Proof of Theorem 2 in the IJCNN submission

Sepp Hochreiter and Klaus Obermayer
Department of Electrical Engineering and Computer Science
Technische Universität Berlin
10587 Berlin, Germany
{hochreit,oby}@cs.tu-berlin.de

Theorem 1 (skipped)

Theorem 2 (Poisson Equation) *Assume that the kernel $k(\boldsymbol{a},\boldsymbol{b}) : T \times T \rightarrow \mathbb{R}$ is two times continuously differentiable, symmetric, and positive definite, and that T is compact. Assume further that forces are symmetric:*

$$\boldsymbol{\nabla}_{\boldsymbol{a}} k(\boldsymbol{a},\boldsymbol{b}) \;=\; -\, \boldsymbol{\nabla}_{\boldsymbol{b}} k(\boldsymbol{a},\boldsymbol{b}) \ .$$

If uniform convergence holds for each ρ then k must be of the following form: k can be partitioned into kernels $k = \sum_l k_{U(\lambda_l)}$, where the $U(\lambda_l)$ form a partition of T and the $k_{U(\lambda_l)}$ obey the following Dirichlet problems on $U(\lambda_l)$ (Poisson equation):

$$\boldsymbol{\nabla}^2_{\boldsymbol{a}} (-\ k_{U(\lambda_l)}(\boldsymbol{a},\boldsymbol{b})) \;=\; \lambda_l \ \delta_{U(\lambda_l)}(\boldsymbol{a}-\boldsymbol{b}) \ \ , \qquad (1)$$

where $\delta_{U(\lambda_l)}$ is the delta function restricted to $U(\lambda_l)$ and $0 \leq \lambda_l$.

Proof.

STEP 1 — We will show that $\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))$ is a Mercer kernel.

$\boldsymbol{\nabla}^2_{\boldsymbol{a}} k\,(\boldsymbol{a},\boldsymbol{b}) \in L^2(T \times T)$ holds. This follows form the fact that $\boldsymbol{\nabla}^2_{\boldsymbol{a}} k\,(\boldsymbol{a},\boldsymbol{b})$ is continuous according to the assumption that k is two times continuously differentiable and the fact that T is compact. Especially the kernel $\boldsymbol{\nabla}^2_{\boldsymbol{a}} k\,(\boldsymbol{a},\boldsymbol{b})$ takes its maximum on $T \times T$: $max_{\boldsymbol{a},\boldsymbol{b} \in T}\ |\boldsymbol{\nabla}^2_{\boldsymbol{a}}(k(\boldsymbol{a},\boldsymbol{b}))| = M$.

According to Theorem **??**:

$$k\,(\boldsymbol{a},\boldsymbol{b}) = \int_T k_d\,(\boldsymbol{a},\boldsymbol{c})\, k_d\,(\boldsymbol{b},\boldsymbol{c})\, d\boldsymbol{c} \ .$$

The symmetry of forces gives

$$\begin{aligned} \boldsymbol{\nabla}^2_{\boldsymbol{a}} k\,(\boldsymbol{a},\boldsymbol{b}) \;&=\; \boldsymbol{\nabla}_{\boldsymbol{a}} \cdot \boldsymbol{\nabla}_{\boldsymbol{a}} k\,(\boldsymbol{a},\boldsymbol{b}) \;=\; -\boldsymbol{\nabla}_{\boldsymbol{a}} \cdot \boldsymbol{\nabla}_{\boldsymbol{b}} k\,(\boldsymbol{a},\boldsymbol{b}) \;= \\ & -\int_T \boldsymbol{\nabla}_{\boldsymbol{a}} \cdot (k_d\,(\boldsymbol{a},\boldsymbol{c})\ \ \boldsymbol{\nabla}_{\boldsymbol{b}} k_d\,(\boldsymbol{b},\boldsymbol{c}))\, d\boldsymbol{c} \;= \\ & -\int_T \boldsymbol{\nabla}_{\boldsymbol{a}} k_d\,(\boldsymbol{a},\boldsymbol{c}) \cdot \boldsymbol{\nabla}_{\boldsymbol{b}} k_d\,(\boldsymbol{b},\boldsymbol{c})\, d\boldsymbol{c} \ , \end{aligned}$$

where "$\cdot$" denotes the dot product. The last equation is valid because it is the divergence operater with respect to $\boldsymbol{a}$ applied to the product of a scalar depending on $\boldsymbol{a}$ and a vector independent of $\boldsymbol{a}$.

