INSTITUT DE FRANCE
Académie des sciences

Comptes Rendus
Mécanique

Stéphane Andrieux

Bregman divergences for physically informed discrepancy measures for learning and computation in thermomechanics

Volume 351 (2023), p. 59-81

Published online: 1 February 2023

https://doi.org/10.5802/crmeca.164

CENTRE MERSENNE

Les Comptes Rendus. Mécanique *sont membres du*
Centre Mersenne pour l'édition scientifique ouverte
www.centre-mersenne.org
e-ISSN : 1873-7234

Comptes Rendus
Mécanique
2023, Vol. 351, p. 59-81
https://doi.org/10.5802/crmeca.164

Short paper / *Note*

Bregman divergences for physically informed discrepancy measures for learning and computation in thermomechanics

Mesures de dissimilarité par divergences de Bregman pour le calcul et l'apprentissage en thermomécanique

Stéphane Andrieux[a]

[a] ONERA, Université Paris-Saclay, Chemin de la Hunière, 91123 Palaiseau, France
E-mail: stephane.andrieux@onera.fr

Abstract. With view on the context of convex thermomechanics, we propose tools based on the concept of Bregman divergence, a notion introduced in the 1960s and used in learning and optimization as well. This study is motivated by the need of "discrepancy measures" between physically constrained fields that are used both in traditional algorithms, analysis methods and data driven modelling or applications as well.

We give also a characterization of symmetrical Bregman divergences through their generating functions which can only be quadratic forms. Some properties of the Bregman divergence and the introduced concept of Bregman Gap between couples of dual quantities are given, and some existing errors in thermomechanics are recovered. Exploiting the framework of Standard Generalized Materials, we give a set of generating functions for a large range of applications, including coupled multi-physics. Finally some results useful for computational geometry are detailed.

Résumé. Dans le contexte de la thermomécanique convexe, nous proposons des outils basés sur le concept de divergence de Bregman, une notion introduite dans les années 1960 et utilisée dans d'autres domaines aussi bien en apprentissage qu'en optimisation. Cette étude est motivée par le besoin de « mesures de dissimilarité » entre des champs physiquement contraints, champs qui sont utilisés à la fois dans les algorithmes traditionnels, les méthodes d'analyse et les applications ou les modélisations basées sur les données.

Nous donnons également une caractérisation des divergences de Bregman symétriques à travers leur fonction génératrice qui ne peut être qu'une fonction quadratique. Certaines propriétés de la divergence de Bregman et du concept introduit d'écart de Bregman entre les couples de quantités duales sont données, et certaines erreurs existantes en thermomécanique sont retrouvées. En exploitant le cadre des matériaux standard généralisés, nous donnons un ensemble de fonctions génératrices pour une large gamme d'applications, y compris la multiphysique couplée. Enfin, certains résultats utiles pour la géométrie computationnelle sont détaillés.

Keywords. Thermomechanics, Convexity, Bregman divergences, Discrepancy measures, Data processing.

Mots-clés. Thermomécanique, Convexité, Divergences de Bregman, Mesures de dissimilarité, Traitement de données.

Manuscript received 6 September 2022, revised 18 November 2022, accepted 20 December 2022.

1. Introduction

In thermomechanics as for many multi-physics applications, the need for metrics, "discrepancy measures" or various concepts derived from a distance in order to compare fields of quantities of interest is growing, not only but mainly with the development of large amount of digitalized data, coming the experiments, computations or in-operation measurements. Indeed, a very large majority of algorithms or data processing methods use, at one time or another, the comparison of fields. We can cite mechanical analysis, inverse problems, model reduction as well as classification or knowledge extraction from massive data.

In order to introduce a correct processing of data, or more generally of fields with different physical dimensions, in order to try to introduce the *a priori* knowledge which we have, thanks to the modeling works and the descriptions of the physics, it is essential to go beyond the usual norms of IR^n or even of functional spaces, used until now mainly to deal with a single physics. The more and more frequent irruption of high dimension spaces or probabilistic tools also raises questions about the usual norms and even about the notions of neighborhood [1, 2]. In the case of thermomechanics, for example, "solution fields" such as displacement, temperature, strain and stress tensors, are structured both locally and spatially through the conservation laws, the thermodynamics requirements, the constitutive equations and the symmetries, and some invariant quantities or global properties are also well established. Even if in practice, data issued from experiments, measurements or even computations are corrupted with some noise, it seems that trying to incorporate as such as possible this *a priori* knowledge can lead to a guarantee of objectivity and performance, especially when mixing the different types or origins of the data that are manipulated.

Coming back to thermomechanics, the general concept of energy has been used for decades and some authors already proposed derived concepts which revealed fruitful in this direction in various applications. To mention only a few: the Kohn–Vogelius functional [3] for distributed parameters identification, the constitutive relation error [4–6] for finite element error estimation and model updating or identification of material parameters, the Reciprocity Gap [7] for crack identification, the Equilibrium Gap [8] used together with image correlation techniques, ...

Let's take a simple illustration. Suppose we want to identify the rigidity of springs distributed along a bar in the traction–compression regime. The bar is clamped at one end and submitted to a prescribed force C at the other. The identification is performed provided the corresponding displacement is measured all along the bar and known. Two situations are considered, in the first one (a), the springs are linear ones, whereas in the second one (b), a supplementary cubic nonlinearity is added. The equations of the two systems are respectively:

$$(a) - u'' + \alpha^2 u = 0, \quad u(0) = 0, u'(1) = C \quad (b) - u'' + \alpha^2 u + \varepsilon\beta u^3 = 0, \quad u(0) = 0, u'(1) = C$$

In the case (a) the rigidity α_2 is sought whereas in the case (b) it is known and the rigidity $\varepsilon\beta$ is sought. A classical approach is to minimize a squared measure of discrepancy between the actual displacement (for the true value of the rigidity, that is α_0 for case (a) and $\varepsilon\beta_0$ for case (b)) and the computed one for a given value of the rigidity. The Figure 1 displays for the cases (a) and (b), the usual least-square function, the energy of the difference of the two fields, and what will be called below the symmetrized Bregman divergence between the two fields (which in case (a) turns about to be exactly the energy of the difference) that is:

$$\begin{aligned}(a)\ G_1(\alpha^2) &= \frac{\alpha_0^2}{2}\int_0^1 (u(\alpha^2) - u(\alpha_0^2))^2 \mathrm{d}x \\ G_2(\alpha^2) &= \frac{1}{2}\int_0^1 \left[(u'(\alpha^2) - u'(\alpha_0))^2 + \alpha^2(u(\alpha^2) - u(\alpha_0^2))^2\right] \mathrm{d}x\end{aligned}$$

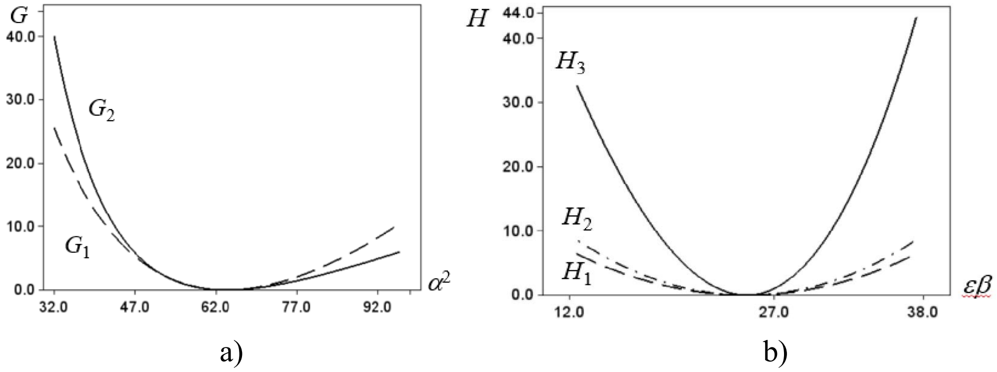


Figure 1. (a) Error functions for the identification of α. (b) Error functions for the identification of β. $\alpha_0 = 8, \beta_0 = 25, \varepsilon = 0.1, C = 50$.

$$
\begin{aligned}
(b)\ H_1(\varepsilon\beta) &= \frac{\alpha^2}{2}\int_0^1 (u(\varepsilon\beta) - u(\varepsilon\beta_0))^2 \mathrm{d}x, \\
H_2(\varepsilon\beta) &= \frac{1}{2}\int_0^1 \left[(u'(\varepsilon\beta) - u'(\varepsilon\beta_0))^2 + \alpha^2(u(\varepsilon\beta) - u(\varepsilon\beta_0))^2 + \varepsilon\beta(u(\varepsilon\beta) - u(\varepsilon\beta_0))^4\right] \mathrm{d}x \\
H_3(\varepsilon\beta) &= \frac{1}{2}\int_0^1 \left[(u'(\varepsilon\beta) - u'(\varepsilon\beta_0))^2 + \alpha^2(u(\varepsilon\beta) - u(\varepsilon\beta_0))^2\right. \\
&\quad \left. + \varepsilon\beta(u(\varepsilon\beta) - u(\varepsilon\beta_0))(u^3(\varepsilon\beta) - u^3(\varepsilon\beta_0))\right] \mathrm{d}x.
\end{aligned}
$$

In both cases, closed form solutions are used, and for case (b) it is obtained by the first two terms of an asymptotic expansion with ε as the small parameter.

It can be observed that, even in this simplistic case, the functions G_2 and H_3 are more suited for the minimization procedure. Indeed the ratios of the Hessians are respectively 1.2 and 5.9 at points α_0 and β_0. The difference being much more pronounced in the nonlinear case.

This paper is devoted to the introduction of various concepts, involving or derived from, the Bregman divergence and leading to various forms of discrepancy measures, in a special sense, for thermomechanics quantities. These concepts encompass some of the concepts cited previously and have been already used for solving nonlinear Cauchy problems in various situations in thermomechanics [9–11]. The paper is organized as follows. First, the Bregman divergence is defined and various properties are described together with the definition of derived concepts such as the Symmetrized Bregman Divergence and the Bregman Gap. Then generating functions for the Bregman divergence and derived concepts are given for a large class of constitutive relations or situations in thermomechanics. Finally, some properties useful for elementary computational geometry are given.

2. The Bregman divergence: main properties and generalization

2.1. *The Bregman divergence*

The Bregman divergence has been introduced by Bregman [12] in convex programming context [13], although the name *Bregman divergence* has been coined later [14]. It has also been studied and used for some years in the field of statistical inference and data processing problems [15], such as estimation, detection, classification, compression, and recognition, in particular for classification or clustering applications [16], and even, more recently, in deep learning applications.

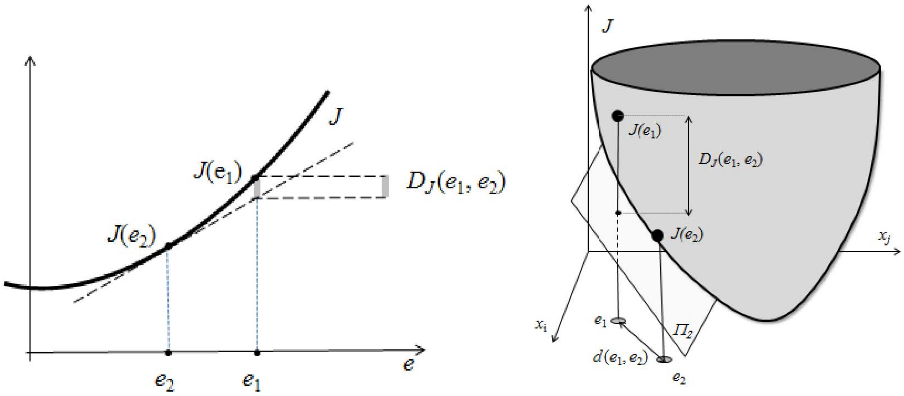


Figure 2. Geometrical interpretation of the Bregman divergence in IR^d, $d = 1$ and 2 for differentiable generating functions.

