

UN ENFOQUE ESTADÍSTICO PARA LA REGULARIZACIÓN DE PROBLEMAS INVERSOS

María J. Carrió^{a,b}, Gisela L. Mazzieri^{a,d} and Karina G. Temperini^{c,d}

^a*Departamento de Matemática, Facultad de Bioquímica y Ciencias Biológicas, Universidad Nacional del Litoral, Santa Fe, Argentina, <http://www.fbc.unl.edu.ar/>.*

^b*Departamento de Matemática, Facultad de Ingeniería Química, Universidad Nacional del Litoral, Santa Fe, Argentina, <http://www.fiq.unl.edu.ar/depto-mate>*

^c*Departamento de Matemática, Facultad de Humanidades y Ciencias, Universidad Nacional del Litoral, Santa Fe, Argentina, <http://www.fhuc.unl.edu.ar/>*

^d*Instituto de Matemática Aplicada del Litoral, IMAL (CONICET-UNL), CCT CONICET Santa Fe, Colectora Ruta Nac. N° 168, Paraje El Pozo, 3000 Santa Fe, Argentina
<http://www.imal.santafe-conicet.gov.ar/>*

Palabras Clave. Problemas Inversos, Estimación Bayesiana, Anomalías.

Resumen. Los problemas inversos mal condicionados aparecen en una amplia variedad de áreas de Ciencia y Tecnología. Con el objetivo de extraer la mayor cantidad posible de información acerca de la solución exacta y mantener la estabilidad del proceso de inversión, surge la necesidad de “regularizar” este tipo de problemas. Mientras que los métodos clásicos de regularización producen una única estimación de la solución, los métodos estadísticos dan como resultado una distribución de probabilidades, la que puede luego utilizarse para obtener diferentes estimaciones de la solución. En muchas aplicaciones, por ejemplo en medicina y astronomía, es de particular interés la “detección de anomalías”. Estos problemas surgen cuando se presume que la variable de estudio (incógnita) está compuesta por dos (o más) componentes, cada una de ellas de diferente naturaleza estadística. La información que se dispone (dato del problema inverso) es una señal (imagen) “borrosa” de la señal (imagen) mixta. El objetivo es entonces, mediante la utilización de distribuciones “a-priori” apropiadas, obtener estimaciones de la variable de interés que reflejen el comportamiento esperado (J. Kaipio and E. Somersalo, *Journal of Computational Physics*, 181:398–406 (2002); J. Kaipio and E. Somersalo, volume 160 of *Applied Mathematical Sciences*, Springer-Verlag, New York, (2005)). En este trabajo se presenta el problema de detección de anomalías bajo degradación de tipo Gaussiana y “prior” de impulso. Asimismo se muestran resultados obtenidos en ejemplos concretos de procesamiento de señales y restauración de imágenes.

1 INTRODUCCIÓN Y PRELIMINARES

El propósito de las técnicas clásicas de regularización es obtener una aproximación “razonable” de la variable de interés basada en los datos de los que se dispone. Por otra parte, desde el punto de vista de la teoría de inversión estadística, la solución de un problema inverso es una distribución de probabilidades que puede utilizarse para obtener estimaciones de la incógnita. Dicha distribución, llamada la distribución de probabilidades “a-posteriori”, describe el grado de conocimiento de la incógnita después que la información contenida en los datos ha sido procesada e incorporada a tal distribución. Por ende, el enfoque estadístico para tratar y resolver un problema inverso es muy diferente, y en principio totalmente disjunto, del enfoque clásico determinístico.

A lo largo de todo este trabajo, por cuestiones de simplicidad, nos restringiremos al estudio de la teoría en \mathbb{R}^n y supondremos que las variables aleatorias involucradas en los distintos modelos son absolutamente continuas y, por ende, representaremos su distribución de probabilidades por medio de una función de densidad. En general tendremos un modelo de la forma: $Y = f(X, E)$ donde X es la variable incógnita, f es la función asociada al modelo, Y es la variable observable y E el error cometido en las mediciones de Y .

Desde el punto de vista Bayesiano, un problema inverso se expresa de la siguiente manera: *dada una observación de $Y = y_{obs}$, hallar la distribución de probabilidad condicional $\pi(x|y_{obs})$ de la variable X .* El siguiente teorema nos provee una manera de determinar dicha distribución.

