Hindawi Publishing Corporation
Advances in Difference Equations
Volume 2011, Article ID 617564, 9 pages
doi:10.1155/2011/617564

Research Article

Solution to a Function Equation and Divergence Measures

Chuan-Lei Dong and Jin Liang

Department of Mathematics, Shanghai Jiao Tong University, Shanghai 200240, China

Correspondence should be addressed to Jin Liang, jinliang@sjtu.edu.cn

Received 1 January 2011; Accepted 11 February 2011

Academic Editor: Toka Diagana

We investigate the solution to the following function equation $f_1(x)g_1(y)+\cdots+f_n(x)g_n(y) = G(x+y)$, which arises from the theory of divergence measures. Moreover, new results on divergence measures are given.

1. Introduction

As early as in 1952, Chernoff [1] used the α-divergence to evaluate classification errors. Since then, the study of various divergence measures has been attracting many researchers. So far, we have known that the Csiszár f-divergence is a unique class of divergences having information monotonicity, from which the dual α geometrical structure with the Fisher metric is derived, and the Bregman divergence is another class of divergences that gives a dually flat geometrical structure different from the α-structure in general. Actually, a divergence measure between two probability distributions or positive measures have been proved a useful tool for solving optimization problems in optimization, signal processing, machine learning, and statistical inference. For more information on the theory of divergence measures, please see, for example, [2–5] and references therein.

Motivated by these studies, we investigate in this paper the solution to the following function equation

$$f_1(x)g_1(y) + \cdots + f_n(x)g_n(y) = G(x+y), \tag{1.1}$$

which arises from the discussion of the theory of divergence measures, and show that for $n > 1$, if $f_i : [a,b] \to R$, $g_i : [a,b] \to R$, $i = 1,2,\ldots,n$, and $G : [2a,2b] \to R$ satisfy

$$\sum_{i=1}^{n} f_i(x)g_i(y) = G(x+y), \tag{1.2}$$

then G is the solution of a linear homogenous differential equation with constant coefficients. Moreover, new results on divergence measures are given.

Throughout this paper, we let R be the set of real numbers and $F \subset R^n$ are a convex set.

Basic notations: $R_+^n := \{x \in R^n : x_i > 0, i = 1, 2, \ldots, n\}$; $\phi : F \to R$ is strictly convex and twice differentiable; $\pi : R_+^n \to F$ is differentiable injective map; D_ϕ^π is the general vector Bregman divergence; $f : (0, +\infty) \to [0, +\infty)$ is strictly convex twice-continuously differentiable function satisfying $f(1) = 0$, $f'(1) = 0$; D_f is the vector f-divergence.

If for every $p, q \in R_+^n$,

$$D_\phi^\pi[p : q] = D_f[p : q], \tag{1.3}$$

then we say the D_ϕ^π or D_f is in the intersection of f-divergence and general Bregman divergence.

For more information on some basic concepts of divergence measures, we refer the reader to, for example, [2–5] and references therein.

2. Main Results

Theorem 2.1. *Assume that there are differentiable functions*

$$f_i : [a, b] \longrightarrow R, \quad g_i : [a, b] \longrightarrow R, \quad i = 1, 2, \ldots, n, \tag{2.1}$$

and $G : [2a, 2b] \to R$ *such that*

$$\sum_{i=1}^{n} f_i(x) g_i(y) = G(x + y), \quad \text{for every } x, y \in [a, b]. \tag{2.2}$$

Then $G \in C^\infty[2a, 2b]$ *and*

$$a_n G^{(n)} + a_{n-1} G^{(n-1)} + \cdots + a_1 G' + a_0 G = 0, \tag{2.3}$$

for some $a_n, a_{n-1}, \ldots, a_0 \in R$.

