Technical Report

Department of Computer Science
and Engineering
University of Minnesota
4-192 EECS Building
200 Union Street SE
Minneapolis, MN 55455-0159 USA

TR 08-034

Conditionally Positive Definite Kernels and Infinitely Divisible Distributions

Arindam Banerjee and Nisheeth Srivastava

October 28, 2008

Conditionally Positive Definite Kernels and Infinitely Divisible Distributions

Arindam Banerjee
Dept of Computer Science & Engineering
University of Minnesota, Twin Cities
banerjee@cs.umn.edu

Nisheeth Srivastava
Dept of Computer Science & Engineering
University of Minnesota, Twin Cities
nsriva@cs.umn.edu

Abstract

We give a precise characterization of two important classes of conditionally positive definite (CPD) kernels in terms of integral transforms of infinitely divisible distributions. In particular, we show that for any stationary CPD kernel $A(x, y) = f(x - y)$, f is the log-characteristic function of a uniquely determined infinitely divisible distribution; further, for any additive CPD kernel $A(x, y) = g(x + y)$, g is the log-moment generating function of a uniquely determined infinitely divisible distribution. The results strengthen the connections between CPD kernels and infinitely divisible distributions.

1 Introduction

Conditionally positive definite (CPD) kernels have gained prominence in recent years in the machine learning and matrix analysis communities as a consequence of their utility in various application contexts under the broad ambit of kernel methods [9, 13]. The class of CPD kernels is known to subsume the better known and much more widely studied class of positive definite kernels [13]. Methods for designing positive definite kernels are well known and have been the focus of intense research in the machine learning community in recent years [13, 9]. A key component of several constructive methods associated with positive definite kernels is the unique probabilistic characterization afforded by Bochner's theorem [4, 2, 5]. An equivalent characterization of CPD kernels is non-existent, although several important related results exist in the literature [8, 10].

In this paper, we present a probabilistic characterization of two important classes of CPD kernels, viz stationary kernels $K(x, y) = f(x - y)$ and additive kernels $K(x, y) = g(x + y)$. We show that such kernels are uniquely determined by suitable integral transforms of infinitely divisible probability distributions [12]. In particular, stationary CPD kernels are the log-characteristic functions and additive CPD kernels are log-moment generating functions of infinitely divisible distributions. Further, we show that the characterization is exhaustive. Finally, using some well known results from the theory of infinitely divisible distributions, we give a closed form for stationary and additive CPD kernels, which can be used to design CPD kernels for real world applications.

2 Background

We first review some results in from the existing literature on harmonic analysis, conditionally positive definite functions, and infinitely divisible distributions. A complex valued function $C : \mathbb{R}^n \times \mathbb{R}^n \mapsto \mathbb{C}$ is called a *positive definite kernel* if and only if

$$\int C(x, y)\ u(x)\ \bar{u}(y)\ dxdy \geq 0\ ,$$

where $\int \|u(x)\|^2 \, dx < \infty$. For any set of finite points $x_1, \ldots, x_n$, the matrix $[C(x_i, x_j)]$ will be a *positive definite matrix*, i.e., $u^T C \bar{u} \geq 0$ for any $u \in \mathbb{C}^n$.[1]

A positive definite kernel is called *stationary* if

$$C(x, y) = F(x - y), \qquad \forall x, y \in \mathbb{R}^n \ ,$$

where $F : \mathbb{R}^n \mapsto \mathbb{C}$ is known as a *positive definite function*. By Bochner's theorem [4, 2, 5], a continuous function F is positive semi-definite iff it is the Fourier transform of a non-negative finite measure, i.e.,

$$F(t) = \int_r e^{i\langle r,t\rangle} \, dP(r) \ , \qquad t \in \mathbb{R}^n \ , \tag{1}$$

where P is a non-negative finite measure on $\mathbb{R}^n$.

A positive definite kernel is called *additive* if

$$C(x, y) = G(x + y), \qquad x, y \in \mathbb{R}^n \ ,$$

where $G : \mathbb{R}^n \mapsto \mathbb{C}$ is known as an *exponentially convex function* [1, 7]. By Devinatz's theorem [6, 2], a continuous function G is exponentially convex iff it is the Laplace transform of a non-negative finite measure, i.e,

$$G(s) = \int_r e^{\langle r,s\rangle} \, dP(r) \ , \qquad s \in \mathbb{R}^n, \tag{2}$$

where P is a non-negative finite measure on $\mathbb{R}^n$.

