HAIT Journal of Science and Engineering C, Volume 4, Issues 1-2, pp. 194-202


# Optimal feature space for representation of finite dimensional data

Inna Kozlov[1,*] and Vladimir Ovchinnikov[2,3]

[1] *Holon Institute of Technology, 52 Golomb St., Holon 58102, Israel*
[2] *Mathematical Institute, Academy of Science of the Czech Republic, Zitna 25, CZ-115 67 Praha 1, Czech Republic*
[3] *Mathematical Department, Voronezh State University, Universitetskaia pl. 1, Voronezh 394006, Russia*
[*] *Corresponding author: innak@hait.ac.il*

Received 5 April 2005, accepted 5 December 2005

**Abstract**

The paper develops a recent Coifman's approach to the problems of Signal and Image Processing, based on spectral theory of operators in Hilbert spaces. New notion of *canvas* is introduced. This allows to use powerful tools of Operator Ideal Theory which, in turn, leads to a numerical algorithm for restoring the original image.

## 1 Introduction

Recently R. Coifman, see in particular, [1], proposed a new approach to reducing dimension of data appeared in image processing. He used for this goal a new function space generated by a few geometric harmonics, i.e., eigenfunctions of some positive integral operator acting in an infinite-dimensional Hilbert space. These eigenfunctions can be found as extreme points of the corresponding quadratic form on the Hilbert space. In the present paper we suggest a modification of the above Coifman idea in order to find the function spaces for data representation with optimal, in some sense, dimension. We make this by solving a series of extreme points problems which increase the require dimension step to step. To avoid technical complications we consider here only a simple model related to the space of functions continuous on $\mathbb{R}^n$.

## 2 A feature space or canvas

**Definition 1.** *A Hilbert space $H$ embedded into the space $C(\mathbb{R}^n)$ of bounded continuous functions on $\mathbb{R}^n$ is called a canvas.*

Every canvas generates a mapping from $\mathbb{R}^n$ to the space $H$ as follows. Each $q \in \mathbb{R}^n$ generates the valuation functional $\delta_q \in C(\mathbb{R}^n)^*$ whose restriction to $H$ can be represented by the F.Riesz theorem as

$$f(q) = \delta_q(f) = \langle f, a(q) \rangle_H, \quad f \in H,$$

here $a : \mathbb{R}^n \mapsto H$ is a weak continuous bounded vector function, which gives the required mapping. Conversely, every such $a$ generates some canvas; namely, the required imbedding of $H$ to $C(\mathbb{R}^n)$ is defined by sending an element $h \in H$ to the continuous function $q \mapsto \langle h, a(q) \rangle_H$. Hence, $a(q)$ may be seen as $\delta$-function on the space $H$.

**Examples of canvases.**

$1^o$. The most natural example of a canvas is the Sobolev space $W_2^s(\mathbb{R}^n)$ with the smoothness $s > n/2$, since $W_2^s(\mathbb{R}^n) \subset C(\mathbb{R}^n)$.

$2^o$. R. Coifman [1] considers the space of bandlimited functions $L_2^\delta(\mathbb{R}^n)$ consisting of square integrable functions on $\mathbb{R}^n$, whose Fourier transforms are supported by the ball of radius $\delta$. Clearly, this space is a canvas.

$3^o$. A finite dimensional subspace in $C(\mathbb{R}^n)$ can be regarded as a canvas. For this end it is necessary to fix some scalar product on such a space.

$4^o$. The following example is also due to R. Coifman. Let $K(s,t)$ be positive definite bounded continuous function on $\mathbb{R}^n \times \mathbb{R}^n$. Consider the linear span $H_K$ of the family of functions $\{K(\cdot, t) : t \in \mathbb{R}^n\}$. This can be related to a bilinear form by the following formula.

If $x(q) = \sum_{i=1}^n \lambda_i K(q, s_i)$ and $y(q) = \sum_{j=1}^m \mu_j K(q, t_j)$ then

$$\langle x, y \rangle = \langle \sum_{i=1}^n \lambda_i K(\cdot, s_i), \sum_{j=1}^m \mu_j K(\cdot, t_j) \rangle$$

$$= \sum_{i,j=1}^{n,m} \lambda_i \mu_j K(s_i, t_j).$$

It is easily seen that this definition is well-defined and that $\langle \cdot, \cdot \rangle_H$ is a semidefinite.