Teorema 1.1. (Fórmula de Bayes para problemas inversos) *Sea X una variable aleatoria en \mathbb{R}^n con distribución de densidad de probabilidad “a-priori” dada por π_{pr} . Sea y_{obs} una observación de la variable aleatoria $Y \in \mathbb{R}^k$ tal que $\pi(y_{obs}) > 0$. Entonces la distribución de probabilidad “a-posteriori” de X , dado el dato y_{obs} está dada por $\pi_{post}(x) \doteq \pi(x|y_{obs}) = \frac{\pi_{pr}(x)\pi(y_{obs}|x)}{\pi(y_{obs})} \propto \pi_{pr}(x)\pi(y_{obs}|x)$.*

El símbolo “ \propto ”; significa “proporcional a”. Teniendo en cuenta el teorema anterior podríamos resumir la resolución de un problema inverso, desde un punto de vista estadístico, en las siguientes tres tareas:

- basado en la información disponible de la variable incógnita X , hallar una distribución a-priori que refleje dicha información;
- hallar la función de verosimilitud $\pi(y|x)$;
- desarrollar métodos para explorar la distribución a-posteriori.

Claramente, la definición abstracta de la solución de un problema inverso como la distribución de probabilidad a-posteriori no es muy útil en la práctica. En general, estamos interesados en determinar distintos estimadores de dicha distribución. Entre los más utilizados se encuentran el estimador “máximo a-posteriori” y la “media condicional” que definimos a continuación:

Definición 1.2. *Dada la densidad de probabilidad a-posteriori $\pi(x|y)$ de la variable aleatoria incógnita $X \in \mathbb{R}^n$, definimos:*

1) *El estimador máximo a-posteriori, y lo denotamos con x_{MAP} , como*

$$x_{MAP} \doteq \arg \max_{x \in \mathbb{R}^n} \pi(x|y). \quad (1)$$

2) El estimador media condicional, y lo denotamos con x_{CM} , como

$$x_{CM} \doteq E(x|y) = \int_{\mathbb{R}^n} x\pi(x|y)dx. \quad (2)$$

El siguiente teorema (Kaipio and Somersalo (2005)) provee una manera de obtener la distribución a-posteriori de la incógnita bajo ciertas hipótesis en el modelo y en las distribuciones involucradas. Es oportuno mencionar que este resultado permite probar la relación entre los enfoques determinístico y estocástico.

Teorema 1.3. Sean las variables aleatorias independientes $X \in \mathbb{R}^n$ e $Y, E \in \mathbb{R}^k$ tales que $X \sim \mathcal{N}(\bar{x}, \Gamma)$, $E \sim \mathcal{N}(0, \sigma^2 I_k)$, donde I_k es la matriz identidad de orden k . Supongamos además que el modelo que relaciona las variables es lineal con ruido aditivo, es decir $Y = AX + E$, donde $A \in \mathbb{R}^{k \times n}$ es una matriz conocida. Entonces la distribución de densidad a-posteriori de X dada una medición de $Y = y$ está dada por

$$\pi(x|y) \propto \exp \left(-\frac{1}{2} \left[\left\| \frac{1}{\sigma} (y - Ax) \right\|^2 + \|R(x - \bar{x})\|^2 \right] \right),$$

donde $\Gamma^{-1} = R^T R$ es la factorización de Cholesky de la inversa de las matriz de covarianza de X .

Del resultado anterior se sigue inmediatamente que obtener el estimador x_{MAP} es equivalente a encontrar una solución regularizada del problema inverso $Ax = y$ con el método de Tikhonov-Phillips generalizado con penalizante $\|Rx\|^2$ y parámetro de regularización σ^2 (Engl et al. (1996), Schuster et al. (2012), Mazzieri et al. (2012)).

2 PROBLEMA DE DETECCIÓN DE ANOMALÍAS

Como mencionamos anteriormente, la distribución a-priori de la incógnita debería reflejar el comportamiento estándar de la misma. Sin embargo, en muchos casos es de particular interés la detección de *anomalías* (desviaciones de lo típico). Por ejemplo, en aplicaciones de imágenes médicas, la distribución a-priori podría basarse en información anatómica estándar mientras que el objetivo es hallar anomalías tales como tumores en tejidos sanos o fracturas óseas, normalmente ausentes desde el punto de vista de estructuras anatómicas “típicas”. Estas anomalías representan “outliers” para una distribución a-priori anatómicamente correcta en el sentido que dicha distribución les asignaría una probabilidad de ocurrencia baja, o incluso podrían ser ignoradas.

