A Characterization of the Domain of Beta-Divergence and Its Connection to Bregman Variational Model

Hyenkyun Woo

School of Liberal Arts, Korea University of Technology and Education, Cheonan 31253, Korea;
hyenkyun@koreatech.ac.kr

Abstract: In image and signal processing, the beta-divergence is well known as a similarity measure between two positive objects. However, it is unclear whether or not the distance-like structure of beta-divergence is preserved, if we extend the domain of the beta-divergence to the negative region. In this article, we study the domain of the beta-divergence and its connection to the Bregman-divergence associated with the convex function of Legendre type. In fact, we show that the domain of beta-divergence (and the corresponding Bregman-divergence) include negative region under the mild condition on the beta value. Additionally, through the relation between the beta-divergence and the Bregman-divergence, we can reformulate various variational models appearing in image processing problems into a unified framework, namely the Bregman variational model. This model has a strong advantage compared to the beta-divergence-based model due to the dual structure of the Bregman-divergence. As an example, we demonstrate how we can build up a convex reformulated variational model with a negative domain for the classic nonconvex problem, which usually appears in synthetic aperture radar image processing problems.

Keywords: beta-divergence; bregman-divergence; sparsity; convex function of legendre type; optimization; synthetic aperture radar; multiplicative noise; bregman proximity operator; convexity; total variation

1. Introduction

In general, the domain of a divergence [1,2] is that confined not by the positiveness of variables but by the positiveness of a divergence (i.e., $D(b|u) \geq 0$). Therefore, the domain of a divergence could be defined to include negative region while keeping positiveness of the divergence. To the best of our knowledge, it is unclear when the domain of the β-divergence (and the corresponding Bregman-divergence) include the negative region. In this article, we systematically explore the domains of the β-divergence [2] and the corresponding Bregman-divergence associated with the convex function of Legendre type [3].

The β-divergence [2,4–7] is a general framework of similarity measures induced from various statistical models, such as Poisson, Gamma, Gaussian, Inverse Gaussian, compound Poisson, and Tweedie distribution. For the connection between the β-divergence and the various statistical distributions, see [8]. Among the diverse statistical distributions, the Tweedie distribution has a unique feature, i.e., the unit deviance of the Tweedie distribution [8] corresponds to the β-divergence with $\beta \in \mathbb{R} \setminus (1,2)$. It is interesting that $(1,2)$ is a vital range of β while defining a convex right Bregman proximity operator [9,10]. We will address this issue in more details in Section 4. In addition, the β-divergence is also used as a distance-like measure in diverse areas, for instance, synthetic aperture radar (SAR) image processing [11,12], audio spectrogram comparison [6,13], and brain EEG signal processing [7].

We note that authors in [7] show the usefulness of the β-divergence with $\beta > 1$ as a robust similarity measure against outliers between two probability distributions. Here, outliers

(rare events) are that have extremely low probability and thus they exist near zero probability. However, the (generalized) Kullback–Leibler-divergence (i.e., β-divergence $D_{\beta=1}(b|u)$), which is a commonly used similarity measure for probability distributions, is undefined at zero ($u = 0$). See Figure 1a. Therefore, it is not easy to obtain robustness against outliers through the (generalized) Kullback–Leibler-divergence. On the contrary, the β-divergence with $\beta > 1$ (i.e., $D_{\beta>1}(b|u)$) is well defined at zero ($u = 0$) and thus it is more robust to outliers than the Kullback–Leibler-divergence. For more details, see [4,5,7]. We also note that if the variables of β-divergence are not probability distributions (i.e., unnormalized) then outliers correspond to the variables that have extremely large values ($\gg 1$) [14]. To detect such kind of outliers under the Gamma distribution assumption, the β-divergence with $\beta \in [-1, 0]$ is used as a distance-like measure in [11]. See also Figure 1c.
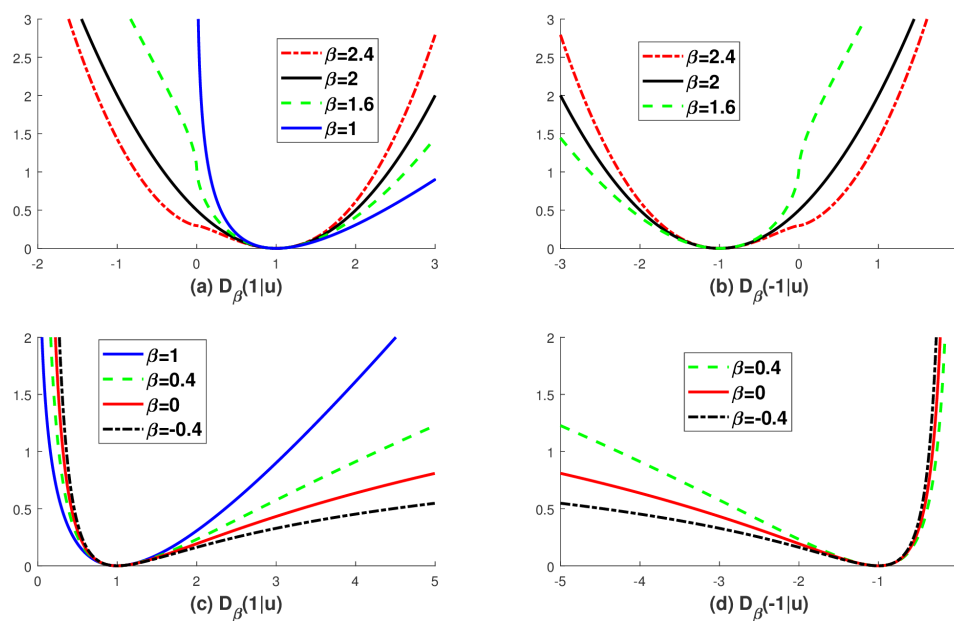
Figure 1. The graphs of the β-divergence $D_\beta(b|u)$, which is based on the proposed extended logarithmic function $\ln_\alpha(u)$ in (4). (**a**) and (**b**) shows $D_\beta(b|u)$ for $\beta \geq 1$ with different choice of b, i.e., $b = 1, -1$; (**c**) and (**d**) shows $D_\beta(b|u)$ for $\beta \leq 1$ with different choice of b i.e., $b = 1, -1$. Note that $D_\beta(b|u)$ with $\beta = 1$ is not defined if $u \in \mathbb{R}_{--}$.

In the case of SAR image data processing, speckle noise is modeled with the Gamma distribution and thus the negative log-likelihood function, which appears in speckle reduction problem, corresponds to the β-divergence with $\beta = 0$, i.e., the Itakura–Saito-divergence. Actually, this model is highly nonconvex [15]. Therefore, various transforms are introduced to relax nonconvexity of the Gamma distribution related speckle reduction model [16–21]. Recently, we have shown that the β-divergence with $\beta \in (0, 1)$ can be used as a transform-less convex relaxation model for SAR speckle reduction problem [12]. Generally, the data captured via a SAR system has extremely high dynamic range [22,23]. Under this harsh environment, β-divergence with $\beta \in (-1, 0)$ is successfully used as a similarity measure for separation of the strong scatterers in SAR data [11]. In addition, the β-divergence is also used for the decomposition of magnitude data of audio spectrograms [6]. In these applications, the domains of data are generally assumed to be positive. However, the domain of the β-divergence can be extended to the negative region. In fact, if $\beta = 2$, then the β-divergence is exactly the square of ℓ_2-distance, the domain of which naturally includes a negative region. Surprisingly, in this article, we show that, under the mild condition on β, there are infinitely many β-divergences that have a negative domain.

It is known that the β-divergence can be reformulated with the Bregman-divergence [2,6]. However, if we restrict the base function of the Bregman-divergence as the convex function of Legendre

type, then some part of the β-divergence cannot be expressed through the Bregman-divergence (see Table 1). Although the Bregman-divergence associated with a convex function of Legendre type does not exactly match with the β-divergence, due to the fruitful mathematical structure of the convex function of Legendre type, the associated Bregman-divergence has many useful properties. For instance, the dual formulation of the Bregman-divergence associated with the convex function of Legendre type can be used as a convex reformulation of some nonconvex problems under the certain condition on its domain [24]. In this article, we demonstrate that, by using the dual Bregman-divergence with the negative convex domain, we can make a convex reformulated Bregman variational model for the classic nonconvex problem that appears in the SAR image noise reduction problem [15]. We also show that we can unify the various variational models appearing in image processing problems as the Bregman variational model having sparsity constraints, e.g., total variation [25,26] (we called it Bregman-TV). Actually, the Bregman variational model corresponds to the right Bregman proximity operator [9,10]. See also [9,10,24,27–30] for theoretical analysis of the Bregman-divergence and related interesting properties of it.

Table 1. We compare the domain of the β-divergence and the domain of the Bregman-divergence associated with the convex function of Legendre type in (19). We note that the domain $\mathbb{R}_+^n \times \mathbb{R}_+^n$ ($\beta > 1$) and the domain $\mathbb{R}_{--}^n \times \mathbb{R}_{--}^n$ ($\beta = 0$) do not exist in the Bregman-divergence. *If we relax the Legendre condition of Φ as a convex and smooth function, then the Bregman-divergence D_Φ also exists in the region $\beta > 1$ with $dom\Phi = \mathbb{R}_+^n$.

Region	β-Divergence		Bregman-Divergence	
entire	$\beta > 1, \beta \in \mathbb{R}_e$	$\Omega_L = \Omega_R = \mathbb{R}^n$	$\beta > 1, \beta \in \mathbb{R}_e$	$dom\Phi = \mathbb{R}^n$
positive	$\boldsymbol{\beta > 1}$	$\Omega_L = \Omega_R = \mathbb{R}_+^n$	-*	-
	$0 < \beta \leq 1$	$\Omega_L = \mathbb{R}_+^n$ and $\Omega_R = \mathbb{R}_{++}^n$	$0 < \beta \leq 1$	$dom\Phi = \mathbb{R}_+^n$
	$\beta \leq 0$	$\Omega_L = \Omega_R = \mathbb{R}_{++}^n$	$\beta \leq 0$	$dom\Phi = \mathbb{R}_{++}^n$
negative	$0 < \beta < 1, \beta \in \mathbb{R}_e$	$\Omega_L = \mathbb{R}_-^n$ and $\Omega_R = \mathbb{R}_{--}^n$	$0 < \beta < 1, \beta \in \mathbb{R}_e$	$dom\Phi = \mathbb{R}_-^n$
	$\boldsymbol{\beta \leq 0}, \beta \in \mathbb{R}_e$	$\Omega_L = \Omega_R = \mathbb{R}_{--}^n$	$\boldsymbol{\beta < 0}, \beta \in \mathbb{R}_e$	$dom\Phi = \mathbb{R}_{--}^n$

1.1. Background

In this section, we review typical examples of the β-divergence, i.e., Itakura–Saito-divergence, Generalized Kullback–Leibler-divergence (I-divergence), and $norm^2$-distance. In addition, we introduce the Bregman-divergence and the corresponding Bregman variational model with sparsity constraints.

Let us start with the β-divergence $D_\beta : \Omega_L \times \Omega_R \to \mathbb{R}_+$ given by

$$D_\beta(b|u) = \langle \int_u^b \left(\frac{b - x}{x^{2-\beta}} \right) dx, \mathbf{1} \rangle, \tag{1}$$

where $\Omega_L \times \Omega_R = \{(b, u) \in \mathbb{R}^n \times \mathbb{R}^n \mid 0 \leq D_\beta(b|u) < \infty\}$ is the domain of the β-divergence. Actually, the domain $\Omega_L \times \Omega_R$ corresponds to the effective domain in optimization [3,31]. We call Ω_L and Ω_R as the left and right domains of the β-divergence. In addition, we assume that the left and right domains, Ω_L and Ω_R, are convex sets, respectively. That is, if $a, b \in \Omega_L$ (or Ω_R), then the line segment between two points also satisfies $\overline{ab} \in \Omega_L$ (or Ω_R). Note that $\langle a, d \rangle = \sum_{i=1}^n a_i d_i$, $a = (a_1, ..., a_n) \in \mathbb{R}^n$, $d = (d_1, ..., d_n) \in \mathbb{R}^n$, $\mathbf{1}$ is the all one vector in $\mathbb{R}^n$, $\mathbb{R}_+ = \{x \in \mathbb{R} \mid x \geq 0\}$, $\mathbb{R}_{++} = \{x \in \mathbb{R} \mid x > 0\}$, $\mathbb{R}_- = \{x \in \mathbb{R} \mid x \leq 0\}$, and $\mathbb{R}_{--} = \{x \in \mathbb{R} \mid x < 0\}$. In addition, integration, multiplication, and division are performed component-wise. Based on a selection of β, we can recover the famous representatives of the β-divergence, i.e., Itakura–Saito-divergence [4,5,13], I-divergence (or generalized Kullback–Leibler-divergence) [20,32], and $norm^2$-distance [25,26]. These three divergences are important examples of the β-divergence, since they show three different types of domains of the β-divergence. We summarize them in the following.

- Itakura–Saito-divergence ($\beta = 0$): $\Omega_L = \mathbb{R}^n_{++}$ and $\Omega_R = \mathbb{R}^n_{++}$:

$$D_\beta(b|u) = \langle \left(\frac{b}{u}\right) - \ln\left(\frac{b}{u}\right) - \mathbf{1}, \mathbf{1} \rangle.$$

Usually, the left and right domain, i.e., Ω_L and Ω_R of Itakura–Saito-divergence, are defined as positive and $\Omega_L = \Omega_R$ [12,13]. However, due to the scale invariance property of it, the variables b and u can be negative at the same time, even within the logarithmic function, i.e., $\Omega_L = \Omega_R = \mathbb{R}^n_{--}$. Based on this keen observation, in this article, we develop a new methodology that systematically detects a domain having the negative region. The Itakura–Saito-divergence is a typical example that can be expressed by the β-divergence and the Bregman-divergence at the same time. However, it has the negative domain in the β-divergence framework, but not in the Bregman-divergence framework (see Table 1).

- Generalized Kullback–Leibler-divergence (I-divergence) ($\beta = 1$): $\Omega_L = \mathbb{R}^n_+$ and $\Omega_R = \mathbb{R}^n_{++}$:

$$D_\beta(b|u) = \langle b \ln\left(\frac{b}{u}\right) - (b - u), \mathbf{1} \rangle,$$

where we naturally assume that $0 \ln 0 = 0$. Interestingly, it has different left and right domains, i.e., $\Omega_L \neq \Omega_R$. Due to the asymmetric structure of the domain of I-divergence, we need to carefully handle the β-divergence at the boundary of each domain. We categorize the class of the β-divergence that has the asymmetric domain structure in Section 2.

