. \quad (3.7)$$

En particular de la desigualdad (3.7) se sigue que

$$\|u_n\|_{H^1(\Omega)}^2 + \|w_n\|_{\mathbb{H}^1(\Omega)}^2 \leq \frac{1}{\mu} \|f\|_{L^2(\Omega)}, \quad (3.8)$$

donde w_n es el óptimo w para u_n . Esto implica que la sucesión (u_n, w_n) es uniformemente acotada en $H^1(\Omega) \times \mathbb{H}^1(\Omega)$ y por lo tanto existe una subsucesión, notada con el mismo nombre $\{(u_n, w_n)\}$, que converge débilmente en $H^1(\Omega) \times \mathbb{H}^1(\Omega)$ a un punto (\hat{u}, \hat{w}) .

Además, $u_n \rightarrow \hat{u}$ converge fuertemente en $L^p(\Omega)$ y $w_n \rightarrow \hat{w}$ en $\mathbb{L}^p(\Omega) := L^p(\Omega; \mathbb{R}^n)$. Recuerde que P y Q son funciones continuas respecto a los parámetros α, β , y la sucesión de parámetros α_n y β_n son uniformemente acotados. Por consiguiente, usando la continuidad del término de fidelidad, es decir, la norma en L^2 con respecto a la convergencia fuerte en L^2 , y la semicontinuidad inferior débil del término H^1 con respecto a la convergencia débil en H^1 y el resultado sobre la regularización de Huber con respecto a la convergencia débil* en \mathcal{M} (Lema 3.1), se obtiene que

$$\begin{aligned} & \frac{1}{2} \|\hat{u} - f\|_{L^2(\Omega)}^2 + \int_{\Omega} P(\hat{\alpha}) |D\hat{u} - \hat{w}|_{\gamma} dx + \int_{\Omega} Q(\hat{\beta}) |E\hat{w}|_{\gamma} dx + \frac{\mu}{2} \left(\|\hat{u}\|_{H^1(\Omega)}^2 + \|\hat{w}\|_{\mathbb{H}^1(\Omega)}^2 \right) \\ & \leq \liminf_n \frac{1}{2} \|u_n - f\|_{L^2(\Omega)}^2 + \int_{\Omega} P(\hat{\alpha}) |Du_n - w_n|_{\gamma} dx + \int_{\Omega} Q(\hat{\beta}) |Ew_n|_{\gamma} dx \\ & \quad + \frac{\mu}{2} \left(\|u_n\|_{H^1(\Omega)}^2 + \|w_n\|_{\mathbb{H}^1(\Omega)}^2 \right), \end{aligned}$$

$$\begin{aligned}
& \frac{1}{2} \|\hat{u} - f\|_{L^2(\Omega)}^2 + \int_{\Omega} P(\hat{\alpha}) |D\hat{u} - \hat{w}|_{\gamma} dx + \int_{\Omega} Q(\hat{\beta}) |E\hat{w}|_{\gamma} dx + \frac{\mu}{2} \left(\|\hat{u}\|_{H^1(\Omega)}^2 + \|\hat{w}\|_{H^1(\Omega)}^2 \right) \\
& \leq \liminf_n \frac{1}{2} \|u_n - f\|_{L^2(\Omega)}^2 + \int_{\Omega} P(\alpha_n) |Du_n - w_n|_{\gamma} dx + \int_{\Omega} Q(\beta_n) |Ew_n|_{\gamma} dx \\
& \quad + \frac{\mu}{2} \left(\|u_n\|_{H^1(\Omega)}^2 + \|w_n\|_{H^1(\Omega)}^2 \right).
\end{aligned}$$

Para el último paso se uso la cota que se obtuvo en la desigualdad (3.7) y la convergencia de (α_n, β_n) en \mathbb{R}^{N+M} . Esto muestra que el punto límite \hat{u} es una solución óptima para $(\hat{\alpha}, \hat{\beta})$. Además, de la semicontinuidad inferior débil del funcional de costo F y ya que el conjunto $\{(\alpha, \beta) : (P(\alpha), Q(\beta)) \geq 0\}$ es cerrado, se tiene que $(\hat{u}, \hat{\alpha}, \hat{\beta})$ es solución óptima de (3.3).

3.4. Multiplicadores de Lagrange

En esta sección, se prueba la existencia de multiplicadores de Lagrange para el problema de optimización de parámetros (3.5) y se establece el sistema de optimalidad asociado. Además, se obtiene una caracterización del gradiente para el funcional de costo reducido, la cual juega un papel importante en el desarrollo de algoritmos de solución rápida para los problemas de aprendizaje. Para esto se prueba en primer lugar que el operador solución es diferenciable.

3.4.1. Diferenciabilidad del operador solución

Recuerde que el modelo de filtrado de ruido considerando $TGV^{2,\gamma,\mu}$ está determinado por:

$$y = (u, w) = \arg \min_{H^1(\Omega) \times H^1(\Omega)} \left\{ \frac{1}{2} a(y, y) + \frac{1}{2} \int_{\Omega} |u - f|^2 + \int_{\Omega} P(\alpha) |Du - w|_{\gamma} + \int_{\Omega} Q(\beta) |Ew|_{\gamma} \right\},$$

donde $a(y, y) = \mu \left(\|u\|_{H^1(\Omega)}^2 + \|w\|_{H^1(\Omega)}^2 \right)$. Como se vio antes, una condición necesaria y suficiente de optimalidad de este problema está dada por la siguiente ecuación variacional,

$$\begin{aligned}
a(y, \Psi) + \int_{\Omega} P(\alpha) h_{\gamma}(Du - w)(D\phi - \varphi) dx \\
+ \int_{\Omega} Q(\beta) h_{\gamma}(Ew) E\varphi dx + \int_{\Omega} (u - f)\phi dx = 0, \text{ para todo } \Psi \in Y, \quad (3.9)
\end{aligned}$$

donde $Y = H^1(\Omega) \times H^1(\Omega)$ y $\Psi = (\phi, \varphi) \in Y$. En adelante se entenderá \mathbb{L}^2 como $L^2(\Omega, \mathbb{R}^2)$ o $L^2(\Omega, \text{Sym}(\mathbb{R}^2))$ dependiendo del contexto.

TEOREMA 3.3. *El operador solución $S : \mathbb{R}^{N+M} \mapsto Y$, que asigna a cada par $(\alpha, \beta) \in \mathbb{R}^{N+M}$ la solución correspondiente al problema de filtrado de ruido (3.9), es Fréchet diferenciable y su derivada está caracterizada por la solución única $z = S'(\alpha, \beta)[\theta_1, \theta_2] \in Y$ de la siguiente*

ecuación linealizada:

$$\begin{aligned}
a(z, \Psi) &+ \int_{\Omega} P'(\alpha) \theta_1 h_{\gamma}(Du - w)(D\phi - \varphi) dx \\
&+ \int_{\Omega} P(\alpha) h'_{\gamma}(Du - w)(Dz_1 - z_2)(D\phi - \varphi) dx + \int_{\Omega} Q'(\beta) \theta_2 h_{\gamma}(Ew) E\varphi dx \\
&[4pt] + \int_{\Omega} Q(\beta) h'_{\gamma}(Ew) Ez_2 E\varphi dx + \int_{\Omega} z_1 \phi dx = 0, \text{ para todo } \Psi \in Y. \quad (3.10)
\end{aligned}$$

DEMOSTRACIÓN 3.3. Gracias a los términos elípticos de $a(\cdot, \cdot)$ y a la monotonía de h_{γ} , la existencia de una única solución de la ecuación linealizada (3.10) se sigue del teorema de Lax-Milgram.

Sea $\xi := y^+ - y - z$, donde $y = S(\alpha, \beta)$, $y^+ = S(\alpha + \theta_1, \beta + \theta_2)$ y $y^+ = (u^+, w^+)$. El objetivo es probar que $\|\xi\|_Y = o(\|\theta\|)$. Combinando las ecuaciones (3.9) evaluada en y^+ , y y (3.10) en z , se obtiene que

$$\begin{aligned}
a(\xi, \Psi) &+ \int_{\Omega} P(\alpha + \theta_1) h_{\gamma}(Du^+ - w^+)(D\phi - \varphi) dx - \int_{\Omega} P(\alpha) h_{\gamma}(Du - w)(D\phi - \varphi) dx \\
&- \int_{\Omega} P'(\alpha) \theta_1 h_{\gamma}(Du - w)(D\phi - \varphi) dx - \int_{\Omega} P(\alpha) h'_{\gamma}(Du - w)(Dz_1 - z_2)(D\phi - \varphi) dx \\
&[4pt] + \int_{\Omega} Q(\beta + \theta_2) h_{\gamma}(Ew^+) E\varphi dx - \int_{\Omega} Q(\beta) h_{\gamma}(Ew) E\varphi dx \\
[4pt] &- \int_{\Omega} Q'(\beta) \theta_2 h_{\gamma}(Ew) E\varphi dx - \int_{\Omega} Q(\beta) h'_{\gamma}(Ew) Ez_2 E\varphi dx + \int_{\Omega} \xi_1 \phi dx = 0, \text{ para todo } \Psi \in Y,
\end{aligned}$$

con $\xi := (\xi_1, \xi_2) \in Y$. Sumando y restando los términos

$$\int_{\Omega} P(\alpha) h'_{\gamma}(Du - w)(D\delta_u - \delta_w)(D\phi - \varphi) dx$$

y

$$\int_{\Omega} Q(\beta) h'_{\gamma}(Ew) E\delta_w E\varphi dx,$$

donde $\delta_u := u^+ - u$ y $\delta_w := w^+ - w$, se obtiene que

$$\begin{aligned}
a(\xi, \Psi) &+ \int_{\Omega} P(\alpha) h'_{\gamma}(Du - w) [D\delta_u - Dz_1 - (\delta_w - z_2)] (D\phi - \varphi) dx \\
&+ \int_{\Omega} Q(\beta) h'_{\gamma}(Ew) [E\delta_w - Ez_2] E\varphi dx + \int_{\Omega} \xi_1 \phi dx = \\
&- \int_{\Omega} P(\alpha + \theta_1) h_{\gamma}(Du^+ - w^+)(D\phi - \varphi) dx - \int_{\Omega} Q(\beta + \theta_2) h_{\gamma}(Ew^+) E\varphi dx \\
&- \int_{\Omega} P(\alpha) [-h_{\gamma}(Du - w) - h'_{\gamma}(Du - w)(D\delta_u - \delta_w)] (D\phi - \varphi) dx \\
&- \int_{\Omega} P'(\alpha) \theta_1 [-h_{\gamma}(Du - w)] (D\phi - \varphi) dx - \int_{\Omega} Q'(\beta) \theta_2 [-h_{\gamma}(Ew)] E\varphi dx \\
&- \int_{\Omega} Q(\beta) [-h_{\gamma}(Ew) - h'_{\gamma}(Ew) E\delta_w] E\varphi, \text{ para todo } \Psi \in Y.
\end{aligned}$$

A partir de que $P \in C^1(\mathbb{R}^N)$ y $Q \in C^1(\mathbb{R}^M)$ se verifica que

$$P(\alpha + \theta_1) = P(\alpha) + P'(\alpha)\theta_1 + o(\|\theta_1\|) \in C(\overline{\Omega}),$$

$$Q(\beta + \theta_2) = Q(\beta) + Q'(\beta)\theta_2 + o(\|\theta_2\|) \in C(\overline{\Omega}),$$

usando esto, se sigue que

$$\begin{aligned} a(\xi, \Psi) &+ \int_{\Omega} P(\alpha) h'_{\gamma}(Du - w)(D\xi_1 - \xi_2)(D\phi - \varphi) dx \\ &+ \int_{\Omega} Q(\beta) h'_{\gamma}(Ew) E\xi_2 : E\varphi dx + \int_{\Omega} \xi_1 \phi dx = \\ &- \int_{\Omega} P(\alpha) [h_{\gamma}(Du^+ - w^+) - h_{\gamma}(Du - w) - h'_{\gamma}(Du - w)(D\delta_u - \delta_w)] (D\phi - \varphi) dx \\ &- \int_{\Omega} o(\|\theta_1\|) h_{\gamma}(Du^+ - w^+)(D\phi - \varphi) \\ &- \int_{\Omega} P'(\alpha)\theta_1 [h_{\gamma}(Du^+ - w^+) - h_{\gamma}(Du - w)] (D\phi - \varphi) dx \\ &- \int_{\Omega} Q(\beta) [h_{\gamma}(Ew^+) - h_{\gamma}(Ew) - h'_{\gamma}(Ew)E\delta_w] E\varphi dx - \int_{\Omega} o(\|\theta_2\|) h_{\gamma}(Ew^+) E\varphi \\ &- \int_{\Omega} Q'(\beta)\theta_2 [h_{\gamma}(Ew^+) - h_{\gamma}(Ew)] E\varphi dx, \text{ para todo } \Psi \in Y. \end{aligned}$$

Tomando $\Psi = \xi$ y usando la monotonía de $h'_{\gamma}(\cdot)$, se verifica que

$$\begin{aligned} \mu \|\xi\|_Y^2 &\leq \|P(\alpha)\|_{L^\infty(\Omega)} \|h_{\gamma}(Du^+ - w^+) - h_{\gamma}(Du - w) \\ &- h'_{\gamma}(Du - w)(D\delta_u - \delta_w)\|_{\mathbb{L}^2} \|D\xi_1 - \xi_2\|_{\mathbb{L}^2} + o(\|\theta_1\|) \|h_{\gamma}(Du^+ - w^+)\|_{\mathbb{L}^2} \|D\xi_1 - \xi_2\|_{\mathbb{L}^2} \\ &+ \|P'(\alpha)\theta_1\|_{L^\infty(\Omega)} \|h_{\gamma}(Du^+ - w^+) - h_{\gamma}(Du - w)\|_{\mathbb{L}^2} \|D\xi_1 - \xi_2\|_{\mathbb{L}^2} \\ &+ \|Q(\beta)\|_{L^\infty(\Omega)} \|h_{\gamma}(Ew^+) - h_{\gamma}(Ew) - h'_{\gamma}(Ew)E\delta_w\|_{\mathbb{L}^2} \|E\xi_2\|_{\mathbb{L}^2} \\ &+ o(\|\theta_2\|) \|h_{\gamma}(Ew^+)\|_{\mathbb{L}^2} \|E\xi_2\|_{\mathbb{L}^2} \\ &+ \|Q'(\beta)\theta_2\|_{L^\infty(\Omega)} \|h_{\gamma}(Ew^+) - h_{\gamma}(Ew)\|_{\mathbb{L}^2} \|E\xi_2\|_{\mathbb{L}^2}. \end{aligned}$$

Considerando la diferenciabilidad de h_{γ} sobre $L^2(\Omega, \mathbb{R}^2)$ y $L^2(\Omega, \text{Sym}(\mathbb{R}^2))$, la Lipschitz continuidad de $h'_{\gamma}(\cdot)$ de \mathbb{L}^2 a \mathbb{L}^2 [34], y $|h_{\gamma}(\cdot)| \leq 1$, se tiene entonces que existe $p > 2$ tal que

$$\begin{aligned} \|\xi\|_Y &\leq \|P(\alpha)\|_{L^\infty(\Omega)} o(\|y^+ - y\|_{1,p}) + \|P'(\alpha)\|_{L^\infty(\Omega)} \|\theta_1\|_{\mathbb{R}^N} \|y^+ - y\|_Y + o(\|\theta_1\|) \\ &+ \|Q(\beta)\|_{L^\infty(\Omega)} o(\|w^+ - w\|_{1,p}) + \|Q'(\beta)\|_{L^\infty(\Omega)} \|\theta_2\|_{\mathbb{R}^M} \|w^+ - w\|_{\mathbb{H}^1(\Omega)} + o(\|\theta_2\|), \quad (3.11) \end{aligned}$$

donde $\|\cdot\|_{1,p}$ es la norma en el espacio $\mathbb{W}^{1,p}(\Omega)$. Gracias a que $P(\alpha), Q(\beta) \in C(\overline{\Omega})$, se

obtiene que

$$\begin{aligned} \|\xi\|_Y \leq C_1(\alpha) o(\|y^+ - y\|_{1,p}) + C_2(\alpha) \|\theta_1\|_{\mathbb{R}^N} \|y^+ - y\|_Y + o(\|\theta\|) \\ + C_3(\beta) o(\|w^+ - w\|_{1,p}) + C_4(\beta) \|\theta_2\|_{\mathbb{R}^M} \|w^+ - w\|_{\mathbb{H}^1(\Omega)}. \end{aligned} \quad (3.12)$$

Por otro lado, el resultado de regularidad para sistemas de segundo orden elípticos semilineales ([25, Teorema 1, Rem. 14]), se sigue que

$$\|y^+ - y\|_{1,p} \leq \tilde{L}\|\theta\|$$

como se demostró en el trabajo de [14]. Reemplazando la última estimación en (3.12), finalmente se tiene que

$$\|\xi\|_Y = o(\|\theta\|).$$

OBSERVACIÓN. El resultado de regularidad extra para sistemas de segundo orden elípticos semilineales usado en la demostración, fue probado por Gröger en ([Teorema 1, Rem. 14],[25]) y se basa en las propiedades del dominio Ω . El resultado se demostró originalmente para los dominios C^2 . Sin embargo, la regularidad del dominio (en el sentido de Gröger) también puede ser verificada para dominios Lipschitz convexos acotados [13], que es precisamente nuestro caso considerando el dominio Ω de la imagen.

