Rotation Invariant Kernels on Spheres

Hongjun Su[1], Hong Zhang[2*]

[1] Department of Electrical and Computer Engineering, Georgia Southern University, Savannah, GA, USA.
[2] Department of Computer Science, Georgia Southern University, Savannah, GA, USA.

* Corresponding author. Tel.: 912-344-3151; email: hzhang@georgiasouthern.edu
Manuscript submitted December 29, 2020; accepted February 1, 2021.
doi: 10.17706/ijapm.2021.11.3.43-49

Abstract: Kernels as similarity measures are key components of machine learning algorithms such as Support Vector Machine and Gaussian Process. Invariant kernels are an effective way to incorporate prior knowledge in applications with the invariance property. In this paper, we characterize all the rotation invariant kernels on spheres. We show that such a kernel is a function of the dot product of the input vectors alone. This function can be expanded as a series of Chebyshev polynomials with non-negative coefficients. In a 2-D space this condition is also sufficient. On a 3-D sphere the function can be expanded as a series of Legendre polynomials with non-negative coefficients. In general, a necessary and sufficient condition for the rotation invariant kernel is that the function on dot products can be expanded as a series of Gegenbauer polynomials with non-negative coefficients.

Key words: Bochner's theorem, Fourier transform, Gegenbauer polynomial, invariant kernel.

1. Introduction

A kernel is a similarity measure that is the key component of support vector machines ([1],[2]) and other machine learning techniques ([3]-[6]). Often the performance of a kernel based machine learning system depends highly on the effectiveness of the kernel to represent the similarity between objects. Capturing the special structural information of the data in the kernels can be useful in improving the system performance.

A kernel is a real valued function on two variables that is symmetric and positive definite.

$$K(x,y) = K(y,x),\ \textstyle\sum_{i,j=1}^{n} K(x_i,x_j)c_i c_j \geq 0$$

A related concept is the function of positive type on a group G. A function $f: G \to \mathbb{C}$ is of positive type or positive definite if

$$\sum_{i,j=1}^{n} f(x_i {x_j}^{-1}) c_i \overline{c_j} \geq 0$$

A positive definite kernel can be defined from a positive definite function by

$$K(x,y) = f(xy^{-1})$$

Constructing kernels invariant to a group of transformations is one of the effective ways to incorporate prior knowledge and to improve the performance of a machine learning system.

Let $K(x,y)$ be a kernel on X and G a group of transformations of X. The kernel is invariant under G if

$$K(Tx,Ty) = K(x,y), \text{ for all } T \in G$$

The structure of an invariant kernel is dependent on the space and the group of transformations. In the Euclidean space, affine transforms such as translations, rotations, and scaling are important in many applications. A very common type of invariant kernels is the kernels invariant to all translations. These kernels are known as a shift invariant or stationary kernel and have the form

$$K(x, y) = k(x - y)$$

Because of the constraints of kernels, it is not always possible to find reasonable invariant kernels, given a symmetry group. For example, it is known that the only kernel invariant to translation, rotation, and scaling is a constant function. In [7], kernels incorporating local invariances were discussed. In [8], we characterized the periodic kernels that are defined on a direct product of circles and invariant to rotations in individual 1-D variables.

In this paper, we extend the results in [8] and study the rotation invariant kernels on spheres. Rotations in n-D spaces are more complex than the 2-D rotations. For example, the group of 3-D rotations, the special orthogonal group $SO(3)$, is non-abelian. We will show that a rotation invariant kernel must be a function of the dot product (inner product) of the two vectors. Consequently, the kernel is reduced to a positive definite function on the circle group. Using Fourier analysis, we show that such a kernel can be expressed as a Chebyshev series with non-negative coefficients. This condition on the series expansion is necessary but not sufficient for $n \geq 3$. A necessary and sufficient condition can be obtained using Schoenberg's theorem.

This paper is organized as follows. Section 2 introduces the rotation groups and their actions on spheres. Section 3 provides a characterization of rotation invariant kernels using Fourier analysis. Numerical results using rotation invariant kernels are presented in Section 4. Finally, Section 5 provides conclusions and proposals for future studies.