Conditionally positive definite (CPD) matrices: A matrix A is called *conditionally positive definite* (CPD) if $u^T A \bar{u} \geq 0$ for any $u \in \mathbb{C}^n$ such that $\sum_{j=1}^n c_j = 0$. The following result demonstrates an important connection between Hermitian CPD matrices and positive definite matrices:

Theorem 1 ([13, 8]) *A matrix $[K_{i,j}]$ is conditionally positive definite and Hermitian if and only if for $A_{i,j} = \exp(K_{i,j})$, the matrix $[A_{i,j}^\beta]$ is positive definite for any $\beta > 0$.*

In Theorem 1, the matrix $[A_{i,j}^\beta]$ is a element-wise exponentiation of each element $A_{i,j}$, as opposed to the more widely studied matrix exponentiation. If the matrices $[A_{i,j}^\beta]$ are positive definite for all $\beta > 0$, then the matrix $[K_{i,j}]$ is known as infinitely divisible [3, 9].

Infinitely divisible distributions: For a probability measure μ, let μ^n denote the n-fold convolution of the probability measure with itself, i.e., $\mu^n = \mu * \cdots * \mu$ (n times). A probability measure μ on $\mathbb{R}^d$ is infinitely divisible if, for any positive integer n, there is a probability measure μ_n on $\mathbb{R}^d$ such that $\mu = \mu_n^n$, i.e., μ is the n-fold convolution of some other measure μ_n, for all $n \in \mathbb{N}$. While the definition of infinitely divisible distributions is based on the n-fold convolution μ^n, the β-fold convolution μ^β is well defined and infinitely divisible for any $\beta \geq 0$ [12, Lemma 7.9]. For the purposes of our analysis, we need the following result concerning the characteristic function of infinitely divisible distributions.[2] The result follows directly from Theorem 8.1 and Corollary 8.3 of [12].

Theorem 2 ([12]) *Let μ be an infinitely divisible distribution with characteristic function $F(z)$. Then, the characteristic function of μ^β is $F(z)^\beta$. Conversely, the characteristic function $F(z)^\beta$ uniquely corresponds to the distribution μ^β.*

---

[1]For the purposes of this article, we do not differentiate between positive definiteness and positive semi-definiteness. In the CPD literature, it is customary to use 'positive definiteness' for both cases. Our results hold for either case.

[2]While the characteristic function (Fourier transform) of the convolution of any probability measure with itself (if well defined) is known to be the product of the original characteristic function, the β-fold convolution need not exist for $\beta \in (0, 1)$ or, more generally, any non-integer β. The β-fold convolution and its characteristic function (Fourier transform) are well-defined for any infinitely divisible distribution.

3 Constructing Conditionally Positive Definite Kernels

In this section, we present constructions for two important classes of CPD kernels on $\mathbb{R}^d$, viz stationary and additive CPD kernels. While the former, briefly discussed in Section 3.1, is obtained entirely using prior results described in Section 2 and is closely connected to known results in the literature [8], the latter requires a novel result characterizing exponentially convex functions as Laplace transforms of infinitely divisible distributions, and is presented in Section 3.2.

3.1 Stationary CPD Kernels

Here, we focus on stationary CPD matrices, i.e., CPD matrices of the form $K(x, y) = f(x - y)$.

Theorem 3 *K is a stationary CPD matrix, i.e., $K(x, y) = f(x - y)$, if and only if*

$$K(x, y) = \log \int_r \exp(i\langle x - y, r\rangle) d\mu(r) \ , \tag{3}$$

for a uniquely determined infinitely divisible distribution μ.

Proof: For the 'if' part, for an infinitely divisible distribution μ, if $F(t) = \int_r \exp(i\langle t, r\rangle) d\mu(r)$ is the characteristic function of μ, then $K(x, y) = \log F(x - y)$, where the complex logarithm is well defined since $\mathrm{Re}(F) \neq 0$.

Now, since μ is infinitely divisible, the characteristic function of $\mu^\beta, \beta > 0$, is simply $F(t)^\beta$, which is positive definite according to Bochner's theorem [4]. Hence, the matrix $[\exp(\beta K(x, y))] = [F(x - y)^\beta]$ is positive definite for all $\beta > 0$, implying $K(x, y)$ is a CPD matrix from Theorem 1.