- $norm^2$-distance ($\beta = 2$): $\Omega_L = \mathbb{R}^n$ and $\Omega_R = \mathbb{R}^n$:

$$D_\beta(b|u) = \frac{1}{2}\|b - u\|^2_2.$$

This divergence is preferable to other divergences, since it has $\mathbb{R}^n$ as its domain for each variable. Unlike the previous two divergences, the domain of it naturally includes a negative region $\mathbb{R}^n_-$. Surprisingly, there are infinitely many β-divergences having $\mathbb{R}^n$ as its domain. We will show it in Section 2.

Additionally, we introduce the Bregman-divergence associated with the convex function of Legendre type [3]. The Bregman-divergence $D_\Phi : \Omega \times int(\Omega) \to \mathbb{R}_+$ is formulated as

$$D_\Phi(b|u) = \Phi(b) - \Phi(u) - \langle b - u, \nabla\Phi(u) \rangle, \tag{2}$$

where the base function Φ is the convex function of Legendre type [3], $\Omega = dom(\Phi) = \{x \in \mathbb{R}^n \mid \Phi(x) \in \mathbb{R}\}$, and $int(\Omega)$ is the interior of Ω. In fact, it is relatively interior of Ω, i.e., $ri(\Omega)$. Note that $ri(\Omega)$ is the interior of Ω relative to its affine hull, which is the smallest affine set including Ω. Therefore, the relative interior $ri(\Omega)$ coincides with the interior $int(\Omega)$ when the affine hull of Ω is $\mathbb{R}^n$. For more details, see Chapter 2.H in [31]. In this article, since the β-divergence (1) is separable in terms of dimension, the affine hull of Ω is always $\mathbb{R}^n$ and thus we simply use $int(\Omega)$ instead of $ri(\Omega)$. Note that the typical examples of the β-divergence in the above (Itakura–Saito-divergence, I-divergence, and $norm^2$-distance) can be reformulated with the Bregman-divergence (2) by using the convex function of Legendre type Φ and the associated domain Ω:

- Itakura–Saito-divergence: $\Phi(x) = \langle -\ln x, \mathbf{1} \rangle$ with $\Omega = \mathbb{R}^n_{++}$,
- I-divergence: $\Phi(x) = \langle x \ln x, \mathbf{1} \rangle$ with $\Omega = \mathbb{R}^n_+$,
- $norm^2$-distance: $\Phi(x) = \langle \frac{1}{2}x^2, \mathbf{1} \rangle$ with $\Omega = \mathbb{R}^n$.

The domain of the second variable of the Bregman-divergence (2) is always open set $int(dom\Phi)$. However, the right domain Ω_R of the second variable of the β-divergence (1) could be a closed set. In the coming section, we thoroughly analyze the relation between the Bregman-divergence

and the β-divergence with regard to its domain. Based on the Bregman-divergence (2), we introduce the *Bregman variational model* that unifies the various minimization problems appearing in image processing:

$$\min_u \{D_\Phi(b|u) + \lambda R(u) \mid u \in int(\Omega)\}, \tag{3}$$

where b is the observed data and $R(u)$ is the sparsity enforcing regularization term, such as total variation [26]. In image processing, (3) corresponds to the denoising problem under the various noise distributions: Poisson, Speckle, Gaussian noise, etc. However, in optimization, it is known as (nonconvex) right Bregman proximity operator under mild conditions. See [9,10,24,30] for more details on the Bregman operator.

1.2. Overview

The article is organized as follows. In Section 2, we analyze the structure of the domain of the β-divergence. In Section 3, we study various mathematical structures of the β-divergence through the Bregman-divergence associated with the convex function of Legendre type. In Section 4, we introduce the Bregman variational model and its dual formulation for convex reformulation of the classic nonconvex problem that appears in the SAR speckle reduction problem. In addition, we introduce the right and left Bregman proximal operator. We give our conclusions in Section 5.

2. A Characterization of the Domain $\Omega_L \times \Omega_R$ of the β-Divergence

In this section, we analyze the structure of the β-divergence and the associated domain $\Omega_L \times \Omega_R$ based on the so-called extended logarithmic function.

Let us start with a definition of the extended logarithmic function that is essential in characterizing the domain of the β-divergence. We note that it corresponds to an equivalence class of Tsallis's generalized logarithmic function [1,33] with an extention to the negative domain.

Definition 1. *Let $\alpha \in \mathbb{R}$, $u = (u_1, ..., u_n) \in dom(\ln_\alpha)$, and*

$$\ln_{\alpha,c}(u) = (\ln_{\alpha,c}(u_1), \ln_{\alpha,c}(u_2), ..., \ln_{\alpha,c}(u_n)),$$

where $\ln_{\alpha,c}(u_i) = \int_c^{u_i} x^{-\alpha} dx$ and $c \in \mathbb{R}_{c,u} = \{c \in \mathbb{R} \mid \ln_{\alpha,c}(u) \in \mathbb{R}^n \ \text{and} \ c \neq u_i, i = 1, ..., n\}$. Then, the extended logarithmic function is defined as an equivalence class

$$[\ln_\alpha(u)]_c = \{x \in \mathbb{R}^n \mid x = \ln_{\alpha,c}(u), \ c \in \mathbb{R}_{c,u}\}.$$

For simplicity, we leave out all constants after integration and then we attain

$$\ln_\alpha(u) = \begin{cases} \ln u, & \text{if } \alpha = 1, \\ \frac{1}{1-\alpha} u^{1-\alpha}, & \text{otherwise,} \end{cases} \tag{4}$$
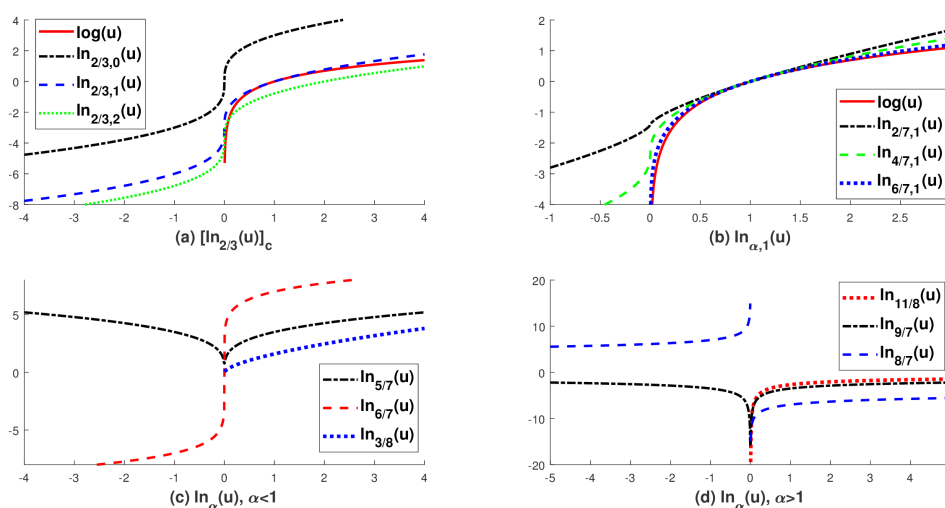
where $dom(\ln_\alpha) = \{x \in \mathbb{R}^n \mid \ln_\alpha(x) \in \mathbb{R}^n\}$. We call (4) as the extended logarithmic function instead of the equivalence class $[\ln_\alpha(u)]_c$, unless otherwise specified.

Note that the domain and range of $\ln_\alpha(u)$ (4) are given in Table 2. In addition, we illustrate the structure of the extended logarithmic function in Figure 2. As noticed in Definition 1, the extended logarithmic function is defined as an equivalence class $[\ln_\alpha(u)]_c$ with respect to c. If we set $c = 1$, then we can recover Tsallis' generalized logarithmic function [1,33] on its positive domain $dom(\ln_\alpha) \subseteq \mathbb{R}_+$. See Figure 2a. However, we cannot use the generalized logarithm (i.e., $\ln_{\alpha,c=1}(u)$) in the negative domain. In fact, if $\alpha > 1$ and $\mathbb{R}_- \subset dom(\ln_\alpha)$ then Tsallis' generalized logarithmic function is undefined, e.g., $\ln_{4,c=1}(-1) = \int_1^{-1} x^{-4} dx \notin \mathbb{R}$. On the other hand, the proposed extended logarithmic function (4) is well defined on $\mathbb{R}$ for all α, since we can choose an appropriate c having

the same sign with u even if $\alpha > 1$, e.g., $\ln_{4,c=-2}(-1) = \int_{-2}^{-1} x^{-4}dx \in \mathbb{R}$. See Figure 2d and Table 2. Indeed, the extended logarithmic function is useful when we simplify the complicated structure of the β-divergence. As described in the following Definition 2, the β-divergence is defined based on the difference of two extended logarithmic functions. In other words, the β-divergence is invariance with respect to a constant function in the extended logarithmic function (4). It is interesting that the Bregman-divergence (2) also has a similar invariance property with respect to an affine function in the base function Φ (see Proposition 1).

Table 2. The domain and range of the extended logarithmic function $\ln_\alpha(x)$ defined in (4).

	$\alpha = 1$	$\alpha < 1$			$\alpha > 1$		
		$\alpha \in \mathbb{R}_o$	$\alpha \in \mathbb{R}_e$	$\alpha \in \mathbb{R}_x$	$\alpha \in \mathbb{R}_o$	$\alpha \in \mathbb{R}_e$	$\alpha \in \mathbb{R}_x$
$dom\ (\ln_\alpha)$	$\mathbb{R}_{++}$	$\mathbb{R}$	$\mathbb{R}$	$\mathbb{R}_+$	$\mathbb{R}_{++}$ or $\mathbb{R}_{--}$	$\mathbb{R}_{++}$ or $\mathbb{R}_{--}$	$\mathbb{R}_{++}$
$range\ (\ln_\alpha)$	$\mathbb{R}$	$\mathbb{R}_+$	$\mathbb{R}$	$\mathbb{R}_+$	$\mathbb{R}_{--}$	$\mathbb{R}_{--}$ or $\mathbb{R}_{++}$	$\mathbb{R}_{--}$

Figure 2. The graphs of the extended logarithmic function in Definition 1. (**a**) shows an equivalence class $[\ln_\alpha(u)]_c = \{\ln_{\alpha,0}(u), \ln_{\alpha,1}(u), \ln_{\alpha,2}(u)\}$ with $\alpha = \frac{2}{3}$; (**b**) shows $\ln_{\alpha,1}(u) = \int_1^u \frac{1}{x^\alpha}dx$ with different choice of $\alpha = \frac{2}{7}, \frac{4}{7}, \frac{6}{7}$; (**c**) and (**d**) show $\ln_\alpha(u)$ in (4) for different choices of α. Note that $\ln_\alpha(u)$ is an extended logarithmic function without a constant term.

Definition 2. *Let* $(b, u) \in \Omega_L \times \Omega_R = \{(b, u) \in \mathbb{R}^n \times \mathbb{R}^n \mid D_\beta(b|u) \in \mathbb{R}_+\}$ *and* $\beta \in \mathbb{R}$. *Then, the* β-*divergence is defined by*

$$
\begin{aligned}
D_\beta(b|u) &= \langle b[\ln_{2-\beta}(b) - \ln_{2-\beta}(u)] - [\ln_{1-\beta}(b) - \ln_{1-\beta}(u)], \mathbf{1} \rangle \quad\quad (5)\\
&= \langle b \int_u^b u^{\beta-2}du - \int_u^b u^{\beta-1}du, \mathbf{1} \rangle.
\end{aligned}
$$

After integration, we get the well-known formula of β-*divergence:*

$$
D_\beta(b|u) = \begin{cases} \langle \left(\frac{b}{u}\right) - \ln\left(\frac{b}{u}\right) - \mathbf{1}, \mathbf{1} \rangle, & \text{if } \beta = 0, \\ \langle b \ln\left(\frac{b}{u}\right) - (b - u), \mathbf{1} \rangle, & \text{if } \beta = 1, \\ \langle \frac{b}{\beta-1}(b^{\beta-1} - u^{\beta-1}) - \frac{1}{\beta}(b^\beta - u^\beta), \mathbf{1} \rangle, & \text{if } \beta \neq 0, 1. \end{cases} \quad (6)
$$

Although the β-divergence has a unified formula (5) via the extended logarithm (4), unfortunately, the determination of the domain $\Omega_L \times \Omega_R$ of the β-divergence heavily depends on β. Before we go

any further, let us introduce the most important equivalence classes in this article. It will simplify complicated notations appearing in the β-divergence and the Bregman-divergence.

$$
\begin{aligned}
\mathbb{R}_e &= \{\, 2k/(2l+1) \mid k,l \in \mathbb{Z}, \,\} \\
\mathbb{R}_o &= \{\, (2k+1)/(2l+1) \mid k,l \in \mathbb{Z}, \,\} \\
\mathbb{R}_x &= \mathbb{R} \setminus \mathbb{R}_o \cup \mathbb{R}_e.
\end{aligned}
\tag{7}
$$

Note that $\mathbb{R}_e$ and $\mathbb{R}_o$ are subsets of the rational number and satisfy $\mathbb{R}_e \cap \mathbb{R}_o = \varnothing$, while $\mathbb{R}_x$ is composed of all irrational numbers with a subset of the rational number that are not in $\mathbb{R}_e \cup \mathbb{R}_o$. For instance, $\mathbb{R}_e = \{0, \pm\frac{2}{3}, \pm\frac{4}{3}, ...\}$, $\mathbb{R}_o = \{\pm 1, \pm\frac{1}{3}, \pm\frac{5}{3}, ...\}$ and $\mathbb{R}_x = \{\pm\frac{1}{4}, \pm\frac{1}{2}, \pm\sqrt{2}, ...\}$.

Since the β-divergence (6) is developed based on the extended logarithmic function (i.e., power functions), inherently, we have to quantify the domain of a power function $p(x) = x^\alpha$ and its inverse function $p^{-1}(y) = y^{1/\alpha}$. Actually, if x is positive, then a power function $p(x)$ and the corresponding inverse function is well defined, irrespective of the choice of an exponent $\alpha \in \mathbb{R} \setminus \{0\}$. On the other hand, in the case of negative domain, e.g., $x < 0$, the domain of a power function $p(x)$ severely depends on the choice of an exponent α. With newly introduced equivalence classes in (7), we can easily categorize the domain of a power function $p(x) = x^\alpha$ and its inverse function $p^{-1}(y) = y^{1/\alpha}$, $\alpha \neq 0$. We summarize it in the following Lemma.

