

# Aprendizaje anticausal en problemas inversos y su aplicación a tomografía optoacústica

Anticausal Learning for Inverse Problems and its Application on Optoacoustic Tomography

Matias Vera\*†<sup>1</sup>, Martín G. González \*†, Leonardo Rey Vega\*†

\*Universidad de Buenos Aires, Facultad de Ingeniería  
Paseo Colón 850, C1063ACV, Buenos Aires, Argentina

†Consejo Nacional de Investigaciones Científicas y Técnicas, (CONICET)  
Godoy Cruz 2290, C1425FQB, Buenos Aires, Argentina

<sup>1</sup>[mvera@fi.uba.ar](mailto:mvera@fi.uba.ar)

Recibido: 09/10/25; Aceptado: 09/11/25

**Resumen**— Los algoritmos de inteligencia artificial habitualmente fallan cuando la distribución de los datos se desvía de la utilizada durante el entrenamiento. Esta vulnerabilidad puede ser corregida post-entrenamiento, pero la misma puede requerir una etapa de ajuste computacionalmente pesada y/o una gran necesidad de nuevos datos. En este contexto, la teoría de causalidad suele ser un excelente paradigma para diferenciar los mecanismos propensos a variaciones de los invariantes. Esto permitiría hacer un ajuste solamente sobre el modelo variable, reduciendo la complejidad del problema. Sin embargo, este paradigma está muy poco estudiado en lo referido a problemas inversos, principalmente porque estos problemas son por definición anticausales. En este trabajo se analiza el desempeño y limitaciones de algoritmos básicos en problemas inversos que cumplan el requisito de aprender de forma anticausal. En particular, se estudian estos algoritmos en el contexto de reconstrucción de imágenes en tomografía optoacústica.

**Palabras clave:** problemas inversos; modelos guiados por la física; teoría de causalidad; tomografía optoacústica

**Abstract**— Artificial intelligence algorithms commonly exhibit poor performance when deployed on data whose distribution deviates from the one utilized during the training phase. While this vulnerability can be addressed post-training, doing so may necessitate a computationally intensive fine-tuning process and/or require a significant acquisition of new data. In this context, causality theory presents an excellent paradigm for distinguishing variation-prone mechanisms from invariant ones. This distinction would permit fitting the model exclusively to the variable components, thereby reducing the complexity of the overall problem. However, this paradigm remains under-explored in relation to inverse problems, primarily because such problems are, by their very definition, anticausal. This work undertakes an analysis of the performance and inherent limitations of fundamental algorithms in inverse problems that satisfy the criteria for anticausal learning. Specifically, these algorithms are

investigated within the context of image reconstruction in optoacoustic tomography.

**Keywords:** inverse problems; physics-guided models; causality theory; optoacoustic tomography

## I. INTRODUCCIÓN

Los problemas inversos constituyen una clase particular de tareas cuyo propósito es inferir causas desconocidas a partir de efectos observados [1]. Se clasifican como tareas anticausales porque las mismas buscan estimar causas a partir de efectos, una dirección opuesta al proceso de generación de las variables. Se presentan de manera natural en numerosos contextos científicos y de ingeniería, entre ellos la reconstrucción de imágenes, la exploración geofísica y el diagnóstico médico. Ejemplos clásicos van desde la reconstrucción de imágenes de alta resolución a partir de mediciones degradadas en diagnóstico por imágenes [2] hasta la recuperación de estructuras subterráneas mediante datos sísmicos [3]. Estas tareas suelen ser mal condicionadas, es decir, su solución puede no ser única o extremadamente sensible al ruido en las mediciones, lo que hace imprescindible la incorporación de conocimiento físico previo para caracterizar al problema. Se denominan entonces modelos guiados por la física.

La tomografía optoacústica (TOA) es un método de obtención de imágenes médicas mediante el uso del efecto optoacústico. Un pulso de luz que incide en el tejido biológico blando se esparcirá por el mismo y una parte será absorbida por moléculas presentes en la muestra biológica, conocidas como cromóforos. La energía del cromóforo excitado se convierte luego en calor, que en el marco de un proceso isocórico, termina generando un aumento de presión [4]. Esto se detecta a través de distintos arreglos de sensores de ultrasonido, generando los llamados *sinogramas*: representaciones gráficas de las señales acústicas en función del tiempo medidas por cada detector. La gran cantidad de configuraciones diferentes de medición en esta tarea, así como la presencia de incertidumbres o el conocimiento parcial de los parámetros, pueden dar lugar a algoritmos de reconstrucción específicamente diseñados para una con-

figuración particular que podría no ser la que se utilizará en una situación práctica final, sufriendo un cambio en la distribución de los datos [5].

El *Principio de Mecanismos Causales Independientes* es una hipótesis heurística proveniente de la teoría de causalidad [6]. Este principio postula que el proceso causal generativo de las variables de un sistema se compone de módulos autónomos que no se informan ni influyen entre sí. En un problema de TOA, esto se traduce en entrenar por separado un modelo para la fuente (representados por imágenes) y otro para el proceso de medición de sinogramas a partir de su fuente (problema directo). Una variación en la configuración experimental solo afectaría al segundo modelo, el cuál podría ser corregido con un proceso de *adaptación de dominio* [7] o *aprendizaje por transferencia* [8]. El hecho de corregir solo una parte del modelo, podría aliviar potencialmente el costo computacional y la gran necesidad de datos del nuevo entorno. Este tipo de aprendizaje podría incluso ayudar a definir algoritmos invariantes a cambios de entorno [5], [9], [10].

El presente trabajo se centra en el estudio de algoritmos básicos de aprendizaje anticausal en el marco de los problemas inversos y su aplicación específica a la tomografía optoacústica. Con ello, se busca realizar una prueba de concepto para evaluar la viabilidad y el potencial de la convergencia de estos campos de estudio.

## II. CONCEPTOS BÁSICOS

### A. Problemas Inversos

A pesar de su generalidad, los problemas inversos siguen un marco matemático bastante unificado. El objetivo es recuperar una muestra desconocida  $y \in \mathbb{R}^{d_y}$  distribuida a partir de  $p(y)$ , suponiendo acceso a mediciones  $x \in \mathbb{R}^{d_x}$  y asumiendo un modelo de la forma:

$$X = \mathcal{A}(Y) + V \quad (1)$$

donde  $X$ ,  $Y$  son variables aleatorias representativas de  $x$  y  $y$ , y  $V \sim \mathcal{N}(0, \Sigma_v)$  es un ruido aleatorio gaussiano independiente de  $Y$ . En otras palabras, el modelo define la relación  $X|Y=y \sim \mathcal{N}(\mathcal{A}(y), \Sigma_v)$ . El predictor óptimo en términos de minimizar el *error cuadrático medio* es  $\mathbb{E}[Y|X=x]$ , el cual se busca estimar:

$$\mathbb{E}[Y|X=x] = \arg \min_{f: \mathbb{R}^{d_x} \rightarrow \mathbb{R}^{d_y}} \mathbb{E}[(f(X) - Y)^2] \quad (2)$$

