C. R. Acad. Sci. Paris, Ser. I ••• (••••) •••–•••

Contents lists available at ScienceDirect

C. R. Acad. Sci. Paris, Ser. I

www.sciencedirect.com

Mathematical analysis/Theory of signals

Best bases for signal spaces

Bases optimales pour des espaces de signaux

Yonathan Aflalo[a], Haïm Brezis[c,b], Alfred Bruckstein[a], Ron Kimmel[a], Nir Sochen[d]

[a] *Computer Science Department, Technion – I.I.T., 32000 Haifa, Israel*
[b] *Department of Mathematics, Technion – I.I.T., 32000 Haifa, Israel*
[c] *Department of Mathematics, Rutgers University, USA*
[d] *Department of Applied Mathematics, Tel Aviv University, Tel Aviv 69978, Israel*

ARTICLE INFO

Article history:
Received 2 October 2016
Accepted 3 October 2016
Available online xxxx

Presented by Haïm Brézis

ABSTRACT

We discuss the topic of selecting optimal orthonormal bases for representing classes of signals defined either through statistics or via some deterministic characterizations, or combinations of the two. In all cases, the best bases result from spectral analysis of a Hermitian matrix that summarizes the prior information we have on the signals we want to represent, achieving optimal progressive approximations. We also provide uniqueness proofs for the discrete cases.

RÉSUMÉ

Dans cette note, nous abordons le problème de la recherche de bases orthonormales optimales en vue de représenter des signaux définis de façon, soit statistique, soit déterministe, ou selon une combinaison des deux. Dans tous les cas, nous montrons que ces bases proviennent de l'analyse spectrale d'une matrice hermitienne qui regroupe l'information émanant des signaux que l'on souhaite représenter. Nous prouvons aussi l'unicité de la base dans le cas discret.

1. Introduction

A fundamental problem in engineering, as well as in mathematics, is the progressive approximation of a signal via a sequence of linearly combined basis signals with optimal weighting coefficients. The setting is as follows. Let $\mathcal{B}$ denote the

E-mail addresses: johnaflalo@gmail.com (Y. Aflalo), brezis@math.rutgers.edu (H. Brezis), freddy@cs.technion.ac.il (A. Bruckstein), ron@cs.technion.ac.il (R. Kimmel), sochen@post.tau.ac.il (N. Sochen).

http://dx.doi.org/10.1016/j.crma.2016.10.002
1631-073X/© 2016 Académie des sciences. Published by Elsevier Masson SAS. All rights reserved.

set of all orthonormal bases of $\mathbb{R}^n$ and for $b \in \mathcal{B}$ write

$$b = (b_1, b_2, .., b_n).$$

Given a class of signals (vectors) in $\mathbb{R}^n$, we want to approximate each signal f by a superposition of k weighted signals selected from an ordered set of given orthonormal vectors $b = (b_1, b_2, \dots, b_n) \in \mathcal{B}$, as follows,

$$\hat{f}_k = \sum_{j=1}^{k} (f, b_j) b_j. \tag{1}$$

This k-approximation of f should be optimal in the sense of minimizing an error measure that evaluates the size of the vectors

$$\mathcal{E}(k) \equiv f - \hat{f}_k, \tag{2}$$

over the class of signals for every k. The meaning of a minimization of the representation error must be made precise depending on the properties that define the specific class of signals we deal with. In case the signals are a realization of a stochastic process, the minimization of a representation error may require the minimization of

$$\mathbf{E}\|\mathcal{E}(k)\|^2 = \mathbf{E}(\|f - \hat{f}_k\|^2), \tag{3}$$

where $\mathbf{E}$ is the ensemble average operator. If, say, f is selected uniformly from a given finite, but possible very large set of N signals, say $F_1, F_2, \dots, F_N \in \mathbb{R}^n$, then, we can minimize the average (squared) error, that is, determine $b \in \mathcal{B}$ that solves

$$\min_{b\in\mathcal{B}} \frac{1}{N} \sum_{i=1}^{N} \left\| F_i - \sum_{j=1}^{k} (F_i, b_j) b_j \right\|^2 . \tag{4}$$

We refer to the solution to (4) as the principal component analysis (PCA) of the set of signals $F_1, F_2, \dots, F_N$. In this paper, we also study the representation error over the class of signals $f \in \mathbb{R}^n$ obeying the condition $\frac{1}{N}\sum_{i=1}^{N}(F_i, f)^2 \le 1$. Hence, we first determine the basis $b \in \mathcal{B}$ that solves for every k the problem

$$\min_{b\in\mathcal{B}} \max_{\substack{f \in \mathbb{R}^n \\ \frac{1}{N}\sum_{i=1}^{N}(F_i, f)^2 \le 1}} \left\| f - \hat{f}_k \right\| . \tag{5}$$

We refer to (5) as the min–max problem. Note that throughout the discussion above, we are searching to minimize an error measure with respect to a basis $b = (b_1, b_2, \dots, b_n)$ that provides an ordered set of orthonormal vectors $b_1, b_2, \dots, b_k, \dots, b_n$, so as to be optimal for all $1 \le k \le n-1$.