3.4.2. Ecuación adjunta

El objetivo de esta sección es caracterizar la ecuación adjunta del problema de optimización de parámetros, usando el método del Lagrangiano. La existencia de una solución única de la ecuación adjunta se sigue del método de transposición, ya que la ecuación linealizada está bien planteada.

Se define el Lagrangiano asociado al problema (3.5) como:

$$\begin{aligned} \mathcal{L}(u, w, \alpha, \beta, p_1, p_2) = F(u) + \mu(u, p_1)_{\mathbb{H}^1(\Omega)} + \mu(w, p_2)_{\mathbb{H}^1(\Omega)} \\ + \int_{\Omega} P(\alpha) h_{\gamma}(Du - w)(Dp_1 - p_2) + \int_{\Omega} Q(\beta) h_{\gamma}(Ew) Ep_2 + \int_{\Omega} (u - f)p_1, \end{aligned}$$

tomando la derivada con respecto a la variable de estado (u, w) en la dirección (δ_u, δ_w) , se obtiene la condición necesaria de optimalidad

$$\begin{aligned} \mathcal{L}'_{(u,w)}(u, w, \alpha, \beta, p_1, p_2)[(\delta_u, \delta_w)] = \nabla_u F(u) \delta_u + \mu(p_1, \delta_u)_{\mathbb{H}^1(\Omega)} + \mu(p_2, \delta_w)_{\mathbb{H}^1(\Omega)} \\ + \int_{\Omega} P(\alpha) h'_{\gamma}(Du - w)(D\delta_u - \delta_w)(Dp_1 - p_2) \\ + \int_{\Omega} Q(\beta) h'_{\gamma}(Ew) E\delta_w Ep_2 + \int_{\Omega} p_1 \delta_u = 0. \end{aligned}$$

Si $\delta_w = 0$, entonces

$$\mu(p_1, \delta_u)_{H^1(\Omega)} + \int_{\Omega} P(\alpha) h'_{\gamma}(Du - w)(Dp_1 - p_2) D\delta_u + \int_{\Omega} p_1 \delta_u = -\nabla_u F(u) \delta_u,$$

si consideramos ahora $\delta_u = 0$, entonces

$$\mu(p_2, \delta_w)_{H^1(\Omega)} - \int_{\Omega} P(\alpha) h'_{\gamma}(Du - w)(Dp_1 - p_2) \delta_w + \int_{\Omega} Q(\beta) h'_{\gamma}(Ew) E p_2 E \delta_w = 0.$$

3.4.3. Condición de optimalidad

Utilizando la diferenciabilidad del operador solución y que la ecuación adjunta esta bien planteada, a continuación se establece el sistema de optimalidad asociado al problema de aprendizaje de dos niveles (3.5), con el objetivo de caracterizar los mínimos locales de dicho problema. Además, se presenta la caracterización del gradiente del funcional de costo reducido, la cual juega un papel importante en el desarrollo de algoritmos de solución rápida para los problemas de aprendizaje.

TEOREMA 3.4. *Sea $(\bar{\alpha}, \bar{\beta}) \in \mathbb{R}^{N+M}$ una solución óptima local del problema (3.5). Si existen $\alpha_0 \in \mathbb{R}^N, \beta_0 \in \mathbb{R}^M$ tales que*

$$P(\bar{\alpha}) + \nabla_{\alpha} P(\bar{\alpha})^T (\alpha_0 - \bar{\alpha}) > 0 \quad y \quad Q(\bar{\beta}) + \nabla_{\beta} Q(\bar{\beta})^T (\beta_0 - \bar{\beta}) > 0, \quad (3.13)$$

entonces existen multiplicadores de Lagrange $\Pi \in Y = \mathbb{H}^1 \times H^1$ y $\lambda_1, \lambda_2 \in \mathcal{M}(\Omega)$, tales que el siguiente sistema se verifica:

$$\begin{aligned} a(y, \Psi) + \int_{\Omega} P(\alpha) h_{\gamma}(Du - w)(D\phi - \varphi) dx \\ + \int_{\Omega} Q(\beta) h_{\gamma}(Ew) E\varphi dx + \int_{\Omega} (u - f)\phi dx = 0, \text{ para todo } \Psi \in Y, \end{aligned} \quad (3.14a)$$

$$\begin{aligned} a(\Pi, \Psi) + \int_{\Omega} P(\alpha) h'_{\gamma}(Du - w)(Dp_1 - p_2)(D\phi - \varphi) dx \\ + \int_{\Omega} Q(\beta) h'_{\gamma}(Ew) E p_2 E \varphi dx + \int_{\Omega} p_1 \phi dx = \nabla_u F(u) \phi, \text{ para todo } \Psi \in Y, \end{aligned} \quad (3.14b)$$

$$P'(\bar{\alpha})^* \lambda_1 - \int_{\Omega} P'(\alpha) h_{\gamma}(Du - w)(Dp_1 - p_2) dx - 2\xi \alpha = 0, \quad (3.14c)$$

$$Q'(\bar{\beta})^* \lambda_2 - \int_{\Omega} Q'(\beta) h_{\gamma}(Ew) E p_2 dx - 2\xi \beta = 0, \quad (3.14d)$$

$$\int_{\Omega} (z_1 - P(\bar{\alpha})) d\lambda_1 \geq 0, \text{ para todo } z_1 \in C(\Omega) : z_1 \geq 0, \quad (3.14e)$$

$$\int_{\Omega} (z_2 - Q(\bar{\beta})) d\lambda_2 \geq 0, \text{ para todo } z_2 \in C(\Omega) : z_2 \geq 0. \quad (3.14f)$$

DEMOSTRACIÓN 3.4. Considere el funcional de costo reducido $\mathcal{F}(\alpha, \beta) = F(u(\alpha, \beta), \alpha, \beta)$. El problema de optimización binivel se puede formular como sigue

$$\min_{\alpha, \beta \in \mathbb{R}^{N+M}} \mathcal{F}(\alpha, \beta)$$

sujeto a:

$$P(\alpha) \geq 0, \text{ en todo punto de } \Omega,$$

$$Q(\beta) \geq 0, \text{ en todo punto de } \Omega,$$

donde $\mathcal{F} : \mathbb{R}^{N+M} \rightarrow \mathbb{R}$, $P : \mathbb{R}^N \rightarrow C(\Omega)$ y $Q : \mathbb{R}^M \rightarrow C(\Omega)$. Se define el cono

$$K := \{t \in C(\Omega) \times C(\Omega) : t(x) \geq 0\}.$$

Las restricciones se pueden escribir como $G(\alpha, \beta) := (P(\alpha), Q(\beta)) \in K$. La existencia de multiplicadores de Lagrange se sigue de [44], si la condición de Slater se verifica, es decir, si

$$\exists(\alpha_0, \beta_0) \text{ tal que } G(\alpha, \beta) + G'(\alpha, \beta)(\alpha_0 - \alpha, \beta_0 - \beta) \in \text{int } K. \quad (3.15)$$

Ya que K tiene interior no vacío y por hipótesis se cumple (3.13), entonces la condición de Slater se verifica. Se sigue así que existen multiplicadores

$$(\lambda_1, \lambda_2) \in K^+ := \{\vartheta \in \mathcal{M}(\Omega) \times \mathcal{M}(\Omega) : \langle \vartheta, z \rangle_{\mathcal{M}(\Omega) \times \mathcal{M}(\Omega), C(\Omega) \times C(\Omega)} \geq 0, \forall z \in K\},$$

tal que

$$\mathcal{F}'(\bar{\alpha}, \bar{\beta}) - G'(\bar{\alpha}, \bar{\beta})^*(\lambda_1, \lambda_2) = 0.$$

Tomando la derivada con respecto a (α, β) y denotando por z la solución de la ecuación linealizada (3.10) se tiene, junto con la ecuación adjunta (3.14b), que

$$\begin{aligned} \mathcal{F}'(\alpha, \beta)[\theta] &= \nabla_u F(u)z + 2\xi \langle (\alpha, \beta), \theta \rangle, \\ &= 2\xi \langle (\alpha, \beta), \theta \rangle - a(\Pi, z) - \int_{\Omega} P(\alpha) h'_{\gamma}(Du - w)(Dp_1 - p_2)(Dz_1 - z_2) \\ &\quad - \int_{\Omega} Q(\beta) h'_{\gamma}(Ew) E p_2 E z_2 - \int_{\Omega} p_1 z_1, \\ &= 2\xi \langle (\alpha, \beta), \theta \rangle - a(z, \Pi) - \int_{\Omega} P(\alpha) h'_{\gamma}(Du - w)(Dz_1 - z_2)(Dp_1 - p_2) \\ &\quad - \int_{\Omega} Q(\beta) h'_{\gamma}(Ew) E z_2 E p_2 - \int_{\Omega} z_1 p_1, \end{aligned}$$

la cual, teniendo en cuenta la ecuación linealizada, se convierte en

$$\begin{aligned} \mathcal{F}'(\alpha, \beta)[\theta] &= 2\xi \langle (\alpha, \beta), \theta \rangle + \int_{\Omega} P'(\alpha) \theta_1 h_{\gamma}(Du - w)(Dp_1 - p_2) \\ &\quad + \int_{\Omega} Q'(\beta) \theta_2 h_{\gamma}(Ew) E p_2. \quad (3.16) \end{aligned}$$

Por otro lado, de la estructura de G se obtiene que

$$\begin{aligned} \langle G'(\bar{\alpha}, \bar{\beta})^*(\lambda_1, \lambda_2), (\theta_1, \theta_2) \rangle &= \langle (\lambda_1, \lambda_2), G'(\bar{\alpha}, \bar{\beta})(\theta_1, \theta_2) \rangle_{\mathcal{M}(\Omega) \times \mathcal{M}(\Omega), C(\Omega) \times C(\Omega)}, \\ &= \langle \lambda_1, P'(\alpha)\theta_1 \rangle_{\mathcal{M}(\Omega), C(\Omega)} + \langle \lambda_2, Q'(\beta)\theta_2 \rangle_{\mathcal{M}(\Omega), C(\Omega)}, \end{aligned}$$

y, por lo tanto se obtiene

$$P'(\bar{\alpha})^*\lambda_1 - \int_{\Omega} P'(\alpha)h_{\gamma}(Du - w)(Dp_1 - p_2) dx - 2\zeta\alpha = 0,$$

y

$$Q'(\bar{\beta})^*\lambda_2 - \int_{\Omega} Q'(\beta)h_{\gamma}(Ew) Ep_2 dx - 2\zeta\beta = 0.$$

OBSERVACIÓN. Si P y Q son lineales con respecto a los parámetros, la condición de Slater (3.15) se reduce a la existencia de (α_0, β_0) , tal que $P(\alpha_0) > 0$ y $Q(\beta_0) > 0$.

OBSERVACIÓN. Se pueden considerar $P(\alpha)(x)$ y $Q(\beta)(x)$ como funciones continuas a trozos. En este caso los multiplicadores λ_1, λ_2 pertenecen al dual del espacio $L^{\infty}(\Omega)$, cuyos elementos no son necesariamente medidas.

A continuación se considera el caso de parámetros escalares, es decir, $P(\alpha) = \alpha \in \mathbb{R}$ y $Q(\beta) = \beta \in \mathbb{R}$, no se requiere verificar una condición de calificación y se obtiene más regularidad de los multiplicadores de Lagrange.

COROLARIO 3.5. Sea $(\bar{\alpha}, \bar{\beta})$ una solución óptima local del problema (3.5), $P(\alpha) = \alpha \in \mathbb{R}$ y $Q(\beta) = \beta \in \mathbb{R}$. Entonces existen multiplicadores de Lagrange $\Pi \in Y$ y $\lambda_1, \lambda_2 \in L^2(\Omega)$, tales que el siguiente sistema se verifica:

$$\begin{aligned} a(y, \Psi) + \alpha \int_{\Omega} h_{\gamma}(Du - w)(D\phi - \varphi) dx \\ + \beta \int_{\Omega} h_{\gamma}(Ew)E\varphi dx + 2 \int_{\Omega} (u - f)\phi dx = 0, \text{ para todo } \Psi \in Y, \end{aligned} \quad (3.17a)$$

$$\begin{aligned} a(\Pi, \Psi) + \alpha \int_{\Omega} h'_{\gamma}(Du - w)(Dp_1 - p_2)(D\phi - \varphi) dx \\ + \beta \int_{\Omega} h'_{\gamma}(Ew) Ep_2 E\varphi dx + 2 \int_{\Omega} p_1\phi dx = 0, \text{ para todo } \Psi \in Y, \end{aligned} \quad (3.17b)$$

$$\lambda_1 = \int_{\Omega} h_{\gamma}(Du - w)(Dp_1 - p_2) - 2\zeta\alpha, \quad (3.17c)$$

$$\lambda_2 = \int_{\Omega} h_{\gamma}(Ew) Ep_2 - 2\zeta\beta, \quad (3.17d)$$

$$\lambda_1 \geq 0, \quad \lambda_2 \geq 0, \quad (3.17e)$$

$$\lambda_1 \cdot \bar{\alpha} = \lambda_2 \cdot \bar{\beta} = 0. \quad (3.17f)$$

DEMOSTRACIÓN 3.5. *El procedimiento es similar que en la demostración del Teorema 3.4, en este caso se aplica el resultado [44, Teorema 3.1] al problema*

$$\min_{(\alpha, \beta) \in U} \mathcal{F}(\alpha, \beta),$$

donde $\mathcal{F} : \mathbb{R}^2 \rightarrow \mathbb{R}$ y U corresponde al ortante positivo en \mathbb{R}^2 . En este caso, nos beneficiamos del hecho de que la restricción de positividad puede ser directamente impuesta a las variables independientes y obtener así la existencia de multiplicadores λ_1, λ_2 tales que

$$\begin{aligned} \lambda_1 &= \nabla_{\alpha} \mathcal{F}(\bar{\alpha}, \bar{\beta}), \\ \lambda_2 &= \nabla_{\beta} \mathcal{F}(\bar{\alpha}, \bar{\beta}), \\ \lambda_1 &\geq 0, \quad \lambda_2 \geq 0, \\ \lambda_1 \cdot \bar{\alpha} &= \lambda_2 \cdot \bar{\beta} = 0, \end{aligned}$$

la cual, junto con la expresión (3.16), conduce al resultado.

