The Mercer-Young Theorem for Matrix-Valued Kernels on Separable Metric Spaces

Eyal Neuman and Sturmius Tuschmann*

Department of Mathematics, Imperial College London

March 28, 2024

Abstract

We generalize the characterization theorem going back to Mercer and Young, which states that a symmetric and continuous kernel is positive definite if and only if it is integrally positive definite. More precisely, we extend the result from real-valued kernels on compact intervals to matrix-valued kernels on separable metric spaces. We also demonstrate the applications of the generalized theorem to the field of convex optimization.

Mathematics Subject Classification (2010/2020): 28A25, 43A35, 47B34

Keywords: positive definite kernel, matrix-valued kernel, measure theory, operator theory, convex optimization

1 Introduction and Main Results

Positive definite kernels are a generalization of positive definite functions, and were first introduced by Mercer in his seminal paper [13] from 1909. Namely, motivated by Hilbert's earlier work on Fredholm integral equations [11], Mercer's original goal was the characterization of all symmetric and continuous kernels $K : [a, b] \times [a, b] \to \mathbb{R}$ on a compact interval $[a, b] \subset \mathbb{R}$ which are strictly integrally positive definite with respect to the space $C([a, b], \mathbb{R})$ of continuous functions on $[a, b]$, i.e. which satisfy

$$\int_a^b \int_a^b f(x)K(x,y)f(y)dxdy > 0, \text{ for all } f \in C([a,b],\mathbb{R}) \text{ with } f \neq 0.$$

*ST is supported by the EPSRC Centre for Doctoral Training in Mathematics of Random Systems: Analysis, Modelling and Simulation (EP/S023925/1)

However, as mentioned by Mercer in the introduction of [13], he quickly figured out that this condition going back to Hilbert (see Chapter V of [11]) was too restrictive to allow a characterization by an equivalent discrete condition on the kernel K. Thus, he investigated kernels K which are integrally positive definite with respect to $C([a,b],\mathbb{R})$, i.e. which satisfy

$$\int_a^b \int_a^b f(x)K(x,y)f(y)dxdy \geq 0, \text{ for all } f \in C([a,b],\mathbb{R}).$$

He was then able to show that a symmetric and continuous kernel $K : [a,b]\times[a,b] \to \mathbb{R}$ is integrally positive definite with respect to $C([a,b],\mathbb{R})$ if and only if K is a positive definite kernel, that is, K satisfies

$$\sum_{i,j=1}^{n} c_i K(x_i,x_j)c_j \geq 0, \text{ for any } n \in \mathbb{N},\ x_1,\ldots,x_n \in [a,b],\ c_1,\ldots,c_n \in \mathbb{R}.$$

Here, positive definite kernels, as they are known today, appeared for the first time in the literature, where the term 'positive definite' is now well-established in spite of the weak inequality in the condition. Just one year later, Mercer's characterization theorem was generalized by Young [17] to the larger function space $L^1([a,b],\mathbb{R})$ of integrable functions on $[a,b]$. The results of Mercer and Young are summarized in the following theorem (see §10 in [13] and §1 in [17]).

Theorem 1.1. (Mercer-Young, 1909-1910) *Let $[a,b] \subset \mathbb{R}$ be a compact interval and $K : [a,b]\times[a,b] \to \mathbb{R}$ be a symmetric and continuous kernel. Then the following statements are equivalent:*

(i) K is positive definite, i.e.

$$\sum_{i,j=1}^{n} c_i K(x_i,x_j)c_j \geq 0, \text{ for any } n \in \mathbb{N},\ x_1,\ldots,x_n \in [a,b],\ c_1,\ldots,c_n \in \mathbb{R}.$$

(ii) K is integrally positive definite with respect to $C([a,b],\mathbb{R})$, i.e.

$$\int_a^b \int_a^b f(x)K(x,y)f(y)dxdy \geq 0, \text{ for all } f \in C([a,b],\mathbb{R}).$$

(iii) K is integrally positive definite with respect to $L^1([a,b],\mathbb{R})$, i.e.

$$\int_a^b \int_a^b f(x)K(x,y)f(y)dxdy \geq 0, \text{ for all } f \in L^1([a,b],\mathbb{R}).$$

The definition of a positive definite kernel may be generalized in various ways, most importantly, to more general domains and codomains. The object of this work is the generalization of Theorem 1.1 to matrix-valued kernels

$$K : X \times X \to \mathbb{R}^{N\times N}$$

on separable metric spaces, in particular allowing for kernels on Polish spaces. Our main result is the following:

