Journal of Machine Learning Research 18 (2017) 1-28 Submitted 1/16; Revised 5/17; Published 10/17

Optimal Dictionary for Least Squares Representation

Mohammed Rayyan Sheriff MOHAMMEDRAYYAN@SC.IITB.AC.IN
Debasish Chatterjee DCHATTER@IITB.AC.IN
Systems and Control Engineering
IIT Bombay
Mumbai 400076, India

Editor: Benjamin Recht

Abstract

Dictionaries are collections of vectors used for the representation of a class of vectors in Euclidean spaces. Recent research on optimal dictionaries is focused on constructing dictionaries that offer sparse representations, i.e., ℓ_0-optimal representations. Here we consider the problem of finding optimal dictionaries with which representations of a given class of vectors is optimal in an ℓ_2-sense: optimality of representation is defined as attaining the minimal average ℓ_2-norm of the coefficients used to represent the vectors in the given class. With the help of recent results on rank-1 decompositions of symmetric positive semidefinite matrices, we provide an explicit description of ℓ_2-optimal dictionaries as well as their algorithmic constructions in polynomial time.

Keywords: ℓ_2-optimal dictionary, rank-1 decomposition, finite tight frames

1. Introduction

A *dictionary* is a collection of vectors in a finite-dimensional vector space over $\mathbb{R}$, with which other vectors of the vector space are represented. A dictionary is a generalization of a basis: While the number of vectors in a basis is exactly equal to the dimension of the vector space, a dictionary may contain more elements. In this article we consider a problem of finding an optimal dictionary, where optimality is interpreted as the minimum expected average size of the coefficients required to represent a certain collection of vectors drawn from a given probability distribution.

We begin with a toy example to motivate the problems treated in this article. Let V be a random vector that attains values 'close' to $\begin{pmatrix}0 & 2\end{pmatrix}^\top$ with high probability; the situation is demonstrated below:

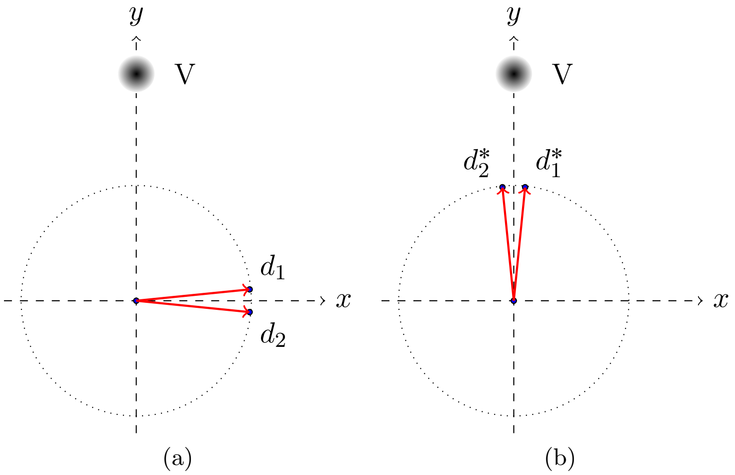

Figure 1: Comparison of two dictionaries.

Suppose that our dictionary consists of the vectors $d_1 = \begin{pmatrix} 1 & -\epsilon \end{pmatrix}^\top$ and $d_2 = \begin{pmatrix} 1 & \epsilon \end{pmatrix}^\top$ in $\mathbb{R}^2$, with a small positive value of ϵ. Since we must represent V using d_1 and d_2, the corresponding coefficients α_1 and α_2 must be such that $\alpha_1 \begin{pmatrix} 1 & \epsilon \end{pmatrix}^\top + \alpha_2 \begin{pmatrix} 1 & -\epsilon \end{pmatrix}^\top = V \approx \begin{pmatrix} 0 & 2 \end{pmatrix}^\top$. A quick calculation shows that the magnitudes of the coefficients α_1 and α_2 should then be approximately equal to $1/(\epsilon)$ with high probability. To wit, the magnitudes of these coefficients are large for small values of ϵ. It is therefore more appropriate in this situation to consider a dictionary consisting of vectors $d_1^* = \begin{pmatrix} \epsilon & 1 \end{pmatrix}^\top$ and $d_2^* = \begin{pmatrix} -\epsilon & 1 \end{pmatrix}^\top$ to represent the samples of V, in which case, the magnitudes of the coefficients of the representations are closer to 1 with high probability. The latter values are comparatively far smaller compared to the values close to $1/(\epsilon)$ obtained with the preceding dictionary. This simple example shows that given some statistical information about the random vectors to be represented, the question of designing a dictionary that minimizes the average cost of representation can be better addressed.

Let us now turn to a situation in which considering the average cost of representations is natural. Our motivation comes from a control theoretic ideas perspective. Consider a linear time-invariant control system modeled by the recursion

$$x(t+1) = Ax(t) + Bu(t), \quad t = 0, 1, \ldots, \tag{1}$$

where the 'system matrix' $A \in \mathbb{R}^{n \times n}$ and the 'control matrix' $B \in \mathbb{R}^{n \times m}$ are given, with the initial boundary condition $x(0) = \bar{x} \in \mathbb{R}^n$ fixed. For an arbitrarily selected $\hat{x} \in \mathbb{R}^n$, consider the standard *reachability problem* for (1), that is:

$$\begin{aligned} &\text{If possible, find a sequence } (u(t))_t \subset \mathbb{R}^m \text{ of control vectors} \\ &\text{that steer the system states to } \hat{x}. \end{aligned} \tag{2}$$

A necessary and sufficient condition for such a sequence to exist for every pair $(\bar{x}, \hat{x})$ is that the rank of the matrix $\mathfrak{R}_K(A,B) := \begin{pmatrix} B & AB & \cdots & A^{n-1}B \end{pmatrix}$ is equal to n, which we impose for the moment. Letting $K := \min \left\{ k \geqslant 0 \,\middle|\, \operatorname{rank}\left(\mathfrak{R}_K(A,B)\right) = n \right\}$ denote the 'reachability index' of (1), we see at once that the control vectors $(u(t))_{t=0}^{K-1}$ needed to execute the transfer of the states of (1) from $\bar{x}$ to $\hat{x}$ must be a solution to the linear equation

$$\hat{x} - A^K \bar{x} = \sum_{t=0}^{K-1} A^t B u(t) = \mathfrak{R}_K(A, B) \begin{pmatrix} u(K-1) \\ \vdots \\ u(1) \\ u(0) \end{pmatrix}.$$

It is now natural to consider the 'control cost' of transferring $\bar{x}$ to $\hat{x}$, for which, a natural candidate is the associated ℓ_2 performance index $\sum_{i=0}^{K-1} \|u(t)\|^2$. Since in practice, the ℓ_2 performance index is analogous to the amount of energy spent to control the system, its practical importance can hardly be overstated in the context of control. Let us list three examples:

- In attitude control/orientation problems of space vehicles, one must execute most of the rapid manoeuvre using the energy from the limited amount of fuel on board, or with the energy available from on-board batteries; minimizing the energy expenditure, therefore, is crucial.
- In controlled automated mobile robots (e.g., automated cars) designed to reach a given location within a certain time, reduction of energy consumption leads directly to reduction in fuel consumed.
- In control of electronic systems such as power electronic drives, the associated ℓ_2 performance index involves information of the amount of power drawn from the electricity grid to control the system, leading directly to minimization of power consumption and thereby heating.

Minimization of control effort has been an integral part of control theory, and is generally studied under the class of Linear Quadratic problems; see, e.g., (Bertsekas, 1995), (Anderson and Moore, 2007), (Clarke, 2013), (Liberzon, 2012), or any standard book on optimal control. It is evident that the task of designing control systems that require minimum control energy for their typical manoeuvres is of great importance.

It is a standard practice to study the reachability problem (2), for $\bar{x} = 0$ and $\hat{x}$ on the unit sphere; due to linearity of (1), this special case provides sufficient insight into the general case. Let us consider the following optimal control problem:

$$\begin{aligned} &\underset{(u(t))_t}{\text{minimize}} && \mathsf{E}\Bigg[\sum_{t=0}^{K-1} \|u(t)\|^2\Bigg] \\ &\text{subject to} && \begin{cases} x(t+1) = Ax(t) + Bu(t) \quad \text{for all } t = 0, \ldots, K-1, \\ x(0) = 0, \\ x(K) = \hat{x} \text{ distributed according to } \mu, \end{cases} \end{aligned} \tag{3}$$

where μ is a probability distribution on $\mathbb{R}^d$. It is known that if $\hat{x}$ is uniformly distributed over the unit sphere, then the optimal control problem (3) admits an unique optimal solution and the optimum value is proportional to $\mathrm{tr}\big(W_{A,B}^{-1}\big)$, where $W_{A,B} := \mathfrak{R}_K(A,B)\mathfrak{R}_K(A,B)^\top$ is the *controllability grammian* of the system; for details see, e.g., (Müller and Weber, 1972) and (Pasqualetti et al., 2014). It can be readily shown that if $\Sigma := \mathsf{E}[\hat{x}\hat{x}^\top]$ is well defined, then the optimum value of (3) is equal to $\mathrm{tr}\big(\Sigma W_{A,B}^{-1}\big)$. Evidently, for a given distribution of

$\hat{x}$, different linear systems (1) — described completely by the pair (A, B) — incur different optimum values $\mathrm{tr}\big(\Sigma W_{A,B}^{-1}\big)$ of (3).