3.5. Problema regularizado

En la sección anterior se evidenció que los multiplicadores de Lagrange asociados a la restricción de desigualdad (3.5c) son medidas, cuya implementación numérica es muy complicada. Como alternativa a esto se introduce la *Regularización de Moreau-Yosida* [32], que consiste en penalizar el funcional de costo con una función C^1 cuando la restricción de positividad no se verifica [15],

$$\frac{1}{2\eta} \|\min(0, \eta P(\alpha))\|_{L^2(\Omega)}^2 + \frac{1}{2\eta} \|\min(0, \eta Q(\beta))\|_{L^2(\Omega)}^2.$$

Esta técnica ha sido ampliamente utilizada en trabajos como [17], [18], [16] donde se consideran problemas de control óptimo con restricciones de desigualdad. Se define la familia de problemas η -penalizados como:

$$\min_{(\alpha, \beta) \in \mathbb{R}^{N+M}} F_{\eta}(u_{\alpha, \beta}) \quad (3.18a)$$

sujeto a

$$u_{\alpha, \beta} \in \arg \min_{u \in H^1(\Omega)} J^{\gamma, \mu}(u; P(\alpha), Q(\beta)), \quad (3.18b)$$

con $\eta > 0$ el parámetro de regularización y el funcional $F_\eta(u_{\alpha,\beta})$ está determinado por:

$$F_\eta(u_{\alpha,\beta}) := F(u_{\alpha,\beta}) + \frac{1}{2\eta} \|\min(0, \eta P(\alpha))\|_{L^2(\Omega)}^2 + \frac{1}{2\eta} \|\min(0, \eta Q(\beta))\|_{L^2(\Omega)}^2,$$

donde,

$$F(u_{\alpha,\beta}) = \frac{1}{2} \|f_0 - u\|_{L^2(\Omega)}^2 + \xi (\|\alpha\|_{\mathbb{R}^N}^2 + \|\beta\|_{\mathbb{R}^M}^2).$$

Reemplazando (3.18b) por la condición necesaria y suficiente de optimalidad, se obtiene el siguiente problema sin restricciones de desigualdad

$$\min_{(\alpha,\beta) \in \mathbb{R}^{N+M}} F_\eta(u_{\alpha,\beta}) \quad (3.19a)$$

sujeto a

$$\begin{aligned} a(y, \Psi) + \int_{\Omega} P(\alpha) h_\gamma(Du - w)(D\phi - \phi) dx + \int_{\Omega} Q(\beta) h_\gamma(Ew) E\phi dx \\ + \int_{\Omega} (u - f)\phi dx = 0, \text{ para todo } \Psi \in Y. \end{aligned} \quad (3.19b)$$

En el siguiente teorema, se estudia la convergencia de las soluciones del problema η -regularizado a las soluciones del problema original.

TEOREMA 3.6. *La sucesión $\{(u_\eta, w_\eta, \alpha_\eta, \beta_\eta)\}_{\eta>0}$ de soluciones del problema η -regularizado (3.19) contiene una subsucesión que converge fuertemente en $H^1(\Omega) \times \mathbb{H}^1(\Omega) \times \mathbb{R}^{N+M}$ a una solución óptima $(\bar{u}, \bar{w}, \bar{\alpha}, \bar{\beta})$ del problema original (3.5).*

DEMOSTRACIÓN 3.6. *Sea $(\bar{u}, \bar{w}, \bar{\alpha}, \bar{\beta}) \in H^1(\Omega) \times \mathbb{H}^1(\Omega) \times \mathbb{R}^{N+M}$ una solución global del problema (3.5). De las propiedades del funcional de costo regularizado (3.18a) se verifica que,*

$$F_\eta(u_{\alpha_\eta, \beta_\eta}) = F(u_{\alpha_\eta, \beta_\eta}) + \underbrace{\frac{1}{2\eta} \|\min(0, \eta P(\alpha_\eta))\|_{L^2(\Omega)}^2 + \frac{1}{2\eta} \|\min(0, \eta Q(\beta_\eta))\|_{L^2(\Omega)}^2}_{\geq 0},$$

además,

$$F_\eta(u_{\alpha_\eta, \beta_\eta}) \leq F_\eta(\bar{u}) = F(\bar{u}). \quad (3.20)$$

Consecuentemente, puesto que $\xi > 0$, se sigue que la sucesión $\{(\alpha_\eta, \beta_\eta)\}_{\eta>0}$ es uniformemente acotada en \mathbb{R}^{N+M} , de donde existe una subsucesión, notada con el mismo nombre $\{(\alpha_\eta, \beta_\eta)\}_{\eta>0}$, tales que $\alpha_\eta \rightarrow \hat{\alpha}$ en \mathbb{R}^N y $\beta_\eta \rightarrow \hat{\beta}$ en \mathbb{R}^M .

Por otro lado, $\{(u_\eta, w_\eta)\}_{\eta>0}$ es uniformemente acotado en $H^1(\Omega) \times \mathbb{H}^1(\Omega)$ por la desigualdad (3.8) antes probada, es decir,

$$\|u_n\|_{H^1(\Omega)}^2 + \|w_n\|_{\mathbb{H}^1(\Omega)}^2 \leq \frac{1}{\mu} \|f\|_{L^2(\Omega)}.$$

Así, existe una subsucesión notada con el mismo nombre $\{u_\eta\}_{\eta>0}$ y $\{w_\eta\}_{\eta>0}$, tales que

$$u_\eta \rightharpoonup \hat{u} \text{ débilmente en } H^1(\Omega),$$

$$w_\eta \rightharpoonup \hat{w} \text{ débilmente en } \mathbb{H}^1(\Omega).$$

Ahora, se debe probar que el punto limite $(\hat{\alpha}, \hat{\beta})$ es factible, es decir, satisface la restricción de positividad del problema original (3.5c). De (3.20) los siguientes términos

$$\frac{1}{2\eta} \|\text{mín}(0, \eta P(\alpha_\eta))\|_{L^2(\Omega)}^2 \text{ y } \frac{1}{2\eta} \|\text{mín}(0, \eta Q(\beta_\eta))\|_{L^2(\Omega)}^2$$

son uniformemente acotados con respecto a η . Así,

$$\lim_{\eta \rightarrow \infty} \|\text{mín}(0, P(\alpha_\eta))\|_{L^2(\Omega)}^2 = 0 \text{ y } \lim_{\eta \rightarrow \infty} \|\text{mín}(0, Q(\beta_\eta))\|_{L^2(\Omega)}^2 = 0.$$

En efecto en el primer caso, aplicando el lema de Fatou se obtiene que

$$\int_{\Omega} \liminf_{\eta \rightarrow \infty} \text{mín}(0, P(\alpha_\eta))^2 \leq \liminf_{\eta \rightarrow \infty} \int_{\Omega} \text{mín}(0, P(\alpha_\eta))^2 = 0,$$

entonces

$$\int_{\Omega} \text{mín}(0, P(\hat{\alpha}))^2 \leq 0$$

lo que implica que $\text{mín}(0, P(\hat{\alpha})) = 0$. Entonces, se verifica que $P(\hat{\alpha}), Q(\hat{\beta}) \geq 0$ en todo punto de Ω . Considerando además que el funcional de costo es débilmente semicontinuo inferior, es decir,

$$F(\hat{u}) \leq \liminf_{\eta \rightarrow \infty} F(u_{\alpha_\eta, \beta_\eta}),$$

donde \hat{u} denota la solución de (3.19b) asociada a $(\hat{\alpha}, \hat{\beta})$. Además, se verifica que

$$\liminf_{\eta \rightarrow \infty} F(u_{\alpha_\eta, \beta_\eta}) \leq \liminf_{\eta \rightarrow \infty} F_\eta(u_{\alpha_\eta, \beta_\eta}) \leq F(\bar{u}),$$

de donde, se sigue que $(\hat{u}, \hat{w}, \hat{\alpha}, \hat{\beta})$ es solución de (3.5). Se denota al óptimo $(\hat{u}, \hat{w}, \hat{\alpha}, \hat{\beta})$ por $(\bar{u}, \bar{w}, \bar{\alpha}, \bar{\beta})$.

Ahora, para verificar la convergencia fuerte se emplea la diferencia de la ecuación variacional (3.19b) en $y_\eta := (u_\eta, w_\eta)$ y $\bar{y} := (\bar{u}, \bar{w})$, donde $u_\eta := u_{\alpha_\eta, \beta_\eta}$ y se sigue que,

$$\begin{aligned} a(y_\eta - \bar{y}, \Psi) + \int_{\Omega} [P(\alpha_\eta)h_\gamma(Du_\eta - w_\eta) - P(\bar{\alpha})h_\gamma(D\bar{u} - \bar{w})] (D\phi - \varphi) dx \\ + \int_{\Omega} [Q(\beta_\eta)h_\gamma(Ew_\eta) - Q(\bar{\beta})h_\gamma(E\bar{w})] E\varphi dx + \int_{\Omega} (u_\eta - \bar{u})\phi dx = 0, \text{ para todo } \Psi \in Y. \end{aligned}$$

Tomando $\Psi = y_\eta - \bar{y}$, $(\phi, \varphi) = (\delta_u, \delta_w)$ donde $\delta_u := u_\eta - \bar{u}$ y $\delta_w := w_\eta - \bar{w}$, sumando y restando los términos:

$$\int_{\Omega} P(\alpha_\eta)h_\gamma(D\bar{u} - \bar{w}) \text{ y } \int_{\Omega} Q(\beta_\eta)h_\gamma(E\bar{w}),$$

se obtiene

$$\begin{aligned}
& \mu \|y_\eta - \bar{y}\|_Y^2 + \int_\Omega P(\alpha_\eta) [h_\gamma(Du_\eta - w_\eta) - h_\gamma(D\bar{u} - \bar{w})] (D\delta_u - \delta_w) dx \\
& \quad + \int_\Omega Q(\beta_\eta) [h_\gamma(Ew_\eta) - h_\gamma(E\bar{w})] (E\delta_w) dx + \int_\Omega (u_\eta - \bar{u})^2 dx, \\
& = - \int_\Omega [P(\alpha_\eta) - P(\bar{\alpha})] h_\gamma(D\bar{u} - \bar{w})(D\delta_u - \delta_w) dx \\
& \quad - \int_\Omega [Q(\beta_\eta) - Q(\bar{\beta})] h_\gamma(E\bar{w})(E\delta_w) dx,
\end{aligned}$$

por la monotonía de $h_\gamma(\cdot)$ se sigue que

$$\begin{aligned}
\mu \|y_\eta - \bar{y}\|_Y^2 & \leq \|P(\alpha_\eta) - P(\bar{\alpha})\|_{L^\infty(\Omega)} \|h_\gamma(D\bar{u} - \bar{w})\|_{\mathbb{L}^2} \|D\delta_u - \delta_w\|_{\mathbb{L}^2} \\
& \quad + \|Q(\beta_\eta) - Q(\bar{\beta})\|_{L^\infty(\Omega)} \|h_\gamma(E\bar{w})\|_{\mathbb{L}^2} \|E\delta_w\|_{\mathbb{L}^2},
\end{aligned}$$

considerando que $\|y\|_Y^2 = \|u\|_{H^1(\Omega)}^2 + \|Ew\|_{H^1(\Omega)}^2$ se llega a la siguiente estimación

$$\begin{aligned}
\mu \|y_\eta - \bar{y}\|_Y^2 & \leq (\|P(\alpha_\eta) - P(\bar{\alpha})\|_{L^\infty(\Omega)} \|h_\gamma(D\bar{u} - \bar{w})\|_{\mathbb{L}^2} \\
& \quad + \|Q(\beta_\eta) - Q(\bar{\beta})\|_{L^\infty(\Omega)} \|h_\gamma(E\bar{w})\|_{\mathbb{L}^2}) \|y_\eta - \bar{y}\|_Y.
\end{aligned}$$

Dado que $|h_\gamma(\cdot)| \leq 1$, se verifica que

$$\|y_\eta - \bar{y}\|_Y \leq C_1 \left(\|P(\alpha_\eta) - P(\bar{\alpha})\|_{L^\infty(\Omega)} + \|Q(\beta_\eta) - Q(\bar{\beta})\|_{L^\infty(\Omega)} \right), \quad (3.21)$$

y se concluye que $u_\eta \rightarrow \bar{u}$ y $w_\eta \rightarrow \bar{w}$, a partir de $\alpha_\eta \rightarrow \bar{\alpha}$ y $\beta_\eta \rightarrow \bar{\beta}$ cuando $\eta \rightarrow \infty$.

3.5.1. Sistema de optimalidad

En esta sección, se busca caracterizar el sistema de optimalidad del problema η -regularizado (3.19) usando el método del Lagrangiano. Esto es posible gracias al resultado de existencia de la ecuación linealizada. El Lagrangiano asociado al problema regularizado (3.19) se define como:

$$\begin{aligned}
\mathcal{L}_\eta(u, w, \alpha, \beta, p_1, p_2) & := F_\eta(u) + \mu(u, p_1)_{H^1(\Omega)} + \mu(w, p_2)_{H^1(\Omega)} \\
& \quad + \int_\Omega P(\alpha) h_\gamma(Du - w)(Dp_1 - p_2) + \int_\Omega Q(\beta) h_\gamma(Ew) Ep_2 + \int_\Omega (u - f)p_1.
\end{aligned}$$

Tomando la derivada con respecto a la variable de estado (u, w) en la dirección (δ_u, δ_w) , se obtiene la condición necesaria de optimalidad

$$\begin{aligned}
\nabla_{(u,w)} \mathcal{L}_\eta(u, w, \alpha, \beta, p_1, p_2)[(\delta_u, \delta_w)] & = \nabla_u F_\eta(u) \delta_u + \mu(p_1, \delta_u)_{H^1(\Omega)} + \mu(p_2, \delta_w)_{H^1(\Omega)} \\
& \quad + \int_\Omega P(\alpha) h'_\gamma(Du - w)(D\delta_u - \delta_w)(Dp_1 - p_2) \\
& \quad + \int_\Omega Q(\beta) h'_\gamma(Ew) E\delta_w Ep_2 + \int_\Omega p_1 \delta_u = 0, \quad (3.22)
\end{aligned}$$

que se conoce como ecuación adjunta. Por otro lado, tomando la derivada parcial con respecto a la variable de control (α, β) en la dirección $(\delta_\alpha, \delta_\beta)$ se obtiene:

$$\begin{aligned} \nabla_{(\alpha, \beta)} \mathcal{L}_\eta(u, w, \alpha, \beta, p_1, p_2)[(\delta_\alpha, \delta_\beta)] &= \nabla_\alpha F_\eta(u) \delta_\alpha + \nabla_\beta F_\eta(u) \delta_\beta + \\ &\int_\Omega P'(\alpha) h_\gamma(Du - w)(Dp_1 - p_2) \delta_\alpha + \int_\Omega Q'(\beta) h_\gamma(Ew) E p_2 E \delta_\beta = 0, \end{aligned}$$

que se conoce como ecuación del gradiente.

OBSERVACIÓN. Como resultado inmediato se obtiene la caracterización del gradiente del funcional de costo reducido:

$$\begin{aligned} \nabla \mathcal{F}_\eta(\alpha, \beta) &:= \nabla_\alpha F(u_{\alpha, \beta}) + \nabla_\beta F(u_{\alpha, \beta}) + \int_\Omega \min(0, \eta P(\alpha)) P'(\alpha) dx \\ &+ \int_\Omega \min(0, \eta Q(\beta)) Q'(\beta) dx + \int_\Omega P'(\alpha) h_\gamma(Du - w)(Dp_1 - p_2) dx \\ &+ \int_\Omega Q'(\beta) h_\gamma(Ew) E p_2 dx. \quad (3.23) \end{aligned}$$

Este resultado desempeña un rol importante en el diseño de algoritmos de optimización para determinar de forma eficiente la solución numérica del problema (3.19).

3.6. Tratamiento numérico

La solución numérica del problema de optimización de parámetros se puede abordar desde dos puntos de vista, el primero consiste en utilizar un método de Newton para resolver el sistema de optimalidad, y el segundo, usar un algoritmo iterativo para la solución del problema de optimización, con la caracterización del gradiente del funcional de costo reducido. En este trabajo nos enfocaremos en los algoritmos iterativos para la resolución de problemas de optimización binivel no lineal, específicamente el método quasi-Newton BFGS.