Lemma 1. *Let $\alpha \in \mathbb{R}$ and $dom_-(p) = \{x \in \mathbb{R}_{--} \mid p(x) \in \mathbb{R}\} = dom\, p \cap \mathbb{R}_{--}$ be the negative domain of a power function $p(x) = x^\alpha$. Then, $p : dom_-(p) \to range_-(p)$ has its negative domain*

$$
dom_-(p) = \begin{cases} \mathbb{R}_{--}, & \text{if } \alpha \in \mathbb{R}_e \cup \mathbb{R}_o, \\ \varnothing, & \text{if } \alpha \in \mathbb{R}_x, \end{cases}
$$

and the corresponding range

$$
range_-(p) = \begin{cases} \mathbb{R}_{--}, & \text{if } \alpha \in \mathbb{R}_o, \\ \mathbb{R}_{++}, & \text{if } \alpha \in \mathbb{R}_e. \end{cases}
$$

In addition, if $\alpha \in \mathbb{R}_o$, then the inverse function of p is well defined and transparent on $dom_-(p)$, i.e., $p^{-1} \circ p(x) = x$ for all $x \in dom_-(p)$. However, if $\alpha \in \mathbb{R}_e \setminus \{0\}$, then $p^{-1} \circ p(x) \neq x$ for all $x \in dom_-(p)$.

Proof. For any $x \in \mathbb{R}_{--}$, let $x = (-1)|x|$, then the power function p is expressed as $p(x) = (-1)^\alpha |x|^\alpha$, $\forall \alpha \in \mathbb{R}$. We note that the negative real domain of $p(x)$ is well-defined only if $(-1)^\alpha \in \{-1, +1\}$. To clarify the evaluation of $(-1)^\alpha$, let us express it in a polar form:

$$
(-1)^\alpha = e^{i\alpha(2l+1)\pi}, \quad \forall l \in \mathbb{Z}.
\tag{8}
$$

Then, we get

$$
(-1)^\alpha = \begin{cases} 1, & \text{if } \alpha \in \mathbb{R}_e, \\ -1, & \text{if } \alpha \in \mathbb{R}_o, \\ \delta, & \text{if } \alpha \in \mathbb{R}_x, \end{cases}
$$

where $\delta \in \mathbb{C} \setminus \mathbb{R}$. Regarding the inverse function $p^{-1}(y) = y^{1/\alpha}$, we have $(-1)^{1/\alpha} \in \mathbb{C} \setminus \mathbb{R}$ for all $\alpha \in \mathbb{R}_e \setminus \{0\}$. However, if $\alpha \in \mathbb{R}_o$, then $(-1)^{1/\alpha} = -1$. That is, we have $p^{-1} \circ p(x) = x$, for all $x \in \mathbb{R}_{--}$. $\square$

Now, with the equivalence classes (7) and Lemma 1, we classify domains of the β-divergence. The details are given in the following Theorem and Table 3. See also Figure 1 for the overall structure of the β-divergence on its domain.

Table 3. A classification of the domain $\Omega_L \times \Omega_R = \{(b, u) \in \mathbb{R}^n \times \mathbb{R}^n \mid D_\beta(b|u) \in \mathbb{R}_+\}$ of the β-divergence in terms of β.

β	$\Omega_L \times \Omega_R$
$\beta > 1$ and $\beta \notin \mathbb{R}_e$	$\mathbb{R}^n_+ \times \mathbb{R}^n_+$
$\beta > 1$ and $\beta \in \mathbb{R}_e$	$\mathbb{R}^n \times \mathbb{R}^n$
$0 < \beta \leq 1$ and $\beta \notin \mathbb{R}_e$	$\mathbb{R}^n_+ \times \mathbb{R}^n_{++}$
$0 < \beta \leq 1$ and $\beta \in \mathbb{R}_e$	$\mathbb{R}^n_+ \times \mathbb{R}^n_{++}$ or $\mathbb{R}^n_- \times \mathbb{R}^n_{--}$
$\beta \leq 0$ and $\beta \notin \mathbb{R}_e$	$\mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$
$\beta \leq 0$ and $\beta \in \mathbb{R}_e$	$\mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$ or $\mathbb{R}^n_{--} \cup \mathbb{R}^n_{--}$

Theorem 1. *Let us consider the domain of the β-divergence*

$$\Omega_L \times \Omega_R = \{(b, u) \in \mathbb{R}^n \times \mathbb{R}^n \mid D_\beta(b|u) \in \mathbb{R}_+\},$$

where

$$D_\beta(b|u) = \langle \int_u^b (b - x)G(x)dx, \mathbf{1} \rangle$$

and $G(x) = x^{\beta-2} = (x_1^{\beta-2}, x_2^{\beta-2}, ..., x_n^{\beta-2})$. In addition, let us assume that the minimum value of $G(x)$ on its domain is nonnegative, i.e.,

$$0 \leq M_G = min_x\{G(x) \mid x_i \in [\min\{u_i, b_i\}, \max\{u_i, b_i\}], \; i = 1, ..., n \}, \tag{9}$$

where $(b, u) \in \Omega_L \times \Omega_R$. Then, the domain of the β-divergence $\Omega_L \times \Omega_R$ is classified as in Table 3. Note that $\Omega_R \subseteq \Omega_L$ for all $\beta \in \mathbb{R}$. In particular, if $\beta \in (0, 1]$, then $\Omega_R \subsetneq \Omega_L$.

Proof. Due to the assumption $M_G \geq 0$, we can easily obtain the positiveness of the β-divergence by the following inequality

$$D_\beta(b|u) = \langle \int_u^b (b - x)x^{\beta-2}dx, \mathbf{1} \rangle = \langle \int_u^b (b - x)G(x)dx, \mathbf{1} \rangle \geq \frac{1}{2}\|b - u\|_2^2 M_G \geq 0. \tag{10}$$

Consequently, we only need to fulfill the following two conditions: (1) Is the domain of the β-divergence determined to satisfy (9)? (2) Is the β-divergence well-defined on its domain?

- **Case 1**: $M_G \geq 0$, for all $(b, u) \in \Omega_L \times \Omega_R$.
  If $(b, u) \in \mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$ then it is trivial to show $M_G \geq 0$ and thus it is always true that $\mathbb{R}^n_{++} \times \mathbb{R}^n_{++} \subseteq \Omega_L \times \Omega_R$. Now, we will find β such that $\mathbb{R}_{--} \times \mathbb{R}_{--} \subset \Omega_L \times \Omega_R$ is satisfied. From Lemma 1, if $x_i < 0$ then, we get the following

$$G(x_i) = \begin{cases} |x_i|^{\beta-2}, & \text{if } \beta \in \mathbb{R}_e, \\ -|x_i|^{\beta-2}, & \text{if } \beta \in \mathbb{R}_o, \\ \varnothing, & \text{if } \beta \in \mathbb{R}_x, \end{cases}$$

  Based on $\beta \in \mathbb{R}$, we have two different cases regarding the domain $\Omega_L \times \Omega_R$.

  - If $\beta \notin \mathbb{R}_e$, then, due to (9), the negative region cannot be included into the domain of the β-divergence. Therefore, we have

  $$\Omega_L \times \Omega_R \subseteq \mathbb{R}^n_+ \times \mathbb{R}^n_+.$$

  - If $\beta \in \mathbb{R}_e$, then due to (9), the domain of the β-divergence can be defined to include negative region. That is, we have

  $$\mathbb{R}^n_{--} \times \mathbb{R}^n_{--} \subseteq \Omega_L \times \Omega_R.$$

- **Case 2**: $D_\beta(b|u) < +\infty$, for all $(b, u) \in \Omega_L \times \Omega_R$.
  Basically, the β-divergence can be expressed as

$$D_\beta(b|u) = \left\langle b \int_u^b x^{\beta-2} dx - \int_u^b x^{\beta-1} dx, \mathbf{1} \right\rangle.$$

That is, it is based on integrations of power functions of real variables u and b in $\mathbb{R}^n$. Therefore, we only need to see whether or not the integration in $D_\beta(b|u)$ is well defined at $\{\mathbf{0}\}$. We note that, after integration, the exponents of $b \in \Omega_L$ and $u \in \Omega_R$ are different and thus the corresponding domains Ω_R and Ω_L could be different as well. Hence, we should consider the following three different cases:

- $\boldsymbol{\beta > 1}$: We do not have any singularity at $\{\mathbf{0}\}$ with respect to $b \in \Omega_L$ and $u \in \Omega_R$. Therefore, we have $\{\mathbf{0}\} \subset \Omega_L = \Omega_R$.
- $\boldsymbol{0 < \beta \leq 1}$: After integration, $b \in \Omega_L$ does not have any singularity at $\{\mathbf{0}\}$. However, $u \in \Omega_R$ has a singularity at $\{\mathbf{0}\}$. Therefore, we have $\{\mathbf{0}\} \subset \Omega_L$ but $\{\mathbf{0}\} \not\subset \Omega_R$ and thus $\Omega_L \neq \Omega_R$ in this region.
- $\boldsymbol{\beta \leq 0}$: In this case, both $b \in \Omega_L$ and $u \in \Omega_R$ have singularity at $\{\mathbf{0}\}$. Thus, $\{\mathbf{0}\} \not\subset \Omega_L = \Omega_R$.

Based upon the analysis in Cases 1 and 2, we have six different choices of domain $\Omega_L \times \Omega_R$ for the β-divergence. It is summarized in Table 3 and illustrated in Figure 1. Since we only consider a convex domain, Ω_L and Ω_R should be selected as a convex set for each variable. In addition, due to the inherent integral formulation of β-divergence between b and u, the domain of both variables should be determined to have the same sign. $\quad\square$

As observed in Table 3, if $\beta \in \mathbb{R}_e \cap \mathbb{R}_-$, then there is a symmetry in the selection of the domain of the β-divergence. That is, $\Omega_L \times \Omega_R = \mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$ or $\mathbb{R}^n_{--} \times \mathbb{R}^n_{--}$. Especially, if $\beta = 0$, then the β-divergence corresponds to the Itakura–Saito-divergence $D_\beta(b|u) = \left\langle \left(\frac{b}{u}\right) - \log\left(\frac{b}{u}\right) - \mathbf{1}, \mathbf{1} \right\rangle$, where the domain of it can be $\mathbb{R}^n_{++} \times \mathbb{R}^n_{++}$ or $\mathbb{R}^n_{--} \times \mathbb{R}^n_{--}$. The positive domain is generally preferable, since it is related to real applications, e.g., intensity data type in the SAR system [11,12]. However, if we reformulate the β-divergence with the Bregman-divergence, then the negative domain commonly appears in the dual Bregman-divergence. In addition, we note that, due to Theorem 1, the β-divergence with the domain defined in Table 3 satisfies the following distance-like properties of the generic divergence [1,2]:

$$D_\beta(b|u) \in \mathbb{R}_+, \tag{11}$$

$$D_\beta(b|u) = 0 \text{ if and only if } b = u, \tag{12}$$

where (11) is followed from the Definition of the domain of the β-divergence and (10). Note that (12) is satisfied, if we restrict the domain of the β-divergence as $\Omega_R \times \Omega_R$. In fact, let us assume that $(b, u) \in \Omega_R \times \Omega_R$. It is trivial to show that $b = u \Rightarrow D_\beta(b|u)$. Therefore, we only need to show $D_\beta(b|u) = 0 \Rightarrow b = u$. Letting $b \neq u$, we then get $D_\beta(b|u) \geq \frac{1}{2}\|b - u\|_2^2 M_G > 0$ from (10).

Since the β-divergence (5) with its domain defined in Table 3 has the distance-like properties, (11) and (12), we can make a variational model with the β-divergence and a regularization term for smoothness constraint of the given data b. The following is an example of a variational model based on the β-divergence [12]

$$u_\beta^* = \arg\min_{u \in B} F_\beta(u) = D_\beta(b|u) + \lambda R(u), \tag{13}$$

where $\lambda > 0$ and $B \subseteq \Omega_R$ is a domain of F_β for a given $b \in B$. Note that B is an open convex set induced from the physical constraints of the observed data b. In the case of a prior $R(u)$, it can be a sparsity-enforcing function such as total variation (TV) $TV(u) = \|\nabla u\|_1$ [25,26] and frame [34]. We call (13) the β-*sparse model* or β-*TV* [12], if TV is used as a prior. Under the domain restriction in

Table 3, actually, we have lots of freedom in choosing $\beta \in \mathbb{R}$ of the β-sparse model (13). However, if we add additional constraints, such as convexity onto the β-divergence, then interestingly, the possible choice of β is dramatically reduced to a small set. For example, $D_\beta(b|u)$ with respect to u is convex on its domain only if $\beta \in [1, 2]$ [6]. Outside of this region, i.e., $\beta \in \mathbb{R} \setminus [1, 2]$, the convexity of it depends on the given data b [12]. In Section 4.1, we analyze the convexity of the β-divergence via the right Bregman proximity operator [28].

Although the proposed β-sparse model $F_\beta(u)$ in (13) is not convex in general, $F_\beta(u)$ has an interesting global optimum property [11,12], in case $\lambda = 0$. See also [35,36]. For completeness of the article, we add it below.

Theorem 2 ([11,36]). *For a given observed data $\{b^1, .., b^{|N|}\}$, let B be an open convex set in Ω_R, $\beta \in \mathbb{R}$, $b^j \in B$, and $\mu = \frac{1}{|N|} \sum_{j=1}^{|N|} b^j$. Then, we obtain that*

$$\sum_{j=1}^{|N|} D_\beta(b^j|u) \geq \sum_{j=1}^{|N|} D_\beta(b^j|\mu) \tag{14}$$

is always satisfied, regardless of the choice of $u \in B$.

Note that μ in (14) corresponds to the β-centroid in segmentation problem and is related to the Bregman centroid, which is extensively studied in [37]. In SAR image processing, if $\beta = 0$, then (14) corresponds to the multi-looking process, which is commonly used to reduce speckle noise in SAR data [12,22,23].

3. The Bregman-Divergence Associated with the Convex Function of Legendre Type for the β-Divergence

In this section, we study the Bregman-divergence associated with the convex function of Legendre type and its connection to the β-divergence. Although there is partial equivalence between two divergences, the Bregman-divergence has an important mathematical dual formulation. With a negative domain, the dual Bregman-divergence is unambiguously useful for convex reformulation of the nonconvex β-sparse model (13) (see Section 4).

Before we proceed further, let us first review the convex function of Legendre type (see Section 26 [3]). See also [27,29].