Against the above backdrop, consider the question of *designing* the linear control system (1) such that the value of (3) is as low as possible. Since most control problems involve designing control sequences to execute a class of desired manoeuvres, for a given distribution of $\hat{x}$ it is then natural to design the linear systems in order to minimize the optimum value $\mathrm{tr}\big(\Sigma W_{A,B}^{-1}\big)$ of the optimal control problem (3). In this case, the system design problem is similar to the one of finding an ℓ_2-optimal dictionary as described above: here the matrices A and B are to be designed, within a feasible region, such that the column vectors constituting the matrix $\mathfrak{R}_K(A, B)$ lead to minimal expected average cost of reachability, i.e., minimal value of (3). Such problems routinely arise in networked control, where the pair (A, B) is a function of the constituent systems and the connectivity of the network. From an operational standpoint, it is good for a networked system to have its components connected in a way such that the resulting system incurs small expected average state transfer costs. Indeed, control systems are typically designed (Müller and Weber, 1972) by optimizing a *figure of merit / measure of quality / measure of controllability*; in particular, networked control systems are designed in (Pasqualetti et al., 2014) using a measure of quality defined there. Based on this work on ℓ_2-optimal dictionaries, we have proposed a novel measure of quality in (Sheriff and Chatterjee, 2017), and further developments for algorithmic synthesis of large-scale control systems will be reported elsewhere. Besides these applications in control theory and practice, one of the key objective of our work here is to investigate and understand the physical nature of the ℓ_2-optimal dictionaries independent of their connection with control theory. Such a study will shed light on other control theoretic properties of observability and estimation.

There has been significant recent research into finding optimal dictionaries, briefly outlined in (Tošić and Frossard, 2011); current research centers around the development of learning algorithms for finding optimal dictionaries. Much of the thrust is on arriving at dictionaries that offer sparse representations of sample vectors. One of the first learning algorithms to develop a dictionary that offers sparse representation of images was given in (Olshausen and Field, 1997). Since then many learning algorithms have been developed to obtain dictionaries that offer sparse representation along with other special properties such as online computation capability (Mairal et al., 2009a), better classification property (Mairal et al., 2009b; Yang et al., 2011), better adaptive properties (Skretting and Engan, 2010); several other algorithms are given in (K. Delgado et al., 2003; Yaghoobi et al., 2009; Mallat and Zhang, 1993).

The problem addressed in this article differs from the mainstream research of finding dictionaries offering sparse (ℓ_0-optimal) representations in the sense that our objective is to find dictionaries that give minimum average ℓ_2-norm of the coefficient vector used for representation. Intuitively, optimization of the ℓ_2-norm of the representation vector tends to 'distribute' the information of the data being represented among all components of the representation vector; this makes the representation robust to accidental changes in the coefficients.

- An advantage of considering the ℓ_2-cost is that it involves a norm arising from an inner product; consequently, it comes with a rich set of properties associated with it. These properties are crucially employed in this article to modify the intrinsically non-convex problem of finding an ℓ_2-optimal dictionary into an *equivalent* convex optimization problem,[1] allowing us to compute an optimal dictionary in *polynomial time* and arrive at *analytical expressions of the optimal costs*. We provide these algorithms in Sections 4.1 and 4.3.

- One more advantage of considering optimization in the ℓ_2-sense is related to the fact that the ℓ_2-cost involves the natural notion of *energy* which is extremely important in practice, especially in control theoretic applications.
- The results presented here also add to the recent developments in the advantages of representing signals/vectors using tight frames for finite-dimensional Hilbert spaces.

This article unveils as follows: In Section 2 we formally introduce our problem of finding an optimal dictionary which offers least square representation. Section 2 is the heart of this article, where we solve the problem of finding an ℓ_2-optimal dictionary, and arrive at an explicit solution. Algorithms to construct ℓ_2-optimal dictionaries are given in Section 4, where we present the proofs of our main results. The case of representing random vectors distributed uniformly on the unit sphere is treated in Subsection 2.4; we demonstrate that the ℓ_2-optimal dictionaries in this case are *finite tight frames*. The intermediate Section 3 contains results related to rank-1 decomposition of positive semidefinite matrices; these constitute essential tools for the solutions of our main results. We conclude in Section 5 with a summary of this work and future directions.

Notations

We employ standard notations in this article. As usual, $\|\cdot\|$ is the standard Euclidean norm. The $n \times n$ identity and $m \times n$ zero matrices are denoted by I_n and $O_{m\times n}$, respectively. For a matrix M we let $\mathrm{tr}(M)$ and M^+ denote its trace and Moore-Penrose pseudo-inverse, respectively. The set of $n \times n$ symmetric and positive (semi-)definite matrices with real entries is denoted by $\mathbb{S}^{n\times n}_{++}$ ($\mathbb{S}^{n\times n}_{+}$), and the set of $n \times n$ symmetric matrices with real entries is denoted by $\mathbb{S}^{n\times n}$. For a Borel probability measure μ defined on $\mathbb{R}^n$, we let $\mathsf{E}_\mu[\cdot]$ denote the corresponding mathematical expectation. The image of a map f is written as $\mathrm{image}(f)$. The gradient of a continuously differentiable function f is denoted by ∇f. For finite ordered sets A and B, we let $A \uplus B$ denote the ordered set consisting of the elements (in their order) of A followed by the elements (in their order) of B; for instance, if $A = (1, 2)$ and $B = (-5, -7)$, then $A \uplus B = (1, 2, -5, -7)$. Suppose that A and B are two ordered sets such that $B \subset A$ as sets, then $A \backslash B$ is the ordered sub-collection in A after deleting the elements of the set B. Finally, given an ordered collection of vectors $(x_i)_{i=1}^n$ in $\mathbb{R}^\nu$ with $\nu \geqslant n$ and equipped with the standard inner product, $\mathrm{Ortho}\big((x_i)_{i=1}^n\big)$ gives the result of Gram-Schmidt orthonormalization of the collection $(x_i)_{i=1}^n$ considered in the order in which they appear i.e., $x_1, x_2, \ldots, x_n$.

2. The ℓ_2-optimal dictionary problem and its solution

Let V denote an $\mathbb{R}^n$-valued random vector defined on some probability space, and having distribution (i.e., Borel probability measure,) μ. We assume that V has finite variance. Let

[1] By equivalence of two optimization problems we mean that an optimal solution to either of the problems can be obtained from an optimal solution to the other problem.

R_V denote the support of μ,[2] and let X_V be the smallest subspace of $\mathbb{R}^n$ containing R_V. Our goal is to represent the instances/samples of V with the help of a *dictionary* of vectors:

$$D_K := \bigl\{ d_i \in \mathbb{R}^n \bigm| \lVert d_i \rVert = 1 \text{ for } i = 1, \ldots, K \bigr\} \quad \text{with a given } K \geqslant n,$$

in an optimal fashion. A *representation* of an instance v of the random vector V is given by the coefficient vector $\alpha = (\alpha_1 \ \ldots \ \alpha_K)^\top$, such that

$$v = \sum_{i=1}^{K} \alpha_i d_i. \tag{4}$$

A *reconstruction* of the sample v from the representation α is carried out by taking the linear combination $\sum_{i=1}^{K} \alpha_i d_i$. We define the *cost* associated with representing v in terms of the coefficient vector α as $\sum_{i=1}^{K} \alpha_i^2$. Since the dictionary vectors $\{d_i\}_{i=1}^K$ must be able to represent any sample of V, the property that $\operatorname{span}\{d_i\}_{i=1}^K \supset R_V$ is essential. A dictionary $D_K = \{d_i\}_{i=1}^K \subset \mathbb{R}^n$ is said to be *feasible* if $\operatorname{span}\{d_i\}_{i=1}^K \supset R_V$. We denote by $\mathcal{D}_K$ the set of all feasible dictionaries.

For a feasible dictionary $D_K = \{d_i\}_{i=1}^K$, with $m := \dim \bigl(\operatorname{span}\{d_i\}_{i=1}^K\bigr)$, and for any $v \in R_V$, the linear equation (4) is satisfied by infinitely many values of α whenever $K > m$. In fact, the solution set of (4) constitutes a $(K-m)$-dimensional affine subspace of $\mathbb{R}^K$. Therefore, in order to represent a given v uniquely, one must define a mechanism of selecting a particular point from this affine subspace, thus making the coefficient vector $\alpha = (\alpha_1 \ \ldots \ \alpha_K)^\top$ a function of v. Let f denote such a function; to wit, $f(v) := \alpha$ is the coefficient vector used to represent the sample v. We call such a map $R_V \ni v \longmapsto f(v) \in \mathbb{R}^K$ a *scheme of representation.* Representation of samples of the random vector V using a dictionary D_K and a scheme f is said to be *proper* if any vector $v \in R_V$ can be uniquely represented and then exactly reconstructed back. It is clear that for proper representation of V with a dictionary D_K consisting of vectors $\{d_i\}_{i=1}^K$, the mapping $R_V \ni v \longmapsto f(v) \in \mathbb{R}^K$ should be an injection that satisfies

$$V = \begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix} f(V) \quad \mu\text{-almost surely.} \tag{5}$$

A scheme f of representation is said to be *feasible* if for some feasible dictionary $D_K := \{d_i\}_{i=1}^K \in \mathcal{D}_K$ the equality $\begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix} f(V) = V$ is satisfied almost surely. We denote by $\mathcal{F}$ the set of all feasible schemes of representation.

Given a scheme f of representation, the (random) cost associated with representing V is given by $\lVert f(V) \rVert^2$. The problem of finding an ℓ_2-optimal dictionary can now be posed as:

> Find a pair consisting of a dictionary $D_K^* \in \mathcal{D}_\mathcal{K}$ and a feasible scheme f^* of representation such that the average cost $\mathsf{E}_\mu\bigl[\lVert f^*(V) \rVert^2\bigr]$ of representation is minimal.