A continuación se presentan dos modelos de detección de anomalías aplicados al procesamiento de señales e imágenes.

2.1 Caso unidimensional: Aplicaciones a restauración de señales

Un modelo matemático básico para la convolución de señales está dado por la ecuación integral de Fredholm de primera clase:

$$g(s) = \int_0^1 a(s-t)f(t)dt, \quad (3)$$

donde $a(s) = \frac{1}{\sqrt{2\pi}\sigma_b} \exp\left(-\frac{s^2}{2\sigma_b^2}\right)$ es un núcleo Gaussiano con media 0 y desviación estándar $\sigma_b > 0$, f es la señal original que se desea estimar en $[0, 1]$ y g es el dato del problema. La

ecuación (3) fue discretizada de manera usual (usando colocación y cuadratura), resultando en un modelo discreto de la forma

$$y = Ax,$$

donde A es una matriz de orden $n \times k$ que resulta de la discretización del núcleo a , $x \in \mathbb{R}^n$ e $y \in \mathbb{R}^k$. El dato y fue contaminado con un ruido aditivo Gaussiano e de media 0 y desviación estándar del orden de $\sigma_e\%$ de $\|y\|_\infty$, es decir $y = Ax + e$. Para la simulación numérica se utilizaron los siguientes parámetros: $\sigma_b = 0.1$, $n = k = 200$, $\sigma_e = 0.5$.

Supongamos que se sabe a-priori que la señal incógnita consiste de una parte regular (que se obtiene bajo condiciones normales) y una parte anómala que puede o no estar presente. Consideramos el modelo estocástico

$$X = X_r + WX_a,$$

donde X_r y X_a son variables aleatorias mutuamente independientes que representan las partes regular y anómala, respectivamente. La variable aleatoria W toma los valores 0 y 1 con probabilidad $1 - \alpha$ y α , respectivamente, es decir, α es la probabilidad de ocurrencia de la anomalía. Dado que en este caso no se dispone de información acerca de las distribuciones de probabilidad de cada una de las componentes X_r y X_a , supondremos que la densidad a-priori de X es Gaussiana con media \bar{x} y matriz de covarianza Γ , ambas desconocidas. Asimismo, la información previa disponible acerca de la parte regular de la señal que se desea estimar, nos permite representarla mediante

$$f_r(t) = h \chi_{[a,b]}(t), \quad (4)$$

donde $\chi_{[a,b]}$ es la función característica sobre el intervalo $[a, b] \subset [0, 1]$. Además creemos que es posible representar la parte anómala por la suma de dos Gaussianas de pequeño soporte, cada uno de los cuales está contenido en $[a, b]$

$$f_a(t) = C_1 e^{-\frac{1}{2\gamma_1}(t-t_1)^2} + C_2 e^{-\frac{1}{2\gamma_2}(t-t_2)^2}. \quad (5)$$

Con el objetivo de obtener los estadísticos muestrales de X (media \hat{x} y matriz de covarianza $\hat{\Gamma}$), generamos una muestra aleatoria de tamaño 20000 para las partes regular y anómala suponiendo que $\alpha = 0.3$ y que los parámetros involucrados en (4) y (5) son variables aleatorias que se distribuyen de la siguiente manera: $h \sim \text{Unif}([1, 1.2])$, $a \sim \text{Unif}([0.2, 0.25])$, $b \sim \text{Unif}([0.65, 0.7])$, $C_1 \sim \text{Unif}([0.3, 0.5])$, $C_2 \sim \text{Unif}([0.3, 0.6])$, $\gamma_1 \sim \text{Unif}([0.009, 0.015])$, $\gamma_2 \sim \text{Unif}([0.012, 0.018])$, $t_1 \sim \text{Unif}([0.3, 0.4])$ y $t_2 \sim \text{Unif}([0.5, 0.6])$.

La señal original se calculó a partir de (4) y (5) utilizando como parámetros los promedios de los valores obtenidos en la simulación de las correspondientes variables aleatorias con distribución uniforme. En la Figura 1 se muestran la señal original (con y sin anomalía) y los correspondientes datos con ruido.