For the 'only if' part, let $K(x, y) = f(x - y)$ be a stationary CPD matrix. Let $F(t) = \exp(f(t))$. Since for any $\beta > 0$, from Theorem 1, $[\exp(\beta K(x, y))] = [F(x - y)^\beta]$ is positive definite, from Bochner's theorem [4] it follows for each $\beta > 0$ there exists an unique probability measure μ_β such that

$$F(t)^\beta = \int_r \exp(i\langle t, r\rangle) d\mu_\beta(r) \ .$$

Since the above holds for all β, from Theorem 2 μ_β must be an infinitely divisible distribution. In particular, for $\beta = 1$,

$$K(x, y) = \log F(x - y) = \log \int_r \exp(i\langle x - y, r\rangle) d\mu(r) \ .$$

That completes the proof. ■

Theorem 3 gives a method to construct stationary CPD kernels: (i) Choose an infinitely divisible distribution μ, (ii) Compute its characteristic function, $F(t) = \int_r \exp(i\langle t, r\rangle) d\mu(r)$, and (iii) Construct kernel $K(x, y) = \log F(x - y)$. Then, K is a CPD kernel by construction.

Example 1 Consider the infinitely divisible distribution $\exp(-\|r\|^2/2)$. The characteristic function can be computed as

$$F(t) = \int_r \exp(i\langle t, r\rangle) \exp(-\|r\|^2/2) dr = \exp(-\|t\|^2/2) \ ,$$

so that $K(x, y) = \log F(x - y) = -\|x - y\|^2/2$ is a CPD kernel since $[\exp(\beta K(x, y))] = [\exp(-\beta\|x - y\|^2/2)]$ is always positive definite for $\beta \geq 0$. ■

It is well-known that the characteristic function of infinitely divisible distributions on $\mathbb{R}^n$ can be expressed in closed form as the *Levy-Khintchine formula* [12], where the representation is found to be uniquely determined by a *generator triplet* (A, ν, γ) where,

- A is a positive definite $n \times n$ matrix
- ν is the Levy measure on $\mathbb{R}^d$ satisfying $\nu(\{0\}) = 0$ and $\int_r \min(\|r\|^2, 1)\ d\nu(r) < \infty$,
- $\gamma \in \mathbb{R}^n$.

For a specific generator triplet (A, ν, γ) and with $D = \{r : \|r\| \leq 1\}$, it is possible to write the resultant characteristic function $F(t)$ in the form

$$F(t) = \exp\left[-\frac{1}{2}\ \langle t, At\rangle + i\langle \gamma, t\rangle + \int_r (e^{i\langle t,r\rangle} - 1 - \imath\langle t, r\rangle \mathbf{1}_D(r))d\nu(r)\right], \quad t \in \mathbb{R}^d\ . \tag{4}$$

Thus, from the construction detailed above, we can determine an explicit form for stationary CPD matrices as shown below. The result follows immediately by replacing the characteristic function computation with the Levy-Khintchine formula above.

Corollary 1 *A valid generator triplet (A, ν, γ) corresponds to a stationary CPD kernel $K(x, y)$ of the form*

$$K(x, y) = -\frac{1}{2}(x - y)^T A(x - y) + i\langle \gamma, x - y\rangle + \int_r (e^{i\langle x-y,r\rangle} - 1 - \imath\langle x - y, r\rangle \mathbf{1}_D(r))d\nu(r)\ . \tag{5}$$

Further, any real valued stationary CPD kernel is of the form

$$K(x, y) = Re(K(x, y)) = -\frac{1}{2}(x - y)^T A(x - y) - \int_r (1 - \cos(\langle x - y, r\rangle))d\nu(r)\ . \tag{6}$$

We note that Corollary 1 implies that kernels of the form $K(x, y) = -(x - y)^T A(x - y)$ will be CPD, and Example 1 can be seen as a special case of this result.

3.2 Additive CPD Kernels

While the characterization of stationary CPD kernels follow straightforwardly from existing results, an equivalent characterization of additive CPD matrices requires a new result which we state below.

Theorem 4 *Let G be an exponentially convex function, implying G can be uniquely expressed as the Laplace transform of a non-negative finite measure μ. Then, for any $\beta > 0$, G^β will be an exponentially convex function if and only if μ is infinitely divisible.*

Proof: For the 'if' part, we know that μ is infinitely divisible, and, by Devinatz's theorem [6], $G(s) = \int_r \exp(\langle r, s\rangle)d\mu(r)$ is the moment generating function (or Laplace transform) of μ well defined over $\mathcal{S} = \mathrm{dom}(G)$. Let $F(t) = \int_r \exp(i\langle r, t\rangle)d\mu(r)$ be the characteristic function (or Fourier transform) of μ. Now, the characteristic function of the base measure $dm(r) = \exp(\langle r, s\rangle)d\mu(r)$ is given by

$$H(t) = \int_r \exp(i\langle t, r\rangle)dm(r) = \int_r \exp(i\langle r, t - is\rangle)d\mu(r)\ ,$$

