

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

Maribel Kateryn Herrera Terán<sup>†</sup>; Juan Carlos De los Reyes Bueno<sup>‡</sup>

<sup>†</sup>*Escuela Politécnica Nacional, Centro de Modelización Matemática, Quito, Ecuador*

e-mail: kateryn.herrera@epn.edu.ec

<sup>‡</sup>*Escuela Politécnica Nacional, Centro de Modelización Matemática, Quito, Ecuador*

e-mail: juan.delosreyes@epn.edu.ec

---

**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.

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. 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.

**Palabras claves:** Restauración de imágenes, variación total, variación total generalizada, optimización binivel, ruido no uniforme.

---

**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. 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 a 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.

**Keywords:** Denoising, total variation, generalized total variation, nonlinear bilevel optimization, non-uniform noise.

---

## 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* 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.

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 [9], el modelo basado en la ecuación de calor [10], y el de Bredies, Kunish y Pock [1] 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.

## 2. MODELO DE VARIACIÓN TOTAL GENERALIZADA

La regularización de variación total (TV) [9] 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 [8]. 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<sup>2</sup> [1].

Sean  $\Omega$  un subconjunto acotado de  $\mathbb{R}^2$ ,  $f: \Omega \rightarrow \mathbb{R}$  la imagen observada. El modelo TGV<sup>2</sup> para el filtrado de ruido consiste en encontrar  $u \in BV(\Omega)$ , es decir una función con variación acotada, solución del problema de minimización,

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

donde  $\alpha, \beta$  son parámetros positivos y

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

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 modelo de variación total TV está determinado por:

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

Los parámetros de regularización  $\alpha, \beta$  determinan la influencia de los términos en (1), por lo que su valor es crítico. Es así que nuestro objetivo es determinar el valor óptimo de los parámetros de regularización en el caso de imágenes que presentan ruido no uniforme.

## 3. ESTUDIO DEL PARÁMETRO DE REGULARIZACIÓN

Como se analizó antes, el parámetro de regularización 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. La pregunta que surge inmediatamente es: ¿existe un método sistemático para elegir el parámetro de regularización?. 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.

Para evidenciar que es importante conocer la distribución y localización del ruido, a continuación en la Figura 1 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.

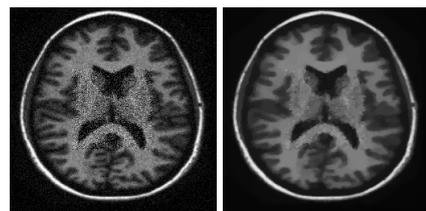


Figura 1. A la izquierda, la imagen con ruido no uniforme y a la derecha la imagen filtrada.

#### 4. PLANTEAMIENTO DEL PROBLEMA

Los parámetros  $\alpha, \beta$  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 (2) y (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 TV y TGV<sup>2</sup> para el filtrado de ruido considerando conjuntos de entrenamiento, fue propuesta en [6] y [2], respectivamente. 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 TV o TGV<sup>2</sup> y será:

$$\min_{(\alpha, \beta) \in \mathbb{R}^{N+M}} F(u_{\alpha, \beta}) \quad (3a)$$

sujeto a

$$u_{\alpha, \beta} \in \arg \min_{u \in BV(\Omega)} \mathcal{J}(u, P(\alpha), Q(\beta)) + \frac{1}{2} \|u - f\|_{L^2(\Omega)}^2 \quad (3b)$$

$$P(\alpha), Q(\beta) \geq 0, \text{ en todo punto de } \Omega \quad (3c)$$

donde

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

$\mathcal{J}(u, P(\alpha), Q(\beta))$  determina el modelo que deseamos emplear TV o TGV<sup>2</sup> y  $f_0$  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.