We next provide a general recipe for finding optimal progressive representation bases and prove their uniqueness for different signal representation problems. Specifically, we will show that the best basis for the expected (average) mean squared error representation problem (PCA)

$$\min_{b\in\mathcal{B}} \frac{1}{N} \sum_{i=1}^{N} \left\| F_i - \sum_{j=1}^{k} (F_i, b_j) b_j \right\|^2 , \tag{6}$$

$\forall k = 1, 2, \dots, n-1$, coincides, after inverting the order of the basis vectors, with the optimal basis that solves the progressive min–max error representation problem defined by

$$\min_{b\in\mathcal{B}} \max_{\substack{f \in \mathbb{R}^n \\ (\mathcal{R}f, f) \le 1}} \left\| f - \sum_{j=1}^{k} (f, b_j) b_j \right\| , \tag{7}$$

where the operator $\mathcal{R}$ is $\mathcal{R}f = \frac{1}{N}\sum_{i=1}^{N}(F_i, f)F_i$, hence,

$$(\mathcal{R}f, f) = \frac{1}{N} \sum_{i=1}^{N} (F_i, f)^2. \tag{8}$$

Note that the eigenvectors $e_1, e_2, \dots, e_n$ of the symmetric non-negative operator $\mathcal{R}$, defined by $\mathcal{R}e_i = \rho_i e_i$, ordered to correspond to the decreasing values of the eigenvalues ρ_i, provide the solution to the PCA problem (4) of best k-term approximation. That is, the optimal b is given by $b_j = \pm e_j$. This is the well-known Karhunen–Loève or Principal Component Analysis solution to the average least squares k-term approximation (4).

The approach outlined below allows us to also analyze the regularized PCA presented in [1], and provide uniqueness conditions for the construction of related representation spaces. The main idea behind the analysis is defining the signal class via a self-adjoint operator whose eigenvectors provide the optimal and unique basis.

Please cite this article in press as: Y. Aflalo et al., Best bases for signal spaces, C. R. Acad. Sci. Paris, Ser. I (2016), http://dx.doi.org/10.1016/j.crma.2016.10.002

2. Optimality and uniqueness for a min–max problem

Let D be a linear operator from $\mathbb{R}^n$ to $\mathbb{R}^N$, where $n \le N$. Assume that D is injective, that is

$$N(D) = \{f \in \mathbb{R}^n,\ Df = 0\} = 0. \tag{9}$$

Set $T = (D^{\mathrm{T}}D)^{1/2} : \mathbb{R}^n \to \mathbb{R}^n$, so that T is symmetric. Denote by

$$0 \le \lambda_1 \le \lambda_2 \le \cdots \le \lambda_n \tag{10}$$

its eigenvalues, with corresponding eigenvectors $e_1, e_2, .., e_n$. Because of assumption (9), $\lambda_1 > 0$. Assume, for simplicity, that

$$\lambda_1 < \lambda_2 < \cdots < \lambda_n. \tag{11}$$

Note that $\forall f \in \mathbb{R}^n$,

$$\|Df\|^2 = (D^{\mathrm{T}}Df, f) = (T^2 f, f) = \|Tf\|^2. \tag{12}$$

Our first result is,

Theorem 2.1. *For every $1 \le k \le n-1$ we have*

$$\alpha_k \equiv \min_{b \in \mathcal{B}} \max_{\substack{f \in \mathbb{R}^n \\ \|Df\| \le 1}} \left\| f - \sum_{i=1}^{k} (f, b_i) b_i \right\| = \frac{1}{\lambda_{k+1}}, \tag{13}$$

where the $\min_{b\in\mathcal{B}}$ in (13) is taken over all orthonormal bases $(b_1, b_2, .., b_n)$ of $\mathbb{R}^n$. Moreover, the only orthonormal basis that is a minimizer of (13) for every $1 \le k \le n-1$ is $(e_1, e_2, .., e_n)$ modulo $\pm$.

Proof. In view of (12), we may replace the constraint $\|Df\| \le 1$ in (13) by $\|Tf\| \le 1$.
Step 1. We have

$$\alpha_k \le \frac{1}{\lambda_{k+1}}, \qquad \forall k < n. \tag{14}$$

Choose $b_i = e_i$, $\forall i = 1, 2, .., n$ in (13) and note that $\forall f$,

$$\left\| f - \sum_{i=1}^{k} (f, e_i) e_i \right\|^2 = \left\| \sum_{i=k+1}^{n} (f, e_i) e_i \right\|^2 = \sum_{i=k+1}^{n} |(f, e_i)|^2, \tag{15}$$

while

$$Tf = \sum_{i=1}^{n} (f, e_i) \lambda_i e_i,$$

so that,

$$\|Tf\|^2 \ge \sum_{i=k+1}^{n} |(f, e_i)|^2 \lambda_i^2 \ge \lambda_{k+1}^2 \sum_{i=k+1}^{n} |(f, e_i)|^2,$$

and hence,

$$\left\| f - \sum_{i=1}^{k} (f, e_i) e_i \right\|^2 \le \frac{1}{\lambda_{k+1}^2} \|Tf\|^2. \tag{16}$$

Step 2. Proof of the main theorem by induction on n.

