Proceeding Series of the Brazilian Society of Computational and Applied Mathematics, v. 10, n. 1, 2023.

Trabalho apresentado no XLII CNMAC, Universidade Federal de Mato Grosso do Sul - Bonito - MS, 2023

Proceeding Series of the Brazilian Society of Computational and Applied Mathematics

Seleção de Parâmetros de Regularização por Minimização de Distâncias Esperadas de Bregman

Elias S. Helou[1]
ICMC-USP, São Carlos, SP
Lucas E. A. Simões[2] Sandra A. Santos[3]
IMECC-UNICAMP, Campinas, SP

Resumo. Técnicas de regularização são ferramentas importantes para a solução numérica de problemas mal-postos (no sentido de Hadamard). O presente trabalho apresenta um método para a seleção do parâmetro de regularização através da minimização de um estimador não tendencioso para o erro preditivo do regularizador. A técnica proposta generaliza métodos existentes através da possibilidade da utilização de preditores mais gerais baseados em divergências de Bregman. O método é aplicável a uma ampla gama de técnicas de regularização lineares ou não-lineares e a diversos modelos estocásticos para o ruído.

Palavras-chave. Regularização, Problemas Inversos, Divergências de Bregman

1 Introdução

Muitos problemas científicos ou de engenharia podem ser formulados como um sistema aproximado de equações não-lineares da forma

$$\boldsymbol{A}(\boldsymbol{x}) \approx \boldsymbol{b}, \tag{1}$$

onde $\boldsymbol{x} \in \mathbb{R}^n$ é o vetor de incógnitas, $\boldsymbol{A} : \mathbb{R}^n \mapsto \mathbb{R}^m$ é a função do sistema que surge do modelo matemático para o problema em mãos e $\boldsymbol{b} \in \mathbb{R}^m$ é o vetor de dados observados, que contém ruído, ou seja, é expresso por

$$\boldsymbol{b} = \boldsymbol{A}(\boldsymbol{x}^*) + \boldsymbol{\epsilon}, \tag{2}$$

onde $\boldsymbol{x}^* \in \mathbb{R}^n$ é a solução exata e $\boldsymbol{\epsilon}$ um vetor desconhecido de variáveis aleatórias. A metodologia proposta a seguir pode ser aplicada a diversos modelos de ruído, incluindo poissoniano, gaussiano, a soma de poissoniano e gaussiano, com distribuição dada pela família exponencial, dentre outros. Aplicações incluem reconstrução tomográfica de imagens [10, 12, 16] e redução de ruído e de borragem de imagens [4, 17].

Problemas inversos mal-condicionados aparecem com grande frequência em aplicações e, portanto, diversos métodos foram desenvolvidos para obter resultados razoáveis a partir de dados contaminados por ruído nestas situações. Tais técnicas são conhecidas como **métodos de regularização** e sempre exigem a escolha de um **parâmetro de regularização** por parte do usuário. O presente trabalho apresenta técnicas aplicáveis ao problema de selecionar o parâmetro de regularização para abordagens não-lineares para regularização.

[1]elias@icmc.usp.br
[2]simoes.lea@gmail.com
[3]sandra@ime.unicamp.br

2 Risco de Bregman

Muitas técnicas de seleção de parâmetro de regularização são baseadas em Estimadores Preditivos Não-tendenciosos de Risco (UPREs, da sigla em inglês) ou em Estimadores Não-tendenciosos de Risco (UREs, da sigla em inglês), ou seja, estimadores computáveis não-tendenciosos para

$$\mathbb{E}\|\boldsymbol{A}(\boldsymbol{x}^\gamma) - \boldsymbol{A}(\boldsymbol{x}^*)\|_2^2 \quad \text{ou} \quad \mathbb{E}\|\boldsymbol{x}^\gamma - \boldsymbol{x}^*\|_2^2, \tag{3}$$

onde $\mathbb{E} := \mathbb{E}_{\boldsymbol{b}}$ é a esperança probabilística sobre a variável aleatória $\boldsymbol{b}$, da qual depende $\boldsymbol{x}^\gamma$. Aqui consideramos que

$$\boldsymbol{x}^\gamma := \boldsymbol{B}_\gamma(\boldsymbol{b}), \tag{4}$$

onde $\boldsymbol{B}_\gamma : \mathbb{R}^m \mapsto \mathbb{R}^n$ é um funcional definido implicitamente pelo método de regularização.

