Trans. Natl. Acad. Sci. Azerb. Ser. Phys.-Tech. Math. Sci.
38 (2018), no. 4, Mathematics, 39-45.

Cauchy functional equation and representation by ridge functions

Aysel A. Asgarova

Received: 15.05.2018 / Revised: 03.08.2018 / Accepted: 01.10.2018

Abstract. *In the current paper, we show the relevance of some classical results on the Cauchy functional equation to the problem of representation by sums of ridge functions with finitely many directions. We prove that if a sum of ridge functions belongs to a class of functions with difference property then under suitable conditions each summand also belongs to this class. We also give some practical corollaries of this result.*

Keywords. Cauchy functional equation, difference property, ridge function

Mathematics Subject Classification (2010): 26B40, 39B22

1 Introduction

A *ridge function* is a multivariate function of the form

$$g\left(\mathbf{a}{\cdot}\mathbf{x}\right)=g\left(a_1x_1+\ldots+a_nx_n\right),$$

where $g:\mathbb{R}\rightarrow\mathbb{R}$, $\mathbf{a}=(a_1,...,a_n)$ is a fixed vector (direction) in $\mathbb{R}^n\backslash\{\mathbf{0}\}$, $\mathbf{x}=(x_1,...,x_n)$ is the variable and $\mathbf{a}{\cdot}\mathbf{x}$ is the usual inner product in $\mathbb{R}^n$. In the theory of partial differential equations, ridge functions have been known under the name of *plane waves (*see, e.g., [18]). They appear as general solutions of some homogeneous hyperbolic type equations. For example, assume that $(a_i,b_i),\ i=1,...,r$, are pairwise linearly independent vectors in $\mathbb{R}^2$. Then the general solution to the equation

$$\prod_{i=1}^{r}\left(a_i\frac{\partial}{\partial x}+b_i\frac{\partial}{\partial y}\right)f\left(x,y\right)=0,$$

where the derivatives are understood in the sense of distributions, are all functions of the form

$$f(x,y)=\sum_{i=1}^{r}g_i\left(b_ix-a_iy\right)$$

for arbitrary continuous univariate functions g_i, $i=1,...,r$.

A.A. Asgarova
Azerbaijan University of Languages, AZ 1014, Baku, Azerbaijan
E-mail: asgarova2016@mail.ru

The term “ridge function” was devised by Logan and Shepp in their pioneering paper [26] dedicated to the mathematics of computerized tomography (see also [19,20,28,29]). After a 1981 paper by Friedman and Stuetzle [11] ridge functions started to appear also in statistics, especially, in the theory of projection pursuit and projection regression (see, e.g., [6,7,10–12]). The general idea therein was to reduce “dimension” and thus bypass the “curse of dimensionality”.

Ridge functions are used in many models in neural network theory. For example, in one of the popular models called MLP (multilayer feedforward perceptron) model, the simplest case considers functions of the form

$$\sum_{i=1}^{r} c_i\sigma(\mathbf{w}^i{\cdot}\mathbf{x}-\theta_i).$$

Here the weights $\mathbf{w}^i$ are vectors in $\mathbb{R}^n$, the thresholds θ_i and the coefficients c_i are real numbers and the activation function σ is a univariate function. Note that for each $\theta \in \mathbb{R}$ and $\mathbf{w} \in \mathbb{R}^n\backslash\{\mathbf{0}\}$ the function

$$\sigma(\mathbf{w}\cdot\mathbf{x}-\theta)$$

is a ridge function. For an extensive study of approximation properties of the MLP model see [34].

Ridge functions are interesting also to approximation theorists. In approximation theory, these functions are implemented as an effective and convenient tool for approximating complicated multivariate functions (see, e.g., [13–17,23,25,27,30,33]).