2. Rotation Groups

The special orthogonal group $SO(n)$ is a subgroup of $n \times n$ nonsingular matrices:

$$SO(n) = \{R \mid RR^T = I, \det(R) = 1\}$$

Acting on the Euclidean space $\mathbb{R}^n$, the group represents the rotations about the origin in the space. The group of 2-D rotations $SO(2)$ is isomorphic to the circle group, which is a compact abelian group. When $n \geq 3$, the group is non-abelian. The group induces an action on spheres centered at the origin.

The transitivity of a group acting on a space is important to the structure of the invariant functions. A group G acting on X is transitive if for any $x, y \in X$, there exists $g \in G$ such that $g(x) = y$. A group action is doubly transitive if for any two pairs of points $(x, y), (z, w) \in X \times X$, there exists $g \in G$ such that $g(x) = y$ and $g(y) = w$.

$SO(n)$ is transitive on the unit sphere S. To prove the transitivity, it suffices to show that the north pole (1,0,...,0) can be mapped to an arbitrary point on the sphere $x = (x_1, x_2, \dots, x_n)$ by a rotation. Using the Gram-Schmidt process, x can be extended to an orthonormal basis $\langle x, v_2, \dots, v_n \rangle$. Write the vectors as columns of a matrix U and we have

$$U\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} = \begin{bmatrix}x_1 & v_{21} & \cdots & v_{n1}\\ x_2 & v_{22} & \cdots & v_{n2}\\ \vdots & \vdots & \ddots & \vdots\\ x_n & v_{2n} & \cdots & v_{nn}\end{bmatrix}\begin{bmatrix}1\\0\\\vdots\\0\end{bmatrix} = \begin{bmatrix}x_1\\x_2\\\vdots\\x_n\end{bmatrix}$$

The matrix U is orthogonal. If $\det(U) = -1$, multiplying one of the columns by -1 would make $U \in SO(n)$.

Because a rotation is an orthogonal transform which preserves distances, it will not be doubly transitive on S. Nevertheless, we will show that $SO(n)$ is transitive on pairs of points that are equal-distant.

Theorem 1 Let $(x, y), (z, w)$ be two pairs of points on the sphere S. There exists $g \in SO(n)$ such that $g(x) = y$ and $g(y) = w$, if and only if $d(x, y) = d(z, w)$ (in either geodesic or Euclidean distances).

Proof. Since the rotations in $SO(n)$ are orthogonal transforms, they preserve distances. The pairs of points must have the same distances if there is a rotation mapping one pair to the other.

Conversely, if $d(x, y) = d(z, w)$, we need to show that there is a rotation g that maps (x, y) to (z, w). Without loss of generality, we may assume that $x = z$, since $SO(n)$ is transitive on S. Consider all the points on S that has a fixed distance r from the point x.

$$S' = \{y \in S \mid d(x, y) = r\}$$

The set S' is a sphere of dimension $n - 1$. The rotations of $SO(n)$ that fix the point x is $SO(n-1)$ acting on S'. Since the action of the group $SO(n-1)$ on S' is transitive, we have a rotation g that leaves x fixed and $g(y) = w$.

3. Invariant Kernals on Spheres

Let K be a kernel on S that is invariant to rotations in $SO(n)$. By Theorem 1, the kernel must be a function of the distance of the two points.

$$K(x, y) = k(d(x, y))$$

Since the points on the unit sphere, the distance is determined by the dot product between the two vectors. $d(x, y) = \sqrt{2 - 2\cos(\theta)} = \sqrt{2 - 2(x \cdot y)}$, where θ is the angle between the two vectors. Consequently, the kernel is a function of the dot product, or the cos of the angle between the vectors.

$$K(x, y) = k(x \cdot y) = k(\cos\theta)$$

This reduces the kernel to a function on the circle group. To characterize such a kernel, we will apply Fourier transform ([9]) and Bochner's theorem ([10], [11]).

Chebyshev Polynomial. A Chebyshev polynomial of degree n is defined as

$$T_n(x) = cos(n \arccos(x))$$

Chebyshev polynomials are orthogonal over the interval $[-1,1]$ with respect to the weight function $1/\sqrt{1 - x^2}$.