In learning applications, the Bregman divergence is used as a discrepancy measure and in some optimization algorithms [17] it is recognized as superior to the usual Euclidian distance.

We give here a slight generalization of the Bregman divergence, which includes the case of non-differentiable generating functions, because they are to be encountered in thermomechanics.

Definition. Let J be a proper convex function, $J(e) : \mathrm{IR}^n \mapsto \mathrm{IR}$, the Bregman divergence, generated by J, between two points (e_1, e_2) belonging to $\mathrm{dom}(J)^2$ is the non-negative scalar:

$$\begin{aligned} D_J(e_1, e_2) &= J(e_1) - J(e_2) - \langle p_{21}, e_1 - e_2 \rangle \\ p_{21} &= \frac{1}{2}\left[\underset{p \in \partial J(e_2)}{\arg\max} \langle p, e_1 - e_2 \rangle + \underset{p \in \partial J(e_2)}{\arg\min} \langle p, e_2 - e_1 \rangle \right] \end{aligned} \tag{1}$$

where $\partial J(e)$ is the subdifferential of J at e: $\partial J(e) = \{p, J(d) \geq J(e) + \langle p, d - e \rangle \ \forall d \in \mathrm{dom}(J)\}$.

If the generating function J is differentiable at point e_2, then the Bregman divergence reduces to:

$$D_J(e_1, e_2) = J(e_1) - J(e_2) - \langle \nabla J(e_2), e_1 - e_2 \rangle \tag{2}$$

Roughly speaking, the Bregman divergence is the tail of the Taylor expansion of J. An extension to infinite dimensional spaces of the Bregman divergence has been provided [18]. The positivity of the Bregman divergence results simply from the genuine definition of a convex function which lies “above its tangent planes” as illustrated on Figure 2. The Bregman divergence can be used as a discrepancy measure between two vectors as it takes strictly positive values if the two vectors are different (when the generating function is strictly convex), and is zero on any pair (e, e).

The definition (1) for non-differentiable generating functions is chosen as to obtain a “medium” discrepancy measure (see Figure 3), and leads also to a non-zero divergence between zero and other points for generating functions that are positively homogeneous with degree one. The geometrical interpretation is slightly different than in the case of differentiable generating functions, and the value p_{12}, which is an element of $\partial J(e)$, by convexity of the subdifferential, turns out to be also the mean value of the directional derivatives in the directions $(e_1 – e_2)$ and $(e_2 – e_1)$.

The choice retained within the subdifferential of J in e_2 does not however guarantees the directional continuity of the Bregman divergence around the point of non-differentiability e_2. Consider for example $J(e) = e^2 + \alpha e^+ - \beta e^-, \alpha, \beta \geq 0$, then $\partial J(0) = [-\beta, \alpha]$ and $D_J(e, 0) = J(e) - (1/2)(\alpha - \beta)e$.

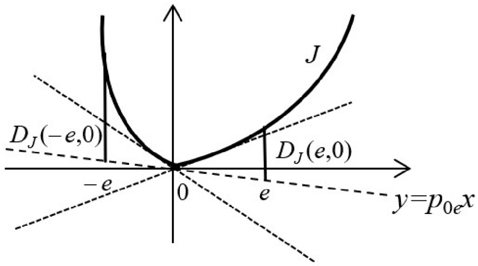


Figure 3. Geometrical interpretation of the Bregman divergence in IR for non-differentiable generating function at $e = 0$.

Table 1. Examples of generating functions and associated Bregman divergences

Domain	Generating function $J(x)$	Bregman divergence $D_J(x, y)$	Name
IR^n	$\|x\|^2$	$\|x-y\|^2$	Euclidian distance
IR^n	$J(x) = x^T Ax$ *A symmetric positive*	$(x-y)^T A(x-y)$	Mahalanobis distance
IR^{+*n}	$\sum x_i \log x_i - x_i$	$\sum x_i \log \dfrac{x_i}{y_i} - x_i + y_i$	Kullback–Leibler divergence or relative entropy
IR^{+*n}	$\sum -\log x_i$	$\sum \dfrac{x_i}{y_i} - \log \dfrac{x_i}{y_i} - 1$	Itakura–Saito discrete distance
$]0,1[$	$x\log x + (1-x)\log(1-x)$	$x\log \dfrac{x}{y} + (1-x)\log \dfrac{1-x}{1-y}$	Logistic loss

There are situations in the field of learning, for example in image classification, which do not require finer concepts such as true metrics that by the way do not really make sense, for example for the triangle inequality [19]. Nevertheless, even if $D_J(e_1, e_2)$ vanishes when $e_1 = e_2$ (and reciprocally if the generating function J is strictly convex), Bregman divergences are not distances as they are neither symmetric nor satisfying the triangle inequality in general. In Table 1, some Bregman divergences are displayed, the two first ones are genuine distances whereas the last ones, which are not, are used in learning applications and statistics.

Nevertheless, the following interesting properties [16] which are useful in optimization as well as in computational geometry have been established and have been extended to the infinite dimensional context [18].

Convexity. $D_J(e_1, e_2)$ is always convex in the first argument (but not necessarily in the second argument, take for example $J(e) = e^3$ in IR^+, $D_{e^3}(e_1, e_2)$ is not convex in e_2 when $e_2 < e_2$).

Linearity. The Bregman divergence is a linear operator with respect to generating functions

$$D_{\lambda J+\mu F}(e_1, e_2) = \lambda D_J(e_1, e_2) + \mu D_F(e_1, e_2) \quad (J, F \textit{ convex } \lambda, \mu \textit{ positive}).$$

Projection onto convex sets. There exists a Bregman projection onto convex sets C with uniqueness if J is differentiable at e

$$\forall C\ \exists! \tilde{e}\ \textit{such as}\ \tilde{e} = \operatorname*{Arg\,Min}_{g\in C} D_J(g, e).$$

Note that the order of arguments in the definition of the projection is crucial as the Bregman divergence is generally not symmetric and not always convex with respect to its second argument.

Hyperplane separation. The set of points equidistant, in terms of Bregman divergence, to a pair of distinct points is an hyperplane.

Weak form of Pythagoras theorem (for differentiable J).

$$D_J(e_1, e_3) = D_J(e_1, e_2) + D_J(e_2, e_3) + \langle \nabla J(e_2) - \nabla J(e_3), e_1 - e_2 \rangle$$

then

when C is a convex set

$$\forall e_2 \in \mathrm{IR}^n \ et \quad \forall e_1 \in C, \quad D_J(e_1, e_2) \geq D_J(e_1, \tilde{e}_2) + D_J(\tilde{e}_2, e_2)$$

when C is an affine subspace

$$\forall e_2 \in \mathrm{IR}^n \ et \quad \forall e_1 \in C, \quad D_J(e_1, e_2) = D_J(e_1, \tilde{e}_2) + D_J(\tilde{e}_2, e_2).$$

As the generating function J is defined up to an affine function because of the linearity property:

$$D_{\langle a,e\rangle + b}(e_1, e_2) \equiv 0, \quad D_{J+F}(e_1, e_2) = D_J(e_1, e_2) + D_F(e_1, e_2)$$

we can suppose, or prescribe, that $J(0)$ and $\nabla J(0)$ are zero: $D_J \equiv D_{J - J(0) - \nabla J(0)\cdot e}$. In this case, it can easily be seen that the following properties hold:

$$D_J(e, 0) \equiv J(e) \leq \langle \nabla J(e), e \rangle. \tag{3}$$

As the Bregman divergence can nevertheless be symmetric, even it is generally not the case, it is interesting to derive a characterization of the symmetric Bregman divergences, because the preceding properties are clearly improved.

Proposition (Characterization of symmetric Bregman divergences). *The Bregman divergence generated by the convex differentiable function J is symmetric iff J is a (positive) quadratic form.*

The proof is provided in Appendix A. This condition is stronger than the condition of positive homogeneity of degree two. For example in IR, the Bregman divergence generated by the function positively homogeneous of degree two $H(x) = a(\langle x\rangle^+)^2 + b(\langle x\rangle^-)^2$, with $a \neq b$, cannot be symmetric. Indeed for any $x > 0$:

$$D_H(x, -x) = ax^2 + 3bx^2 \neq bx^2 + 3ax^2 = D_H(-x, x).$$

This result, perhaps disappointing by the restriction it imposes on the generating functions J, has nevertheless as a corollary that symmetric Bregman divergences are actually distances (namely Mahalanobis distances in finite dimension).

Corollary. *Symmetric Bregman divergences satisfy the triangle inequality and are therefore distances.*

2.2. *Symmetrized Bregman divergences, Bregman gaps and ε-Bregman gaps*

For application in thermomechanics, it can be desirable to have concepts of dissimilarity that first give the same role to both components of the pair (e_1, e_2) to be analyzed, and second involve primal and dual variables as well, because their dual product has a strong energetic sense. That is the reason why we introduce first the notion of symmetrized Bregman divergence and second the notion of Bregman Gap. The symmetrized Bregman divergence, as illustrated on Figure 4, is simply taken as the sum of the Bregman divergences (the first idea seems to come back to 1946, [20]):

$$D_J^s(e_1, e_2) = D_J(e_1, e_2) + D_J(e_2, e_1). \tag{4}$$

Another definitions have been given: the Jensen–Bregman divergence [21]:

$$JB_J(e_1, e_2) = D_J\left(e_1, \frac{e_1 + e_2}{2}\right) + D_J\left(e_2, \frac{e_1 + e_2}{2}\right) \tag{5}$$

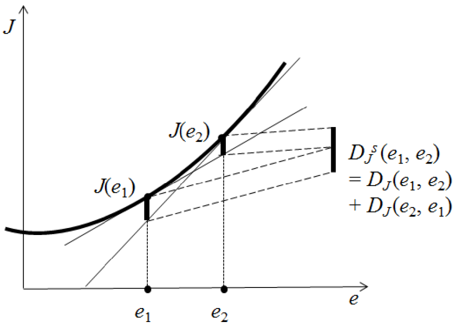


Figure 4. Geometric interpretation of the symmetrized Bregman divergence in IR for differentiable generating functions.

and the symmetrized divergence of Chen *et al.* [22] (a positive quantity because J is convex):

$$m_J(e_1, e_2) = \frac{1}{2}[J(e_1) + J(e_2)] - J\left(\frac{e_1 + e_2}{2}\right) \tag{6}$$

grounded on the property: $m_J(e_1, e_2) = \min_z (1/2)[D_J(e_1, z) + D_J(e_2, z)]$. The properties of the symmetrized divergence (6), mainly for the scalar case, rely strongly on the regularity of the generating function, at least a C^2. Nevertheless, in the context of thermomechanics, potentials useful for defining generating functions can be not differentiable (see part 3).