In this paper, we consider the problem of representation by sums of ridge functions with r, $r \geq 1$, fixed directions. Let the directions $\mathbf{a}^i \in \mathbb{R}^n\backslash\{\mathbf{0}\}$, $i = 1, ..., r$, be given and pairwise linearly independent. Assume we know that a function $f(\mathbf{x})$ can be represented in the form

$$f(\mathbf{x}) = \sum_{i=1}^{r} g_i(\mathbf{a}^i\cdot\mathbf{x}). \tag{1.1}$$

Assume in addition that f is of the class $C^k(\mathbb{R}^n)$. What can we say about g_i? Can we say that $g_i \in C^k(\mathbb{R})$? The case $r = 1$ is obvious. In this case, if $f \in C^k(\mathbb{R}^n)$, then for $\mathbf{c} \in \mathbb{R}^n$ satisfying $\mathbf{a}^1\cdot\mathbf{c} = 1$ we have that $g_1(t) = f(t\mathbf{c})$ is in $C^k(\mathbb{R})$. The same argument can be carried out for the case $r = 2$. In this case, since the vectors $\mathbf{a}^1$ and $\mathbf{a}^2$ are linearly independent, there exists a vector $\mathbf{c} \in \mathbb{R}^n$ satisfying $\mathbf{a}^1\cdot\mathbf{c} = 1$ and $\mathbf{a}^2\cdot\mathbf{c} = 0$. Therefore, we obtain that the function $g_1(t) = f(t\mathbf{c}) - g_2(0)$ is in the class $C^k(\mathbb{R})$. Similarly, one can verify that $g_2 \in C^k(\mathbb{R})$ (see [2]).

The above questions becomes quite difficult if the number of directions $r \geq 3$. For $r = 3$, there are many smooth functions which decompose into sums of very badly behaved ridge functions. This is a consequence of the classical Cauchy Functional Equation (CFE). This equation is defined as

$$h(x+y) = h(x) + h(y),\ h:\mathbb{R}\to\mathbb{R}, \tag{1.2}$$

which has a class of simple solutions $h(x) = cx$, $c \in \mathbb{R}$. However, it easily follows from the Hamel basis theory that CFE has also a large class of badly behaved solutions. These solutions are called “badly behaved” because they are weird over reals. They are, for example, not continuous at a point, not monotone at an interval, not bounded at any set of positive measure (see, e.g., [1]). Let h_1 be any such solution of the equation (1.2). Then the zero function can be written as

$$0 = h_1(x) + h_1(y) - h_1(x+y). \tag{1.3}$$

Note that the functions involved in (1.3) are ridge functions with the directions $(1,0)$, $(0,1)$ and $(1,1)$ respectively. This particular example shows that for smoothness of the representation (1.1) one must impose additional conditions on the functions g_i, $i = 1, ..., r$.

It was first proved by Buhman and Pinkus [5] that if in (1.1) $f \in C^k(\mathbb{R}^n)$, $k \geq r-1$ and $g_i \in L^1_{loc}(\mathbb{R})$ for each i, then $g_i \in C^k(\mathbb{R})$ for $i = 1, ..., r$. In [32] Pinkus found a strong relationship between CFE and the problem of smoothness in ridge function representation. He generalized extensively the previous result of Buhman and Pinkus [5]. He showed that the solution is quite simple and natural if the functions g_i are taken from a class $\mathcal{B}$ of real-valued functions u defined on $\mathbb{R}$. By definition, u is in $\mathcal{B}$ if for any function $v \in C(\mathbb{R})$ for which $u - v$ satisfies CFE, $u - v$ is linear, i.e. $u(x) - v(x) = cx$, where $c \in \mathbb{R}$ (see [32]). The result of Pinkus states that if in (1.1) $f \in C^k(\mathbb{R}^n)$ and each $g_i \in \mathcal{B}$, then necessarily $g_i \in C^k(\mathbb{R})$ for $i = 1, ..., r$.

The above representation problem was also considered by Konyagin and Kuleshov [21, 22] and by Kuleshov [24]. They mainly analyze the continuity of the representation, that is, the question if and when continuity of f in (1.1) guarantees the continuity of g_i. There are also other results concerning the smoothness of ridge function representation generalizing the above result of Pinkus (see [24]). The results in [21,22,24] involve certain subsets (convex open sets, convex bodies, etc.) of $\mathbb{R}^n$ instead of only $\mathbb{R}^n$ itself.