Fourier Transform. Let G be a locally compact abelian group. A character $\chi(x)$ is a continuous group homomorphism from G to the circle group $\mathbb{T}$. The dual group is the set of all characters on G, denoted by $\hat{G}$, along with the multiplication operation. Let f be a function on G. The Fourier transform of f is defined as a function on $\hat{G}$:

$$\hat{f}(\chi) = \int_G f(x)\overline{\chi(x)}d\nu(x)$$

where ν is the Haar measure on G.

Bochner's Theorem. For a locally compact abelian group G with dual group $\hat{G}$, and a positive definite function f

$$f(x) = \int_{\hat{G}} \chi(x) d\mu(\chi)$$

where μ is a positive finite measure.

Theorem 2. A rotation invariant kernel $K(x, y)$ defined on S has the following form.

$$K(x,y) = k(x \cdot y) = \sum_{n \geq 0} a_n T_n(x \cdot y) \quad a_n \geq 0$$

where T_n is the Chebyshev polynomial of degree n.

Proof. Because of the rotation invariance, the kernel is a function of the dot product

$$K(x,y) = k(x \cdot y) = k(\cos\theta)$$

We show that $k(\cos\theta)$ as a function of θ is a function of positive type on the circle group $\mathbb{T}$. Let $\theta_1, \theta_2, \ldots, \theta_n \in \mathbb{T}$ and $x_i = (\cos\theta_i, \sin\theta_i, 0, 0, \ldots) \in S$. Then

$$x_i \cdot x_j = \cos\theta_i \cos\theta_j + \sin\theta_i \sin\theta_j = cos(\theta_i - \theta_j)$$

Since K is positive definite,

$$\sum_{i,j=1}^{n} k\left(cos(\theta_i - \theta_j)\right) c_i c_j = \sum_{i,j=1}^{n} K(x_i, x_j) c_i c_j \geq 0$$

The dual group of the circle group $\mathbb{T}$ is the group of integers $\mathbb{Z}$. By using the Fourier inversion formula, the function $k(\cos\theta)$ can be expressed as a series over the dual group. Since k is real-valued, the Fourier series can be written as a series of sin and cos functions.

$$k(\cos\theta) = \sum_{n=0}^{\infty} a_n \cos(n\theta) + b_n \sin(n\theta)$$

Since $k(\cos\theta)$ is an even function, the Fourier series contains only cos terms.

$$k(\cos\theta) = \sum_{n=0}^{\infty} a_n \cos n\theta$$

Now we only need to show that $a_n \geq 0$. Since $k(\cos\theta)$ is positive definite, by Bochner's theorem,

$$k(\cos\theta) = \int_{\hat{G}} \chi(\theta) d\mu(\chi)$$

In this case, the dual group $\hat{G}$ is the integer group $\mathbb{Z}$. The positive measure μ corresponds to the Fourier coefficients a_n. Therefore, $a_n \geq 0$.

On the sphere S, $\cos\theta = x \cdot y$. We have

$$K(x,y) = k(x \cdot y) = k(\cos\theta) = \sum_{n=0}^{\infty} a_n \cos n\theta = \sum_{n=0}^{\infty} a_n \cos(n \arccos(x \cdot y))$$

Therefore, by definition of Chebyshev polynomials,

$$K(x,y) = \sum_{n=0}^{\infty} a_n T_n(x \cdot y) \quad a_n \geq 0$$

The Chebyshev series expansion of the dot products with non-negative coefficients is a necessary condition for the invariant kernels. For 2-dimensional space, the condition is also sufficient. This is due to the fact that the 2-D sphere can be identified as the circle group and the positive definite kernel corresponds to the positive definite function directly.

For $n \geq 3$, the condition in Theorem 2 may not be sufficient. The spheres do not have an abelian group structure as in the 2-D case. We may apply Schoenberg's theorem to obtain a necessary and sufficient condition for invariant kernels on n-D spheres.

Legendre Polynomial. A Legendre polynomial $P_n(x)$ of degree n can be defined by the generating function.

$$\frac{1}{\sqrt{1-2xt+t^2}} = \sum_{n=0}^{\infty} P_n(x)t^n$$

Legendre polynomials are orthogonal over the interval $[-1,1]$ with respect to the weight function 1.