Theorem 1.2. *Let X be a separable metric space equipped with a locally finite measure μ on $\mathcal{B}(X)$ that has support X. Let $N \in \mathbb{N}$ and $K : X \times X \to \mathbb{R}^{N\times N}$ be a bounded and continuous kernel satisfying $K(x, y) = K(y, x)^\top$ for all $x, y \in X$. Then the following statements are equivalent:*

(i) K is positive definite, i.e.

$$\sum_{i,j=1}^{n} c_i^\top K(x_i, x_j)c_j \geq 0, \text{ for any } n \in \mathbb{N},\ x_1, \ldots, x_n \in X,\ c_1, \ldots, c_n \in \mathbb{R}^N. \quad (1.1)$$

(ii) K is integrally positive definite with respect to $C(X, \mathbb{R}^N) \cap L^1(X, \mathbb{R}^N)$, i.e.

$$\int_X \int_X f(x)^\top K(x, y)f(y)d\mu(x)d\mu(y) \geq 0, \text{ for all } f \in C(X, \mathbb{R}^N) \cap L^1(X, \mathbb{R}^N). \quad (1.2)$$

(iii) K is integrally positive definite with respect to $L^1(X, \mathbb{R}^N)$, i.e.

$$\int_X \int_X f(x)^\top K(x, y)f(y)d\mu(x)d\mu(y) \geq 0, \text{ for all } f \in L^1(X, \mathbb{R}^N). \quad (1.3)$$

Remark 1.3. *Theorem 1.2 simplifies to Theorem 1.1 when choosing $X = [a, b]$, $N = 1$, and μ to be the Lebesgue measure on $\mathcal{B}([a, b])$, as all continuous functions on the compact inteval $[a, b]$ are bounded and thus Lebesgue-integrable.*

Remark 1.4. *If X is a separable metric space equipped with a locally finite measure μ on $\mathcal{B}(X)$ as in the assumptions of Theorem 1.2, then μ is also σ-finite. Namely, since μ is locally finite, for every $x \in X$ there exists an open neighborhood $U_x \subset X$ of x with $\mu(U_x) < \infty$. As a separable metric space, X is also a Lindelöf space, that is, every open cover of X has a countable subcover. In particular, there exists a countable subcover of $(U_x)_{x\in X}$, which yields that μ is σ-finite.*

Applications of Theorem 1.2. Since their introduction more than 100 years ago, positive definite kernels and their generalizations have found numerous applications in the fields of functional, numerical and stochastic analysis, approximation theory, computer graphics, engineering design, fluid dynamics, geostatistics, probability theory, statistics and machine learning (see the survey article by Fasshauer [8]). Therefore, the result of Theorem 1.2 is not only of theoretical interest, but also of relevance for applications in the aforementioned fields. In particular, the minimization of the energy integral

$$\mathcal{E}(\mu) = \int_{\mathbb{R}} \int_{\mathbb{R}} K(x-y) d\mu(x) d\mu(y),$$

over all probability measures μ supported on a fixed compact set $A \subset \mathbb{R}$ plays an important role in potential theory, where the minimizing measure μ^* is called a capacity measure (see e.g. [10]).

In recent years, a variant of this problem has gained significant interest in the areas of convex optimization and control theory, see [1, 2, 3, 9, 15] and references therein. In this class of optimization problems the cost functionals consist of quadratic terms containing kernels along with lower-order terms, i.e. they are of the form

$$V^1 := \min_{u \in \mathcal{A}} \mathbb{E}\left[\int_{\mathbb{R}_+} \int_{\mathbb{R}_+} u(t)^\top K(t,s) u(s) ds dt + ...\right], \tag{1.4}$$

where optimization is performed over a class $\mathcal{A}$ of admissible stochastic or deterministic controls in $L^1(\mathbb{R}_+, \mathbb{R}^N)$, and the expectation is omitted in the deterministic case. Results such as Theorem 1.2 are of central importance for approximating these problems and reducing their dimension. In particular, one may consider the following minimization problem,

$$V^2 := \inf_{\mathcal{P} \in \mathbb{P}} \min_{v \in \Xi(\mathcal{P})} \mathbb{E}\left[\sum_{i,j \geq 1} v_i^\top K(t_i, t_j) v_j + ...\right], \tag{1.5}$$

where $\mathbb{P}$ is the class of all locally finite partitions $\mathcal{P}$ of $\mathbb{R}_+$ and $\Xi(\mathcal{P})$ is the class of piecewise constant stochastic or deterministic controls in $\mathcal{A}$ varying only on the partition $\mathcal{P} \in \mathbb{P}$. Then (1.5) provides an approximation to the stochastic or deterministic control problem (1.4) by reducing it to a finite-dimensional matrix optimization problem (see Theorem 3 in [4] for a special case). In this setting, conditions (1.1) and (1.2) are essential for characterizing the convexity of the cost functionals in (1.5) and (1.4), respectively (see e.g [2, 4, 9]).