Let $(b_1, b_2, .., b_n)$ be an orthonormal basis which is a minimizer in (13) for all $k = 1, 2, .., n-1$. Choosing in particular $k = n-1$ gives

$$|(f, b_n)| \le \alpha_{n-1} \|Tf\| \qquad \forall f. \tag{17}$$

Taking $f = b_n$ in (17), and writing

$$b_n = \sum_{i=1}^{n} (b_n, e_i) e_i,$$

yields

$$1 \le \alpha_{n-1}^2 \sum_{i=1}^{n} |\lambda_i|^2 |(b_n, e_i)|^2.$$

But

$$1 = \|b_n\|^2 = \sum_{i=1}^{n} |(b_n, e_i)|^2,$$

implies

$$\sum_{i=1}^{n} (1 - \lambda_i^2 \alpha_{n-1}^2) |(b_n, e_i)|^2 \le 0. \tag{18}$$

On the other hand,

$$1 - \lambda_i^2 \alpha_{n-1}^2 > 0 \qquad \forall i = 1, 2, .., n-1, \tag{19}$$

and

$$1 - \lambda_n^2 \alpha_{n-1}^2 \ge 0 \tag{20}$$

since $\lambda_i < \lambda_n \le \frac{1}{\alpha_{n-1}}$, by Step 1.

It follows from (18), (19), and (20), that

$$(b_n, e_i) = 0 \qquad \forall i = 1, 2, .., n-1, \tag{21}$$

and

$$1 - \lambda_n^2 \alpha_{n-1}^2 = 0, \tag{22}$$

as otherwise we would deduce that $(b_n, e_n) = 0$, and thus $b_n = 0$, that contradicts our assumption about b_n. Therefore,

$$\alpha_{n-1} = \frac{1}{\lambda_n}. \tag{23}$$

Hence,

$$b_n = \sum_{i=1}^{n} (b_n, e_i) e_i = (b_n, e_n) e_n,$$

and $|(b_n, e_n)| = 1$, so that $b_n = \pm e_n$.

The space $M = (e_n)^\perp$ has dimension $(n-1)$ and $T(M) \subset M$. Assume that the theorem holds up to $n-1$. The eigenvalues of $T|_M$ $(= T$ restricted to $M)$ are $\lambda_1 < \lambda_2 < \cdots < \lambda_{n-1}$ and $b_1, b_2, .., b_{n-1}$ is an orthonormal basis of M. By the induction assumption applied in M to $T|_M$ we deduce that $\alpha_k = \frac{1}{\lambda_{k+1}}$ $\forall k \le n-2$, and that $b_i = \pm e_i$ for $i = 1, 2, .., n-1$. □

3. PCA revisited

Consider a linear symmetric operator $\mathcal{R} : \mathbb{R}^n \to \mathbb{R}^n$. Denote

$$\rho_1 \ge \rho_2 \ge \cdots \ge \rho_n \tag{24}$$

its eigenvalues (no assumptions on their signs), and

$$e = (e_1, e_2, .., e_n)$$

the corresponding orthonormal basis of eigenvectors, that is,

$$\mathcal{R} e_i = \rho_i e_i \quad \forall i = 1, 2, .., n. \tag{25}$$

Assume, for simplicity, that

$$\rho_1 > \rho_2 > \cdots > \rho_n. \tag{26}$$

As in Section 2, let $\mathcal{B}$ denote the class of all orthonormal bases of $\mathbb{R}^n$ and for $b \in \mathcal{B}$ write

$$b = (b_1, b_2, .., b_n).$$

Given $b \in \mathcal{B}$, and $k = 1, 2, .., n$, set

$$Y_k(b) = \sum_{j=1}^{k} (\mathcal{R} b_j, b_j). \tag{27}$$

Note that for every $b \in \mathcal{B}$,

$$Y_n(b) = \rho_1 + \rho_2 + \cdots + \rho_n. \tag{28}$$

Indeed, write for every $j = 1, 2, \ldots, n$

$$b_j = \sum_{i=1}^{n} (b_j, e_i) e_i, \tag{29}$$

so that

$$\mathcal{R} b_j = \sum_{i=1}^{n} \rho_i (b_j, e_i) e_i, \tag{30}$$

$$(\mathcal{R} b_j, b_j) = \sum_{i=1}^{n} \rho_i (b_j, e_i)^2, \tag{31}$$

and

$$Y_n(b) = \sum_{j=1}^{n} \sum_{i=1}^{n} \rho_i (b_j, e_i)^2 = \sum_{i=1}^{n} \rho_i. \tag{32}$$