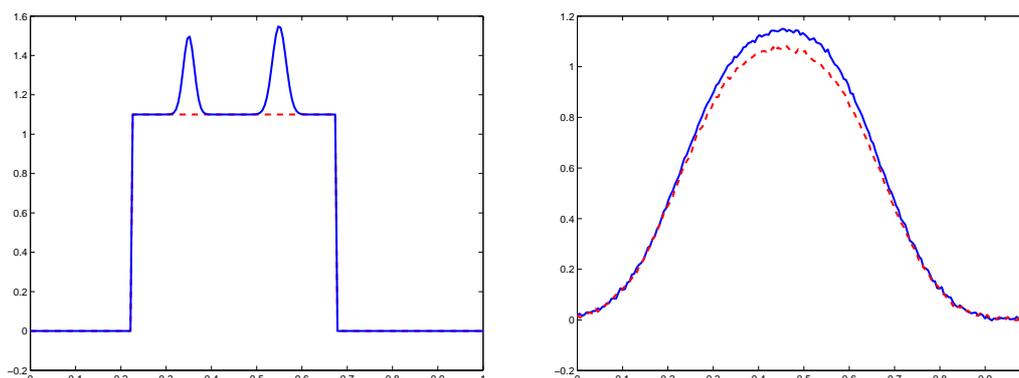


Figura 1. Señal original anómala (—) y sin anomalía (- -) (izquierda), datos con ruido generados de la señal anómala (—) y sin anomalía (- -) (derecha).

Dado que $X \sim \mathcal{N}(\bar{x}, \Gamma)$ y la variable aleatoria ruido $E \sim \mathcal{N}(0, \tilde{\sigma}_e^2 I_n)$, con $\tilde{\sigma}_e = \frac{\sigma_e}{100} \|y\|_\infty$, del Teorema 1.3 y (1) se sigue que

$$x_{\text{MAP}} = \arg \min_{x \in \mathbb{R}^n} [\|y - Ax\|^2 + \tilde{\sigma}_e^2 \|R(x - \bar{x})\|^2].$$

Es importante observar aquí que se aproximó la matriz de covarianza muestral $\hat{\Gamma}$ por la que resulta de su descomposición en valores singulares truncada (con umbral 10^{-10}), a la cual seguiremos denotando con $\hat{\Gamma}$. Así, la matriz R necesaria para el cálculo del estimador x_{MAP} , se obtiene de la factorización de Cholesky de $\hat{\Gamma}^{-1}$.

Para obtener el estimador x_{MAP} se utilizó el método Gradiente Conjugado de Mínimos Cuadrados (CGLS) (Calvetti and Somersalo (2007)) utilizando como criterio de parada el principio de discrepancia de Morozov con $\tau = 1.1$ (Morozov (1966), Engl et al. (1996)). En la Figura 2 (izquierda) se muestran la señal original con anomalía y las 3 iteraciones obtenidas con CGLS a partir del correspondiente dato con ruido. Análogamente, a la derecha se presentan los resultados obtenidos para la señal original sin anomalía.

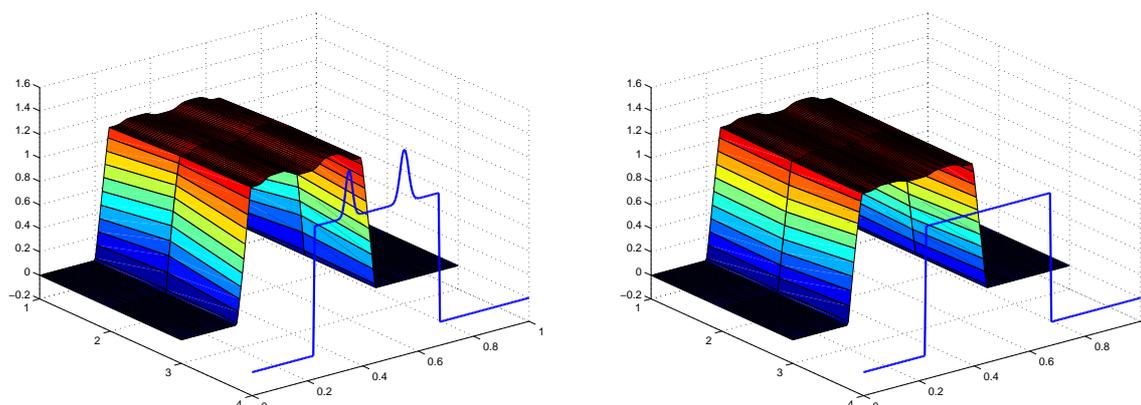


Figura 2. Señal original e iteraciones obtenidas con CGLS para dato anómalo (izquierda) y sin anomalía (derecha).