The extension of Theorem 1.1 to separable metric spaces is of particular interest when dealing with kernel methods in machine learning, where positive definite kernels defined on general input spaces are often employed. We also refer to [5, 6, 12], where the positive definite property of a matrix-valued kernel is associated with its reproducing property and applications in multi-task learning are studied. In this class of machine learning problems the associated loss functions are often quadratic and can be compared to (1.4), with general input spaces as domains of integration.

2 Proof of the Main Theorem

This section is dedicated to the proof of Theorem 1.2, which will be given by showing the following three implications,

$$(1.3) \Rightarrow (1.2), \quad (1.2) \Rightarrow (1.1), \quad (1.1) \Rightarrow (1.3).$$

In order to prove the third implication, the following auxiliary lemma will be needed:

Lemma 2.1. *Let X be an arbitrary set and $K : X \times X \to \mathbb{R}^{N \times N}$ be a mapping with $K(x,y) = K(y,x)^\top$ for all $x, y \in X$. Then condition (1.1) is equivalent to*

$$\sum_{i,j=1}^{n} \overline{z_i}^\top K(x_i, x_j) z_j \geq 0, \text{ for any } n \in \mathbb{N},\ x_1, \dots, x_n \in X,\ z_1, \dots, z_n \in \mathbb{C}^N. \tag{2.1}$$

Proof of Lemma 2.1. Let $n \in \mathbb{N}$, $x_1, \dots, x_n \in X$, $z_1, \dots, z_n \in \mathbb{C}^N$, and assume that condition (1.1) holds. Then a direct computation using (1.1) yields,

$$\begin{aligned}
&\sum_{i,j=1}^{n} \overline{z_i}^\top K(x_i, x_j) z_j = \sum_{i,j=1}^{n} (\mathrm{Re}(z_i) - i\,\mathrm{Im}(z_i))^\top K(x_i, x_j)(\mathrm{Re}(z_j) + i\,\mathrm{Im}(z_j)) \\
&\geq -i \sum_{i,j=1}^{n} \mathrm{Im}(z_i)^\top K(x_i, x_j) \,\mathrm{Re}(z_j) + i \sum_{i,j=1}^{n} \mathrm{Re}(z_i)^\top K(x_i, x_j) \,\mathrm{Im}(z_j) \\
&= -i \sum_{i,j=1}^{n} \mathrm{Im}(z_i)^\top K(x_i, x_j) \,\mathrm{Re}(z_j) + i \sum_{i,j=1}^{n} \mathrm{Im}(z_j)^\top K(x_i, x_j)^\top \,\mathrm{Re}(z_i) = 0,
\end{aligned}$$

since $K(x_i, x_j)^\top = K(x_j, x_i)$ by assumption. The reverse direction is trivial. □

Proof of Theorem 1.2. We proceed by showing the three implications,

$$(1.3) \Rightarrow (1.2), \quad (1.2) \Rightarrow (1.1), \quad (1.1) \Rightarrow (1.3).$$

The first implication (1.3) ⇒ (1.2) is trivial.

In order to prove (1.2) ⇒ (1.1), we will generalize Mercer's original proof (see Part II of [13]). Assume that condition (1.2) holds. Fix $n \in \mathbb{N}$ and n points $x_1, \dots, x_n \in X$. Without loss of generality, we can assume the n points to be distinct, because for $n \geq 2$, $x_{n-1} = x_n$ and any $c_1, \dots, c_n \in \mathbb{R}^N$, we can define $\widetilde{c}_{n-1} := c_{n-1} + c_n$ and $\widetilde{c_i} := c_i$ for $1 \leq i \leq n-3$ so that the following holds,

$$\sum_{i,j=1}^{n} c_i^\top K(x_i, x_j) c_j = \sum_{i,j=1}^{n-1} \widetilde{c_i}^\top K(x_i, x_j) \widetilde{c_j}.$$

This assumption allows us to choose real numbers $\delta, \varepsilon > 0$ sufficiently small such that

$$B_{\delta+\varepsilon}(x_i) \cap B_{\delta+\varepsilon}(x_j) = \emptyset, \quad \text{for all } i, j = 1, \dots, n \text{ with } i \neq j. \tag{2.2}$$

Since μ is locally finite, we can moreover assume that

$$\mu(\overline{B_{\delta+\varepsilon}(x_i)}) < \infty, \quad \text{for all } i = 1, \dots, n. \tag{2.3}$$

Here $B_r(x) \subset X$ denotes the open ball of radius $r > 0$ centred at $x \in X$. As a metric space, X is in particular a normal topological space. Therefore, using Urysohn's Lemma (see Lemma 15.6 in [16]), for $i = 1, \dots, n$, there exist continuous functions illustrated in Figure 1 given by,

$$f_{x_i,\delta,\varepsilon} : X \to [0, 1], \quad \text{with } f_{x_i,\delta,\varepsilon}(x) = \begin{cases} 1 & \text{for } x \in \overline{B_\delta(x_i)}, \\ 0 & \text{for } x \in X \setminus B_{\delta+\varepsilon}(x_i), \end{cases} \tag{2.4}$$

where the closed subsets $\overline{B_\delta(x_i)}$ and $X \setminus B_{\delta+\varepsilon}(x_i)$ of X are disjoint by definition.