where the integral is convergent and analytic as a function of $(t - is), t \in \mathbb{R}^n, s \in \mathcal{S}$ [11, Theorem 2.7.1]. In other words, $F(t)$ has an analytic extension to $\mathbb{R}^n - i\mathcal{S} \subseteq \mathbb{C}^n$. As a result, following [7], $G(s) = F(-is)$, i.e., the Laplace transform can be obtained from the Fourier transform by a simple plug-in. Since μ is infinitely divisible, the characteristic function of μ^β, the β-fold convolution of μ where $\beta \geq 0$, is simply $F_\beta(t) = F(t)^\beta$, from Theorem 2. Let $G_\beta(s) = \int_r \exp(\langle r, s\rangle)d\mu^\beta(r)$ be the corresponding Laplace transform. Now, the characteristic function of the base measure $dm^\beta(r) = \exp(\langle r, s\rangle)\ d\mu^\beta(r)$ is given by

$$H_\beta(t) = \int_r \exp(i\langle t, r\rangle)dm^\beta(r) = \int_r \exp(i\langle r, t - is\rangle)\ d\mu^\beta(r)\ ,$$

where, as before, the integral is convergent and analytic as a function of $(t - is), t \in \mathbb{R}^n, s \in \mathcal{S}$ [11, Theorem 2.7.1]. In other words, $F_\beta(t)$ has an analytic extension to $\mathbb{R}^n - i\mathcal{S} \subseteq \mathbb{C}^d$. As a result, following Theorem 2,

$$G_\beta(s) = F_\beta(-is) = F(-is)^\beta = G(s)^\beta \ .$$

Since $G_\beta(s)$ is the Laplace transform of a probability measure, it is exponentially convex following [6]. Hence, $G(s)^\beta$ is exponentially convex for $\beta \geq 0$.

For the 'only if' part, from Devinatz's theorem [6] it follows that $\forall \beta > 0$, there exists a non-negative measure μ_β such that

$$G(s)^\beta = \int_x \exp(\langle x, s\rangle) \ d\mu_\beta(x) \ .$$

Then, from [7], it follows that the characteristic function $F_\beta(t)$ of μ_β can be obtained using a simple plug-in $F_\beta(t) = G(it)^\beta$. Choosing $\beta = 1$ and $\beta = n$, we note that $F_n(t) = G(it)^n = F(t)^n$ so that the characteristic function of μ_n is the n-fold product of that of μ. Since this holds for all n, μ is infinitely divisible by definition. ∎

As shown by the following result, additive CPD matrices, i.e. CPD matrices of the form $K(x, y) = f(x+y)$, can be exhaustively characterized as the log moment generating function (or cumulant function) of infinitely divisible distributions.

Theorem 5 *K is an additive CPD matrix, i.e., $K(x, y) = g(x + y)$, if and only if*

$$K(x, y) = \log \int_r \exp(\langle x + y, r\rangle) d\mu(r) \ , \tag{7}$$

for a uniquely determined infinitely divisible distribution μ.

Proof: For the 'if' part, for an infinitely divisible distribution μ, if $G(s) = \int_r \exp(\langle s, r\rangle) d\mu(r)$ is the moment generating function, then $K(x, y) = \log G(x + y)$, where the logarithm is well-defined since $G(s) > 0$. Now, since μ is infinitely divisible, from Theorem 4, $G(s)^\beta$ is exponentially convex, implying $[G(x+y)^\beta]$ is positive definite. As a result, the matrix $[\exp(\beta K(x, y))]$ is positive definite, implying K is conditionally positive definite by Theorem 1.

For the 'only if' part, let $K(x, y) = g(x+y)$ be a conditionally positive definite kernel. Let $G(s) = \exp(g(s))$, so that the matrix $[\exp(\beta K(x, y))] = G(x+y)^\beta$. Since for any $\beta > 0$, $[\exp(\beta K(x, y))] = G(x+y)^\beta$ is positive definite, we note that $G(s)^\beta$ is an exponentially convex function for all $\beta > 0$. Then, from Theorem 4 it follows that

$$G(s)^\beta = \int_r \exp(\langle r, s\rangle) d\mu^\beta(r) \ ,$$

for some uniquely determined infinitely divisible distribution μ^β. Then, for $\beta = 1$, it follows that μ is an infinitely divisible distribution. Finally, from the fact that $\exp(K(x, y)) = G(x + y)$, we have

$$K(x, y) = \log \int_r \exp(\langle r, x + y\rangle) d\mu(r) \ ,$$

where μ is an infinitely divisible distribution. That completes the proof. ∎

Theorem 5 gives a method to construct additive CPD kernels: (i) Choose an infinitely divisible distribution μ, (ii) Compute its moment generating function, $G(s) = \int_r \exp(\langle s, r\rangle) d\mu(r)$, and (iii) Construct kernel $K(x, y) = \log G(x + y)$. Then, K will provably be an additive CPD kernel.