Here the subscript μ indicates the distribution of random vector V with respect to which the expectation is evaluated. In other words, we have the following optimization problem:

$$\begin{aligned} &\underset{D_K, f}{\text{minimize}} && \mathsf{E}_\mu\bigl[\lVert f(V) \rVert^2\bigr] \\ &\text{subject to} && \begin{cases} D_K \in \mathcal{D}_K, \\ f \in \mathcal{F}. \end{cases} \end{aligned} \tag{6}$$

[2] Recall (Parthasarathy, 2005, Theorem 2.1, Definition 2.1, pp. 27-28) that the support of μ is the set of points $z \in \mathbb{R}^n$ such that the μ-measure of every open neighbourhood of z is positive.

The problem given in (6) will be referred to as the ℓ_2-*optimal dictionary* problem. It should be noted that the ℓ_2-optimal dictionary problem is non-convex due to the constraint that the dictionary vectors $\{d_i\}_{i=1}^K$ of a feasible dictionary must be of unit length. Even if we change this constraint to $\{\|d_i\| \leqslant 1\}$ from $\{\|d_i\| = 1\}$, which makes the feasible region of dictionary vectors convex, the set of feasible schemes of representation is not known to be a convex set a priori.

In this article we solve the ℓ_2-optimal dictionary problem given in (6) in two steps:
(Step I) We let $X_V = \mathbb{R}^n$.
(Step II) We let X_V be any proper nontrivial subspace of $\mathbb{R}^n$.[3]
The remainder of this section is devoted to describing Steps I and II by exposing our main results, followed by discussions, a numerical example, and a treatment of the important case of the uniform distribution on the unit sphere of $\mathbb{R}^n$.

2.1 Step I: $X_V = \mathbb{R}^n$

If $X_V = \mathbb{R}^n$, a dictionary of vectors $D_K = \{d_i\}_{i=1}^K \subset \mathbb{R}^n$ is feasible if and only if $\|d_i\| = 1$ for all $i = 1, \ldots, K$, and $\operatorname{span}\{d_i\}_{i=1}^K = \mathbb{R}^n$. Thus, the ℓ_2-optimization problem (6) reduces to:

$$\begin{aligned} &\underset{\{d_i\}_{i=1}^K, f}{\text{minimize}} && \mathsf{E}_\mu\big[\|f(V)\|^2\big] \\ &\text{subject to} && \begin{cases} \|d_i\| = 1 \text{ for all } i = 1, \ldots, K, \\ \operatorname{span}\{d_i\}_{i=1}^K = \mathbb{R}^n, \\ \begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix} f(V) = V \quad \mu\text{-almost surely.} \end{cases} \end{aligned} \tag{7}$$

Let $\Sigma_V := \mathsf{E}_\mu[VV^\top]$. We claim that Σ_V is positive definite. Indeed, if not, then there exists a nonzero vector $x \in \mathbb{R}^n$ such that $x^\top V = 0$ almost surely, which contradicts the assumption that $X_V = \mathbb{R}^n$.

Existence and characterization of the optimal solutions to (7) is done by the following:

Theorem 1. *Consider the optimization problem* (7)*, and let* $\Sigma_V := \mathsf{E}_\mu\big[VV^\top\big]$.
○ (7) *admits an optimal solution.*
○ *The optimal value corresponding to* (7) *is* $\dfrac{\big(\operatorname{tr}(\Sigma_V^{1/2})\big)^2}{K}$.
○ *Optimal solutions of* (7) *are characterized by:*
▻ *a dictionary* $D_K^* = \{d_i^*\}_{i=1}^K$ *that is feasible for* (7) *and that satisfies*

$$\sum_{i=1}^K d_i^* {d_i^*}^\top = M^* := \frac{K}{\operatorname{tr}\big(\Sigma_V^{1/2}\big)}\, \Sigma_V^{1/2}, \tag{8}$$

and
▻ *a scheme* $f^*_{D_K^*}(v) := \begin{pmatrix} d_1^* & d_2^* & \cdots & d_K^* \end{pmatrix}^+ v$.

Moreover, all optimal dictionary-scheme pairs can be obtained via the procedure described in Algorithm 2 on p. 22.

[3] The trivial case of $X_V = \{0\}$ is discarded because then there is nothing to prove; we therefore limit ourselves to 'nontrivial' subspaces of $\mathbb{R}^n$.

2.2 Step II: X_V is a strict nontrivial subspace of $\mathbb{R}^n$

Let X_V be any proper nontrivial subspace of $\mathbb{R}^n$. In this situation it is reasonable to expect that no optimal dictionary that solves (6) contains elements that do not belong to X_V. That this indeed happens is the assertion of the following Lemma, whose proof is provided in Section 4:

Lemma 2. *Optimal solutions, if any exists, of problem* (6) *are such that the optimal dictionary vectors $\{d_i^*\}_{i=1}^K$ satisfy $d_i^* \in X_V$ for all $i = 1, \ldots, K$.*

Lemma 2 guarantees that if the problem (6) admits a solution, then the corresponding optimal dictionary vectors must be elements of X_V. This means that it is enough to optimize over dictionaries with their elements in X_V instead of the whole of $\mathbb{R}^n$. Therefore, the constraint $\operatorname{span}\{d_i\}_{i=1}^K \supset R_V$ can be equivalently stated as $\operatorname{span}\{d_i\}_{i=1}^K = X_V$.

Let the dimension of X_V be m with $m < n$, and let $\mathcal{B} = \{b_i\}_{i=1}^m$ be a basis for X_V. It should be noted that $X_V = \operatorname{image}(\Sigma_V)$, and therefore, a basis of X_V can be obtained by computing a basis of the subspace $\operatorname{image}(\Sigma_V)$. An example of such a basis of X_V is the collection of unit eigenvectors of Σ_V corresponding to its non-zero eigenvalues.

Fix a basis $\mathcal{B} = \{b_i\}_{i=1}^m$ of X_V. Let B be a matrix containing the vectors $\{b_i\}_{i=1}^m$ as its columns:

$$B := \begin{pmatrix} b_1 & b_2 & \cdots & b_m \end{pmatrix}.$$

If δ_i is the representation of the dictionary vector d_i in the basis $\mathcal{B}$, i.e., $d_i = B\delta_i$, then the constraints on the family $\{d_i\}_{i=1}^K$ get transformed to the following ones on $\{\delta_i\}_{i=1}^K$:

○ $\|d_i\|^2 = 1 \quad \Rightarrow \quad \delta_i^\top \big(B^\top B\big)\delta_i = 1$, and

○ $\operatorname{span}\{d_i\}_{i=1}^K \supset R_V \quad \Rightarrow \quad \operatorname{span}\{d_i\}_{i=1}^K = X_V \quad \Rightarrow \quad \operatorname{span}\{\delta_i\}_{i=1}^K = \mathbb{R}^m$.

We define the random vector

$$V_X := \big((B^\top B)^{-1} B^\top\big) V.$$

Then V_X is an $\mathbb{R}^m$ valued random vector which is the representation of random vector V in the basis $\mathcal{B}$. For every scheme f that is feasible for (6), let us define an associated scheme for representing samples of the random vector V_X by

$$\mathbb{R}^m \ni v \longmapsto f_X(v) := f(Bv) \in \mathbb{R}^K.$$

The conditions on feasibility of f in (6) imply that the scheme f_X is feasible if for a feasible dictionary of vectors $\{\delta_i\}_{i=1}^K$,

$$\begin{pmatrix} \delta_1 & \delta_2 & \cdots & \delta_K \end{pmatrix} f_X(V_X) = V_X \quad \mu\text{-almost surely.}$$

In other words, in contrast to the problem (6), where the optimization is carried out over vectors in $\mathbb{R}^n$, we can equivalently consider the same problem in $\mathbb{R}^m$, but with the following modified constraints:

$$\begin{aligned} &\underset{\{\delta_i\}_{i=1}^K, f_X}{\text{minimize}} && \mathsf{E}_\mu\big[\|f_X(V_X)\|^2\big] \\ &\text{subject to} && \begin{cases} \delta_i^\top \big(B^\top B\big)\delta_i = 1 \text{ for all } i = 1, \ldots, K, \\ \operatorname{span}\{\delta_i\}_{i=1}^K = \mathbb{R}^m, \\ \begin{pmatrix} \delta_1 & \delta_2 & \cdots & \delta_K \end{pmatrix} f_X(V_X) = V_X \ \ \mu\text{-almost surely.} \end{cases} \end{aligned} \tag{9}$$

In relation to the problem (9) let us define the following quantities

$$\begin{cases} \Sigma_V := \mathsf{E}_\mu[VV^\top] \\ \Sigma := (B^\top B)^{-1/2}\big(B^\top \Sigma_V B\big)(B^\top B)^{-1/2} \\ H^* := \dfrac{K}{\operatorname{tr}\big(\Sigma^{1/2}\big)}\big((B^\top B)^{-1/2}\Sigma^{1/2}(B^\top B)^{-1/2}\big). \end{cases} \tag{10}$$

Since the support of V_X is m-dimensional, we conclude from previous discussion that $\Sigma_{V_X} := \mathsf{E}_\mu\big[V_X V_X^\top\big]$ is positive definite. Since $\Sigma = (B^\top B)^{1/2}\Sigma_{V_X}(B^\top B)^{1/2}$, it follows that Σ is positive definite, which in turn implies that H^* is positive definite.