Motivated by [13], we introduce several types of subsets of $X \times X$, which are illustrated in Figure 2. For all $i, j = 1, \dots, n$ let

$$Q_{ij} := \overline{B_{\delta+\varepsilon}(x_i)} \times \overline{B_{\delta+\varepsilon}(x_j)}, \quad q_{ij} := \overline{B_\delta(x_i)} \times \overline{B_\delta(x_j)}, \quad r_{ij} := Q_{ij} \setminus q_{ij}. \tag{2.5}$$

Now let $c_1, \dots, c_n \in \mathbb{R}^N$ be arbitrary vectors. Recalling (2.4), define the continuous and integrable function,

$$f_{\delta,\varepsilon} : X \to \mathbb{R}^N, \quad f_{\delta,\varepsilon}(x) := \sum_{i=1}^{n} c_i \, \frac{1}{\mu(\overline{B_\delta(x_i)})} f_{x_i,\delta,\varepsilon}(x), \tag{2.6}$$

where $0 < \mu(\overline{B_\delta(x_i)}) < \infty$ due to the assumption of Theorem 1.2 that the support of μ is X and (2.3).

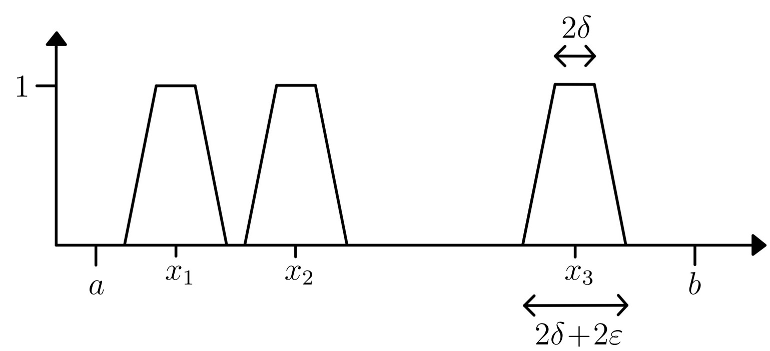

Figure 1: Functions $f_{x_i,\delta,\varepsilon}$ corresponding to given points $x_i \in X$ for $i \in \{1, \dots, n\}$ and given $\delta, \varepsilon > 0$ in the case where $X = [a, b]$ and $n = 3$.

We denote by $\mu^2 = \mu \times \mu$ the product measure on $\mathcal{B}(X \times X)$ and by $\|\cdot\|$ the Frobenius norm. It follows from (2.2), (2.4), (2.5), (2.6), and the submultiplicativity of the Frobenius norm that,

$$\begin{aligned}
&f_{\delta,\varepsilon}(x)^\top K(x,y) f_{\delta,\varepsilon}(y) = \frac{1}{\mu^2(q_{ij})} c_i^\top K(x,y) c_j, && \text{for } (x,y) \in q_{ij},\\
&|f_{\delta,\varepsilon}(x)^\top K(x,y) f_{\delta,\varepsilon}(y)| \le \frac{\|c_i\|\|c_j\|}{\mu^2(q_{ij})} \sup_{x,y\in X} \|K(x,y)\|, && \text{for } (x,y) \in r_{ij},\\
&f_{\delta,\varepsilon}(x)^\top K(x,y) f_{\delta,\varepsilon}(y) = 0, && \text{for } (x,y) \notin \bigcup_{i,j=1}^{n} Q_{ij}.
\end{aligned} \tag{2.7}$$

Therefore due to (2.7),

$$\begin{aligned}
&\int_X \int_X f_{\delta,\varepsilon}(x)^\top K(x,y) f_{\delta,\varepsilon}(y) d\mu(x) d\mu(y) \qquad (2.8)\\
&= \sum_{i,j=1}^{n} \frac{1}{\mu^2(q_{ij})} \int_{q_{ij}} c_i^\top K(x,y) c_j d\mu^2(x,y) + \sum_{i,j=1}^{n} \int_{r_{ij}} f_{\delta,\varepsilon}(x)^\top K(x,y) f_{\delta,\varepsilon}(y) d\mu^2(x,y).
\end{aligned}$$

