Int. J. Anal. Appl. (2023), 21:57

International Journal of Analysis and Applications

A PDE Approach to the Problems of Optimality of Expectations

Mahir Hasanov*

Department of Mathematics, Istanbul Beykent University, Türkiye

**Corresponding author:* hasanov61@yahoo.com

Abstract. Let (X, Z) be a bivariate random vector. A predictor of X based on Z is just a Borel function $g(Z)$. The problem of "least squares prediction" of X given the observation Z is to find the global minimum point of the functional $E[(X - g(Z))^2]$ with respect to all random variables $g(Z)$, where g is a Borel function. It is well known that the solution of this problem is the conditional expectation $E(X|Z)$. We also know that, if for a nonnegative smooth function $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $arg\ min_{g(Z)} E[F(X, g(Z))] = E[X|Z]$, for all X and Z, then $F(x, y)$ is a Bregmann loss function. It is also of interest, for a fixed φ to find $F(x, y)$, satisfying, $arg\ min_{g(Z)} E[F(X, g(Z))] = \varphi(E[X|Z])$, for all X and Z. In more general setting, a stronger problem is to find $F(x, y)$ satisfying $arg\ min_{y \in \mathbb{R}} E[F(X, y)] = \varphi(E[X])$, $\forall X$. We study this problem and develop a partial differential equation (PDE) approach to solution of these problems.

1. Introduction and Preliminary Facts

Best approximation problems in Mathematics have long history of study. It is known that for every given x in a Hilbert space H and every given closed subspace L of H there is a unique best approximation to x out of L (namely, $y = Px$, where P is the orthogonal projection of H onto L) (see [8] and [11]). Theorem 1.1 below, regarding the optimality of conditional expectations with respect to L_2 loss function $F(x, y) = (x - y)^2$ follows from this result.

Theorem 1.1. *(see [1], [9], [13]) Let (X, Z) be a bivariate random vector and $L_Z = \{g(Z)|g(Z) \in L_2(\Omega),\ g$ is a Borel function$\}$. Let $E[X^2] < \infty$. Then there exists a Borel function $g_0 : \mathbb{R} \to \mathbb{R}$ with $E[(g_0(Z)^2] < \infty$, such that $E[(X - g_0(Z))^2] = inf\{E[(X - g(Z))^2 \big| g(Z) \in L_Z\}$. Moreover, $g_0(Z) = E[X|Z]$.*

Received: Apr. 14, 2023.

2020 *Mathematics Subject Classification.* 60A05, 49K45 .

Key words and phrases. expectation; conditional expectation; random variables; Bregman loss functions: partial differential equations.

https://doi.org/10.28924/2291-8639-21-2023-57
© 2023 the author(s).
ISSN: 2291-8639

This theorem means that the distance function $||X-Y||_2^2$ attains its minimum value at $Y=\psi(Z)=E[X|Z]$. Thus,

$$arg\ min_{Y\in L_Z}||X-Y||_2^2=E[X|Z]. \tag{1.1}$$

We recall some basic notions and facts from probability theory in the form we use in this paper ([1], [9], [13]).

Expectation. Let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space and $X:\Omega\to\mathbb{R}$ be a random variable. By the definition, a random variable is measurable, i.e., $X^{-1}(\sigma_B)\subset\mathcal{F}$, where σ_B is the Borel algebra, consisting of all Borel sets in $\mathbb{R}$. The expectation of a random variable X is defined by the following integral, which is Lebesgue integral with respect to the probability measure.

$$E[X]=\int_\Omega X\,d\mathbb{P}.$$

Particularly, for a simple random variable $X(w)=\sum_{i=1}^n a_i\chi_{A_i}(w)$,

$$E[X]=\sum_{i=1}^n a_iP(A_i). \tag{1.2}$$

$$L_2(\Omega)=\{X|\ \int_\Omega|X|^2\,d\mathbb{P}<\infty\}.$$

The norm in $L_2(\Omega)$ is defined by $||X||_2=\left(\int_\Omega|X|^2\,d\mathbb{P}\right)^{\frac{1}{2}}$.