$\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))$ induces a Hilbert-Schmidt operator and for all $\rho \in L^2(T)$ we get

$$\begin{aligned}(*)\text{: } 0 \;\; \leq \;\; & \int_T \int_T \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ \rho(\boldsymbol{b})\ \rho(\boldsymbol{a})\ d\boldsymbol{b}\ d\boldsymbol{a} \ = \\ & \int_T \rho(\boldsymbol{a}) \ \left(\int_T \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ \rho(\boldsymbol{b})\ d\boldsymbol{b} \right) d\boldsymbol{a} \ = \\ & \int_T \left(\int_T \rho(\boldsymbol{a}) \boldsymbol{\nabla}_{\boldsymbol{a}} k_d\,(\boldsymbol{a},\boldsymbol{c})\ d\boldsymbol{a} \right)^2 \ d\boldsymbol{c}\ .\end{aligned}$$

(*) allows to apply Mercers theorem to the kernel $\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))$: the following sum converges absolutely and uniformly:

$$\forall \boldsymbol{a},\boldsymbol{b} \in T:\ -\ \boldsymbol{\nabla}^2_{\boldsymbol{a}} k(\boldsymbol{a},\boldsymbol{b}) \ = \ \sum_{n=1}^{\infty} \lambda_n\ e_n(\boldsymbol{a})\ e_n(\boldsymbol{b})$$

and $\forall n:\ \lambda_n \ \geq \ 0$. The $e_1, e_2, \ldots$ are an orthonormal eigenfunction system. Further $L^2(T) \ = \ \overline{\operatorname{span}\{e_n \mid 1 \leq n\}}$, where $\overline{A}$ denotes the closure of a set A.

STEP 2 – For a given $\boldsymbol{a}$ we assume that $\lambda_i \neq \lambda_j$ and $e_i(\boldsymbol{a})\ e_j(\boldsymbol{a})\ \neq 0$. From this assumption we will deduce that $0 < m \leq \int_{T\setminus\{\boldsymbol{a}\}} \left(\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\right)^2\ d\boldsymbol{b}$.

Let $\boldsymbol{a}$ be given. Let us assume that $\lambda_i \neq \lambda_j$ and $e_i(\boldsymbol{a}) \neq 0$ and $e_j(\boldsymbol{a}) \neq 0$, i.e., $e_i(\boldsymbol{a})\ e_j(\boldsymbol{a})\ \neq 0$ holds.

We define

$$\hat{\rho}(\boldsymbol{b}) \ = \ e_i(\boldsymbol{b}) \ + \ \left(-\frac{e_i(\boldsymbol{a})}{e_j(\boldsymbol{a})} \right) \ e_j(\boldsymbol{b})\ .$$

We constructed $\hat{\rho}(\boldsymbol{b})$ so that $\hat{\rho}(\boldsymbol{a}) \ = \ 0$.

$$\int_T \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ \hat{\rho}(\boldsymbol{b})\ d\boldsymbol{b} \ = \ (\lambda_i - \lambda_j)\ e_i(\boldsymbol{a}) \ \neq \ 0\ .$$

We found that $\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b})) \ \neq \ 0$ on a mesurable nonzero set which does not contain $\boldsymbol{a}$. Therefore we have

$$0 < m \leq \int_{T\setminus\{\boldsymbol{a}\}} \left(\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\right)^2\ d\boldsymbol{b}\ .$$

STEP 3 – Assumption of step 2. We will construct function $\tilde{\rho}$ which violates $\operatorname{sign}(\boldsymbol{\nabla}\cdot(\mathbf{E}(\boldsymbol{a}))) \ = \ \operatorname{sign}\left(\int_T \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ \tilde{\rho}(\boldsymbol{b})\ d\boldsymbol{b}\right) \ = \ \operatorname{sign}(\tilde{\rho}(\boldsymbol{a}))$ and which as is absolute maximum value at $\boldsymbol{a}$.

We use the assumptions of step 2.