To summarize, an ℓ_2-optimal dictionary-scheme pair that solves the optimization problem (6) is equivalently obtained from an optimal solution of the problem (9), and is characterized by the following:

Theorem 3. *Consider the optimization problem* (9).
○ (9) *admits an optimal solution.*
○ *The optimal value corresponding to* (9) *is* $\dfrac{\big(\operatorname{tr}(\Sigma^{1/2})\big)^2}{K}$.
○ *Optimal solutions of* (9) *are characterized by:*
▻ *a dictionary* $D_K^* = \{\delta_i^*\}_{i=1}^K$ *that is feasible for* (9) *and that satisfies*

$$\sum_{i=1}^{K} \delta_i^* \delta_i^{*\top} = H^*, \tag{11}$$

and
▻ *a scheme* $f_X^*(u) := \begin{pmatrix}\delta_1^* & \delta_2^* & \cdots & \delta_K^*\end{pmatrix}^+ u$.
Consequently, an optimal solution of the ℓ_2-optimal dictionary problem (6) *consisting of an ℓ_2-optimal dictionary-scheme pair is given by*
○ *A collection of vectors* $\{d_i^*\}_{i=1}^K$ *defined as* $d_i^* := B\delta_i^*$ *for* $i = 1, 2, \ldots, K$, *and*
○ *the scheme* $f^*(v) := \begin{pmatrix}d_1^* & d_2^* & \cdots & d_K^*\end{pmatrix}^+ v$.
Moreover, all optimal dictionary-scheme pairs can be obtained via the procedure given in Algorithm 3 on p. 26.

2.3 Discussion and a numerical example

Remark 4. The problem (6) does not a priori hypothesize an affine/linear structure of candidate schemes. The fact that linear schemes are optimal in (6) is one of the crucial assertions of both Theorem 1 and Theorem 3.

Remark 5. Algorithmic computation of an ℓ_2-optimal dictionary relies on the second moment Σ_V of the random vector V. Complete knowledge of the distribution μ is, therefore, unnecessary. This is an advantage since in practical situations, learning/estimating Σ_V from data is comparatively less demanding than getting a description of the distribution μ itself.

Remark 6. Let $M \in \mathbb{S}_+^{n\times n}$ be such that $\operatorname{image}(M) = X_V$, let $\mathcal{B} = \{b_i\}_{i=1}^m$ be a basis for X_V evaluated as a basis for $\operatorname{image}(M)$. Let

$$B := \begin{pmatrix} b_1 & b_2 & \cdots & b_m \end{pmatrix}$$

$$
\begin{aligned}
\Sigma(M) &:= (B^\top B)^{-1/2}\big(B^\top M B\big)(B^\top B)^{-1/2} \\
H(M) &:= \frac{K}{\operatorname{tr}\Big(\big(\Sigma(M)\big)^{1/2}\Big)}\big((B^\top B)^{-1/2}\big(\Sigma(M)\big)^{1/2}(B^\top B)^{-1/2}\big).
\end{aligned}
$$

Suppose that $\{d_i\}_{i=1}^K$ and $f(\cdot)$ are the dictionary and the scheme obtained using the procedure given in Algorithm 3 using M and K as inputs. By simplifying the pseudo-inverse $\begin{pmatrix} d_1 & d_2 & \cdots & d_K\end{pmatrix}^+$ in $f(\cdot)$, the average cost $J(M)$ of representing V using the scheme $f(\cdot)$ turns out to be

$$
\begin{aligned}
J(M) &= \mathsf{E}_\mu\left[V^\top B(B^\top B)^{-1}\big(H(M)\big)^{-1}(B^\top B)^{-1}B^\top V\right] \\
&= \operatorname{tr}\Big(\big(H(M)\big)^{-1}(B^\top B)^{-1}B^\top \Sigma_V B(B^\top B)^{-1}\Big) \\
&= \operatorname{tr}\Big(\big(H(M)\big)^{-1}(B^\top B)^{-1/2}\ \Sigma\ (B^\top B)^{-1/2}\Big) \\
&= \operatorname{tr}\Big((B^\top B)^{-1/2}\big(H(M)\big)^{-1}(B^\top B)^{-1/2}\ \Sigma\Big) \\
&= \frac{1}{K}\ \operatorname{tr}\Big(\big(\Sigma(M)\big)^{1/2}\Big)\ \operatorname{tr}\Big(\big(\Sigma(M)\big)^{-1/2}\Sigma\Big).
\end{aligned} \tag{12}
$$

Let $S := \big\{T \in \mathbb{S}_+^{n\times n} \ \big|\ \operatorname{image}(T) = X_V\big\}$. Since the sequence of maps

$$
\begin{aligned}
S &\ni T \longmapsto \Sigma(T) \in \mathbb{S}_{++}^{m\times m}, \\
\mathbb{S}_{++}^{m\times m} &\ni T \longmapsto T^{1/2} \in \mathbb{S}_{++}^{m\times m}, \\
\mathbb{S}_{++}^{m\times m} &\ni T \longmapsto T^{-1}\Sigma \in \mathbb{S}_{++}^{m\times m}, \\
\mathbb{S}_{+}^{m\times m} &\ni T \longmapsto \operatorname{tr}(T) \in \mathbb{R},
\end{aligned}
$$

are, evidently, continuous, it follows at once that the map $S \ni M \longmapsto J(M) \in \mathbb{R}$ is also continuous. If $\widehat{\Sigma}_V$ denotes the estimated second moment of V, and the estimation is carried out with a large enough number of samples of V, with probability one we have $\operatorname{image}(\widehat{\Sigma}_V) = X_V$. Therefore, by continuity of $M \longmapsto J(M)$, we see at once that

$$
J(\widehat{\Sigma}_V) \xrightarrow[\widehat{\Sigma}_V \longrightarrow \Sigma_V]{} J(\Sigma_V) = \frac{\big(\operatorname{tr}(\Sigma^{1/2})\big)^2}{K}.
$$

Remark 7. The optimal average cost of representation of a random vector V is inversely proportional to the size K of the optimal dictionary, as is evident from the optimal costs in Theorems 1 and 3. To wit, the optimal average cost of representation decreases monotonically with K, which is expected.

Remark 8. ℓ_2-optimal dictionaries for representing a random vector V are also optimal for representing any scalar multiple αV of V for any $0 \neq \alpha \in \mathbb{R}$. Indeed, it is clear that H^* defined in (10) is invariant under nonzero scalar multiplications of V. Therefore, ℓ_2-optimal dictionaries are also invariant under nonzero scalar multiplications of the random vector V. This fact also follows from the observation made in Remark 4.

Remark 9. An ℓ_2-optimal dictionary as characterized by Theorem 3 appears there in the form of what is known as a *rank*-1 *decomposition* of the positive definite matrix H^*. Elements of the theory of rank-1 decompositions of positive definite matrices is discussed below in Section 3. This particular decomposition plays a crucial rôle in transforming the search space of the ℓ_2-optimal dictionary problem (7) from the set of dictionaries to the set of symmetric positive definite matrices with real entries, and translating the non-convex ℓ_2-optimal dictionary problem into a tractable convex one.

Remark 10. All ℓ_2-optimal dictionaries are unique upto rank-1 decompositions of a unique positive definite matrix that is obtained from the second moment $\mathsf{E}[VV^\top]$ of the random vector V. That is, for a given random vector whose samples are to be optimally represented, every ℓ_2-optimal dictionary is obtained from a rank-1 decomposition of a unique positive definite matrix.

Remark 11. Looking ahead at Algorithm 3, it becomes evident that non-uniqueness of optimal dictionaries can be attributed to the non-uniqueness in the selection of C in Step 5 of Algorithm 3, and the element of choice associated to the selection of p_j and p_k in Step 2 of Algorithm 1. The number of optimal solutions may be infinite depending on the distribution of the random vector V. For instance, if V is uniformly distributed over the unit sphere of $\mathbb{R}^n$ and $K = n$, then the elements in an ℓ_2-optimal dictionary form an orthonormal basis of $\mathbb{R}^n$. (The special case of uniform distribution of V over spheres is discussed in Section 2.4.) Of course, there are infinitely many orthonormal bases of $\mathbb{R}^n$ for $n \geqslant 2$.

Remark 12. From Algorithm 3 on p. 26 we can infer that by calculating the matrix B there, consisting of the eigenvectors of Σ_V corresponding to its non-zero eigenvalues, the computations of $(B^\top B)^{-1/2}$, $\Sigma^{1/2}$, and C in the decomposition given in Step 5 become straightforward. Therefore, the chief computational load in Algorithm 3 consists of eigen-decomposition of Σ_V and that in Algorithm 1 (in Step 6), both of which can be performed in polynomial time.