Conditional Expectation. Let (X,Z) be a bivariate random vector. The conditional expectation of X given Z is denoted by $E[X|Z]$, which is a random variable, defined by

$$\psi(Z)(w)=\psi(Z(w))=E[X|Z=Z(w)],\forall w\in\Omega.$$

The following problem is a natural generalization of the problem (1.1), which has very important applications (see [2] and references therein); find a loss function $F(x,y)$ satisfying the following condition

$$arg\ min_{y\in\mathbb{R}}E[F(X,y)]=\varphi(E[X]),\ \forall X, \tag{1.3}$$

where φ is a Borel function. In this paper our main concern will be the problem (1.3). Such problems arise in different contexts of statistics and probability theory (see [4]). In the case of $\varphi(x)=x$; $F(x,y)=C(x-y)$ and $F(x,y)=(x-y)^2$ the optimality of conditional expectations have been studied by many authors (see [1], [9], [10], [13]). For $\varphi(x)=x$ and arbitrary function $F(x,y)$ the Bregman loss functions play an important role ([5], [6], [7]). Particularly, it was proved in [2] (see Theorem 1.2 below) that if for a nonnegative smooth function $F:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$, $arg\ min_{g(Z)}E[F(X,g(Z))]=E[X|Z]$, for all X and Z, then $F(x,y)$ is a Bregmann loss function.

Definition 1.1. *Let $f:\mathbb{R}\to\mathbb{R}$ be a strictly convex differentiable function. Then the Bregman Loss Function (BLF) $D_f:\mathbb{R}\times\mathbb{R}\to\mathbb{R}$ is defined as*

$$D_f(x,y)=f(x)-f(y)-f'(y)(x-y)$$

In general, Bregman loss functions are defined by using strictly convex differentiable functions $f : \mathbb{R}^n \to \mathbb{R}$. In this paper, for convenience we consider the case $n = 1$. All results can easily be extended to the case $n > 1$. For more information on Bregman loss functions see [3] and [12].

The following theorem contains the most general result, regarding problem 1.3 in the case of $\varphi(x) = x$.

Theorem 1.2. *([2]) Let $D_f : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$ be a BLF. Then,*

$$arg\ min_{Y \in L_Z} E[D_f(X, Y)] = E[X|Z].$$

Moreover, if $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $F \geq 0$, $F(x, x) = 0$, F and F_x are continuous functions and for all X and Z, $arg\ min_{Y \in L_Z} E[D_f(X, Y)] = E[X|Z]$ then F is a BLF.

The rest of this paper will be organized as follows. In Section 2 we present a theorem about optimality of expectations. Section 3 consists of two subsections. In subsection 3.1 we develop a partial differential equation approach for critical points of $E[F(X, y)]$. The main problem studied in this subsection is: when $y = \varphi(E[X])$ is a critical point of the function $E[F(X, y)]$ for every $X \in L_1(\Omega)$? We present a partial differential equation approach for solving this problem and give a necessary and sufficient condition. In subsection 3.2 we study extreme problems. Our main goal is to find the class of all F such that $y = \varphi(E[X])$ is *a unique* extremum point for $E[F(X, y)]$, for all $X \in L_1(\Omega)$.

2. On the Optimality of Expectations

We start with a slightly stronger version of Theorem 1.2.

Theorem 2.1. *Let $F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $F \geq 0$, $F(x, x) = 0$, F_x and F_y are continuous. Suppose that there exists a function $\varphi : \mathbb{R} \to \mathbb{R}$ such that $\varphi(E[X])$ is a unique minimizer for $E[F(X, y)]$ in $\mathbb{R}$ for all $X \in L_1(\Omega)$,i.e.,*

$$arg\ min_{y \in \mathbb{R}} E[F(X, y)] = \varphi(E[X]), \forall X \in L_1(\Omega),$$

provided that $F(X, y) \in L_1(\Omega)$. Then $F(x, y)$ is a BLF if and only if $\varphi(x) = x$.