Se puede observar que la performance del estimador x_{MAP} es muy buena cuando el dato proviene de una fuente con anomalía. Asimismo, cuando el dato es no-anómalo, se obtiene

una buena aproximación de la componente regular de X . Para una mejor visualización de este comportamiento, en la Figura 3 se muestran la señal original y los estimadores máximo a-posteriori obtenidos con dato anómalo (izquierda) y sin anomalía (derecha).

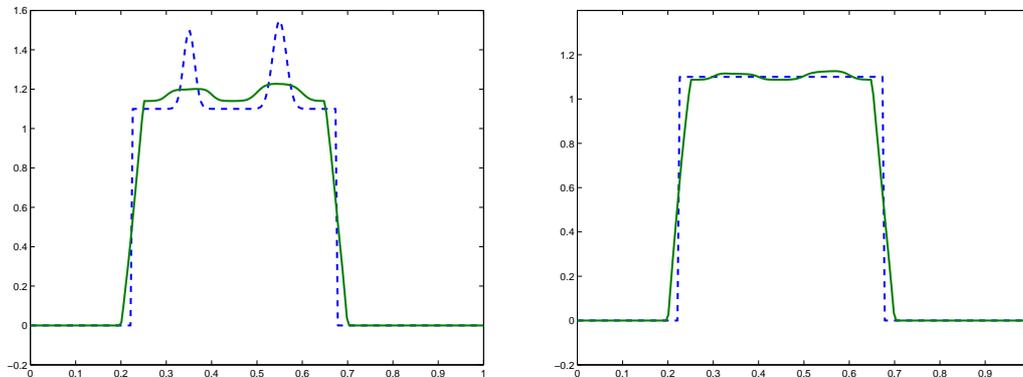


Figura 3. Señal original (- -) y estimador MAP (—) para dato anómalo (izquierda) y sin anomalía (derecha).

2.2 Caso bidimensional: Aplicaciones a restauración de imágenes

El objetivo de esta sección es el estudio de la detección de anomalías en un problema de restauración de imágenes. Análogamente al caso unidimensional, un modelo matemático muy general para el proceso de degradación de imágenes está dado por la siguiente ecuación integral

$$g(s, t) = \int_0^1 \int_0^1 a(s, t, s', t') f(s', t') ds' dt',$$

donde f y g son las imágenes original y degradada, respectivamente y el núcleo a está dado por la Gaussiana bidimensional

$$a(s, t, s', t') = \frac{1}{2\pi\sigma_h\sigma_v} \exp\left(-\frac{1}{2}\left(\frac{s-s'}{\sigma_h}\right)^2 - \frac{1}{2}\left(\frac{t-t'}{\sigma_v}\right)^2\right),$$

donde las desviaciones estándar σ_h y σ_v caracterizan la degradación en la dirección horizontal y vertical, respectivamente. En este ejemplo se utilizaron $\sigma_h = 0.04$ y $\sigma_v = 0.05$. Para obtener el dato del problema inverso se contaminó la imagen borrosa con ruido aditivo Gaussiano de media 0 y desviación estándar $\tilde{\sigma}_e = 0.01 \|y\|_\infty$, donde y es la discretización del dato g . El correspondiente modelo estocástico está dado por $Y = AX + E$, donde $X, Y, E \in \mathbb{R}^N$ son vectores aleatorios, $N \doteq n^2$, siendo la imagen original de tamaño $n \times n$. La matriz A de orden $N \times N$ está asociada a la discretización del núcleo a .

Supongamos que la información disponible a-priori de la variable X , en el caso estándar, se codifica en la densidad de probabilidad π_1 . Además creemos que X presenta anomalías, es decir, características que se desvían considerablemente de un caso típico. Asimismo conocemos, en términos estadísticos, qué tipo de desviaciones son posibles. Esta información se expresa en una densidad de probabilidad que denotaremos con π_2 . En este ejemplo, al igual que en el caso unidimensional, consideramos un modelo aditivo dado por

$$X = U + V, \text{ con } U \sim \pi_1(u), V \sim \pi_2(v),$$

donde U representa la parte regular de X y V , la parte anómala. De esta descomposición resulta $Y = AU + AV + E$. Suponiendo que las variables U y V son independientes, este modelo conduce a la siguiente densidad conjunta a-priori