Como as quantidades nas esperanças em (3) não são diretamente computáveis, o Lema de Stein [11, 21] é a ferramenta que permite a obtenção de estimadores úteis para estas quantidades, inicialmente para o caso gaussiano, mas que posteriormente foi generalizada para diversos outros modelos de erro [1, 7, 9, 13–15].

A norma ao quadrado da diferença entre dois vetores é uma medida comum de discrepância que pode ser generalizada através das **divergências de Bregman** $D_f(\boldsymbol{x}, \boldsymbol{y})$, definidas como se segue, para $f : \mathbb{R}^n \to \mathbb{R}$ estritamente convexa:

$$D_f(\boldsymbol{x}, \boldsymbol{y}) = f(\boldsymbol{x}) - f(\boldsymbol{y}) - \nabla f(\boldsymbol{y})^T(\boldsymbol{x} - \boldsymbol{y}), \tag{5}$$

onde $\boldsymbol{x} \in \mathbb{R}^n$ e $\boldsymbol{y} \in \mathbb{R}^n$. O objetivo do presente trabalho é estender os UPREs para divergências de Bregman gerais. Tal objetivo pode ser atingido como se segue.

3 Estimadores Não-tendenciosos de Risco de Bregman

Utilizaremos estimadores não-tendenciosos para quantidades da forma

$$\mathbb{E}\left[\boldsymbol{h}(\boldsymbol{b})^T\boldsymbol{\beta}\right], \tag{6}$$

onde $\boldsymbol{h} : \mathbb{R}^m \mapsto \mathbb{R}^m$ e $\mathbb{E}[\boldsymbol{b}] = \boldsymbol{\beta}$. Para tanto, é necessário o conhecimento das leis probabilísticas que regem $\boldsymbol{b}$. Por exemplo, a aplicação repetida do Lema de Stein [21, Lemma 2] resulta no seguinte, onde $\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2\mathbb{I})$ denota um vetor $\boldsymbol{b}$ de variáveis aleatórias independentes tal que cada uma de suas componentes b_i possui distribuição normal com média β_i e variância σ^2 e $\mathbb{I}$ denota a matriz identidade com dimensões apropriadas ao contexto:

Lema 1. *Sejam $\boldsymbol{\beta} \in \mathbb{R}^m$ e $\boldsymbol{b} \sim \mathcal{N}(\boldsymbol{\beta}, \sigma^2\mathbb{I})$ e considere $\boldsymbol{h} : \mathbb{R}^m \mapsto \mathbb{R}^m$ tal que $\boldsymbol{h}$ é fracamente diferenciável e, para $i \in \{1, 2, \ldots, m\}$, $\mathbb{E}\left|\frac{\partial h_i}{\partial b_i}(\boldsymbol{b})\right| < \infty$. Então*

$$\mathbb{E}\left[\boldsymbol{h}(\boldsymbol{b})^T(\boldsymbol{b} - \boldsymbol{\beta})\right] = \sigma^2\mathbb{E}\left[\sum_{i=1}^m \frac{\partial h_i}{\partial b_i}(\boldsymbol{b})\right]. \tag{7}$$

Agora podemos aplicar este resultado ao caso não-linear (2) e (4) com a divergência de Bregman D_f esperada como medida de risco. Primeiro reescrevemos:

$$\begin{aligned}
&D_f\big(\boldsymbol{A}(\boldsymbol{x}^*), \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) \\
&\quad = f\big(\boldsymbol{A}(\boldsymbol{x}^*)\big) - f\big(\boldsymbol{A}(\boldsymbol{x}^\gamma)\big) - \nabla f\big(\boldsymbol{A}(\boldsymbol{x}^\gamma)\big)^T\big(\boldsymbol{A}(\boldsymbol{x}^*) - \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) \\
&\quad = f\big(\boldsymbol{A}(\boldsymbol{x}^*)\big) - f\big(\boldsymbol{A}(\boldsymbol{x}^\gamma)\big) - \nabla f\big(\boldsymbol{A}(\boldsymbol{x}^\gamma)\big)^T\big(\boldsymbol{b} - \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) \\
&\qquad\qquad + \nabla f\big(\boldsymbol{A}(\boldsymbol{x}^\gamma)\big)^T\big(\boldsymbol{b} - \boldsymbol{A}(\boldsymbol{x}^*)\big) \\
&\quad = f\big(\boldsymbol{A}(\boldsymbol{x}^*)\big) - f(\boldsymbol{b}) + D_f\big(\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) + \nabla f\big(\boldsymbol{A}(\boldsymbol{x}^\gamma)\big)^T\big(\boldsymbol{b} - \boldsymbol{A}(\boldsymbol{x}^*)\big).
\end{aligned} \tag{8}$$