Proof. Since f_i, g_i is differentiable functions, it is clear that

$$f_i, g_i \in L^2[a, b], \quad i = 1, 2, \ldots, n. \tag{2.4}$$

Let

$$M = \operatorname{span}\{f_1, f_2, \ldots, f_n\}. \tag{2.5}$$

Then M is a finite dimension space. So we can find differentiable functions

$$s_1, s_2, \ldots, s_m \in M \tag{2.6}$$

as the orthonormal bases of M, where $m \le n$. Observing that

$$\begin{aligned}\sum_{i=1}^{n} f_i(x)g_i(y) &= \sum_{i=1}^{n}\left[\sum_{j=1}^{m} a_{ij}s_j(x)g_i(y)\right]\\ &= \sum_{j=1}^{m} s_j(x)\sum_{i=1}^{n} a_{ij}g_i(y)\\ &= \sum_{j=1}^{m} s_j(x)t_j(y),\end{aligned} \tag{2.7}$$

where

$$a_{ij} \in R, \quad t_j(y) = \sum_{i=1}^{n} a_{ij}g_i(y), \quad i = 1,2,\ldots,n, \quad j = 1,2,\ldots,m, \tag{2.8}$$

we have

$$G(x+y) = \sum_{i=1}^{n} f_i(x)g_i(y) = \sum_{j=1}^{m} s_j(x)t_j(y), \text{ for every } x, y \in [a,b]. \tag{2.9}$$

Clearly,

$$t_j \in L^2[a,b], \quad j = 1,\ldots,m. \tag{2.10}$$

Next we prove that

$$t_j \in M, \quad j = 1,\ldots,m. \tag{2.11}$$

It is easy to see that we only need to prove the following fact:

$$\operatorname{span}\{s_1, s_2, \ldots, s_m, t_1, t_2, \ldots, t_m\} = M. \tag{2.12}$$

Actually, if this is not true, that is,

$$\operatorname{span}\{s_1, s_2, \ldots, s_m, t_1, t_2, \ldots, t_m\} \ne M, \tag{2.13}$$

then there exists $t \ne 0$ such that

$$t \in \operatorname{span}\{s_1, s_2, \ldots, s_m, t_1, t_2, \ldots, t_m\}, \quad t \perp M. \tag{2.14}$$

Therefore

$$\begin{aligned}
\int_a^b G(x+y)t(x)dx &= \int_a^b \sum_{i=1}^m s_i(x)t(x)t_i(y)dx \\
&= \sum_{i=1}^m \int_a^b s_i(x)t(x)dx\, t_i(y) \\
&= 0, \quad \text{for every } y \in [a,b],
\end{aligned} \tag{2.15}$$

$$\begin{aligned}
\int_a^b G(y+x)t(y)dy &= \int_a^b \sum_{i=1}^m s_i(x)t(y)t_i(y)dy \\
&= \sum_{i=1}^m \int_a^b t_i(y)t(y)dy\, s_i(x), \quad \text{for every } x \in [a,b].
\end{aligned}$$

Because

$$\int_a^b G(x+y)t(x)dx = 0, \quad \text{for every } y \in [a,b], \tag{2.16}$$

we get

$$\int_a^b G(y+x)t(y)dy = 0, \quad \text{for every } x \in [a,b], \tag{2.17}$$

that is,

$$\sum_{i=1}^m \int_a^b t_i(y)t(y)dy\, s_i(x) = 0, \quad \text{for every } x \in [a,b]. \tag{2.18}$$

Since $s_1, s_2, \ldots, s_m$ is linearly independent, we see that

$$\int_a^b t_i(y)t(y)dy = 0. \tag{2.19}$$

So

$$t \perp \operatorname{span}\{s_1, s_2, \ldots, s_m, t_1, t_2, \ldots, t_m\}. \tag{2.20}$$

This is a contradiction. Hence (2.12) holds, and so does (2.11). Thus, there are $b_{ij} \in R\,(i = 1,2,\ldots,m,\ j = 1,2,\ldots,m)$ such that

$$t_i = b_{ij}s_j, \quad i = 1,2,\ldots,m,\ j = 1,2,\ldots,m. \tag{2.21}$$

Therefore,

$$\begin{aligned} G(x+y) &= \sum_{i=1}^{m} s_i(x)t_i(y) = \sum_{i,j=1}^{m} b_{ij}s_i(x)s_j(y), \quad \text{for every } x,y \in [a,b], \\ G(y+x) &= \sum_{i=1}^{m} s_i(y)t_i(x) = \sum_{i,j=1}^{m} b_{ij}s_i(y)s_j(x), \quad \text{for every } x,y \in [a,b]. \end{aligned} \tag{2.22}$$