**Observación 1:** Muchas aplicaciones prácticas pueden aproximarse con un modelo lineal no invertible,  $\mathcal{A}(y) = A \cdot y$  donde  $A$  es una matriz de dimensión  $d_x \times d_y$ . Esta matriz  $A$  suele estar fuertemente mal condicionada, lo que vuelve esencial diferentes estrategias de regularización para aproximar su inversión. En este trabajo se tiene en cuenta esta hipótesis, entonces se obtiene el siguiente modelo directo de (1):

$$x = Ay + v \quad (3)$$

### B. Fuera de Distribución

En aplicaciones de sensado es razonable suponer que la distribución de los datos puede cambiar porque las condiciones de adquisición y los factores experimentales rara vez son idénticos entre mediciones. Por ejemplo, en imágenes de TOA, la posición del sensor, la velocidad del sonido o incluso pequeñas variaciones en la muestra pueden alterar la relación entre la fuente imagen original y la observación  $x$  (sinograma).

Sea  $p^e(x, y)$  la distribución de los datos, denotamos  $e \in \mathcal{E}$  a las posibles variaciones que puede sufrir. La cuestión clave a abordar es la representación de los entornos en el contexto de un problema inverso. Un ejemplo paradigmático de problema inverso en ingeniería involucra un escenario en el cual  $y$  representa una variable física a sensar, y  $x$  corresponde a su medición indirecta. Si se supone que las posibles variaciones se deben a cambios en las condiciones experimentales donde se lleva a cabo el sensado, es razonable suponer que la distribución  $p(y)$  permanece fija mientras que el componente dependiente del entorno queda capturado por  $p^e(x|y)$ .

En la teoría de la causalidad, este tipo de problema se conoce como *aprendizaje anticausal*, dado que la meta es predecir la causa a partir del efecto. En estos problemas, el enfoque recomendado a seguir es el que se describe a continuación [6]:

**Observación 2:**  $p^e(x|y)$  representa el mecanismo causal que genera  $X$  a partir de  $Y$ , y es independiente de la distribución de la causa,  $p(y)$ . Por otro lado,  $p^e(y|x)$  es sensible al cambio en la distribución de  $p(y)$ . Por lo tanto, en términos generales, al estimar  $p^e(y|x)$  conviene modelar  $p^e(x|y)$  y  $p(y)$  por separado y luego construir  $p^e(y|x)$  usando la regla de Bayes.

En este contexto, se estudian diversas técnicas para estimar  $\mathbb{E}[Y|X=x]$ , con el objetivo de desacoplar el modelado de  $p(x|y)$  de  $p(y)$ , aprovechando el conocimiento de la distribución  $X|Y=y \sim \mathcal{N}(Ay, \Sigma_v)$ . La estimación se efectuará en dos etapas:

1. **Modelo directo:** estimar los parámetros  $(A, \Sigma_v)$  a partir de un conjunto de datos aleatorio  $\{(X_i, Y_i)\}_{i=1}^n$ . En este trabajo, para el modelo directo se supondrá: 1) una matriz de covarianza esférica para el ruido  $\Sigma_v = \sigma_v^2 \cdot \mathbf{I}$ , donde  $\sigma_v^2$  es la estimación empírica de la varianza; y 2) la matriz  $A$  será determinada con la metodología estándar de TOA (la cuál será explicada a continuación).
2. **Componente invariante:** desarrollar un procedimiento capaz de realizar este modelado desacoplado de  $p(x|y)$  (conocido en este paso) y  $p(y)$  (desconocido), sin comprometer significativamente el rendimiento en comparación con los métodos clásicos (diseñados para un único entorno). En esta etapa se utiliza un conjunto de datos simple de objetivos  $\{Y_i\}_{i=1}^n$ . Nos enfocaremos en algoritmos no neuronales, con el fin de compararlos con métodos clásicos como la *proyección lineal inversa* (LBP, por sus siglas en inglés).

### C. Tomografía Optoacústica

Es bien sabido que, tras la excitación de una muestra biológica por un pulso electromagnético  $\delta(t)$ , la presión acústica  $p(\mathbf{r}, t)$  en la posición  $\mathbf{r} \in \mathbb{R}^3$  y tiempo  $t$  satisface la ecuación diferencial [11]:

$$\left( \frac{\partial^2}{\partial t^2} - v_s^2 \nabla^2 \right) p(\mathbf{r}, t) = 0 \quad (4)$$

con las condiciones iniciales:

$$p(\mathbf{r}, 0) = p_0(\mathbf{r}), \quad \frac{\partial p}{\partial t}(\mathbf{r}, 0) = 0 \quad (5)$$

donde  $p_0(\mathbf{r})$  es la presión optoacústica inicial y  $v_s$  representa la velocidad del sonido en el medio, el cual se supone homogéneo y sin absorción acústica. Bajo la hipótesis usual de confinamiento térmico y acústico [12], es decir, cuando la duración del pulso láser es lo suficientemente corta como para que se pueda despreciar la conducción de calor y la propagación acústica hacia regiones vecinas de la zona iluminada, la presión inducida inicialmente  $p_0(\mathbf{r})$  es proporcional a la densidad de energía óptica total absorbida. Usando la formalización de la función de Green, la presión recibida por un detector puntual ideal en la posición  $\mathbf{r}_d$  puede escribirse como:

$$p_d(\mathbf{r}_d, t) = \frac{1}{4\pi v_s^2} \frac{\partial}{\partial t} \iiint_V p_0(\mathbf{r}) \frac{\delta(t - |\mathbf{r}_d - \mathbf{r}|/v_s)}{|\mathbf{r}_d - \mathbf{r}|} d\mathbf{r} \quad (6)$$

El objetivo del problema inverso en TOA es reconstruir  $p_0(\mathbf{r})$  a partir del sinograma  $p_d(\mathbf{r}_d, t)$  medido en varias posiciones  $\mathbf{r}_d$ , que típicamente se encuentran sobre una superficie  $S$  que contiene el volumen de interés [13].

Varios enfoques, como los algoritmos de retroproyección [14], [15], son de los más populares y utilizados en el problema de reconstrucción de imágenes en TOA. Dichos métodos proporcionan fórmulas de reconstrucción en forma cerrada en términos de las señales detectadas sobre la superficie de detección. Sin embargo, estos métodos suponen que los detectores son puntuales, sin limitaciones de ancho de banda y con respuesta angular isotrópica [16]. En la práctica, los transductores tienen tamaño finito, ancho de banda limitado y su respuesta espacial no es constante [17]. Además, las señales detectadas son ruidosas. Estas desviaciones del escenario ideal supuesto por las fórmulas exactas de reconstrucción pueden generar artefactos e imágenes distorsionadas.