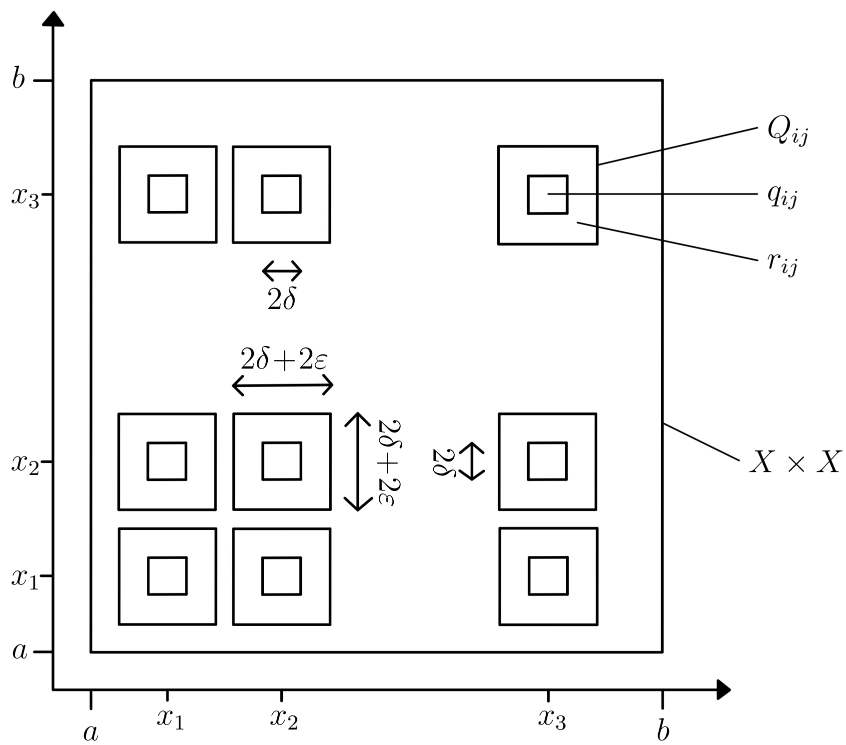

Figure 2: The subsets Q_{ij}, q_{ij}, $r_{ij} \subset X \times X$ associated with given points $x_i, x_j \in X$ for $i, j \in \{1, \ldots, n\}$ in the case where $X = [a, b]$ and $n = 3$.

Using (2.7), the second term on the right-hand side of (2.8) can be bounded as follows,

$$\Big|\sum_{i,j=1}^{n}\int_{r_{ij}} f_{\delta,\varepsilon}(x)^\top K(x,y) f_{\delta,\varepsilon}(y)d\mu^2(x,y)\Big| \le \sum_{i,j=1}^{n}\frac{\mu^2(r_{ij})}{\mu^2(q_{ij})}\|c_i\|\|c_j\|\sup_{x,y\in X}\|K(x,y)\|. \tag{2.9}$$

We prove (1.2) ⇒ (1.1) by contradiction. Assume (1.2). Then if (1.1) does not hold, there exists a real number $C>0$, depending on $x_1,\dots,x_n$ and $c_1,\dots,c_n$, with

$$\sum_{i,j=1}^{n} c_i^\top K(x_i,x_j)c_j \le -C. \tag{2.10}$$

We construct the functions $f_{\delta,\varepsilon}$ in (2.6) as follows. Due to the continuity of K, we can choose $\delta>0$ and thereby the sets q_{ij} sufficiently small such that

$$\big|c_i^\top K(x,y)c_j - c_i^\top K(x_i,x_j)c_j\big| \le \frac{C}{2n^2}, \quad \text{for all } (x,y)\in q_{ij},\ i,j=1,\dots,n. \tag{2.11}$$

Note that $\varepsilon>0$ in $f_{\delta,\varepsilon}$ will be fixed later.

It follows from (2.10) and (2.11) that the first term on the right-hand side of (2.8) is bounded from above as follows,

$$\begin{aligned}
&\sum_{i,j=1}^{n}\frac{1}{\mu^2(q_{ij})}\int_{q_{ij}} c_i^\top K(x,y)c_j d\mu^2(x,y)\\
&\le \sum_{i,j=1}^{n}\frac{1}{\mu^2(q_{ij})}\int_{q_{ij}}\Big(c_i^\top K(x_i,x_j)c_j+\frac{C}{2n^2}\Big)d\mu^2(x,y)\\
&= \sum_{i,j=1}^{n}\Big(c_i^\top K(x_i,x_j)c_j+\frac{C}{2n^2}\Big)\\
&\le -C+\frac{C}{2} = -\frac{C}{2}.
\end{aligned} \tag{2.12}$$