The results of Pinkus [32] give rise to the following natural and important problem. Assume in the representation (1.1) $f \in C^k(\mathbb{R}^n)$, but the functions g_i are arbitrarily behaved (that is, we allow very badly behaved functions). Can we write f as a sum $\sum_{i=1}^r f_i(\mathbf{a}^i \cdot \mathbf{x})$ but with the $f_i \in C^k(\mathbb{R})$, $i = 1, ..., r$? This problem was posed in [5] and [31]. In [2], Aliev and Ismailov gave a partial solution to this problem. Their solution comprises the cases in which $r-1$ directions of given r directions are linearly independent. Note that this condition is satisfied by default if we are given three directions, as it is assumed that all the directions are pairwise linearly independent. The representation problem in the case of three directions was initially considered in [4]. For bivariate functions having the degree of smoothness $k \geq r-2$, the problem was completely solved in [3].

In this paper we generalize the result of Pinkus in such a way that instead of the pair $C^k(\mathbb{R}^n)$ and $\mathcal{B}$, one would be able to take many other important pairs of function classes. As a practical example, for each natural number k we suggest a special function class $\mathcal{B}_k$, which is wider than $\mathcal{B}$, and show that if a function $f \in C^k(\mathbb{R}^n)$ has the representation (1.1) and $g_i \in \mathcal{B}_k$, $i = 1, ..., r$, then $g_i \in C^k(\mathbb{R})$, $i = 1, ..., r$.

2 Main result

In [32], A.Pinkus considered the problem of smoothness in ridge function representation. For a given function f represented by (1.1), he posed and answered the following question. If f belongs to some smoothness class, what can we say about the smoothness of the functions g_i? He proved that for a large class of representing functions, the representation is smooth. That is, if a priori assume that in the representation (1.1), the functions g_i is of a certain class of "quite well behaved functions", then they have the same degree of smoothness as the function f. As the mentioned class of "quite well behaved functions" one may take, e.g., the set of continuous functions, the set of Lebesgue measurable functions, etc. All these classes come from the class $\mathcal{B}$ considered by Pinkus [32] and the classical theory of CFE (Cauchy Functional Equation). In [32], $\mathcal{B}$ denotes any translation invariant (that is, $f(\cdot + t) \in \mathcal{B}$ for any $t \in \mathbb{R}$ if $f \in \mathcal{B}$) linear space of real-valued functions u defined on $\mathbb{R}$ such that if there is a function $v \in C(\mathbb{R})$ for which $u - v$ satisfies CFE, then $u - v$ is necessarily linear, i.e. $u(x) - v(x) = cx$, for some constant $c \in \mathbb{R}$. Such definition of $\mathcal{B}$ is required in the proof of the following theorem.

Theorem 2.1 *(Pinkus [32]). Assume $f \in C^k(\mathbb{R}^n)$ is of the form (1.1). Assume, in addition, that each $g_i \in \mathcal{B}$. Then necessarily $g_i \in C^k(\mathbb{R})$ for $i = 1, ..., r$.*

Below we prove more general version of the result of Pinkus. Let $\mathcal{A}$ denote any class of functions $f : \mathbb{R} \to \mathbb{R}$ with the property that if f satisfies CFE, then $f = cx$. These classes (namely, in this form) appear almost everywhere in the literature on CFE (see, e.g., [1]). Besides, we assume that $\mathcal{A}$ is difference invariant, that is, $f(\cdot + t) - f(\cdot) \in \mathcal{A}$ for any $t \in \mathbb{R}$ if $f \in \mathcal{A}$. Simple examples of $\mathcal{A}$ are the sets of continuous, bounded and Lebesgue measurable functions.

Along with the classes $\mathcal{A}$, we consider the classes of functions having the difference property. Let $\mathcal{D}$ denote any class of functions with the property that if $\triangle_t f = f(\cdot + t) - f(\cdot) \in \mathcal{D}$ for all $t \in \mathbb{R}$, then $f - s \in \mathcal{D}$, for some s satisfying CFE. Several classes with the difference property are investigated in de Bruijn [8,9]. Some of these classes are:

1) $C(\mathbb{R})$, continuous functions;
2) $C^k(\mathbb{R})$, functions with continuous derivatives up to order k;
3) $C^\infty(\mathbb{R})$, infinitely differentiable functions;
4) analytic functions;
5) functions which are absolutely continuous on any finite interval;
6) functions having bounded variation over any finite interval;
7) algebraic polynomials;
8) trigonometric polynomials;
9) Riemann integrable functions.