Un enfoque diferente pero relacionado al problema de reconstrucción está dado por los algoritmos basados en matrices [18]. En esta técnica, la solución directa de (6) se discretiza. Como resultado, se obtiene la ecuación matricial (3) que se usa para resolver el problema inverso, donde  $x$  es un vector columna que representa las presiones medidas en un conjunto de posiciones de detectores  $\mathbf{r}_{dl}$  ( $l = 1 \dots N_d$ ) y en instantes de tiempo  $t_k$  ( $k = 1 \dots N_t$ );  $y$  es un vector columna que representa los valores de la presión acústica inicial, y que típicamente se denomina la imagen de referencia;  $y$  y  $A$  es la matriz del modelo. El  $j$ -ésimo elemento ( $j = 1 \dots N$ ) en  $y$  contiene el valor promedio de la presión inicial dentro de un elemento de volumen de tamaño  $\Delta V$  en la posición  $\mathbf{r}_j$ . Una de las

ventajas de este enfoque es que cualquier efecto lineal en el sistema puede ser considerado fácilmente (por ejemplo, la respuesta espacial y temporal de los sensores) [19]. Una vez establecida la formulación discreta, el problema inverso se reduce al problema algebraico de invertir (3). La matriz  $A$  puede escribirse como la multiplicación de dos matrices  $A^{oa} A^s$ , donde  $A^s$  representa la función de respuesta del sistema de imagen para un sensor puntual ideal y  $A^{oa}$  es la forma matricial de un operador de derivada temporal. La matriz  $A^s$  se define como [20]:

$$A_{lkj}^s = \frac{1}{4\pi v_s^2} \frac{\Delta V}{\Delta t^2} \frac{d(t_k, \mathbf{r}_j, \mathbf{r}_{dl})}{|\mathbf{r}_{dl} - \mathbf{r}_j|} \quad (7)$$

$$d(t_k, \mathbf{r}_j, \mathbf{r}_{dl}) = \begin{cases} 1 & \text{si } |t_k - \frac{|\mathbf{r}_{dl} - \mathbf{r}_j|}{v_s}| < \Delta t/2 \\ 0 & \text{en otro caso} \end{cases} \quad (8)$$

donde  $\Delta t$  es el paso temporal en el que se muestrean las señales  $p_d(\mathbf{r}_d, t)$ . No es difícil ver que (7) constituye una discretización del integrando en (6), mientras que (8) indica el instante de tiempo en que el efecto de la presión inicial en la posición  $\mathbf{r}_j$  es capturado por el sensor  $\mathbf{r}_{dl}$ . En el caso de un detector de tamaño finito, la respuesta impulsiva espacial del sensor se tiene en cuenta dividiendo el área del sensor en elementos superficiales (tratados como detectores puntuales) que luego se suman [20], [21].

*Observación 3:* Es razonable suponer que, en una aplicación real, la matriz del modelo no se conoce completamente. En este trabajo, consideraremos que la diferencia entre ambas reside en la incertidumbre de la posición del sensor y de la velocidad del sonido al calcularla.

### III. MODELOS ANTICAUSALES

Nos enfocamos en modelos que definen una función de aproximación  $\hat{y}(x)$  basada en los parámetros  $(A, \sigma_v^2)$  y un conjunto de datos objetivo  $\{Y_i\}_{i=1}^n$ . Estos modelos buscan abordar los desafíos de los problemas inversos aprovechando representaciones anticausales desacopladas y estrategias computacionales eficientes.

#### A. Proyección Lineal Inversa (LBP)

El problema de inversión puede formularse usando un criterio cuadrático combinado con un término de regularización de Tikhonov:

$$\hat{y}_{TIK}(x) = \min_{y \in \mathbb{R}^{d_y}} \|Ay - x\|^2 + \lambda \|y\|^2 \quad (9)$$

donde  $\lambda \geq 0$  es un parámetro de regularización que mejora la estabilidad del problema inverso, el cual suele ser mal condicionado. Además, este término de regularización mitiga los efectos del ruido en las señales medidas. La solución al problema regularizado de Tikhonov puede derivarse analíticamente como [22]:

$$\hat{y}_{TIK}(x) = (A^T A + \lambda I)^{-1} A^T x \quad (10)$$

Sin embargo, calcular esta inversa puede ser computacionalmente costoso, particularmente en problemas a gran escala. Para abordar esto, una simplificación común es

considerar valores grandes de  $\lambda$ . En el régimen asintótico cuando  $\lambda \rightarrow \infty$ , la solución de Tikhonov se vuelve proporcional al operador adjunto  $\hat{y}_{\text{TIK}}(x) \propto A^T x$ , y esta aproximación conduce a una solución computacionalmente eficiente conocida como *proyección lineal inversa* (LBP) [23]:

$$\hat{y}_{\text{LBP}}(x) = A^T x \quad (11)$$

La solución LBP proporciona un método sencillo y computacionalmente económico para obtener una reconstrucción inicial de la imagen. Sin embargo, es importante notar que las reconstrucciones basadas en LBP suelen exhibir limitaciones y artefactos, particularmente en escenarios de visión limitada [24]. A pesar de estas desventajas, LBP sigue siendo ampliamente utilizada debido a su simplicidad y facilidad de implementación. Cabe destacar que esta aproximación depende únicamente de la matriz  $A$ , independiente de  $\sigma_v^2$  y del conjunto de datos.

#### B. Estimación Monte Carlo

El cálculo de  $\mathbb{E}[Y|X = x]$  mediante la descomposición causal  $p(x, y) = p(y)p(x|y)$  para este modelo puede escribirse como:

$$\mathbb{E}[Y|X = x] = \int_{\mathbb{R}^d y} y \frac{p(x|y)p(y)}{p(x)} dy \quad (12)$$

$$= \int_{\mathbb{R}^d y} y e^{-\frac{1}{2}(x-Ay)^T \Sigma_v^{-1}(x-Ay)} p(y) dy \quad (13)$$

Una estimación Monte Carlo puede realizarse de manera directa como:

$$\hat{y}_{\text{MC}}(x) = \frac{\sum_{i=1}^n y_i e^{-\frac{1}{2}(x-Ay_i)^T \Sigma_v^{-1}(x-Ay_i)}}{\sum_{i=1}^n e^{-\frac{1}{2}(x-Ay_i)^T \Sigma_v^{-1}(x-Ay_i)}}$$

#### C. Modelo de Mezcla Gaussiana (GMM)

El enfoque propuesto consiste en aprender  $p(y)$  mediante un Modelo de Mezcla Gaussiana (GMM) entrenado a través del algoritmo de Expectativa-Maximización (EM) con  $m$  gaussianas diagonales. Para integrar este modelo con  $X|Y = y$ , se emplea un marco de análisis de factores.