Theorem 3.1. *Assume (26) holds. For every $k = 1, 2, .., n-1$, we have*

$$Y_k(e) = \rho_1 + \rho_2 + \cdots + \rho_k, \tag{33}$$

and for every $b \in \mathcal{B}$,

$$Y_k(b) \le \rho_1 + \rho_2 + \cdots + \rho_k. \tag{34}$$

Moreover,

$$Y_k(b) = \rho_1 + \rho_2 + \cdots + \rho_k \quad \forall k = 1, 2, .., n-1, \tag{35}$$

if and only if

$$b_j = \pm e_j \quad \forall j = 1, 2, .., n. \tag{36}$$

Proof. In view of (31) we have

$$Y_k(b) = \sum_{j=1}^{k} \sum_{i=1}^{n} \rho_i (b_j, e_i)^2. \tag{37}$$

Thus, Theorem 3.1 is an immediate consequence of the following

Lemma 3.1. *Let $e, b \in \mathcal{B}$ and let $(\rho_i)_{1 \le i \le n}$ be a sequence in $\mathbb{R}$ satisfying (26). Then, for every $k = 1, 2, \ldots, n-1$*

$$Y_k \equiv \sum_{j=1}^{k} \sum_{i=1}^{n} \rho_i (b_j, e_i)^2 \le \rho_1 + \rho_2 + \cdots + \rho_k. \tag{38}$$

Moreover, equality in (38) holds for every $k = 1, 2, \ldots, n-1$, if and only if

$$b_j = \pm e_j \qquad \forall j = 1, 2, \ldots, n.$$

Proof. Write

$$Y_k = U_k + V_k, \tag{39}$$

where

$$U_k = \sum_{i=1}^{k}\sum_{j=1}^{n} \rho_i (b_j, e_i)^2 \tag{40}$$

and

$$V_k = \sum_{i=k+1}^{n}\sum_{j=1}^{k} \rho_i (b_j, e_i)^2. \tag{41}$$

From (41) and (24) we have

$$\begin{aligned} V_k &\le \rho_{k+1} \sum_{i=k+1}^{n}\sum_{j=1}^{k} (b_j, e_i)^2 \\ &= \rho_{k+1} \sum_{j=1}^{k} \left(1 - \sum_{i=1}^{k} (b_j, e_i)^2\right) \\ &= \rho_{k+1} \left(k - \sum_{i=1}^{k}\sum_{j=1}^{k} (b_j, e_i)^2\right) \\ &= \rho_{k+1} \sum_{i=1}^{k} \left(1 - \sum_{j=1}^{k} (b_j, e_i)^2\right). \end{aligned} \tag{42}$$

On the other hand, by (40),

$$\sum_{i=1}^{k} \rho_i - U_k = \sum_{i=1}^{k} \rho_i \left(1 - \sum_{j=1}^{k} (b_j, e_i)^2\right).$$

Applying (24) once more yields

$$\begin{aligned} \sum_{i=1}^{k} \rho_i - U_k &\ge \rho_{k+1} \sum_{i=1}^{k} \left(1 - \sum_{j=1}^{k} (b_j, e_i)^2\right) \\ &\ge V_k, \end{aligned}$$

by (42). From (42) and (39) we deduce (38).

Assume now that equality in (38) holds for every $k = 1, 2, \dots, n-1$. Taking $k = 1$, we have

$$\sum_{i=1}^{n} \rho_i (b_1, e_i)^2 = \rho_1,$$

so that

$$\sum_{i=1}^{n} (\rho_i - \rho_1)(b_1, e_i)^2 = 0.$$

From (26) we deduce that

$$(b_1, e_i) = 0 \qquad \forall i = 2, \dots, n$$

and thus, $b_1 = \pm e_1$. We are now reduced to the same question for two bases $e' = (e_2, \dots, e_n)$ and $b' = (b_2, \dots, b_n)$ of $\mathbb{R}^{n-1}$, and we may conclude by induction. □

Remark 3.1. If we assume only (24) instead of (26), then (34) holds. Moreover, equality in (35) holds if and only if $\mathcal{R} b_j = \rho_j b_j \ \forall j = 1, 2, .., n$. The proof is an easy adaptation of the one presented above.

Remark 3.2. Assertion (38) in Lemma 3.1 can also be derived from a lemma in Mirsky [4]. Applying Mirsky's lemma (and following his notations) to the doubly stochastic matrix $d_{ij} = (b_j, e_i)^2$, $x_i = \rho_i$ and $y_i = 1$ for $1 \le i \le k$, $y_i = 0$ for $k+1 \le i \le n$, yields

$$\sum_{i=1}^{n}\sum_{j=1}^{k}\rho_i(b_j, e_i)^2 \le \sum_{i=1}^{k}\rho_i, \tag{43}$$

which is precisely (38).

We are now going to apply Theorem 3.1 in a PCA-type setting. Let $F_1, F_2, .., F_N \in \mathbb{R}^n$ be given. Consider the linear operator $\mathcal{R} : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$\mathcal{R}f = \frac{1}{N}\sum_{i=1}^{N}(F_i, f)F_i, \qquad f \in \mathbb{R}^n, \tag{44}$$

so that

$$(\mathcal{R}f, g) = \frac{1}{N}\sum_{i=1}^{N}(F_i, f)(F_i, g), \tag{45}$$

and thus $\mathcal{R}$ is symmetric and nonnegative. As above, denote

$$\rho_1 \ge \rho_2 \ge \cdots \ge \rho_n \ge 0 \tag{46}$$

its eigenvalues, and

$$e = (e_1, e_2, .., e_n)$$

the corresponding orthonormal basis of eigenvectors. Assume, for simplicity, that

$$\rho_1 > \rho_2 > \cdots > \rho_n. \tag{47}$$

Let $\mathcal{B}$ denote the class of all orthonormal bases of $\mathbb{R}^n$, and for $b \in \mathcal{B}$ write

$$b = (b_1, b_2, .., b_n).$$

Given $b \in \mathcal{B}$, and $k = 1, 2, .., n$, set

$$X_k(b) \equiv \frac{1}{N}\sum_{i=1}^{N}\left\| F_i - \sum_{j=1}^{k}(F_i, b_j)b_j \right\|^2. \tag{48}$$

Clearly, $X_n(b) = 0 \ \forall b \in \mathcal{B}$.