Let us show that actually we obtain a scalar product on $H_K$ and that $H_K$ is continuously imbedded into $C(\mathbb{R}^n)$.

Consider a vector-valued function $a : \mathbb{R}^n \to H_K$ defined by

$$a(q)(\cdot) = K(q, \cdot),$$

where $q \in \mathbb{R}^n$. By the definition, $\|a(q)\|^2_{H_K} = \langle K(q, \cdot), K(q, \cdot) \rangle = K(q, q) \leq C$. The space $H_K$ is embedded into $C(\mathbb{R}^n)$ by $x \mapsto \langle x, a(q) \rangle$ with the estimate

$$\|\langle x, a(\cdot) \rangle\|_{C(\mathbb{R}^n)} \leq C\|x\|_{H_K}.$$

which follows from the Schwartz inequality. Moreover, the map $x \mapsto \langle x, a(q) \rangle$ is the identity. Indeed,

$$\langle x, a(q) \rangle = \langle \sum_{i=1}^{n} \lambda_i K(\cdot, s_i), K(\cdot, q) \rangle = \sum_{i=1}^{n} \lambda_i K(q, s_i) = x(q).$$

Hence, this bilinear form is not degenerate and actually defines the scalar product on $H_K$. We denote the completion of $H_K$ with respect to the corresponding norm by the same symbol $H_K$.

The example $4^o$ is actually universal because any continuous bounded vector-valued function $a : \mathbb{R}^n \to H$ determines the function $K(s, t) = \langle a(s), a(t) \rangle$ which is a positive definite continuous and bounded on $\mathbb{R}^n \times \mathbb{R}^n$. The corresponding space $H_K$ is isometrically isomorphic to the linear span of the family of vectors $\{a(q)\}$ in $H$, and corresponding vectors are identified with the same functions of $C(\mathbb{R}^n)$ as follows. Since

$$\|\sum_{i=1}^{n} \lambda_i a(s_i)\|^2_H = \sum_{i,j=1}^{n} \lambda_i \lambda_j \langle a(s_i), a(s_j) \rangle$$

$$= \sum_{i,j=1}^{n} \lambda_i \lambda_j K(s_i, s_j)$$

$$= \langle \sum_{i=1}^{n} \lambda_i K(\cdot, s_i), \sum_{j=1}^{n} \lambda_i K(\cdot, s_j) \rangle = \|\sum_{i=1}^{n} \lambda_i K(\cdot, s_i)\|^2_{H_K},$$

we identify $\sum_{i=1}^{n} \lambda_i a(s_i)$ with $\sum_{i=1}^{n} \lambda_i K(\cdot, s_i)$.

## 3 Representation of data

We present the data as the Radon measure $\mu$ on the set $\Gamma$.

Let $\Gamma$ be a compact set in $\mathbb{R}^n$ and $\mu$ be a finite Radon measure on $\Gamma$. As usual we denote by $L_2(\mu)$ the Hilbert space of measurable functions on $\Gamma$ such that

$$\int_\Gamma |f(q)|^2 d\mu(q) < \infty.$$

Let $H$ be a canvas. The canvas $H$ is naturally mapped into $L_2(\mu)$ through $L_\infty(\mu)$. We denote this map by $I_\mu : H \to L_2(\mu)$. If we identify the elements of $H$ with the corresponding functions from $C(\mathbb{R}^n)$, then $I_\mu$ is the operator of restriction of functions to $\Gamma$.

Thus we have

$$I_\mu : H \to L_\infty(\mu) \subset L_2(\mu) \tag{1}$$

The embedding (1) implies that $I_\mu$ is a Hilbert–Schmidt operator (see [3]).

The adjoint operator

$$I_\mu^* : L_2(\mu) = L_2(\mu)^* \subset L_\infty(\mu)^* \to H^* = H$$

is also a Hilbert–Schmidt operator. Hence the product

$$H_\mu = I_\mu I_\mu^* : L_2(\mu) \subset L_\infty(\mu)^* \to H \subset L_\infty(\mu) \to L_2(\mu)$$

is a positive trace-class operator in $L_2(\mu)$.