El GMM supone la existencia de una variable *mezcladora* categórica  $K \sim \text{Cat}(\omega_1, \dots, \omega_m)$  cuyo vínculo con el resto de las variables está definido por la cadena de Markov  $K \rightarrow Y \rightarrow X$  (donde las flechas definen la relación de causalidad supuesta). Es decir, el modelo supone por un lado la distribución  $Y|K = k \sim \mathcal{N}(\mu_k, \Lambda_k)$  (con  $\Lambda_k$  una matriz diagonal) y por el otro  $X|Y = y \sim \mathcal{N}(Ay, \sigma_v^2 \cdot \mathbf{I})$ .

El algoritmo EM entrena  $\{\omega_k, \mu_k, \Lambda_k\}_{k=1}^m$  a partir de un conjunto de datos  $\{Y_i\}_{i=1}^n$ . Tras el entrenamiento, la inferencia puede realizarse con el siguiente lema.

**Lema 1:** En el modelo descrito anteriormente, la esperanza condicional puede calcularse como:

$$\mathbb{E}[Y|X = x] = \sum_{k=1}^m \mathbb{P}(K = k|X = x) \mathbb{E}[Y|X = x, K = k] \quad (14)$$

donde

$$\mathbb{E}[Y|X = x, K = k] \quad (15)$$

$$= \mu_k + \Lambda_k A^T (\sigma_v^2 \mathbf{I} + A \Lambda_k A^T)^{-1} (x - A \mu_k)$$

y

$$\mathbb{P}(K = k|X = x) = \frac{\omega_k \cdot \mathcal{N}_x(A\mu_k, \sigma_v^2 \mathbf{I} + A\Lambda_k A^T)}{\sum_{l=1}^m \omega_l \cdot \mathcal{N}_x(A\mu_l, \sigma_v^2 \mathbf{I} + A\Lambda_l A^T)} \quad (16)$$

con  $\mathcal{N}_x(\mu, \Sigma)$  la densidad de probabilidad asociada a una distribución normal de media  $\mu$  y covarianza  $\Sigma$  evaluada en  $x$ .

La demostración puede verse en el Apéndice A. Dentro de este marco, definimos el estimador  $\hat{y}_{\text{GMM}}(x) = \mathbb{E}[Y|X = x]$ , calculado usando el Lema 1.

#### D. Fórmula de Tweedie

El operador de esperanza condicional es una herramienta fundamental en numerosos campos que dependen del análisis estadístico. Existen diversas identidades derivadas que establecen relaciones entre la esperanza condicional y otras cantidades estadísticas, como la varianza condicional. Entre estas, una de las identidades más relevantes es la fórmula de Tweedie [25], que proporciona un método para calcular la esperanza condicional a través de la medida marginal. En este contexto, la fórmula de Tweedie se formaliza en el siguiente lema.

**Lema 2:** La esperanza condicional  $\mathbb{E}[Y|X = x]$  satisface la siguiente identidad:

$$A \cdot \mathbb{E}[Y|X = x] = x + \sigma_v^2 \cdot \psi(x) \quad (17)$$

donde  $\psi(x) = \nabla_x \log p(x)$  denota la función *score*, con  $\nabla_x$  representando el gradiente respecto de  $x$ . La función *score* también puede expresarse como:

$$\psi(x) = \nabla_x \log \mathbb{E} \left[ e^{-\frac{1}{2\sigma_v^2} \|x - Ay\|^2} \right] \quad (18)$$

La demostración se encuentra en el Apéndice B. En este contexto, definimos la aproximación de Tweedie como:

$$\hat{y}_{\text{TW}}(x) = A^T x + \sigma_v^2 A^T \cdot \psi(x) \quad (19)$$

Dentro de este marco, la función de Tweedie puede verse como una generalización de LBP (11), donde la desviación respecto de la formulación estándar de LBP está gobernada por el parámetro  $\sigma_v^2$ , estimado en el *modelo directo*. La aproximación de la función *score*, sin embargo, sigue siendo un problema abierto. Nosotros utilizaremos una aproximación de Taylor de primer orden de la siguiente manera:

$$\begin{aligned} e^{-\frac{1}{2\sigma_v^2} \|x - Ay\|^2} &\approx 1 - \frac{1}{2\sigma_v^2} (x^T x - 2y^T A^T x + y^T A^T A y) \\ &= 1 - \frac{1}{2\sigma_v^2} (x^T x - 2y^T A^T x + \text{Tr}(A^T A y y^T)) \end{aligned} \quad (20)$$

donde  $\text{Tr}(\cdot)$  es la *traza* de la matriz. La aproximación de la

función *score* se da entonces por:

$$\begin{aligned} \psi(x) &= \frac{\nabla_x \mathbb{E} \left[ e^{-\frac{1}{2\sigma_v^2} \|x - AY\|^2} \right]}{\mathbb{E} \left[ e^{-\frac{1}{2\sigma_v^2} \|x - AY\|^2} \right]} \\ &\approx \frac{-\frac{1}{\sigma_v^2} (x - A\mathbb{E}[Y])}{1 - \frac{1}{2\sigma_v^2} (x^T x - 2\mathbb{E}[Y]^T A^T x + \text{Tr}(A^T A \mathbb{E}[YY^T]))} \end{aligned} \quad (21)$$

Definimos la función de aproximación de Taylor como:

$$\begin{aligned} \hat{y}_{\text{TAY}}(x) &= \\ &A^T x - \frac{(A^T x - A^T A \mathbb{E}[Y])}{1 - \frac{1}{2\sigma_v^2} (x^T x - 2\mathbb{E}[Y]^T A^T x + \text{Tr}(A^T A \mathbb{E}[YY^T]))} \end{aligned} \quad (22)$$

Cuando  $\sigma_v^2 \rightarrow 0$ , esta aproximación converge a la solución LBP  $\hat{y}_{\text{TAY}}(x) = \hat{y}_{\text{LBP}}(x) = A^T x$ , pero, por el contrario, cuando  $\sigma_v^2 \rightarrow \infty$ , la aproximación se vuelve constante  $\hat{y}_{\text{TAY}}(x) = A^T A \mathbb{E}[Y]$ . En este contexto,  $\sigma_v^2$  sirve como medida de la calidad de la matriz  $A$ : si el ruido es pequeño, la reconstrucción se basa principalmente en  $A$ ; si el ruido es significativo, la reconstrucción también incorpora información del conjunto de datos a través de  $\mathbb{E}[Y]$  y  $\mathbb{E}[YY^T]$  (estimados de forma empírica).