Corollary 3.1. *For every $k = 1, 2, .., n-1$ we have*

$$X_k(e) = \rho_{k+1} + \rho_{k+2} + \cdots + \rho_n, \tag{49}$$

and for every $b \in \mathcal{B}$,

$$X_k(b) \ge \rho_{k+1} + \rho_{k+2} + \cdots + \rho_n. \tag{50}$$

Moreover,

$$X_k(b) = \rho_{k+1} + \rho_{k+2} + \cdots + \rho_n \quad \forall k = 1, 2, .., n-1, \tag{51}$$

if and only if

$$b_j = \pm e_j \quad \forall j = 1, 2, .., n. \tag{52}$$

Please cite this article in press as: Y. Aflalo et al., Best bases for signal spaces, C. R. Acad. Sci. Paris, Ser. I (2016), http://dx.doi.org/10.1016/j.crma.2016.10.002

Proof. From (48) we have

$$\begin{aligned} X_k(b) &= \frac{1}{N}\sum_{i=1}^{N}\|F_i\|^2 - \frac{1}{N}\sum_{i=1}^{N}\sum_{j=1}^{k}(F_i,b_j)^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\|F_i\|^2 - \sum_{j=1}^{k}(\mathcal{R}b_jb_j) && \text{by (45)} \\ &= \frac{1}{N}\sum_{i=1}^{N}\|F_i\|^2 - Y_k(b) && \text{by (48).} \end{aligned} \tag{53}$$

Applying (34) we deduce that

$$X_k(b) \geq \frac{1}{N}\sum_{i=1}^{N}\|F_i\|^2 - (\rho_1+\rho_2+\cdots+\rho_k). \tag{54}$$

On the other hand, by (45) and (25) we have

$$\begin{aligned} \sum_{j=1}^{n}(\mathcal{R}e_j,e_j) = \sum_{j=1}^{n}\rho_j &= \frac{1}{N}\sum_{j=1}^{n}\sum_{i=1}^{N}(F_i,e_j)^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}\|F_i\|^2. \end{aligned} \tag{55}$$

Combining (54) and (55) yields

$$\begin{aligned} X_k(b) &\geq \sum_{j=1}^{n}\rho_j - (\rho_1+\rho_2+\ldots\rho_k) \\ &= \rho_{k+1}+\rho_{k+2}+\cdots+\rho_n, \end{aligned}$$

which is precisely (50). On the other hand,

$$\begin{aligned} X_k(e) &= \sum_{i=1}^{N}\|F_i\|^2 - Y_k(e) \\ &= \sum_{j=1}^{n}\rho_j - Y_k(e) \\ &= \rho_{k+1}+\rho_{k+2}+\cdots+\rho_n \end{aligned}$$

by (34), that is, (49) holds. The last assertion in Corollary 3.1 follows from the fact that

$$X_k(b)+Y_k(b) = \sum_{j=1}^{n}\rho_j. \quad \square$$

4. Returning to min–max

Here, we apply Theorem 2.1 in the above (PCA-type) setting, that is, we start with $F_1, F_2, \ldots, F_N \in \mathbb{R}^n$, and consider the operator $\mathcal{R}$ defined by (44). Set

$$D = \mathcal{R}^{1/2}, \tag{56}$$

so that D is also symmetric and nonnegative, moreover $T = (D^{\mathrm{T}}D)^{1/2} = D$, and by (56),

$$\|Df\|^2 = (Df,Df) = (D^2f,f) = (\mathcal{R}f,f). \tag{57}$$

Thus, by (45) and (57)

$$\|Df\|^2 = \frac{1}{N}\sum_{i=1}^{N}(F_i,f)^2. \tag{58}$$

We denote again by

$$\rho_1 \geq \rho_2 \geq \cdots \geq \rho_n \geq 0$$

the eigenvalues of $\mathcal{R}$ with corresponding eigenvectors $e_1, e_2, \ldots, e_n$. Hence, the eigenvalues of $T = D$

$$0 \leq \lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n,$$

in non-decreasing order (as in Section 2) are given by

$$\lambda_k = \rho_{n-k+1}^{1/2}.$$

Assume for simplicity that

$$\rho_1 > \rho_2 > \cdots > \rho_n > 0. \tag{59}$$

Applying Theorem 2.1 yields

Corollary 4.1. *For every $1 \leq k \leq n-1$ we have*

$$\alpha_k^{\min\max} = \min_{b \in \mathcal{B}} \max_{\substack{f \in \mathbb{R}^n \\ \frac{1}{N}\sum_{i=1}^N (F_i, f)^2 \leq 1}} \left\| f - \sum_{j=1}^k (f, b_j) b_j \right\| = \rho_{n-k}^{-1/2}. \tag{60}$$

Moreover, the only orthonormal basis b that is a minimizer of (60) for every $1 \leq k \leq n-1$ is $(e_n, e_{n-1}, \ldots, e_1)$ modulo ±.