#### 5. 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 (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  $\mathcal{J}(u, P(\alpha), Q(\beta))$  de la restricción (3b) por su versión regularizada  $\mathcal{J}^{\gamma, \mu}(u, P(\alpha), Q(\beta))$ , con parámetro  $\gamma > 0$  asociado a la regularización de Huber y parámetro  $\mu > 0$  asociado al término elíptico. La regularización de Huber  $C^1 |\cdot|_{\gamma}: \mathbb{R}^n \rightarrow \mathbb{R}$  está definida por,

$$|u|_{\gamma} = \begin{cases} \|u\| + \frac{\gamma}{2} t_1^2 - t_2 + A & \text{si } \mathcal{A}^{\gamma}, \\ B\|u\| + C\|u\|^2 - \frac{\gamma^3}{6} \|u\|^3 + D & \text{si } \mathcal{S}^{\gamma}, \\ \frac{\gamma}{2} \|u\|^2 & \text{si } \mathcal{T}^{\gamma}, \end{cases}$$

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

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

Se definen los conjuntos activo, semi-activo e inactivo como sigue,

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

En el caso del modelo de variación total tenemos que

$$TV_{\alpha}^{\gamma, \mu}(u) = \int_{\Omega} |Du|_{\gamma} + \frac{\mu}{2} \|u\|_{H^1(\Omega)}^2 \quad (4)$$

y para el modelo de variación total generalizada tenemos

$$\begin{aligned} TGV_{\alpha, \beta}^{2, \gamma, \mu}(u) &:= \min_{w \in H^1(\Omega)} \int_{\Omega} P(\alpha) |Du - w|_{\gamma} \\ &+ \int_{\Omega} Q(\beta) |Ew|_{\gamma} + \frac{\mu}{2} \left( \|u\|_{H^1(\Omega)}^2 + \|w\|_{H^1(\Omega)}^2 \right), \quad (5) \end{aligned}$$

donde  $H^1(\Omega) = H^1(\Omega; \mathbb{R}^N)$ .

Considerar  $\mathcal{J}^{\gamma, \mu}(u, P(\alpha), Q(\beta))$  permite que el problema de optimización inferior (3b) se pueda reemplazar por su condición necesaria y suficiente de optimalidad, se obtiene así un problema de optimización con restricciones. Para establecer el sistema de optimalidad de este problema de optimización con restricciones se deben verificar la diferenciabilidad del operador solución, que la ecuación adjunta esta bien planteada y a continuación demostrar la existencia de los multiplicadores de Lagrange para finalmente caracterizar del gradiente del funcional de costo reducido, el cual juega un papel importante en el desarrollo de algoritmos de solución rápida para los problemas de aprendizaje. Estos resultados no se presentan en este escrito pues son extensos, pero pueden ser consultados en el trabajo de titulación.

## 6. PROBLEMA REGULARIZADO

Se demostró que los multiplicadores de Lagrange asociados a la restricción de desigualdad (3c) son medidas, cuya implementación numérica es muy complicada. Como alternativa a esto se introduce la *Regularización de Moreau-Yosida* [7], que consiste en penalizar el funcional de costo con una función  $C^1$  cuando la restricción de positividad no se verifica [3],

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

Esta técnica ha sido ampliamente utilizada en trabajos como [4], [5] donde se consideran problemas de control óptimo con restricciones de desigualdad. Se define la familia de problemas  $\eta$ -penalizados como:

$$\min_{(\alpha, \beta) \in \mathbb{R}^{N+M}} F_\eta(u_{\alpha, \beta}) \quad (6a)$$

sujeto a

$$u_{\alpha, \beta} \in \arg \min_{u \in H^1(\Omega)} J^{\gamma, \mu}(u; P(\alpha), Q(\beta)), \quad (6b)$$

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,

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

En el siguiente teorema, se estudia la convergencia de las soluciones del problema  $\eta$ -regularizado a las soluciones del problema original en el caso TGV.