Example 1. Let $V = \begin{pmatrix} V_1 \\ V_2 \end{pmatrix}$ be a random vector taking values in $\mathbb{R}^2$, with V_1 and V_2 being independent random variables. Let the density functions of V_1 and V_2 be

$$\rho_{V_1}(v) = 2(v-1)\mathbb{1}_{[1,2]}(v) \quad \text{and} \quad \rho_{V_2}(v) = 2(2-v)\mathbb{1}_{[1,2]}(v),$$

respectively. The support of V is, therefore, the square $[1,2] \times [1,2]$. Elementary calculations lead to $\Sigma_V := \mathsf{E}_\rho[VV^\top] = \begin{pmatrix} 17/6 & 20/9 \\ 20/9 & 11/6 \end{pmatrix}$. We employed the procedure described in Algorithm 2 for the given matrix Σ_V and $K = 3$ in MATLAB. An optimal dictionary $\{y_1^*, y_2^*, y_3^*\}$ was obtained, with

$$y_1^* = \begin{pmatrix} 0.9789 \\ 0.2045 \end{pmatrix}, \quad y_2^* = \begin{pmatrix} 0.6792 \\ 0.7339 \end{pmatrix}, \quad y_3^* = \begin{pmatrix} 0.5870 \\ 0.8096 \end{pmatrix};$$

the optimum value of the objective function was reported to be 1.8930. This collection $\{y_i^*\}_{i=1}^3$ of optimal vectors are marked with crosses on the circumference of the unit circle shown in Figure 2. A second optimal dictionary $\{z_1^*, z_2^*, z_3^*\}$ was obtained, also using Algorithm 2, with dictionary vectors

$$z_1^* = \begin{pmatrix} 0.4214 \\ 0.9069 \end{pmatrix}, \quad z_2^* = \begin{pmatrix} 0.9284 \\ 0.3717 \end{pmatrix}, \quad z_3^* = \begin{pmatrix} 0.8513 \\ 0.5247 \end{pmatrix},$$

with an identical optimal value as in the former case. The vectors $\{z_i^*\}_{i=1}^3$ are marked with dark circles on the circumference of the unit circle in Figure 2.

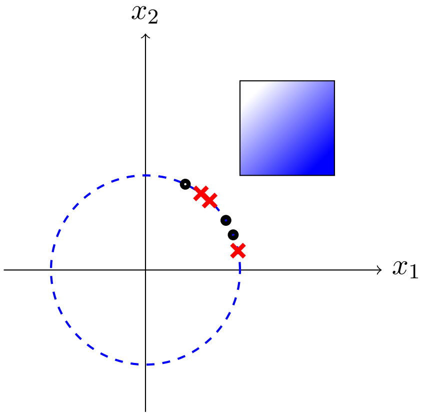

Figure 2: The two optimal dictionaries in Example 1.

It is expected that the optimal dictionary vectors are concentrated towards the bottom right corner of the support $[1,2] \times [1,2]$ (the region with strong shading in figure 2). In the optimal solution $\{z_i^*\}_{i=1}^3$, two vectors z_2^* and z_3^* point to the region where density of V is concentrated the most. Also, for the solution $\{y_i^*\}_{i=1}^3$, two vectors y_2^* and y_3^* are oriented towards the center of the square $[1,2] \times [1,2]$, with the remaining vector pointing towards the region of higher density. These results correlate positively with what may be expected out of ℓ_2-optimal dictionaries.

2.4 Uniform distribution over the unit sphere

We shall test our results on the important case of μ being the uniform distribution on the unit sphere. Note that due to (rigid) rotational symmetry of the distribution, it follows that rigid rotations of optimal dictionaries in this case are also optimal.

Let us consider a dictionary consisting of (unit) vectors that are 'close' to each other, i.e., the inner product between any two elements of the dictionary is close to 1. It is quite evident that such a dictionary is not optimal for representing uniformly distributed samples due to the fact that samples of V that are almost orthogonal to the dictionary vectors carry equal priority as any other vector but require large coefficients for their representation. It is, therefore, more natural to search for dictionaries in which the constituent vectors are 'maximally spaced out'.

Several examples of collections of vectors that are 'maximally spaced out' may be found in (Benedetto and Fickus, 2003, Section 4). Collections of vectors that are maximally far apart from each other are known to attain 'equilibria' under the actions of different kinds of forces defined and explained in (Benedetto and Fickus, 2003, Section 4) and (Saff and Kuijlaars, 1997, p. 6). Such collections of vectors are generalized by the ideas of *tight frames* as explained in (Benedetto and Fickus, 2003); see also (Christensen, 2016; Daubechies et al., 1986; Benedetto and Fickus, 2003; Zimmermann, 2001) for related information.

We recall here some standard definitions for completeness and to provide the necessary substratum for our next result. Let n, K be positive integers such that $K \geqslant n$. We say that a collection of vectors $\{x_i\}_{i=1}^K$ is a *frame* for $\mathbb{R}^n$ if there exist some constants $c, C > 0$ such that

$$c\left\|x\right\|^2 \leqslant \sum_{i=1}^{K} \langle x_i, x\rangle^2 \leqslant C\left\|x\right\|^2 \quad \text{for all } x \in \mathbb{R}^n.$$

We say that a frame $\{x_i\}_{i=1}^K \subset \mathbb{R}^n$ is *tight* if $c = C$. In addition, if $\{x_i\}_{i=1}^K \subset \mathbb{R}^n$ is a tight frame and $\|x_i\| = 1$ for all $i = 1, 2, \ldots, K$, we say that the collection $\{x_i\}_{i=1}^K$ is a *c-unit norm tight frame* (a c-UNTF).

We have the following connection between ℓ_2-optimal dictionaries and UNTFs:

Proposition 13. *A dictionary $D_K = \{d_i\}_{i=1}^K$ is optimal for representing samples of a random vector V that is uniformly distributed over the surface of the unit sphere of $\mathbb{R}^n$ if and only if the collection $\{d_i\}_{i=1}^K$ of vectors constitute a $\frac{K}{n}$-UNTF.*

Proof If V is uniformly distributed over the unit sphere, we have $\Sigma_V = \mathsf{E}[VV^\top] = \frac{1}{n}I_n$. According to Theorem 1 the collection $\{d_i\}_{i=1}^K$ is an optimal dictionary if and only if

$$\sum_{i=1}^{K} d_i d_i^\top = \frac{K}{\operatorname{tr}\left(\frac{1}{\sqrt{n}}I_n\right)}\left(\frac{1}{\sqrt{n}}I_n\right) = \frac{K}{n}I_n. \tag{13}$$

Since the family $\{d_i\}_{i=1}^K$ must span $\mathbb{R}^n$ by definition, it is a frame. The *frame operator* for the frame $\{d_i\}_{i=1}^K$ is given by (Benedetto and Fickus, 2003, Section 2)

$$\mathbb{R}^n \ni y \longmapsto S(y) := \sum_{i=1}^{K} \langle d_i, y\rangle\, d_i = \left(\sum_{i=1}^{K} d_i d_i^\top\right) y \in \mathbb{R}^n,$$

where $\langle v, w\rangle = v^\top w$ is the standard inner product in $\mathbb{R}^n$. (Benedetto and Fickus, 2003, Theorem 3.1) asserts that a collection of unit norm vectors $\{d_i\}_{i=1}^K$ forms a tight frame in $\mathbb{R}^n$ if and only if the collection is a $\frac{K}{n}$-UNTF. From (Benedetto and Fickus, 2003, Theorem 2.1) it follows that a collection of vectors $\{d_i\}_{i=1}^K$ is a $\frac{K}{n}$-UNTF if and only if

$$S = \sum_{i=1}^{K} d_i d_i^\top = \frac{K}{n} I_n. \tag{14}$$

The assertion follows from (13) and (14). ■

3. A particular class of rank-1 decompositions of matrices

We collect and establish here some results on the theory of rank-1 decompositions of matrices. While these facts will be needed for our main results, they are also of independent interest.

A standard result in matrix theory (Bhatia, 2009, p. 2) states that a symmetric positive semidefinite matrix with real entries $M \in \mathbb{S}_+^{n\times n}$, can be decomposed as $YY^\top$ for some

$Y \in \mathbb{R}^{n \times r}$, where $r := \operatorname{rank}(M)$. Let y_i indicate the i th column of the matrix Y. Then the equality $M = YY^\top$ is equivalent to

$$M = \sum_{i=1}^{r} y_i y_i^\top .$$

More generally for $K \geqslant r$, let

$$\overline{M} := \begin{pmatrix} M & O_{n \times (K-r)} \\ O_{(K-r)\times n} & I_{K-r} \end{pmatrix},$$

where O is a zero matrix of order $n \times (K-r)$. If we consider the decomposition of $\overline{M}$ as $\overline{M} = \overline{Y}\,\overline{Y}^\top$ with $\overline{Y} \in \mathbb{R}^{(n+K-r)\times K}$, and indicate by Y the upper $n \times K$ matrix block of $\overline{Y}$, we get $M = YY^\top$. In other words

$$M = \sum_{i=1}^{K} y_i y_i^\top . \tag{15}$$

There are numerous ways of decomposing positive semidefinite matrices; some of them are discussed in (Zhang, 2011, Theorem 7.3). The speciality of a particular decomposition lies in the characteristics exhibited by the vectors y_i's. A particular rank-1 decomposition which we will use to solve the ℓ_2-optimal dictionary problem is the one where for every $M \in \mathbb{S}_+^{n\times n}$ and $K \geqslant r := \operatorname{rank}(M)$ there exists a collection of vectors $\{y_i\}_{i=1}^{K} \subset \mathbb{R}^n$ that satisfy

$$M = \sum_{i=1}^{K} y_i y_i^\top \quad \text{and} \quad y_i^\top y_i = \frac{\operatorname{tr}(M)}{K} \quad \text{for all } i = 1, \ldots, K. \tag{16}$$

We are now in a position to present Algorithm 1 and its associated Theorem 14, whose corollaries will give us the needed rank-1 decomposition of (16). We mention that Algorithm 1 is, in principle, similar to Procedure 1 of (Sturm and Zhang, 2003), and in particular, the assertions of Theorem 14 and its corollaries can be obtained by applying (Sturm and Zhang, 2003, Proposition 3 and Corollary 4) via some straightforward modifications. However, we provide the complete proofs here for the sake of completeness.