5. Relating the min–max problem to the PCA

Given $F_1, F_2, \ldots, F_N \in \mathbb{R}^n$, denote by $\mathcal{B}$ the class of all orthonormal bases of $\mathbb{R}^n$, and for $b \in \mathcal{B}$ write $b = (b_1, b_2, \ldots, b_n)$. Then, the optimal and unique bases obtained for the two following minimization problems, namely the PCA and the min–max, are intimately related.

Corollary 4.1 provides the solution to

$$\begin{aligned} b^{\min\max} &= \operatorname*{argmin}_{b \in \mathcal{B}} \max_{\substack{f \in \mathbb{R}^n \\ \frac{1}{N}\sum_{i=1}^N (F_i, f)^2 \leq 1}} \left\| f - \sum_{j=1}^k (f, b_j) b_j \right\| \\ &= \operatorname*{argmin}_{b \in \mathcal{B}} \max_{\substack{f \in \mathbb{R}^n \\ (\mathcal{R}f, f) \leq 1}} \left\| f - \sum_{j=1}^k (f, b_j) b_j \right\|, \end{aligned} \tag{61}$$

$\forall k = 1, 2, \ldots, n-1$, with $\mathcal{R}$ defined by (44). On the other hand, the PCA-based progressive representation problem searches for the optimal basis such that $\forall k = 1, 2, \ldots, n-1$,

$$\begin{aligned} b^{\mathrm{PCA}} &= \operatorname*{argmin}_{b \in \mathcal{B}} \frac{1}{N} \sum_{i=1}^N \left\| F_i - \sum_{j=1}^k (F_i, b_j) b_j \right\|^2 \\ &= \operatorname*{argmin}_{b \in \mathcal{B}} \sum_{j=1}^k \left(-(\mathcal{R} b_j, b_j) \right) \\ &= \operatorname*{argmax}_{b \in \mathcal{B}} \sum_{j=1}^k \left(\mathcal{R} b_j, b_j \right). \end{aligned} \tag{62}$$

In both cases, the optimal basis is given (modulo ±) by the eigenvectors $(e_1, e_2, \ldots, e_n)$ of the operator $\mathcal{R}$ and eigenvectors are ordered according to the values of their corresponding eigenvalues. Specifically, we have shown that $b_i^{\mathrm{PCA}} = \pm b_{n+1-i}^{\min\max}$, where in both cases the solution is given by the ordered eigenvectors of $\mathcal{R}$, one corresponding to the descending and the other to the ascending sizes of the eigenvalues, for $\mathcal{R}$'s having a simple spectrum. Moreover, in both cases, the solution is also unique.

At this point, we might be puzzled by the change in ordering of the basis vectors for the min–max error solution and the optimal squared average error (PCA) solution for the problems discussed so far. The reason of this result is the fact that the optimization is carried out for two different classes of signals. Indeed, in the first case we are minimizing the maximal error for the partial k-approximation of a signal f from the class of signals defined by

$$\left\{ f \;\middle|\; \frac{1}{N}\sum_{i=1}^{N}(F_i, f)^2 \le 1 \right\}, \tag{63}$$

whereas in the second PCA-case we are minimizing the average squared error of the k-approximation of signals selected from the set

$$\{F_1, F_2, \dots, F_N\}. \tag{64}$$

Suppose however we would like to do a PCA analysis for the class of signals defined by (63). To do so we need the positive definite symmetric autocorrelation matrix of the set of vectors F uniformly drawn from the (bounded but continuous region of $\mathbb{R}^n$) given by (63), which is defined by

$$R_f = Eff^{\mathrm{T}}, \qquad f \sim \text{uniform}\left\{ f \middle| \frac{1}{N}\sum_{i=1}^{N}(F_i, f)^2 \le 1 \right\}.$$

We have from the condition defining the class of f's that

$$(\mathcal{R}f, f) = \sum_{i=1}^{N} \rho_i (e_i, f)^2 \le 1;$$

hence the vectors

$$s(f) = \Lambda^{1/2} e^{\mathrm{T}} f,$$

have length ≤ 1 and are uniformly distributed in the unit sphere in $\mathbb{R}^n$, see [3]. Here, e denotes the eigenvectors unitary operator $e = (e_1, e_2, \dots, e_n)$, and $\Lambda^{1/2} = \operatorname{diag}(\rho_1^{1/2}, \rho_2^{1/2}, \dots, \rho_n^{1/2})$. Therefore, we have $Ess^{\mathrm{T}} = \mathcal{I}$ (identity) implying that

$$\Lambda^{1/2} e^{\mathrm{T}} Eff^{\mathrm{T}} e \Lambda^{1/2} = \mathcal{I},$$

or

$$Eff^{\mathrm{T}} = e\Lambda^{-1} e^{\mathrm{T}}.$$

This shows that the PCA of the class of signals defined by (63) has the eigenvectors of $\mathcal{R}$ and the eigenvalues equal to $\rho_1^{-1}, \rho_2^{-1}, \dots, \rho_n^{-1}$. Hence, we have

$$0 < \rho_1^{-1} < \rho_2^{-1} < \dots < \rho_n^{-1},$$

and the ordering of the eigenvectors is exactly the same for the min–max and the PCA problems for the class of signals determined by (63).