Now (2.8) and (2.12) yield the following,

$$\int_X\int_X f_{\delta,\varepsilon}(x)^\top K(x,y) f_{\delta,\varepsilon}(y)d\mu(x)d\mu(y) \le \sum_{i,j=1}^{n}\int_{r_{ij}} f(x)^\top K(x,y) f_{\delta,\varepsilon}(y)d\mu^2(x,y) - \frac{C}{2},$$

which together with assumption (1.2) applied to $f_{\delta,\varepsilon}$ and (2.9) implies,

$$\frac{C}{2} \le \sum_{i,j=1}^{n}\frac{\mu^2(r_{ij})}{\mu^2(q_{ij})}\|c_i\|\|c_j\|\sup_{x,y\in X}\|K(x,y)\|. \tag{2.13}$$

Note that the right-hand side of (2.13) is finite as K is bounded by assumption. Finally, it follows from (2.3) and the definition of the sets r_{ij} in (2.5) that

$$\lim_{\varepsilon\to 0}\mu^2(r_{ij}) = \lim_{\varepsilon\to 0}\mu^2(\overline{B_{\delta+\varepsilon}(x_i)}\times\overline{B_{\delta+\varepsilon}(x_j)}) - \mu^2(\overline{B_\delta(x_i)}\times\overline{B_\delta(x_j)}) = 0,$$

which implies that if $\varepsilon > 0$ is chosen sufficiently small, (2.13) does not hold and we get a contradiction. This finishes the proof of the implication (1.2) $\Rightarrow$ (1.1).

For the third implication (1.1) $\Rightarrow$ (1.3), assume that (1.1) holds.

Case 1: First assume that μ is a finite measure. The key idea of the proof is to apply a generalized version of Mercer's theorem for matrix-valued kernels (see [7]). In order to apply this result, note that (1.1) implies (2.1) by Lemma 2.1. Moreover, since K is bounded and μ is finite by assumption, it holds that

$$\int_X \operatorname{Tr} K(x,x)d\mu(x) \leq N\mu(X) \sup_{1\leq \ell\leq N} \sup_{x\in X} K_{\ell\ell}(x,x) < \infty, \tag{2.14}$$

where Tr denotes the trace of a matrix in $\mathbb{R}^{N\times N}$. Note that G must be nonnegative on the diagonal, which can be seen from condition (1.1) by choosing $n=1$ and $c_1 = e_l$ for example. Next, define the linear integral operator

$$\mathbf{K} : L^2(X,\mathbb{C}^N) \to L^2(X,\mathbb{C}^N), \quad (\mathbf{K}f)(x) := \int_X K(x,y)f(y)d\mu(y),$$

which is well-defined by Theorem A.1 in [7], and denote by $\ker(\mathbf{K})$ its kernel. Since (2.1) and (2.14) hold, we can apply the generalized version of Mercer's theorem for matrix-valued kernels (see Theorem 4.1 in [7]). In combination with the auxiliary Theorem A.1 in [7], it implies the existence of a countable orthonormal basis $\{\varphi_k\}_{k\in I}$ of $\ker(\mathbf{K})^\perp \subset L^2(X,\mathbb{C}^N)$ of continuous eigenfunctions of $\mathbf{K}$ with a corresponding family $\{\sigma_k\}_{k\in I} \subset (0,\infty)$ of positive eigenvalues such that

$$K_{\ell m}(x,y) = \sum_{k\in I} \sigma_k \overline{\varphi_k^\ell(x)}\varphi_k^m(y), \quad \text{for all } (x,y)\in X\times X,\ \ell,m\in\{1,\dots,N\}, \tag{2.15}$$

where the series converges uniformly on $X\times X$. Now let $f\in L^1(X,\mathbb{R}^N)$. From (2.15) it follows that,

$$\begin{aligned}
&\int_X\int_X f(x)^\top K(x,y)f(y)d\mu(x)d\mu(y) \\
&= \int_X\int_X \sum_{\ell,m=1}^N f^\ell(x)\Big(\sum_{k\in I}\sigma_k\overline{\varphi_k^\ell(x)}\varphi_k^m(y)\Big)f^m(y)d\mu(x)d\mu(y) \\
&= \int_X\int_X \sum_{k\in I}\sigma_k \sum_{\ell,m=1}^N f^\ell(x)\overline{\varphi_k^\ell(x)}\varphi_k^m(y)f^m(y)d\mu(x)d\mu(y) \\
&= \int_X\int_X\int_I \sigma_k \sum_{\ell,m=1}^N f^\ell(x)\overline{\varphi_k^\ell(x)}\varphi_k^m(y)f^m(y)d\nu(k)d\mu(x)d\mu(y),
\end{aligned} \tag{2.16}$$

where ν is the counting measure on the countable index set of the eigenfunctions I. In particular, ν is σ-finite.