So we have

$$G(x+y) = \sum_{i,j=1}^{m} \frac{b_{ij}+b_{ji}}{2} s_i(x)s_j(y), \quad \text{for every } x,y \in [a,b]. \tag{2.23}$$

Define

$$c_{ij} := \frac{b_{ij}+b_{ji}}{2}, \quad i = 1,2,\dots,m, \quad j = 1,2,\dots,m. \tag{2.24}$$

Then

$$G(x+y) = \sum_{i,j=1}^{m} c_{ij}s_i(x)s_j(y), \quad \text{for every } x,y \in [a,b]. \tag{2.25}$$

Let $S = (s_1, s_2, \dots, s_m)$, and

$$C = \begin{pmatrix} c_{11} & c_{12} & \cdots & c_{1m} \\ c_{21} & c_{22} & \cdots & c_{2m} \\ \vdots & \vdots & \vdots & \vdots \\ c_{n1} & c_{m1} & \cdots & c_{mm} \end{pmatrix}. \tag{2.26}$$

Then

$$G(x+y) = \sum_{i,j=1}^{m} c_{ij}s_i(x)s_j(y) = S(x)CS(y)^T, \quad \text{for every } x,y \in [a,b]. \tag{2.27}$$

Since C is a symmetric matrix, we have

$$C = Q\Lambda Q^T. \tag{2.28}$$

for an orthogonal matrix Q, and a diagonal matrix

$$\Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{pmatrix}. \tag{2.29}$$

Write

$$W = (r_1, r_2, \dots, r_m) = (s_1, s_2, \dots, s_m)Q. \tag{2.30}$$

Then

$$G(x+y) = S(x)CS(y)^T = W(x)\Lambda W(y)^T, \quad \text{for every } x, y \in [a,b]. \tag{2.31}$$

So, for all $x, y \in [a,b]$,

$$G(x+y) = \begin{pmatrix} r_1(x) & \cdots & r_m(x) \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{pmatrix} \begin{pmatrix} r_1(y) \\ \vdots \\ r_m(y) \end{pmatrix}. \tag{2.32}$$

Without loss the generalization, we can assume that

$$\lambda_1, \lambda_2, \dots, \lambda_m \neq 0. \tag{2.33}$$

Thus, for all $x, y \in [a,b]$,

$$\frac{\partial G(x+y)}{\partial x} = \begin{pmatrix} r_1(x) & \cdots & r'_m(x) \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{pmatrix} \begin{pmatrix} r_1(y) \\ \vdots \\ r_m(y) \end{pmatrix}. \tag{2.34}$$

By the similar arguments as above, we can prove

$$\operatorname{span}\{r_1, \dots, r_m, r'_1, \dots, r'_m\} = \operatorname{span}\{r_1, \dots, r_m\}. \tag{2.35}$$

So there is a matrix A satisfying

$$\begin{pmatrix} r'_1 & \cdots & r'_m \end{pmatrix} = \begin{pmatrix} r_1 & \dots & r_m \end{pmatrix} A. \tag{2.36}$$

Thus,

$$G'(x+y) = \frac{\partial G(x+y)}{\partial x} = R(x)A\Lambda R(y)^T. \tag{2.37}$$

By mathematical induction we obtain

$$G^{(i)}(x+y) = R(x)A^i\Lambda R(y)^T, \quad \forall i = 0,1,\ldots \tag{2.38}$$

So $G \in C^\infty[2a,2b]$.