Definition 3. *Let $\Phi : \Omega \to \mathbb{R}$ be lower semicontinuous, convex, and proper function defined on $\Omega = dom\Phi \subseteq \mathbb{R}^n$. Then, Φ is the convex function of Legendre type (or Legendre), if the following conditions are satisfied (i.e., Φ is essentially strictly convex and essentially smooth):*

1. *$int(\Omega) \neq \emptyset$.*
2. *Φ is differentiable on $int(\Omega)$.*
3. *$\forall x \in bd(\Omega)$ and $\forall y \in int(\Omega)$,*

$$\lim_{t \downarrow 0} \langle \nabla\Phi(x + t(y - x)), y - x \rangle = -\infty.$$

4. *Φ is strictly convex on $int(\Omega)$.*

Here, $\Omega = dom\Phi = \{x \in \mathbb{R}^n \mid \Phi(u) \in \mathbb{R}\}$ is a convex set and $bd(\Omega)$ is the boundary of the domain Ω.

The main advantage of the convex function of Legendre type is that the inverse function of the gradient of it has an isomorphism with the gradient of its conjugate function as described below. This is a useful property when we characterize the dual structure of the Bregman-divergence associated with the convex function of Legendre type.

Theorem 3 ([3,27]). *Let $\Omega = dom\Phi$ and $\Omega^* = dom\Phi^* = \{x^* \in \mathbb{R}^n \mid \Phi^*(x^*) \in \mathbb{R}\}$. Then, the function Φ is the convex function of Legendre type if and only if its conjugate*

$$\Phi^*(x) = \sup_{\xi \in \Omega}\langle x, \xi\rangle - \Phi(\xi)$$

is the convex function of Legendre type. In this case, the gradient mapping

$$\nabla\Phi : int(\Omega) \to int(\Omega^*) \tag{15}$$

is an isomorphism with its inverse mapping $(\nabla\Phi)^{-1} = \nabla\Phi^*$.

For more details on Theorem 3, see Theorem 26.5 in [3] and Fact 2.9 in [27]. Let us assume that Φ be the convex function of Legendre type. Then, we can define the Bregman-divergence $D_\Phi : \Omega \times int(\Omega) \to \mathbb{R}_+$ associated with Legendre Φ:

$$D_\Phi(b|u) = \Phi(b) - \Phi(u) - \langle b - u, \nabla\Phi(u)\rangle, \tag{16}$$

where $b \in \Omega$ and $u \in int(\Omega)$. Several functions we are interested in are in a category of the convex function of Legendre type. For instance, Shannon entropy function $\Phi(x) = \langle x \log x, \mathbf{1}\rangle$ is a typical example of Legendre. the Bregman-divergence associated with it corresponds to the β-divergence with $\beta = 1$, i.e., Generalized Kullback–Leibler-divergence. We note that there is the convex function of Legendre type which does not have a corresponding β-divergence. For instance, Fermi–Dirac entropy function $\Phi(x) = \langle x \ln x + (1 - x)\ln(1 - x), \mathbf{1}\rangle$ is Legendre and the associated Bregman divergence is the Logistic loss function $D_\Phi(b|u) = \langle b \ln\left(\frac{b}{u}\right) + (1 - b)\ln\left(\frac{1-b}{1-u}\right), \mathbf{1}\rangle$. See [27,36] for more details on the Bregman-divergence. In the following, we summarize various useful features of the Bregman-divergence associated with Legendre Φ.

Theorem 4. *Let $\Omega = dom\Phi$ and $\Omega^* = dom\Phi^*$. Then, the Bregman-divergence associated with the convex function of Legendre type Φ satisfies the following.*

1. *$D_\Phi(b|u)$ is strictly convex with respect to b on $int(\Omega)$.*
2. *For any $u \in int(\Omega)$, $D_\Phi(b|u)$ is coercive with respect to b, i.e., $\lim_{\|b\|\to\infty} D_\Phi(b|u) \to \infty$.*
3. *For some $b \in int(\Omega)$, $D_\Phi(b|u)$ is coercive with respect to u if and only if $\Omega^* = int(\Omega^*)$.*
4. *$D_\Phi(b|u) = 0$ if and only if $b = u$, where $b \in int(\Omega)$*
5. *For any $b, u \in int(\Omega)$,*
$$D_\Phi(b|u) = D_{\Phi^*}(\nabla\Phi(u), \nabla\Phi(b)), \tag{17}$$

   where Φ^* is the conjugate function of Φ.
6. *For any $b^*, u^* \in int(\Omega^*)$,*

$$D_{\Phi^*}(b^*|u^*) = D_\Phi(\nabla\Phi^*(u^*), \nabla\Phi^*(b^*)).$$

Proof. Since Φ is the convex function of Legendre type, Φ is strictly convex on $int(\Omega)$. Hence, (1) is trivial. The proofs of (2)–(6) are in Theorem 3.7, Theorem 3.9 and Corollary 3.11 in [27]. $\square$

In the above Theorem, the dual formulation (17) is a unique feature of the Bregman-divergence with Legendre. Unfortunately, the β-divergence does not have a corresponding dual concept. Later, we will show how to use the dual Bregman-divergence (17) to make a convex reformulated model of the nonconvex β-sparse model (13). In addition, we note that the β-divergence (5) is established based on the extended logarithmic function (4), which is an equivalence class in terms of a constant function. Therefore, we can say that the β-divergence is invariant with respect to a constant function in the extended logarithmic function. Interestingly, the base function Φ of the Bregman-divergence also

has such kind of invariance property with respect to an affine function. For this, Φ does not need to be Legendre. However, for simplicity, we assume that Φ is Legendre. The details are following.

Proposition 1. *Let us define an equivalence class of the convex function of Legendre type Φ in terms of affine function as follows:*

$$[\Phi]_{\mathcal{A}} = \{\Phi'(x) \in \mathbb{R} \mid \Phi'(x) = \Phi(x) + \langle c, x \rangle + \langle d, \mathbf{1} \rangle, \; c, d \in \mathbb{R}^n\}, \tag{18}$$

where $\mathcal{A} = \{\langle c, x \rangle + \langle d, \mathbf{1} \rangle \mid c, d \in \mathbb{R}^n \}$. Then, the Bregman-divergence $D_{[\Phi]_{\mathcal{A}}}$ associated with an equivalence class $[\Phi]_{\mathcal{A}}$ is equal to the Bregman-divergence D_{Φ} associated with Φ, irrespective of the choice of an affine function $\mathcal{A}$.

Proof. We have the following equivalence with respect to an arbitrary affine function $\mathcal{A}$:

$$
\begin{aligned}
D_{[\Phi]_{\mathcal{A}}}(b|u) &= [\Phi(b)]_{\mathcal{A}} - [\Phi(u)]_{\mathcal{A}} - \langle b - u, \nabla[\Phi(u)]_{\mathcal{A}} \rangle \\
&= (\Phi(b) + \langle c, b \rangle + \langle d, \mathbf{1} \rangle) - (\Phi(u) + \langle c, u \rangle + \langle d, \mathbf{1} \rangle) - \langle b - u, \nabla\Phi(u) + c \rangle \\
&= \Phi(b) - \Phi(u) - \langle b - u, \nabla\Phi(u) \rangle \\
&= D_{\Phi}(b|u),
\end{aligned}
$$

where $(b, u) \in \Omega \times int(\Omega)$ and $\Omega = dom\Phi$. Therefore, Φ with an arbitrary affine function $\mathcal{A}$ (i.e., an equivalence class $[\Phi]_{\mathcal{A}}$ in terms of affine function $\mathcal{A}$) does not change the structure of the Bregman-divergence $D_{\Phi}(b|u)$ at all. $\square$

To connect the β-divergence and the Bregman-divergence associated with Legendre, we need to find a specialized convex function of Legendre type. Based on the comments in [2], we use an integral formula of the extended logarithmic function for the special convex function of Legendre type. Through this connection, we can reformulate the β-divergence into the Bregman-divergence associated with the convex function of Legendre type. The details are following.

Theorem 5. *Let $x \in \Omega \subseteq \mathbb{R}^n$ and*

$$\Phi(x) = \left\langle \int_{\mathbf{d}}^{x} \ln_{2-\beta}(t)dt, \mathbf{1} \right\rangle = \begin{cases} \langle -\ln x, \mathbf{1} \rangle, & \text{if } \beta = 0, \\ \langle x \ln x, \mathbf{1} \rangle, & \text{if } \beta = 1, \\ \langle \frac{1}{\beta(\beta-1)} x^{\beta}, \mathbf{1} \rangle, & \text{if } \beta \neq 0, 1, \end{cases} \tag{19}$$

where $\Omega = dom\Phi = \{x \in \mathbb{R}^n \mid \Phi(x) \in \mathbb{R}\}$, $\ln_{\alpha}(t)$ is the extended logarithmic function in (4), $\mathbf{d}$ is an arbitrary constant vector in $\mathbb{R}^n$ and be selected to satisfy the condition $\Phi(x) \in \mathbb{R}$. For simplicity, by Proposition 1, we dropped all affine function in $\Phi(x)$. Then, Φ in (19) is the convex function of Legendre type with the domain Ω given below:

$$\begin{cases} \text{I. entire region:} \\ \quad \beta > 1, \beta \in \mathbb{R}_e & \text{and} & \Omega = \mathbb{R}^n, \\ \text{II. positive region:} \\ \quad 0 < \beta \leq 1 & \text{and} & \Omega = \mathbb{R}_+^n, \\ \quad \beta \leq 0 & \text{and} & \Omega = \mathbb{R}_{++}^n, \\ \text{III. negative region:} \\ \quad 0 < \beta < 1, \beta \in \mathbb{R}_e & \text{and} & \Omega = \mathbb{R}_-^n, \\ \quad \beta < 0, \beta \in \mathbb{R}_e & \text{and} & \Omega = \mathbb{R}_{--}^n. \end{cases} \tag{20}$$

Proof. For simplicity, all affine functions are left out based on Proposition 1. In addition, it is trivial to show that $\Phi(x)$ is the Burg entropy ($\langle -\ln x, \mathbf{1} \rangle$ with $\Omega = \mathbb{R}_{++}^n$) if $\beta = 0$ and the Shannon entropy ($\langle x \ln x, \mathbf{1} \rangle$ with $\Omega = \mathbb{R}_+^n$) if $\beta = 1$. They are well-known examples of Legendre. As noticed

in (2), the corresponding Bregman-divergences are Itakura–Saito-divergence and Generalized Kullback–Leibler-divergence.

Now, we only need to check whether $\Phi(x) = \langle \frac{1}{\beta(\beta-1)} x^\beta, \mathbf{1} \rangle$ $(\beta \neq 0, 1)$ is Legendre or not. Among four conditions in Definition 3, it is trivial to show that $\Phi(x)$ satisfies the conditions 1 and 2. In the end, $\Phi(x)$ with $\beta \neq 0, 1$ and two Legendre conditions 3 and 4 are left.

I. Condition 3 in Definition 3:

Since $\Phi(x) = \langle \frac{1}{\beta(\beta-1)} x^\beta, \mathbf{1} \rangle$ $(\beta \neq 0, 1)$ is a power function, the only possible boundary of Φ is $\{\mathbf{0}\}$. Therefore, we search $\beta \in \mathbb{R} \setminus \{0, 1\}$ to find where condition 3 is satisfied. Since $\nabla\Phi(x) = \frac{1}{\beta-1} x^{\beta-1}$, we get the following at the potential boundary $\{\mathbf{0}\}$:

- $y \in \mathbb{R}^n_{++}$:

$$\lim_{t \downarrow 0} \langle \nabla\Phi(0 + ty), y \rangle = \begin{cases} -\infty & , \text{ if } \beta < 1 \text{ and } \beta \neq 0, \\ 0 & , \text{ if } \beta > 1. \end{cases}$$

- $y \in \mathbb{R}^n_{--}$:

$$\lim_{t \downarrow 0} \langle \nabla\Phi(0 + ty), y \rangle = \begin{cases} -\infty & , \text{ if } \beta < 1 \text{ and } \beta \in \mathbb{R}_e \setminus \{0\}, \\ +\infty & , \text{ if } \beta < 1 \text{ and } \beta \in \mathbb{R}_o, \\ 0 & , \text{ if } \beta > 1. \end{cases}$$

In summary, the following are the possible domains and the corresponding β for condition 3. We note that, if $\beta > 1$, then $\mathbf{0} \notin bd(dom\Phi)$:

- $\mathbb{R}^n_{++} \subseteq \Omega$ and $\beta \in (-\infty, 1) \setminus \{0\}$,
- $\mathbb{R}^n_{--} \subseteq \Omega$ and $\beta \in ((-\infty, 1) \setminus \{0\}) \cap \mathbb{R}_e$.

II. Condition 4 in Definition 3:

It can be easily checked by the fact that Φ is strictly convex on $int(\Omega)$ if and only if $\nabla\Phi$ is strictly monotone, that is, the following is satisfied [38] :

$$\langle \nabla\Phi(x) - \nabla\Phi(x'), x - x' \rangle > 0 \quad \text{whenever } x \neq x'.$$

Since $\nabla\Phi$ is separable in terms of dimension, we only need to show that $\nabla\Phi(x)$ is a strictly increasing function on $int(\Omega) \cap \mathbb{R}$. Note that if $\nabla^2\Phi(x) = x^{\beta-2} > 0$ on an open region, then Φ is strictly convex (i.e., $\nabla\Phi$ is strictly increasing) in that region:

- $\nabla^2\Phi(x) > 0$ if $\beta \in \mathbb{R} \setminus \{0, 1\}$ and $x \in \mathbb{R}^n_{++}$,
- $\nabla^2\Phi(x) > 0$ if $\beta \in (\mathbb{R} \setminus \{0, 1\}) \cap \mathbb{R}_e$ and $x \in \mathbb{R}^n_{--}$.

Note that at $\{\mathbf{0}\}$, we need to directly show that $\nabla\Phi$ is strictly increasing.

Now, we integrate the information in the above Legendre conditions 3 and 4 for the decision of the domain $\Omega = dom\Phi$ based on β. The details are the following:

- $\beta > 1$

  - $\beta \in \mathbb{R}_e : dom\Phi = \mathbb{R}^n$ and $\forall x \in \mathbb{R}^n_-$, $x^{\beta-1} = -|x|^{\beta-1}$. Therefore, $\nabla\Phi(x)$ is strictly increasing function on $int(dom\Phi)$ since $\nabla\Phi(x) = -\nabla\Phi(-x)$.
  - $\beta \in \mathbb{R}_o : dom\Phi = \mathbb{R}^n$ but Φ is an odd function with respect to zero and thus it is not a convex function.
  - $\beta \in \mathbb{R}_x : dom\Phi = \mathbb{R}^n_+$ but it does not satisfy condition 3 at $\mathbf{0}$.