From step 1 we know that $\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{a}))$ is bounded, i.e., $\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{a})) \leq M$

We define for $T \in \mathbb{R}^d$:

$$\delta_n(\boldsymbol{a}-\boldsymbol{b}) \ := \ \begin{cases} n\left(1 \ - \ 2\ \left(\frac{n}{d+1}\right)^{\frac{1}{d}}\ \max_i |a_i \ - \ b_i|\right) & \forall\ \boldsymbol{b} \in \mathcal{B}_n(\boldsymbol{a}) \\ 0 & \text{otherwise} \end{cases} ,$$

where

$$\mathcal{B}_n(\boldsymbol{a}) \ := \ \left\{ \boldsymbol{b} \mid \forall_i :\ a_i \ - \ 0.5\ \left(\frac{d+1}{n}\right)^{\frac{1}{d}} \ \leq \ b_i \ \leq \ a_i \ + \ 0.5\ \left(\frac{d+1}{n}\right)^{\frac{1}{d}} \right\}$$

the d-dimensional hypercube with edge length $\left(\frac{d+1}{n}\right)^{\frac{1}{d}}$ and center $\boldsymbol{a}$. $(\boldsymbol{b}, \delta_n(\boldsymbol{a}-\boldsymbol{b}))$ is the surface of $(d+1)$-dimensional hyperpyramid with base $\mathcal{B}_n(\boldsymbol{a})$ and altitude n, which is reached at a the single peak at $\boldsymbol{a}$.

We ensured that $\int_T \delta_n(\boldsymbol{a}-\boldsymbol{b})d\boldsymbol{b} \ = \ 1$:

$$\begin{aligned} \int_T \delta_n(\boldsymbol{a}-\boldsymbol{b})d\boldsymbol{b} \ \ &= \ \ \int_{\mathcal{B}_n(\boldsymbol{a})} n \left(1 \ - \ 2\ \left(\frac{n}{d+1}\right)^{\frac{1}{d}} \max_i |a_i \ - \ b_i|\right)\ d\boldsymbol{b} \ = \\ & \left(\left(\frac{d+1}{n}\right)^{\frac{1}{d}}\right)^d\ n \int_{\mathcal{B}_{d+1}(\boldsymbol{0})} \left(1 \ - \ 2\ \max_i |c_i|\right)\ d\boldsymbol{c} \ = \\ & (d+1)\ \left(1 \ - \ 2^{d+1}\ d \int_0^{0.5} c_1 \int_0^{c_1} dc_2 \ldots \int_0^{c_1} dc_d\ dc_1\right) \ = \\ & (d+1)\ \left(1 \ - \ 2^{d+1}\ d \int_0^{0.5} c_1^d dc_1\right) \ = \\ & (d+1)\ \left(1 \ - \ \frac{d}{d+1}\right) \ = \ 1 \ . \end{aligned}$$

We defined $c_i \ := \ \left(\frac{n}{d+1}\right)^{\frac{1}{d}} (a_i \ - \ b_i)$ and $\mathcal{B}_{d+1}(\boldsymbol{0})$ as the d-dimensional hypercube with edge length 1 and center $\boldsymbol{0}$. The factor 2^d resulted from replacing the d integrals $\int_{-0.5}^{0.5}$ by $2\ \int_0^{0.5}$. The factor d stems from the fact that only one of the d components of $\boldsymbol{c}$ is the maximum, therefore, we only considered the case where c_1 is the maximum and multiplied this case by d.

The function δ_n allows us to construct

$$\tilde{\rho}(\boldsymbol{b}) \ := \ \begin{cases} \delta_n(\boldsymbol{a}-\boldsymbol{b}) & \text{for } \boldsymbol{b} \in \mathcal{B}_n(\boldsymbol{a}) \\ -\ \left(\frac{\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{a}))}{\int_{T \setminus \mathcal{B}_n(\boldsymbol{a})} (\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b})))^2\ d\boldsymbol{b}} \ + \ \epsilon\right) \left(\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\right) & \text{otherwise} \end{cases} .$$

(I): for large enough n the following holds:

$$\text{argmax}_{\boldsymbol{b} \in T}\ |\tilde{\rho}(\boldsymbol{b})| \ = \ \boldsymbol{a}$$

and

$$\max_{\boldsymbol{b}\in T}\ |\tilde{\rho}(\boldsymbol{b})|\ =\ n\ >\ 0\ .$$

$\boldsymbol{a}$ is the global maximum of $|\tilde{\rho}|$.