Denote by $G_\mu$ the product

$$G_\mu = I_\mu^* I_\mu : H \to L_\infty(\mu) \subset L_2(\mu) \subset L_\infty(\mu)^* \to H,$$

which is also a trace-class operator. Note that if $I_\mu$ is an embedding, then Ker $G_\mu = 0$ .

Recall that any canvas is generated by a vector-valued function $a(q)$ or a positive definite kernel $K(s,t) = \langle a(s), a(t)\rangle = a(s)(t)$. Our aim now is to present the operators $H_\mu$ and $G_\mu$ in terms of these functions.

By definition $I_\mu$ takes $h$ to $\langle h, a(q)\rangle$ in $L_2(\mu)$, hence $I_\mu^*$ takes each function $f(q) \in L_2(\mu)$ to the element which generates the functional

$$\alpha(h) = \int_\Gamma \langle h, a(q)\rangle f(q) d\mu.$$

Hence

$$I_\mu^*(f) = \int_\Gamma f(q)a(q)d\mu(q) \qquad \text{(weak convergence).}$$

Thereby

$$H_\mu(f)(t) = I_\mu I_\mu^*(f) = \langle \int_\Gamma f(s)a(s)d\mu(s), a(t)\rangle \tag{2}$$

$$= \int_\Gamma \langle a(s), a(t)\rangle f(s)d\mu(s) = \int_\Gamma K(s,t)f(s)d\mu(s).$$

From the other side

$$\langle G_\mu(h), g\rangle = \langle I_\mu^* I_\mu(h), g\rangle = \langle I_\mu(h), I_\mu(g)\rangle_{L_2(\mu)}$$

$$= \int_\Gamma \langle a(s), h\rangle\langle a(s), g\rangle d\mu(s)$$

$$= \int_\Gamma \langle (a(s)\otimes a(s))h, g\rangle d\mu(s),$$

where $a \otimes b$ denotes the rank one operator in the space $H$ defined by $(a \otimes b)(h) = \langle h, a\rangle b$.

In other words

$$G_\mu = \int_\Gamma a(s)\otimes a(s)d\mu(s) \qquad \text{(weak convergence),} \tag{3}$$

i.e, $G_\mu$ is an integral of rank one operators.

Since $h(s) = \langle h, a(s)\rangle$ and $g(s) = \langle g, a(s)\rangle$ we also have

$$\langle G_\mu(h), g\rangle = \int_\Gamma h(s)\overline{g(s)}d\mu(s). \tag{4}$$

It is well known that $H_\mu = I_\mu I_\mu^*$ and $G_\mu = I_\mu^* I_\mu$ are similar or metrically equivalent (non-zero eigenvalues of $H_\mu$ and $G_\mu$ are equal and have the same multiplicity).

Both these operators may be used for estimation of the quality of representation of the data given by the measure $\mu$ on the canvas $H$. However if we want to compare two representations on two different canvases $H'$ and $H''$, we compare the operators $H'_\mu$ and $H''_\mu$. If we want to compare representations of two different data $\mu'$ and $\mu''$, we use $G_{\mu'}$ and $G_{\mu''}$.

## 4 Geometric harmonics

The eigenvectors of the operator $G_\mu$, corresponding to non-zero eigenvalues, were called by R. Coifman by geometric harmonics. Denote by $\lambda_j$ the eigenvalues of the operator $G_\mu$ indexed by the standard manner such that $\lambda_1 \geq \cdots \geq \lambda_j \geq \lambda_{j+1} \geq \ldots$ including multiplicity. The corresponding orthogonal unit eigenvectors will be denoted by $\psi_j$. Recall that $\psi_j$ are functions from $C(\mathbb{R}^n)$ since $H \subset C(\mathbb{R}^n)$. Since $G_\mu$ is a trace class operator, we have

$$\sum_{j=1}^{\infty} \lambda_j = \mathrm{tr}(G_\mu) = \int_\Gamma \mathrm{tr}(a(s) \otimes a(s)) d\mu(s) = \int_\Gamma K(s,s) d\mu(s) < \infty$$

because of (3).

The operator $H_\mu$ has the same non-zero eigenvalues $\lambda_j$ and the corresponding orthogonal unit eigenvectors will be denoted by $\varphi_j \in L_2(\mu)$.