$$\pi(u, v) = \pi_1(u)\pi_2(v).$$

Consideraremos que la parte regular es suave, lo cual se expresa en términos de una densidad de suavidad de segundo orden (Kaipio and Somersalo (2005), Calvetti and Somersalo (2007)), es decir

$$\pi_1(u) \propto \exp\left(-\frac{\alpha}{2} \|Lu\|^2\right),$$

donde L es la aproximación por diferencias finitas del Laplaciano de u con condición de borde nula y el parámetro α se supone conocido. Así, la matriz de covarianza de la variable aleatoria U está dada por $\Gamma_u \doteq \frac{1}{\alpha}(L^T L)^{-1}$. Respecto de la parte anómala, esperamos que la imagen pueda contener unos pocos píxeles que se diferencian (en intensidad) notablemente de sus vecinos. Se supone además que las componentes del vector aleatorio V toman valores positivos. Para modelar este tipo de anomalías, utilizamos una prior de impulso ℓ^1 dada por

$$\pi_2(v) \propto \pi_+(v) \exp(-\gamma \|v\|_1), \quad \|v\|_1 = \sum_{j=1}^N |v_j|,$$

donde el parámetro γ se supone conocido y π_+ es la restricción de positividad, es decir, $\pi_+(v) = 1$ si todas las componentes de v son positivas y cero en otro caso (Kaipio and Somersalo (2005)). Por último, supondremos que el ruido E es Gaussiano de media cero y matriz de covarianza $\Gamma_e \doteq \sigma_e^2 I_N$ e independiente de U y V .

De las densidades dadas de los vectores aleatorios involucrados en el modelo se deduce fácilmente la siguiente densidad de probabilidad conjunta

$$\begin{aligned} \pi(u, y, v) &= \pi(u, y|v)\pi_2(v) \\ &\propto \exp\left(-\frac{1}{2}[u^T(y - Av)^T]\Gamma^{-1}\begin{bmatrix} u \\ y - Av \end{bmatrix}\right) \pi_+(v) \exp(-\gamma \|v\|_1) \\ &= \pi_+(v) \exp\left(-\frac{1}{2}[u^T(y - Av)^T]\Gamma^{-1}\begin{bmatrix} u \\ y - Av \end{bmatrix} - \gamma \|v\|_1\right), \end{aligned} \quad (6)$$

donde $\Gamma = \begin{pmatrix} \Gamma_u & \Gamma_u A^T \\ A\Gamma_u & A\Gamma_u A^T + \Gamma_e \end{pmatrix} \doteq \begin{pmatrix} \Gamma_{11} & \Gamma_{12} \\ \Gamma_{21} & \Gamma_{22} \end{pmatrix}$.

Para calcular los estimadores U_{CM} y V_{CM} definidos en (2) es necesario obtener las densidades condicionales $\pi(u|v, y)$ y $\pi(v|u, y)$. Del Teorema 1.1, la ecuación (6) y el Teorema 3.5 en Kaipio and Somersalo (2005) se obtiene

$$\pi(u|v, y) \propto \exp\left(-\frac{1}{2}(u - u_0)^T \tilde{\Gamma}_{22}^{-1}(u - u_0)\right), \quad (7)$$

donde $u_0 = \Gamma_{12}\Gamma_{22}^{-1}(y - Av)$ y $\tilde{\Gamma}_{22}$ es el complemento de Schur de Γ_{22} (ver Kaipio and Somersalo (2005)), y

$$\begin{aligned} \pi(v|u, y) &\propto \pi_+(v) \exp\left(-\frac{1}{2}(v^T A^T R_{22} Av - 2v^T (A^T R_{21} u + A^T R_{22} y)) - \gamma \|v\|_1\right) \\ &= \pi_+(v) \exp\left(-\frac{1}{2}(v^T B v - 2v^T q)\right), \end{aligned} \quad (8)$$

donde $B = A^T R_{22} A$, $q = A^T R_{21} u + A^T R_{22} y - \gamma \mathbf{1}$, $\mathbf{1} = [1, \dots, 1]^T$ y R_{ij} denotan los bloques de la matriz $\Gamma^{-1} = \begin{pmatrix} R_{11} & R_{12} \\ R_{21} & R_{22} \end{pmatrix}$.