- $0 < \beta < 1$

- $\beta \in \mathbb{R}_e$: $dom\Phi = \mathbb{R}^n$ but $dom\nabla\Phi = \mathbb{R}^n \setminus \{\mathbf{0}\}$ is not a convex set. Since $int(dom\Phi) \subset dom\nabla\Phi$ [3], to keep convexity of the domain $int(dom\Phi)$, we need to select $int(dom\Phi) = \mathbb{R}^n_{--}$ or $\mathbb{R}^n_{++}$. That is, we have $dom\Phi = \mathbb{R}^n_-$ or $dom\Phi = \mathbb{R}^n_+$. In both cases, we know that $\nabla\Phi$ is strictly increasing on $int(dom\Phi)$, since $\nabla^2\Phi = x^{\beta-2} > 0$, $\forall x \in int(dom\Phi)$.
- $\beta \in \mathbb{R}_o$: Following the case $\beta \in \mathbb{R}_e$, we have $dom\Phi = \mathbb{R}^n_+$ or $\mathbb{R}^n_-$. If $dom\Phi = \mathbb{R}^n_+$, then $\nabla\Phi$ is strictly increasing on $int(dom\Phi)$ ($\nabla^2\Phi > 0$). However, if $x < 0$, then $\nabla^2\Phi(x) < 0$ and thus it is not convex on its negative domain.
- $\beta \in \mathbb{R}_x$: $dom\Phi = \mathbb{R}^n_+$ and $\nabla\Phi$ is strictly increasing function on $int(dom\Phi)$ since $\nabla^2\Phi = x^{\beta-2} > 0$ for all $x \in int(dom\Phi)$.
- Condition 3 is satisfied on the above selected domain.

- $\beta < 0$

  - $\beta \in \mathbb{R}_e$: $dom\Phi = \mathbb{R}^n \setminus \{\mathbf{0}\}$ is not a convex set. Therefore, we need to restrict $dom\Phi$ as a convex set, i.e., $dom\Phi = \mathbb{R}^n_{--}$ or $dom\Phi = \mathbb{R}^n_{++}$. On both domains, $\nabla\Phi$ is strictly increasing ($\nabla^2\Phi > 0$, $\forall x \in dom\Phi$).
  - $\beta \in \mathbb{R}_o$: Following the case $\beta \in \mathbb{R}_e$, we have $dom\Phi = \mathbb{R}^n_{++}$ or $\mathbb{R}^n_{--}$. If $dom\Phi = \mathbb{R}^n_{++}$ then $\nabla\Phi$ is strictly increasing on $int(dom\Phi)$.
  - $\beta \in \mathbb{R}_x$: $dom\Phi = \mathbb{R}^n_{++}$ and $\nabla\Phi$ is strictly increasing function on $int(dom\Phi)$ since $\nabla^2\Phi = x^{\beta-2} > 0$ for all $x \in int(dom\Phi)$.
  - Condition 3 is satisfied on the above selected domain.

$\square$

Remark 1. *Note that Φ in* (19) *should be an equivalence class*

$$[\Phi(x)]_{\mathcal{A}} = \langle \int_{\mathbf{d}}^{x} [\ln_{2-\beta}(t)]_{\mathbf{c}} \, dt, \mathbf{1} \rangle$$

with an affine function $\mathcal{A} = \{\langle \alpha, x \rangle + \langle \gamma, \mathbf{1} \rangle \mid \alpha = f(\mathbf{c}), \gamma = g(\mathbf{c}, \mathbf{d})\}$. Here, $[\ln_{2-\beta}(t)]_{\mathbf{c}}$ is an equivalence class of an extended logarithmic function in (4). *As observed in* (18), *we have $D_{[\Phi]_{\mathcal{A}}}(b|u) = D_{\Phi}(b|u)$ for any affine function $\mathcal{A}$. Therefore, we can drop all affine function in $[\Phi]_{\mathcal{A}}$.*

Since Φ in (19) is Legendre under the domain condition (20), we can establish a new Bregman-divergence associated with Φ in (19). Interestingly, it corresponds to the β-divergence (1) [2]. However, there is a mismatch between the domain of the Bregman-divergence in (20) and the domain of the β-divergence in Table 3. We summarize it in Table 1. As a matter of fact, the positive domain with $\beta > 1$ is not defined in Φ (19) due to the Legendre condition. In addition, in the case of $\beta = 0$, the negative domain $\mathbb{R}^n_{--} \times \mathbb{R}^n_{--}$ is not defined in the Bregman-divergence with Φ (19). In the following Theorem, we show that under the restriction of the domain of the β-divergence to the domain of the Bregman-divergence, we can get an equivalence between the β-divergence and the Bregman-divergence associated with Legendre Φ (19).

Theorem 6. *Let us consider the β-divergence* (1) *and the Bregman-divergence* (16) *associated with Legendre Φ* (19). *If we restrict the domain of the β-divergence with the domain of the Bregman-divergence associated with Φ* (19), *which is $\Omega_L \times \Omega_R = dom\Phi \times int(dom\Phi)$ (see Table 1), then the β-divergence is equal to the Bregman-divergence associated with Legendre Φ* (19).

Proof. Since the domain of the β-divergence $\Omega_L \times \Omega_R$ is set up with $dom\Phi \times int(dom\Phi)$, the β-divergence is well defined with the restricted domain. In the following, we show an equivalence between the β-divergence and the Bregman-divergence under the domain condition of the Bregman-divergence:

$$
\begin{aligned}
D_\beta(b|u) &= \langle \int_u^b \frac{b-t}{t^{2-\beta}} dt, \mathbf{1} \rangle = \langle -\int_u^b t^{\beta-1} dt, \mathbf{1} \rangle + \langle b, \int_u^b t^{\beta-2} dt \rangle \\
&= \langle (1-\beta) \int_u^b \nabla\Phi(t) dt, \mathbf{1} \rangle + \langle b, \frac{1}{\beta-1}(b^{\beta-1} - u^{\beta-1}) \rangle \\
&= \Phi(b) - \Phi(u) - \frac{1}{\beta-1} u^{\beta-1}(b-u) \\
&= \Phi(b) - \Phi(u) - \langle b-u, \nabla\Phi(u) \rangle.
\end{aligned}
$$

Note that we do not use any $\nabla\Phi(b)$ information in the above derivation and thus the above equivalence is well satisfied within the domain of the Bregman-divergence associated with Legendre Φ (19). $\square$

In the following Theorem, we calculate the conjugate function Φ^* and the corresponding domain $dom\Phi^*$ of the convex function of Legendre type Φ defined in (19). The computation of the domain of Φ^* is useful in determining the structure of $D_\Phi(b|u)$. For instance, as noticed in Theorem 4 (3), if $dom\Phi^*$ is open, then the corresponding Bregman-divergence $D_\Phi(b|u)$ is coercive with respect to $u \in int(dom\Phi)$. Surprisingly, when $\beta \in [0,1)$, $D_\Phi(b|u)$ is not convex but coercive with respect to u. This fact is importantly used in SAR speckle reduction problems [12,15,21].

Theorem 7. *Let Φ (19) be the convex function of Legendre type with $dom\Phi$ (20). Then, Φ^*, the conjugate of Φ, and the corresponding domain $dom\Phi^*$ is calculated as follows:*

- $\beta = 0$ *and* $dom\Phi = \mathbb{R}_{++}^n$: $dom\Phi^* = \mathbb{R}_{--}^n$

$$
\Phi^*(x) = -\langle \mathbf{1} + \ln(-x), \mathbf{1} \rangle.
$$

- $\beta = 1$ *and* $dom\Phi = \mathbb{R}_+^n$: $dom\Phi^* = \mathbb{R}^n$

$$
\Phi^*(x) = \langle \exp(x - \mathbf{1}), \mathbf{1} \rangle.
$$

- $\beta \notin \{0,1\}$:

$$
\Phi^*(x) = \langle \frac{1}{\beta}((\beta-1)x)^{\frac{\beta}{\beta-1}}, \mathbf{1} \rangle.
$$

In this case, $dom\Phi^*$ depends on β:

$$
\begin{cases}
\beta > 1, \ \beta \in \mathbb{R}_e, & and \ dom\Phi = \mathbb{R}^n : dom\Phi^* = \mathbb{R}^n, \\
0 < \beta < 1, & and \ dom\Phi = \mathbb{R}_+^n : dom\Phi^* = \mathbb{R}_{--}^n, \\
0 < \beta < 1, \beta \in \mathbb{R}_e, & and \ dom\Phi = \mathbb{R}_-^n : dom\Phi^* = \mathbb{R}_{++}^n, \\
\beta < 0, & and \ dom\Phi = \mathbb{R}_{++}^n : dom\Phi^* = \mathbb{R}_-^n, \\
\beta < 0, \beta \in \mathbb{R}_e, & and \ dom\Phi = \mathbb{R}_{--}^n : dom\Phi^* = \mathbb{R}_+^n.
\end{cases}
\tag{21}
$$

Proof. Since Φ is Legendre, from Theorem 3, if $x \in int(dom\Phi^*)$, then we have

$$
\Phi^*(x) = \langle x, \nabla\Phi^{-1}(x) \rangle - \Phi(\nabla\Phi^{-1}(x)).
\tag{22}
$$

As noticed in (20), the domain of Φ (19) depends on β and thus the domain of its conjugate function Φ^* also depends on β. We categorize $dom\Phi^*$ below, by using (22):

- $\beta = 0 : \Phi(x) = \langle -\ln x, \mathbf{1} \rangle$ and $dom\Phi = \mathbb{R}_{++}^n$. From (22), the conjugate function Φ^* is calculated as

$$
\Phi^*(x) = \langle x\left(-\frac{1}{x}\right) + \ln\left(-\frac{1}{x}\right), \mathbf{1} \rangle = \langle -\mathbf{1} - \ln(-x), \mathbf{1} \rangle.
$$

Therefore, the domain of Φ^* becomes $dom\Phi^* = \mathbb{R}^n_{--}$.

- $\beta = 1 : \Phi(x) = \langle x\ln x, \mathbf{1}\rangle$ and $dom\Phi = \mathbb{R}^n_+$. From (22), the conjugate function Φ^* is calculated as

$$
\begin{aligned}
\Phi^*(x) &= \langle x\exp(x-1) - \exp(x-1)\ln(\exp(x-1)), \mathbf{1}\rangle \\
&= \langle x\exp(x-1) - (x-1)\exp(x-1), \mathbf{1}\rangle = \langle \exp(x-1), \mathbf{1}\rangle.
\end{aligned}
$$

It is trivial to show that $dom\Phi^* = \mathbb{R}^n$.

- $\beta \neq 0, 1 : \Phi(x) = \langle \frac{1}{\beta(\beta-1)}x^\beta, \mathbf{1}\rangle$ and $dom\Phi$ is given in (20). By simple calculation, we get

$$
\nabla\Phi^{-1}(x) = ((\beta-1)x)^{\frac{1}{(\beta-1)}}, \quad x \in int(dom\Phi^*). \tag{23}
$$

and from (22), the conjugate function Φ^* is derived as follows:

$$
\begin{aligned}
\Phi^*(x) &= \langle x((\beta-1)x)^{1/(\beta-1)} - \frac{1}{\beta(\beta-1)}((\beta-1)x)^{\beta/(\beta-1)}, \mathbf{1}\rangle \\
&= \langle \frac{1}{\beta}((\beta-1)x)^{\beta/(\beta-1)}, \mathbf{1}\rangle.
\end{aligned} \tag{24}
$$

Now, we need to decide the domain

$$
dom\Phi^* = \{x \in \mathbb{R}^n \mid \Phi^*(x) \in \mathbb{R}\}.
$$

While we identify $dom\Phi^*$, it should be selected based on the following isomorphism (in Theorem 3)

$$
\nabla\Phi : int(dom\Phi) \to int(dom\Phi^*),
$$

where $\nabla\Phi(x) = \frac{1}{\beta-1}x^{\beta-1} = (\nabla\Phi^*)^{-1}$ and the following estimation

$$
\beta/(\beta-1) \in \mathbb{R}_e, \quad \text{if } \beta \in \mathbb{R}_e.
$$

With the above information and the classification of $dom\Phi$ in (20), we are going to decide $dom\Phi^*$ based on β.

- $\beta > 1$ and $\beta \in \mathbb{R}_e$ and $dom\Phi = \mathbb{R}^n$:
  We have $\beta/(\beta-1) \in \mathbb{R}_e \cap \mathbb{R}_{++}$ and thus $dom\Phi^* = \mathbb{R}^n$. In addition, for all $x \in \mathbb{R}^n$, $\nabla\Phi(x) = \frac{1}{\beta-1}x^{\beta-1} \in \mathbb{R}^n$ is well defined.
- $0 < \beta < 1$ and $dom\Phi = \mathbb{R}^n_+$:
  In this case, $\beta/(\beta-1) < 0$ and $\beta - 1 < 0$. Therefore, for all $\beta \in (0,1)$, we have $dom\Phi^* = \mathbb{R}^n_{--}$. In addition, for all $x \in \mathbb{R}^n_{++}$, we have $\nabla\Phi(x) \in \mathbb{R}^n_{--}$. That is, the isomorphism between $int(dom\Phi)$ and $int(dom\Phi^*)$ is well defined.
- $0 < \beta < 1$ and $\beta \in \mathbb{R}_e$ and $dom\Phi = \mathbb{R}^n$:
  In this case, $\beta/(\beta-1) < 0$ and $\beta/(\beta-1) \in \mathbb{R}_e$ and $\beta - 1 < 0$. Therefore, the possible $dom\Phi^*$ is $\mathbb{R}^n_{--}$ or $\mathbb{R}^n_{++}$. However, we need to choose $dom\Phi^* = \mathbb{R}^n_{++}$ from the isomorphic mapping $\nabla\Phi : int(dom\Phi) \to int(dom\Phi^*)$.
- $\beta < 0$ and $dom\Phi = \mathbb{R}^n_{++}$:
  In this case, $0 < \beta/(\beta-1) < 1$. Therefore, we have $dom\Phi^* = \mathbb{R}^n_-$. Actually, this domain is well matched with the bijective mapping $\nabla\Phi(x) = \frac{1}{\beta-1}x^{\beta-1} < \mathbf{0}$ for all $x > \mathbf{0}$.
- $\beta < 0$ and $\beta \in \mathbb{R}_e$ and $dom\Phi = \mathbb{R}^n_{--}$:
  In this case, $\beta/(\beta-1) \in (0,1) \cap \mathbb{R}_e$. The possible domain is $dom\Phi^* = \mathbb{R}^n$. However, $dom\nabla\Phi^* = \mathbb{R}^n_{--} \cup \mathbb{R}^n_{++}$. From Theorem 23.4 in [3], $int(dom\Phi^*) \subset dom\nabla\Phi^*$. Due to the convexity constraint of the domain, $dom\Phi^* = \mathbb{R}^n_+$ or $dom\Phi^* = \mathbb{R}^n_-$. Through the isomorphic mapping $\nabla\Phi(x) = \frac{1}{\beta-1}x^{\beta-1} > \mathbf{0}, \forall x < \mathbf{0}$, we select $dom\Phi^* = \mathbb{R}^n_+$.