Consider the Schmidt expansion of $I_\mu$

$$I_\mu = \sum_{j=1}^{\mathrm{rank}\ I_\mu} \lambda_j^{1/2} \langle \cdot, \psi_j \rangle \varphi_j, \tag{5}$$

where $\psi_j$ is an orthogonal unit sequence in $H$, and $\varphi_j$ is orthogonal unit sequence in $L_2(\mu)$ (see [2]) .

Hence $I_\mu(\psi_j) = \lambda_j^{1/2} \varphi_j$, which means that the restriction of $\psi_j$ on $\Gamma$ is multiple of $\varphi_j$. Thus $\psi_j / \lambda_j$ may be considered as an extension of $\varphi_j$ from $\Gamma$ to $\mathbb{R}^n$.

Since the Schmidt expansion for the adjoint operator $I_\mu^*$ has a form

$$I_\mu^* = \sum_{j=1}^{\mathrm{rank}\ I_\mu} \lambda_j^{1/2} \langle \cdot, \varphi_j \rangle_{L_2(\mu)} \psi_j, \tag{6}$$

we have $I_\mu^*(\varphi_j) = \lambda_j^{1/2} \psi_j$. Hence

$$\psi_j = \frac{1}{\lambda_j^{1/2}} I_\mu^*(\varphi_j) = \frac{1}{\lambda_j^{1/2}} \int_\Gamma \varphi_j(s) a(s) d\mu(s),$$

or

$$\psi_j(q) = \frac{1}{\lambda_j^{1/2}} \int_\Gamma \varphi_j(s) a(s)(q) d\mu(s) = \frac{1}{\lambda_j^{1/2}} \int_\Gamma \varphi_j(s) \langle a(s), a(q) \rangle d\mu(s)$$

$$= \frac{1}{\lambda_j^{1/2}} \int\limits_{\Gamma} \varphi_j(s) K(s,q) d\mu(s),$$

which gives the formula for the extension of $\varphi_j(s)$.

# 5 Reduction of the canvas dimension

In this section we turn to the problem of low dimension representation of the data. Given a fixed representation on the canvas $H$ we try to find a new canvas $F$ of the smallest dimension such that distortion between the new representation and the old one doesn't exceed fixed number $\varepsilon > 0$. We are going to measure the distance between two representations as the operator norm of the difference $H_\mu - F_\mu$.

**Theorem 1.** *Let $H$ be an arbitrary canvas and $\varepsilon > 0$. The linear span $F$ of the first geometric harmonics $\psi_j$, corresponding to the eigenvalues $\lambda_j > \varepsilon$, is the canvas of the smallest dimension such that*

$$\|H_\mu - F_\mu\|_{L(L_2(\mu))} \le \varepsilon.$$

*Proof.* Denote by $n_\varepsilon$ the smallest $j$ such that $\lambda_{n_\varepsilon+1} \le \varepsilon$. Then $\lambda_1 \ge \cdots \lambda_{n_\varepsilon} > \varepsilon$ and $F$ is the linear span of $\psi_1, \cdots, \psi_{n_\varepsilon}$. Denote by $a_F(q)$ representing function, corresponding to the canvas $F$.

We evidently have

$$a_F(q) = \sum_{j=1}^{n_\varepsilon} \langle a_F(q), \psi_j \rangle \psi_j,$$

and since $\langle \psi_j, a_F(q) \rangle = \psi_j(q)$ we conclude

$$a_F(q)(s) = \sum_{j=1}^{n_\varepsilon} \langle a_F(q), \psi_j \rangle \psi_j(s) = \sum_{j=1}^{n_\varepsilon} \psi_j(q) \psi_j(s).$$

Thus the kernel, corresponding to the canvas $F$, denoted by $K_F(s,t)$ is equal to

$$\sum_{j=1}^{n_\varepsilon} \psi(t) \psi_j(s),$$

and

$$F_\mu(f)(q) = \int\limits_{\Gamma} \sum_{j=1}^{n_\varepsilon} \psi_j(q) \psi_j(s) f(s) d\mu(s),$$

because of (1).