Theorem 14. *For any matrix $\Lambda \in \mathbb{R}^{n\times n}$ there exists an orthonormal collection $(x_i)_{i=1}^{n} \subset \mathbb{R}^n$ of vectors satisfying*

$$x_i^\top \Lambda x_i = \frac{\operatorname{tr}(\Lambda)}{n} \qquad \textit{for all } \; i = 1, \ldots, n,$$

Moreover, such a collection can be obtained from Algorithm 1.

Proof First we establish that the collection of vectors $(x_i)_{i=1}^{n-1}$ contained in S_{n-1} (recall that S_{n-1} is generated in the **for** loop in the Algorithm 1,) are orthonormal, and satisfy $x_i^\top \Lambda x_i = \frac{\operatorname{tr}(\Lambda)}{n}$ for $i = 1, \ldots, n-1$. We shall prove this by induction on i.

The induction base: For $i = 1$, we have $P_1 = (e_1, e_2, \ldots, e_n)$. Since $\sum_{m=1}^{n} e_m^\top \Lambda e_m = \operatorname{tr}(\Lambda)$, vectors $p_j, p_k \in P_1$ exist such that $p_j^\top \Lambda p_j \leqslant \frac{\operatorname{tr}(\Lambda)}{n} \leqslant p_k^\top \Lambda p_k$. We solve for θ in the equation

$$g_{p_j;p_k}(\theta) := \frac{\big((1-\theta)p_j + \theta p_k\big)^\top \Lambda \big((1-\theta)p_j + \theta p_k\big)}{\big((1-\theta)^2 + \theta^2\big)} = \frac{\operatorname{tr}(\Lambda)}{n}. \tag{17}$$

Algorithm 1: Calculation of orthonormal bases à la Theorem 14

```
Input: A matrix $\Lambda \in \mathbb{R}^{n\times n}$.
Output: An orthonormal collection of vectors $(x_i)_{i=1}^n \subset \mathbb{R}^n$ such that $x_i^\top \Lambda x_i = \frac{\mathrm{tr}(\Lambda)}{n}$
        for all $i = 1, \dots, n$.
Initialize quantities by $S_0 = \varnothing$, $i = 1$.
for $i$ from 1 to $(n-1)$
do
    $S_i' = S_{i-1} \uplus (e_1, e_2, \dots, e_n)$.
    $P_i = \mathrm{Ortho}(S_i') \backslash S_{i-1}$.
    Find $p_j, p_k \in P_i$ such that $p_j^\top \Lambda p_j \leqslant \frac{\mathrm{tr}(\Lambda)}{n} \leqslant p_k^\top \Lambda p_k$.
    Let $\Theta \in [0,1]$ be a solution of the equation (in $\theta$)
        $$\big((1-\theta)p_j + \theta p_k\big)^\top \Lambda \big((1-\theta)p_j + \theta p_k\big) = \frac{\mathrm{tr}(\Lambda)}{n}\big((1-\theta)^2 + \theta^2\big)$$
    Define $x_i := \frac{(1-\Theta)p_j + \Theta p_k}{((1-\Theta)^2+\Theta^2)^{1/2}}$.
    Define $S_i := S_{i-1} \uplus (x_i)$.
end for loop
$S_n' = S_{n-1} \uplus (e_1, e_2, \dots, e_n)$.
Output $S_n := \mathrm{Ortho}(S_n')$.
```

We know that a solution exists in $[0,1]$ because for $\theta = 0$ we have

$$g_{p_j;p_k}(0) = \left[\frac{((1-\theta)p_j + \theta p_k)^\top \Lambda ((1-\theta)p_j + \theta p_k)}{((1-\theta)^2 + \theta^2)}\right]_{\theta=0} = p_j^\top \Lambda p_j \leqslant \frac{\mathrm{tr}(\Lambda)}{n},$$

for $\theta = 1$ we have

$$g_{p_j;p_k}(1) = \left[\frac{((1-\theta)p_j + \theta p_k)^\top \Lambda ((1-\theta)p_j + \theta p_k)}{((1-\theta)^2 + \theta^2)}\right]_{\theta=1} = p_k^\top \Lambda p_k \geqslant \frac{\mathrm{tr}(\Lambda)}{n},$$

and $g_{p_j;p_k}(\cdot)$ is a continuous function of θ. Let Θ be such a solution. Then, following the notation in Algorithm 1, we have

$$x_1 := \frac{(1-\Theta)p_j + \Theta p_k}{\sqrt{(1-\Theta)^2 + \Theta^2}}.$$

Since p_j, p_k are elements of P_1, they are orthonormal; therefore,

$$\|x_1\| = \frac{\sqrt{(1-\Theta)^2 \|p_j\|^2 + \Theta^2 \|p_k\|^2}}{\sqrt{(1-\Theta)^2 + \Theta^2}} = 1,$$

and since Θ is a solution of equation (17) we have

$$x_1^\top \Lambda x_1 = \frac{\mathrm{tr}(\Lambda)}{n}.$$

Induction hypothesis: Assume that for some i between 1 and $n-1$ the collection $S_i = (x_\ell)_{\ell=1}^{i}$ is orthonormal, and satisfies

$$x_\ell^\top \Lambda x_\ell = \frac{\operatorname{tr}(\Lambda)}{n} \quad \text{for all } \ell = 1, \ldots, i.$$

Induction step: In view of the induction hypothesis, we define

$$S'_{i+1} \coloneqq S_i \uplus (e_1, e_2, \ldots, e_n) = (x_1, x_2, \ldots x_i, e_1, e_2, \ldots, e_n),$$

and compute

$$\begin{aligned} \operatorname{Ortho}(S'_{i+1}) &= (x_1, x_2, \ldots, x_i, p_1, p_2, \ldots, p_{n-i}), \\ P_{i+1} &= (p_1, p_2, \ldots, p_{n-i}) \end{aligned}$$

as in Algorithm 1. Since the collection $(x_\ell)_{\ell=1}^{i} \uplus (p_\ell)_{\ell=1}^{n-i}$ is an orthonormal basis for $\mathbb{R}^n$, we have

$$\sum_{\ell=1}^{i} x_\ell^\top \Lambda x_\ell + \sum_{\ell=1}^{n-i} p_\ell^\top \Lambda p_\ell = \operatorname{tr}(\Lambda),$$

leading to

$$\sum_{\ell=1}^{n-i} p_\ell^\top \Lambda p_\ell = \frac{(n-i)}{n} \operatorname{tr}(\Lambda).$$

Thus, there exist vectors $p_j, p_k \in P_{i+1}$ such that $p_j^\top \Lambda p_j \leqslant \frac{\operatorname{tr}(\Lambda)}{n} \leqslant p_k^\top \Lambda p_k$. Let us consider the equation

$$g_{p_j,p_k}(\theta) \coloneqq \frac{((1-\theta)p_j + \theta p_k)^\top \Lambda ((1-\theta)p_j + \theta p_k)}{((1-\theta)^2 + \theta^2)} = \frac{\operatorname{tr}(\Lambda)}{n} \tag{18}$$

in θ. From arguments given in the case of $i = 1$, we know that a solution Θ of (18) exists on $[0, 1]$. We define

$$x_{i+1} \coloneqq \frac{(1-\Theta)p_j + \Theta p_k}{\sqrt{(1-\Theta)^2 + \Theta^2}}.$$

Since p_j, p_k are orthogonal to the vectors $(x_\ell)_{\ell=1}^{i}$, so is any linear combination of p_j, p_k. Therefore, x_{i+1} is orthogonal to the vectors $(x_\ell)_{\ell=1}^{i}$, which, along with the fact that

$$\|x_{i+1}\| = \frac{\sqrt{(1-\Theta)^2 \|p_j\|^2 + \Theta^2 \|p_k\|^2}}{\sqrt{(1-\Theta)^2 + \Theta^2}} = 1,$$

makes the collection $(x_\ell)_{\ell=1}^{i+1}$ orthonormal. Also, since Θ is a solution of (18), we get

$$x_{i+1}^\top \Lambda x_{i+1} = \frac{\operatorname{tr}(\Lambda)}{n}.$$

Therefore, by mathematical induction, we conclude that the collection $(x_i)_{i=1}^{n-1}$ contained in S_{n-1} has the required properties.

Finally, in the 4th and 5th steps of Algorithm 1, we get

$$S'_n = (x_1, x_2, \ldots, x_{n-1}, e_1, e_2, \ldots, e_n),$$

and

$$\mathrm{Ortho}(S'_n) = (x_1, x_2, \ldots, x_{n-1}, x_n).$$

By construction, $(x_\ell)_{\ell=1}^n$ is an orthonormal collection, implying that $\sum_{i=1}^n x_i^\top \Lambda x_i = \mathrm{tr}(\Lambda)$. In turn, this leads to

$$\begin{aligned} x_n^\top \Lambda x_n &= \sum_{i=1}^n x_i^\top \Lambda x_i - \sum_{i=1}^{n-1} x_i^\top \Lambda x_i \\ &= \mathrm{tr}(\Lambda) - \left(\frac{n-1}{n}\right) \mathrm{tr}(\Lambda) \\ &= \frac{\mathrm{tr}(\Lambda)}{n}. \end{aligned}$$

Thus, Algorithm 1 yields a collection of orthonormal vectors $(x_i)_{i=1}^n$ such that

$$x_i^\top \Lambda x_i = \frac{\mathrm{tr}(\Lambda)}{n} \quad \text{for all } i = 1, 2, \ldots, n,$$

thereby completing the proof. ■

Corollary 15 (Rank-1 decomposition). *Let $X \in \mathbb{S}_+^{n\times n}$, define $r := \mathrm{rank}(X)$, and let $T \in \mathbb{S}^{n\times n}$. There exists a collection of vectors $\{x_i\}_{i=1}^r \subset \mathbb{R}^n$ such that*

$$X = \sum_{j=1}^r x_j x_j^\top, \quad \textit{and} \quad x_i^\top T x_i = \frac{1}{r}\,\mathrm{tr}(XT) \ \textit{for all} \ i = 1, \ldots, r.$$