In another direction, it was suggested in [1] to use

$$\begin{aligned} b^{\mathrm{RPCA}} &= \operatorname*{argmin}_{b \in \mathcal{B}} \left(\frac{1}{N}\sum_{i=1}^{N} \left\| F_i - \sum_{j=1}^{k}(F_i, b_j) b_j \right\|^2 + \mu \sum_{j=1}^{k} \|Db_j\|^2 \right) \\ &= \operatorname*{argmin}_{b \in \mathcal{B}} \sum_{j=1}^{k} \left(\mu(D^{\mathrm{T}}Db_j, b_j) - \frac{1}{N}\sum_{i=1}^{N}(F_i, b_j)^2 \right), \end{aligned} \tag{65}$$

$\forall k$ such that $1 \le k \le n-1$, as a regularized PCA model for data analysis, which was extremely useful for efficient shape representation. Here μ is a positive constant and $D : \mathbb{R}^n \to \mathbb{R}^N$ is a linear operator as in Section 2. Consider the operator $\mathcal{Q} : \mathbb{R}^n \to \mathbb{R}^n$ defined by

$$\mathcal{Q}f = \frac{1}{N}\sum_{i=1}^{N}(F_i, f)F_i - \mu D^{\mathrm{T}}Df.$$

Clearly $\mathcal{Q}$ is symmetric and (65) becomes

$$b^{\mathrm{RPCA}} = \operatorname*{argmax}_{b \in \mathcal{B}} \sum_{j=1}^{k} \left(\mathcal{Q}b_j, b_j\right). \tag{66}$$

We are now in a position to apply Theorem 3.1 (provided the eigenvalues of $\mathcal{Q}$ are simple). The solution to (66) is given by $(\pm e_i)$ where the e_i's are the eigenvectors ordered in descending order (see (26)).

The optimality and uniqueness proofs for the subspaces introduced in this note provide novel perspective on signal representation. It allowed us to justify some classical tools and relate the PCA to a specific min–max problem. Finally, the presented framework enabled us to support the construction of the regularized PCA suggested in [1] for an appropriate smoothness term.

Acknowledgements

This work was supported by ERC advanced grant 267414 (NORDIA). HB was partially supported by NSF grant DMS-1207793 and also by grant number 238702 of the European Commission (ITN, project FIRST).

Appendix A. Min–max in the infinite dimensional case

We now return to the setting of Section 2, this time in the infinite dimensional case. Let H be a Hilbert space equipped with the scalar product (u, v) and corresponding norm $\|u\| = (u, u)^{1/2}$. Let $V \subset H$ be another Hilbert space, equipped with the norm $\| \, \|_V$, such that V is a dense subspace of H and $V \subset H$ with compact injection. Let $D : V \to H$ be a bounded linear operator such that

$$\|u\|_V \le C(\|Du\| + \|u\|) \qquad \forall u \in V,$$

for some constant $C > 0$.

Set $T = (D^{\mathrm{T}}D)^{1/2} : V \to H$. It is easy to see that $T + \gamma\mathcal{I}$ is bijective from V onto H for every $\gamma > 0$. In particular $(T + \gamma\mathcal{I})^{-1}$ is a self-adjoint compact operator from H into itself. Thus, $(T + \gamma\mathcal{I})^{-1}$ admits a spectral decomposition. Returning to T, we obtain an orthonormal basis $e = (e_1, e_2, \ldots)$ of H which consists of eigenvectors of T, that is,

$$e_i \in V \quad \forall i \qquad \text{and} \qquad Te_i = \lambda_i e_i \quad \forall i$$

with

$$0 \le \lambda_1 \le \lambda_2 \le \ldots \tag{A.1}$$

and

$$\lambda_i \to \infty \quad \text{as} \quad i \to \infty. \tag{A.2}$$

Example. A standard example is $H = L^2(\Omega)$ and $V = H_0^1(\Omega)$, where Ω is domain in $\mathbb{R}^N$, and $H_0^1(\Omega) = \{u \in L^2(\Omega);\ \nabla u \in L^2(\Omega);$ and $u = 0$ on $\partial\Omega\}$. Taking $Du = \nabla u (= \text{grad } u)$ we have $D^{\mathrm{T}}D = -\Delta$. Then, $T = (-\Delta)^{1/2}$ and the e_i's are the eigenfunctions of $-\Delta$ with zero Dirichlet boundary condition on $\partial\Omega$; moreover, $\lambda_i = \mu_i^{1/2}$ $\forall i$ where the μ_i's are the eigenvalues of $-\Delta$.