$\square$

In the following, the global minimization property of the β-divergence in Theorem 2 is reformulated with the Bregman-divergence. For more details, see [35,36].

Theorem 8 ([36]). *For all $b^i \in int(dom\Phi)$ with an index set $i \in N$, the following inequality is always satisfied, irrespective of the choice of $u \in int(dom\Phi)$:*

$$\sum_{i \in N} D_\Phi(b^i|u) \geq \sum_{i \in N} D_\Phi(b^i|\mu), \tag{25}$$

where $\mu = \frac{1}{|N|} \sum_{i \in N} b^i$ and $|N|$ is the cardinality of the set N.

Proof. Let us start with the generalized Pythagoras Theorem [35] of the Bregman-divergence:

$$D_\Phi(b|u) - D_\Phi(b|\mu) = D_\Phi(\mu|u) + \langle \nabla\Phi(\mu) - \nabla\Phi(u), b - \mu \rangle. \tag{26}$$

For all $b_i \in int(dom\Phi)$ with the index set N, let $\mu = \frac{1}{|N|} \sum_{i \in N} b^i$. Then, from (26), we get

$$\frac{1}{|N|} \sum_{i \in N} \left[D_\Phi(b^i|u) - D_\Phi(b^i|\mu) = D_\Phi(\mu|u) + \langle \nabla\Phi(\mu) - \nabla\Phi(u), b^i - \mu \rangle \right].$$

Since $D_\Phi(\mu|u) \geq 0$, we have

$$\sum_{i \in N} D_\Phi(b^i|u) \geq \sum_{i \in N} D_\Phi(b^i|\mu),$$

irrespective of the choice of $u \in int(dom\Phi)$. $\square$

Note that μ in (25) corresponds to the Bregman centroid, which is extensively studied in [37]. Now, we are ready to jump into the variational model having the Bregman-divergence as its fitting term. Many various important variational models induced from the statistical distribution are in this category.

4. Bregman Variational Model—Bregman-TV

In this section, we study the β-sparse model (13) with TV regularization via Bregman divergence (16) associated with Legendre (19) under the domain condition in (20). First, we introduce Bregman proximity operators in Section 4.1 and then we demonstrate how to use dual Bregman-divergence with the negative domain for convex reformulation of the nonconvex β-TV model [12] in Section 4.2.

The image data in general is observed in 2D array and have a limited dynamic range, due to physical constraints of the image capturing system. Therefore, let us assume that the observed image data is bounded and also column-wise stacked. That is, $b \in B \subset \mathbb{R}^n_+$, where B is an open and bounded convex set. Now, we start with the following Bregman variational model with total variation, i.e., Bregman-TV.

$$\min_{u \in B} D_\Phi(b|Lu) + \lambda TV(u), \tag{27}$$

where $TV(u) = \sum_{i=1}^n \|\nabla u_i\|$ is a typical sparsity constraint in image processing, $L : \mathbb{R}^n_+ \to \mathbb{R}^n_+$ is a linear mapping, $\lambda > 0$, and $B \subseteq int(dom\Phi) \subset \mathbb{R}^n_+$. Although the domain B is nonnegative in real applications, through the dual formulation of the Bregman-divergence (17), the nonpositive domain is very common and sometimes is useful for convex reformulation of nonconvex variational models appearing in SAR image enhancement problems. See Theorem 7 for the negative domain of the conjugate function Φ^*.

We note that L is a matrix with nonnegative entries and it is designed based on various applications of image processing, e.g., for the image deblurring problem, L is a blur or a convolution matrix; for an image inpainting problem, L is a binary mask matrix; for an image denoising problem, L is an identity matrix. See [39,40] for more details on image denoising, deblurring, and inpainting problems with total variation or other sparsity constraints such as wavelet frames. The following are typical

examples of the Bregman-TV induced from the various physical noise sources, e.g., Gaussian, Poisson, and Speckle noise:

- $\beta = 2$: Image restoration problems (e.g., denoising, deblurring, inpainting) under Gaussian noise [25,26,41]

$$\min_{u \in B} \frac{1}{2} \|b - Lu\|_2^2 + \lambda TV(u). \tag{28}$$

- $\beta = 1$: Image restoration problems (e.g., denoising, deblurring, inpainting) under Poisson noise [32]

$$\min_{u \in B} \langle Lu - b \ln (Lu), \mathbf{1} \rangle + \lambda TV(u). \tag{29}$$

- $\beta = 0$: Image restoration problems (e.g., denoising, deblurring, inpainting) under Gamma multiplicative noise (or speckle noise) [15]

$$\min_{u \in B} \langle \left(\frac{b}{Lu} \right) + \ln (Lu), \mathbf{1} \rangle + \lambda TV(u). \tag{30}$$

- $\beta \in (0,1)$: A convex relaxed model [12] for the above SAR image restoration model (30). Additionally, this region is related to the compound Poisson distribution [8]

$$\min_{u \in V} \langle \frac{b(Lu)^{\beta-1}}{1 - \beta} + \frac{(Lu)^\beta}{\beta}, \mathbf{1} \rangle + \lambda TV(u). \tag{31}$$

For the remainder of this article, we only consider the image denoising problem ($L = I$). It is also known as the (nonconvex) right Bregman proximity operator [10].

4.1. Bregman Proximity Operators

In this section, we introduce the right and left Bregman proximity operator [9,10,30] based on the Bregman-divergence associated with Φ (19). In this section, let us assume that Φ (19) is convex, smooth, and (dimensionally) separable function (not necessarily Legendre). That is, the Bregman-divergence $D_\Phi(b|u)$ associated with Φ (19) also exists in the positive domain $dom\Phi = \mathbb{R}_+^n$ with $\beta > 1$. See Table 1.

We note that the Bregman-divergence $D_\Phi(b|u)$ associated with Φ (19) is strictly convex with respect to b (see Theorem 4). On the other hand, convexity of $D_\Phi(b|u)$ with respect to u strongly depends on the observed data b and β in Φ (19). Based on [9,12,28], we present three different convexities of the Bregman-divergence associated with Φ. Let $\Omega = dom\Phi$. Then, we have the following:

- The Bregman-divergence D_Φ (16) is *jointly convex* if $D_\Phi(b|u)$ is convex with respect to (b, u) on $int(\Omega) \times int(\Omega)$.
- The Bregman-divergence D_Φ (16) is *separately convex* if $D_\Phi(b|u)$ is convex with respect to $u \in int(\Omega)$ for all $b \in int(\Omega)$.
- The Bregman-divergence D_Φ (16) is *conditionally convex* if $D_\Phi(b|u)$ is convex with respect to $u \in B$ for all $b \in B$, where $B \subseteq int(\Omega)$ is an open convex set and depends on b.

We note that the *conditional convexity* is first introduced in this article based on the previous analysis of the β-TV model [12]. The reason we are interested in conditional convexity is that, in real applications, the dynamic range of the observed data is very limited. For instance, the observed image data via an optical camera have 8-bit resolution (i.e., $b \in [0, 2^8]$) [41] and the intensity level of the backscattered radar signal in SAR system is 32-bit resolution (i.e., $b \in [0, 2^{32}]$) at most [21]. Therefore, it is natural to consider convexity depending on the given data b.

The following Theorem, mostly based on Theorem 3.3 in [28], is useful in characterizing convexity of the Bregman-divergence $D_\Phi(b|u)$ associated with Φ (19). We note that $Diag(A)$ is a vector with diagonal entries of a matrix (or tensor) A. Also, a function f is concave if and only if $f(\alpha u + (1 - \alpha)v) \geq \alpha f(u) + (1 - \alpha)f(v)$, for all $\alpha \in [0, 1]$.

Theorem 9. *Let $D_\Phi(b|u)$ be the Bregman-divergence associated with convex, smooth, and (dimensionally) separable function Φ. In addition, we assume that $h = Diag(\nabla^2\Phi) > \mathbf{0}$, then we have the following useful criterion for the convexity of $D_\Phi(b|u)$. Here, $\Omega = dom\Phi$.*

(i) *$D_\Phi(b|u)$ is jointly convex if and only if $1/h$ is concave. Note that, since $h = (h_1, ..., h_n)$ is (dimensionally) separable, $1/h$ is defined as $1/h(u) = (1/h_1(u), ..., 1/h_n(u))$, it is concave if and only if h satisfy the following inequality:*

$$h(u) + \nabla h(u)(u - b) \geq \frac{h^2(u)}{h(b)}, \quad \forall b, u \in int(\Omega). \tag{32}$$

Moreover, if $\nabla^2 h$ exists, then $D_\Phi(b|u)$ is jointly convex if and only if

$$h(u)Diag(\nabla^2 h(u)) \geq 2(Diag(\nabla h))^2. \tag{33}$$

(ii) *$D_\Phi(b|u)$ is separately convex if and only if*

$$h(u) + \nabla h(u)(u - b) \geq \mathbf{0}, \quad \forall b, u \in int(\Omega). \tag{34}$$

(iii) *$D_\Phi(b|u)$ is conditionally convex if and only if*

$$h(u) + \nabla h(u)(u - b) \geq \mathbf{0}, \quad \forall b, u \in B, \tag{35}$$

where B is an open convex set in $int(\Omega)$ and depends on b [12,42].

Proof. The proof of the first two convexities are given in Theorem 3.3 in [28]. For the conditional convexity, let us take second derivatives of $D_\Phi(b|u)$ with respect to u. Then, we get

$$h(u) + \nabla h(u)(u - b) \geq \mathbf{0}.$$

For each $b \in B_b \subset int(dom\Phi)$, we can find the domain of $u \in B_u \subset int(dom\Phi)$ satisfying the above condition. Let B be a convex and open set satisfying $B \subseteq B_b \cap B_u$. Then, we have the conditional convexity condition in (35). □

The following Theorem shows an interesting result that $D_\Phi(b|u)$ associated with Φ (19) is convex on its whole domain with respect to u in a very limited region $\beta \in [1, 2]$. From a statistical point of view, this region is a little bit curious. In fact, if $\beta \in (1, 2)$ then the Bregman-divergence $D_\Phi(b|u)$ does not have the corresponding statistical Tweedie distribution [8].

Theorem 10. *Let Φ be a convex and smooth function in (19) (not necessarily Legendre). Then, $D_\Phi(b|u)$ is separately convex (and also jointly convex) with the following domain conditions:*

$$int(\Omega) = \begin{cases} \mathbb{R}^n_{++}, & if \ \beta \in [1, 2), \\ \mathbb{R}^n_{--}, & if \ \beta \in [1, 2) \cap \mathbb{R}_e, \\ \mathbb{R}^n, & if \ \beta = 2. \end{cases} \tag{36}$$

Due to the physical constraints of the observed data b, if we further restrict the domain of b, then we have conditional convexity of $D_\Phi(b|u)$. Let $\alpha = (\beta - 2)/(\beta - 1)$, if $\beta \neq 1$ and $B^\pm_{m/M}$ be a constant vector in $\mathbb{R}^n$ representing $B^\pm_{m/M}\mathbf{1}$. Then, we have the following:

• **Case I**: *Let us assume that the given data b be positive and have the following limitation in measurement*

$$b \in int(\Omega^+_b) = \{b \in \mathbb{R}^n_{++} \mid b \in (B^+_m, B^+_M)\}. \tag{37}$$

Then, $D_\Phi(b|u)$ is conditionally convex on B, which is given below:

- $\beta \in (2, +\infty) : B = \{u \in \mathbb{R}^n_{++} \mid u \in (max(B^+_m, \alpha B^+_M), B^+_M),\}$ *where* $\alpha \in (0, 1)$.
- $\beta \in (-\infty, 1) : B = \{u \in \mathbb{R}^n_{++} \mid u \in (B^+_m, min(\alpha B^+_m, B^+_M)),\}$ *where* $\alpha \in (1, +\infty)$.

- **Case II**: *Let us assume that the given data b is negative and has the following limitation in measurement*

$$b \in int(\Omega^-_b) = \{b \in \mathbb{R}^n_{--} \mid b \in (B^-_m, B^-_M)\}. \tag{38}$$

   Then, $D_\Phi(b|u)$ *is conditionally convex on B, which is given below:*

- $\beta \in (2, +\infty) \cap \mathbb{R}_e :$ $B = \{u \in \mathbb{R}^n_{--} \mid u \in (B^-_m, min(\alpha B^-_m, B^-_M)),\}$ *where* $\alpha \in (0, 1)$.
- $\beta \in (-\infty, 1) \cap \mathbb{R}_e \setminus \{0\} : B = \{u \in \mathbb{R}^n_{--} \mid u \in (max(B^-_m, \alpha B^-_M), B^-_M),\}$ *where* $\alpha \in (1, 2) \cup (2, +\infty)$.

Proof. Since Φ in (19) is sufficiently smooth, we use (34) with $h(u) = u^{\beta-2}$. To find separately convex region, we need to find β satisfying

$$h(u) + \nabla h(u)(u - b) = u^{\beta-3}[(\beta - 1)u - (\beta - 2)b] \geq \mathbf{0}, \quad \forall b, u \in int(\Omega), \tag{39}$$

where we need to decide the corresponding domain Ω based on Table 1. In the case of conditional convexity, let us assume that the domain of b is limited as (37) or (38). In the following, we summarize the separate and conditional convexity of D_Φ.

- $int(\Omega) = \mathbb{R}^n$ with $\beta > 1$ and $\beta \in \mathbb{R}_e$:
  We simplify (39) as

$$u \geq 0 \text{ and } u \geq \alpha b \quad or \quad u \leq 0 \text{ and } u \leq \alpha b.$$

  Then, we have

  - $\beta = 2 : \alpha = 0$ and thus u does not depend on b in (39). In fact, for any $b \in int(\Omega)$, we can select arbitrary $u \in int(\Omega)$.
  - $\beta \neq 2$: In this case, the domain of u depends on b. For instance, if $\beta = 3$ and $b > 0$, then the domain of u bounded below, i.e., $u \geq 0.5b > 0$. Therefore, the domain of u cannot be the whole region $int(\Omega) = \mathbb{R}^n$. This restriction is related to conditional convexity of D_Φ and is given in the following cases; $int(\Omega) = \mathbb{R}^n_{++}$ ($\beta \in \mathbb{R}$) and $\mathbb{R}^n_{--}$ ($\beta \in \mathbb{R}_e \setminus \{0\}$).