Proof We know (Bhatia, 2009, p. 2) that any symmetric positive semidefinite matrix X with real entries and of rank r can be decomposed as $CC^\top$ where $C \in \mathbb{R}^{n\times r}$. Let us define $\Lambda \in \mathbb{R}^{r\times r}$ as $\Lambda := C^\top T C$. According to Theorem 14 a collection of orthonormal vectors $\{y_i\}_{i=1}^r \subset \mathbb{R}^r$ can be obtained such that

$$y_i^\top C^\top T\, C y_i = y_i^\top \Lambda y_i = \frac{\mathrm{tr}(\Lambda)}{r}.$$

We define a collection $\{x_i\}_{i=1}^r \subset \mathbb{R}^r$ by $x_i := C y_i$ for $i = 1, \ldots, r$. Then

$$\sum_{i=1}^r x_i x_i^\top = C\left(\sum_{i=1}^r y_i y_i^\top\right) C^\top = C I_r C^\top = X.$$

Moreover, for every $i = 1, \ldots, r$,

$$x_i^\top T x_i = y_i^\top C^\top T C y_i = \frac{\mathrm{tr}(\Lambda)}{r} = \frac{1}{r}\,\mathrm{tr}(C^\top T C) = \frac{1}{r}\,\mathrm{tr}(XT).$$

The assertion follows. ■

Corollary 15 is generalized slightly by the following one; we shall employ this particular form to solve the ℓ_2-optimal dictionary problem in Theorem 1.

Corollary 16. *Let $M \in \mathbb{S}_+^{n\times n}$ and define $r := \operatorname{rank}(M)$. Let $A \in \mathbb{S}^{n\times n}$ and $K \geqslant r$ be given. There exists a collection of vectors $\{y_i\}_{i=1}^K \subset \mathbb{R}^n$ such that*

$$M = \sum_{j=1}^{K} y_j y_j^\top, \quad \text{and} \quad y_i^\top A y_i = \frac{1}{K}\operatorname{tr}(MA) \;\text{ for all }\; i = 1, \ldots, K. \tag{19}$$

Proof Let us consider the square matrices X, T of order $K + n - r$ in Corollary 15 to be

$$X := \begin{pmatrix} M & O_{n\times(K-r)} \\ O_{(K-r)\times n} & I_{K-r} \end{pmatrix} \quad \text{and} \quad T := \begin{pmatrix} A & O_{n\times(K-r)} \\ O_{(K-r)\times n} & O_{(K-r)\times(K-r)} \end{pmatrix}.$$

Then $\operatorname{rank}(X) = K$ by construction. Therefore, vectors $\{x_i\}_{i=1}^K \subset \mathbb{R}^{n+K-r}$ exist satisfying the properties in Corollary 15. Let us denote $\mathbb{R}^n \ni y_i := \begin{pmatrix} x_{i1} & \ldots & x_{in} \end{pmatrix}^\top$ for $i = 1, \ldots, K$; in other words, y_i is the vector formed by the first n components of x_i. Then

$$\sum_{i=1}^{K} y_i y_i^\top = M,$$

and for any $i = 1, \ldots, K$,

$$y_i^\top A y_i = x_i^\top T x_i = \frac{1}{K}\operatorname{tr}(XT) = \frac{1}{K}\operatorname{tr}(MA).$$

The assertion follows at once. ■

4. Proofs of Theorem 1, Lemma 2, and Theorem 3

4.1 Proof of Theorem 1

Proof For a given dictionary $D_K \in \mathcal{D}_K$ of vectors $\{d_i\}_{i=1}^K$ that is feasible for (7), let us define a scheme of representation

$$\mathbb{R}^n \ni v \longmapsto f^*_{D_K}(v) := \begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix}^+ v \in \mathbb{R}^K.$$

Quite clearly, $\begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix} f^*_{D_K}(v) = v$ for any $v \in \mathbb{R}^n$ by the definition of the pseudo-inverse because if $\operatorname{span}\{d_i\}_{i=1}^K = \mathbb{R}^n$, then $\begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix}^+ v$ solves the equation $\begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix} x = v$. Therefore,

$$\begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix} f^*_{D_K}(V) = V \quad \mu\text{-almost surely.}$$

We know that $f^*_{D_K}(v) = \begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix}^+ v$ is the solution of the least squares problem

$$\begin{aligned} &\underset{x\in\mathbb{R}^K}{\text{minimize}} && \|x\|^2 \\ &\text{subject to} && \begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix} x = v. \end{aligned}$$

Therefore, for an arbitrary $f \in \mathcal{F}$ such that $\begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix} f(v) = v$ for all $v \in \mathbb{R}^n$, we must have

$$\left\| f^*_{D_K}(v) \right\|^2 \leqslant \| f(v) \|^2 \quad \text{for all } v \in \mathbb{R}^n.$$

Therefore,

$$\left\| f^*_{D_K}(V) \right\|^2 \leqslant \| f(V) \|^2 \quad \mu\text{-almost surely,}$$

and hence,

$$\mathsf{E}_\mu\big[\left\| f^*_{D_K}(V) \right\|^2 \big] \leqslant \mathsf{E}_\mu\big[\| f(V) \|^2 \big].$$

Minimizing over all feasible dictionaries and schemes, we get

$$\inf_{D_K \in \mathcal{D}_\mathcal{K}} \mathsf{E}_\mu\big[\left\| f^*_{D_K}(V) \right\|^2 \big] \leqslant \inf_{\substack{D_K \in \mathcal{D}_K, \\ f \in \mathcal{F}}} \mathsf{E}_\mu\big[\| f(V) \|^2 \big] \tag{20}$$

The problem on the left-hand side of the inequality (20) is

$$\begin{aligned} & \underset{\{d_i\}_{i=1}^K}{\text{minimize}} && \mathsf{E}_\mu\big[\left\| f^*_{D_K}(V) \right\|^2 \big] \\ & \text{subject to} && \begin{cases} \|d_i\| = 1 \text{ for all } i = 1, \ldots, K, \\ \operatorname{span}\{d_i\}_{i=1}^K = \mathbb{R}^n. \end{cases} \end{aligned} \tag{21}$$

From (20) we can conclude that the optimal value, if it exists, of problem (7) is bounded below by the optimal value, if it exists, of the one given in (21). Our strategy is to demonstrate that optimization problem (21) admits a solution, and we shall furnish a feasible solution of (7) that achieves a value of the objective function that is equal to the optimal value of the problem (21). This will solve (7).

Let $D := \begin{pmatrix} d_1 & d_2 & \cdots & d_K \end{pmatrix}$. The objective function in (21) can be computed as

$$\begin{aligned} \mathsf{E}_\mu\big[\left\| f^*_{D_K}(V) \right\|^2 \big] &= \mathsf{E}_\mu\big[\left\| D^+ V \right\|^2 \big] \\ &= \mathsf{E}_\mu\big[V^\top (D^+)^\top D^+ V \big] \\ &= \mathsf{E}_\mu\big[V^\top \big(D^\top (DD^\top)^{-1} \big)^\top \big(D^\top (DD^\top)^{-1} \big) V \big] \\ &= \mathsf{E}_\mu\big[V^\top (DD^\top)^{-1} DD^\top (DD^\top)^{-1} V \big] \\ &= \mathsf{E}_\mu\big[V^\top (DD^\top)^{-1} V \big] \\ &= \mathsf{E}_\mu\big[\operatorname{tr}(V^\top (DD^\top)^{-1} V) \big] \\ &= \mathsf{E}_\mu\big[\operatorname{tr}(VV^\top (DD^\top)^{-1}) \big] \\ &= \operatorname{tr}\big(\mathsf{E}_\mu\big[VV^\top \big] (DD^\top)^{-1} \big). \end{aligned}$$

Letting $\Sigma_V := \mathsf{E}_\mu\big[VV^\top \big]$ and writing $DD^\top = \sum_{i=1}^K d_i d_i^\top$ the optimization problem (21) is rephrased as

$$\begin{aligned} & \underset{\{d_i\}_{i=1}^K}{\text{minimize}} && \operatorname{tr}\left(\Sigma_V \left(\sum_{i=1}^K d_i d_i^\top \right)^{-1} \right) \\ & \text{subject to} && \begin{cases} \|d_i\| = 1 \text{ for all } i = 1, \ldots, K, \\ \operatorname{span}\{d_i\}_{i=1}^K = \mathbb{R}^n. \end{cases} \end{aligned} \tag{22}$$

Let S be the feasible set for the problem in (22). At first (22) appears to be non-convex. Let us demonstrate that the objective function of (22) is convex in $DD^\top$. We know that whenever Σ_V is a positive definite matrix,

$$\mathrm{tr}(\Sigma_V M^{-1}) = \mathrm{tr}\Big(\Sigma_V^{1/2} M^{-1} \Sigma_V^{1/2}\Big) = \mathrm{tr}\Big(\big(\Sigma_V^{-1/2} M \Sigma_V^{-1/2}\big)^{-1}\Big).$$