We denote by $\mathcal{B}$ the class of all orthonormal bases of H. If $b \in \mathcal{B}$ write $b = (b_1, b_2, \ldots)$ and set for any $k = 1, 2, \ldots$

$$\begin{aligned} \alpha_k(b) &= \sup_{\substack{f \in V \\ \|Df\| \le 1}} \left\| f - \sum_{i=1}^{k} (f, b_i) b_i \right\| \\ &= \sup_{\substack{f \in V \\ \|Tf\| \le 1}} \left\| f - \sum_{i=1}^{k} (f, b_i) b_i \right\|, \end{aligned} \tag{A.3}$$

using the fact that $\|Df\| = \|Tf\|$, $\forall f \in V$.

Our main result in this section is

Theorem A.1. *For every $k \ge 1$, we have*

$$\alpha_k(e) = \frac{1}{\lambda_{k+1}}, \tag{A.4}$$

and for every $b \in \mathcal{B}$

$$\alpha_k(b) \ge \frac{1}{\lambda_{k+1}}. \tag{A.5}$$

Please cite this article in press as: Y. Aflalo et al., Best bases for signal spaces, C. R. Acad. Sci. Paris, Ser. I (2016), http://dx.doi.org/10.1016/j.crma.2016.10.002

The proof consists of two steps.

Step 1. We have, for every $k \geq 1$,

$$\alpha_k(e) \leq \frac{1}{\lambda_{k+1}}. \tag{A.6}$$

The proof of Step 1 is identical to the proof of Step 1 in the finite-dimensional case (Section 2) and we will not repeat it.

Step 2. $\forall b \in \mathcal{B}$, we have

$$\alpha_k(b) \geq \frac{1}{\lambda_{k+1}} \qquad \forall k = 1, 2, \ldots \tag{A.7}$$

Proof. We follow the same arguments as in [1], that is, we use the Courant–Fischer min–max principle, see, e.g., [2], page 517. It implies in particular that $\forall k = 1, 2, \ldots$, and every closed subspace Λ of V of codimension k, we have

$$\min_{\substack{g \in \Lambda \\ g \neq 0}} \frac{\|Tg\|}{\|g\|} \leq \lambda_{k+1}. \tag{A.8}$$

Fix any basis $b \in \mathcal{B}$ and set

$$\Lambda_0 = \{f \in V;\ (f, b_i) = 0\ \forall i = 1, 2, \ldots, k\}$$

so that Λ_0 is a closed subspace of V, of codimension k. Applying (A.8) with $\Lambda = \Lambda_0$, we obtain some $g \in V$ such that

$$(g, b_i) = 0 \qquad \forall i = 1, 2, \ldots, k \tag{A.9}$$

$$\|g\| = 1 \tag{A.10}$$

$$\|Tg\| \leq \lambda_{k+1}. \tag{A.11}$$

Set

$$f_\epsilon = \frac{g}{\lambda_{k+1} + \epsilon}, \qquad \epsilon > 0, \tag{A.12}$$

so that $f_\epsilon \in V$ and, by (A.11), $\|Tf_\epsilon\| \leq 1$. Inserting $f = f_\epsilon$ in (A.3) yields

$$\alpha_k(b) \geq \left\| f_\epsilon - \sum_{i=1}^{k} (f_\epsilon, b_i) \right\|. \tag{A.13}$$

From (A.9), (A.12) and (A.13), we deduce that

$$\alpha_k(b) \geq \|f_\epsilon\| = \frac{1}{\lambda_{k+1} + \epsilon},$$

by (A.10) and (A.12). Since $\epsilon > 0$ is arbitrary, we obtain (A.7).

Proof of Theorem A.1 completed. Combining Step 1 and Step 2, we readily have that

$$\alpha_k(e) = \frac{1}{\lambda_{k+1}},$$

and consequently, $\forall k = 1, 2, \ldots$

$$\min_{b \in \mathcal{B}} \sup_{\substack{f \in V \\ \|Tf\| \leq 1}} \left\| f - \sum_{i=1}^{k} (f, b_i) b_i \right\| = \frac{1}{\lambda_{k+1}}, \tag{A.14}$$

and the min in (A.14) is achieved when $b = e$. □

Remark A.1. We still have the standing problem: is it true that if strict inequality holds in (A.1) and if $b \in \mathcal{B}$ satisfies

$$\alpha_k(b) = \frac{1}{\lambda_{k+1}} \qquad \forall k = 1, 2, \ldots \tag{A.15}$$

then $b = e$ modulo ±? In the finite-dimensional case, we have a proof of Step 2 without the Courant–Fischer min–max principle and we were able to use induction combined with dimensional reduction to settle uniqueness.

References

[1] Y. Aflalo, H. Brezis, R. Kimmel, On the optimality of shape and data representation in the spectral domain, SIAM J. Imaging Sci. 8 (2) (2015) 1141–1160.

[2] H. Brezis, Functional Analysis, Sobolev Spaces and Partial Differential Equations, Springer Universitext Series, 2010.

[3] J.D. Gammell, T.D. Barfoot, The probability density function of a transformation-based hyperellipsoid sampling technique, arXiv:1404.1347, 2014.

[4] L. Mirsky, A trace inequality of John von Neumann, Monatsh. Math. 79 (1975) 303–306.

Please cite this article in press as: Y. Aflalo et al., Best bases for signal spaces, C. R. Acad. Sci. Paris, Ser. I (2016), http://dx.doi.org/10.1016/j.crma.2016.10.002