Então podemos provar o seguinte:

Proposição 1. *Suponha que* $\boldsymbol{b} \sim \mathcal{N}\big(\boldsymbol{A}(\boldsymbol{x}^*), \sigma^2\mathbb{I}\big)$ *e sejam* $f : \mathbb{R}^n \to \mathbb{R}$, $\boldsymbol{A} : \mathbb{R}^n \mapsto \mathbb{R}^m$, $\boldsymbol{B}_\gamma : \mathbb{R}^m \mapsto \mathbb{R}^n$, $\boldsymbol{x}^* \in \mathbb{R}^n$ *dados. Defina* $\boldsymbol{x}^\gamma := \boldsymbol{B}_\gamma(\boldsymbol{b})$ *e denote*

$$\boldsymbol{g}_\gamma := \nabla f \circ \boldsymbol{A} \circ \boldsymbol{B}_\gamma. \tag{9}$$

Suponha que f, $\boldsymbol{A}$ *e* $\boldsymbol{B}_\gamma$ *são tais que* $\boldsymbol{g}_\gamma$ *como definido em* (9) *é fracamente diferenciável,* $\mathbb{E}f(\boldsymbol{b}) < \infty$, $\mathbb{E}D_f\big(\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) < \infty$ *e que para* $i \in \{1, 2, \ldots, m\}$, $\mathbb{E}_{b_i}\left|\frac{\partial g_i}{\partial b_i}(\boldsymbol{b})\right| < \infty$. *Então temos:*

$$\mathbb{E}D_f\big(\boldsymbol{A}(\boldsymbol{x}^*), \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) = K + \mathbb{E}D_f\big(\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) + \sigma^2\mathbb{E}\left[\sum_{i=1}^m \frac{\partial g_i}{\partial b_i}(\boldsymbol{b})\right], \tag{10}$$

onde K *é uma constante independente de* γ.

Demonstração. Seja $K := f\big(\boldsymbol{A}(\boldsymbol{x}^*)\big) - \mathbb{E}f(\boldsymbol{b})$, então ao calcularmos esperanças em ambos os lados de (8) obtemos

$$\begin{aligned}\mathbb{E}D_f\big(\boldsymbol{A}(\boldsymbol{x}^*), \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) &= K + \mathbb{E}D_f\big(\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) + \mathbb{E}\left[\nabla f\big(\boldsymbol{A}(\boldsymbol{x}^\gamma)\big)^T\big(\boldsymbol{b} - \boldsymbol{A}(\boldsymbol{x}^*)\big)\right]\\ &= K + \mathbb{E}D_f\big(\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) + \mathbb{E}\left[\boldsymbol{g}_\gamma(\boldsymbol{b})^T\big(\boldsymbol{b} - \boldsymbol{A}(\boldsymbol{x}^*)\big)\right].\end{aligned}$$

Usando o Lema 1 para substituir o último termo do lado direito obtemos o resultado desejado. □

A proposição acima mostra que do ponto de vista da obtenção de um estimador para o risco médio de métodos não-lineares de regularização, aplicar o lema de Stein a uma função de Bregman geral não é mais difícil do que aplicá-lo ao risco médio do erro quadrático. Além do mais, parecem existir motivos convincentes para utilizar outras medidas de divergência [6]. De fato, uma variedade de divergências de Bregman vem sendo utilizada com sucesso em diversas aplicações, tais como análise de componentes principais [5]; estimação de densidade **on-line** [2]; aprendizado de máquina [3, 20] e processamento de fala [8].