Si $\nabla \mathcal{F}_\eta$ es localmente Lipschitz continua, el algoritmo BFGS está bien definido y genera una sucesión $\{(\alpha_k, \beta_k)\}_{k \in \mathbb{N}}$ que converge a un punto estacionario $(\hat{\alpha}, \hat{\beta})$. Esto gracias a que, la Lipschitz continuidad local implica la continuidad uniforme en los conjunto de nivel. Para probar que el gradiente del funcional de costo reducido es localmente Lipschitz continuo se emplean estimaciones de la variable de estado y del estado adjunto que se muestran en los siguientes resultados.

LEMA 3.7. Si $f \in L^\infty(\Omega)$ y $u_{\alpha, \beta} \in C^{1, \rho}(\Omega)$, para algún $\rho \in (0, 1)$ entonces para $(\alpha^*, \beta^*) \in \mathbb{R}^{N+M}$ arbitrario pero fijo, y $(\alpha, \beta), (\hat{\alpha}, \hat{\beta}) \in V_\delta((\alpha^*, \beta^*))$ se verifica:

- i) $\|u\|_{H^1(\Omega)} \leq K_1$,
- ii) $\|p\|_Y \leq K_2$,
- iii) $\|p - \hat{p}\|_Y \leq K_3(\alpha^*, \beta^*) \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}}$.

DEMOSTRACIÓN 3.7. Se notará $u := u_{\alpha,\beta}$ como la solución de la ecuación de estado (3.19b) correspondiente a (α, β) y $p := p_{\alpha,\beta}$ como la solución de la ecuación adjunta (3.22) correspondiente a (α, β) . Empezaremos con la estimación de $\|u\|_{H^1(\Omega)}$, a continuación se presenta la estimación de $\|p\|_Y$ y finalizaremos con la estimación de $\|p - \hat{p}\|_Y$, donde $Y = H^1(\Omega) \times \mathbb{H}^1(\Omega)$.

Para estimar $\|u\|_{H^1(\Omega)}$ se usa la ecuación de estado, es decir,

$$a(y, \Psi) + \int_{\Omega} P(\alpha)h_{\gamma}(Du - w)(D\phi - \varphi) dx + \int_{\Omega} Q(\beta)h_{\gamma}(Ew)E\varphi dx + \int_{\Omega} (u - f)\phi dx = 0,$$

donde $\Psi = (\phi, \varphi)$. Tomando $\Psi = (u, w)$, se obtiene que

$$\mu\|y\|_Y^2 + \int_{\Omega} P(\alpha)h_{\gamma}(Du - w)(Du - w) dx + \int_{\Omega} Q(\beta)h_{\gamma}(Ew)(Ew) dx + \int_{\Omega} u^2 dx = \int_{\Omega} f u dx, \forall \Psi = (\phi, \varphi) \in Y,$$

gracias a la monotonía de $h_{\gamma}(\cdot)$ y ya que $P(\alpha)$ y $Q(\beta)$ son positivas, se sigue que

$$\begin{aligned} \mu\|y\|_Y^2 &\leq \|f\|_{L^2(\Omega)}\|u\|_{L^2(\Omega)}, \\ &\leq \|f\|_{L^2(\Omega)}\|y\|_Y, \\ \|u\|_{H^1(\Omega)} &\leq \|y\|_Y \leq \frac{1}{\mu}\|f\|_{L^2(\Omega)} := C_1, \end{aligned}$$

para la última desigualdad se consideró que $\|y\|_Y^2 = \|u\|_{H^1(\Omega)}^2 + \|w\|_{\mathbb{H}^1(\Omega)}^2$. Ahora, para la estimación del estado adjunto se usa un proceso similar, considerando la ecuación adjunta (3.22), es decir,

$$\begin{aligned} \nabla_u F(u)\phi + \mu(p_1, \phi)_{H^1(\Omega)} + \mu(p_2, \varphi)_{\mathbb{H}^1(\Omega)} + \int_{\Omega} P(\alpha)h'_{\gamma}(Du - w)(D\phi - \varphi)(Dp_1 - p_2) \\ + \int_{\Omega} Q(\beta)h'_{\gamma}(Ew)E\varphi Ep_2 + \int_{\Omega} p_1\phi = 0, \forall \Psi = (\phi, \varphi) \in Y. \end{aligned}$$

tomando $\Psi = (\phi, \varphi) = (p_1, p_2)$, se sigue que

$$\begin{aligned} \mu\|p\|_Y^2 + \int_{\Omega} P(\alpha)h'_{\gamma}(Du - w)(Dp_1 - p_2)(Dp_1 - p_2) \\ + \int_{\Omega} Q(\beta)h'_{\gamma}(Ew)Ep_2Ep_2 + \int_{\Omega} p_1^2 = - \int_{\Omega} (u - f_0)p_1. \end{aligned}$$

Gracias a la positividad de $h'_\gamma(\cdot)$ y ya que $P(\alpha)$ y $Q(\beta)$ son positivas, se sigue que

$$\begin{aligned}\mu \|p\|_Y^2 &\leq \|u - f_0\|_{L^2(\Omega)} \|p_1\|_{L^2(\Omega)}, \\ &\leq \left(\|u\|_{L^2(\Omega)} + \|f_0\|_{L^2(\Omega)} \right) \|p\|_Y, \\ \|p\|_Y &\leq \frac{1}{\mu} \left(\|f_0\|_{L^2(\Omega)} + \|u\|_{H^1(\Omega)} \right) := K_1.\end{aligned}$$

Finalmente, para hallar la estimación de $\|p - \hat{p}\|_Y$, se usa exactamente el proceso anterior considerando la diferencia de la ecuación adjunta para p y \hat{p} ,

$$\begin{aligned}(\nabla_u F(u) - \nabla_u F(\hat{u}))\phi + \mu(p_1 - \hat{p}_1, \phi)_{H^1(\Omega)} + \mu(p_2 - \hat{p}_2, \varphi)_{H^1(\Omega)} \\ + \int_{\Omega} (P(\alpha)h'_\gamma(Du - w)(Dp_1 - p_2) - P(\hat{\alpha})h'_\gamma(D\hat{u} - \hat{w})(D\hat{p}_1 - \hat{p}_2)) (D\delta_Y - \varphi) \\ + \int_{\Omega} (Q(\beta)h'_\gamma(Ew)Ep_2 - Q(\hat{\beta})h'_\gamma(E\hat{w})E\hat{p}_2) E\varphi + \int_{\Omega} (p_1 - \hat{p}_1)\phi = 0, \forall \Psi = (\phi, \varphi) \in Y.\end{aligned}$$

Sumando y restando los términos

$$\int_{\Omega} P(\alpha)h'_\gamma(D\hat{u} - \hat{w})(Dp_1 - p_2), \quad \int_{\Omega} P(\hat{\alpha})h'_\gamma(D\hat{u} - \hat{w})(Dp_1 - p_2),$$

y

$$\int_{\Omega} Q(\beta)h'_\gamma(E\hat{w})Ep_2, \quad \int_{\Omega} Q(\hat{\beta})h'_\gamma(E\hat{w})E\hat{p}_2$$

y tomando $\Psi = (\phi, \varphi) = (p_1 - \hat{p}_1, p_2 - \hat{p}_2)$, se obtiene

$$\begin{aligned}\mu \|p - \hat{p}\|_Y^2 + \int_{\Omega} P(\hat{\alpha})h'_\gamma(D\hat{u} - \hat{w}) [D(p_1 - \hat{p}_1) - (p_2 - \hat{p}_2)] (D(p_1 - \hat{p}_1) - (p_2 - \hat{p}_2)) \\ + \int_{\Omega} (p_1 - \hat{p}_1)^2 + \int_{\Omega} Q(\hat{\beta})h'_\gamma(E\hat{w})E(p_2 - \hat{p}_2)E(p_2 - \hat{p}_2) = - \int_{\Omega} (u - \hat{u})(p_1 - \hat{p}_1) \\ - \int_{\Omega} \left\{ [P(\alpha) - P(\hat{\alpha})]h'_\gamma(D\hat{u} - \hat{w})(Dp_1 - p_2) + P(\alpha)[h'_\gamma(Du - w) - h'_\gamma(D\hat{u} - \hat{w})] \right. \\ \left. (Dp_1 - p_2) \right\} (D(p_1 - \hat{p}_1) - (p_2 - \hat{p}_2)) - \int_{\Omega} \left\{ (Q(\beta)[h'_\gamma(Ew) - h'_\gamma(E\hat{w})]Ep_2 \right. \\ \left. - (Q(\beta) - Q(\hat{\beta}))h'_\gamma(E\hat{w})E\hat{p}_2 \right\} E(p_2 - \hat{p}_2),\end{aligned}$$

gracias a la positividad de $h'_\gamma(\cdot)$ se verifica que,

$$\begin{aligned}\mu \|p - \hat{p}\|_Y^2 &\leq \|u - \hat{u}\|_{L^2(\Omega)} \|p_1 - \hat{p}_1\|_{L^2(\Omega)} + \left\{ \|P(\alpha) - P(\hat{\alpha})\|_{L^\infty(\Omega)} \|h'_\gamma(D\hat{u} - \hat{w})(Dp_1 - p_2)\|_{L^2(\Omega)} \right. \\ &\quad \left. + \|P(\alpha)\|_{L^\infty(\Omega)} \|(h'_\gamma(Du - w) - h'_\gamma(D\hat{u} - \hat{w}))(Dp_1 - p_2)\|_{L^2(\Omega)} \right\} \\ &\quad + \|D(p_1 - \hat{p}_1) - (p_2 - \hat{p}_2)\|_{L^2(\Omega)} + \left\{ \|Q(\beta)\|_{L^\infty(\Omega)} \|[h'_\gamma(Ew) - h'_\gamma(E\hat{w})]Ep_2\|_{L^2(\Omega)} \right. \\ &\quad \left. + \|Q(\beta) - Q(\hat{\beta})\|_{L^\infty(\Omega)} \|h'_\gamma(E\hat{w})E\hat{p}_2\|_{L^2(\Omega)} \right\} \|E(p_2 - \hat{p}_2)\|_{L^2(\Omega)}.\end{aligned}$$

Se deben estimar

$$\|(h'_\gamma(Du - w) - h'_\gamma(D\hat{u} - \hat{w}))(Dp_1 - p_2)\|_{\mathbb{L}^2} \text{ y } \|[h'_\gamma(Ew) - h'_\gamma(E\hat{w})]Ep_2\|_{\mathbb{L}^2},$$

para lo cual se requiere regularidad extra tanto del estado como del estado adjunto. A partir de la hipótesis $f \in L^\infty(\Omega)$ y $u_{\alpha,\beta} \in C^{1,\rho}(\Omega)$, para algún $\rho \in (0, 1)$, se obtiene

$$\|y_{\alpha,\beta} - y_{\hat{\alpha},\hat{\beta}}\|_{W^{1,r}} \leq C\|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}}, \quad (3.24)$$

para todo $(\alpha, \beta) \in V(\alpha^*, \beta^*)$, y para algún $r > 2$. Este resultado de regularidad para sistemas de segundo orden elípticos semilineales puede ser probada bajo ciertos supuestos sobre el dominio (ver [25, Teorema 1, Rem. 14]).

En el caso del estado adjunto, bajo la misma hipótesis y usando el resultado para sistemas elípticos lineales [42], se verifica para cualquier $s > 2$ que

$$\|p_{\alpha,\beta}\|_{W^{1,s}} \leq \hat{C}. \quad (3.25)$$

Observe que,

$$\begin{aligned} & \|(h'_\gamma(Du - w) - h'_\gamma(D\hat{u} - \hat{w}))(Dp_1 - p_2)\|_{\mathbb{L}^2} \\ & \leq \|h'_\gamma(Du - w) - h'_\gamma(D\hat{u} - \hat{w})\|_{L^p} \|Dp_1 - p_2\|_{L^q} \end{aligned}$$

donde $1/p + 1/q = 1/2$. Ya que $h'_\gamma(\cdot)$ es Lipschitz continuo de $L^p(\Omega)$ a $L^p(\Omega)$ (cf. [34]), y por los resultados (3.24) y (3.25) con $r = p, s = q = 2p/(p-2)$ respectivamente se obtiene que

$$\begin{aligned} & \|(h'_\gamma(Du - w) - h'_\gamma(D\hat{u} - \hat{w}))(Dp_1 - p_2)\|_{\mathbb{L}^2} \\ & \leq M_\gamma (\|D(u - \hat{u}) - (w - \hat{w})\|_{L^p \times \mathbb{L}^p}) \|p\|_{W^{1,s}}, \\ & \leq M_\gamma M_1 \|y - \hat{y}\|_{W^{1,r}} \|p\|_{W^{1,s}}, \\ & \leq M_2 \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}}. \end{aligned}$$

Siguiendo un esquema similar, se obtiene

$$\begin{aligned} \|[h'_\gamma(Ew) - h'_\gamma(E\hat{w})]Ep_2\|_{L^2(\Omega)} & \leq \|h'_\gamma(Ew) - h'_\gamma(E\hat{w})\|_{L^p} \|Ep_2\|_{L^q}, \\ & \leq K_\gamma \|E(w - \hat{w})\|_{L^p} \|p\|_{W^{1,q}}, \\ & \leq K_1 K_\gamma \|y - \hat{y}\|_{W^{1,r}}, \\ & \leq K_2 \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}}. \end{aligned}$$

Así, usando (3.21) y $|h'_\gamma(\cdot)| < 1$ se verifica que

$$\begin{aligned} \mu \|p - \hat{p}\|_Y &\leq C_1 \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}} + \|P(\alpha) - P(\hat{\alpha})\|_{L^\infty(\Omega)} \\ &\quad + \|P(\alpha)\|_{L^\infty(\Omega)} M_2 \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}} \\ &\quad + \|Q(\beta)\|_{L^\infty(\Omega)} K_2 \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}} + \|Q(\beta) - Q(\hat{\beta})\|_{L^\infty(\Omega)}. \end{aligned}$$

Por otro lado, P, Q son continuamente diferenciables en \mathbb{R}^N y \mathbb{R}^M , respectivamente, de donde se sigue que P, Q son localmente Lipschitz continuas. Esto se emplea para acotar $\|P(\alpha)\|_{L^\infty(\Omega)}$ y $\|P'(\alpha)\|_{L^\infty(\Omega)}$, como se muestra a continuación

$$\begin{aligned} \|P(\alpha)\|_{L^\infty(\Omega)} &= \|P(\alpha) - P(\alpha^*) + P(\alpha^*)\|_{L^\infty(\Omega)}, \\ &\leq \|P(\alpha) - P(\alpha^*)\|_{L^\infty(\Omega)} + \|P(\alpha^*)\|_{L^\infty(\Omega)}, \\ &\leq L_1(\alpha^*) \|\alpha - \alpha^*\|_{\mathbb{R}^N} + M_3(\alpha^*), \\ &\leq L_1(\alpha^*) \delta + M_3(\alpha^*), \end{aligned}$$

ya que $(\alpha, \beta) \in V_\delta(\alpha^*, \beta^*)$. Además se supone que P, Q tienen derivada Lipschitz continua local, de donde

$$\begin{aligned} \|P'(\alpha)\|_{L^\infty(\Omega)} &= \|P'(\alpha) - P'(\alpha^*) + P'(\alpha^*)\|_{L^\infty(\Omega)}, \\ &\leq \|P'(\alpha) - P'(\alpha^*)\|_{L^\infty(\Omega)} + \|P'(\alpha^*)\|_{L^\infty(\Omega)}, \\ &\leq L_2(\alpha^*) \|\alpha - \alpha^*\|_{\mathbb{R}^N} + M_4(\alpha^*), \\ &\leq L_2(\alpha^*) \delta + M_4(\alpha^*). \end{aligned}$$

Así se obtiene,

$$\begin{aligned} \mu \|p - \hat{p}\|_Y &\leq C_1 \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}} + L_3 \|\alpha - \hat{\alpha}\|_{\mathbb{R}^{N+M}} \\ &\quad + (L_1(\alpha^*) \delta + M_3(\alpha^*)) M_2 \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}} \\ &\quad + (L_4(\beta^*) \delta + M_5(\alpha^*)) K_2 \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}} + L_5(\beta^*) \|\beta - \hat{\beta}\|_{\mathbb{R}^M}, \\ \|p - \hat{p}\|_Y &\leq K_3(\alpha^*, \beta^*) \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}}. \end{aligned}$$

que corresponde al resultado deseado.