#### IV. RESULTADOS EXPERIMENTALES

Los siguientes ejemplos se realizaron en una aplicación de TOA, donde  $x$  es el sinograma e  $y$  la imagen original. Utilizamos un conjunto de entrenamiento de 5652 imágenes  $\{y_i\}_{i=1}^{5652}$ . Dicho conjunto está conformado por datos sintéticos generados a partir de imágenes de vasculatura retiniana obtenidas de bases de datos públicas [26]–[30]. Tanto la matriz  $A$  como la naturaleza del ruido aditivo se conocerán parcialmente. Las incertidumbres en la posición del sensor y la velocidad del sonido en la matriz  $A$  son del 0,5 %. Las imágenes son creadas con ruido blanco gaussiano de varianza aleatoria, elegida para que el valor de la SNR del sinograma varíe uniformemente entre 20 dB y 80 dB ( $\sigma_v^2$  será estimado empíricamente y se considerará fijo).

Por último, se aplica *aumento de datos* para incrementar el conjunto de entrenamiento a 50000 muestras utilizando datos sintéticos generados por el modelo de difusión presentado en [31].

##### A. Estimación Monte-Carlo

El primer experimento utiliza la estimación Monte-Carlo  $\hat{y}_{\text{MC}}(x)$ . Los resultados fueron interesantes: aproximadamente el 70 % de las estimaciones resultaron ser solo ruido, mientras que el 30 % restante tuvo un desempeño muy bueno. Los resultados exitosos se muestran en la Fig. 1. Al aplicar aumento de datos en este experimento, los resultados fueron desalentadores: ninguna de las reconstrucciones fue exitosa, observándose ruido en todos los casos. Se concluye entonces que este método es muy sensible a la falta de muestras en un entorno del punto de inferencia. El espacio imagen posee una complejidad inherente tal, que la cantidad de las muestras con las que se cuenta no alcanza para hacer converger a la ley de los grandes números.

##### B. Modelo de Mezcla de Gaussianas

El segundo experimento utiliza el GMM previamente descrito  $\hat{y}_{\text{GMM}}(x)$  con  $m = 100$ . Una verificación inicial de la calidad de aprendizaje del GMM es usarlo como método generativo. Sin aumento de datos, las imágenes generadas fueron muy ruidosas y presentaban pocas venas. Sin embargo, con aumento de datos, el desempeño fue satisfactorio. Los ejemplos se presentan en la Fig. 2. Aunque las imágenes generadas parecen razonables, el desempeño del GMM en reconstrucción de imágenes no fue satisfactoria, produciendo salidas ruidosas como puede verse en la segunda fila de la Fig. 3.

Un análisis de los parámetros optimizados revela que las gaussianas exhiben covarianzas relativamente estrechas. Este resultado implica que la región del espacio latente correctamente caracterizada se concentra principalmente en torno a las  $m$  medias de las gaussianas. Este fenómeno podría atribuirse a que el aumento de datos sesgó el proceso de aprendizaje hacia las muestras reales disponibles. Es plausible que la estrategia de aumento de datos empleada no sea la más adecuada para la estimación anticausal propuesta, considerando, además, su efecto perjudicial previamente observado en la evaluación mediante Monte Carlo. No obstante, dado que el algoritmo no lograba converger a los patrones esperados sin la aplicación del aumento de datos, se infiere una doble limitación: por un lado, el número de muestras disponibles es insuficiente para el número de gaussianas utilizadas (justificando el uso del aumento de datos); por otro lado, la complejidad intrínseca de las imágenes impide una reducción del tamaño del conjunto de datos. En esencia, nos enfrentamos a un compromiso metodológico o una limitación fundamental del conjunto de datos.

##### C. Fórmula de Tweedie

El tercer experimento utiliza la aproximación de Taylor de la fórmula de Tweedie  $\hat{y}_{\text{TAY}}(x)$ . La Fig. 3 compara el desempeño de este método con la aproximación LBP, mostrando resultados similares. En la Fig. 4, se puede observar el desempeño de  $\hat{y}_{\text{TAY}}(x)$  en función de  $\sigma_v^2$  para diferentes métricas populares: Índice de Similitud Estructural (SSIM), Correlación de Pearson (PC), Error Cuadrático Medio (RMSE) y Relación Señal-Ruido de Pico (PSNR) [5]. Mientras que SSIM, RMSE y PSNR son óptimos cuando  $\sigma_v^2 = 0$  (aproximación LBP), en términos de PC existe una destacable ventaja al usar la aproximación de Taylor.

#### V. CONCLUSIONES Y TRABAJOS FUTUROS

En el presente trabajo se propuso y se evaluó una metodología para la resolución de problemas inversos desde una perspectiva anticausal, la cual facilita la adaptación a posteriores cambios en la distribución de los datos (cambios de entorno). Se exploraron métodos básicos, incluyendo LBP, Monte Carlo, GMM y Tweedie. Si bien LBP demostró ser el método más consistentemente robusto, fue superado en rendimiento en escenarios específicos. El método de Monte Carlo mostró la capacidad de alcanzar una alta precisión en regiones puntuales del espacio de solución, y el estimador de Tweedie evidenció mejoras significativas con respecto a métrica PC.

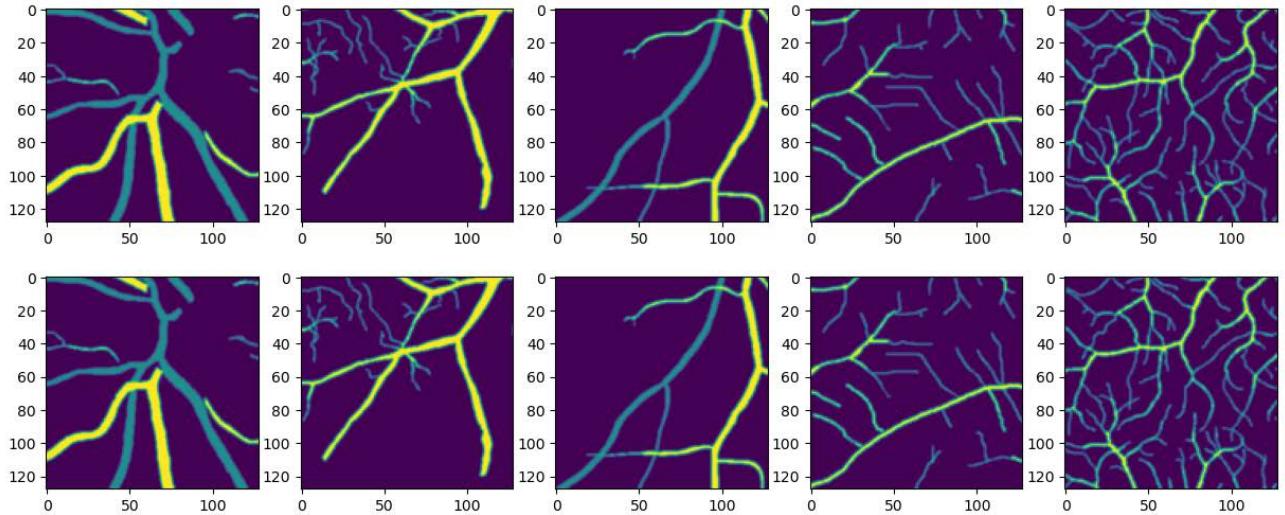


Figura 1: Ejemplos de las imágenes reconstruidas con Monte-Carlo para los casos donde el método funciona. La primera fila muestra la imagen original y la segunda la reconstrucción Monte-Carlo.