Examinemos agora o caso de Poisson. Suponhamos que b possui distribuição de Poisson com média β. Denotamos isso por $b \sim \mathcal{P}(\beta)$. Além do mais, se $\boldsymbol{b}$ é um vetor de variáveis aleatórias independentes tais que $b_i \sim \mathcal{P}(\beta_i)$, simplificamos a notação por $\boldsymbol{b} \sim \mathcal{P}(\boldsymbol{\beta})$. Agora, sejam $b \sim \mathcal{P}(\beta)$ e $h : \mathbb{R} \mapsto \mathbb{R}$ tal que $\mathbb{E}_b[h(b)] < \infty$, então temos [18]:

$$\mathbb{E}_b[\beta h(b)] = \mathbb{E}_b[bh(b-1)]. \tag{11}$$

Esta equação pode ser utilizada para provar o seguinte resultado [14, Propriedade 2]:

Lema 2. *Seja* $\boldsymbol{\beta} \in \mathbb{R}^m_+$, $\boldsymbol{b} \sim \mathcal{P}(\boldsymbol{\beta})$ *e considere* $\boldsymbol{h} : \mathbb{R}^m \mapsto \mathbb{R}^m$ *tal que para* $i \in \{1, 2, \ldots, m\}$, $\mathbb{E}_{b_i}[h_i(\boldsymbol{b})] < \infty$ *e* $\mathbb{E}\left[\boldsymbol{h}(\boldsymbol{b})^T\boldsymbol{b}\right] < \infty$. *Então*

$$\mathbb{E}\left[\boldsymbol{h}(\boldsymbol{b})^T(\boldsymbol{b} - \boldsymbol{\beta})\right] = \mathbb{E}\left[\boldsymbol{b}^T\big(\boldsymbol{h}(\boldsymbol{b}) - \boldsymbol{h}^{[-1]}(\boldsymbol{b})\big)\right], \tag{12}$$

com $\boldsymbol{h}^{[\xi]}$, *para* $\xi \in \mathbb{R}$, *com componentes dadas por*

$$h_i^{[\xi]}(\boldsymbol{b}) := h_i(\boldsymbol{b} + \xi\boldsymbol{e}^i), \tag{13}$$

onde $\boldsymbol{e}^i$ *denota a i-ésima coluna da matriz identidade* $m \times m$.

Agora, utilizando o Lema 2 ao invés do Lema 1, obtemos o resultado a seguir, a prova do qual omitimos por causa de sua similaridade com a prova da Proposição 1.

Proposição 2. *Suponha que* $\boldsymbol{b} \sim \mathcal{P}\big(\boldsymbol{A}(\boldsymbol{x}^*)\big)$ *e sejam* f, $\boldsymbol{A}$, $\boldsymbol{x}^*$, $\boldsymbol{x}^\gamma$ *e* $\boldsymbol{g}_\gamma$ *tais como na Proposição 1. Assuma que* $\boldsymbol{g}_\gamma^{[-1]}$ *siga a notação de* (13)*. Assuma também que* f, $\boldsymbol{A}$ *e* $\boldsymbol{B}_\gamma$ *sejam tais que* $\mathbb{E}f(\boldsymbol{b}) < \infty$, $\mathbb{E}D_f\big(\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) < \infty$, $\mathbb{E}\left[\boldsymbol{g}_\gamma(\boldsymbol{b})^T\boldsymbol{b}\right] < \infty$ *e que para* $i \in \{1, 2, \ldots, m\}$, $\mathbb{E}_{b_i}[g_i(\boldsymbol{b})] < \infty$*. Então temos*

$$\mathbb{E}D_f\big(\boldsymbol{A}(\boldsymbol{x}^*), \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) = K + \mathbb{E}D_f\big(\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) + \mathbb{E}\left[\boldsymbol{b}^T\big(\boldsymbol{g}_\gamma(\boldsymbol{b}) - \boldsymbol{g}_\gamma^{[-1]}(\boldsymbol{b})\big)\right], \tag{14}$$

onde K *é uma constante independente de* γ*.*

Neste ponto, o padrão deve ter se tornado evidente para o leitor. A ideia é que, dado um estimador não-tendencioso, computável através dos dados, para

$$\mathbb{E}\left[\nabla f\big(\boldsymbol{A}(\boldsymbol{x}^\gamma)\big)^T(\boldsymbol{b} - \boldsymbol{A}(\boldsymbol{x}^*))\right], \tag{15}$$

é possível obter de forma simples, a menos de uma constante, um estimador para o valor de $\mathbb{E}D_f\big(\boldsymbol{A}(\boldsymbol{x}^*), \boldsymbol{A}(\boldsymbol{x}^\gamma)\big)$ levando (8) em consideração. Existem estimadores práticos para quantidades como (15) para uma variedade de modelos de ruído estocástico. Por exemplo, podemos mencionar os artigos [14, 15] para o caso da soma de ruído gaussiano com poissoniano; [7] para o caso da família exponencial (que inclui as distribuições gaussiana, de Poisson, binomial, gama e gaussiana inversa) e [9, 13] para distribuições elípticas.