Since $\lambda_j^{1/2}\psi_j|_\Gamma = \varphi_j$ we obtain

$$F_\mu(f)(q) = \int\limits_\Gamma \sum_{j=1}^{n_\varepsilon} \lambda_j \varphi_j(q)\varphi_j(s) f(s) d\mu(s).$$

Therefore

$$[H_\mu - F_\mu](f)(q) = \int\limits_\Gamma \sum_{j=n_\varepsilon+1}^{\text{rank } I_\mu} \lambda_j \varphi_j(q)\varphi_j(s) f(s) d\mu(s),$$

and we conclude that

$$\|H_\mu - F_\mu\|_{L(L_2(\mu))} \le \varepsilon.$$

Recall now that eigenvalues of the operator $H_\mu$ satisfy to the approximation property due to G.Allahverdiev (see [2])

$$\lambda_{j+1} = \min \|H_\mu - K\|, \tag{7}$$

where minimum is taken over all bounded linear operators $K$ with the rank less than or equal to $j$.

Let now $\widetilde{F}$ be a canvas of dimension $n$ such that $\|H_\mu - \widetilde{F}_\mu\| \le \varepsilon$. The rank of $\widetilde{F}_\mu$ is less than or equal to $n$, therefore (7) yields $\lambda_{n+1} \le \varepsilon$. Hence by our definition of $n_\varepsilon$ we obtain $n \ge n_\varepsilon$. Thus $n_\varepsilon$ is the smallest dimension for representation of the measure $\mu$ with the distortion $\varepsilon$. Theorem is proved.

Thus we see that the problem of finding an optimal finite dimensional representation is reduced to the problem of finding the geometric harmonics and the corresponding eigenvalues of the operator $H_\mu$.

If we have opportunity to find the norm of a positive compact operator and the corresponding extreme point we can find the optimal canvas $F$ such that $\|H_\mu - F_\mu\| \le \varepsilon$ for any $\varepsilon > 0$ by the following procedure.

Step 1. Set $F = 0$.

Step 2. Find

$$m := \max_{f \in H \ominus F,\ \|f\|=1} \langle G_\mu f, f\rangle = \|f\|^2_{L_2(\mu)}$$

and a maximum point $g \in H \ominus F$.

Step 3. If $m \le \varepsilon$, then STOP. If $m > \varepsilon$, then $F := F \oplus \{\lambda g\}$.

Step 4. GOTO Step 1.

The final space $F$ is generated by the first harmonics $\psi_1, \cdots \psi_{n_\varepsilon}$, since $m$ consequently gives us $\lambda_j$, and $g$ gives us the corresponding $\psi_j$.

If we have an opportunity for any given $\varepsilon$ to find a unit vector $g$ such that $\|Ag\| > \varepsilon$, or to show that

$$\max_{f \in H,\ \|f\|=1} \|Af\| \leq \varepsilon,$$

then we can find the dimension of the optimal canvas with the help of the following algorithm.

Again we denote by $F$ a finite dimensional subspace in $H$.

Step 1. Set $F = 0$.

Step 2. If we find unit vector $g \in H \ominus F$ such that

$$\|g\|_{L_2(\mu)} > \varepsilon,$$

then $F := F \oplus \{\lambda g\}$, if for all $f \in H \ominus F$, we have $\|f\|_{L_2(\mu)} \leq \varepsilon$, then STOP.

Step 3. Find

$$m := \min_{f \in G,\ \|f\|=1} \|f\|_{L_2(\mu)}$$

and a minimal point $g$.

Step 4. If $m > \varepsilon$ GOTO Step 1. If $m \leq \varepsilon$, then $F := F \ominus \{\lambda g\}$.

Step 5. GOTO Step 1.

Thus for the final space $F$ we have $\|x\| > \varepsilon$ if $\|x\| = 1,\ x \in F$, and $\|y\| \leq \varepsilon$ for all unit vectors $y$ from the orthogonal complement of $F$. The same property has the space generated by the first geometric harmonics. Therefore neither of these two spaces contains vectors orthogonal to the other space. Hence these spaces are isomorphic. Thus we find the dimension of the optimal canvas.

## References


[1] S. Lafon, *Diffusion Maps and Geometric Harmonics*, Ph.D. dissertation, Yale University (May 2004).

[2] I. Gohberg and M. Krein, *Introduction to the Theory of Linear Non-selfadjoint Operators* (Amer. Math. Soc., Providence, 1988).

[3] A. Grothendieck, *Produits Tensoriels Topologiques et Espaces Nucléaires*, Mem. Amer. Math. Soc. n.16 (1955).