Proof. let $F(x, y)$ be a BLF. Then,

$$F(x, y) = f(x) - f(y) - f'(y)(x - y).$$

We can write

$$F(X, y) = f(X) - f(y) - f'(y)(X - y)$$

and

$$F(X, E[X]) = f(X) - f(E[X]) - f'(E[X])(X - E[X]).$$

Hence,

$$F(X, y) - F(X, E[X]) = f(E[X]) - f(y) + f'(E[X])(X - E[X]) - f'(y)(X - y).$$

Obviously,

$$E\big[f'(E[X])(X - E[X])\big] = 0 \text{ and } E\big[f'(y)(X - y)\big] = f'(y)(E[X] - y).$$

Then,

$$E\big[F(X, y) - F(X, E[X])\big] = f(E[X]) - f(y) - f'(y)(E[X] - y).$$

Consequently,

$$E\big[F(X, y) - F(X, E[X])\big] = D_f(E[X], y) \geq 0. \tag{2.1}$$

Since $F(x, y) = D_f(x, y)$ is a BLF, $D_f(E[X], y) = 0 \Leftrightarrow y = E[X]$. Thus $y = E[X]$ is a minimum point of $E[F(X, y)]$. By the condition $\varphi(E[X])$ is a unique minimizer. Then, it follows immediately that $\varphi(x) = x$.

Now let $\varphi(x) = x$. and

$$arg\ min_{y\in\mathbb{R}} E[F(X, y)] = E[X], \forall X \in L_1(\Omega).$$

Then it follows from this condition that F is a BLF. This case was proved in [2] (see Theorem 3). □

3. A PDE Approach to Optimality Problems

3.1. **Critical Points.** In this section we develop a partial differential equation (PDE) approach for critical points of $E[F(X, y)]$. More precisely, the main question is: when $y = \varphi(E[X])$ is a critical point of the function $E[F(X, y)]$ for every X? We give a necessary and sufficient condition for this question.

The following assumption will be needed throughout this section.
$F : \mathbb{R} \times \mathbb{R} \to \mathbb{R}$, $F(x, x) = 0$, and the function F has first and second derivatives. Now we prove a critical point theorem.

Theorem 3.1. *Let $\varphi : \mathbb{R} \to \mathbb{R}$ be an invertible function. Then, $y = \varphi(E[X])$ is a critical point of the function $E[F(X, y)]$ for all $X \in L_1(\Omega)$,* **if and only if** *$F(x, y)$ is a solution of the following PDE*

$$F_{xy}(\varphi^{-1}(y) - x) + F_y = 0. \tag{3.1}$$

Proof. Let $y = \varphi(E[X])$ be a critical point of the function $E[F(X, y)]$ for all $X \in L_1(\Omega)$. Consider a simple random variable X such that $P(X = a) = p$, $P(X = b) = q$ and $p + q = 1$.

By (1.2)

$$E[F(X, y)] = pF(a, y) + qF(b, y).$$

and

$$\varphi(E[X]) = \varphi(pa + qb).$$

Then

$$pF_y(a, \varphi(pa + qb)) + pF_y(b, \varphi(pa + qb)) = 0.$$

It means that

$$\frac{F_y(a,\varphi(pa+qb))}{q} = -\frac{F_y(b,\varphi(pa+qb))}{p} \Leftrightarrow$$

$$\frac{F_y(a,\varphi(pa+qb))}{q(b-a)} = -\frac{F_y(b,\varphi(pa+qb))}{p(b-a)}. \tag{3.2}$$

$y = \varphi(E[X]) \Rightarrow y = \varphi(pa+qb) \Rightarrow pa+qb = \varphi^{-1}(y)$. Note that

$$pa+qb-a = q(b-a) \text{ and } pa+qb-b = -p(b-a).$$

Hence,

$$\varphi^{-1}(y) - a = q(b-a) \text{ and } \varphi^{-1}(y) - b = -p(b-a).$$

It follows from equation (3.2) that

$$\frac{F_y(a,y)}{\varphi^{-1}(y)-a} = \frac{F_y(b,y)}{\varphi^{-1}(y)-b}.$$

Therefore, the function $\frac{F_y(x,y)}{\varphi^{-1}(y)-x}$ does not depend on x. Then

$$\frac{\partial}{\partial x}\Big[\frac{F_y(x,y)}{\varphi^{-1}(y)-x}\Big] = 0$$

and

$$\frac{F_{xy}(\varphi^{-1}(y)-x)+F_y}{(\varphi^{-1}(y)-x)^2} = 0.$$

Consequently,

$$F_{xy}(\varphi^{-1}(y)-x)+F_y = 0.$$

To finish the proof of this theorem, we need to show that the (3.1) implies $y = \varphi(E[X])$ is a critical point of the function $E[F(X,y)]$ for all $X \in L_1(\Omega)$. Thus, by (3.1)