De posse de um estimador cuja esperança seja igual, a menos de uma constante, de um UPRE generalizado para divergências de Bregman, uma regra prática de seleção de parâmetro de regularização é obtida pela minimização deste estimador. Há questões pendentes relacionadas ao cálculo das quantidades $\sum_{i=1}^m \frac{\partial g_i}{\partial b_i}(\boldsymbol{b})$ e $\boldsymbol{b}^T\big(\boldsymbol{g}_\gamma(\boldsymbol{b}) - \boldsymbol{g}_\gamma^{[-1]}(\boldsymbol{b})\big)$, mas existem formas práticas baseadas em ideias estocásticas para aproximar este valor com precisão suficiente.

Uma ideia apresentada em [15, 19] é utilizar a seguinte igualdade:

$$\mathbb{E}_{\boldsymbol{\omega}}\left[\lim_{\epsilon\downarrow 0}\frac{1}{\epsilon}\boldsymbol{\omega}^T\operatorname{diag}(\boldsymbol{z})\big(\boldsymbol{g}_\gamma(\boldsymbol{b} + \epsilon\boldsymbol{\omega}) - \boldsymbol{g}_\gamma(\boldsymbol{b})\big)\right] = \boldsymbol{z}^T\boldsymbol{\partial}\boldsymbol{g}_\gamma(\boldsymbol{b}), \tag{16}$$

onde $\boldsymbol{z} \in \mathbb{R}^m$, $\boldsymbol{\omega} \in \mathbb{R}^m$ é tal que $\mathbb{E}_{\boldsymbol{\omega}}\boldsymbol{\omega} = \boldsymbol{0}$ e $\mathbb{E}_{\boldsymbol{\omega}}\boldsymbol{\omega}\boldsymbol{\omega}^T = \mathbb{I}$, e $\boldsymbol{\partial}\boldsymbol{g}_\gamma$ é definido por componentes como

$$\boldsymbol{\partial}_i\boldsymbol{g}_\gamma = \frac{\partial g_i}{\partial b_i}. \tag{17}$$

Portanto, se o modelo de ruído é gaussiano, podemos definir o seguinte estimador

$$\text{G-UPBRE}_\epsilon^f(\gamma) := D_f\big(\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) + \frac{\sigma^2}{\epsilon}\boldsymbol{\omega}^T\big(\boldsymbol{g}_\gamma(\boldsymbol{b} + \epsilon\boldsymbol{\omega}) - \boldsymbol{g}_\gamma(\boldsymbol{b})\big), \tag{18}$$

onde UPBRE é a a sigla em inglês para para Estimador Não-tendencioso de Risco Preditivo de Bregman. Logo, graças a (10) e (16), temos

$$\mathbb{E}_{\boldsymbol{b},\boldsymbol{\omega}}\left[\lim_{\epsilon\downarrow 0}\text{G-UPBRE}_\epsilon^f(\gamma)\right] = \mathbb{E}D_f\big(\boldsymbol{A}(\boldsymbol{x}^*), \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) - K, \tag{19}$$

onde K não depende de γ.

De forma semelhante, se assumirmos um modelo de Poisson, podemos proceder como a seguir. Com $\boldsymbol{\omega}$ como anteriormente, definamos então

$$\text{P-UPBRE}_\epsilon^f(\gamma) := D_f\big(\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) + \frac{1}{\epsilon}\boldsymbol{\omega}^T\operatorname{diag}(\boldsymbol{b})\big(\boldsymbol{g}_\gamma(\boldsymbol{b} + \epsilon\boldsymbol{\omega}) - \boldsymbol{g}_\gamma(\boldsymbol{b})\big). \tag{20}$$