Example 2 Consider the infinitely divisible distribution $\exp(-\|r\|^2/2)$. The Laplace transform can be computed as

$$G(s) = \int_r \exp(\langle s, r\rangle) \exp(-\|r\|^2/2) dr = \exp(\|s\|^2/2) \ ,$$

so that $K(x, y) = \log G(x + y) = \|x + y\|^2/2$ is a CPD matrix. ∎

As shown in the proof of Theorem 4, if μ is infinitely divisible, the Laplace transform $G(s) = \int_r \exp(\langle s, r\rangle)d\mu(r)$ can be obtained from the Fourier transform $F(t) = \int_r \exp(i\langle t, r\rangle)d\mu(r)$ by a simple plug-in procedure, i.e., $G(s) = F(-is)$. As a result, we can obtain the moment generating function of infinitely divisible distributions by applying the plug-in procedure to the Levy-Khintchine formula, so that

$$G(s) = \exp\left[\frac{1}{2}\langle s, As\rangle + \langle \gamma, s\rangle + \int_r (e^{\langle s,r\rangle} - 1 - \langle s, r\rangle \mathbf{1}_D(r))d\nu(r)\right], \quad s \in \mathbb{R}^n \ . \tag{8}$$

Corollary 2 *A valid generator triplet (A, ν, γ) corresponds to an additive CPD matrix $K(x, y)$ of the form*

$$K(x, y) = (x + y)^T A(x + y) + \langle \gamma, x + y\rangle + \int_r (e^{\langle x+y,r\rangle} - 1 - \langle x + y, r\rangle \mathbf{1}_D(r))d\nu(r) \ . \tag{9}$$

We note that Corollary 2 implies that kernels of the form $K(x, y) = (x + y)^T A(x + y)$ will be CPD, and Example 2 can be seen as a special case of this result.

4 Conclusions

We have presented an exhaustive probabilistic characterization of stationary and additive CPD kernels. In particular, our results give a way to construct CPD kernels from infinitely divisible distributions. The relationship between infinitely divisible distributions and Levy processes [12] also imply close connections between CPD kernels and Levy processes. Further, through a direct application of the closed form for the characteristic function of infinitely divisible distributions [12] as well as some recent advances in harmonic analysis [7], we have given explicit forms of stationary and additive CPD kernels. Broadly, our results are analogous to the widely studied Bochner's theorem [4, 5] for positive definite kernels. As Bochner's theorem has found wide applications in the context of approximation theory [5] and machine learning [13], we expect our characterization to find important applications in such domains, particularly in the context of infinitely divisible kernels [3, 9].

References

[1] A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh. Clustering with Bregman divergences. *Journal of Machine Learning Research*, 6:1705–1749, 2005.

[2] C. Berg, J. Christensen, and P. Ressel. *Harmonic Analysis on Semigroups: Theory of Positive Definite and Related Functions*. Springer-Verlag, 1984.

[3] R. Bhatia. Infinitely divisible matrices. *American Mathematical Monthly*, 2005.

[4] S. Bochner. Monotone funktionen, stieltjes integrale und harmonische analyse. *Math. Ann.*, 108:378–410, 1933.

[5] E. W. Cheney and W. A. Light. *A course in approximation theory*. Brooks Cole, 1999.

[6] A. Devinatz. The representation of functions as Laplace-Stieltjes integrals. *Duke Mathematical Journal*, 24:481–498, 1955.

[7] W. Ehm, M. G. Genton, and T. Gneiting. Stationary covariances associated with exponentially convex functions. *Bernoulli*, 9(4):607–615, 2003.

[8] A. Guichardet. *Symmetric Hilbert Spaces and Related Topics*. Springer, 1972.

[9] D. Haussler. Convolution kernels on discrete structures. Technical report, UC Santa Cruz, 2000.

[10] S. I. Karpushev. Conditionally positive-definite functions on locally compact groups and the Levy-Khinchin formula. *Journal of Mathematical Sciences*, 28:489–498, 1985.

[11] E. Lehmann and J. Romano. *Testing Statistical Hypothesis*. Springer, 2005.

[12] K. Sato. *Levy Processes and Infinitely Divisible Distributions*. Cambridge University Press, 1999.

[13] B. Schölkopf and A. Smola. *Learning with Kernels*. MIT Press, 2001.