En el siguiente resultado, se prueba para η fijo la Lipschitz continuidad local en el caso del modelo TGV. Se emplea la definición 2.1 de Lipschitz continuidad que se introdujo en el capítulo anterior.

TEOREMA 3.8. Si $f \in L^\infty(\Omega)$ y $u_{\alpha, \beta} \in C^{1,\rho}(\Omega)$, para algún $\rho \in (0, 1)$ entonces $\nabla \mathcal{F}_\eta$ es localmente Lipschitz continua.

DEMOSTRACIÓN 3.8. Sea $(\alpha^*, \beta^*) \in \mathbb{R}^{N+M}$ arbitrario pero fijo, $(\alpha, \beta), (\hat{\alpha}, \hat{\beta}) \in V_\delta((\alpha^*, \beta^*))$. Se debe hallar una constante $L(\alpha^*, \beta^*) > 0$ tal que,

$$\|\nabla \mathcal{F}(\alpha, \beta) - \nabla \mathcal{F}(\hat{\alpha}, \hat{\beta})\| \leq L(\alpha^*, \beta^*) \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|.$$

Se consideran $u := u_{\alpha, \beta}$, $\hat{u} := u_{\hat{\alpha}, \hat{\beta}}$, $p := p(\alpha, \beta)$ y $\hat{p} := p(\hat{\alpha}, \hat{\beta})$, donde $u_{\alpha, \beta}$ denota la solución de (3.19b) asociada con (α, β) y p el estado adjunto solución de (3.22) para $u_{\alpha, \beta}$. Se tiene entonces que

$$\begin{aligned} \|\nabla \mathcal{F}_\eta(\alpha, \beta) - \nabla \mathcal{F}_\eta(\hat{\alpha}, \hat{\beta})\| &\leq 2\zeta \|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\| + \left\| \int_{\Omega} \min(0, \eta P(\alpha)) P'(\alpha) - \right. \\ &\quad \left. \int_{\Omega} \min(0, \eta P(\hat{\alpha})) P'(\hat{\alpha}) \right\| + \left\| \int_{\Omega} \min(0, \eta Q(\beta)) Q'(\beta) - \int_{\Omega} \min(0, \eta Q(\hat{\beta})) Q'(\hat{\beta}) \right\| \\ &\quad + \left\| \int_{\Omega} P'(\alpha) h_\gamma(Du - w)(Dp_1 - p_2) - \int_{\Omega} P'(\hat{\alpha}) h_\gamma(D\hat{u} - \hat{w})(D\hat{p}_1 - \hat{p}_2) \right\| \\ &\quad + \left\| \int_{\Omega} Q'(\beta) h_\gamma(Ew) Ep_2 - \int_{\Omega} Q'(\hat{\beta}) h_\gamma(E\hat{w}) E\hat{p}_2 \right\|. \end{aligned}$$

En el caso del segundo término de $\|\nabla \mathcal{F}_\eta(\alpha, \beta) - \nabla \mathcal{F}_\eta(\hat{\alpha}, \hat{\beta})\|$ se suma y resta el término

$$\int_{\Omega} \min(0, \eta P(\hat{\alpha})) P'(\alpha).$$

Ya que \min es una función Lipschitz continua con constante uno, se sigue que

$$\begin{aligned} &\left\| \int_{\Omega} \min(0, \eta P(\alpha)) P'(\alpha) - \int_{\Omega} \min(0, \eta P(\hat{\alpha})) P'(\hat{\alpha}) \right\| \\ &\leq \left\| \min(0, \eta P(\alpha)) - \min(0, \eta P(\hat{\alpha})) \right\|_{L^2(\Omega)} \|P'(\alpha)\|_{L^2(\Omega)} \\ &\quad + \left\| \min(0, \eta P(\hat{\alpha})) \right\|_{L^2(\Omega)} \|P'(\alpha) - P'(\hat{\alpha})\|_{L^2(\Omega)}, \\ &\leq \eta \|P(\alpha) - P(\hat{\alpha})\|_{L^2(\Omega)} \|P'(\alpha)\|_{L^2(\Omega)} + (\|P(\hat{\alpha}) - P(\alpha^*)\|_{L^2(\Omega)} \\ &\quad + \|P(\alpha^*)\|_{L^2(\Omega)}) L_3(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}, \\ &\leq L_4(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} (L_2(\alpha^*)\delta + M_4(\alpha^*)) + (L_5(\alpha^*)\delta + M_1(\alpha^*)) L_3(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}, \\ &\leq M(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}. \end{aligned}$$

Usando el mismo proceso para el caso del término asociado a Q , se obtiene

$$\left\| \int_{\Omega} \min(0, \eta Q(\beta)) Q'(\beta) - \int_{\Omega} \min(0, \eta Q(\hat{\beta})) Q'(\hat{\beta}) \right\| \leq N(\beta^*) \|\beta - \hat{\beta}\|_{\mathbb{R}^M}.$$

Ahora, para los dos últimos términos de $\|\nabla \mathcal{F}(\alpha, \beta) - \nabla \mathcal{F}(\hat{\alpha}, \hat{\beta})\|$, sumando y restando el término

$$\int_{\Omega} P'(\alpha) h_\gamma(D\hat{u} - \hat{w})(D\hat{p}_1 - \hat{p}_2)$$

se sigue que,

$$\begin{aligned} & \left\| \int_{\Omega} P'(\alpha) h_{\gamma}(Du - w)(Dp_1 - p_2) - \int_{\Omega} P'(\hat{\alpha}) h_{\gamma}(D\hat{u} - \hat{w})(D\hat{p}_1 - \hat{p}_2) \right\| \\ & \leq \underbrace{\int_{\Omega} \|P'(\alpha) (h_{\gamma}(Du - w)(Dp_1 - p_2) - h_{\gamma}(D\hat{u} - \hat{w})(D\hat{p}_1 - \hat{p}_2))\|}_{a} \\ & \quad + \underbrace{\int_{\Omega} \|(P'(\alpha) - P'(\hat{\alpha})) h_{\gamma}(D\hat{u} - \hat{w})(D\hat{p}_1 - \hat{p}_2)\|}_{b}. \end{aligned}$$

Empecemos acotando el término más simple, (b)

$$\begin{aligned} b & \leq \|P'(\alpha) - P'(\hat{\alpha})\|_{L^{\infty}(\Omega)} \|h_{\gamma}(D\hat{u} - \hat{w})\|_{\mathbb{L}^2} \|D\hat{p}_1 - \hat{p}_2\|_{\mathbb{L}^2}, \\ & \leq L_3(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} \|D\hat{p}_1 - \hat{p}_2\|_{\mathbb{L}^2}, \\ & \leq L_3(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} \|\hat{p}\|_Y = C_1(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N}. \end{aligned}$$

En el caso del término (a), se suma y resta el término

$$P'(\alpha) h_{\gamma}(D\hat{u} - \hat{w})(Dp_1 - p_2)$$

para obtener,

$$\begin{aligned} a & \leq \|P'(\alpha)\|_{L^{\infty}(\Omega)} \{ \|h_{\gamma}(Du - w) - h_{\gamma}(D\hat{u} - \hat{w})\|_{\mathbb{L}^2} \|Dp_1 - p_2\|_{\mathbb{L}^2} \\ & \quad + \|h_{\gamma}(D\hat{u} - \hat{w})\|_{\mathbb{L}^2} \|Dp_1 - p_2 - (D\hat{p}_1 - \hat{p}_2)\|_{\mathbb{L}^2} \}, \\ & \leq (L_2(\alpha^*)\delta + M_4(\alpha^*)) \{ \|h_{\gamma}(Du - w) - h_{\gamma}(D\hat{u} - \hat{w})\|_{\mathbb{L}^2} \|p\|_Y + \|p - \hat{p}\|_Y \}, \\ & \leq (L_2(\alpha^*)\delta + M_4(\alpha^*)) \{ L_3 \|D(u - \hat{u}) - (w - \hat{w})\|_{\mathbb{L}^2} + \|p - \hat{p}\|_Y \}, \\ & \leq (L_2(\alpha^*)\delta + M_4(\alpha^*)) (L_3 \|y - \hat{y}\|_Y + \|p - \hat{p}\|_Y), \\ & \leq C_2(\alpha^*) \|(\alpha, \hat{\alpha}) - (\beta, \hat{\beta})\|_{\mathbb{R}^{N+M}}, \end{aligned}$$

por tanto, se sigue que

$$\begin{aligned} & \left\| \int_{\Omega} P'(\alpha) h_{\gamma}(Du - w)(Dp_1 - p_2) - \int_{\Omega} P'(\hat{\alpha}) h_{\gamma}(D\hat{u} - \hat{w})(D\hat{p}_1 - \hat{p}_2) \right\| \\ & \leq C_1(\alpha^*) \|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} + C_2(\alpha^*) \|(\alpha, \hat{\alpha}) - (\beta, \hat{\beta})\|_{\mathbb{R}^{N+M}}. \end{aligned}$$

Usando un proceso similar, se obtiene que

$$\begin{aligned} & \left\| \int_{\Omega} Q'(\beta) h_{\gamma}(Ew) Ep_2 - \int_{\Omega} Q'(\hat{\beta}) h_{\gamma}(E\hat{w}) E\hat{p}_2 \right\| \leq \|Q'(\beta) - Q'(\hat{\beta})\|_{\mathbb{L}^2} \|h_{\gamma}(Ew) Ep_2\|_{\mathbb{L}^2} \\ & \quad + \|Q'(\hat{\beta})\|_{L^{\infty}(\Omega)} \|h_{\gamma}(Ew) - h_{\gamma}(E\hat{w})\|_{\mathbb{L}^2} \|Ep_2\|_{\mathbb{L}^2} + \|Ep_2 - E\hat{p}_2\|_{\mathbb{L}^2} \|h_{\gamma}(E\hat{w})\|_{\mathbb{L}^2}, \\ & \leq L_3(\beta^*) \|\beta - \hat{\beta}\| + (L_4(\beta^*)\delta + M_5(\beta^*)) \|(\alpha, \hat{\alpha}) - (\beta, \hat{\beta})\|_{\mathbb{R}^{N+M}} + \|p - \hat{p}\|_Y, \\ & \leq M_2(\beta^*) \|(\alpha, \hat{\alpha}) - (\beta, \hat{\beta})\|_{\mathbb{R}^{N+M}}. \end{aligned}$$

Finalmente,

$$\begin{aligned}
\|\nabla \mathcal{F}_\eta(\alpha, \beta) - \nabla \mathcal{F}_\eta(\hat{\alpha}, \hat{\beta})\| &\leq 2\zeta\|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}} + M(\alpha^*)\|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} + N(\beta^*)\|\beta - \hat{\beta}\|_{\mathbb{R}^M} \\
&\quad + C_1(\alpha^*)\|\alpha - \hat{\alpha}\|_{\mathbb{R}^N} + C_2(\alpha^*)\|(\alpha, \hat{\alpha}) - (\beta, \hat{\beta})\|_{\mathbb{R}^{N+M}} \\
&\quad + M_2(\beta^*)\|(\alpha, \hat{\alpha}) - (\beta, \hat{\beta})\|_{\mathbb{R}^{N+M}}, \\
&\leq L(\alpha^*, \beta^*)\|(\alpha, \beta) - (\hat{\alpha}, \hat{\beta})\|_{\mathbb{R}^{N+M}}
\end{aligned}$$

que corresponde al resultado.

3.7. Implementación numérica

En esta sección se abordará la implementación numérica del problema de optimización de parámetros, para lo cual empezaremos por presentar de forma resumida el trabajo desarrollado hasta este punto. A continuación se presentan los esquemas de discretización de las ecuaciones de estado, adjunta y del gradiente para finalizar, se discuten los resultados y conclusiones obtenidos en los experimentos numéricos, a partir de la implementación de los algoritmos propuestos para resolver el problema de optimización del parámetro con dependencia espacial presente en el modelo de variación total generalizada para el filtrado de ruido.

Gracias a los resultados expuestos en la sección 3.5 se puede abordar el problema de optimización binivel (3.3) como un problema de optimización sin restricciones de desigualdad usando la regularización de Moreau-Yosida, el mismo que está determinado por:

$$\min_{(\alpha, \beta) \in \mathbb{R}^{N+M}} F(u_{\alpha, \beta}) + \frac{1}{2\eta} \|\min(0, \eta P(\alpha))\|_{L^2(\Omega)}^2 + \frac{1}{2\eta} \|\min(0, \eta Q(\beta))\|_{L^2(\Omega)}^2, \quad (3.26a)$$

sujeto a

$$u_{\alpha, \beta} \in \arg \min_{u \in BV(\Omega)} E_{TGV} = TGV_{\alpha, \beta}^{2, \gamma, \mu}(u) + \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2, \quad (3.26b)$$

donde

$$F(u_{\alpha, \beta}) = \frac{1}{2} \|f_0 - u\|_{L^2(\Omega)}^2 + \zeta(\|\alpha\|_{\mathbb{R}^N}^2 + \|\beta\|_{\mathbb{R}^M}^2),$$

$$\begin{aligned}
TGV_{\alpha, \beta}^{2, \gamma, \mu}(u) := \min_{w \in H^1(\Omega)} \int_{\Omega} P(\alpha) |Du - w|_{\gamma} + \int_{\Omega} Q(\beta) |Ew|_{\gamma} \\
+ \frac{\mu}{2} \left(\|u\|_{H^1(\Omega)}^2 + \|w\|_{H^1(\Omega)}^2 \right).
\end{aligned}$$

Para la solución numérica de este problema de optimización se propone usar el método quasi-Newton BFGS, esto gracias a la caracterización que se encontró del gradiente del funcional de costo reducido (3.23). El esquema del método de BFGS requiere resolver en cada iteración la ecuación de estado y la ecuación adjunta hasta que se verifique el criterio de

parada, por lo cual empezaremos por presentar los algoritmos empleados para resolver estas ecuaciones. La ecuación de estado corresponde a la condición de optimalidad de primer orden del problema de optimización para el filtrado de ruido (3.26b), determinada por

$$E'_{TGV} = a(y, \Psi) + \int_{\Omega} P(\alpha) h_{\gamma}(Du - w)(D\phi - \varphi) dx + \int_{\Omega} Q(\beta) h_{\gamma}(Ew) E\varphi dx + \int_{\Omega} (u - f)\phi dx = 0, \text{ para todo } \Psi = (\phi, \varphi) \in Y,$$

donde $y = (u, w) \in Y = H^1(\Omega) \times \mathbf{H}^1(\Omega)$ y $a(y, y) = \mu \left(\|u\|_{H^1(\Omega)}^2 + \|w\|_{\mathbf{H}^1(\Omega)}^2 \right)$. Para encontrar su solución numérica se resolverá el problema de optimización empleando el método de Newton Semi-Smooth y para reducir el número de variables del sistema se considera un proceso de reducción que se explicará más adelante.