These definitions are not equivalent even for symmetric Bregman divergences and the first definition (4) is chosen here.

If the generating function of the Bregman divergence is differentiable, then the symmetrized Bregman divergence takes a very simple form, which can be used to calculate it:

$$D_J^s(e_1, e_2) = \langle \nabla J(e_1) - \nabla J(e_2), e_1 - e_2 \rangle. \tag{7}$$

This expression provides another justification for the positivity the symmetrized Bregman divergence by the property of monotonicity of the gradient of a convex function. It also provides a convenient procedure for calculation of the symmetrized Bregman divergence; the following example is obtained in this way:

$$J(e) = \frac{1}{2}\left[e^4 + \frac{e^2}{2} - e^2 \ln e\right] \Rightarrow D_J^s(e_1, e_2) = 2(e_1^3 - e_2^3 + e \ln e_1 - e_2 \ln e_2)(e_1 - e_2).$$

In the framework of linear elasticity in small perturbations, then if e is the linearized strain tensor $\boldsymbol{\varepsilon}$, and J is taken as φ the elastic energy density, the symmetric Bregman divergence is the so-called error in (elastic) constitutive equation, involving also the stress tensor $\boldsymbol{\sigma}$, or twice the energy of the difference, or even up to a multiplication coefficient, the Kohn–Vogelius functional integrand:

$$D_\varphi^s(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2) = \langle \nabla\varphi(\boldsymbol{\varepsilon}_1) - \nabla\varphi(\boldsymbol{\varepsilon}_2), \boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_2 \rangle = (\boldsymbol{\sigma}_1 - \boldsymbol{\sigma}_2) : (\boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_2) = 2\varphi(\boldsymbol{\varepsilon}_1 - \boldsymbol{\varepsilon}_2). \tag{8}$$

Moreover when the generating function J of the Bregman divergence is twice differentiable, it can be given a quadratic expression of the symmetric divergence using the Hessian value at some point e_{12} of the segment $[e_1, e_2]$, thanks to the intermediate value theorem, but obviously the point e_{12} depends on the pair (e_1, e_2):

$$D_J^s(e_1, e_2) = {}^t(e_1 - e_2)\nabla^2 J(e_{12})(e_1 - e_2) \quad e_{12} \in [e_1, e_2].$$

The symmetric Bregman divergence, although symmetric by construction, still does not possess generically the triangle property; nevertheless it can be shown that the violation of triangle inequality is limited by a quantity related to the Hessian bounds [23]:

$$\exists m, M > 0 \quad m^t\xi\cdot\xi \leq {}^t\xi\cdot\nabla^2 J\cdot\xi \leq M^t\xi\cdot\xi \Rightarrow$$
$$\sqrt{D_J^s(e_1,e_3)} \leq \sqrt{D_J^s(e_1,e_2)} + \sqrt{D_J^s(e_2,e_3)} + (\sqrt{M}-\sqrt{m})[D_J^s(e_1,e_2)D_J^s(e_2,e_3)]^{1/4}.$$

It is worth noting that the symmetrized Bregman divergence, as a function of (e_1, e_2), is generally not separately convex. This can be illustrated by the simple counter example of the generating function $J(e) = e - \sin e$ on $[0,\pi]$. The symmetrized divergence with the origin $D_J(0,e) = \nabla J(e)\cdot e = e(1-\cos e)$, is not convex over the entire segment $[0,\pi]$. The same counter-example applies to the Jensen–Bregman divergence (5) and the Chen *et al.* symmetrized divergence (6) as well.

With a view to application in thermomechanics, it is of interest to introduce a concept that enjoys symmetry while simultaneously involving dual and primal variables. More precisely, it is useful in applications concerning data structured by an underlying physics (experimental data or data from numerical simulations involving partial differential equations), to have divergences or discrepancy measures between couples of dual variables (e.g. stress and strain tensors, temperature gradient and heat flux, . . .).

Definition (Bregman Gap). Let J be a convex, not necessarily differentiable function, the Bregman gap BG_J generated by J between e_1 and the couple of dual quantities (e_2, p_2), $p_2 \in \partial J(e_2)$, is the non-negative quantity:

$$BG_J(e_1,[e_2,p_2]) = J(e_1) - J(e_2) - \langle p_2, e_1 - e_2\rangle. \tag{9}$$

Here in contrast with the definition (1) of Bregman divergences for non-differentiable functions, no choice is made for the element in the subdifferential of J at e_2 because it is itself an argument of the Bregman Gap. In turn, this notion allows to define a symmetric gap between any couples of dual variables $[e_1, p_1]$, and $[e_2, p_2]$.

Definition (Symmetrized Bregman Gap). Generated by the convex function J, the Symmetrized Bregman Gap BG_J^s between the two couples of dual quantities (e_1, p_1) and (e_2, p_2), $p_1 \in \partial J(e_1)$, $p_2 \in \partial J(e_2)$, is the nonnegative scalar:

$$BG_J^s([e_1,p_1],[e_2,p_2]) = BG_J(e_1,[e_2,p_2]) + BG_J(e_2,[e_1,p_1]).$$

A straightforward calculation gives the following expression for the Symmetrized Bregman Gap:

$$BG_J^s([e_1,p_1],[e_2,p_2]) = \langle p_1 - p_2, e_1 - e_2\rangle \geq 0. \tag{10}$$

Note that the Symmetrized Bregman Gap is separately convex (the proof is provided in Appendix B). The generating function does not appears anymore in this expression, but its positivity relies strongly on the existence of a convex function J such that $p_i \in \partial J(e_i)$, because this expression is nothing but the monotony of the differential of convex functions. In applications, the determination of J is then not necessary as soon as the p_i are available, although its existence is mandatory (an interesting example is given by elasticity with large transformations, cf. part 3 and Appendix C). Going back to the example of linear elasticity where the elastic energy density φ is quadratic (provided the reference state is stress-free), the Symmetrized Bregman Gap between two states $(\boldsymbol{\sigma}_1, \boldsymbol{\varepsilon}_1)$ and $(\boldsymbol{\sigma}_2, \boldsymbol{\varepsilon}_2)$, or their discrepancy measure, is simply twice the elastic energy of the difference of the strain tensors $2\varphi(\varepsilon_1 - \varepsilon_2)$, or twice the elastic complementary energy of the difference of the stress tensors $2\varphi*(\sigma_1 - \sigma_2)$, provided that $\sigma_i \in \partial_\varepsilon\varphi(\varepsilon_i), \varepsilon_i \in \partial_\sigma\varphi*(\sigma_i)$. Again, an expression already existing in the literature for application with elastic materials or structures with quadratic potentials is recovered.

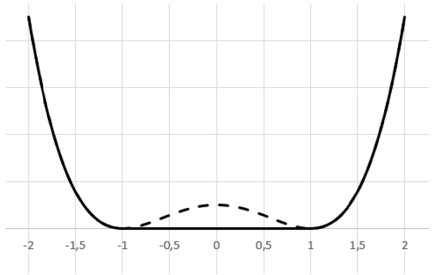


Figure 5. Original function J (dotted line) and its biconjugate (solid line) that can be used as generating function for a Bregman divergence and associated concepts.

More generally, the convex framework [24, 25] used here leads to two properties involving the convex conjugate J^* of the generating function J.

The Bregman Gap generated by the convex function J is nothing but the Legendre–Fenchel residual.

If J^* is the Legendre–Fenchel conjugate function of J then: $D_J(e_1,e_2) = D_{J^*}(p_{21},p_{12})$ *or if the function J is differentiable:* $D_J(e_1,e_2) = D_{J^*}(\nabla J(e_2),\nabla J(e_1))$

These results come from the property of the Legendre–Fenchel transform:

$$J(e)+J^*(p) = \langle p,e\rangle \Leftrightarrow e \in \partial J^*(p), p \in \partial J(e).$$

So that:

$$\begin{aligned} D_J(e_1,e_2) &= J(e_1) - \langle p_{21},e_1\rangle - (J(e_2) - \langle p_{21},e_2\rangle) \\ &= J(e_1) - \langle p_{21},e_1\rangle + J^*(p_{21}) \\ &= \langle p_{12},e_1\rangle - J^*(p_{12}) - \langle p,e_1\rangle + J^*(p) \quad \forall p \in \partial J(e_1) \\ &= J^*(p_{21}) - J^*(p_{12}) - \langle p_{21}-p_{12},e_1\rangle \\ &= D_{J^*}(p_{21},p_{12}). \end{aligned}$$

This expression makes particularly sense in a thermodynamic framework were the thermodynamic potentials (cf. part 3) are used as generating functions: coherently, dual potentials are in correspondence in the Bregman divergence through the dual variables. Because the biconjugate J^{**} of any function J is convex (like its conjugate J^*), one can use it in order to derive a Bregman divergence for a nonconvex function. And then we have again (because $J^{***} = J^*$):

$$D_{J^{**}}(e_1,e_2) = D_{J^*}(p_{21},p_{12}).$$

As an example, the following double-well free energy potential, used in phase transformation models, is not convex, but its conjugate is convex, although not strictly convex, and can be calculated by taking advantage from the fact that the biconjugate of J is also the largest closed convex function smaller than J.

$$J(e) = \frac{1}{2}(e^2-1)^2, \quad J^{**}(e) = \left| \begin{array}{ll} \frac{1}{2}(e^2-1)^2 & \text{if } |e|<1 \\ 0 & \text{if } |e|<1. \end{array}\right.$$

However, the separation property of the generated Bregman divergence fails in the interphase segment $[-1,1]$ as illustrated on Figure 5.

$$D_{J^{**}}(e_1,e_2) = \tfrac{1}{2}(e_1^2-1)^2 I_{[-1,1]}(e_1) - \tfrac{1}{2}(e_2^2-1)^2 I_{[-1,1]}(e_2) - 2e_2(e_2^2-1)(e_1-e_2) I_{[-1,1]}(e_2).$$

In applications, the dual quantities p, a gradient or a sub-gradient, are often prone to errors, coming either from computations or measurements. If an estimation of the approximation level, say ε, is known, we can use the notion of ε-subdifferential, coming from the optimization community [26, 27] and proved to be the right approximation notion for subdifferential, and incorporate it in a definition of ε-Bregman Gaps.

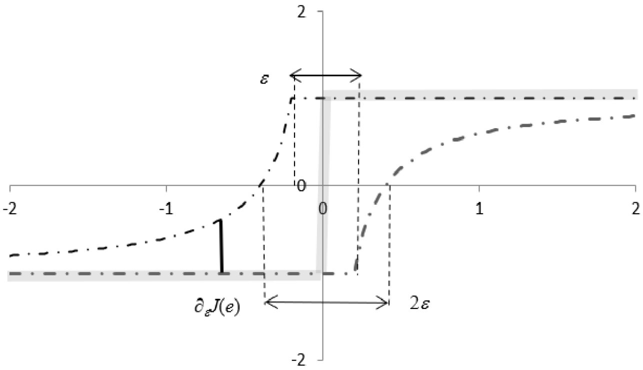


Figure 6. ε-sub-differential of $J(e) = |e|$ as a function of e along the real axis for $\varepsilon = 0.4$. For $\varepsilon = 0$, the subdifferential (heavy grey line) is the segment $[-1,1]$ at $e = 0$, and reduces to $\{1\}$ and $\{-1\}$ for $e > 0$ and $e < 0$ respectively.