Gegenbauer Polynomial. A Gegenbauer polynomial of order α and degree n is defined by the generating function.

$$\frac{1}{(1-2xt+t^2)^{\alpha}} = \sum_{n=0}^{\infty} P_n^{\alpha}(x)t^n$$

Gegenbauer polynomials are orthogonal over the interval $[-1,1]$ with respect to the weight function $(1-x^2)^{\alpha-1/2}$.

Note that Chebyshev polynomials and Legendre polynomials are special cases of Gegenbauer polynomials with orders 0 and 1/2 respectively.

Schoenberg's Theorem. ([12]) A positive definite function $k(x \cdot y)$ on the unit sphere is invariant under $SO(d)$ if and only if it has the following form.

$$k(x \cdot y) = \sum_{n=0}^{\infty} a_n P_n^{\frac{d}{2}-1}(x \cdot y) \quad a_n \geq 0$$

where $P_n^{\frac{d}{2}-1}(x)$ is the Gegenbauer polynomial of order $d/2-1$ and degree n.

When $d=2$ the order of the Gegenbauer polynomials is 0 and the series is a Chebyshev polynomial expansion. When $d=3$ the order of the Gegenbauer polynomials is 1/2 and the series is a Legendre polynomial expansion.

4. Numerical Results and Discussions

A 3-dimensional example is used to illustrate the rotation invariance of the kernels on a sphere as defined in Section III. A random sample of 8 points on the sphere and its rotated version are shown in Fig. 1. The red dots are the random sample and the blue dots are the rotated version.

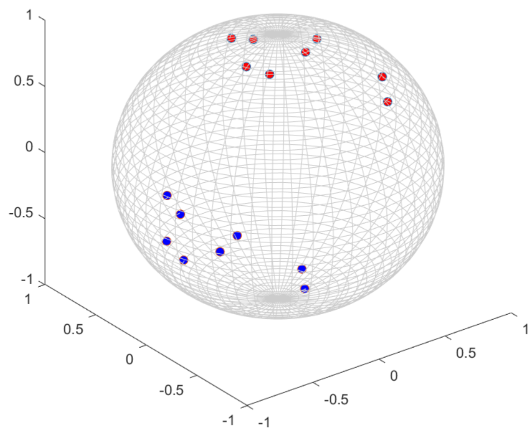

Fig. 1. Rotated data points on a sphere.

In the 3-D space, the rotation invariant kernels are represented as a series of Legendre polynomials. The following kernel is constructed with a Legendre polynomial on the dot product. It is applied to the two data sets. $K(x,y) = P_3(x \cdot y) = 2.5(x \cdot y)^3 - 1.5(x \cdot y)$

To show the rotation invariance, the Gram matrices on the two data sets are calculated. The two Gram matrices are identical:

$$\begin{bmatrix} 4.0000 & 2.0679 & 3.4466 & 3.1025 & 3.0814 & 3.8844 & 3.6012 & 1.9971 \\ 2.0679 & 4.0000 & 1.7508 & 1.2609 & 3.1086 & 1.6103 & 2.9660 & 3.8239 \\ 3.4466 & 1.7508 & 4.0000 & 3.9006 & 3.3748 & 3.6801 & 3.4990 & 1.3846 \\ 3.1025 & 1.2609 & 3.9006 & 4.0000 & 2.9331 & 3.5043 & 3.0304 & 0.9329 \\ 3.0814 & 3.1086 & 3.3748 & 2.9331 & 4.0000 & 2.9158 & 3.8540 & 2.5867 \\ 3.8844 & 1.6103 & 3.6801 & 3.5043 & 2.9158 & 4.0000 & 3.3856 & 1.4676 \\ 3.6012 & 2.9660 & 3.4990 & 3.0304 & 3.8540 & 3.3856 & 4.0000 & 2.6413 \\ 1.9971 & 3.8239 & 1.3846 & 0.9329 & 2.5867 & 1.4676 & 2.6413 & 4.0000 \end{bmatrix}$$

Kernel based machine learning methods such as support vector machine (SVM) and Gaussian process (GP) have the kernel (and specifically, the Gram matrix) as the oracle. The rotation invariance of the kernel will be propagated to the learning system.