**TEOREMA 1.** *La sucesión  $\{(u_\eta, w_\eta, \alpha_\eta, \beta_\eta)\}_{\eta>0}$  de soluciones del problema  $\eta$ -regularizado (6) contiene una subsecuencia 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.*

Finalmente, se establece el sistema de optimalidad asociado al problema regularizado (6) y gracias al Teorema 1 se garantiza la convergencia a la solución del problema original.

$$a(y, \Psi) + \int_{\Omega} P(\alpha) h_\gamma(Du - w)(D\phi - \varphi) dx + \int_{\Omega} Q(\beta) h_\gamma(Ew) E\varphi dx + \int_{\Omega} (u - f)\phi dx = 0 \quad (7a)$$

$$F'_\eta(u)\delta_u + \mu(p_1, \delta_u)_{H^1} + \mu(p_2, \delta_w)_{H^1} + \int_{\Omega} P(\alpha) h'_\gamma(Du - w)(D\delta_u - \delta_w)(Dp_1 - p_2) + \int_{\Omega} Q(\beta) h'_\gamma(Ew) E\delta_w E p_2 + \int_{\Omega} p_1 \delta_u = 0 \quad (7b)$$

$$\nabla F(u_{\alpha, \beta}) = \int_{\Omega} P'(\alpha) h_\gamma(Du - w)(Dp_1 - p_2) dx + \int_{\Omega} \min(0, P(\alpha)) P'(\alpha) dx + \int_{\Omega} \min(0, Q(\beta)) Q'(\beta) dx + \int_{\Omega} Q'(\beta) h_\gamma(Ew) E p_2 dx = 0. \quad (7c)$$

En la siguiente sección se abordará la implementación numérica del problema de optimización de parámetros, para lo cual empezaremos por presentar el problema en una dimensión. Los esquemas de discretización de las ecuaciones de estado, adjunta y del gradiente no se incluyen pues son extensos. El algoritmo para la optimización de parámetros se presenta a continuación el algoritmo 1,

---

### Algorithm 1 BFGS optimización de parámetros

---

- 1: Dados  $f, f_0, \mu, \gamma, \eta, \xi$
- 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 2,  $(u_k, w_k)$
- 6: Resolver la ecuación adjunta usando el algoritmo 3,  $p_k = (p_k^1, p_k^2)$
- 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**

$$B_{k+1} = B_k + \frac{(s_k - B_k z_k) s_k^T + s_k (s_k - B_k z_k)^T}{s_k^T z_k} - \frac{(s_k - B_k z_k)^T z_k}{(s_k^T z_k)^2} s_k s_k^T. \quad (8)$$

- 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  $\lambda$  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**  $(\alpha_k, \beta_k)$
-

**Algorithm 2** 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**  $\overline{H}'(u, q, w, r)(\Delta u, \Delta w) = -\overline{H}(u, q, w, r)$ , con

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

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

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

- 5: **Actualizar**

$$\begin{aligned} \Delta q &= -I_{\max}(\max(1, \gamma|\mathbb{G}u - w|)q - \gamma(\mathbb{G}u - w) + H_\gamma \hat{q} \mathbb{G} \Delta u - \gamma \mathbb{G} \Delta u + H_\gamma \hat{q} \Delta w + \gamma \Delta w) \\ \Delta r &= -I_{e_{\max}}(\max(1, \gamma|\mathbb{E}w|)r - \gamma(\mathbb{E}w) + H_{e_\gamma} \hat{r} \mathbb{E} \Delta w - \gamma \mathbb{E} \Delta w). \end{aligned}$$

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

$$k = k + 1$$

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

**Algorithm 3** Ecuación adjunta

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

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

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

## 7. FILTRADO DE RUIDO DE SEÑALES

### 7.1 Comparación modelo TV y TGV

Sabemos que la regularización de variación total (TV) [9] 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 [8]. 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  $(\alpha, \beta)$ , esto para mostrar la importancia del problema de optimización de parámetros.

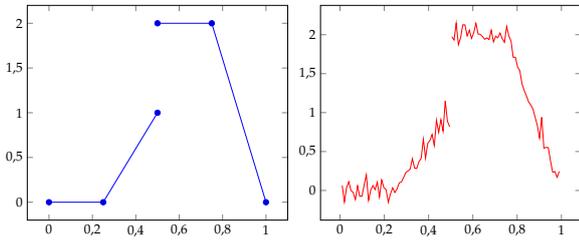


Figura 2. A la izquierda, la función original  $f_0$  y a la derecha la función con ruido  $f$ .