Definition (ε-subgradients, ε-sub differentials [26, 27]). Let J be any convex function and pick any e in dom J, then for a positive scalar ε, a vector p is called an ε-subgradient of J at e, when the following property holds:

$$J(v) - J(e) - \langle p, v - e\rangle \geq -\varepsilon \quad \forall v \quad \text{or equivalently} \quad J(e) + J^*(p) - \langle p, e\rangle \leq \varepsilon.$$

The set of all ε-subgradients at e is called the ε-sub differential of J at e, and denoted by $\partial_\varepsilon J(e)$.

If J is a quadratic function, $J(x) = (1/2)x^t Qx$, then the ε-subdifferential is simply the set $\partial_\varepsilon J(x) = \{Qx + \tau : \tau^t Q^{-1}\tau \leq 2\varepsilon\}$. In the context of linear elasticity, the ε–stress tensors associated to a strain tensor e, are then the sum of the actual stress tensor σ associated by the constitutive equation to e and any stress tensor with elastic energy lesser than ε. As an example in one dimension, the ε-subdifferential of the function $J(e) = |e|$ is illustrated on Figure 6.

Definition (ε-Bregman Gaps). Let J be a convex, not necessarily differentiable function, and ε a non-negative scalar:

(i) the *ε-Bregman Gap ε-BG_J generated by J between e_1 and the couple of approximate dual quantities* (e_2, p_2), $p_2 \in \partial_\varepsilon J(e_2)$, *is the non-negative quantity:*

$$\varepsilon BG_J(e_1, [e_2, p_2]) = J(e_1) - J(e_2) - \langle p_2, e_1 - e_2\rangle + \varepsilon, \quad p_2 \in \partial_\varepsilon J(e_2).$$

(ii) the ε-Symmetrized Bregman Gap εBG_J^s generated by J between the couple of approximate dual quantities (e_1, p_1) and (e_2, p_2), $p_1 \in \partial_\varepsilon J(e_1)$, $p_2 \in \partial_\varepsilon J(e_2)$ is the non-negative quantity:

$$\varepsilon BG_J^s([e_1, p_1], [e_2, p_2]) = \varepsilon BG_J(e_1, [e_2, p_2]) + \varepsilon BG_J(e_2, [e_1, p_1]) = \langle p_1 - p_2, e_1 - e_2\rangle + 2\varepsilon$$
$$p_1 \in \partial_\varepsilon J(e_1),\ p_2 \in \partial_\varepsilon J(e_2).$$

The interest of this definition is that, is leads to a positive discrepancy measure without knowing the generating function but only provided it exists with an approximation ε for the relation $p \in \partial J(e)$. Unfortunately, any property of separated convexity does not make sense for BG^s nor for ε-BG^s because their domains are not convex unless the generating function is quadratic. Indeed, the convexity would necessitate the convexity of the domains, and therefore prescribe the following implication, which is not ensured except for quadratic functions:

$$p_1 \in \partial J(e_1), p_2 \in \partial J(e_2) \Rightarrow (\lambda p_1 + (1-\lambda)p_2) \in \partial J(\lambda e_1 + (1-\lambda)e_2).$$

3. Thermodynamic potentials as Bregman generating functions in thermomechanics

The objective of section is to build suitable Bregman divergences and Bregman Gaps for the processing of data that originate either from physical experiments or from simulation computations of these same physical phenomena. As mentioned in the introduction, it is believed that taking into account the specific structure of these data that is brought by the physical constraints, conservations laws or constitutive equations can lead to better performance of the algorithms used for processing this kind of data. Indeed metrics and scalar products are at the heart of processing methods such as Proper Order Decomposition used for the sparse representation of the data, identification of structures buried in it or model reductions for the PDE governing it, or such as classification, clustering and regression algorithms in learning, often by the way of minimizing discrepancy functions between data. The application to nonlinear cases, emanating notably from the nonlinearity of the constitutive equation (for small deformation), are appealing for the definition of non-quadratic pseudo (squared) metrics.

While the merits of various metrics or pseudo-metrics, scalar or pseudo scalar products, can be considered at the very least to be more or less the same when applied to data corresponding to a single physics, the situation is very different in the case of multi-physics data relating to coupled phenomena. This is due on the one hand to the question of the physical units of the data, and on the other hand to the need to make the best use of *a priori* information contained in the coupled equations satisfied by the data. Even with uncertainties, it is desirable to construct more elaborate metrics than the usual Euclidean metrics of IR^n, and the concepts associated with the Bregman divergence seem fruitful in that respect. The main reason for this is the flexibility they allow in the choice of generating functions.

In mechanics, in thermics or more generally as soon as thermodynamic phenomena are involved, it is natural, when choosing generating functions for Bregman divergences, to turn to thermodynamic potentials or dissipation functions and potentials. Indeed, in a large majority of cases, these are convex functions of their arguments. Furthermore, these potentials were fundamentally conceived to introduce an energetic notion but also to define the relations between dual variables, and it is precisely this last particularity that can be exploited with Bregman divergences. It appears then the gap functions defined by the Symmetric Bregman Gaps coincide with some forms of error in constitutive equations. In multi-physics situations, advantage can be taken of the additivity property of the Bregman divergence (cf. part 2.1):

$$D_{\lambda J+\mu F}(e_1,e_2)=\lambda D_J(e_1,e_2)+\mu D_F(e_1,e_2)\quad (J,F \textit{ convex } \lambda,\mu \textit{ positive scalars})$$

for combining various potentials either coupled or involving different physical quantities, in order to build pseudo-metrics covering all the involved data and taking into account all the couplings between them. In this section, advantage is taken from the convex framework of Generalized Standard Materials to build ad hoc Bregman Gaps, and extension is made for the generalization of this framework with Implicit Generalized Standard Materials. Nevertheless, the classical framework of linear elliptic problems first is analyzed, where the (linear) constitutive equations are incorporated directly in the conservation laws.

3.1. *Potentials for data from physical fields*

When the data to be processed come from vector fields, issued from experiments or from numerical simulations, more or less accurately obeying a system of PDEs, energy quantities such as the potentials from which these equations derive can be used for the building of generating functions for Bregman divergences. An important first field of application is provided by the framework of linear symmetric PDEs. If, in the quasi-static assumption, the data field $u(x)$ on

Table 2. Quadratic energies or dissipations associated with various linear physics

Physics	U space	Energy or dissipation density A
Scalar conduction or diffusion (thermal, electrical, Darcean flow, incompressible Stokes flow)	$H^1(\Omega)$	$\frac{1}{2}\underline{\underline{k}}.\nabla u.\nabla u$
Elasticity	$H^1(\Omega)^3$	$\frac{1}{2}\underline{\underline{\underline{\underline{C}}}}:\underline{\underline{\varepsilon}}(u):\underline{\underline{\varepsilon}}(u)$
Plane elasticity on elastic support	$H^1(\Omega)^2$	$\frac{1}{2}\underline{\underline{\underline{\underline{C}}}}:\underline{\underline{\varepsilon}}(u):\underline{\underline{\varepsilon}}(u)+\frac{1}{2}ku^2$

a spatial domain Ω is supposed to obey such a system of equations, the variational form of these equations reads:

$$u \in V, \quad a(u,v) = l(v) \quad \forall v \in V_0$$

where a is a symmetric bilinear form defined on the product of functions spaces $U(\Omega) \times U(\Omega)$, l is a linear form on U, $V \subset U$ is the space of admissible fields (with respect to boundary conditions on parts of $\partial\Omega$) and V_0 its tangent space. If a is continuous and coercive and l is continuous, the solutions of the system are characterized by the optimization property of a potential P:

$$u = \arg\min_V P(v) \equiv \frac{1}{2}a(v,v) - l(v).$$

The convexity and positivity of a lead to consider the global Bregman divergence, i.e. having as argument the field u over Ω, and generated by the quadratic form associated with a.

$$J(u) = a(u,u)$$

because J is positive, convex and, in addition, $J(u) = 0$ implies u equals zero as well. J being a quadratic function, the Bregman divergence generated by J is symmetric and takes the simple form:

$$D_J(u,v) = a(u,u) - a(v,v) - 2a(v,u-v) = a(u-v,u-v).$$

The local Bregman divergence is generated by the integrand A of a:

$$a(u,v) = \int_\Omega A(u(x),v(x))\mathrm{d}x, \quad D_A(u(x),v(x)) = A(u(x)-v(x),u(x)-v(x)).$$

The previous expression is an abuse of notation because in general A is also a function of the derivatives of u and v: $A([u,\partial u/\partial x_i,\ldots],[v,\partial v/\partial x_i,\ldots])$. $A(u, u)$ can be often interpreted as the density of a global energy or a dissipation $V(u) = 1/2a(u,u)$ as illustrated in the Table 2.

In some situations where A is only a function of the derivatives of u, and not of the values of u, attention must be paid to the space V, which is usually a quotient space, in order to recover the strict convexity by eliminating the "rigid modes" with no energy. Table 2 displays some examples.

The discrepancy measure generated by the Bregman divergence is of particular interest in relation with usual Euclidean metrics in situation where the medium is anisotropic and/or heterogenous [28]. Another important field of application arises in fluid mechanics. For Navier–Stokes equations where the variable of interest is the velocity field $\boldsymbol{v}$, two quadratic functions can be used for generating Bregman divergences: the kinetic energy: $\rho\|\boldsymbol{v}\|^2$ and the enstrophy $\rho\|\mathit{rot}\,\boldsymbol{v}\|^2$. The enstrophy is the square of the vorticity for incompressible flows. Then, using the associated Bregman divergence allows to compare two velocity fields (or more precisely their fluctuations) in terms of the vorticity contents, this can be of interest for example in POD in order to identify coherent structures like vortexes in a flow.

3.2. *Generating functions in thermomechanics for Generalized Standard Materials*

The theory of Generalized Standard Materials (GSM) [29] is a coherent thermodynamic framework for modelling constitutive equations using only convex potentials. In addition to the usual description of the thermomechanical state constituted by the state variables (ε, σ, T), that is (mechanical deformation, stress tensor, temperature), some internal (hidden) variables collectively denoted by α are introduced.

Then complementing the free energy (Helmholtz energy), $\varphi(\varepsilon, T, \alpha)$, a pseudo potential of dissipation appears as a function of the rates of the internal variables $\mathscr{D}(\dot\alpha)$, and the constitutive equations are split into two groups:

$$\sigma = \frac{\partial\varphi}{\partial\varepsilon}, \quad S = -\frac{\partial\varphi}{\partial T}, \quad A = -\frac{\partial\varphi}{\partial\alpha} \quad \text{and} \quad A = \frac{\partial\mathscr{D}}{\partial\dot\alpha}.$$

The first group expresses the relation between dual variables (S is the entropy, A the thermodynamic force), the second group gives the evolution equation of the internal variable α, and is extended for non-differentiable pseudo-potentials thanks to the notion of subdifferential: $A \in \partial\mathscr{D}(\dot\alpha)$. This extension is necessary when modeling constitutive equation involving yield functions, which are associated with positively homogeneous dissipation pseudo-potentials with degree one.