Portanto, aplicação de (16) leva a

$$\mathbb{E}_{\boldsymbol{\omega}}\left[\lim_{\epsilon\downarrow 0}\text{P-UPBRE}_\epsilon^f(\gamma)\right] = D_f\big(\boldsymbol{b}, \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) + \boldsymbol{b}^T\boldsymbol{\partial}\boldsymbol{g}_\gamma(\boldsymbol{b}). \tag{21}$$

Além disso, notemos que uma aproximação de Taylor de primeira ordem leva a

$$\mathbb{E}\left[\boldsymbol{b}^T\big(\boldsymbol{g}_\gamma(\boldsymbol{b}) - \boldsymbol{g}_\gamma^{[-1]}(\boldsymbol{b})\big)\right] \approx \mathbb{E}\left[\boldsymbol{b}^T\boldsymbol{\partial}\boldsymbol{g}_\gamma(\boldsymbol{b})\right]. \tag{22}$$

Finalmente, calculando a esperança com relação a $\boldsymbol{b}$ em ambos os lados de (21), levando (22) em consideração e então utilizando (14), obtemos:

$$\mathbb{E}_{\boldsymbol{b},\boldsymbol{\omega}}\left[\lim_{\epsilon\downarrow 0}\text{P-UPBRE}_\epsilon^f(\gamma)\right] \approx \mathbb{E}D_f\big(\boldsymbol{A}(\boldsymbol{x}^*), \boldsymbol{A}(\boldsymbol{x}^\gamma)\big) - K, \tag{23}$$

que é um resultado aproximado, diferentemente de (19). O erro na aproximação deve ser relativamente pequeno, uma vez que, para variáveis aleatórias de Poisson em aplicações de imageamento, perturbações de tamanho 1 são provavelmente pequenas em comparação com o valor das variáveis sendo perturbadas. De fato, experimentação numérica mostra que a aproximação (22) é precisa o suficiente para aplicações práticas [15].

4 Considerações Finais

O presente trabalho demonstrou que, do ponto de vista teórico, não é difícil generalizar a ideia de escolher o parâmetro de regularização através da minimização do erro preditivo quadrático pela escolha de tal parâmetro via minimização do erro preditivo de Bregman. Experimentação numérica por nós realizada mostrou que, de fato, utilizar

$$\underset{\gamma\in\mathbb{R}_+}{\arg\min}\,\text{G-UPBRE}_\epsilon^f(\gamma) \quad \text{ou} \quad \underset{\gamma\in\mathbb{R}_+}{\arg\min}\,\text{P-UPBRE}_\epsilon^f(\gamma) \tag{24}$$

com f distinta da que resulta em $D_f(\boldsymbol{x},\boldsymbol{y}) = \|\boldsymbol{x}-\boldsymbol{y}\|_2^2$ pode ser vantajoso em aplicações de reconstrução de imagens tomográficas. Tais experimentos não foram apresentados por serem inconclusivos e porque nossa intenção foi destacar a teoria atualmente desenvolvida.

Além de realizar experimentação numérica mais conclusiva, outro objetivo de pesquisa futura é estudar, teórica e experimentalmente, a ocorrência de fenômenos de concentração da medida na aplicação dos métodos propostos em problemas de grande porte.

Agradecimentos

O presente trabalho foi financiado pelos projetos CNPq 305010/2020-4 e 308004/2022-1 e FAPESP 2018/24293-0 e 2013/07375-0.

Referências

[1] R. Averkamp e C. Houdré. “Stein Estimate for Infinitely Divisible Laws”. Em: **ESAIM: Probability and Statistics** 10 (2006), pp. 269–276. DOI: 10.1051/ps:2006011. URL: http://journals.cambridge.org/abstract_S1292810006000115.

[2] K. S. Azoury e M. K. Warmuth. “Relative Loss Bounds for On-Line Density Estimation with the Exponential Family of Distributions”. Em: **Machine Learning** 43.3 (2001), pp. 211–246. DOI: 10.1023/A:1010896012157. URL: http://link.springer.com/article/10.1023/A:1010896012157.

[3] A. Banerjee et al. “Clustering with Bregman Divergences”. Em: **Journal of Machine Learning Research** 6 (2005), pp. 1705–1749. URL: http://dl.acm.org/citation.cfm?id=1194902.

[4] A. Bovik, ed. **Handbook of Image & Video Processing**. San Diego (CA): Elsevier Academic Press, 2005.