Let

$$b_0 + b_1\lambda + \cdots + b_m\lambda^m \tag{2.39}$$

be the annihilation polynomial of A. Then

$$\begin{aligned} &b_0G(x+y) + b_1G'(x+y) + \cdots + b_mG^{(m)}(x+y) \\ &= \sum_{i=0}^{m} b_i R(x)A^i\Lambda R(y)^T \\ &= R(x)\sum_{i=0}^{m} b_i A^i\Lambda R(y) \\ &= 0. \end{aligned} \tag{2.40}$$

Since $n \geq m$, we can find $a_n,\ a_{n-1},\ldots,\ a_0 \in R$ such that

$$a_nG^{(n)} + a_{n-1}G^{(n-1)} + \cdots + a_1G' + a_0G = 0. \tag{2.41}$$

The proof is then complete. □

Theorem 2.2. *Let the f-divergence D_f be in the section of f-divergence and general Bregman divergence. Then $G(x) = f''(e^x)$ satisfies*

$$\sum_{i=0}^{n} a_i G^{(i)} = 0, \tag{2.42}$$

for some $a_n,\ldots,a_0 \in R$.

Proof. If D_f, D_ϕ^π are in the intersection of f-divergence and general Bregmen divergence, then we have

$$xf\left(\frac{y}{x}\right)n = \phi(\pi(X)) - \phi(\pi(Y)) - \sum_{i=1}^{n}\frac{\partial\phi(\pi(Y))}{\partial x_i}(\pi_i(X) - \pi_i(Y)), \quad \forall x,y \in (0,+\infty), \tag{2.43}$$

where

$$X = (x, x, \dots, x) \in R^n, \qquad Y = (y, y, \dots, y) \in R^n. \tag{2.44}$$

Let

$$\frac{\partial \phi(\pi(Y))}{\partial x_i} = s_i(y), \quad \pi_i(X) = t_i(x). \tag{2.45}$$

Then

$$\frac{\partial^2 x f(y/x) n}{\partial x \partial y} = \frac{\partial^2 \left[\phi(\pi(X)) - \phi(\pi(Y)) - \sum_{i=1}^{n} s_i(y)\left(t_i(x) - t_i(y)\right)\right]}{\partial x \partial y}. \tag{2.46}$$

Hence

$$\frac{y}{x^2} f''\left(\frac{y}{x}\right) = \sum_{i=1}^{n} s_i'(y) t_i'(x). \tag{2.47}$$

Let

$$G(x) = f''(e^x), \qquad f_i(x) = \frac{s_i'(e^x)}{e^x}, \qquad g_i(x) = t_i(e^{-x}) e^{-2x}. \tag{2.48}$$

Then

$$G(x+y) = \sum_{i=1}^{n} f_i(x) g_i(y). \tag{2.49}$$

Thus, a modification of Theorem 2.1 implies the conclusion. □

Moreover, it is not so hard to deduce the following theorem.

Theorem 2.3. *Let a vector f-divergence is are the intersection of vector f-divergence and general Bregman divergence and π satisfy*

$$\pi(x) = (\pi_1(x_1), \dots, \pi_n(x_n)), \quad \forall x \in R_+^n, \tag{2.50}$$

where $\pi_1, \dots, \pi_n$ is strictly monotone twice-continuously differentiable functions. Then the f divergence is α-divergence or vector α-divergence times a positive constant c.

Acknowledgments

This work was supported partially by the NSF of China and the Specialized Research Fund for the Doctoral Program of Higher Education of China.

References

[1] H. Chernoff, "A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations," *Annals of Mathematical Statistics*, vol. 23, pp. 493–507, 1952.

[2] S. Amari, "Information geometry and its applications: convex function and dually flat manifold," in *Emerging Trends in Visual Computing*, F. Nielsen, Ed., vol. 5416 of *Lecture Notes in Computer Science*, pp. 75–102, Springer, Berlin, Germany, 2009.

[3] L. M. Brègman, "The relaxation method of finding a common point of convex sets and its application to the solution of problems in convex programming," *Computational Mathematics and Mathematical Physics*, vol. 7, pp. 200–217, 1967.

[4] A. Cichocki, R. Zdunek, A. H. Phan, and S. Amari, *Non-Negative Matrix and Tensor Factorizations: Applications to Explanatory Multi-Way Data Analysis and Blind Source Separation*, Wiley, New York, NY, USA, 2009.

[5] I. Csiszár, "Why least squares and maximum entropy? An axiomatic approach to inference for linear inverse problems," *The Annals of Statistics*, vol. 19, no. 4, pp. 2032–2066, 1991.