In the sequel, we assume that the considered class $\mathcal{D}$ contains linear functions and forms a linear space.

We define the following relation between the classes $\mathcal{A}$ and $\mathcal{D}$.

Definition 2.1 *Let $\mathcal{A}$ and $\mathcal{D}$ be two classes of functions defined above. We say that the classes $\mathcal{A}$ and $\mathcal{D}$ are compatible if for any pair $f \in \mathcal{A}$ and $g \in \mathcal{D}$, we have $f - g \in \mathcal{A}$.*

For example, the classes $\mathcal{A}$ =$C^m(\mathbb{R})$ and $\mathcal{D}$ =$C^k(\mathbb{R})$ are compatible if $m \leq k$, the class $\mathcal{A}$ of functions continuous on the real axis is compatible with the class $\mathcal{D}$ of algebraic polynomials, etc.

The following theorem is valid. Its proof is based on the ideas set forth in [32].

Theorem 2.2 *Let $\mathcal{A}$ and $\mathcal{D}$ be any two compatible classes of functions. Assume $f(\mathbf{x}) = \sum_{i=1}^{r} g_i(\mathbf{a}^i \cdot \mathbf{x})$, where $f(\mathbf{c}t) \in \mathcal{D}$ for any $\mathbf{c} \in \mathbb{R}^n$ and $g_i \in \mathcal{A}$. Then $g_i \in \mathcal{D}$.*

Proof. We prove this theorem by induction on r. The result is valid when $r = 1$. Indeed, for any direction $\mathbf{c}$ orthogonal to $\mathbf{a}^1$, we can write that $g_1(t) = f(\mathbf{c}t) \in \mathcal{D}$. Assume that the result is valid for $r - 1$.

Chose any vector $\mathbf{e} \in \mathbb{R}^n$ satisfying $\mathbf{e} \cdot \mathbf{a}^r = 0$ and $\mathbf{e} \cdot \mathbf{a}^i = b_i \neq 0$, for $i = 1, ..., r - 1$. Clearly, there exists a vector with this property. The property of $\mathbf{e}$ enables us to write that

$$f(\mathbf{x} + \mathbf{e}t) - f(\mathbf{x}) = \sum_{i=1}^{r-1} g_i(\mathbf{a}^i \cdot \mathbf{x} + b_i t) - g_i(\mathbf{a}^i \cdot \mathbf{x}). \tag{2.1}$$

Consider the following functions

$$F(\mathbf{x}) = f(\mathbf{x} + \mathbf{e}t) - f(\mathbf{x})$$

and

$$h_i(y) = g_i(y + b_i t) - g_i(y), \; i = 1, ..., r - 1.$$

Then (3.1) can be written as

$$F(\mathbf{x}) = \sum_{i=1}^{r-1} h_i(\mathbf{a}^i \cdot \mathbf{x}).$$

Note that $F(\mathbf{c}t) \in \mathcal{D}$ for any $\mathbf{c} \in \mathbb{R}^n$ and $h_i \in \mathcal{A}$ for all $t \in \mathbb{R}$. Thus it follows by our induction assumption that $h_i \in \mathcal{D}$ for all $t \in \mathbb{R}$.