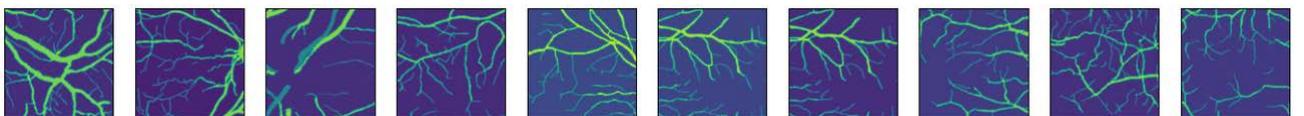


Figura 2: Desempeño del GMM con aumento de datos como modelo generativo.

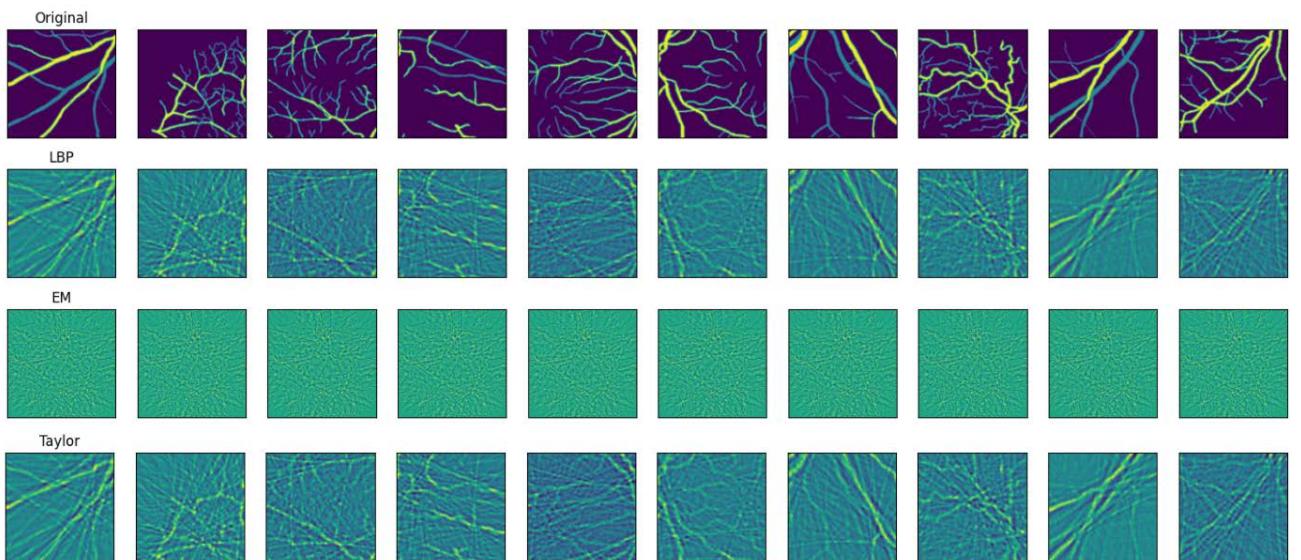
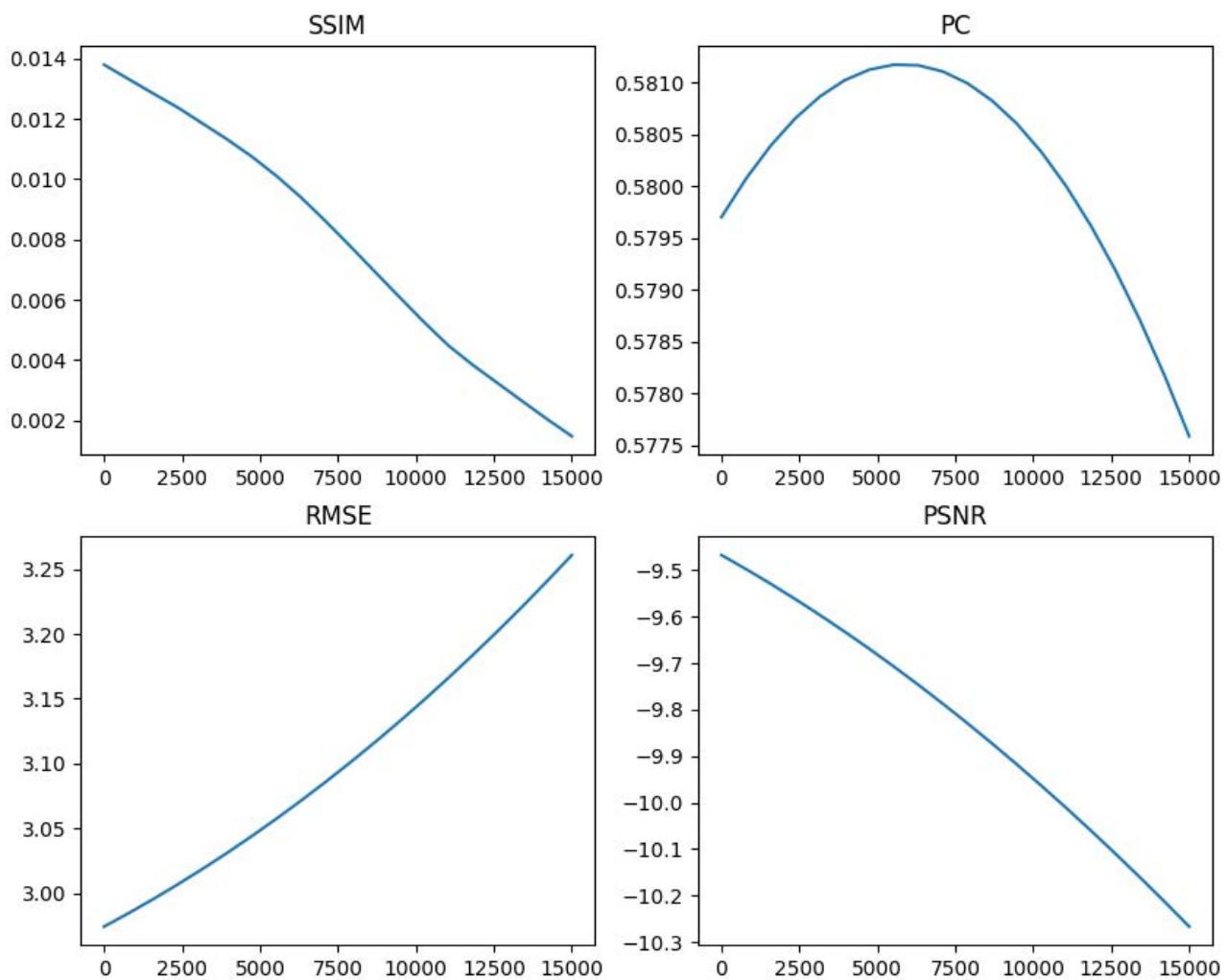


Figura 3: Ejemplos de imágenes reconstruidas. La primera fila muestra la imagen original, la segunda la reconstrucción LBP, la tercera GMM (algoritmo EM) entrenada con aumento de datos, y la cuarta utilizando aproximación de Taylor de la fórmula de Tweedie.

