Supplementary Material for Spectral Regularization for Support Estimation

Ernesto De Vito
DSA, Univ. di Genova, and
INFN, Sezione di Genova, Italy
devito@dima.ungie.it

Lorenzo Rosasco
CBCL - MIT, - USA, and
IIT, Italy
lrosasco@mit.edu

Alessandro Toigo
Politec. di Milano, Dept. of Math., and
INFN, Sezione di Milano, Italy
toigo@ge.infn.it

The supplementary material contains two sections. The first collect all the proofs in the paper. The second, gives further developments and examples of completely regular RKHS.

1 Proofs

In this section we collect all the proofs.

Proof of Proposition 1. We prove that (i) implies (ii). Given $x_0 \notin C$, by contradiction, assume there is $f \in \mathcal{H}$ such that $f(x_0) \neq 0$, $f(x) = 0$ for all $x \in C$ and $K_{x_0} \in \operatorname{Ran} P_C$. Clearly, $f \in \ker P_C = \operatorname{Ran} P_C^{\perp}$, so that $f(x_0) = \langle f, K_{x_0} \rangle = 0$, which is a contradiction.
We prove that (ii) implies (iii). If $x \in C$, then $K_x \in \operatorname{Ran} P_C$ by definition of P_C, so that $F_C(x) = K(x,x)$, hence $C \subset \{x \in X \mid K(x,x) = F_C(x)\}$. If $x \notin C$, by assumption $(I - P_C)K_x \neq 0$, thus $K(x,x) - F_C(x) = \|(I - P_C)K_x\|^2 \neq 0$ and $C \supset \{x \in X \mid K(x,x) = F_C(x)\}$.
We prove that (iii) implies (i). If $x_0 \notin C$, define $f = (I - P_C)K_{x_0} \in \ker P_C$, so that $f(x) = 0$ for all $x \in C$. Furthermore, $f(x_0) = K(x_0, x_0) - F_C(x_0) \neq 0$.
Clearly, (ii) implies that $K(x,x) \neq 0$ for all $x \notin C$. $\qquad\square$

Proof of Proposition 2. To prove that the sets separated by $\mathcal{H}$ are closed with respect to d_K, note that by definition of the metric d_K, the map $x \mapsto K_x$ is continuous from X, endowed with the metric d_K, into $\mathcal{H}$, so that $x \mapsto K(x,x) - F_C(x)$ is a continuous map, hence its zero level set is a closed subset of X.
To show that if $\mathcal{H}$ is separable and the kernel is measurable, then the sets separated by $\mathcal{H}$ are measurable, we first observe that the map $x \mapsto K_x$ is measurable from X into $\mathcal{H}$ since $\mathcal{H}$ is separable and K is a measurable kernel (Proposition 3.1 in [1]). As above, it follows that $x \mapsto K(x,x) - F_C(x)$ is a measurable map, so that its zero level set is a measurable subset of X. $\qquad\square$

Proof of Corollary 1.   (i) This follows from Proposition 12 and Corollary 3 in [1].

    (ii) Condition (a) states that, for any sequence $(x_j)_{j \in \mathbb{N}}$ such that $\lim_j d_X(x_j, x) = 0$ for some $x \in X$, then $\lim_j d_K(x_j, x) = 0$. Hence, the d_K-closed subsets are d_X-closed, too. Conversely, if C is d_X-closed, (b) implies that $C = \{x \in X \mid K(x,x) - F_C(x) = 0\}$, which is a d_K-closed subset by (i) of Proposition 2.

    (iii) It follows from (ii) and (b). Note that, since the points are closed sets for d_X, condition (b) implies that $K_x \neq K_t$ if $x \neq t$.