Now it is not difficult to prove that the result is valid also for r. From the definition of $\mathcal{D}$ and the functions h_i we obtain that $g_i - s_i \in \mathcal{D}$, for some s_i satisfying CFE. Put $u_i = g_i - s_i$, $i = 1, ..., r-1$. Since the classes $\mathcal{A}$ and $\mathcal{D}$ are compatible, one can see that the functions $s_i = g_i - u_i$ are in the class $\mathcal{A}$. But then by the definition of $\mathcal{A}$, $s_i = c_i x$, for some constants c_i, $i = 1, ..., r-1$. Thus we conclude that the functions $g_i = u_i + s_i$ are in the class $\mathcal{D}$. This is valid for $i = 1, ..., r-1$. It is not difficult to see that g_r is also in the class $\mathcal{D}$. (To see this instead of the vector $\mathbf{e}$ take any vector $\mathbf{d}$ such that $\mathbf{d} \cdot \mathbf{a}^1 = 0$ and $\mathbf{d} \cdot \mathbf{a}^i \neq 0$, for $i = 2, ..., r$, and repeat the above process). Thus from the assumption that the result is valid for $r-1$ we has derived that it is valid for r. Theorem 2.2 has been proved.

One can easily formulate many useful and applicable corollaries of Theorem 2.2. Below we formulate three of them. The first corollary extends the class of $\mathcal{B}$ considered by Pinkus. Let for each $k \in \mathbb{N}$, $\mathcal{B}_k$ be any class of difference invariant functions u defined on $\mathbb{R}$ such that if there is a function $v \in C^k(\mathbb{R})$ for which $u - v$ satisfies CFE, then $u(x) - v(x) = cx$, for some constant $c \in \mathbb{R}$. Clearly, $\mathcal{B}_k$ may contain more functions than $\mathcal{B}$. Let, for example, s be any badly behaved solution of CFE, t be a nowhere differentiable continuous function and $u = s + t$. Then u cannot be in $\mathcal{B}$, but it can be in $\mathcal{B}_k$, since there is no function $v \in C^k(\mathbb{R})$ such that $u - v$ satisfies CFE.

Corollary 2.1 *Assume $f \in C^k(\mathbb{R}^n)$ is of the form (1.1). Assume, in addition, that each $g_i \in \mathcal{B}_k$. Then necessarily $g_i \in C^k(\mathbb{R})$ for $i = 1, ..., r$.*

The next corollary is about representation of polynomials by sums of ridge functions.

Corollary 2.2 *Assume a multivariate polynomial function f is of the form (1.1). Assume, in addition, that each g_i, $i = 1, ..., r$, is continuous. Then g_i are univariate polynomials.*

The next corollary follows from the fact that the class of continuous functions has the difference property (see [8]) and is compatible with the class of Lebesgue measurable functions.

Corollary 2.3 *Assume a continuous function f is of the form (1.1). Assume, in addition, that each g_i, $i = 1, ..., r$, is Lebesgue measurable. Then all the functions g_i are continuous.*

References

1. Aczél, J.: Functional Equations and their Applications, *Academic Press, New York* (1966)
2. Aliev, R.A., Ismailov, V.E.: *On a smoothness problem in ridge function representation*, Adv. in Appl. Math. **73** (2016), 154–169.
3. Aliev, R.A., Ismailov, V.E.: *On the representation by bivariate ridge functions*, arXiv preprint https://arxiv.org/abs/1606.07940.
4. Aliev, R.A., Ismailov, V.E., Shahbalayeva, T.M.: *On the representation by sums of ridge functions*, Proc. Inst. Math. Mech. Natl. Acad. Sci. Azerb. **41**(2) (2015), 106–118.
5. Buhmann, M.D., Pinkus, A.: *Identifying linear combinations of ridge functions*, Adv. in Appl. Math. **22** (1999), 103-118.