Sean q, r variables auxiliares, se define el operador H como

$$H(u, q, w, r) = \begin{pmatrix} \mu(u, \phi)_{H^1(\Omega)} + (P(\alpha)q, D\phi) + \int_{\Omega} (u - f)\phi dx \\ q - h_{\gamma}(Du - w) \\ \mu(w, \varphi)_{\mathbf{H}^1(\Omega)} - (P(\alpha)q, \varphi) + (Q(\beta)r, E\varphi) \\ r - h_{\gamma}(Ew) \end{pmatrix},$$

así la ecuación de estado será $H(u, q, w, r) = 0$. El esquema del método de Newton Semi-Smooth para resolver esta ecuación no lineal Newton diferenciable es

$$H'(u, q, w, r)(\delta u, \delta q, \delta w, \delta r) = -H(u, q, w, r), \quad (3.27)$$

donde $(\delta u, \delta q, \delta w, \delta r)$ corresponde a la dirección en la cual se deriva el operador H ,

$$H'(u, q, w, r)(\delta u, \delta q, \delta w, \delta r) = \begin{pmatrix} \mu(\delta u, \phi)_{H^1(\Omega)} + \int_{\Omega} \delta u \phi dx + (P(\alpha)\delta q, D\phi)_{L^2(\Omega)} \\ -h'_{\gamma}(Du - w)D\delta u + \delta q + h'_{\gamma}(Du - w)\delta w \\ -(P(\alpha)\delta q, \varphi) + \mu(\delta w, \varphi)_{\mathbf{H}^1(\Omega)} + (Q(\beta)\delta r, E\varphi) \\ -h'_{\gamma}(Ew)E\delta w + \delta r \end{pmatrix},$$

con $h_{\gamma}(\cdot)$ la derivada de la regularización de Huber $|\cdot|_{\gamma}$ definida por:

$$|w|_{\gamma} = \begin{cases} \frac{\gamma}{2}|w|^2 & \text{si } |w| \leq \frac{1}{\gamma}, \\ |w| - \frac{1}{2\gamma} & \text{si } |w| > \frac{1}{\gamma}, \end{cases}$$

y su derivada es:

$$h_{\gamma}(w) = \frac{\gamma w}{\max(1, \gamma|w|)},$$

se considera esta regularización por ser la más natural para el proceso de reducción de variables. Además, $h'_{\gamma}(\cdot)$ es la derivada de Newton de $h_{\gamma}(\cdot)$. La matriz H' deberá ser modificada con una proyección para garantizar que la matriz $H'(u, q, w, r)$ sea definida positiva. Como

se muestra en el trabajo [7] la proyección consiste en reemplazar

$$q \quad \text{por} \quad \hat{q} := \frac{q}{\max(1, |q|)}.$$

A continuación, se define la *Newton diferenciabilidad*, pues la presencia de la función $\max(1, \cdot)$ proveniente de la regularización, hace que no se pueda aplicar el método de Newton clásico.

3.7.1. Método de Newton Semi-Smooth

Para resolver el problema de filtrado de ruido usando el modelo de variación total generalizada TGV, se propone emplear el método de Newton Semi-Smooth, esto por la presencia del término no diferenciable $\max(1, \cdot)$.

DEFINICIÓN 3.1. (*Newton-diferenciabilidad*) Sean D un subconjunto abierto de un espacio de Banach u y $F: D \rightarrow Z$ un funcional. F es Newton diferenciable en el subconjunto abierto $V \subset D$ si existe $G: V \rightarrow \mathcal{L}(U, Z)$, una derivada generalizada tal, que:

$$\lim_{h \rightarrow 0} \frac{1}{\|h\|_U} \|E(u+h) - E(u) - G(u+h)h\|_Z = 0,$$

para todo $u \in V$.

EJEMPLO 1. Considere la función valor absoluto f :

$$\begin{aligned} E: \mathbb{R} &\rightarrow \mathbb{R} \\ u &\mapsto |u|. \end{aligned}$$

La función f es no diferenciable en 0. Sin embargo, mediante el uso de la derivada generalizada

$$g(u) = \begin{cases} 1, & \text{si } u < 0; \\ -1, & \text{si } u \geq 0. \end{cases}$$

para el caso $u = 0$ se obtiene:

- si $h > 0$: $\|u+h\| - \|u\| - \|h\| = 0$,
- si $h < 0$: $\|u+h\| - \|u\| + \|h\| = |-u-h - u+h| = 0$.

Así,

$$\lim_{h \rightarrow 0} \frac{1}{|h|} |E(u+h) - E(u) - g(u+h)h| = 0$$

y f es Newton diferenciable.

TEOREMA 3.9. (*Iteración de Newton*) Sea \tilde{u} una solución de $E(u) = 0$, con F Newton diferenciable en una vecindad abierta de V de \tilde{u} . Si

$$\|G(u)^{-1}\|_{\mathcal{L}(Z,U)} \leq C,$$

para alguna constante $C > 0$ y para todo $u \in V$, entonces la iteración Newton

$$u_{k+1} = -G(u_k)^{-1}E(u_k)$$

converge superlinealmente a \tilde{u} siempre que $\|u_0 - \tilde{u}\|_U$ sea suficientemente pequeño.

PROPOSICIÓN 3.10. La función $\text{máx}(1, \cdot): \mathbb{R}^n \rightarrow \mathbb{R}^n$ es Newton diferenciable en \mathbb{R}^n , con derivada generalizada:

$$G_{\text{máx}}: \mathbb{R}^n \rightarrow \mathbb{R}^n$$

$$(G_{\text{máx}}(u))_i = \begin{cases} 1, & (u)_i > 1, \\ 0, & (u)_i \leq 1. \end{cases}$$

Retomando el esquema de Newton Semi-Smooth (3.27) presentaremos a continuación el proceso de eliminación de variables. En primer lugar, se igualan la segunda y tercera componente de (3.27), de donde se obtienen las siguientes ecuaciones:

$$H_\gamma \hat{q} D\delta u - \gamma D\delta u + \text{máx}(1, \gamma|Du - w|)\delta q + H_\gamma \hat{q} \delta w + \gamma \delta w$$

$$= -(\text{máx}(1, \gamma|Du - w|)q - \gamma(Du - w))$$

$$He_\gamma \hat{r} E\delta w - \gamma E\delta w + \text{máx}(1, \gamma|Ew|)\delta r = -(\text{máx}(1, \gamma|Ew|)r - \gamma(Ew)),$$

donde

$$H_\gamma = G_{\text{máx}}(Du - w)\gamma \frac{Du - w}{|Du - w|} \quad \text{y} \quad He_\gamma = G_{\text{máx}}(Ew)\gamma \frac{Ew}{|Ew|},$$

despejando δq y δr obtenemos:

$$\delta q = -I_{\text{máx}}(\text{máx}(1, \gamma|Du - w|)q - \gamma(Du - w) + H_\gamma \hat{q} D\delta u - \gamma D\delta u + H_\gamma \hat{q} \delta w + \gamma \delta w)$$

$$\delta r = -I_{e_{\text{máx}}}(\text{máx}(1, \gamma|Ew|)r - \gamma(Ew) + He_\gamma \hat{r} E\delta w - \gamma E\delta w).$$

donde $I_{\text{máx}} = \text{diag}(1/\text{máx}(1, \gamma|Du - w|))$ y $I_{e_{\text{máx}}} = \text{diag}(1/\text{máx}(1, \gamma|Ew|))$. Como siguiente paso, reemplazamos δq y δr para obtener el sistema reducido

$$\overline{H}'(u, q, w, r)(\delta u, \delta w) = -\overline{H}(u, q, w, r).$$

Para la implementación numérica se empleó esquemas de discretización con diferencias finitas, ampliamente utilizado en el procesamiento matemático de imágenes debido a la estructura de la imagen digital. Se consideran $\Omega = (0,1) \times (0,1)$ el dominio de las imágenes en escala de grises, $n \times n$ la dimensión de las mismas, $h = 1/(n+1)$ el paso de discretización de la malla, \mathbb{G} matriz de discretización del gradiente, \mathbb{A} la matriz de discretización del laplaciano, $\overrightarrow{\mathbb{A}}$ la matriz de discretización del laplaciano vectorial, $\overrightarrow{\text{div}}$ matriz de discretización de la divergencia, \mathbb{E} la discretización del tensor simetrizado y $\overrightarrow{\text{Div}}$ matriz de discretización de la divergencia vectorial. El algoritmo para resolver la ecuación de estado para α y β fijos se presenta a continuación:

Algorithm 4 Newton Semi-Smooth–ecuación de estado

- 1: Dados $f, \alpha, \beta, \mu, \gamma$
- 2: Evaluar $P(\alpha), Q(\beta)$
- 3: **Inicializar** $k = 0, u_0, q_0, w_0, r_0$
- 4: **Repetir** $\bar{H}'(u, q, w, r)(\Delta u, \Delta w) = -\bar{H}(u, q, w, r)$, con

$$\bar{H}(u_k, q_k, w_k, r_k) = \begin{pmatrix} -\mu \mathbf{A}u + \mu u + (u - f) - \gamma \operatorname{div} P(\alpha) I_{\max}(\mathbf{G}u - w) \\ -\mu \vec{\mathbf{A}}w + \mu w - \gamma P(\alpha) I_{\max}(\mathbf{G}u - w) - \gamma \vec{\operatorname{Div}} Q(\beta) I_{e_{\max}} \mathbf{E}w \end{pmatrix}$$

$$\bar{H}'(u_k, q_k, w_k, r_k) =$$

$$\begin{pmatrix} -\mu \mathbf{A} + (\mu u + 1)I - \mathbf{G}'P(\alpha) I_{\max}(H_\gamma \hat{q} \mathbf{G} - \gamma \mathbf{G}) & -\mathbf{G}'P(\alpha) I_{\max}(-H_\gamma \hat{q} + \gamma I) \\ P(\alpha) I_{\max}(H_\gamma \hat{q} \mathbf{G} - \gamma \mathbf{G}) & -\mu \vec{\mathbf{A}}w + \mu \mathbf{I} + P(\alpha) I_{\max}(-H_\gamma \hat{q} + \gamma I) - \vec{\operatorname{Div}} Q(\beta) I_{e_{\max}}(H_{e_\gamma} \hat{r} - \gamma \mathbf{E}) \end{pmatrix}$$

- 5: **Actualizar**

$$\Delta q = -I_{\max}(\max(1, \gamma |\mathbf{G}u - w|)q - \gamma(\mathbf{G}u - w) + H_\gamma \hat{q} \mathbf{G} \Delta u - \gamma \mathbf{G} \Delta u + H_\gamma \hat{q} \Delta w + \gamma \Delta w)$$

$$\Delta r = -I_{e_{\max}}(\max(1, \gamma |\mathbf{E}w|)r - \gamma(\mathbf{E}w) + H_{e_\gamma} \hat{r} \mathbf{E} \Delta w - \gamma \mathbf{E} \Delta w).$$

$$(u_{k+1}, q_{k+1}, w_{k+1}, r_{k+1}) = (u_k, q_k, w_k, r_k) + (\Delta u, \Delta q, \Delta w, \Delta r)$$

$$k = k + 1$$

- 6: **Hasta** verificar criterio de parada $\|(u_{k+1}, q_{k+1}, w_{k+1}, r_{k+1}) - (u_k, q_k, w_k, r_k)\| < tol$

- 7: **Retornar** (u_k, w_k)
-

La ecuación adjunta (3.22), corresponde a una ecuación lineal respecto a la variable $p = (p_1, p_2) \in Y$, y está determinada como se presenta a continuación:

$$\mu(p_1, \phi)_{H^1(\Omega)} + (P(\alpha)h'_\gamma(Du - w)(Dp_1 - p_2), D\phi) + (p_1, \phi) = -(u - f_o, \phi),$$

$$\mu(p_2, \varphi)_{H^1(\Omega)} - (P(\alpha)s, \varphi) + (Q(\beta)h'_\gamma(Ew)Ep_2, E\varphi) = 0,$$

donde f_o es la imagen original o una aproximación de la misma y $h'_\gamma(\cdot)$ determinado por

$$h'_\gamma(Du)D\phi = \begin{cases} \frac{D\phi}{|Du|} - \frac{\langle Du, D\phi \rangle Du}{|Du|^3} & A\gamma, \\ (1 - \frac{\gamma}{2}\theta^2(Du)) \left(\frac{D\phi}{|Du|} - \frac{Du \langle D\phi, Du \rangle}{|Du|^3} \right) + \gamma^2 \theta(Du) \frac{\langle Du, D\phi \rangle Du}{|Du|^2} & S\gamma, \\ \gamma D\phi & I\gamma. \end{cases}$$

En este caso no se utilizan variables auxiliares a diferencia del proceso que se realizó para el caso del modelo TV, ya que el tamaño del sistema que se debe resolver usando variables auxiliares es de $9n^2 \times 9n^2$ y sin usarla es de, $5n^2 \times 5n^2$. Usando esquemas de discretización

en diferencias finitas se obtiene el siguiente algoritmo para determinar el estado adjunto $p = (p_1, p_2)$, para α y β fijos

Algorithm 5 Ecuación adjunta

- 1: Dados $f_0, u, w, \alpha, \beta, \mu, \gamma$
- 2: Evaluar $P(\alpha), Q(\beta)$
- 3: **Resolver**

$$\begin{pmatrix} -\mu\mathbb{A} + (\mu + 1)\mathbb{I} - \text{div}P(\alpha)h'_\gamma(\mathbb{G}u - w)\mathbb{G} & \text{div}P(\alpha)h'_\gamma(\mathbb{G}u - w) \\ -P(\alpha)h'_\gamma(\mathbb{G}u - w)\mathbb{G} & -\mu\vec{\mathbb{A}} + \mu\mathbb{I} + P(\alpha)h'_\gamma(\mathbb{G}u - w) - \vec{\text{Div}}Q(\beta)h'_\gamma(\mathbb{E}w)\mathbb{E} \end{pmatrix} \begin{pmatrix} p_1 \\ p_2 \end{pmatrix} = \begin{pmatrix} -u + f_0 \\ \mathbf{0} \end{pmatrix}$$

- 4: **Retornar** $p = (p_1, p_2)$
-

Gracias a la caracterización del gradiente del funcional de costo reducido obtenida en (3.23), es decir,

$$\begin{aligned} \nabla \mathcal{F}_\eta(\alpha, \beta) = & \nabla_\alpha F(u_{\alpha, \beta}) + \nabla_\beta F(u_{\alpha, \beta}) + \int_\Omega \text{mín}(0, \eta P(\alpha)) P'(\alpha) dx \\ & + \int_\Omega \text{mín}(0, \eta Q(\beta)) Q'(\beta) dx + \int_\Omega P'(\alpha) h_\gamma(Du - w)(Dp_1 - p_2) dx \\ & + \int_\Omega Q'(\beta) h_\gamma(Ew) Ep_2 dx \end{aligned}$$

se propone utilizar el método BFGS para resolver el problema de optimización de parámetros con dependencia espacial. Las integrales dobles se aproximaron usando la regla de punto medio (RPM) determinada por:

$$\int_a^b \int_c^d f(x, y) dy dx \approx \text{RPM}(f, h)$$

donde,

$$\begin{aligned} \text{RPM}(f, h) = & \frac{1}{4} h^2 \{ f(a, c) + f(b, c) + f(a, d) + f(b, d) + \\ & 2 \sum_{i=1}^{n-2} f(x_i, c) + 2 \sum_{i=1}^{n-2} f(x_i, d) + 2 \sum_{i=1}^{n-2} f(a, y_i) + \\ & 2 \sum_{i=1}^{n-2} f(b, y_i) + 4 \sum_{i=1}^{n-2} \sum_{j=1}^{n-2} f(x_i, y_j) \}. \end{aligned}$$

Así se obtiene la aproximación del gradiente,

$$\begin{aligned} \nabla \mathcal{F}_\eta(\alpha, \beta) \approx & 2\zeta(\alpha + \beta) + \text{RPM}(\text{mín}(0, \eta P(\alpha))P'(\alpha)) \\ & + \text{RPM}(\text{mín}(0, \eta Q(\beta))Q'(\beta)) + \text{RPM}(P'(\alpha)h_\gamma(\mathbb{G}u - w)(\mathbb{G}p_1 - p_2)) \\ & + \text{RPM}(Q'(\beta)h_\gamma(\mathbb{E}w)\mathbb{E}p_2), \end{aligned}$$

donde $P'(\cdot)$ y $Q'(\cdot)$ corresponden a la derivada de P y Q respectivamente. En 6, se presenta el algoritmo que nos permitirá determinar los parámetros α y β óptimos. Siguiendo las ideas presentadas en el trabajo [14] se eligió como criterio de parada

$$\frac{\|(\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)\|_{\mathbb{R}^{N+M}}}{\|(\alpha_{k+1}, \beta_{k+1})\|_{\mathbb{R}^{N+M}}} < \text{tol}$$

y se fijo la tolerancia como $\text{tol} = 10^{-6}$.