Next, we would like to switch the order of integration on the right-hand side of (2.16) using Fubini's theorem. In order to do that, we show that the following integral is finite, using Young's inequality, (2.14), (2.15) and the fact that $f \in L^1(X, \mathbb{R}^N)$,

$$\begin{aligned}
&\int_X \int_X \int_I \left|\sigma_k \sum_{\ell,m=1}^N f^\ell(x)\overline{\varphi_k^\ell(x)}\varphi_k^m(y) f^m(y)\right| d\nu(k) d\mu(x) d\mu(y) \\
&\leq \int_X \int_X \sum_{\ell,m=1}^N \left|f^\ell(x)\right| \int_I \sigma_k \left(\left|\varphi_k^\ell(x)\right|^2 + \left|\varphi_k^m(y)\right|^2\right) d\nu(k) \left|f^m(y)\right| d\mu(x) d\mu(y) \\
&= \int_X \int_X \sum_{\ell,m=1}^N \left|f^\ell(x)\right| \left(K_{\ell\ell}(x,x) + K_{mm}(y,y)\right) \left|f^m(y)\right| d\mu(x) d\mu(y) \\
&\leq 2\Big(\sup_{1\leq \ell \leq N} \sup_{x\in X} K_{\ell\ell}(x,x)\Big) \sum_{\ell,m=1}^N \int_X \left|f^\ell(x)\right| d\mu(x) \int_X \left|f^m(y)\right| d\mu(y) \\
&< \infty.
\end{aligned}$$

Let $\mathbf{1}_N \in \mathbb{R}^{N\times N}$ denote the matrix of ones in all entries, which is symmetric positive semidefinite as its eigenvalues are N with multiplicity 1 and 0 with multiplicity $N-1$. Moreover, for every $k \in I$ we define

$$g_k : X \to \mathbb{C}^N, \quad g_k(x) := \begin{pmatrix} \varphi_k^1(x) f^1(x) \\ \vdots \\ \varphi_k^N(x) f^N(x) \end{pmatrix},$$

which is the Hadamard product of φ_k and f. Then (2.16) and Fubini's theorem imply

$$\begin{aligned}
&\int_X \int_X f(x)^\top K(x,y) f(y) d\mu(x) d\mu(y) \\
&= \int_I \int_X \int_X \sigma_k \sum_{\ell,m=1}^N f^\ell(x)\overline{\varphi_k^\ell(x)}\varphi_k^m(y) f^m(y) d\mu(x) d\mu(y) d\nu(k) \\
&= \int_I \int_X \int_X \sigma_k\, \overline{g_k(x)}^\top \mathbf{1}_N\, g_k(y) d\mu(x) d\mu(y) d\nu(k) \\
&= \sum_{k\in I} \sigma_k \overline{\Big(\int_X g_k(x) d\mu(x)\Big)}^\top \mathbf{1}_N \Big(\int_X g_k(y) d\mu(y)\Big) \geq 0.
\end{aligned}$$

This shows the implication (1.1) $\Rightarrow$ (1.3) in the case where μ is finite.

Case 2: Now assume that μ is a locally finite measure. Then μ is also σ-finite due to Remark 1.4. Let $(X_s)_{s\in\mathbb{N}}$ be an increasing sequence of $\mathcal{B}(X)$-measurable subsets of X with $\mu(X_s) < \infty$ for all $s \geq 1$ and $\bigcup_{s\geq 1} X_s = X$. Let $f \in L^1(X, \mathbb{R}^N)$. Define the sequence of functions $(F_s)_{s\geq 1}$ by

$$F_s : X \times X \to \mathbb{R}, \quad F_s(x,y) := \mathbb{1}_{X_s}(x)\mathbb{1}_{X_s}(y) f(x)^\top K(x,y) f(y), \quad \text{for } s \geq 1.$$