6. Candes, E.J.: *Ridgelets: estimating with ridge functions*, Ann. Statist. **31** (2003), 1561–1599.
7. Candes, E.J.: Ridgelets: theory and applications. *Ph.D. Thesis, Technical Report, Department of Statistics, Stanford University* (1998).
8. de Bruijn, N.G.: *Functions whose differences belong to a given class*, Nieuw Arch. Wiskd. **23** (1951), 194–218.
9. de Bruijn, N.G.: *A difference property for Riemann integrable functions and for some similar classes of functions*, Indag. Math. **14** (1952), 145–151.
10. Donoho, D.L., Johnstone, I.M.: *Projection-based approximation and a duality method with kernel methods*, Ann. Statist. **17** (1989), 58-106.
11. Friedman, J.H., Stuetzle, W.: *Projection pursuit regression*, J.Amer. Statist. Assoc. **76** (1981), 817-823.
12. Huber, P.J.: *Projection pursuit*, Ann. Statist. **13** (1985), 435-475.
13. Ismailov, V.E.: *Approximation by sums of ridge functions with fixed directions*, (Russian); translated from Algebra i Analiz **28**(6) (2016), 20–69, St. Petersburg Math. J. **28**(6) (2017), 741-772.
14. Ismailov, V.E.: *A note on the equioscillation theorem for best ridge function approximation*, Expo. Math. **35**(3) (2017), 343–349.
15. Ismailov, V.E.: *Approximation by ridge functions and neural networks with a bounded number of neurons*. Appl. Anal. **94** (11) (2015), 2245–2260.
16. Ismailov, V.E.: *On the best L_2 approximation by ridge functions*, Appl. Math. E-Notes **7** (2007), 71-76.
17. Ismailov, V.E.: *Characterization of an extremal sum of ridge functions*, J. Comp. Appl. Math. **205** (2007), 105–115.
18. John, F.: Plane Waves and Spherical Means Applied to Partial Differential Equations, *Interscience, New York* (1955).
19. Kazantsev, I.: *Tomographic reconstruction from arbitrary directions using ridge functions*, Inverse Problems **14** (1998), 635–645.
20. Kazantsev, I., Lemahieu, I.: *Reconstruction of elongated structures using ridge functions and natural pixels*, Inverse Problems **16** (2000), 505–517.
21. Konyagin, S.V., Kuleshov, A.A.: *On the continuity of finite sums of ridge functions* (Russian), translated from Mat. Zametki, **98** (2) (2015), 308–309, Math. Notes **98**(1-2) (2015), 336–338.
22. Konyagin, S.V., Kuleshov, A.A.: *On some properties of finite sums of ridge functions defined on convex subsets of $\mathbb{R}^n$.* (Russian) Tr. Mat. Inst. Steklova **293**, Funktsionalnye Prostranstva, Teoriya Priblizheniĭ, Smezhnye Razdely Matematicheskogo An. (2016), 193–200.
23. Kroo, A.: *On approximation by ridge functions*, Constr. Approx. **13** (1997), 447–460.
24. Kuleshov, A.A.: *On some properties of smooth sums of ridge functions* (Russian), Tr. Mat. Inst. Steklova **294**, Sovremennye Problemy Matematiki, Mekhaniki i Matematicheskoĭ Fiziki. II (2016), 99–104.
25. Lin, V.Ya, Pinkus, A.: *Fundamentality of ridge functions*, J. Approx. Theory **75** (1993), 295-311.
26. Logan, B.F., Shepp, L.A.: *Optimal reconstruction of a function from its projections*, Duke Math.J. **42** (1975), 645–659.
27. Maiorov, V.E.: *On best approximation by ridge functions*, J. Approx. Theory **99** (1999), 68–94.
28. Marr, R.B.: *On the reconstruction of a function on a circular domain from a sampling of its line integrals*, J. Math. Anal. Appl. **45** (1974), 357–374.
29. Natterer, F.: The Mathematics of Computerized Tomography, *Wiley, New York* (1986).
30. Petrushev, P.P.: *Approximation by ridge functions and neural networks*, SIAM J.Math. Anal. **30** (1998), 155-189.

31. Pinkus, A.: Ridge Functions, Cambridge Tracts in Mathematics, *205. Cambridge University Press, Cambridge* (2015).
32. Pinkus, A.: *Smoothness and uniqueness in ridge function representation*, Indag. Math. (N.S.) **24**(4) (2013), 725–738.
33. Pinkus, A.: Approximating by ridge functions, in: Surface Fitting and Multiresolution Methods (A.Le Méhauté, C.Rabut and L.L.Schumaker, eds), *Vanderbilt Univ.Press (Nashville)*, 279-292 (1997).
34. Pinkus, A.: *Approximation theory of the MLP model in neural networks*, Acta Numerica **8** (1999), 143-195.