$\qquad\square$

Proof of Lemma 1. Denote by $\mathcal{H}'$ the reproducing kernel Hilbert space with kernel K', and define the map $\Phi : X \to \mathcal{H}$, $\Phi(x) = K_x / \|K_x\|$. It is simple to check that $\langle \Phi(y), \Phi(x) \rangle = K'(x, y)$ and $\Phi(X)^\perp = \{0\}$, so that the map $\Phi_* : \mathcal{H} \to \mathcal{H}'$

$$(\Phi_* f)(x) = \langle f, \Phi(x) \rangle$$

is a unitary operator with $K'_x = \Phi_* \Phi(x)$. Clearly, for any $f \in \mathcal{H}$ and $x \in X$

$$\langle \Phi_* f, K'_x \rangle_{\mathcal{H}'} = \langle \Phi_* f, \Phi_* \Phi(x) \rangle_{\mathcal{H}'} = \frac{\langle f, K_x \rangle_{\mathcal{H}}}{\|K_x\|}.$$

The above equality shows that $\mathcal{H}$ and $\mathcal{H}'$ separate the same sets. $\qquad\square$

Proof of Theorem 1. Without loss of generality, we can assume that the sequence (λ_n) is nondecreasing. For any n define $G_n : X \to \mathbb{R}$ as $G_n(x) = \langle (T + \lambda_n I)^{-1} T K_x, K_x \rangle$. Clearly, G_n is continuous. The idea of the proof is to split the error in an approximation error controlling the deviation of G_n from F_ρ and a sample error controlling the deviation of F_n from G_n.

We first deal with the approximation error. Since T is compact and positive, there exists an orthonormal basis (e_j) of eigenvectors of T with corresponding sequence (σ_j) of positive eigenvalues. Hence

$$G_n(x) = \sum_k (\frac{\sigma_j}{\sigma_j + \lambda_n}) |\langle K_x, e_j \rangle|^2.$$

It follows that $G_n(x)$ converges to $F_\rho(x)$ for all $x \in X$ and the sequence $(G_n(x))$ is nondecreasing. Dini's theorem implies that $\sup_{x \in C} |G_n(x) - F_\rho(x)|$ converges to zero for every compact set C. For the sample error we can note that

$$(T + \lambda_n I)^{-1} T - (T_n + \lambda_n I)^{-1} T_n = \lambda_n (T + \lambda_n I)^{-1} (T_n - T)(T_n + \lambda_n I)^{-1},$$

and $\|(T + \lambda_n I)^{-1} K_x\| \leq 1/\lambda_n$ as well as $\|(T_n + \lambda_n I)^{-1} K_x\| \leq 1/\lambda_n$. Then $\sup_{x \in X} |G_n(x) - F_n(x)| \leq \frac{1}{\lambda_n} \|T - T_n\|_{HS}$ and we can use the concentration results for $\|T - T_n\|_{HS}$ and the proposed regularization parameter choice to prove convergence of F_n to F_ρ. $\qquad\square$

Proof of Theorem 2. Without loss of generality, we assume that X is itself compact and we prove the statement for $C = X$. The proof is split into two steps. First we show that

$$\lim_{n \to +\infty} \sup_{x \in X_\rho} d_K(x, X_n) = 0$$

Indeed the considered choice of τ_n implies that there exists $n_0 \in \mathbb{N}$ such that for all $n \geq n_0$

$$|F_n(x) - F_\rho(x)| \leq \tau_n \qquad \forall x \in X.$$

If $x \in X_\rho$,

$$F_n(x) - 1 = F_n(x) - F_\rho(x) \geq -|F_n(x) - F_\rho(x)| \geq -\tau_n,$$

so $x \in X_n$ and, hence, $d_K(x, X_n) = 0$ for all $n \geq n_0$.