For the simplicity of the presentation, we now limit ourselves to isothermal evolutions. The Euler implicit time incremental form of these constitutive equations reads:

$$\sigma + \Delta\sigma = \frac{\partial\varphi}{\partial\varepsilon}[\varepsilon + \Delta\varepsilon, \alpha + \Delta\alpha], \quad A + \Delta A = -\frac{\partial\varphi}{\partial\alpha}[\varepsilon + \Delta\varepsilon, \alpha + \Delta\alpha], \quad A + \Delta A = \frac{\partial\mathscr{D}}{\partial\dot\alpha}\left(\frac{\Delta\alpha}{\Delta t}\right) \tag{11}$$

showing that the (thermodynamically) conjugate couples for describing the thermodynamic state are $[(\sigma + \Delta\sigma, \varepsilon + \Delta\varepsilon), (A + \Delta A, \alpha + \Delta\alpha)]$. So, denoting:

$$\hat\varphi(\Delta\varepsilon, \Delta\alpha) = \varphi(\varepsilon + \Delta\varepsilon, \alpha + \Delta\alpha), \quad \hat{\mathscr{D}}(\Delta\alpha) = \mathscr{D}\left(\frac{\Delta\alpha}{\Delta t}\right) \tag{12}$$

a Bregman divergence on the increments of the primal state variables $\Delta e = (\Delta\varepsilon, \Delta\alpha)$ can be built using $J_\chi = (1-\chi)\hat\varphi + \chi\Delta t\hat{\mathscr{D}}$ as the generating function, where χ is a non-dimensional weighting factor between the free energy potential and the dissipation pseudo-potential ($\chi \in [0,1]$ and $\chi < 1$ if $\mathscr{D}$ is homogeneous of degree one, for preserving the strict convexity). This leads to the following Symmetrized Bregman Gap on incremental couples of dual state variables $[\Delta e, \Delta p]$, with $\Delta e = (\Delta\varepsilon, \Delta\alpha)$ and $\Delta p = (\Delta\sigma, \Delta A)$:

$$BG^s_{J_\chi}([\Delta e_1, \Delta p_1], [\Delta e_2, \Delta p_2]) = (1-\chi)(\Delta\sigma_1 - \Delta\sigma_2) : (\Delta\varepsilon_1 - \Delta\varepsilon_2) + (1-2\chi)\langle \Delta A_1 - \Delta A_2, \Delta\alpha_1 - \Delta\alpha_2\rangle. \tag{13}$$

This reduces to the so called Drücker mechanical error [30, 31] when χ takes the value $1/2$, that is when the free energy and dissipation are equally balanced in building the discrepancy measure.

$$BG^s_{J_{1/2}}([\Delta e_1, \Delta p_1], [\Delta e_2, \Delta p_2]) = \tfrac{1}{2}(\Delta\sigma_1 - \Delta\sigma_2) : (\Delta\varepsilon_1 - \Delta\varepsilon_2). \tag{14}$$

For perfect isotropic elasto-plasticity for example, the primal state is described by the couple $(\varepsilon, \varepsilon^p)$ where ε^p is the plastic strain, the free energy is simply $\varphi(\varepsilon, \varepsilon^p) = 1/2C : (\varepsilon - \varepsilon^p) : (\varepsilon - \varepsilon^p)$ and the pseudo-potential of dissipation is the non-differentiable positively homogeneous function of degree one $\mathscr{D}(\dot\varepsilon^p) = \sigma_0\|\dot\varepsilon^p\|_{\text{dev}}$. The Symmetrized Bregman Gap reads then (for $\chi = 0$, the elastic energy gap is of course retrieved):

$$\begin{aligned}&BG^s_J([(\Delta\varepsilon_1, \Delta\varepsilon^p_1), (\Delta\sigma_1, -\Delta\sigma_1)], [(\Delta\varepsilon_2, \Delta\varepsilon^p_2), (\Delta\sigma_2, -\Delta\sigma_2)])\\ &\quad = (\Delta\sigma_1 - \Delta\sigma_2) : [(1-\chi)(\Delta\varepsilon_1 - \Delta\varepsilon_2) - (1-2\chi) : (\Delta\varepsilon^p_1 - \Delta\varepsilon^p_2)].\end{aligned}$$

By a reverse approach to the situations in Section 3.1, Global Symmetrized Bregman Gaps are simply built by integrating over the geometrical domain of interest Ω. It is worth noting, that a direct approach on the domain Ω is also possible following [32]. Because the Bregman Gap takes exactly the value of the Fenchel residual [25] of its generating function J, involving its conjugate J^*:

$$r_F(e,p) = J^*(p) + J(e) - \langle p, e\rangle$$

the Symmetrized Bregman gaps are closely related to the errors in constitutive equations largely used in thermomechanics. Furthermore, it is possible to define a dual Bregman Gap with J^* as generating function with the property:

$$BG_J(e_1,[e_2,p_2]) = BG_{J^*}(p_2,[p_1,e_1]).$$

This result just reflects that conjugate thermodynamic potentials can equivalently be used for building the Bregman Gaps within the framework of Generalized Standard Materials. Nonetheless, it has to be kept in mind that even if the generating function J does not appear anymore in the final expression (7) or (8) of the Bregman gaps, existence of a convex generating function is the essential condition for these expressions being positive. In finite elasticity for example, the Coleman–Noll error which involves the gradient of the deformation F and the first Piola–Kirchhoff stress tensor T, and which is exactly the expression of the Bregman Gap, is not a positive quantity. Fortunately here, it is possible to turn to a weaker notion of convexity, namely the poly-convexity [33], and to use then as primal variables (F, Cof F, det F) in a correctly build, and consequently positive, Bregman Gap, details are given in the Appendix C.

3.3. *Bregman gaps generated by bi-potentials*

Unfortunately, while the GSM framework is very wide, there are still many constitutive laws that do not fit into it, such as Armstrong–Frederick hardening plasticity, Drücker–Prager elastoplasticity, Clam-Clay models, some damage models or even Coulomb friction. An extension of the framework of Generalized Standard Materials has been proposed [34] in order to take into account these behaviors that do not obey the normality law in the complementary equation. In this formulation, the pseudo-dissipation potential becomes a function of both the primal variable e and its dual variable p, justifying the name of bi-potential, or implicit potential since the relationship between e and p is expressed by a subdifferential implicit equation:

$$e \in \partial_p b(e,p) \quad or \quad p \in \partial_e b(e,p).$$

Definition (Bi-potentials [34]). Let e and p be two vectors of variables in duality. A function b of $\mathrm{IR}^n \times \mathrm{IR}^n$ with value in $\mathrm{IR} \cup \{+\infty\}$ is a bi-potential for (e,p) if and only if:

(i) b is separately convex and lower semi-continuous for e and p
(ii) b satisfies the generalized Fenchel inequality: $b(e,p) \geq \langle e,p\rangle \; \forall e,p$
(iii) $b(e,p) = \langle e,p\rangle \Leftrightarrow e \in \partial_p b(e,p) \Leftrightarrow p \in \partial_e b(e,p)$.

The material models defined using a bi-potential for the dissipation pseudo-potential are called Implicit Generalized Standard Materials. Indeed, in the particular case where b is a separate variables function, $b(e,p) = f(e) + g(p)$, then necessarily g and f are mutually conjugate. Therefore, Generalized Standard Materials are a special cases of Implicit Generalized Standard Materials for which the bi-potential takes the special form: $b(e,p) = f(e) + f*(p)$. General bi-potentials can be used as generating functions for a Bregman Gap provided that the definition of the latter is adapted as follows.

Definition (Bregman Gap generated by a bi-potential). The Bregman gap generated by the bi-potential $b(e,p)$ is defined by the positive quantity:

$$BG_b(e_1,[e_2,p_2]) = b(e_1,p_2) - b(e_2,p_2) - \langle p_2, e_1 - e_2\rangle \quad \text{with } p_2 \in \partial_p b(e_2,p_2) \tag{15}$$

or equivalently:

$$BG_b(e_1,[e_2,p_2]) = b(e_1,p_2) - \langle p_2,e_1\rangle. \tag{16}$$

This definition generalizes the initial definition of Bregman Gaps: if the bi-potential is a separate variables function, then the genuine definition is retrieved. Similarly, Symmetrized Bregman Gaps can be derived.

Definition (Symmetrized Bregman Gap generated by a bi-potential). Let b be a bi-potential, the Symmetrized Bregman Gap generated by b is:

$$\begin{aligned} BG^s_b([e_1,p_1],[e_2,p_2]) &= BG_b(e_1,[e_2,p_2]) + BG_b(e_2,[e_1,p_1]) \\ &= b(e_1,p_2) + b(e_2,p_1) - \langle p_2,e_1\rangle - \langle p_1,e_2\rangle. \end{aligned} \tag{17}$$

As soon as the bi-potential b is (separately) strictly convex, nullity of the Symmetric Bregman Gap generated by b implies the equality between primal state variables e_1 and e_2 and the associated dual variables p_1 and p_2. However, the normality law, characteristic of Generalized Standard Materials, being deduced from the inequality $\langle p_1 - p_2, e_1 - e_2\rangle \geq 0$, it can be then expected that, since it is violated for non-standard materials, the reduction to (10) of the Symmetrized Bregman Gap to this expression is no longer verified for Symmetrized Bregman Gaps generated by a bi-potential. Indeed,

$$\begin{gathered} BG^s_b([e_1,p_1],[e_2,p_2]) = \langle p_1 - p_2, e_1 - e_2\rangle + R \\ R = b(e_1,p_2) + b(e_2,p_1) - \langle p_1,e_1\rangle - \langle p_2,e_2\rangle \neq 0. \end{gathered}$$

However, the following calculation ensures the positivity of BG^s_b:

$$R = b(e_1,p_2) + b(e_2,p_1) - b(e_1,p_1) - b(e_2,p_2) \geq \langle p_1, e_2 - e_1\rangle - \langle p_2, e_1 - e_2\rangle = -\langle p_1 - p_2, e_1 - e_2\rangle.$$

This results shows that for constitutive relations not entering the Generalized Standard Materials framework but the implicit potential framework, convex generating functions for Bregman Gaps can be built with bi-potentials. This extends considerably the application field of the Symmetrized Bregman Gaps.

3.4. *Expression of Symmetrized Bregman Gaps for various modelling in thermomechanics*

Table 3 gathers some Symmetrized Bregman Gaps associated with phenomena (or models) encountered in thermo-mechanics. Global Symmetrized Bregman Gaps are obtained by spatial integration over the domain occupied by the solid or over the interface between two contacting solids.