(II):

$$\begin{aligned}
&\boldsymbol{\nabla}\cdot(\mathbf{E}(\boldsymbol{a}))\ =\\
&\int_T \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ \tilde{\rho}(\boldsymbol{b})\ d\boldsymbol{b}\ =\\
&\int_{B_n(\boldsymbol{a})} \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ \delta_n(\boldsymbol{a}-\boldsymbol{b})\ d\boldsymbol{b}\ -\\
&\left(\frac{\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{a}))}{\int_{T\setminus\mathcal{B}_n(\boldsymbol{a})}\left(\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\right)^2\ d\boldsymbol{b}}\ +\ \epsilon\right)\left(\int_{T\setminus\mathcal{B}_n(\boldsymbol{a})}\left(\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\right)^2\ d\boldsymbol{b}\right)\ =\\
&\int_{B_n(\boldsymbol{a})} \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ \delta_n(\boldsymbol{a}-\boldsymbol{b})\ d\boldsymbol{b}\ -\ \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{a}))\ -\\
&\epsilon\int_{T\setminus\mathcal{B}_n(\boldsymbol{a})}\left(\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\right)^2\ d\boldsymbol{b}
\end{aligned}$$

For

$$\epsilon\ >\ q_n\ =\ \frac{\int_{B_n(\boldsymbol{a})} \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ \delta_n(\boldsymbol{a}-\boldsymbol{b})\ d\boldsymbol{b}\ -\ \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{a}))}{\int_{T\setminus\mathcal{B}_n(\boldsymbol{a})}\left(\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\right)^2}$$

we get

$$\boldsymbol{\nabla}\cdot(\mathbf{E}(\boldsymbol{a}))\ =\ \int_T \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ \tilde{\rho}(\boldsymbol{b})\ d\boldsymbol{b}\ < 0\ .$$

The later inequality can be ensured for large enough n because $\lim_{n\to\infty} q_n\ =\ 0$. In order to prove thsi limit we show two facts. Firstly, we need that (a) the denominator is bounded from below. Secondly, we show that (b) the numerator goes to zero with increasing n.

Ad (a): we deduce from step 2 that even if we subtract a small enough neighborhood $\mathcal{B}_n(\boldsymbol{a})$ of $\boldsymbol{a}$ from T we obtain

$$0 < m_n \leq \int_{T\setminus\mathcal{B}_n(\boldsymbol{a})}\left(\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\right)^2\ d\boldsymbol{b}\ .$$

With increasing n the neighborhood $\mathcal{B}_n(\boldsymbol{a})$ contracts around $\boldsymbol{a}$ and m_n increases, i.e. there exists a m_{n_0} so that $n > n_0$ implies $m_n > m_{n_0}$.

Ad (b): the following limit holds:

$$\lim_{n\to\infty}\int_{B_n(\boldsymbol{a})} \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ \delta_n(\boldsymbol{a}-\boldsymbol{b})\ d\boldsymbol{b}\ =\ \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{a}))\ .$$

This limit follows from

$$\left| \int_{B_n(\boldsymbol{a})} \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ \delta_n(\boldsymbol{a}-\boldsymbol{b})\ d\boldsymbol{b}\ -\ \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{a})) \right| \ =$$

$$\left| \int_{B_n(\boldsymbol{a})} \big(\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ -\ \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{a}))\big)\ \delta_n(\boldsymbol{a}-\boldsymbol{b})\ d\boldsymbol{b} \right| \ \leq$$

$$\int_{B_n(\boldsymbol{a})} \big|\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))\ -\ \boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{a}))\big|\ \delta_n(\boldsymbol{a}-\boldsymbol{b})\ d\boldsymbol{b}\ \leq$$

$$\int_{B_n(\boldsymbol{a})} \tau_n(\boldsymbol{a})\, \|\boldsymbol{a}\ -\ \boldsymbol{b}\|\ \delta_n(\boldsymbol{a}-\boldsymbol{b})\ d\boldsymbol{b}\ \leq$$

$$\tau_n(\boldsymbol{a})\ \int_{B_n(\boldsymbol{a})} \left(\frac{d+1}{n}\right)^{\frac{1}{d}}\ \delta_n(\boldsymbol{a}-\boldsymbol{b})\ d\boldsymbol{b}\ =\ \tau_n(\boldsymbol{a})\ \left(\frac{d+1}{n}\right)^{\frac{1}{d}}\ .$$

The factor $\tau_n(\boldsymbol{a})$ exists because $\boldsymbol{\nabla}^2_{\boldsymbol{a}}(-\ k(\boldsymbol{a},\boldsymbol{b}))$ is continuous (see step 1). Both factors $\tau_n(\boldsymbol{a})$ and $\big(\frac{d+1}{n}\big)^{\frac{1}{d}}$ vanish with increasing n.