$$F_{xy}(\varphi^{-1}(y)-x)+F_y = 0.$$

Multiplying, this equation by the integrating factor $\mu(x,y) = \frac{1}{(\varphi^{-1}(y)-x)^2}$ we get

$$\frac{1}{\varphi^{-1}(y)-x}F_{xy} + \frac{1}{(\varphi^{-1}(y)-x)^2}F_y = 0.$$

Then,

$$\Big(\frac{1}{\varphi^{-1}(y)-x}F_y\Big)_x = 0 \text{ and } \frac{F_y}{\varphi^{-1}(y)-x} = C(y) \Rightarrow F_y = (\varphi^{-1}(y)-x)C(y).$$

Setting $y = \varphi(E[X])$ we get

$$F_y(X,\varphi(E[X])) = \big(E[X]-X\big)C(\varphi(E[X]))$$

and

$$E\big[F_y(X,\varphi(E[X])\big] = \big(E[X]-E[X]\big)C(\varphi(E[X])) = 0.$$

□

We next give an application of this theorem.

Example 3.1. *Let us find a general solution of the following problem*

$$F_{xy}(\varphi^{-1}(y)-x)+F_y=0,\ \ F(x,x)=0$$

in the case of $\varphi(y)=y$.

Solution. We can write the equation in the form

$$F_{xy}+\frac{1}{y-x}F_y=0.$$

Multiplying, this equation by the integrating factor $\mu(x,y)=\frac{1}{y-x}$ we get

$$\frac{1}{y-x}F_{xy}+\frac{1}{(y-x)^2}F_y=0.$$

Then,

$$\Big(\frac{1}{y-x}F_y\Big)_x=0 \text{ and } \frac{F_y}{y-x}=C(y).$$

Let $C(y)=f''(y)$. By using integration by parts we obtain that

$$\int_x^y F_y(x,t)\,dt=\int_x^y f''(t)(t-x))dt=\Big[f'(t)(t-x)\Big]_{t=x}^{t=y}-\int_x^y f(t)\,dt.$$

Consequently,

$$F(x,y)=f(x)-f(y)-f'(y)(x-y).$$

The following corollary immediately follows from this example and Theorem 3.1.

Corollary 3.1. *If* $F(x,x)=0$ *and* $y=E[X]$ *is a critical point of the function* $E[F(X,y)]$ *for all* $X\in L_1(\Omega)$*, then* $F(x,y)$ *can be written in the form* $F(x,y)=f(x)-f(y)-f'(y)(x-y)$ *for a differentiable function* f.

Not. By imposing additional conditions: $F(x,y)\geq 0$ and $E[X]$ is the unique minimizer, it was proved in [2] that F is a BLF.

3.2. **Extreme Points.** Let $\varphi:\mathbb{R}\to\mathbb{R}$ be an invertible function. In this subsection the main problem is to find the class of all F such that $y=\varphi(E[X])$ is *a unique* extremum point for $E[F(X,y)]$, for all $X\in L_1(\Omega)$. We first prove the following theorem.

Theorem 3.2. *Let*

$$arg\ min_{y\in\mathbb{R}}E[F(X,y)]=\varphi(E[X]),\forall X\in L_1(\Omega).$$

then

$$F(x,y)=\big(\varphi^{-1}(y)-x\big)f'(y)-\big(\varphi^{-1}(x)-x\big)f'(x)-\int_x^y f'(t)\big(\varphi^{-1}(t)\big)'\,dt, \tag{3.3}$$

where f is a differentiable function satisfying the following condition

$$\big(\varphi^{-1}(y)-x\big)f'(y)+\int_y^{\varphi(x)} f'(t)\big(\varphi^{-1}(t)\big)'dt>0,\ \forall y\neq\varphi(x), \tag{3.4}$$