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 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  $\alpha, \beta$  por lo cual se emplea la heurística propuesta en [2], el mejor resultado se obtuvo con el parámetro  $(\alpha, \beta) = (0.13, 0.3)$ .

Modelo TV			Modelo TGV			
Parámetro $\alpha$	Número de iteraciones	SNR	Parámetro $\alpha/n$	Parámetro $\beta/n^2$	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 1. Resultados obtenidos con distintos valores del parámetro de regularización.

A continuación, se presentan en la Figura 3 en color rojo las mejores soluciones usando los modelos de variación total y de variación total generalizada y en color azul la función original.

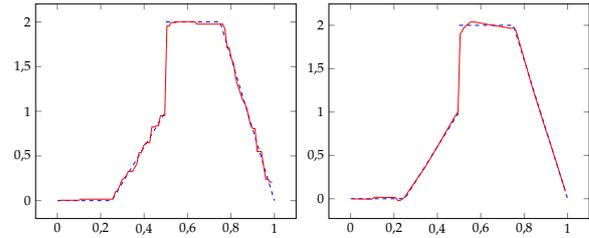


Figura 3. A la izquierda, la solución del modelo TV con  $\alpha = 490$  y a la derecha la solución del modelo TGV con  $(\alpha, \beta) = (0.13, 0.3)$ .

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 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 las características de la función a ser reconstruida y de los detalles que se quieran preservar.

## 7.2 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 presentan en la implementación numérica y entender la naturaleza del problema. Empecemos por realizar la siguiente observación

En el trabajo [2] 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:  $\xi = 1e - 8$  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 2 se muestran los resultados ob-

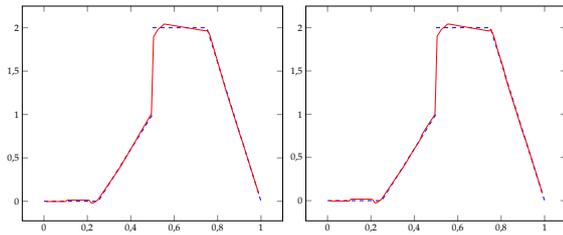
tenidos, además se presentan el número de iteraciones, el valor del funcional objetivo y el valor de las medidas de comparación SNR y PSNR.

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 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 4 se presentan los resultados obtenidos usando el parámetro óptimo, determinado 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.



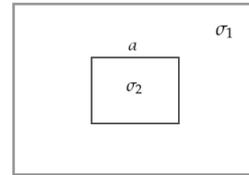
**Figura 4.** A la izquierda, la solución del modelo TGV usando el parámetro  $(\alpha/n, \beta/n^2)=(0.13,0.3)$  y a la derecha la solución usando el parámetro óptimo  $(\alpha^*/n, \beta^*/n^2)=(0.1559,0.4413)$ .

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 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.

## 8. PROCESAMIENTO DE IMÁGENES

En esta sección, se presentan experimentos sobre la optimización de parámetros del modelo de variación total y 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.

Se emplean imágenes que presentan ruido no uniforme, estas se obtuvieron usando simulación de ruido, que consiste en: dada una imagen original se perturbada sumando dos variables aleatorias con distribución Gaussiana de media 0 y desviación estándar  $\sigma_1$  y  $\sigma_2$ , 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 5.



**Figura 5.** Simulación de ruido no uniforme.

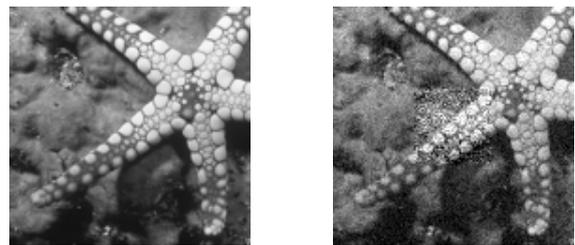
Se utilizaron imágenes  $f_0$  con distintas características como: brillo, bordes, contenido y textura, lo que permitió analizar el comportamiento de los algoritmos propuestos en este trabajo. La implementación numérica se realizó haciendo uso del software de programación Matlab, en un equipo con procesador de Intel(R) Core(TM) i5 de 2.5 GHz con 4GB de memoria RAM.