Then, we prove that

$$\lim_{n \to +\infty} \sup_{x \in X_n} d_K(x, X_\rho) = 0$$

by contradiction. Assume the opposite, then there exists $\epsilon > 0$ such that for all $k \in \mathbb{N}$ there is $n_k \geq k$ and $\sup_{x \in X_{n_k}} d_K(x, X_\rho) \geq 2\epsilon$. Hence there is $x_k \in X_{n_k}$ such that

$$d_K(x_k, x) \geq \epsilon \qquad \text{for all } x \in X_\rho. \tag{1}$$

Since X is compact, possibly passing to a subsequence we can assume that $(x_k)_{k \in \mathbb{N}}$ converges to a $x_0 \in X$. We claim that $x_0 \in X_\rho$. Indeed

$$|F_\rho(x_0) - 1| \leq |F_\rho(x_0) - F_\rho(x_k)| + |F_\rho(x_k) - F_{n_k}(x_k)| + |F_{n_k}(x_k) - 1|$$
$$\leq |F_\rho(x_0) - F_\rho(x_k)| + \sup_{x \in X} |F_\rho(x) - F_{n_k}(x)| + \tau_{n_k}$$

where the third term is due to the fact that $x_k \in X_{n_k}$ so that

$$1 + \tau_{n_k} \geq 1 \geq F_{n_k}(x_k) \geq 1 - \tau_{n_k}.$$

Since n_k goes to $+\infty$, F_ρ is continuous in x_0 and F_n converges to F_ρ uniformly, it follows that $F_\rho(x_0) = 1$, that is $x_0 \in X_\rho$. However, (1) implies that $d(x_0, x) \geq \epsilon$ for all $x \in X_\rho$, so that there is a contradiction. $\qquad\square$

2 Complete Regularity: sufficient conditions and examples

In this section we provide some sufficient conditions characterizing completely regular reproducing kernel Hilbert spaces and we give some examples of such spaces.

The first result is about translation invariant kernels on $\mathbb{R}^d$. We show that if Fourier transform of the kernel satisfies a suitable growth condition, then the corresponding reproducing kernel Hilbert space is completely regular. In the following $L^1(\mathbb{R}^d)$ is the space of integrable functions with respect to the Lebesgue measure dx in $\mathbb{R}^d$ and $\hat{f}$ is the Fourier transform of $f \in L^1(\mathbb{R}^d)$, defined as

$$\hat{f}(\omega) = \int e^{-2\pi i \omega \cdot x} f(x) dx$$

Theorem. *Let $K : \mathbb{R}^d \to \mathbb{C}$ be a continuous function in $L^1(\mathbb{R}^d)$ such that*

$$\hat{K}(\omega) \geq \frac{a}{(1 + b\|\omega\|^m)^n} \qquad \forall \omega \in \mathbb{R}^d \tag{2}$$

for suitable $m, n \in \mathbb{N}$ and $a, b > 0$. Then,

    (i) the translation invariant kernel $K(x,t) = K(x - t)$ is positive definite and continuous;

    (ii) the corresponding reproducing kernel Hilbert space $\mathcal{H}$ is completely regular.

The proof depends on an explicit characterization of the reproducing kernel Hilbert space $\mathcal{H}$ given by the following result.

Proposition. *Let $K : \mathbb{R}^d \to \mathbb{C}$ be a continuous function in $L^1(\mathbb{R}^d)$ such that $\hat{K} = \hat{K}$ is strictly positive, then the kernel $K(x,t) = K(x - t)$ is positive definite and the corresponding reproducing kernel is*

$$\mathcal{H} = \left\{ f \in L^2 \mid \int \hat{K}(\omega)^{-1} |\hat{f}(\omega)|^2 d\omega < \infty \right\} \tag{3}$$

Proof. Denote by $L^2 = L^2(\mathbb{R}^d)$ the space of square integrable functions with scalar product $\langle \cdot, \cdot \rangle_2$ and recall that $\mathcal{F}$ is a unitary operator in L^2. We claim that K is a positive definite kernel and $\hat{K} \in L^1(\mathbb{R}^d)$. Let L_K be the integral operator of kernel K with respect to the Lebesgue measure, namely

$$(L_K f)(t) = \int K(t - x) f(x)\, dx = (K * f)(t),$$

which is a bounded operator on L^2 since $K \in L^1(\mathbb{R}^d)$. Since L_K is a convolution operator, the Fourier transform makes L_K unitary equivalent to the multiplicative operator by $\hat{K}$. It follows that