Proof. By Theorem 3.1

$$F_{xy}(\varphi^{-1}(y)-x)+F_y=0,\ F(x,x)=0.$$

Then,

$$\Big(\frac{1}{\varphi^{-1}(y)-x}F_y\Big)_x=0 \text{ and } \frac{F_y}{\varphi^{-1}(y)-x}=C(y).$$

Setting $C(y)=f''(y)$ we can write

$$F_y=\big(\varphi^{-1}(y)-x\big)f''(y). \tag{3.5}$$

Using integration by parts in (3.5) we obtain that

$$\int_x^y F_y(x,t)\,dt=\int_x^y(\varphi^{-1}(t)-x))\,df'(t)=\Big[f'(t)(\varphi^{-1}(t)-x)\Big]_{t=x}^{t=y}-\int_x^y f'(t)\big(\varphi^{-1}(t)\big)'\,dt.$$

Consequently,

$$F(x,y)=\big(\varphi^{-1}(y)-x\big)f'(y)-\big(\varphi^{-1}(x)-x\big)f'(x)-\int_x^y f'(t)\big(\varphi^{-1}(t)\big)'\,dt$$

and (3.3) holds.

Now we use the condition

$$arg\ min_{y\in\mathbb{R}}E[F(X,y)]=\varphi(E[X]).$$

This condition means that

$$E\Big[F(X,y)-F(X,\varphi\big(E[X]\big)\Big]>0,$$

provided that $y\neq\varphi\big(E[X]\big)$. Using (3.3) we obtain that

$$E\Big[F(X,y)-F(X,\varphi\big(E[X]\big)\Big]=\big(\varphi^{-1}(y)-E[X]\big)f'(y)+\int_y^{\varphi(E[X])} f'(t)\big(\varphi^{-1}(t)\big)'\,dt>0.$$

Thus,

$$\big(\varphi^{-1}(y)-x\big)f'(y)+\int_y^{\varphi(x)} f'(t)\big(\varphi^{-1}(t)\big)'dt>0,\ \forall y\neq\varphi(x).$$

□

Note. In case of $\varphi(x)=x$,

$$\big(\varphi^{-1}(y)-x\big)f'(y)+\int_y^{\varphi(x)} f'(t)\big(\varphi^{-1}(t)\big)'dt>0,\ \forall y\neq\varphi(x)\Rightarrow$$

$$f(x)-f(y)-f'(y)(x-y)>0,\ x\neq y.$$

and

$$F(x,y) = \big(\varphi^{-1}(y) - x\big) f'(y) - \big(\varphi^{-1}(x) - x\big) f'(x) - \int_x^y f'(t)\big(\varphi^{-1}(t)\big)' \, dt \Rightarrow$$
$$F(x,y) = f(x) - f(y) - f'(y)(x-y).$$

Therefore, in case of $\varphi(x) = x$, the condition (3.4) means that f is a is strictly convex function and (3.3) means simply that $F(x,y)$ is a Bregman loss function.

Corollary 3.2. *Let*

$$arg\ max_{y\in\mathbb{R}} E[F(X,y)] = \varphi(E[X]), \forall X \in L_1(\Omega).$$

Then

$$F(x,y) = \big(\varphi^{-1}(y) - x\big) f'(y) - \big(\varphi^{-1}(x) - x\big) f'(x) - \int_x^y f'(t)\big(\varphi^{-1}(t)\big)' \, dt$$

and

$$\big(\varphi^{-1}(y) - x\big) f'(y) + \int_y^{\varphi(x)} f'(t)\big(\varphi^{-1}(t)\big)' dt < 0, \ \forall y \neq \varphi(x).$$

Finally, we discus the condition (3.4), which is a generalization of the strictly convexity condition. The main question is: are there functions satisfying the following inequality

$$\big(\varphi^{-1}(y) - x\big) f'(y) + \int_y^{\varphi(x)} f'(t)\big(\varphi^{-1}(t)\big)' dt > 0, \ \forall y \neq \varphi(x).$$

Regarding this question, we prove the following theorem.