[5] M. Collins, S. Dasgupta e R. E. Schapire. "A Generalization of Principal Components Analysis to the Exponential Family". Em: **Advances in Neural Information Processing Systems** 14 (2001). URL: http://papers.nips.cc/paper/2078-a-generalization-of-principal-components-analysis-to-the-exponential-family.

[6] I. Csiszár. "Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems". Em: **The Annals of Statistics** 19.4 (1991), pp. 2032–2066. URL: http://www.jstor.org/stable/2241918.

[7] Y. C. Eldar. "Generalized SURE for Exponential Families: Applications to Regularization". Em: **IEEE Transactions on Signal Processing** 57.2 (2009), pp. 471–481. DOI: 10.1109/TSP.2008.2008212. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4663926.

[8] R. M. Gray et al. "Distortion Measures for Speech Processing". Em: **IEEE Transactions on Acoustics, Speech and Signal Processing** 28.4 (1980), pp. 367–376. DOI: 10.1109/TASSP.1980.1163421. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=1163421.

[9] M. Hamada e E. A. Valdez. "CAPM and Option Pricing with Elliptically Contoured Distributions". Em: **The Journal of Risk and Insurance** 75.2 (2008), pp. 387–409. DOI: 10.1111/j.1539-6975.2008.00265.x. URL: http://onlinelibrary.wiley.com/doi/10.1111/j.1539-6975.2008.00265.x/full.

[10] G. T. Herman. **Image Reconstruction from Projections: The Fundamentals of Computerized Tomography**. Boston (MA): Academic Press, 1980.

[11] W. James e C. Stein. "Estimation with Quadratic Loss". Em: **Proceedings of the Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Contributions to the Theory of Statistics**. University of California Press, 1961, pp. 361–379. URL: http://projecteuclid.org/euclid.bsmsp/1200512173.

[12] A. C. Kak e M. Slaney. **Principles of Computerized Tomographic Imaging**. New York: IEEE Press, 1988.

[13] Z. Landsman e J. Nešlehová. "Stein's Lemma for Elliptical Random Vectors". Em: **Journal of Multivariate Analysis** 99 (2008), pp. 912–927. DOI: 10.1016/j.jmva.2007.05.006. URL: http://www.sciencedirect.com/science/article/pii/S0047259X07000814.

[14] F. Luisier, T. Blu e M. Unser. "Image Denoising in Mixed Poisson–Gaussian Noise". Em: **IEEE Transactions on Image Processing** 20.3 (2011), pp. 696–708. DOI: 10.1109/TIP.2010.2073477. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5570958.

[15] Y. L. Montagner, E. D. Angelini e J.-C. Olivo-Marin. "An Unbiased Risk Estimator for Image Denoising in the Presence of Mixed Poisson–Gaussian Noise". Em: **IEEE Transactions on Image Processing** 23.3 (2014), pp. 1255–1268. DOI: 10.1109/TIP.2014.2300821. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=6714502.

[16] F. Natterer. **The Mathematics of Computerized Tomography**. Chichester, New York, Brisbane, Toronto, Singapore: John Wiley & Sons Ltd, 1986.

[17] F. Natterer e F. Wübbeling. **Mathematical Methods in Image Reconstruction**. Philadelphia (PA): SIAM, 2001.

[18] J. C.-M. Peng. **Simultaneous Estimation of the Parameters of Independent Poisson Distribution**. Rel. técn. EFS NSF 78. Stanford, dez. de 1975. URL: https://purl.stanford.edu/gj419gc5599.

[19] S. Ramani, T. Blu e M. Unser. "Monte-Carlo SURE: A Black-Box Optimization of Regularization Parameters for General Denoising Algorithms". Em: **IEEE Transactions on Image Processing** 17.9 (2008), pp. 1540–1554. DOI: 10.1109/TIP.2008.2001404. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4598837.

[20] S. Si, D. Tao e B. Geng. "Bregman Divergence-Based Regularization for Transfer Subspace Learning". Em: **IEEE Transactions on Knowledge and Data Engineering** 22.7 (2010), pp. 929–942. DOI: 10.1109/TKDE.2009.126. URL: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4967588&tag=1.

[21] C. M. Stein. "Estimation of the Mean of a Multivariate Normal Distribution". Em: **The Annals of Statistics** 9.6 (1981), pp. 1135–1151. URL: http://www.jstor.org/stable/2240405.