$$\langle L_K f, f \rangle_2 = \left\langle \hat{K} \cdot \hat{f}, \hat{f} \right\rangle_2 \geq 0 \qquad \forall\, f \in L^2(\mathbb{R}^d)$$

since $\hat{K} \geq 0$ by assumption. Hence L_K is a positive operator, so that K is positive definite. Indeed, let $\{\varphi_n\}_{n \in \mathbb{N}}$ be a Dirac sequence in 0 and, for each $x \in X$ define φ_n^x as $\varphi_n^x(t) = \varphi_n(t - x)$. Fixed $\{x_i\}_{i=1,2\ldots N} \subset \mathbb{R}^d$ and $\{c_i\}_{i=1,2\ldots N} \subset \mathbb{C}$, set $\phi_n = \sum_{i=1}^n c_i \varphi_n^{x_i}$, then

$$0 \leq \langle L_K \phi_n, \phi_n \rangle_2 = \sum_{i,j=1}^n c_i \overline{c_j} \langle L_K \varphi_n^{x_i}, \varphi_n^{x_j} \rangle \underset{n \to \infty}{\longrightarrow} \sum_{i,j=1}^n c_i \overline{c_j} K(x_i, x_j),$$

where the last equality is due to the continuity of K. Since K is a positive definite function, the Fourier inversion theorem in $L^1(\mathbb{R}^d)$ implies that $\hat{K} \in L^1(\mathbb{R}^d)$, see [2].

Finally, let $\mathcal{H}$ be the corresponding reproducing kernel Hilbert space. Since the support of the Lebesgue measure is $\mathbb{R}^d$, a generalization of the Mercer theorem [1] implies that $L_K^{1/2}$ is a unitary isomorphism of L^2 onto $\mathcal{H}$. Clearly $\widehat{L_K^{1/2} f} = \hat{K}^{1/2} \hat{f}$, so that the (3) follows. $\qquad\square$

The first part of the above proposition is a converse result of Bochner theorem [2]. Given the above proposition, the proof of the Theorem is straightforward.

Proof of the Theorem. Condition (2) implies that $\hat{K}$ is strictly positive, so that (i) is proved by Proposition 2. In particular, from (3) we see that $f \in \mathcal{H}$ if and only if $f \in L^2$ and

$$\int \hat{K}(\omega)^{-1}|\hat{f}(\omega)|^2 d\omega < \infty.$$

Clearly, if $f \in C_c^\infty(\mathbb{R}^d)$, then $\hat{f}$ is a Schwartz function on $\mathbb{R}^d$, so that $f \in \mathcal{H}$ by (2). □

A second result gives a very simple tool to construct completely regular reproducing kernel Hilbert spaces on high dimensional spaces.

Proposition. *If X_i, $i = 1, 2 \ldots d$, are sets and $\mathcal{H}_{K_i}$ are completely regular reproducing kernel Hilbert spaces on X_i for all $i = 1, 2 \ldots d$, then $\mathcal{H}_K$ is completely regular on the product space $X = X_1 \times X_2 \ldots X_d$, where K is the product kernel $K = K_1 K_2 \ldots K_d$.*

Proof. Each set X_i and X are endowed with the metric d_K induced by the corresponding kernel. We claim that in this way X is the topological product of the X_i's. Indeed, $\mathcal{H}_K = \mathcal{H}_{K_1} \otimes \ldots \otimes \mathcal{H}_{K_d}$, and, if $x = (x_1, \ldots, x_d) \in X$, then $K_x = K_{x_1} \otimes \ldots \otimes K_{x_d}$. It follows that, if $\{x_{i,k}\}_{k \in \mathbb{N}}$ is a sequence in X_i such that $\lim_k x_{i,k} = x_i$, then

$$\begin{aligned} \lim_k d_K \left((x_{1,k}, \ldots, x_{d,k}), (x_1, \ldots, x_d) \right)^2 &= \lim_k \|K_{(x_{1,k}, \ldots, x_{d,k})} - K_{(x_1, \ldots, x_d)}\|^2 \\ &= \lim_k K(x_{1,k}, x_{1,k}) \ldots K(x_{d,k}, x_{d,k}) - 2\mathrm{Re}\left(K(x_{1,k}, x_1) \ldots K(x_{d,k}, x_d)\right) \\ &\quad + K(x_1, x_1) \ldots K(x_d, x_d) \end{aligned}$$