Para calcular los estimadores U_{CM} y V_{CM} se obtuvo una muestra de tamaño 35000 (de la cual se descartaron los primeros 3000 elementos) a partir de las densidades condicionales dadas en (7) y (8) utilizando el Método de Monte Carlo con cadenas de Markov (MCMC) llamado muestreador de Gibbs (Ross (1999), Kaipio and Somersalo (2005)). En el ejemplo numérico consideramos una imagen original de 30×30 píxeles cuya parte regular consiste en la suma de tres Gaussianas y su parte anómala, de tres fuentes focales. En las densidades a-priori de U y V se utilizaron los parámetros $\alpha = 0.001$ y $\gamma = 8$. En la Figura 4 se muestran las imágenes original y borrosa con ruido.

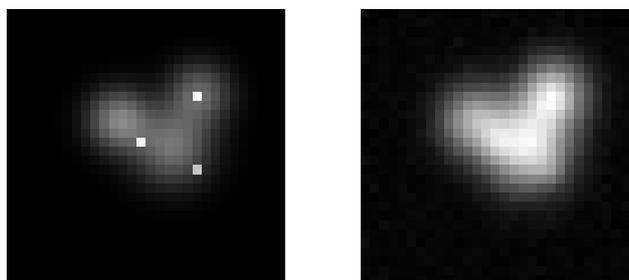


Figura 4. Imagen original (izquierda) e imagen borrosa con ruido (derecha).

En la Figura 5 se muestran las medias condicionales para las partes regular y anómala obtenidas con el algoritmo MCMC. Podemos concluir claramente que en el estimador V_{CM} pueden localizarse las anomalías presentes en la imagen original, mientras que el estimador U_{CM} refleja el comportamiento suave esperado en la componente regular de X .

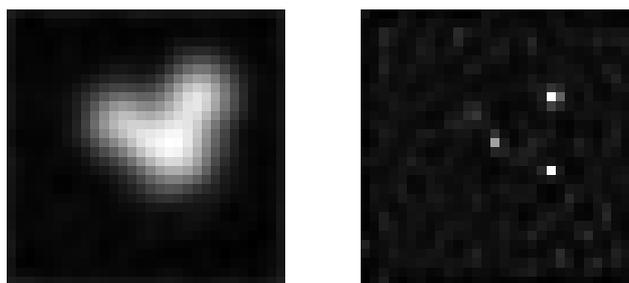


Figura 5. Medias condicionales para la parte regular (izquierda) y anómala (derecha).

3 CONCLUSIONES

En este trabajo se estudió la detección de anomalías en el problema inverso de deconvolución. Se mostró un caso aplicado al procesamiento de señales y otro bidimensional en restauración de imágenes. Se obtuvieron estimadores apropiados en cada uno de los casos que permitieron explorar las densidades a-posteriori obtenidas que evidencian la buena performance de los algoritmos y métodos utilizados.

4 AGRADECIMIENTOS

Este trabajo ha sido subvencionado en parte por el Consejo Nacional de Investigaciones Científicas y Técnicas, CONICET, a través del proyecto PIP 2014-2016 Nro. 11220130100216-CO, por la Universidad Nacional del Litoral, a través de los proyectos CAI+D PJov 2011 Nro. 50020110100055 y CAI+D PI 2011 Nro. 50120110100294 y por la Air Force Office of Scientific Research, AFOSR/SOARD, USA, a través de la Grant FA9550-14-1-0130.

REFERENCES

- Calvetti D. and Somersalo E. *Introduction to Bayesian Scientific Computing. Ten Lectures on Subjective Computing*, volume 2 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, New York, 2007.
- Engl H.W., Hanke M., and Neubauer A. *Regularization of inverse problems*, volume 375 of *Mathematics and its Applications*. Kluwer Academic Publishers Group, Dordrecht, 1996.
- Kaipio J. and Somersalo E. *Statistical and computational inverse problems*, volume 160 of *Applied Mathematical Sciences*. Springer-Verlag, New York, 2005.
- Mazzieri G.L., Spies R.D., and Temperini K.G. Existence, uniqueness and stability of minimizers of generalized Tikhonov-Phillips functionals. *Journal of Mathematical Analysis and Applications*, 396:396–411, 2012.
- Morozov V.A. On the solution of functional equations by the method of regularization. *Soviet Math. Dokl.*, 7:414–417, 1966.
- Ross S.M. *Simulación*. Prentice Hall, México, 1999.
- Schuster T., Kaltenbacher B., Hofmann B., and Kazimierski K. *Regularization Methods in Banach Spaces*. de Gruyter, Berlin, New York, 2012.