## 9. EXPERIMENTOS MODELO DE VARIACIÓN TOTAL

### 9.1 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  $P(\alpha) = \alpha_1 x + \alpha_2 y + \alpha_3$ .

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 6,



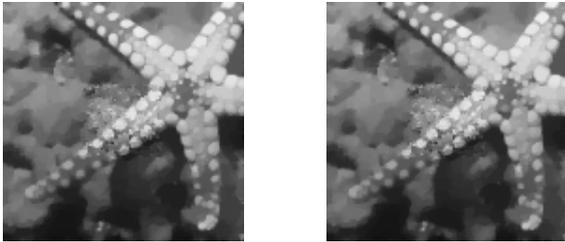
**Figura 6.** Imagen original e imagen con ruido no uniforme.

En la Tabla 3 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 $\alpha^*$	F. objetivo
$\alpha$	39.71	25.987	0.803	2054.0	1.618e-3
$\alpha_1 x + \alpha_2 y + \alpha_3$	40.21	26.234	0.814	(886.59, 734.35, 1554.0)	1.487e-3

**Tabla 3.** Comparación del parámetro escalar y con dependencia espacial.



**Figura 7.** A la izquierda imagen solución con  $\alpha$  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 8,



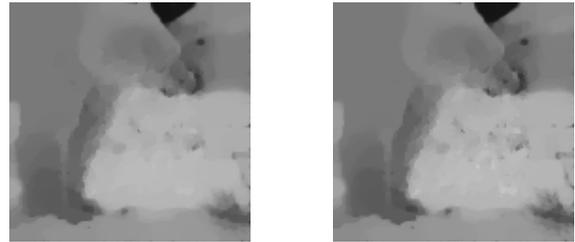
**Figura 8.** Imagen original e imagen con ruido no uniforme.

En la Tabla 4 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 $\alpha^*$	F. objetivo
$\alpha$	49.2651	27.432	0.8088	13618.0	6.979e-3
$\alpha_1 x + \alpha_2 y + \alpha_3$	50.0049	27.802	0.8151	(670.65, 649.36, 797.19)	6.213e-4

**Tabla 4.** 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 9.** A la izquierda imagen solución con  $\alpha$  escalar y a la derecha la solución considerando dependencia espacial.

## 9.2 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 10,



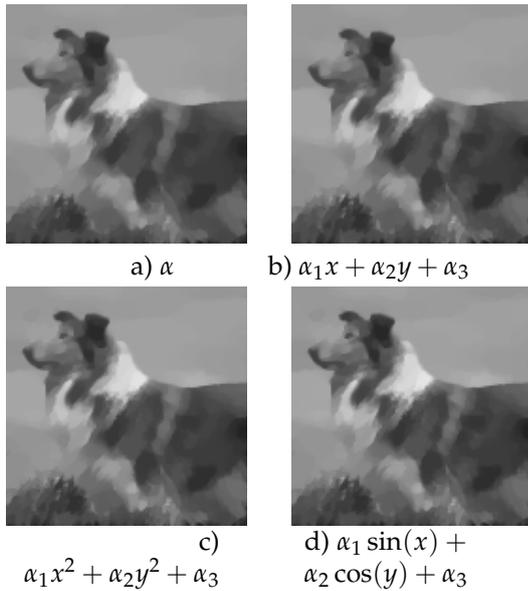
**Figura 10.** Imagen original e imagen con ruido no uniforme.