Algorithm 6 BFGS optimización de parámetros

- 1: Dados $f, f_0, \mu, \gamma, \eta, \zeta$
- 2: **Inicializar** $k = 0, B_0, \alpha_0, \beta_0$
- 3: **Repetir**
- 4: Calcular $P(\alpha_k), Q(\beta_k), P'(\alpha_k)$ y $Q'(\beta_k)$
- 5: Resolver la ecuación de estado usando el algoritmo 4, (u_k, w_k)
- 6: Resolver la ecuación adjunta usando el algoritmo 5, $p_k = (p_k^1, p_k^2)$
- 7: Calcular $\nabla \mathcal{F}_\eta(\alpha_k, \beta_k)$,
- 8: $s_k = (\alpha_{k+1}, \beta_{k+1}) - (\alpha_k, \beta_k)$,
- 9: $z_k = \nabla \mathcal{F}_\eta(\alpha_{k+1}, \beta_{k+1}) - \nabla \mathcal{F}_\eta(\alpha_k, \beta_k)$
- 10: **If** $s_k^T z_k > 0$ **then**

$$B_{k+1} = B_k + \frac{(s_k - B_k z_k) s_k^T + s_k (s_k - B_k z_k)^T}{s_k^T z_k} - \frac{(s_k - B_k z_k)^T z_k}{(s_k^T z_k)^2} s_k s_k^T.$$

- 11: Calcular la dirección de descenso como
 - 12: $d_k = -B_{k+1} \nabla \mathcal{F}_\eta(\alpha_k, \beta_k)$
 - 13: **else**
 - 14: Calcular la dirección de descenso como
 - 15: $d_k = -B_k \nabla \mathcal{F}_\eta(\alpha_k, \beta_k)$
 - 16: **End**
 - 17: Determinar el paso de descenso λ usando una de búsqueda lineal
 - 18: **Actualizar**
 - 19: $(\alpha_{k+1}, \beta_{k+1}) = (\alpha_k, \beta_k) + \lambda d_k$
 - 20: $k = k + 1$
 - 21: **Hasta** verificar criterio de parada.
 - 22: **Retornar** (α_k, β_k)
-

3.7.2. Filtrado de ruido de señales

Comparación modelo TV y TGV

Sabemos que la regularización de variación total (TV) [39] preserva los bordes y se comporta muy bien si la imagen a ser reconstruida vista es constante a trozos. Sin embargo, si se consideran funciones lineales a trozos se produce un efecto de escalonamiento [38]. Una posibilidad para contrarrestar tales efectos es introducir derivadas de orden superior en la regularización. Así, la *regularización de variación total generalizada (TGV)* fue introducida como una variante de orden superior de la regularización de variación total en el procesamiento digital de imágenes. Para evidenciar este efecto a continuación se presenta el problema de filtrado de ruido de señales, es decir, el problema en una dimensión, considerando los modelos TV y TGV.

En este experimento se considera una función lineal y constante a trozos con discontinuidades, con el fin de mostrar los efectos de los modelos TV y TGV para el filtrado de ruido.

Se define f_0 como la señal original sin ruido, como sigue

$$f_0 = \begin{cases} 0 & 0 \leq x \leq 1/4 \\ 1 + 2(x - 1/4) & 1/4 \leq x \leq 2/4 \\ 2 & 2/4 \leq x \leq 3/4 \\ 2 - 4(x - 3/4) & 3/4 \leq x \leq 1. \end{cases}$$

Para generar la señal con ruido se suma a la función $f_0: [0, 1] \rightarrow \mathbb{R}$ una variable aleatoria con distribución Gaussiana de media 0 y desviación estándar 0.1, a la que llamaremos f . Nuestro objetivo es filtrar el ruido y mantener las principales características de la función original f_0 , para lo cual, se prueban distintos valores de los parámetros de regularización (α, β) , esto para mostrar la importancia del problema de optimización de parámetros.

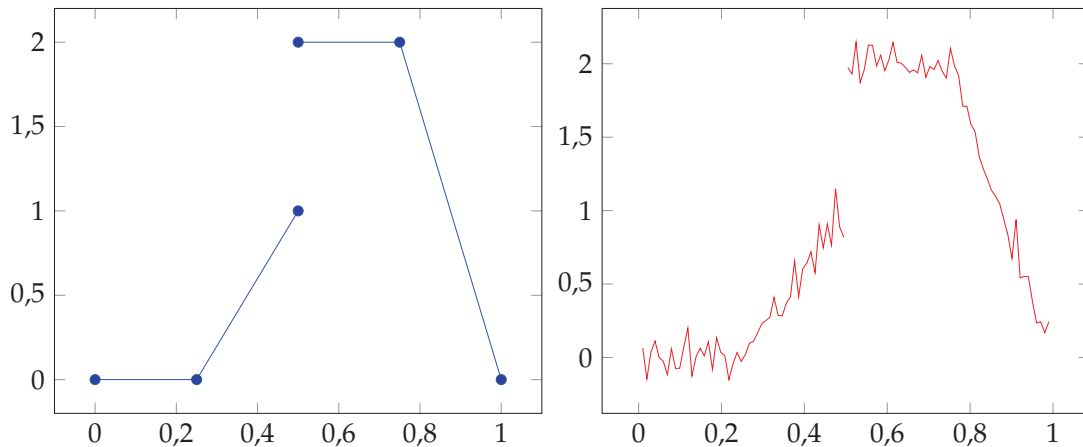


Figura 3.1. A la izquierda, la función original f_0 y a la derecha la función con ruido f .

Para comparar los resultados obtenidos, con diferentes valores de parámetros de los regularización escalar, se calcula la relación señal ruido SNR (Signal to noise ratio), que es una medida usada en ciencia e ingeniería para medir el nivel de una señal deseada con el nivel de ruido. Los parámetros empleados para este experimento fueron: número de puntos de la malla $n = 100$, tamaño de la malla $h = 1/(n + 1)$, regularización de Huber $\gamma = 1e3$, peso del término elíptico $\mu = 1e - 9$.

En la Tabla 3.1 se muestran los resultados obtenidos variando el valor del parámetro de regularización, además se presentan el número de iteraciones y el valor de SNR. En el caso del modelo de variación total TV el mejor resultado se obtuvo con $\alpha = 490$, mientras que en el caso del modelo de variación total generalizada TGV es mucho más difícil hallar el valor de los parámetros α, β por lo cual se emplea la heurística propuesta en [14], el mejor resultado se obtuvo con el parámetro $(\alpha, \beta) = (0.13, 0.3)$.

Modelo TV			Modelo TGV			
Parámetro α	Número de iteraciones	SNR	Parámetro α/n	Parámetro β/n^2	Número de iteraciones	SNR
10	25	9.4035	0.3		1615	31.9961
300	9	30.2541	0.05		14	29.1265
400	9	30.4900	0.1	0.3	13	34.4104
480	9	30.5280	0.12		15	35.3119
490	9	30.5282	0.13		15	35.4592
495	9	30.5280		0.31	14	34.3812
500	9	30.5277	0.1	0.1	29	33.2615
1000	12	30.2490		0.25	15	34.3812

Tabla 3.1. Resultados obtenidos con distintos valores del parámetro de regularización.

A continuación, se presentan en la Figura 3.2 en color rojo las mejores soluciones usando los modelos de variación total y de variación total generalizada y en color azul la función original.

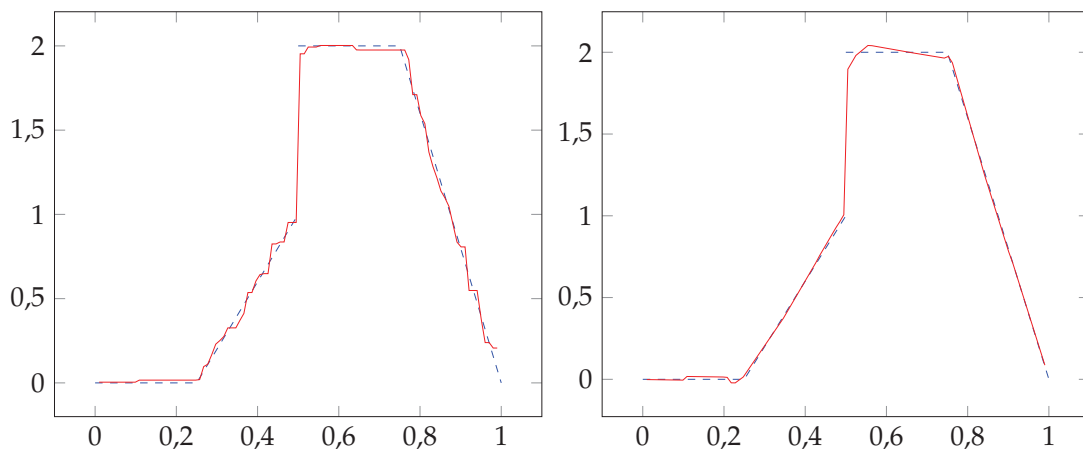


Figura 3.2. A la izquierda, la solución del modelo TV con $\alpha = 490$ y a la derecha la solución del modelo TGV con $(\alpha, \beta) = (0.13, 0.3)$.

Como se esperaba la regularización de variación total produce un efecto de escalera en los intervalos donde la función es lineal, mientras que en los intervalos donde la función es constante se acerca mucho a la señal original y por último hay que destacar que preserva los saltos. Por otro lado, la regularización de variación total generalizada se acerca mucho en los intervalos donde la función es lineal, mientras que en los intervalos donde la función es constante no se aproxima a la señal original y preserva los saltos. En el contexto de las imágenes, los saltos representan el cambio en la intensidad o color por presencia de bordes y los intervalos donde la señal es lineal representa el cambio de tono de un mismo color. Así, el modelo que se emplee dependerá de la características de la función a ser reconstruida y de los detalles que se quieran preservar.

Optimización de parámetros en una dimensión

A continuación, utilizamos al algoritmo de optimización de parámetros propuesto, para determinar el peso óptimo en el modelo de filtrado de ruido de señales usando el modelo TGV. El objetivo de resolver el problema en una dimensión es superar los inconvenientes que se presenten en la implementación numérica y entender la naturaleza del problema. Empecemos por realizar la siguiente observación

En el trabajo [14] se menciona la necesidad de usar estrategias de *warm initialisation* para garantizar la convergencia del algoritmo. Así, la resolución de los experimentos se realiza considerando un esquema de arranque en caliente, el cual consiste en usar la solución obtenida empleando el modelo TV como condición inicial para resolver el modelo TGV. Es decir, tomaremos la mejor solución obtenida usando el modelo variación total para inicializar los algoritmos del modelo de variación total generalizada TGV.

Se fijan los parámetros presentes en el problema binivel como: $\zeta = 1e - 8$ parámetro de Tikhonov, $\eta = 1e4$ parámetro de regularización de Moreau-Yosida y la tolerancia $tol = 1e - 8$. Considerando la idea del arranque en caliente, se emplea $(1/\alpha_{TV}, 2/(\alpha_{TV}n))$ como punto inicial, esto pues en el último experimento se determinó que $\alpha_{TV} = 490$ era el mejor valor del parámetro en el caso del modelo TV. En la Tabla 3.2 se muestran los resultados obtenidos, además se presentan el número de iteraciones, el valor del funcional objetivo y el valor de las medidas de comparación SNR y PSNR.

Punto inicial ($\alpha_0/n, \beta_0/n^2$)	SNR	PSNR	Número de iteraciones	Parámetro óptimo ($\alpha^*/n, \beta^*/n^2$)	F. objetivo
$(n/\alpha_{TV}, 2n/\alpha_{TV})$	35.7479	34.1923	4	(0.1559, 0.4413)	1.9043e-2

Tabla 3.2. Resultados de l problema de optimización de parámetros en una dimensión.

Las dificultades que se presentaron en este experimento fueron determinar el punto inicial, pues existen puntos que no llevan a una especie de planicie donde la función objetivo no mejora su valor, para superar esto se decidió usar el arranque en caliente.

En la Figura 3.3 se presentan los resultados obtenidos usando el parámetro óptimo, de-

terminado a través del problema binivel y la solución usando $(0.13, 0.1)$, este parámetro se determinó en el experimento anterior con un proceso de inspección. A simple vista no es notoria la mejoría en la aproximación de la función original por lo que se destaca el uso de medidas como SNR y PSNR para la comparación cuantitativa de las soluciones.

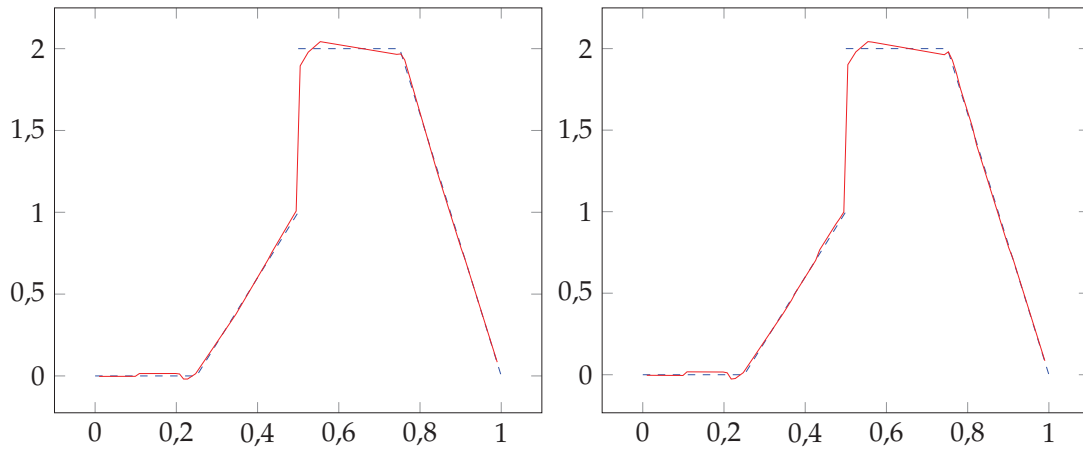


Figura 3.3. A la izquierda, la solución del modelo TGV usando el parámetro $(\alpha/n, \beta/n^2)=(0.13,0.3)$ y a la derecha la solución usando el parámetro óptimo $(\alpha^*/n, \beta^*/n^2)= (0.1559,0.4413)$.

Usando el parámetro $(0.13, 0.1)$ que se determinó con un proceso de inspección, se obtuvieron los valores $SNR=35.4592$ y $PSNR=33.9035$, como observamos en la tabla 3.2, usando el valor del parámetro óptimo $(\alpha^*/n, \beta^*/n^2)$ y comparando las medidas SNR y PSNR el valor obtenido es más alto por lo que es valioso el aporte del problema de optimización de parámetros.

3.7.3. Procesamiento de imágenes

En esta sección, se presentan experimentos sobre la optimización de parámetros del modelo de variación total generalizada para el filtrado de ruido de imágenes. Se deben aclarar: el origen de las imágenes con ruido no uniforme que se emplearon y las medidas que nos permitirán comparar las imágenes resultantes, por lo que, a continuación se presentan las siguientes observaciones.

Al igual que en los experimentos presentados para el modelo de variación total TV, se emplean imágenes que presentan ruido no uniforme, estas se obtuvieron usando simulación de ruido, que consiste en: dada una imagen original se perturba sumando dos variables aleatorias con distribución Gaussiana de media 0 y desviación estándar σ_1 y σ_2 , en distintas localizaciones para obtener ruido no uniforme. Para que sea un reto mayor el filtrado de ruido, se consideran regiones de tamaño a , y el esquema se resume en la Figura 3.4.