The expression of Global Symmetrized Bregman Gaps can sometimes be drastically simplified by using in supplement the equilibrium equation satisfied by the stress field and the compatibility of the strain tensor field. This is the case for the Drücker error (14), where domain integration can be replaced with boubdary integration:

$$\int_\Omega (\Delta\sigma_1 - \Delta\sigma_2) : (\Delta\varepsilon_1 - \Delta\varepsilon_2)\mathrm{d}\Omega \equiv \int_{\partial\Omega} (\Delta\sigma_1\cdot n - \Delta\sigma_2\cdot n) : (\Delta u_1 - \Delta u_2)\mathrm{d}S. \tag{18}$$

This leads to significant reduction of the burden computations

4. Computational geometry with Bregman concepts

Various learning techniques like SVM, clustering or partitioning, manifold learning, topological data analysis, classification etc. rely on basic problems of computational geometry,where a

Table 3. Bregman Gaps from various models in thermomechanics χ are non-dimensional scalars in [0,1], D is a characteristic time

Phenomenon	Generating function	Symmetrized Bregman Gap
Scalar conduction or diffusion	Dissipation function	$K:\nabla(u_1-u_2)\cdot\nabla(u_1-u_2)$
Linear elasticity	Elastic energy	$C:\varepsilon(u_1-u_2):\varepsilon(u_1-u_2)$
Non-linear elasticity	Elastic energy (if convex)	$(\Delta\sigma_1-\Delta\sigma_2):(\Delta\varepsilon(u_1)-\Delta\varepsilon(u_2))$
Hyper-elasticity (Finite strain)	Polyconvex elastic energy (variables: cofactors of F)	$(T_1-T_2):(M_1-M_2)+(cT_1-cT_2):(N_1-N_2)$ $+(p_1-p_2)(d_1-d_2)$
Standard elastoplasticity	Free energy dissipation pseudo-potential	$(1-\chi)(\Delta\sigma_1-\Delta\sigma_2):(\Delta\varepsilon(u_1)-\Delta\varepsilon(u_2))$ $+\chi(\Delta A_1-\Delta A_2):(\Delta\alpha_1-\Delta\alpha_2)$
Contact/Friction (contact zone) [11]	Dissipation bi-potential	$(\sigma^1_{nt}-\sigma^2_{nt}):(u^1_t-u^2_t)-(\rho\sigma^1_{nn}-\rho\sigma^2_{nn})(\lvert u^t_1\rvert-\lvert u^t_2\rvert)$
Non-standard elastoplasticity	Elastic energy dissipation bi-potential	$(1-\chi)(\Delta\sigma_1-\Delta\sigma_2):(\Delta\varepsilon(u_1)-\Delta\varepsilon(u_2))$ $+\chi(\Delta A_1-\Delta A_2):(\Delta\alpha_1-\Delta\alpha_2)$
Thermo-elasticity	Elastic energy thermal dissipation	$(1-\chi)C:[\varepsilon(u_1-u_2)-\alpha(T_1-T_2)Id]:[\varepsilon(u_1-u_2)$ $-\alpha(T_1-T_2)Id]+\chi DK\nabla(u_1-u_2)\cdot\nabla(u_1-u_2)$

distance or a metric usually named loss function in learning application, plays a central role. In order to use Bregman divergences or Bregman Gaps as loss functions, hopefully better adapted to the processing of data structured by physics, we give below some essential geometric properties. Methods for dimension reduction like the Proper Orthogonal Decomposition methods or generalizations of them can also benefit from more physically informed scalar products or, even more importantly, inner-products on product-spaces of different physical quantities. Computational geometry with Bregman divergences can be challenging because of the non-symmetry of the original Bregman divergence but also and over all because of the violation of the triangle inequality, except for quadratic Bregman divergences (Mahalanobis distances). Nevertheless, the convex basis of the Bregman divergence leads to properties that allow almost all the operations of computational geometry, sometimes at the cost of slightly altered concepts. In this section are given some geometric properties and indications on the computation of interesting geometric quantities regarding applications.

4.1. *Centroid properties and Bregman balls*

The first and essential geometric property of the Bregman divergence, especially for applications in clustering is the following. It expresses that the mean of a finite set S of N points minimizes the sum of Bregman divergences or Bregman Gaps between this point and any other points of the set S, whatever the choice of the Bregman divergence or gap.

Property (Centroid property for Bregman divergences [16] and for Symmetrized Bregman Gaps). *For any set S of N points in IR^n, and any Bregman divergence D_J, one has:*

$$\mu=\frac{1}{N}\sum_{i=1,N}x_i=\underset{s\in\mathrm{IR}^n}{\arg\min}\,C_J(s)\equiv\frac{1}{N}\sum_{i=1,N}D_J(x_i,s).\tag{19}$$

For any set S of N couples [e_i, p_i] , dual points with respect to any convex function J in IR^{2n}, and for the symmetrized Bregman gap BG_J^sgenerated by J, one has:

$$\begin{bmatrix} \mu_e \\ \mu_p \end{bmatrix} = \frac{1}{N}\sum_{i=1,N}\begin{bmatrix} e_i \\ p_i \end{bmatrix} = \underset{\left[\begin{smallmatrix} x \\ y \end{smallmatrix}\right]\in \mathrm{IR}^n}{\arg\min}\, C_J([x,y]) \equiv \frac{1}{N}\sum_{i=1,N} BG_J^s([e_i,p_i],[x,y]). \tag{20}$$

Proof. Using the expression of BG_J^s, we can calculate the derivatives of C_J with respect to x and y:

$$C_J([x,y]) = \frac{1}{N}\sum_{i=1,N}\langle e_i - x, p_i - y\rangle = \frac{1}{N}\sum_{i=1,N}\langle e_i, p_i\rangle - \left\langle \frac{1}{N}\sum_{i=1,N} e_i, y\right\rangle - \left\langle x, \frac{1}{N}\sum_{i=1,N} p_i\right\rangle + \langle x, y\rangle$$

$$\frac{\partial}{\partial x}C_J([e,p]) = p - \frac{1}{N}\sum_{i=1,N} p_i, \quad \frac{\partial}{\partial y}C_J([e,p]) = e - \frac{1}{N}\sum_{i=1,N} e_i.$$

Equating to zero the two derivatives leads to the announced result because the Symmetrized Bregman Gap, hence C_J , is separately convex. This also shows that the minimum is unique. □

The property extends to the probabilistic context with the expectation replacing the mean,

$$E_\nu[f(x)] = \frac{1}{N}\sum_{i=1,N}\nu_i f(x_i)$$

and an exhaustiveness property [35] states that the centroid property is in fact a characterization of Bregman divergences when the minimization is extended over every probability distributions. If the centroid property holds for a positive function F, such that $F(x,x) = 0$, then F is necessarily a Bregman divergence. By the same arguments, this can be established also for the Symmetrized Bregman Gap.

The generalization of the ball and sphere concepts, usually defined by means of a distance, is straightforward for both Bregman divergences and gaps. Nevertheless, due to the non-symmetry of D_J, two distinct kinds of balls or spheres appear.

Definition (Bregman balls [36,37] and Bregman Gap balls). The Bregman ball BB_J, respectively the second type Bregman Ball BB'_J, associated to the generating function J, with center c and pseudo-radius ρ is the set:

$$\begin{aligned} BB_J(c,\rho) &= \{e, D_J(e,c) \le \rho^2\} \\ BB'_J(c,\rho) &= \{e, D_J(c,e) \le \rho^2\} \end{aligned} \tag{21}$$

the associated Bregman spheres are respectively $\partial BB_J(c,\rho)$ and $\partial BB'_J(c,\rho)$.

The Bregman Gap ball, associated to the generating function J, with center μ and radius ρ is the set:

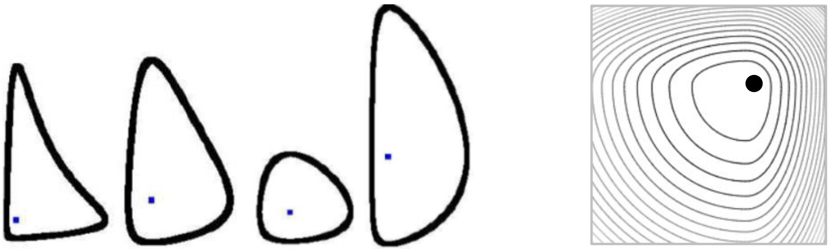
$$BBG_J^s(\mu,\rho) = \{e, BG_J^s(e,\mu) \le \rho^2\} = \{e, BG_J^s(\mu,e) \le \rho^2\} \tag{22}$$

the associated Bregman sphere is $\partial BBG_J^s(c,\rho)$.

When the generating function is the Euclidean norm, the usual notions of balls and spheres are recovered. As the Bregman divergences and symmetrized Bregman gaps are convex with respect to its first variable, Bregman balls and Bregman Gap balls are convex sets (whereas the second type Bregman balls BB'_J are generally not). To get some insights into the concepts of Bregman balls, it is interesting to look at origin centered balls, although these balls are not invariant by translation of their center. Recalling that the generating function is chosen with $J(0) = \nabla J(0) = 0$, the simple expressions are obtained:

$$\begin{aligned} BB_J(0,\rho) = \{J(e) \le \rho\}, \quad BB'_J(0,\rho) &= \{\nabla J(e)\cdot e - J(e) \le \rho\} \\ BBG_J^s(0,\rho) &= \{\nabla J(e)\cdot e \le \rho\}. \end{aligned} \tag{23}$$

Illustrations of these different balls are displayed on Figure 7 for various generating functions.



(a) *(b)*

Figure 7. Bregman Balls. (a) BB' and BB for Itakura–Saito (two first balls), and Kullback–Leibler Bregman divergences (two last balls), from [NN1]. (b) Symmetrized Bregman Gap origin-centered balls generated by the elastic potential for an isotropic elastic material with different moduli in traction and compression.

If the generating function J is positively homogeneous with degree α (necessarily $\alpha > 1$), these expressions can be further simplified and involve only the value of the function J. This illustrates the strong relation between the Bregman balls and the level-lines of their generating functions:

$$BB_J(0,\rho) = \{J(e) \le \rho\}, \quad BB'_J(0,\rho) = \{(\alpha-1)J(e) \le \rho\}$$
$$BBG^s_J(0,\rho) = \{\alpha J(e) \le \rho\}.$$

These examples show how the physically information one can *a priori* have on the data can be taken into account and have potentially a strong qualitative impact on manipulating and comparing these data. The concept of (convex) ball seems to be the right way of dealing with questions like nearest neighbor determination or minimization of dissimilarity measures between physical fields. If the Bregman divergence is strictly convex then the Smallest Enclosing Bregman Ball, in the sense of the smallest radius ρ, is unique and an efficient algorithm has been given for computing it [38].

4.2. *Computational geometry with Bregman concepts*

This section is devoted to some basic tools in computational geometry that are commonly used in situations where a true distance or a true scalar product exists. Much work has been done with Bregman divergence in the learning community, especially for classification algorithms or building Voronoi diagrams [37–41]. It was necessary for adapting methods that rely on the triangle inequality, a property that is not satisfied by most Bregman concepts, except of course for the Mahalanobis divergences which are norms (cf. Section 2.1). Some results can be extended to Bregman Gaps for use in applications in the field of thermomechanics and more generally in processing physically based data. The first useful concept is the bisector concept. Interestingly, bisectors are hyperplanes independently of the generating function, even strongly nonlinear.

Definition (Bregman bi-sectors [37] and Bregman Gaps bisectors). The Bregman bisector of a pair (e_1, e_2) is the set of equidistant points x with respect to a Bregman divergence:

$$BBs_J(e_1,e_2) = \{x, D_J(x,e_1) = D_J(x,e_2)\}.$$

The Bregman Gap bisector of a couple of dual quantities $([e_1, p_1], ([e_2, p_2])$ is the set of equidistant points x with respect to symmetrized Bregman Gap:

$$BGBs_J([e_1,p_1],[e_2,p_2]) = \{[e,p], BG^s_J([e,p],[e_1,p_1]) = BG^s_J([e,p],[e_2,p_2])\}.$$

Bisectors are hyperplanes, the two points or couples lie on the different sides of this hyperplane.