Since $\lim_k K(x_{i,k}, x_{i,k}) = \lim_k K(x_{i,k}, x_i) = K(x_i, x_i)$, the claim follows.

If $C \subset X$ is closed and $x = (x_1, \ldots, x_d) \in X \setminus C$, let U_i, $i = 1, \ldots d$, be open neighborhoods of x_i in X_i such that $U = U_1 \times \ldots \times U_d$ is disjoint from C. Pick $f_i \in \mathcal{H}_{K_i}$ such that $f_i(x_i) \neq 0$ and $f_i|_{X_i \setminus U_i} = 0$. Then the product function $f(x_1, \ldots, x_d) = f_1(x_1) \ldots f_d(x_d)$ is in $\mathcal{H}_K$, and satisfies $f(x) \neq 0$ and $f|_C = 0$. □

We end the section by presenting wo classes of completely regular reproducing kernel Hilbert spaces with exponential kernels. The first result is about exponential kernels defined by a euclidean metric.

Proposition. *Let*
$$K : \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}, \qquad K(x, y) = e^{-\alpha\|x-y\|},$$
with $\alpha > 0$. Then K is a positive definite kernel and the corresponding reproducing kernel Hilbert space $\mathcal{H}$ is completely regular for all $d \in \mathbb{N} \setminus \{0\}$.

Proof. Since K is a radial function, its Fourier transform is

$$\begin{aligned} \hat{K}(\omega) &= \frac{2\pi}{\|\omega\|^{\frac{d-2}{2}}} \int_0^\infty e^{-\alpha r} r^{\frac{d}{2}} J_{\frac{d-2}{2}}(2\pi\|\omega\|r) dr \\ &= 2^d \pi^{\frac{d-1}{2}} \alpha \Gamma(\frac{d+1}{2}) \left(\alpha^2 + 4\pi^2\|\omega^2\|\right)^{-\frac{d+1}{2}}, \end{aligned} \tag{4}$$

where J_n is the Bessel function of order n, Γ is Euler gamma function, and we used formula 6.623(2) p. 712 in [3] to evaluate the integral. The claim then follows from the above Theorem. □

Eqs. (4) shows that (up to a constant rescaling of norms)

$$\mathcal{H} = W^{\frac{d+1}{2}}(\mathbb{R}^d),$$

where W^s is the Sobolev space of order s.

Finally, we consider the exponential kernel defined by the ℓ_1-norm, that is

$$K(x, y) = e^{-\alpha\|x-y\|_1} \qquad \|(x_1, x_2 \ldots x_d)\|_1 = |x_1| + |x_2| + \ldots + |x_d|.$$

For $d = 1$, the last proposition shows that $\mathcal{H}_K$ is completely regular. The same is true for arbitrary $d \geq 2$ as a consequence of the previous results.

References

[1] C. Carmeli, E. De Vito, and A. Toigo. Vector valued reproducing kernel Hilbert spaces of integrable functions and Mercer theorem. *Anal. Appl. (Singap.)*, 4(4):377–408, 2006.

[2] G. Folland. *A course in abstract harmonic analysis*. Studies in Advanced Mathematics. CRC Press, Boca Raton, FL, 1995.

[3] I. S. Gradshteyn and I. M. Ryzhik. *Table of integrals, series, and products*. Fourth edition prepared by Ju. V. Geronimus and M. Ju. Ceĭtlin. Translated from the Russian by Scripta Technica, Inc. Translation edited by Alan Jeffrey. Academic Press, New York, 1965.