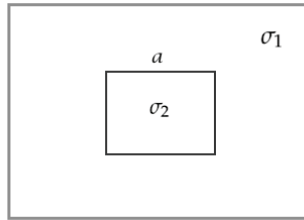


Figura 3.4. Simulación de ruido no uniforme.

Se utilizaron imágenes f_o con distintas características como: brillo, bordes, contenido y textura, lo que permitió analizar el comportamiento de los algoritmos propuestos en este trabajo. La implementación numérica se realizó haciendo uso del software de programación Matlab, en un equipo con procesador de Intel(R) Core(TM) i5 de 2.5 GHz con 4GB de memoria RAM.

Experimento 1

En este experimento se mostrará la importancia de emplear parámetros con dependencia espacial en lugar de parámetros escalares para filtrar imágenes con ruido no uniforme. Para lo cual se resolvió el problema de optimización de parámetros escalares, es decir, $P(\alpha) = \alpha$, $Q(\beta) = \beta$ y para el problema con dependencia espacial $P(\alpha) = \alpha_1x + \alpha_2y + \alpha_3$, $Q(\beta) = \beta_1x + \beta_2y + \beta_3$.

Se considera una imagen con detalles para apreciar visualmente el efecto. La imagen original fue perturbada sumando dos variables aleatorias con distribución Gaussiana de media 0 y desviación estándar $\sigma_1 = 0.002$, $\sigma_2 = 0.02$, en distintas localizaciones para obtener ruido no uniforme como se observa en la Figura 3.5,

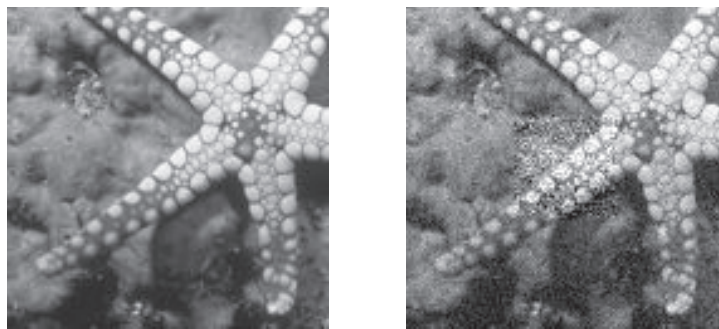


Figura 3.5. Imagen original e imagen con ruido no uniforme.

En la Tabla 3.3 se muestran los resultados obtenidos usando el modelo de variación total generalizada para el filtrado del ruido con parámetros escalares y con dependencia espacial. Se fija como punto inicial $(\alpha_0, \beta_0) = (1/\alpha_{TV}, 1/n * \alpha_{TV})$ en el caso escalar y en el caso espacial $(\alpha_0, \beta_0) = (1/\alpha_{TV}(1, 1, 1), 1/n * \alpha_{TV}(1, 1, 1))$, donde α_{TV} corresponde al valor óptimo

del parámetro del modelo TV determinado en el capítulo anterior. Como se esperaba los resultados obtenidos considerando el parámetro con dependencia espacial son mejores.

Parámetro $P(\alpha)$	Parámetro $Q(\alpha)$	SNR	PSNR	SSIM	Parámetro óptimo α^*/n	Parámetro óptimo β^*/n^2	F. objetivo
α	β	40.3760	26.3156	0.8201	5.729e-2	3.116e-2	9.598e-4
$\alpha_1x + \alpha_2y + \alpha_3$	$\beta_1x + \beta_2y + \beta_3$	40.6241	26.4396	0.8265	(0.0510, 0.0510, 0.0510)	(0.0252, 0.0369, -0.007)	1.091e-3

Tabla 3.3. Comparación del parámetro escalar y con dependencia espacial.

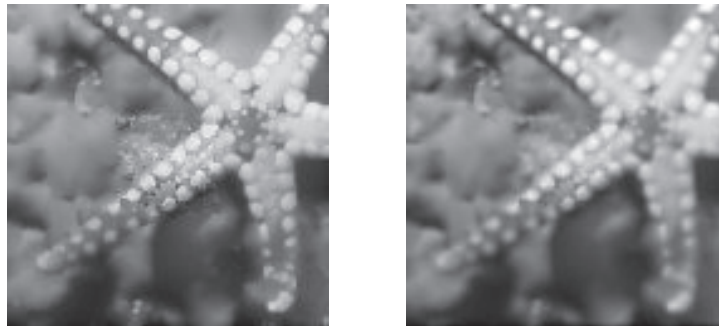


Figura 3.6. A la izquierda imagen solución con α escalar y a la derecha la solución considerando dependencia espacial.

Capítulo 4

Conclusiones

En este trabajo, se propone un enfoque de optimización binivel en espacios funcionales con el objetivo de determinar los parámetros de regularización con dependencia espacial óptimos, presentes en los modelos de filtrado de ruido de imágenes usando variación total TV y variación total generalizada TGV. Se presentó el análisis teórico de este problema de optimización, así como su implementación numérica.

Las dificultades que surgen de inmediato al tratar un problema binivel no lineal en espacios de dimensión infinita son la no diferenciabilidad de la restricción, que corresponde al problema de filtrado de ruido, lo que hace imposible verificar condiciones de calificación para sistemas de Karush-Kuhn-Tucker (KKT). Para tratar el problema de no diferenciabilidad empleamos una regularización de tipo Huber para el término no diferenciable del modelo TV y TGV. Además, para analizar este problema en un espacio de Hilbert incorporamos un término elíptico al modelo original, esto basados en trabajos previos de procesamiento de imágenes donde se muestra la existencia y consistencia de emplear estas técnicas.

Los resultados analíticos que obtuvimos fueron la existencia de soluciones del problema de optimización de parámetros, se demostró la Fréchet diferenciabilidad del operador solución, lo que permite probar la existencia de multiplicadores de Lagrange. Además, la existencia del estado adjunto, que permite obtener una caracterización del gradiente del funcional de costo reducido. Se evidenció que los multiplicadores asociados a las restricciones de positividad son medidas, que computacionalmente hablando son muy costosas de implementar. Para superar esto se propuso introducir la *regularización de Moreau-Yosida*, donde se estableció el sistema de optimalidad asociado al problema regularizado y se mostró que las soluciones de los problemas regularizados convergen hacia la solución del problema original.

Los valores óptimos de los parámetros con dependencia espacial tanto del modelo TV como del TGV, se calculan numéricamente usando un método cuasi-Newton, específicamente el método BFGS, junto con los algoritmos de tipo Newton Semi-Smooth para la resolución del modelo de filtrado de ruido. Para comparar el rendimiento de los modelos, en términos de calidad de imagen, se emplearon las medidas SNR, PSNR y SSIM. El algoritmo propuesto

para la optimización de parámetros puede ser empleado para determinar el peso del parámetro en el modelo de filtrado de ruido de imágenes con ruido uniforme y no uniforme. Este es un proceso complejo, especialmente cuando la intervención del usuario es mínima, pues existen aplicaciones donde el usuario señala la ubicación e intensidad del ruido.

A continuación, se presentan las líneas de trabajo futuras:

- En los experimentos se emplearon imágenes en escala de grises, este estudio se puede extender a imágenes a color o multicanal.
- Se puede utilizar la descomposición de dominios en las imágenes en los algoritmos planteados, si se busca restaurar imágenes con tamaños superiores a los 150×150 píxeles.
- Las aplicaciones mostradas en este trabajo contemplan solo imágenes, sin embargo, se puede extender su aplicación a secuencias de imágenes, es decir, vídeos.
- Para determinar numéricamente el parámetro óptimo se usó como conjunto de entrenamiento la imagen original, el que puede ser reemplazado por un conjunto de imágenes adquiridas por un mismo dispositivo que nos permita aprender el error de este dispositivo.
- Se podría realizar una interfaz gráfica sencilla amigable con el usuario.

Bibliografía

- [1] Hans Wilhelm Alt. *Lineare Funktionalanalysis: Eine anwendungsorientierte Einführung*. Springer-Verlag, 2012.
- [2] Gilles Aubert and Pierre Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations*, volume 147. Springer, 2006.
- [3] Lorenz Biegler, George Biros, Omar Ghattas, Matthias Heinkenschloss, David Keyes, Bani Mallick, Luis Tenorio, Bart van Bloemen Waanders, Karen Willcox, and Youssef Marzouk. *Large-scale inverse problems and quantification of uncertainty*, volume 712. John Wiley & Sons, 2011.
- [4] Kristian Bredies, Yiqiu Dong, and Michael Hintermüller. Spatially dependent regularization parameter selection in total generalized variation models for image restoration. *International Journal of Computer Mathematics*, 90(1):109–123, 2013.
- [5] Kristian Bredies, Karl Kunisch, and Thomas Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.
- [6] Tan Bui-Thanh, Karen Willcox, and Omar Ghattas. Model reduction for large-scale systems with high-dimensional parametric input space. *SIAM Journal on Scientific Computing*, 30(6):3270–3288, 2008.
- [7] Luca Calatroni, Cao Chung, Juan Carlos De Los Reyes, Carola-Bibiane Schönlieb, and Tuomo Valkonen. Bilevel approaches for learning of variational imaging models. *arXiv preprint arXiv:1505.02120*, 2015.
- [8] Luca Calatroni, Juan Carlos De los Reyes, and Carola-Bibiane Schönlieb. Dynamic sampling schemes for optimal noise learning under multiple nonsmooth constraints. In *System Modeling and Optimization*, pages 85–95. Springer Verlag, 2014.
- [9] VC Cao, JC De los Reyes, and CB Schoenlieb. Learning optimal spatially-dependent regularization parameters in total variation image restoration. 2016. To appear.
- [10] Yunjin Chen, Thomas Pock, and Horst Bischof. Learning ℓ_1 -based analysis and synthesis sparsity priors using bi-level optimization. In *Workshop on Analysis Operator Learning vs. Dictionary Learning, NIPS 2012*, 2012.

- [11] Yunjin Chen, René Ranftl, and Thomas Pock. Insights into analysis operator learning: From patch-based sparse models to higher-order mrfs. *Image Processing, IEEE Transactions on*, 2014. to appear.
- [12] Julianne Chung, Malena I Español, and Tuan Nguyen. Optimal regularization parameters for general-form tikhonov regularization. *arXiv preprint arXiv:1407.1911*, 2014.
- [13] Monique Dauge. Neumann and mixed problems on curvilinear polyhedra. *Integral equations and operator theory*, 15(2):227–261, 1992.
- [14] JC De los Reyes, C-B Schönlieb, and T Valkonen. Bilevel parameter learning for higher-order total variation regularisation models. *arXiv preprint arXiv:1508.07243*, 2015.
- [15] Juan Carlos De los Reyes. *Numerical PDE-constrained optimization*. Springer, 2015.
- [16] Juan Carlos De los Reyes and Michael Hintermüller. A duality based semismooth newton framework for solving variational inequalities of the second kind. *Interfaces and Free Boundaries*, 13(4):437–462, 2011.
- [17] Juan Carlos de los Reyes and K Kunisch. A semi-smooth newton method for regularized state-constrained optimal control of the navier-stokes equations. *Computing*, 78(4):287–309, 2006.
- [18] Juan Carlos De Los Reyes and Karl Kunisch. A semi-smooth newton method for control constrained boundary optimal control of the navier–stokes equations. *Nonlinear Analysis: Theory, Methods & Applications*, 62(7):1289–1316, 2005.
- [19] Juan Carlos De los Reyes and Carola-Bibiane Schönlieb. Image denoising: Learning the noise model via nonsmooth pde-constrained optimization. *Inverse Problems & Imaging*, 7(4), 2013.
- [20] Juan Carlos de Los Reyes, Carola-Bibiane Schönlieb, and Tuomo Valkonen. The structure of optimal parameters for image restoration problems. Submitted, 2015.
- [21] Justin Domke. Generic methods for optimization-based modeling. In *International Conference on Artificial Intelligence and Statistics*, pages 318–326, 2012.
- [22] Yiqiu Dong, Michael Hintermüller, and M Monserrat Rincon-Camacho. Automated regularization parameter selection in multi-scale total variation models for image restoration. *Journal of Mathematical Imaging and Vision*, 40(1):82–104, 2011.
- [23] Carl Geiger and Christian Kanzow. *Numerische Verfahren zur Lösung unrestringierter Optimierungsaufgaben*. Springer-Verlag, 2013.
- [24] Rafael C Gonzalez and Richard E Woods. *Digital image processing*, 2002.
- [25] K. Gröger. A $W^{1,p}$ -estimate for solutions to mixed boundary value problems for second order elliptic differential equations. *Math. Ann.*, 283(4):679–687, 1989.

- [26] E Haber and L Tenorio. Learning regularization functionals—a supervised training approach. *Inverse Problems*, 19(3):611, 2003.
- [27] Eldad Haber, Lior Horesh, and Luis Tenorio. Numerical methods for the design of large-scale nonlinear discrete ill-posed inverse problems. *Inverse Problems*, 26(2):025002, 2010.
- [28] Per Christian Hansen. *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*, volume 4. Siam, 1998.
- [29] MICHAEL HINTERMÜLLER. Nonlinear optimization, part i: Unconstrained and box-constrained problems. *Department of Mathematics, Humboldt-University of Berlin, Germany*, 2011.
- [30] Michael Hintermüller, Konstantinos Papafitsoros, and Carlos N Rautenberg. Analytical aspects of spatially adapted total variation regularisation. *arXiv preprint arXiv:1609.01074*, 2016.
- [31] Michael Hintermüller and Georg Stadler. An infeasible primal-dual algorithm for total bounded variation-based inf-convolution-type image restoration. *SIAM Journal on Scientific Computing*, 28(1):1–23, 2006.
- [32] Kazufumi Ito and Karl Kunisch. Semi-smooth newton methods for state-constrained optimal control problems. *Systems & Control Letters*, 50(3):221–228, 2003.
- [33] Johannes Jahn. *Introduction to the theory of nonlinear optimization*. Springer Science & Business Media, 2007.
- [34] MA Krasnoselskii, PP Zabreiko, EL Pustyl'ii, and PE Sobolevskii. Integral operators in spaces of summable functions. *Moscow*, 1976.
- [35] Karl Kunisch and Thomas Pock. A bilevel optimization approach for parameter learning in variational models. *SIAM Journal on Imaging Sciences*, 6(2):938–983, 2013.
- [36] Pascal Thériault Lauzier, Jie Tang, and Guang-Hong Chen. Non-uniform noise spatial distribution in ct myocardial perfusion and a potential solution: statistical image reconstruction. In *SPIE Medical Imaging*, pages 831338–831338. International Society for Optics and Photonics, 2012.
- [37] Pascal Thériault Lauzier, Jie Tang, Michael A Speidel, and Guang-Hong Chen. Noise spatial nonuniformity and the impact of statistical image reconstruction in ct myocardial perfusion imaging. *Medical physics*, 39(7):4079–4092, 2012.
- [38] Wolfgang Ring. Structural properties of solutions to total variation regularization problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 34(4):799–810, 2000.
- [39] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

- [40] Marshall F Tappen. Utilizing variational optimization to learn Markov random fields. In *Computer Vision and Pattern Recognition, 2007. CVPR'07. IEEE Conference on*, pages 1–8. IEEE, 2007.
- [41] Andreĭ Nikolaevich Tikhonov, Vasilii Iakovlevich Arsenin, and Fritz John. *Solutions of ill-posed problems*. Winston Washington, DC, 1977.
- [42] GM Troianiello. *Elliptic equations and obstacle problems. the university series in mathematics*, 1987.
- [43] Fabio Viola, Andrew Fitzgibbon, and Roberto Cipolla. A unifying resolution-independent formulation for early vision. In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pages 494–501. IEEE, 2012.
- [44] J. Zowe and S. Kurcyusz. Regularity and stability for the mathematical programming problem in Banach spaces. *Appl. Math. Optim.*, 5(1):49–62, 1979.