- $int(\Omega) = \mathbb{R}^n_{++}$ and $\beta \in \mathbb{R}$:
  We simplify (39) as

$$(\beta - 1)u \geq (\beta - 2)b. \tag{40}$$

  - $\beta > 1$: (40) is simplified as $u \geq \alpha b$ and we get

    * $\beta \in (1, 2)$: Since $\alpha \in (-\infty, 0)$, $u \geq \alpha b$ is satisfied for all $u, b \in int(\Omega) = \mathbb{R}^n_{++}$.
    * $\beta > 2$: Let us assume that the given data b is bounded, i.e., (37) is satisfied. From $u \geq \alpha b$ and $\alpha \in (0, 1)$, we have to satisfy the condition $u \geq \alpha B^+_M$ and thus the restricted domain B corresponding to (37) is given as

$$B = \{u \in \mathbb{R}^n_{++} \mid u \in (max(B^+_m, \alpha B^+_M), B^+_M)\}.$$

      In fact, for all $b \in B$, $D_\Phi(b|u)$ is convex in terms of $u \in B$.
  - $\beta = 1$: (40) is satisfied for all $u, b \in int(\Omega) = \mathbb{R}^n_{++}$.
  - $\beta < 1$: Let us assume that b satisfies (37). Then, from $u \leq \alpha b$ with $\alpha \in (1, +\infty)$, we get $u \leq \alpha B^+_m$ and thus the restricted domain B corresponding to (37) is given as

$$B = \{u \in \mathbb{R}^n_{++} \mid u \in (B^+_m, min(\alpha B^+_m, B^+_M))\}.$$

- $int(\Omega) = \mathbb{R}^n_{--}$ and $\beta \in \mathbb{R}_e \setminus \{0\}$:
  We simplify (39) as

$$(\beta - 1)u \leq (\beta - 2)b. \tag{41}$$

    – $\beta > 1$: (41) is simplified as $u \leq \alpha b$ and we get

          * $\beta \in (1, 2)$: $\alpha \in (-\infty, 0)$ and $u \leq \alpha b$ is satisfied for all $u, b \in int(\Omega) = \mathbb{R}^n_{--}$
          * $\beta > 2$: Let us assume that b satisfies (38). Then, from $u \leq \alpha b$ with $\alpha \in (0, 1)$, we get $u \leq \alpha B^-_m$ and thus the restricted domain B corresponding to (38) is given as

$$B = \{u \in \mathbb{R}^n_{--} \mid u \in (B^-_m, min(\alpha B^-_m, B^-_M))\}.$$

    – $\beta < 1$: Let us assume that b satisfy (38). Then, from $u \geq \alpha b$ with $\alpha \in (1, 2) \cup (2, +\infty)$, we get $u \geq \alpha B^-_M$ and thus the restricted domain B corresponding to (38) is given as

$$B = \{u \in \mathbb{R}^n_{--} \mid u \in (max(B^-_m, \alpha B^-_M), B^-_M)\}.$$

In addition, we note that it is not difficult to see that $D_\Phi(b|u)$ is jointly convex on its domain $int(\Omega) \times int(\Omega)$, where $int(\Omega)$ is given in (36). From the joint convexity condition in (33), we get the following condition

$$(\beta - 2)(\beta - 1)u^{2(\beta-3)} \leq 0.$$

Since the exponent of u is even, i.e., $u^{2(\beta-3)} \geq 0$ for all $u \in int(\Omega)$ (36), if $\beta \in [1, 2]$, then $(\beta - 2)(\beta - 1) \leq 0$ and thus $D_\Phi(b|u)$ is jointly convex under the domain constraints in (36). $\square$

Based on joint (and separate) convexity of D_Φ associated with Φ (19) on the domain (36), we can define the right Bregman proximity operator $\overrightarrow{P}_{\lambda TV} : int(\Omega) \to int(\Omega)$ associated with Φ (19) as follows:

$$\overrightarrow{P}_{\lambda TV}(b) = \underset{u \in int(\Omega)}{\arg\min} \, D_\Phi(b|u) + \lambda TV(u), \qquad (42)$$

where $D_\Phi(b|u)$ associated with Φ (19) is strictly convex, smooth, and coercive with respect to u and total variation is also convex and a coercive function [40]. We note that $D_\Phi(b|u)$ with the domain $int(\Omega)$ (36) is coercive with respect to u, due to the joint convexity condition in (32) and $\frac{h^2(u)}{h(b)} > 0$. For more details on the right Bregman proximal operator, see [10,24]. We should be cautious that, although the right Bregman proximal operator $\overrightarrow{P}_{\lambda TV}(b)$ (42) is well defined for the given data $b \in int(\Omega)$, its usefulness in real applications is limited, due to the separately convex condition $\beta \in [1, 2]$. Actually, in the case of $\beta = 2$, it is just an ordinary proximal operator [9]. Note that, in real applications such as SAR [12], $D_\Phi(b|u)$ with $\beta \in [0, 1)$ and the positive domain constraints (37) (i.e., $D_\Phi(b|u)$ is conditionally convex) is used. In this case, it corresponds to not convex but coercive right Bregman proximity operator. See Theorem 4 (3) for the coercivity of the operator on its domain $int(\Omega)$.

Now, let us consider Φ (19) with the domain condition in Table 1. Then, $D_\Phi(b|u)$ is coercive and strictly convex in terms of b (see Theorem 4). Hence, we can also define the left Bregman proximity operator $\overleftarrow{P}_{\lambda TV} : int(\Omega) \to int(\Omega)$ as follows:

$$\overleftarrow{P}_{\lambda TV}(u) = \underset{b \in int(\Omega)}{\arg\min} \, D_\Phi(b|u) + \lambda TV(b). \qquad (43)$$

Unlike the right Bregman proximity operator, the left Bregman proximity operator (43) associated with the base function Φ in (19) is strictly convex and coercive for all $\beta \in \mathbb{R}$ on its domain $int(\Omega)$, where $\Omega = dom\Phi$. We note that the left Bregman proximity operator can be characterized in a more simple way as

$$\overleftarrow{P}_{\lambda TV} = (\nabla \Phi + \lambda \partial TV)^{-1} \circ \nabla \Phi,$$

where ∂TV is a subgradient of TV. Actually, $\overleftarrow{P}_{\lambda TV}$ is a maximal monotone operator. See [30,43] for more details on Bregman proximity operator and the corresponding Bregman–Moreau envelopes.

Remark 2. *We could use β-divergence to define proximity operators. For instance, the right β-divergence proximity operator can be defined as*

$$\overset{\rightarrow\beta}{P}_{\lambda TV}(b) = \arg\min_{u \in \Omega_R} D_\beta(b|u) + \lambda TV(u). \tag{44}$$

Instead of TV in (44), *if we use an indicator function*

$$\iota_S(u) = \begin{cases} 0 & \text{, if } u \in S, \\ +\infty & \text{, if } u \notin S, \end{cases}$$

for a convex set S, then we get the right β-divergence projection operator for S as follows:

$$\overset{\rightarrow\beta}{P}_{\iota_S}(b) = \arg\min_{u \in \Omega_R} D_\beta(b|u) + \iota_S(u). \tag{45}$$

It is interesting that the robustness of β-divergence [7] *can be explained through the right β-divergence projection operator* (45)*. Let us assume that b, u in* (45) *are probability distributions (i.e., $\sum_i b_i = 1$ and $\sum_i u_i = 1$) and S is a set of Gaussian distributions. Here, the notation is slightly abused, since the Gaussian distribution is a continuous probability distribution and it is not a convex set. We note that the (generalized) Kullback–Leibler (KL) divergence (i.e., $D_\beta(b|u)$ with $\beta = 1$), which is a commonly used similarity measure between two probability distributions, is undefined at zero probability ($u = 0$). See Figure 1a and Table 1. However, outliers (i.e., rare events) have extremely low probability and thus they exist near zero probability. In this case, KL-divergence amplify the value near zero, i.e., $\lim_{u \to 0} D_{\beta=1}(b|u) = +\infty$. However, when $\beta > 1$, as noticed in Figure 1a and Table 1, $\lim_{u \to 0} D_{\beta>1}(b|u) < +\infty$. Thus, outliers which exist near zero are not weighted too much. Hence, the right β-divergence projection operator* (45) *with $\beta > 1$ is more robust to outliers than the KL-divergence-based operator. For more details, see* [4,5,7]*. Note that we can also define the left β-divergence proximity operator as*

$$\overset{\leftarrow\beta}{P}_{\lambda TV}(u) = \arg\min_{b \in \Omega_L} D_\beta(b|u) + \lambda TV(b). \tag{46}$$

4.2. Dual Bregman-Divergence-Based Left Bregman Operator for a Convex Reformulation of the Bregman-TV with $\beta < 1$

In this section, we introduce a convex reformulation of the nonconvex Bregman-TV (27) ($\beta < 1$ and $L = I$) associated with Φ (19), which is the convex function of Legendre type and its domain is given in Table 1. Note that the problems we study in this section are related to the speckle reduction problem [12,15].

Due to the Theorem 4, we have the following reformulated Bregman-TV:

$$\min_{u \in B} D_{\Phi^*}(\nabla\Phi(u)|\nabla\Phi(b)) + \lambda TV(u), \tag{47}$$

where $B = \{x \in \mathbb{R}^n_{++} \mid x \in (B_m^+, B_M^+)\}$ with $B_m^+ \geq \epsilon\mathbf{1}$ for some $\epsilon > 0$. See [12,42] for real SAR data processing applications where the box constraint B is a critical element of the performance. Now, let $w = \nabla\Phi(u)$ and the corresponding domain $\nabla\Phi(B)$, then we have the left Bregman proximity operator associated with the dual Bregman-divergence:

$$\overset{\leftarrow}{P}_{\lambda TV \circ \nabla\Phi^*}(\nabla\Phi(b)) = \arg\min_{w \in \nabla\Phi(B)} F(w) = D_{\Phi^*}(w|\nabla\Phi(b)) + \lambda TV(\nabla\Phi^*(w)), \tag{48}$$

where $\nabla\Phi(B) \subseteq int(dom\Phi^*)$ and we use $\nabla\Phi \circ \nabla\Phi^*(w) = w$. Since $\beta < 1$, $\nabla\Phi = \frac{1}{\beta-1}x^{\beta-1}$ is strictly increasing on its domain $int(dom\Phi) = \mathbb{R}^n_{++}$. Therefore, we have the transformed domain defined on $\mathbb{R}^n_{--}$ as

$$\nabla\Phi(B) = \{x \in \mathbb{R}^n_{--} \mid x \in (\nabla\Phi(B_m^+), \nabla\Phi(B_M^+))\}, \tag{49}$$

where $-\infty < \nabla\Phi(\epsilon) \leq \nabla\Phi(B_m^+) < \nabla\Phi(B_M^+) < 0$ and it is also a convex set. Moreover, $\nabla\Phi^*$ is also a strictly convex function by the following Lemma.

Lemma 2. *Let us assume that $\beta < 1$ and $int(dom\Phi) = \mathbb{R}^n_{++}$. Then, we have*

$$\nabla\Phi^*(w) = [(\beta-1)w]^{1/(\beta-1)},$$

where $w \in int(dom\Phi^*) = \mathbb{R}^n_{--}$. Note that $\nabla\Phi^*$ is strictly increasing and strictly convex on its domain $int(dom\Phi^*)$. However, it is not coercive but bounded below. In fact, we have $\nabla\Phi^*(w) \geq 0$, $\forall w \in int(dom\Phi^*)$.

Proof. Let $f(w) = \nabla\Phi^*(w)$. Then, since $\beta < 1$ and $w \in \mathbb{R}^n_{--}$, we have

$$\nabla^2 f(w) = (2-\beta)[(\beta-1)w]^{\frac{3-2\beta}{\beta-1}} > 0.$$

Since Φ^* is Legendre, $\nabla\Phi^*$ is strictly increasing on its domain. We note that, although $\nabla\Phi^*$ is strictly convex and strictly increasing, it is not coercive but bounded below, i.e.,

$$\lim_{w\to-\infty} \nabla\Phi^*(w) = 0. \tag{50}$$

$\square$

Finally, by using strict convexity and strictly increasing property of $\nabla\Phi^*$, we have a unique solution of the minimization problem in (48).

Theorem 11. *Let Φ (19) be the convex function of Legendre type and its domain is given in Table 1. In addition, we assume that $\beta < 1$ and $B = \{x \in \mathbb{R}^n_{++} \mid x \in (B_m^+, B_M^+)\}$. Then, for the given data $b \in B$, the left Bregman proximity operator (48) associated with the dual Bregman-divergence is well-defined. That is, there is a unique solution $w^* = \arg\min_{w \in \nabla\Phi(B)} F(w)$.*

Proof. Since Φ (19) is Legendre, Φ^* is also Legendre on the domain $\mathbb{R}^n_{--}(= int(dom\Phi^*))$. Therefore, $D_{\Phi^*}(w|x)$ is strictly convex and coercive in terms of w by Theorem 4 (2). In addition, TV is a composition of $\|\cdot\|_1 \circ \mathcal{D}$, where $\mathcal{D}$ is a linear matrix (i.e, first order difference matrix) [44] and $\|a\|_1 = \sum_i |a_i|$. Hence, we have

$$TV(\nabla\Phi^*(w)) = \|\cdot\|_1 \circ \mathcal{D} \circ \nabla\Phi^*(w),$$

where $\nabla\Phi^*(w)$ is strictly convex (Lemma 2) and $\mathcal{D}$ is a linear operator and $\|\cdot\|_1$ is a simple metric (convex and increasing). Therefore, $TV(\nabla\Phi^*(w))$ is also a convex function (see Section IV.2.1 in [38]). In addition, since $\nabla\Phi^*(w)$ is lower bounded (Lemma 2), $TV(\nabla\Phi^*(w))$ is also lower bound. Then, the objective function $F(w)$ in (48) is coercive (see Lemma 2.12 in [10]) and strictly convex. In the end, the left Bregman proximity operator associated with dual Bregman-divergence has an unique solution (see Proposition 3.5 in [10]) as

$$w^* = \overleftarrow{P}_{TV\circ\nabla\Phi^*}(\nabla\Phi(b)),$$

where $b \in B$ and $w^* = \arg\min_{w \in \nabla\Phi(B)} F(w)$ in (48). Regarding the domain $\nabla\Phi(B)$, since Φ^* is Legendre, $|\nabla\Phi^*(x)| \to \infty$ as $|x| \to 0$. Therefore, we need to keep a distance from $\{0\}$ to assure

$|\nabla\Phi^*(x)| < +\infty$. In fact, as noticed in (49), the transformed domain $\nabla\Phi(B)$ is a convex set and away from $\{\mathbf{0}\}$. $\square$

The above Theorem is quite surprising. We can get a unique solution of the nonconvex Bregman-TV (27) ($\beta < 1$ and $L = I$) through the left Bregman proximity operator (48) with an additional isomorphic transformation mapping $\nabla\Phi^* \circ \nabla\Phi = I$ as

$$u^* \in \nabla\Phi^* \circ \overleftarrow{P}_{(\lambda TV \circ \nabla\Phi^*)} \circ \nabla\Phi(b). \tag{51}$$

However, in general, due to the severe nonlinearity of $\nabla\Phi^*$ within the non-smooth regularizer, i.e., $TV(\nabla\Phi^*(w))$, it is not easy to design a stable numerical algorithm to find a solution u^* in (51). To overcome this drawback, we can directly modify (47) with a constraint $w = \nabla\Phi(u)$ as

$$\min_{w,u} D_{\Phi^*}(w|\nabla\Phi(b)) + \lambda TV(u), \tag{52}$$

subject to the following constraints:

$$\begin{cases} u \in B, \\ w \in \nabla\Phi(B), \\ w = \nabla\Phi(u) \text{ (or } u = \nabla\Phi^*(w)). \end{cases} \tag{53}$$

Since $w = \nabla\Phi(u)$ is a nonlinear constraint and thus we cannot directly apply highly sophisticated augmented Lagrangian-based optimization algorithm. As a heuristic, to remedy these nonlinear constraints, we may consider the following penalty method [45]:

$$(w^{(k+1)}, u^{(k+1)}) = \arg\min_{w \in \nabla\Phi(B), u \in B} D_{\Phi^*}(w|\nabla\Phi(b)) + \tfrac{\rho}{2}\|w - \nabla\Phi(u^{(k)})\|^2 + \tfrac{\tau}{2}\|u - \nabla\Phi^*(w^{(k)})\|^2 + \lambda TV(u). \tag{54}$$

This model is convex in terms of w and u, respectively. However, it is not convex with respect to (u, w). In case of speckle reduction problems (30), nonlinearity of $\nabla\Phi$ could be reduced by using a shifting technique in [42].