The Bregman divergence being not symmetric, there is another notion of Bregman bisector, namely of the second type [37] which is generally not an hyperplane.

$$BBs'_J(e_1,e_2) = \{x, D_J(e_1,x) = D_J(e_2,x)\}.$$

Regarding the Bregman balls, the classical approach for determining an orthogonal projection on a distance-defined ball cannot be followed because it relies strongly on the use of the triangle inequality. Let e be a point in IR^n and a Bregman ball of center c and radius ρ, with e not belonging to the closure of $BB_J(c,\rho)$. The projection $\tilde{e}$ on $BB_J(c,\rho)$ is:

$$\tilde{e} = \underset{x\in BB_J(c,\rho)}{\arg\min}\ D_J(x,e).$$

Because the ball is strictly convex, the projection lies on the boundary of $BB_J(c,\rho)$ or hypersphere $\partial BB_J(c,\rho)$, the projection $\tilde{e}$ can then be determined by the stationary point $(\tilde{e},\lambda)$ of the following Lagrangian:

$$\begin{aligned} L(x,\mu) &= D_J(x,c) + \mu(BB_J(x,c) - \rho) \\ &= (1+\mu)J(x) - J(e) - \mu J(c) - \langle \nabla J(e) + \mu \nabla J(c), x\rangle + \langle \nabla J(e), e\rangle + \mu\langle \nabla J(c), c\rangle. \end{aligned}$$

So that:

$$\begin{aligned} D_x L(\tilde{e},\lambda)\cdot \delta x &= (1+\lambda)\langle \nabla J(\tilde{e}), \delta x\rangle - \langle \nabla J(e) + \lambda \nabla J(c), \delta x\rangle \quad \forall \delta x \\ &= \langle (1+\lambda)\nabla J(\tilde{e}) - \nabla J(e) - \lambda \nabla J(c), \delta x\rangle \quad \forall \delta x \\ &= 0 \quad \forall \delta x \end{aligned}$$

and, setting $\eta = \lambda/(1+\lambda) \in]0,1[$, we obtain the characterization [Ca2]:

$$\nabla J(\tilde{e}) = (1-\eta)\nabla J(e) + \eta \nabla J(c).$$

In the very special case where J is quadratic, that is D_J is a Mahalanobis distance, we get that the projection $\tilde{e}$ of e lies is the intersection of the hypersphere and the line segment joining e and the center c of the ball. This is because the gradient at x is simply Ax with A a symmetric definite positive matrix. But in all other cases, the characterization of $\tilde{e}$ tells us that the alignment only takes place in the gradient space. Because the symmetric Bregman divergences are in fact true (squared) distances, and in view to derive efficient approximation procedures, the following definition of a subclass, relating a Bregman divergence to a symmetric one is of interest:

Definition (μ-similar Bregman divergences [42]). A Bregman divergence D_J defined on a domain K of a vector space V is μ-similar for some $0<\mu<1$ if there exists a symmetric Bregman divergence generated by $Q_\Box$ *such that:*

$$\mu D_{Q_\Box}(e_1,e_2) \le D_J(e_1,e_2) \le D_{Q_\Box}(e_1,e_2) \quad \forall (e_1,e_2)\in K^2.$$

$Q_\Box$ is necessarily a quadratic form, as shown previously, so that the Bregman divergence generated by $Q_\Box$ is a Mahalanobis distance. μ-similar Bregman divergences enjoy the following properties: they are approximately symmetric and satisfy a variant of the double triangle inequality [42]:

$$D_J\ \mu\text{-similar} \Rightarrow D_J(e_1,e_2) \le \frac{1}{\mu} D_J(e_2,e_1), \quad D_J(e_1,e_2) \le \frac{2}{\mu}[D_J(e_1,e_3) + D_J(e_2,e_3)] \quad \forall (e_1,e_2,e_3).$$

This concept can then be used for deriving approximated methods involving the Bregman divergence or to obtain quickly a first result to be further refined, because for example the associated Bregman balls can be bordered by ellipsoids. As quoted by [43], most Bregman divergences that are used in practice for clustering are μ-similar when restricted to a domain X that avoids the possible singularities of J. In the thermomechanics context however, this is not the case. As an example, the Bregman divergence generated by the function positively homogeneous of degree two $J(x) = a(\langle x\rangle^+)^2 + b(\langle x\rangle^-)^2$, with $a<b$, is μ-similar with $Q_\Box(x) = bx^2$ and $m = a/b$. But the Bregman divergence generated by the function $J(x) = a(\langle x\rangle^+)^2 + b(\langle x\rangle^-)^2 + \alpha|x|, \alpha>0$ is not (just examine the function $f(x) = D_J(x,-x)$).

5. Conclusion

In this paper, we try to extend the concept of Bregman divergence in thermomechanics, and we show how the large domain covered by the convex framework allows to systematically build discrepancy measures between fields of thermodynamic variables including couples of dual ones. Most of the already existing "errors" or "gaps" can then be encompassed. This gives rise to discrepancy measures between variables with different dimensions, encountered especially in multi-physic applications, allowing to consider truly coupled approaches. We also give some useful properties related to these concepts for geometrical computation that is at the heart of the algorithms used or already imagined for the processing of massive data produced by experiments, computations or in-service measurements.

The perspectives of direct applications are numerous. In inverse problems, the symmetrized Bregman Gaps have been used, even if not always under this denomination, for nonlinear Cauchy problems in mechanics and the subsequent identification problems (nonlinear elasticity with small transformations, [9], plastic zones identification [10], cracks identification in contact mechanics [11]). As mentioned in the introduction, replacement of usual least-squares functionals in various algorithms or procedures in nonlinear mechanics and data-driven approaches can also be of interest and is seemingly straightforward.

Also, adapted algorithms to Bregman divergence have been developed for some years in the field of statistical inference and data processing problems, such as estimation, detection, classification, compression, and recovery, in particular for classification or clustering applications, and even, more recently, in deep learning applications. The way is then paved for using the tools presented in this paper in new thermomechanics applications.

Future work will focus on the use of the Bregman divergence in (a revised form of) the Proper Orthogonal Decomposition, especially in nonlinear and coupled thermomechanics. The link with processing or using large amount of data being clear either for exploring the data (or efficiently store it, whether digital or experimental in origin), or for building reduced order models. The global objective remaining to use the Bregman tools for physically informed data processing.

Conflicts of interest

The author has no conflict of interest to declare.

Appendix A. Proof of the characterization of symmetric Bregman divergences

The condition for symmetry of D_J reads, for the generating convex function J

$$2(J(e_1)-J(e_2)) = \langle \nabla J(e_1)+\nabla J(e_2), e_1-e_2\rangle \quad \forall (e_1,e_2)\in \mathrm{IR}^n\times\mathrm{IR}^n.$$

Step 1: Let's suppose first that $e_2=0$, then: $2J(e)=\langle\nabla J(0)+\nabla J(e),e\rangle$ $\forall e\in\mathrm{IR}^n$, leading to the necessary condition: $J(e)=1/2\langle\nabla J(e),e\rangle$. Introduce now the real function f_e parameterized with e:

$$f_e(\lambda)=J(\lambda e),\quad \lambda\in\mathrm{IR}^+.$$

The derivative of f_e with respect to λ being $(\mathrm{d}/\mathrm{d}\lambda)f_e(\lambda)=\nabla J(\lambda e)\cdot e$, we have: $f_e(\lambda)=(\lambda/2)(\mathrm{d}/\mathrm{d}\lambda)f_e(\lambda)$. f_e is a positive function because J is convex, $J(0)$ and $\nabla J(0)$ are null, so:

$$f_e(\lambda)=J(\lambda e)-J(0)\geq\nabla J(0)\cdot\lambda e\equiv 0\Rightarrow f_e(\lambda)\geq 0.$$

Integrating, we get the result $f_e(\lambda)=K_e\lambda^2$ $\forall\lambda\in\mathrm{IR}^+$, leading to the following expression for J:

$$J(e)=\frac{\|e\|^2}{2}k\left(\frac{e}{\|e\|}\right)\quad\forall e\in\mathrm{IR}^n.$$

Note that the condition on J for the symmetry of D_J ensures recursively that f_e is infinitely directionally derivable in each direction e. This result also shows that J is necessarily positively homogeneous of degree 2.

Step 2: Going back to the condition for symmetry of J, and using the consequence of the property of positive homogeneity of the degree two ($\nabla J(e)\cdot e = 2J(e)$), we have now for any (e_1, e_2):

$$\begin{aligned} 2(J(e_1) - J(e_2)) &= \langle \nabla J(e_1) + \nabla J(e_2), e_1 - e_2\rangle \\ &= \langle \nabla J(e_1), e_1\rangle - \langle \nabla J(e_1), e_2\rangle + \langle \nabla J(e_2), e_1\rangle - \langle \nabla J(e_2), e_2\rangle \\ &= 2J(e_1) - \langle \nabla J(e_1), e_2\rangle + \langle \nabla J(e_2), e_1\rangle - 2J(e_2) \end{aligned}$$

so that $\langle \nabla J(e_2), e_1\rangle = \langle \nabla J(e_1), e_2\rangle\ \forall e_1, e_2$. Next, for any (e_1, e_2, e_3) and any (λ, μ)

$$\begin{aligned} \langle \nabla J(\lambda e_1 + \mu e_2), e_3\rangle &= \langle \nabla J(e_3), \lambda e_1 + \mu e_2\rangle \\ &= \lambda\langle \nabla J(e_3), e_1\rangle + \mu\langle \nabla J(e_3), e_2\rangle \\ &= \lambda\langle \nabla J(e_1), e_3\rangle + \mu\langle \nabla J(e_2), e_3\rangle \\ &= \langle \lambda\nabla J(e_1) + \mu\nabla J(e_2), e_3\rangle. \end{aligned}$$

So that $\nabla J(\lambda e_1 + \mu e_2) = \lambda\nabla J(e_1) + \mu\nabla J(e_2)$, showing that is $\nabla J(e)$ is linear, $\nabla J(e) = A_{ij}\langle e, E_j\rangle E_i$, with A necessarily a symmetric matrix, because:

$$A_{lk} = \langle \nabla J(E_l), E_k\rangle = \langle \nabla J(E_k), E_l\rangle = A_{kl}.$$

The function J is then necessarily a quadratic form: $J(e) = (1/2)\langle Ae, e\rangle$, $A = A^t$ recalling that $J(0) = \nabla J(0) = 0$.

Appendix B. Proof of separate convexity of the Symmetrized Bregman Gap

The following inequality has to be established.

$$\begin{aligned} &\forall([e_1, p_1], [e_2, p_2], [e_0, p_0]) BG_J^s(\lambda[e_1, p_1] + (1-\lambda)[e_2, p_2], [e_0, p_0]) \\ &\leq \lambda BG_J^s([e_1, p_1], [e_0, p_0]) + (1-\lambda) BG_J^s([e_2, p_2], [e_0, p_0]). \end{aligned}$$

Consider the two functions of λ:

$$\begin{aligned} F(\lambda) &= \langle \lambda e_1 + (1-\lambda)e_2 - e_0, \lambda p_1 + (1-\lambda)p_2 - p_0\rangle \\ G(\lambda) &= \lambda\langle e_1 - e_0, p_1 - p_0\rangle + (1-\lambda)\langle e_2 - e_0, p_2 - p_0\rangle. \end{aligned}$$

We are now led to show that the function $f(\lambda) = F(\lambda) - G(\lambda)$ is negative along the segment [0, 1]. But $f(0) = 0$, so that we can consider the derivative of f which turns out to be expressed by:

$$f'(\lambda) = (2\lambda - 1)\langle e_1 - e_2, p_1 - p_2\rangle = C(2\lambda - 1) \quad C \geq 0.$$

The function f can be calculated as $f(\lambda) = C\lambda(\lambda - 1)$ and is strictly negative for λ in $]0, 1[$ as soon as $C = BG_J^s([e_1, p_1], [e_2, p_2]) = \langle e_1 - e_2, p_1 - p_2\rangle > 0$, and null if $C = 0$.