Theorem 3.3. *If $\varphi(x)$ is an increasing function and $f''(x) > 0, \forall x \in \mathbb{R}$. Then*

$$\big(\varphi^{-1}(y) - x\big) f'(y) + \int_y^{\varphi(x)} f'(t)\big(\varphi^{-1}(t)\big)' dt > 0, \ \forall y \neq \varphi(x).$$

Proof. Let us define

$$G(x,y) = \big(\varphi^{-1}(y) - x\big) f'(y) + \int_y^{\varphi(x)} f'(t)\big(\varphi^{-1}(t)\big)' dt.$$

Then,

$$G_y(x,y) = \big(\varphi^{-1}(y)\big)' f'(y) + \big(\varphi^{-1}(y) - x\big) f''(y) - \big(\varphi^{-1}(y)\big)' f'(y) \Rightarrow$$

$$G_y(x,y) = \big(\varphi^{-1}(y) - x\big) f''(y).$$

We have

$$y > \varphi(x) \Leftrightarrow \varphi^{-1}(y) - x > 0 \Leftrightarrow G_y(x,y) > 0,$$
$$y < \varphi(x) \Leftrightarrow \varphi^{-1}(y) - x < 0 \Leftrightarrow G_y(x,y) < 0$$

and $G_y(x, \varphi(x)) = 0$. Consequently,

$$G(x,y) = \big(\varphi^{-1}(y) - x\big) f'(y) + \int_y^{\varphi(x)} f'(t)\big(\varphi^{-1}(t)\big)' dt > 0, \ \forall y \neq \varphi(x).$$

□

Corollary 3.3. *If $\varphi(x)$ is a decreasing function and $f''(x) > 0, \forall x \in \mathbb{R}$. Then*

$$\big(\varphi^{-1}(y) - x\big)f'(y) + \int_{y}^{\varphi(x)} f'(t)\big(\varphi^{-1}(t)\big)' dt < 0,\ \ \forall y \neq \varphi(x).$$

Conflicts of Interest: The author declares that there are no conflicts of interest regarding the publication of this paper.

References

[1] K.B. Athreya, S.N. Lahiri, Measure Theory and Probability Theory, Springer Texts in Statistics, Springer, New York, 2006.

[2] A. Banerjee, X. Guo, H. Wang, On the Optimality of Conditional Expectation as a Bregman Predictor, IEEE Trans. Inform. Theory. 51 (2005), 2664–2669. https://doi.org/10.1109/tit.2005.850145.

[3] H.H. Bauschke, M.S. Macklem, J.B. Sewell, X. Wang, Klee Sets and Chebyshev Centers for the Right Bregman Distance, J. Approx. Theory. 162 (2010), 1225–1244. https://doi.org/10.1016/j.jat.2010.01.001.

[4] A. Ben-Tal, A. Charnes, M. Teboulle, Entropic Means, J. Math. Anal. Appl. 139 (1989), 537–551. https://doi.org/10.1016/0022-247x(89)90128-5.

[5] Y. Censor, S. Zenios, Parallel Optimization: Theory, Algorithms, and Applications, Oxford University Press, London, 1998.

[6] I. Csiszar, Why Least Squares and Maximum Entropy? An Axiomatic Approach to Inference for Linear Inverse Problems, Ann. Stat. 19 (1991), 2032–2066. https://doi.org/10.1214/aos/1176348385.

[7] I. Csiszar, Generalized Projections for Non-Negative Functions, in: Proceedings of 1995 IEEE International Symposium on Information Theory, IEEE, Whistler, BC, Canada, 1995: p. 6. https://doi.org/10.1109/ISIT.1995.531108.

[8] F. Deutsch, Best Approximation in Inner Product Spaces, Springer-Verlag, New York, 2021.

[9] G. Grimmett, D. Stirzaker, Probability and Random Processes, Oxford University Press, Oxford, 2004.

[10] S. Karlin, H.M. Taylor, A Second Course in Stochastic Processes, 2nd ed. Academic Press, San Diego, 1991.

[11] M. Hasanov, The Spectra of Two-Parameter Quadratic Operator Pencils, Math. Computer Model. 54 (2011), 742–755. https://doi.org/10.1016/j.mcm.2011.03.018.

[12] D. Reem, S. Reich, A. De Pierro, Re-Examination of Bregman Functions and New Properties of Their Divergences, Optimization. 68 (2018), 279–348. https://doi.org/10.1080/02331934.2018.1543295.

[13] D. Williams, Probability with Martingales, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 2001.