La descomposición anticausal  $p(x|y) - p(y)$ , presenta muchas potencialidades una vez aprendidas, pero evidencia un aumento en la dificultad de dicho aprendizaje. Mientras que todos los métodos utilizaron el mismo modelado de  $p(x|y)$  (caracterizados por  $A$  y  $\sigma_v^2$ ), los de mejor desempeño fueron LBP (el cuál evita modelar  $p(y)$ ) y el estimador Tweedie (solamente modela los momentos  $\mathbb{E}[Y]$  y  $\mathbb{E}[YY^T]$ ), dando a entender que la dificultad está en el aprendizaje del espacio de las imágenes. Esto se corrobora al analizar el pobre desempeño del aumento de datos utilizado, el cual ya había demostrado buen desempeño en configuraciones

estándar [31]. Es evidente entonces que la complejidad de aprendizaje en el espacio de las imágenes es superior al del espacio de sinogramas.

El análisis aquí presentado establece una base inicial en esta área de estudio, dejando un amplio espectro para la investigación futura. Una línea de exploración pendiente consiste en optimizar la aproximación de Tweedie para garantizar una utilización más eficiente de los datos, superando la información provista únicamente por los momentos estadísticos. Adicionalmente, sería de gran relevancia incorporar una métrica de desempeño que esté directamente aco-

Figura 4: Figuras de mérito (SSIM, PC, RMSE y PSNR) en función de  $\sigma_v^2$  usando la aproximación de Taylor.

plada con la adaptación efectiva al entorno (trascendiendo la mera potencialidad teórica) para así cuantificar su beneficio real. Finalmente, resulta crucial investigar la adaptación de este enfoque anticausal a arquitecturas avanzadas, como las redes neuronales o los modelos difusivos. Este esfuerzo implicaría no solo la modelización del problema directo, sino también la incorporación explícita de la identidad de Bayes, como se demostró en trabajos previos (por ejemplo, (13)). Sin embargo, la definición de la estrategia de implementación óptima requiere exploración adicional.

#### AGRADECIMIENTOS

Este trabajo fue financiado por la Universidad de Buenos Aires (UBACYT 20020190100032BA), CONICET (PIP 11220200101826CO) y la Agencia I+D+i (PICT 2020-01336).

#### REFERENCIAS

- [1] G. Daras, H. Chung, C.-H. Lai, Y. Mitsufuji, J. C. Ye, P. Milanfar, A. Dimakis, and M. Delbracio, “A survey on diffusion models for inverse problems,” *CoRR*, 2024, arXiv e-prints: 2410.00083.
- [2] Y. Song, L. Shen, L. Xing, and S. Ermon, “Solving inverse problems in medical imaging with score-based generative models,” in *International Conference on Learning Representations (ICLR)*, 2022.
- [3] P. Lailly, “The seismic inverse problem as a sequence of before stack migrations,” in *Conference on Inverse Scattering, Theory and application*, 1983, pp. 206–220.
- [4] M. Xu and L. Wang, “Photoacoustic imaging in biomedicine,” *Rev. Sci. Instrum.*, vol. 77, p. 041101, 2006.
- [5] M. Vera, M. G. González, and L. R. Vega, “Invariant representations in deep learning for optoacoustic imaging,” *Review of Scientific Instruments*, vol. 94, no. 5, p. 054904, 05 2023. [Online]. Available: <https://doi.org/10.1063/5.0139286>
- [6] B. Schölkopf, D. Janzing, J. Peters, E. Sgouritsa, K. Zhang, and J. Mooij, “On causal and anticausal learning,” in *International Conference on Machine Learning (ICML)*, ser. ICML’12. Madison, WI, USA: Omnipress, 2012, p. 459–466.
- [7] H. Daumé III, “Frustratingly easy domain adaptation,” in *Annual Meeting of the Association of Computational Linguistics (ACL)*. Prague, Czech Republic: Association for Computational Linguistics, 2007, pp. 256–263.
- [8] M. Rojas-Carulla, B. Schölkopf, R. Turner, and J. Peters, “Invariant models for causal transfer learning,” *Journal of Machine Learning Research (JMLR)*, vol. 19, no. 36, pp. 1–34, 2018.
- [9] M. Arjovsky, L. Bottou, I. Gulrajani, and D. Lopez-Paz, “Invariant risk minimization,” *CoRR*, Jul. 2019, arXiv e-prints: 1907.02893.
- [10] B. Aubin, A. Slowik, M. Arjovsky, L. Bottou, and D. Lopez-Paz, “Linear unit-tests for invariance discovery,” *CoRR*, Feb. 2021, arXiv e-prints: 2102.10867.
- [11] L. Wang and H. Wu, *Biomedical Optics: Principles and Imaging*, 1st ed. Hoboken, NJ: Wiley-Interscience, May 2007.
- [12] R. Kruger, P. Liu, Y. Fang, and R. Appledorn, “Photoacoustic ultrasound (paus)—reconstruction tomography,” *Medical Physics*, vol. 22, no. 10, p. 1605–1609, 1995.
- [13] C. Lutzweiler and D. Razansky, “Optoacoustic imaging and tomography: reconstruction approaches and outstanding challenges in image performance and quantification,” *Sensors*, vol. 13, pp. 7345–7384, 2013.
- [14] M. Xu and L. Wang, “Universal back-projection algorithm for pho-