Therefore (I) and (II) contradict the Lemma “Maximum point requirement”. The assumption in step 2 must be false which leads to following conclusion.

STEP 4 – Conclusion.

We conclude that $\lambda_i \neq \lambda_j$ implies $\forall_{\boldsymbol{a}\in T}:\ e_i(\boldsymbol{a})\ e_j(\boldsymbol{a})\ =\ 0$.

We define for $\lambda_l > 0$

$$U(\lambda_l)\ :=\ \{\boldsymbol{a}\ \mid\ \exists\, e_i:\ \ \lambda_i\ =\ \lambda_l\ \neq\ 0\ \wedge\ e_i(\boldsymbol{a})\ \neq\ 0\}$$

and for $\lambda_l = 0$

$$U(0)\ :=\ \{\boldsymbol{a}\ \mid\ e_i(\boldsymbol{a})\ \neq\ 0\ \Rightarrow\ \lambda_i\ =\ 0\}\ .$$

The $U(\lambda_l)$ are equivalence classes which partition T.

We will prove the last statement. If $\boldsymbol{a}\ \in\ U(\lambda_l)\ \cap\ U(\lambda_k)$ for $\lambda_l\ \neq\ \lambda_k$ and $\lambda_l, \lambda_k \neq 0$ then $e_i(\boldsymbol{a})\ \neq\ 0$ and $e_j(\boldsymbol{a})\ \neq\ 0$ exist. The beginning of step 4 states that $\lambda_l\ =\ \lambda_i\ =\ \lambda_j\ =\ \lambda_k$ in contradiction to $\lambda_l\ \neq\ \lambda_k$. We treat $U(0)$ next. If $\boldsymbol{a} \in U(0)$ then $\forall_i : e_i(\boldsymbol{a})\ \neq\ 0\ \Rightarrow\ \lambda_i\ =\ 0$ which implies $\boldsymbol{a} \notin U(\lambda_i)$ with $\lambda_i \neq 0$. Hence $U(\lambda_l)\ \cap\ U(\lambda_k)\ =\ \emptyset$.

The constant functions around $\boldsymbol{a}$ with finit support are from $L^2(T)$ and, therefore, have representations through the orthonormal system e_i. This implies that at least for one i inequality $e_i(\boldsymbol{a}) \neq 0$ holds, so that $\boldsymbol{a} \in U(\lambda_i)$. Therefore $\bigcup_l U(\lambda_l)\ =\ T$. This finishes the proof that the $U(\lambda_l)$ are a partition of T.

Next we show that k is zero for arguments from different $U(\lambda_l)$. If $k(\boldsymbol{a},\boldsymbol{b})\ \neq\ 0$ then an e_i exists with $e_i(\boldsymbol{a})\ \neq\ 0$ and $e_i(\boldsymbol{b})\ \neq\ 0$, thus, an l exists with $\boldsymbol{a},\boldsymbol{b}\ \in\ U(\lambda_l)$. Let be $\partial U(\lambda_l)$ the frontier of $U(\lambda_l)$. We see that $\partial U(\lambda_l)\ \cap\ \partial U(\lambda_k)\ \subset\ U(0)$.

We found that the kernel k is a composition of kernels on the $U(\lambda_l)$. This composition also applies to all other functions from $L^2(T)$ (beginning of step 4: $\lambda_i \neq \lambda_j$ implies $\forall_{\boldsymbol{a} \in T} : \; e_i(\boldsymbol{a}) \; e_j(\boldsymbol{a}) \; = \; 0$).

We deduce that $L^2(U(\lambda_l)) \; = \; \overline{\text{span}\{e_i \mid \lambda_i \; = \; \lambda_l\}}$. Let $k_{U(\lambda_l)}$ be k restricted to $U(\lambda_l)$ then we obtain for every $\boldsymbol{b} \in U(\lambda_l)$:

$$\boldsymbol{\nabla}^2_{\boldsymbol{a}}(- \; k_{U(\lambda_l)}(\boldsymbol{a}, \boldsymbol{b})) \; = \; \lambda_l \sum_{n : \lambda_n = \lambda_l} e_n(\boldsymbol{a}) \; e_n(\boldsymbol{b}) \; = \; \lambda_l \; \delta_{U(\lambda_l)}(\boldsymbol{a} - \boldsymbol{b}) \; ,$$

where $\delta_{U(\lambda_l)}$ is the delta function on $U(\lambda_l)$.

■