Then $(F_s)_{s\geq 1}$ converges pointwise to the function $f(x)^\top K(x,y)f(y)$ on $X\times X$, since for every $(x,y)\in X\times X$ there exists an $s_0\in\mathbb{N}$ such that $x,y\in X_s$ for all $s\geq s_0$. Moreover, it holds for all $s\geq 1$ that

$$\begin{aligned}&\int_X\int_X \big|F_s(x,y)\big|d\mu(x)d\mu(y)\\ &\leq \int_X\int_X \mathbb{1}_{X_s}(x)\mathbb{1}_{X_s}(y)\big|f(x)\big|\big\|K(x,y)\big\|\big|f(y)\big|d\mu(x)d\mu(y)\\ &\leq \int_X\int_X \big|f(x)\big|\big|f(y)\big|d\mu(x)d\mu(y)\sup_{x,y\in X}\|K(x,y)\|\\ &<\infty,\end{aligned}$$

where we used the submultiplicativity of the Frobenius norm, the integrability of f and the boundedness of K by the hypothesis. Thus dominated convergence can be applied to get

$$\begin{aligned}\int_X\int_X f(x)^\top K(x,y)f(y)d\mu(x)d\mu(y) &= \lim_{s\to\infty}\int_X\int_X F_s(x,y)d\mu(x)d\mu(y)\\ &= \lim_{s\to\infty}\int_{X_s}\int_{X_s} f(x)^\top K(x,y)f(y)d\mu(x)d\mu(y)\\ &\geq 0,\end{aligned}$$

where the last inequality follows from Case 1, since for all $s\geq 1$ we have $\mu(X_s)<\infty$ and that $(X_s,\mathcal{B}(X_s),\mu|_{X_s})$ and $K|_{X_s\times X_s}$ satisfy the assumptions of Theorem 1.2 and condition (1.1).

This shows the implication (1.1) $\Rightarrow$ (1.3) in the case where μ is locally finite and completes the proof.

□

References

[1] E. Abi Jaber, E. Miller, and H. Pham. Integral operator Riccati equations arising in stochastic volterra control problems. *SIAM Journal on Control and Optimization*, 59(2):1581–1603, 2021.

[2] E. Abi Jaber, E. Neuman, and M. Voß. Equilibrium in functional stochastic games with mean-field interaction. *arXiv preprint arXiv:2306.05433*, 2023.

[3] A. Alfonsi and A. Schied. Capacitary measures for completely monotone kernels via singular control. *SIAM Journal on Control and Optimization*, 51(2): 1758–1780, 2013. doi: 10.1137/120862223. URL https://doi.org/10.1137/120862223.

[4] A. Alfonsi, F. Klöck, and A. Schied. Multivariate transient price impact and matrix-valued positive definite functions. *Mathematics of Operations Research*, 41(3):914–934, 2016. ISSN 0364765X, 15265471. URL http://www.jstor.org/stable/24736389.

[5] C. Brouard, M. Szafranski, and F. d'Alché Buc. Input output kernel regression: Supervised and semi-supervised structured output prediction with operator-valued kernels. *Journal of Machine Learning Research*, 17(176):1–48, 2016.

[6] A. Caponnetto, C. A. Micchelli, M. Pontil, and Y. Ying. Universal multi-task kernels. *The Journal of Machine Learning Research*, 9:1615–1646, 2008.

[7] E. De Vito, V. Umanità, and S. Villa. An extension of Mercer theorem to matrix-valued measurable kernels. *Applied and Computational Harmonic Analysis*, 34 (3):339–351, 2013.

[8] G. E. Fasshauer. Positive definite kernels: past, present and future. *Dolomites Research Notes on Approximation*, 4:21–63, 2011.

[9] J. Gatheral, A. Schied, and A. Slynko. Transient linear price impact and Fredholm integral equations. *Math. Finance*, 22:445–474, 2012.

[10] L. L. Helms. *Potential theory*. Universitext. Springer London, 2009.

[11] D. Hilbert. Grundzüge einer allgemeinen Theorie der linearen Integralgleichungen I. *Göttinger Nachrichten*, 1904:49–91, 1904.

[12] H. Kadri, E. Duflos, P. Preux, S. Canu, A. Rakotomamonjy, and J. Audiffren. Operator-valued kernels for learning from functional response data. *Journal of Machine Learning Research*, 17(20):1–54, 2016.

[13] J. Mercer. Functions of positive and negative type, and their connection the theory of integral equations. *Philosophical transactions of the royal society of London. Series A*, 209:415–446, 1909.

[14] J. Stewart. Positive definite functions and generalizations, an historical survey. *The Rocky Mountain Journal of Mathematics*, 6(3):409–434, 1976.

[15] H. Wang, J. Yong, and C. Zhou. Linear-quadratic optimal controls for stochastic Volterra integral equations: Causal state feedback and path-dependent riccati equations. *SIAM Journal on Control and Optimization*, 61(4):2595–2629, 2024/03/22 2023. doi: 10.1137/22M1492696. URL https://doi.org/10.1137/22M1492696.

[16] S. Willard. *General Topology*. Dover Publications, 2012.

[17] W. H. Young. A note on a class of symmetric functions and on a theorem required in the theory of integral equations. *Messenger of Mathematics*, 40:37–43, 1910.