Appendix C. Dual quantities involved in polyconvex elastic potentials

In the framework of finite transformations, the most used concepts are the gradient of the transformation F and the first Piola–Kirchhoff tensor T [44]. Mimicking the infinitesimal transformations framework an elastic law could be simply stated via an elastic potential as:

$$T = \frac{\partial W^{\mathrm{el}}}{\partial F}.$$

Unfortunately, the potential W^{el}, function of the three invariants of F is generally not convex with respect to F, so that the Coleman–Noll error

$$(T_1(F_1) - T_2(F_2)) : (F_1 - F_2)$$

is not always a positive expression. This error could be formally derived as the Bregman Gap with generating function W^{el}. But this last function being not convex, the formally defined Bregman divergence associate to W^{el} cannot be Bregman divergence. For studying the existence of solution in elasticity J. Ball [33] introduced the following notion of polyconvexity.

The potential W^{el} is polyconvex if it exists a convex function W of F, cof F, det F such that:

$$W^{\mathrm{el}}(F) = W(F, \operatorname{cof} F, \det F) \quad \text{for any } F\ s.t.\ \det F > 0$$

with cof $F = (\det F)F^{-T}$.

Then, introducing the duals quantities of F, cof F[1] et det F, namely T, cT, and p (where p is the actual pressure), the Bregman divergence generated by the now convex function W leads in turn to the Symmetrized Bregman Gap

$$(T_1 - T_2) : (M_1 - M_2) + (cT_1 - c_2 T) : (N_1 - N_2) + (p_1 - p_2)(d_1 - d_2)$$

which is then a positive quantity.

References

[1] C. C. Aggarwal, A. Hinneburg, D. A. Keim, "On the surprising behavior of distance metrics in high dimensional space", in *Database Theory — ICDT 2001* (J. Van den Bussche, V. Vianu, eds.), Lecture Notes in Computer Science, vol. 1973, Springer, Berlin, Heidelberg, 2001, p. 420-434.

[2] K. Beyer, J. Goldstein, R. Ramakrishnan, U. Schaft, "When is "nearest neighbor" meaningful?", in *Database Theory— ICDT'99. ICDT 1999* (C. Beeri, P. Buneman, eds.), Lecture Notes in Computer Science, vol. 1540, Springer, Berlin, Heidelberg, 1999.

[3] R. V. Kohn, M. Vogelius, "Relaxation of a variational method for impedance computed tomography", *Commun. Pure Appl. Math.* **40** (1987), p. 745-777.

[4] P. Ladevèze, D. Leguillon, "Error estimate procedure in the finite element method and applications", *SIAM J. Numer. Anal.* **20** (1983), p. 485-509.

[5] N. Moës, P. Ladevèze, B. Douchin, "Constitutive relation error estimators for (visco)plastic finite element analysis with softening", *Comput. Methods Appl. Mech. Eng.* **176** (1999), p. 247-264.

[6] P. Feissel, O. Allix, "Modified constitutive relation error identification strategy for transient dynamics with corrupted data: the elastic case", *Appl. Mech. Eng.* **196** (2007), p. 1968-1983.

[7] S. Andrieux, A. Ben Adba, H. D. Bui, "Sur identification de fissures planes via le concept d'écart à la réciprocité", *C. R. Acad. Sci. Paris Sér. II* **324** (1997), p. 431-438.

[8] D. Claire, F. Hild, S. Roux, "Identification of damage fields using kinematics measurements", *C. R. Méc.* **330** (2002), p. 729-734.

[9] S. Andrieux, T. N. Baranger, "Solution of nonlinear Cauchy problem for hyperelastic solids", *Inverse Probl.* **31** (2015), no. 11, p. 115003-115022.

[10] S. Andrieux, T. N. Baranger, "On the determination of missing boundary data for solids with nonlinear material behaviors, using displacement fields measured on a part of their boundaries", *J. Mech. Phys. Solids* **97** (2016), p. 140-155.

[11] S. Andrieux, T. N. Baranger, "Nonlinear Cauchy problem and identification in contact mechanics: a solving method based on Bregman-gap", *Inverse Probl.* **36** (2020), no. 11, article no. 115012.

[12] L. Bregman, "The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming", *USSR Comput. Math. Math. Phys.* **7** (1967), no. 3, p. 200-217.

[13] K. C. Kiwiel, "Proximal minimization methods with generalized Bregman functions", *SIAM J. Control Optim.* **35** (1997), no. 4, p. 1142-1168.

[14] Y. Censor, A. Lent, "An iterative row-action method for interval convex programming", *J. Optim. Theory Appl.* **34** (1981), no. 3, p. 321-353.

[1] Cof $A = (\det A)A^{-T}$.

[15] C. Basserville, “Divergence measures for statistical data processing–An annotated bibliography”, *Signal Process.* **93** (2013), p. 621-633.

[16] A. Banerjee, S. Merugu, I. S. Dhillon, J. Ghosh, “Clustering with Bregman divergences”, *J. Mach. Learn. Res.* **6** (2005), p. 1705-1749.

[17] R.-A. Dragomir, “Bregman gradient methods for relatively-smooth optimization”, PhD Thesis, UT1 Capitole, France, 2021.

[18] B. A. Frigyik, S. Srivastava, M. R. Gupta, “Functional Bregman divergence and Bayesian estimation of distributions”, *IEEE Trans. Inf. Theory* **54** (2008), no. 11, p. 1681-1685.

[19] Z. Zhang, B. C. Ooi, S. Parthasarathy, A. K. H. Tung, “Similarity search on Bregman divergence: towards non-metric indexing”, in *35th International Conference on Very Large Data Bases, VLBD09, Lyon, France*, 2009.

[20] H. Jeffreys, “An invariant form of the prior probability in estimation problems”, *Proc. R. Soc. Lond.* **186** (1946), p. 463-461.

[21] F. Nielsen, R. Nock, “Skew Jensen-Bregman Voronoi diagrams”, in *Transactions on Computational Science XIV*, Lecture Notes in Computer Science, vol. 6970, Springer, Berlin, Heidelberg, 2011, p. 102-128.

[22] P. Chen, Y. Chen, M. Rao, “Metrics defined by Bregman divergences”, *Commun. Math. Sci.* **6** (2008), p. 915-926.

[23] L. Wu, C. H. Choi, R. Jin, J. Zhu, N. Yu, “Learning Bregman distance for semi supervised clustering”, *IEEE Trans. Knowl. Data Eng.* **24** (2012), no. 3, p. 478-491.

[24] R. T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970.

[25] I. Ekeland, R. Témam, *Convex Analysis and Variational Problems*, SIAM, Philadephia, 1999.

[26] J. B. Hiriart-Urruty, C. Lemarechal, “Convex analysis and minimization algorithms II”, in *Advanced Theory and Bundle Methods*, Grundlehren der mathematischen Wissenschaften, vol. 306, Springer, Berlin, Heidelberg, 2013.

[27] A. Brøndsted, R. T. Rockafellar, “On the subdifferentiability of convex functions”, *Proc. Amer. Math. Soc.* **16** (1965), p. 605-611.

[28] S. Andrieux, T. N. Baranger, A. Ben Abda, “Solving Cauchy problem by minimizing an energy-like functional”, *Inverse Probl.* **22** (2006), p. 115-133.

[29] B. Halphen, Q. S. Nguyen, “Sur les matériaux standards généralisés”, *J. Méc.* **14** (1975), p. 39-63.

[30] D. C. Drucker, “A more fundamental approach to plastic stress-strain relations”, in *Proceedings of the First US Congress of Applied Mechanics*, ASME, New York, 1951, p. 487-491.

[31] J. J. Marigo, “From Clausius-Duhem and Drücker-Ilyushin inequalities to standard materials”, in *Continuum Thermomechanics: The Art and Science of Modelling Material behaviour, Paul Germain's Anniversary Volume* (A. Maugin, R. Douot, F. Sidoroff, eds.), Solid Mechanics and Its Applications, vol. 76, Kluwer Academic Publishers, Springer, Dordrecht, 2000.

[32] P. Germain, Q. S. Nguyen, P. Suquet, “Continuum thermodynamics”, *Trans. ASME* **50** (1983), p. 1010-1020.

[33] J. M. Ball, “Convexity conditions and existence theorems in nonlinear elasticity”, *Arch. Rat. Mech. Anal.* **65** (1977), p. 193-201.

[34] G. de Saxcé, Z. Q. Feng, “New inequation and functional for contact with friction: the implicit standard material approach”, *Mech. Struct. Mach.* **19** (1991), no. 3, p. 301-325.

[35] A. Banerjee, X. Guo, H. Wang, “On the optimality condition expectations as a Bregman predictor”, *IEEE Trans. Inf. Theory* **51** (2005), p. 2664-2669.

[36] R. Nock, F. Nielsen, “Fitting the smallest enclosing Bregman ball”, in *Machine Learning: ECML 2005* (J. Gama, R. Camacho, P. B. Brazdil, A. M. Jorge, L. Torgo, eds.), Lecture Notes in Computer Science, vol. 3720, Springer, Berlin, 2005.

[37] J. D. Boissonnat, F. Nielsen, R. Nock, “Bregman Voronoi diagrams”, *Discrete Comput. Geom.* **44** (2010), p. 281-307.

[38] F. Nielsen, R. Nock, “On the smallest enclosing information disk”, *Inf. Process. Lett.* **105** (2005), p. 93-97.

[39] A. Fischer, “Quantization and clustering with Bregman divergences”, *J. Multivar. Anal.* **101** (2010), no. 9, p. 2207-2221.

[40] M. Luci, O. Bachem, A. Krause, “String coresets for hard and soft Bregman Clustering with application to exponential family mixture”, in *Proceedings of 19th International Conference on Artificial Intelligence and Statistics (AISTATS), Cadiz, Spain*, 2016.

[41] L. Cayton, “Efficient Bregman range search”, in *Advances in Neural Information Processing Systems 22–Proceedings of the 2009 Conference*, 2009, p. 243-251.

[42] M. R. Ackermann, J. Blömer, “Coresets and approximate clustering for Bregman divergences”, in *International Symposium on Discrete Algorithms (SODA)*, SIAM, 2009, p. 1088-1097.

[43] L. Cayton, “Fast nearest neighbor retrieval for Bregman divergences”, in *Proceedings of the 25th International Conference on Machine Learning, Helsinki, Finland*, 2008.

[44] P. G. Ciarlet, *Mathematical Elasticity. Vol. I: Three-Dimensional Elasticity*, North Holland, Amsterdam, 1988.