By incorporating the natural symmetries of the sphere, the invariant kernel is benefited in measuring the similarity based on simpler and more essential structure information. The reduced capacity in the learning system will lead to less overfitting and better performance in generalization.

5. Conclusions and Future Work

In this paper, we studied the kernels on spheres that are invariant under the actions of the rotation group. By studying the transitivity of the rotation groups on spheres, we concluded that such an invariant kernel must be a function of the dot product of the two vectors. By using the Fourier transform on locally compact abelian groups and the Bochner's theorem, we further proved that such a kernel has a Chebyshev polynomial expansion with non-negative coefficients regardless of the dimension.

The Chebyshev series expansion with non-negative coefficients of the positive definite function of dot products is a necessary condition for the invariant kernels. For 2-dimensional space, the condition is also sufficient. For $n \geq 3$, the condition may not be sufficient. Schoenberg's theorem can be applied to derive a necessary and sufficient condition in terms of Gegenbauer polynomial expansions.

For future work, we will consider invariant kernels related to other spaces and symmetry groups, by exploring modern harmonic analysis methods in relation to kernel structures. We will also study the efficacy of the invariant kernels on practical machine learning applications.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Both authors conducted the research. Hong Zhang wrote the paper. Both authors approved the final version.

References

[1] Bishop, C. (2006). *Pattern Recognition and Machine Learning*. New York: Springer.

[2] Boser, B. E., Guyon, I. M., & Vapnik, V. N. (1992). A training algorithm for optimal margin classifiers. *Proceedings of the Fifth Annual Workshop on Computational Learning Theory - COLT '92* (p. 144).

[3] MacKay, D. J. C. (1998). Introduction to gaussian processes. In Bishop, C. M. (Ed.), *Neural Networks and Machine Learning*. Springer-Verlag.

[4] Rasmussen, C. E., & Christopher, K. I. W. (2006). *Gaussian Processes for Machine Learning*. The MIT Press.
[5] Su, H., & Zhang, H. (2014). Distances and kernels based on cumulative distribution functions. *Emerging Trends in Image Processing, Computer Vision and Pattern Recognition* (Chapter 36). Morgan Kaufmann.
[6] Su, H., & Zhang, H. (2015). Efficient computation of distribution kernels and distances. *Proceedings of the 2015 International Conference on Image Processing, Computer Vision, & Pattern Recognition, IPCV'15*.
[7] Burges, C. (1999). Geometry and invariance in kernel based methods*. Advances in Kernel Methods: Support Vector Learning* (pp. 89–116). Cambridge, MA, USA: MIT Press.
[8] Su, H., & Zhang, H. (2019). On stationary periodic kernels. *The 2019 International Conference on Image Processing, Computer Vision, & Pattern Recognition* (pp. 43-46).
[9] Rudin, W. (1962). *Fourier Analysis on Groups*. Wiley-Interscience.
[10] Folland, G. B. (1995). *A Course in Abstract Harmonic Analysis*. CRC.
[11] Gihman, I. I., & Skorohod, A. V. (1974). *The Theory of Stochastic Processes, Vol. 1*. Springer-Verlag.
[12] Schoenberg, I. J. (1942). Positive definite functions on spheres. *Duke Mathematical Journal*, *9(1)*, 96-108.

Hongjun Su was born in Chengdu, China. He received his PhD in electrical and computer engineering (2001) from University of Dayton, USA, and his M.Sc. in electro-optics (1994) and B.Sc in physics/optics (1991) from Sichuan University, Chengdu, China. He is an associate professor of electrical and computer engineering at Georgia Southern University, Savannah, Georgia, USA. His research interests include IoT and its applications, high-speed network protocols and architecture design, QoS of networks, wireless and sensor networks, optical fiber communication and networks, pattern recognition, image processing, 3-D sensing and measurement, machine vision, big data processing, smart machine learning algorithms.

Hong Zhang was born in Shanghai, China. He received his PhD (1989) and MA in mathematics (1985), MSEE in electrical engineering (1984) from University of Pittsburgh, USA, and his BS in computer science (1982) from Fudan University, China. He is a Professor of Computer Science at Georgia Southern University, Savannah, Georgia, USA. His current research interests include image processing, machine learning and biomedical applications.