- toacoustic computed tomography," *Phys. Rev. E*, vol. 71, p. 016706, 2005.
- [15] P. Burgholzer, J. Bauer-Marschallinger, H. Grün, M. Haltmeier, and G. Paltauf, "Temporal back-projection algorithms for photoacoustic tomography with integrating line detectors," *Inverse Problems*, vol. 23, no. 6, p. S65–S80, Nov 2007.
- [16] A. Rosenthal, V. Ntziachristos, and D. Razansky, "Acoustic inversion in optoacoustic tomography: A review," *Current Medical Imaging Reviews*, vol. 9, pp. 318–336, 2013.
- [17] C. Tian, M. Pei, K. Shen, S. Liu, Z. Hu, and T. Feng, "Impact of system factors on the performance of photoacoustic tomography scanners," *Phys. Rev. Applied*, vol. 13, p. 014001, 2020.
- [18] A. Rosenthal, D. Razansky, and V. Ntziachristos, "Fast semi-analytical model-based acoustic inversion for quantitative optoacoustic tomography," *IEEE Transactions on Medical Imaging*, vol. 29, no. 6, p. 1275–1285, Jun 2010.
- [19] L. Hirsch, M. G. González, and L. Rey Vega, "On the robustness of model-based algorithms for photoacoustic tomography: Comparison between time and frequency domains," *Review of Scientific Instruments*, vol. 92, no. 11, p. 114901, Nov 2021.
- [20] G. Paltauf, P. R. Torke, and R. Nuster, "Modeling photoacoustic imaging with a scanning focused detector using monte carlo simulation of energy deposition," *Journal of biomedical optics*, vol. 23, no. 12, 2018.
- [21] A. Rosenthal, V. Ntziachristos, and D. Razansky, "Model-based optoacoustic inversion with arbitrary-shape detectors," *Medical Physics*, vol. 38, no. 7, pp. 4285–4295, 2011.
- [22] A. D. Cezaro and F. T. D. Cezaro, "Regularization approaches for quantitative photoacoustic tomography using the radiative transfer equation," *Journal of Computational and Applied Mathematics*, vol. 292, pp. 1–14, 2015.
- [23] S. R. Arridge, M. M. Betcke, B. T. Cox, F. Lucka, and B. E. Treeby, "On the adjoint operator in photoacoustic tomography," *Inverse Problems*, vol. 32, no. 11, p. 115012, Oct 2016.
- [24] M. G. González, M. Vera, and L. J. R. Vega, "Combining band-frequency separation and deep neural networks for optoacoustic imaging," *Optics and Lasers in Engineering*, vol. 163, p. 107471, 2023.
- [25] A. Hyvärinen, "Estimation of non-normalized statistical models by score matching," *Journal of Machine Learning Research*, vol. 6, no. 24, pp. 695–709, 2005.
- [26] DRIVE, "DRIVE: Digital retinal images for vessel extraction," 2020. [Online]. Available: <https://drive.grand-challenge.org/>
- [27] ARIA, "Automated retinal image analysis," 2006. [Online]. Available: <http://www.damianjfarnell.com/>
- [28] RITE, "Retinal images vessel tree extraction," 2013. [Online]. Available: <https://medicine.uiowa.edu/eye/rite-dataset>
- [29] STARE, "Structured analysis of the retina," 2000. [Online]. Available: <https://cecas.clemson.edu/~ahoover/stare/>
- [30] A. Hatamizadeh, H. Hosseini, N. Patel, J. Choi, C. Pole, C. Hoeferlin, S. Schwartz, and D. Terzopoulos, "RAVIR: A dataset and methodology for the semantic segmentation and quantitative analysis of retinal arteries and veins in infrared reflectance imaging," *IEEE Journal of Biomedical and Health Informatics*, 2022.
- [31] M. González, M. Vera, A. Dreszman, and L. Rey Vega, "Diffusion assisted image reconstruction in optoacoustic tomography," *Optics and Lasers in Engineering*, vol. 178, p. 108242, 2024.
- [32] K. B. Petersen and M. S. Pedersen, "The matrix cookbook," 2008.

## APÉNDICE A DEMOSTRACIÓN DEL LEMA 1

Mientras que (14) es una identidad estándar, las identidades (15) y (16) requieren un análisis de factores. Es sencillo notar que la distribución de  $X|K = k$  es normal debido a las hipótesis del modelo. Su media y la covarianza pueden calcularse como:

$$\mathbb{E}[X|K = k] = \mathbb{E}[\mathbb{E}[X|Y]|K = k] = A \cdot \mu_k \quad (23)$$

$$\begin{aligned} \text{cov}(X|K = k) &= \mathbb{E}[\text{cov}(X|Y)|K = k] + \text{cov}(\mathbb{E}[X|Y]|K = k) \\ &= \sigma_v^2 \cdot \mathbf{I} + A \cdot \Lambda_k \cdot A^T \end{aligned} \quad (24)$$

De este modo, (16) puede obtenerse mediante la regla de Bayes. Con esta técnica es fácil hallar la distribución

conjunta:

$$\begin{bmatrix} X \\ Y \end{bmatrix} \Big|_{K=k} \sim \mathcal{N} \left( \begin{bmatrix} A\mu_k \\ \mu_k \end{bmatrix}, \begin{bmatrix} \sigma_v^2 \mathbf{I} + A\Lambda_k A^T & A\Lambda_k \\ \Lambda_k A^T & \Lambda_k \end{bmatrix} \right) \quad (25)$$

La ecuación (15) se prueba utilizando propiedades de variables normales multivariadas [32].

## APÉNDICE B DEMOSTRACIÓN DEL LEMA 2

La función *score* se define como  $\psi(x) = \nabla_x \log p(x)$ . Por un lado, puede escribirse intercambiando el orden de diferenciación e integración como:

$$\psi(x) = \frac{\nabla_x p(x)}{p(x)} = \frac{1}{p(x)} \int_{\mathbb{R}^{d_y}} p(y) \nabla_x p(x|y) dy \quad (26)$$

En nuestro modelo  $X|Y = y \sim \mathcal{N}(Ay, \sigma_v^2 \mathbf{I})$ , el gradiente puede calcularse como:

$$\nabla_x p(x|y) = \frac{Ay - x}{\sigma_v^2} p(x|y) \quad (27)$$

y la función *score* es

$$\begin{aligned} \psi(x) &= \frac{1}{p(x)} \int_{\mathbb{R}^{d_y}} p(y) \frac{Ay - x}{\sigma_v^2} p(x|y) dy \\ &= \frac{A \cdot \mathbb{E}[Y|X = x] - x}{\sigma_v^2} \end{aligned} \quad (28)$$

De esta manera, (17) se demuestra resolviendo (28). Por otro lado, la densidad de probabilidad  $p(x)$  en este modelo puede escribirse como:

$$\begin{aligned} p(x) &= \frac{1}{(2\pi\sigma_v^2)^{\frac{d_x}{2}}} \int_{\mathbb{R}^{d_y}} p(y) e^{-\frac{1}{2\sigma_v^2} \|x - Ay\|^2} dy \\ &= \frac{1}{(2\pi\sigma_v^2)^{\frac{d_x}{2}}} \mathbb{E} \left[ e^{-\frac{1}{2\sigma_v^2} \|x - AY\|^2} \right] \end{aligned} \quad (29)$$

Por lo tanto, la función *score* también puede expresarse como

$$\psi(x) = \nabla_x \log \mathbb{E} \left[ e^{-\frac{1}{2\sigma_v^2} \|x - AY\|^2} \right] \quad (30)$$

Finalmente, el lema queda demostrado.