En la Tabla 5 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 $\alpha^*$	F. objetivo
$\alpha$	44.467	27.736	0.7816	1680.2	1.011e-3
$\alpha_1 x + \alpha_2 y + \alpha_3$	45.849	28.427	0.8091	(530.84, 796.85, 1332.5)	8.928e-4
$\alpha_1 x^2 + \alpha_2 y^2 + \alpha_3$	45.408	28.207	0.7992	(-112.41, 196.46, 1862.0)	1.008e-3
$\alpha_1 \sin(x) + \alpha_2 \cos(y) + \alpha_3$	45.953	28.479	0.8196	(482.87, 888.43, 1209.4)	8.641e-4

**Tabla 5.** Número de iteraciones para distintos valores del parámetro.

En la Figura 11 se comparan las imágenes obtenidas para las diferentes funciones  $P(\cdot)$ .



**Figura 11.** Resultados obtenidos para distintos  $P(\cdot)$ .

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.

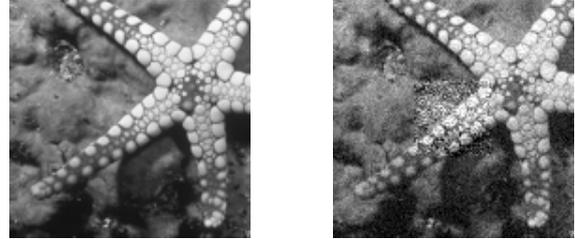
## 10. EXPERIMENTOS MODELO DE VARIACIÓN TOTAL GENERALIZADA

### 10.0.1 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 lo cual se resolvió el problema de optimización de parámetros escalares, es decir,  $P(\alpha) =$

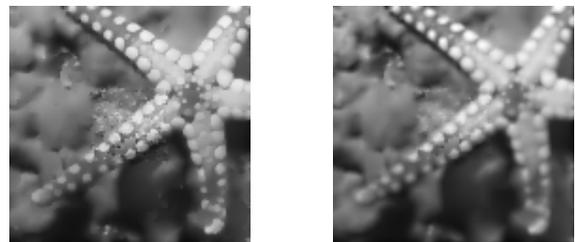
$\alpha$ ,  $Q(\beta) = \beta$  y para el problema con dependencia espacial  $P(\alpha) = \alpha_1 x + \alpha_2 y + \alpha_3$ ,  $Q(\beta) = \beta_1 x + \beta_2 y + \beta_3$ .

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 12,



**Figura 12.** Imagen original e imagen con ruido no uniforme.

En la Tabla 6 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  $\alpha_{TV}$  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.



**Figura 13.** A la izquierda imagen solución con  $\alpha$  escalar y a la derecha la solución considerando dependencia espacial.

## 10. 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.

Parámetro $P(\alpha)$	Parámetro $Q(\alpha)$	SNR	PSNR	SSIM	Parámetro óptimo $\alpha^*/n$	Parámetro óptimo $\beta^*/n^2$	F. objetivo
$\alpha$	$\beta$	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 6.** Comparación del parámetro escalar y con dependencia espacial.

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  $150 \times 150$  píxeles.

- Las aplicaciones mostradas en este trabajo contemplan solo imágenes, sin embargo, se puede extender su aplicación a secuencias de imágenes, es decir, videos.

## 11. REFERENCIAS

- [1] Kristian Bredies, Karl Kunisch, and Thomas Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.
- [2] JC De los Reyes, C-B Schönlieb, and T Valkonen. Bi-level parameter learning for higher-order total variation regularisation models. *arXiv preprint arXiv:1508.07243*, 2015.
- [3] Juan Carlos De los Reyes. *Numerical PDE-constrained optimization*. Springer, 2015.
- [4] 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.
- [5] 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.
- [6] 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.
- [7] Kazufumi Ito and Karl Kunisch. Semi-smooth newton methods for state-constrained optimal control problems. *Systems & Control Letters*, 50(3):221–228, 2003.
- [8] Wolfgang Ring. Structural properties of solutions to total variation regularization problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, 34(4):799–810, 2000.
- [9] Leonid I Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.
- [10] Andreĭ Nikolaevich Tikhonov, Vasilii IAKovlevich Arsenin, and Fritz John. *Solutions of ill-posed problems*. Winston Washington, DC, 1977.