In the following example, we show how (51) can be applied to relax nonconvex speckle reduction problems (30) with $L = I$.

Example 1. *Let us consider the following nonconvex minimization problems. For a given $b \in B$,*

$$\min_{u \in B} D_\Phi(b|u) + \lambda TV(u), \tag{55}$$

where $B = \{x \in \mathbb{R}^n_{++} \mid x \in (B^+_m, B^+_M)\}$ is an open convex set with $B^+_m \geq \epsilon\mathbf{1}$ for some $\epsilon > 0$. Note that $\Phi(u) = \langle -\ln u, \mathbf{1}\rangle$ is the convex function of Legendre type ($\beta = 0$) and

$$D_\Phi(b|u) = \left\langle \left(\frac{b}{u}\right) - \ln\left(\frac{b}{u}\right) - \mathbf{1}, \mathbf{1}\right\rangle$$

is the Bregman-divergence associated with Φ (Burg entropy) function. This model is known as AA-model [15]. It is well known that it is not easy to find a global minimizer of (55), due to the severe nonconvexity of $D_\Phi(b|u)$ in terms of u [15,21]. Therefore, various transform-based convex relaxation approaches are introduced [12,16–21,42]. In this example, we are going to use dual Bregman-divergence to find a solution of (55). We note that $\Phi(u) = \langle -\ln u, \mathbf{1}\rangle$ is the convex function of Legendre type on its domain $int(dom\Phi) = \mathbb{R}^n_{++}$. Hence, by Theorem 7, we get the following corresponding conjugate function:

$$\Phi^*(w) = \langle -\mathbf{1} - \ln(-w), \mathbf{1}\rangle.$$

This function is also the convex function of Legendre type on its domain $int(dom\Phi^*) = \mathbb{R}^n_{--}$. Now, by using the dual Bregman-divergence, we get a convex reformulated version of (55) as

$$
\begin{cases}
w^* = \arg\min_{w \in B_w} F(w), \\
u^* = -\frac{1}{w^*},
\end{cases}
\tag{56}
$$

where

$$
F(w) = -\langle \ln(-w), \mathbf{1} \rangle - \langle w, b \rangle + TV\left(-\frac{\mathbf{1}}{w}\right)
$$

and

$$
B_w = \left\{ x \in \mathbb{R}^n_{--} \mid x \in \left(-\frac{\mathbf{1}}{B^+_m}, -\frac{\mathbf{1}}{B^+_M}\right) \right\}.
$$

We note that $B_w = \nabla\Phi(B)$ is a convex set and $TV(-\frac{1}{w})$ is also convex for all $w \in \mathbb{R}^n_{--}$. Therefore, the objective function $F(w)$ in (56) is strictly convex on its domain B_w. In addition, due to the Theorem 4 (2), $F(w)$ is coercive in the domain $\mathbb{R}^n_{--}$. Therefore, we have a unique solution u^* of (55). A similar inverse transformation on the positive domain $\mathbb{R}^n_{++}$ itself is introduced in [45,46].

5. Conclusions

In this article, we introduced the extended logarithmic function and, based on that, we could redefine the domain of the β-divergence. In fact, we have found that if β is in the class $\mathbb{R}_e = \{x \in \mathbb{R} \mid x = 2k/(2l+1), k, l \in \mathbb{Z}\}$, then the negative region $\mathbb{R}^n_{--}$ should be included into the domain of the β-divergence. In addition, if we use the integral of the extended logarithmic function as a base function of the Bregman-divergence, then we have a partial equivalence between the β-divergence and the Bregman-divergence associated with the Legendre base function. Last but not least, by using dual formulation of the Bregman-divergence associated with convex function of Legendre type and the negative domain of it, we have shown that we could make a convex reformulated model of the nonconvex variational model that appears in the SAR speckle reduction problem. The approaches in this article could be extended to other divergences, such as α- and γ-divergences [2]. In addition, we could plug the presented model into various segmentation problems [11,47,48].

Conflicts of Interest: The author declares no conflict of interest.

References

1. Amari, S.; Nagaoka, H. *Methods of Information Geometry*; American Mathematical Society: Washington, DC, USA, 2000.
2. Cichocki, A.; Amari, S. Families of Alpha- Beta- and Gamma- divergences: Flexible and robust measures of similarities. *Entropy* **2010**, *12*, 1532–1568.
3. Rockafellar, R.T. *Convex Analysis*; Princeton University Press: Princeton, NJ, USA, 1970.
4. Basu, A.; Harris, I.R.; Hjort, N.L.; Jones, M.C. Robust and efficient estimation by minimizing a density power divergence. *Biometrika* **1998**, *85*, 549–559.
5. Eguchi, S.; Kano, Y. Robustifying Maximum Likelihood Estimation. Available online: https://www.researchgate.net/profile/Shinto_Eguchi/publication/228561230_Robustifing_maximum_likelihood_estimation_by_psi-divergence/links/545d65910cf2c1a63bfa63e6/Robustifing-maximum-likelihood-estimation-by-psi-divergence.pdf (accessed on 7 September 2017).
6. Fevotte, C.; Idier, J. Algorithm for Nonnegative Matrix Factorization with the beta-divergence. *Neural Comput.* **2011**, *23*, 2421–2456.

7. Samek, W.; Blythe, D.; Müller, K.-R.; Kawanabe, M. Robust Spatial Filtering with Beta Divergence. In Proceedings of the 27th Annual Conference on Neural Information Processing Systems 2013, Lake Tahoe, NV, USA, 5–10 December 2013; pp. 1007–10015.

8. Jorgensen, B. *The Theory of Dispersion Models*; Chapman & Hall: London, UK, 1997.

9. Bauschke, H.H.; Combettes, P.L. Iterating Bregman Retractions. *SIAM J. Optim.* **2003**, *13*, 1159–1173.

10. Bauschke, H.H.; Combettes, P.L.; Noll, D. Joint minimization with alternating Bregman proximity operators. *Pac. J. Optim.* **2006**, *2*, 401–424.

11. Woo, H. Beta-divergence based two-phase segmentation model for synthetic aperture radar images. *Electron. Lett.* **2016**, *52*, 1721–1723.

12. Woo, H.; Ha, J. Besta-divergence-based variational model for speckle reduction. *IEEE Signal Proc. Lett.* **2016**, *23*, 1557–1561.

13. Fevotte, C.; Bertin, N.; Durrieu, J.-L. Nonnegative Matrix Factorization with the Itakura–Saito Divergence: with application to music analysis. *Neural Comput.* **2009**, *21*, 793–830.

14. Lobry, S.; Denis, L.; Tupin, F. Multitemporal SAR image decomposition into strong scatterers, background, and speckle. *IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens.* **2016**, *9*, 3419–3429.

15. Aubert, G.; Aujol, J.F. A variational approach to remove multiplicative noise. *SIAM J. Appl. Math.* **2008**, *68*, 925–946.

16. Bioucas-Dias, J.M.; Figueiredo, M.A.T. Multiplicative noise removal using variable splitting and constrained optimization. *IEEE Trans. Image Process.* **2010**, *19*, 1720–1730.

17. Huang, Y.M.; Ng, M.K.; Wen, Y.W. A new total variation method for multiplicative noise removal. *SIAM J. Imaging Sci.* **2009**, *2*, 20–40.

18. Kang, M.; Yun, S.; Woo, H. Two-level convex relaxed variational model for multiplicative denoising. *SIAM J. Imaging Sci.* **2011**, *6*, 875–903.

19. Shi, J.; Osher, S. A nonlinear inverse scale space method for a convex multiplicative noise model. *SIAM J. Imaging Sci.* **2008**, *1*, 294–321.

20. Steidl, G.; Teuber, T. Removing multiplicative noise by Douglas-Rachford splitting. *J. Math Imaging Vis.* **2010**, *36*, 168–184.

21. Yun, S.; Woo, H. A new multiplicative denoising variational model based on m-th root transformation. *IEEE Trans. Image Process.* **2012**, *21*, 2523–2533.

22. Bamler, R. Principles of synthetic aperture radar. *Surv. Geophys.* **2000**, *21*, 147–157.

23. Oliver, C.; Quegan, S. *Understanding Synthetic Aperture Radar Imaging*; SciTech Publishing: Raleigh, NC, USA, 2004.

24. Bauschke, H.H.; Wang, X.; Ye, J.; Yuan, X. Bregman distances and Chebyshev sets. *J. Approx. Theory* **2009**, *159*, 3–25.

25. Chambolle, A. An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **2004**, *20*, 89–97.

26. Rudin, L.; Osher, S.; Fatemi, E. Nonlinear total variation based noise removal algorithms. *Phys. D* **1992**, *60*, 259–268.

27. Bauschke, H.H.; Borwein, J.M. Legendre functions and the method of random Bregman projections. *J. Convex Anal.* **1997**, *4*, 27–67.

28. Bauschke, H.H.; Borwein, J.M. Joint and separate convexity of the Bregman distance. In *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*; Burnariou, D., Censor, Y., Reich, S., Eds.; Elsevier: Amsterdam, The Netherlands, 2001.

29. Bauschke, H.H.; Borwein, J.M.; Combettes, P.L. Essential smoothness, essential strict convexity and Legendre functions in Banach spaces. *Commun. Contemp. Math.* **2001**, *3*, 615–647.

30. Bauschke, H.H.; Dao, M.N.; Lindstrom, S.B. Regularizing with Bregman-Moreau envelopes. *arXiv* **2017**, arXiv:1705.06019v1.

31. Rockafellar, R.T.; Wets, R.J.-B. *Variational Analysis*; Springer: New York, NY, USA, 1998.

32. Setzer, S.; Steidl, G.; Teuber, T. Deblurring Poissonian images by split Bregman techniques. *J. Vis. Commun. Image Represent.* **2010**, *21*, 193–199.

33. Ashok, M.; Sundaresan, R. Minimization problems based on relative α-entropy II: Reverse projection. *IEEE Trans. Inf. Theory* **2015**, *61*, 5081–5095.

34. Durand, S.; Fadili, J.; Nikolova, M. Multiplicative noise removal using L1 fidelity on frame coefficients. *J. Math Imaging Vis.* **2010**, *36*, 201–226.

35. Teboulle, M. A unified continuous optimization framework for center-based clustering methods. *J. Mach. Learn. Res.* **2007**, *8*, 65–102.

36. Banerjee, A.; Merugu, S.; Dhillon, I.S.; Ghosh, J. Clustering with Bregman Divergences. *J. Mach. Learn. Res.* **2005**, *6*, 1705–1749.

37. Nielsen, F.; Nock, R. Sided and symmetrized Bregman centroids. *IEEE Trans. Inf. Theory* **2009**, *55*, 2882–2904.

38. Hiriart-Urruty, J.-B.; Lemarechal, C. *Convex Analysis and Minimization Algorithms: Part 1: Fundamentals*; Springer: New York, NT, USA, 1996.

39. Chan, T.F.; Shen, J. *Image Processing and Analysis*; SIAM: Philadelphia, PA, USA, 2005.

40. Aubert, G.; Kornprobst, P. *Mathematical Problems in Image Processing*; Springer: New York, NY, USA, 2006.

41. Yun, S.; Woo, H. Linearized proximal alternating minimization algorithm for motion deblurring by nonlocal regularization. *Pattern Recognit.* **2011**, *44*, 1312–1326.

42. Woo, H.; Yun, S. Alternating minimization algorithm for speckle reduction with a shifting technique. *IEEE Tran. Image Process.* **2012**, *21*, 1701–1714.

43. Kan, C.; Song, W. The Moreau envelope function and proximal mapping in the sense of the Bregman distance. *Nonlinear Anal.* **2012**, *75*, 1385–1399.

44. Micchelli, C.A.; Shen, L.; Xu, Y. Proximity algorithms for image models: denoising. *Inverse Probl.* **2011**, *27*, 045009.

45. Nie, X.; Qiao, H.; Zhang, B. A Variational model for PolSAR data speckle reduction based on the Wishart distribution. *IEEE Trans. Image Process.* **2015**, *24*, 1209–1222.

46. Oh, A.K.; Willett, R.M. Regularized Non-Gaussian Image Denoising. *arXiv* **2015**, arXiv:1508.02971v1.

47. Chan, T.F.; Vese, L. Active contours without edges. *IEEE Trans. Image Process.* **2001**, *10*, 266–277.

48. Chan, T.F.; Esedoglu, S.; Nikolova, M. Algorithms for finding global minimizers of image segmentation and denoising models. *SIAM J. Appl. Math.* **2006**, *66*, 1632–1648.