From (Bhatia, 1997, p. 113 and Exercise V.1.15, p. 117) we know that inversion of a matrix is a *matrix convex* map on the set of positive definite matrices. Therefore, for any $\theta \in [0,1]$ and $M_1, M_2 \in \mathbb{S}_{++}^{n\times n}$ we have

$$\begin{aligned}&\Big(\Sigma_V^{-1/2}\big((1-\theta)M_1 + \theta M_2\big)\Sigma_V^{-1/2}\Big)^{-1}\\ &\qquad = \Big((1-\theta)\left(\Sigma_V^{-1/2} M_1 \Sigma_V^{-1/2}\right) + \theta\left(\Sigma_V^{-1/2} M_2 \Sigma_V^{-1/2}\right)\Big)^{-1}\\ &\qquad\qquad \preceq (1-\theta)\left(\Sigma_V^{-1/2} M_1 \Sigma_V^{-1/2}\right)^{-1} + \theta\left(\Sigma_V^{-1/2} M_2 \Sigma_V^{-1/2}\right)^{-1},\end{aligned} \tag{23}$$

where $A \preceq B$ implies that $B - A$ is positive semidefinite. Since $\mathrm{tr}(\cdot)$ is a *linear functional* over the set of $n \times n$ matrices we have

$$\begin{aligned}\mathrm{tr}\Big(\Sigma_V\big((1-\theta)M_1 + \theta M_2\big)^{-1}\Big) &= \mathrm{tr}\Big(\Big(\Sigma_V^{-1/2}\big((1-\theta)M_1 + \theta M_2\big)\Sigma_V^{-1/2}\Big)^{-1}\Big)\\ &\leqslant (1-\theta)\,\mathrm{tr}\Big(\Big(\Sigma_V^{-1/2} M_1 \Sigma_V^{-1/2}\Big)^{-1}\Big) + \theta\,\mathrm{tr}\Big(\Big(\Sigma_V^{-1/2} M_2 \Sigma_V^{-1/2}\Big)^{-1}\Big)\\ &\leqslant (1-\theta)\,\mathrm{tr}(\Sigma_V M_1^{-1}) + \theta\,\mathrm{tr}(\Sigma_V M_2^{-1}).\end{aligned}$$

In other words, the function $M \longmapsto \mathrm{tr}(\Sigma_V M^{-1})$ is a convex function on the set of symmetric and positive definite matrices. Moreover, we know that for a collection $\{d_i\}_{i=1}^K$ that is feasible for (22),

$$\mathcal{D}_K \ni \{d_i\}_{i=1}^K \longmapsto h(d_1, \ldots, d_K) \coloneqq \sum_{i=1}^K d_i d_i^\top$$

maps into the set of positive definite matrices. Therefore, the objective function in (22) is a convex function on image(h). This allows us to translate the feasible set of the optimization problem (22) to the set of matrices M formed by all feasible collections $\{d_i\}_{i=1}^K$, i.e., on $h(\mathcal{D}_K)$.

Let $R \coloneqq \big\{M \in \mathbb{S}_{++}^{n\times n} \;\big|\; \mathrm{tr}(M) = K\big\}$. On the one hand, from Corollary 16 with $A = I_n$, we know that any symmetric and positive definite matrix $M \in R$ can be decomposed as

$$M = \sum_{i=1}^K d_i d_i^\top \quad \text{with } \|d_i\| = \sqrt{\frac{\mathrm{tr}(M)}{K}} = 1 \text{ for all } i = 1, \ldots, K.$$

The fact that M is positive definite implies that $\mathrm{span}\{d_i\}_{i=1}^K = \mathbb{R}^n$. Therefore, $\{d_i\}_{i=1}^K \in \mathcal{D}_K$ and $M = h(d_1, \ldots, d_K)$, which implies that

$$R \subset h(S). \tag{24}$$

On the other hand, for any collection of vectors $\{d_i\}_{i=1}^K \in \mathcal{D}_K$, we have $h(d_1, \ldots, d_K) = \sum_{i=1}^K d_i d_i^\top \in \mathbb{S}_{++}^{n\times n}$ and $\operatorname{tr}\bigl(h(d_1, \ldots, d_K)\bigr) = \sum_{i=1}^K d_i^\top d_i = K$. Therefore, by definition of R,

$$h(S) \subset R. \tag{25}$$

From (24) and (25) we conclude that $h(\mathcal{D}_K) = R$. The optimization problem (22) is, therefore, equivalent to the one where the feasible set is the set of positive definite matrices with trace K, i.e., from (22),

$$\begin{aligned} &\underset{M\in\mathbb{S}_{++}^{n\times n}}{\text{minimize}} && \operatorname{tr}\bigl(\Sigma_V M^{-1}\bigr) \\ &\text{subject to} && \operatorname{tr}(M) - K = 0. \end{aligned} \tag{26}$$

The optimization problem in (26) is convex since its objective function is convex (as a function of M) and the feasible region is the intersection of a convex cone $\mathbb{S}_{++}^{n\times n}$ and the affine space $\bigl\{M \in \mathbb{R}^{n\times n} \bigm| \operatorname{tr}(M) - K = 0\bigr\}$. In the light of (Boyd and Vandenberghe, 2004, p. 244) it follows that (26) can be solved by considering just the first order optimality conditions. These first order optimality conditions are expressed in terms of a Lagrangian

$$L(M, \gamma) := \operatorname{tr}(M^{-1}\Sigma_V) + \gamma\bigl(\operatorname{tr}(M) - K\bigr),$$

containing a KKT multiplier γ at an optimal point M^* as

$$\begin{aligned} 0 = \nabla_M L(M^*, \gamma) &= \nabla_M \Bigl(\operatorname{tr}(M^{-1}\Sigma_V) + \gamma\bigl(\operatorname{tr}(M) - K\bigr)\Bigr)\Big|_{M=M^*} \\ &= -\bigl((M^*)^{-1}\Sigma_V(M^*)^{-1}\bigr)^\top + \gamma I_n. \end{aligned} \tag{27}$$

But since $M^*, \Sigma_V \in \mathbb{S}_{++}^{n\times n}$, by symmetry it follows that $(M^*)^{-1}\Sigma_V(M^*)^{-1} = \gamma I_n$, leading to

$$\Sigma_V = \gamma (M^*)^2. \tag{28}$$

Since $\Sigma_V \neq O_{n\times n}$, we get $\gamma \neq 0$, and write M^* as

$$M^* = \frac{1}{\sqrt{\gamma}} \Sigma_V^{1/2}.$$

To evaluate γ we use the fact that by construction $K = \operatorname{tr}(M^*) = \frac{1}{\sqrt{\gamma}} \operatorname{tr}\bigl(\Sigma_V^{1/2}\bigr)$, which gives

$$\gamma = \left(\frac{\operatorname{tr}\bigl(\Sigma_V^{1/2}\bigr)}{K}\right)^2.$$

In other words, the final expression of the optimizer M^* in the problem (26) is

$$M^* = \frac{K}{\operatorname{tr}\bigl(\Sigma_V^{1/2}\bigr)} \Sigma_V^{1/2}. \tag{29}$$

It follows that the optimal value of the problem (26) (and therefore of (22)) is $\dfrac{\bigl(\operatorname{tr}(\Sigma_V^{1/2})\bigr)^2}{K}$. Therefore, this value must be a lower bound of the optimal value, if it exists, for the problem (7).

Employing Corollary 16 with $A = I_n$, we decompose M^* as

$$M^* = \sum_{i=1}^{K} d_i^* d_i^{*\top} \quad \text{with } \|d_i^*\| = 1 \text{ for each } i = 1, \ldots, K. \tag{30}$$

Let us consider the dictionary D_K^* consisting of the vectors $\{d_i^*\}_{i=1}^K$ obtained above. Since $X_V = \mathbb{R}^n$, the matrices $\Sigma_V, \Sigma_V^{1/2}$, and M^* are of rank n, and therefore, $\mathrm{span}\{d_i^*\}_{i=1}^K = \mathbb{R}^n$. Along with the fact that $\|d_i^*\| = 1$, we see that the dictionary D_K^* of vectors $\{d_i^*\}_{i=1}^K$ is feasible for the problem (7).

Let us define the scheme

$$\mathbb{R}^n \ni v \longmapsto f_{D_K^*}^*(v) := \begin{pmatrix} d_1^* & d_2^* & \cdots & d_K^* \end{pmatrix}^+ v \in \mathbb{R}^K.$$

It is evident that this scheme $f_{D_K^*}^*$ is feasible for (7). But then the objective function in (7) evaluated at $D_K = D_K^*$ and $f = f_{D_K^*}^*$ must be equal to $\dfrac{\left(\mathrm{tr}(\Sigma_V^{1/2})\right)^2}{K}$. Since this particular value is also a lower bound for the optimal value of (7), the problem (7) is solvable. An optimal dictionary-scheme pair is given by

$$\begin{cases} D_K^* = \{d_i^*\}_{i=1}^K \text{ obtained from the decomposition (30), and} \\ \mathbb{R}^n \ni v \longmapsto f^*(v) := \begin{pmatrix} d_1^* & d_2^* & \cdots & d_K^* \end{pmatrix}^+ v \in \mathbb{R}^K. \end{cases} \tag{31}$$

The proof is now complete. ■

We provide the Algorithm 2 that computes optimal dictionary-scheme pairs for the case $X_V = \mathbb{R}^n$. The inputs to the algorithm are the matrix Σ_V and the size K of a dictionary:

```
Algorithm 2: $\ell_2$-optimal dictionary for the case $X_V = \mathbb{R}^n$.
Input: A matrix $\Sigma_V \in \mathbb{S}_{++}^{n\times n}$ and a number $K \geqslant n$.
Output: An $\ell_2$-optimal dictionary-scheme pair $\left(\{d_i^*\}_{i=1}^K, f^*\right)$.
Define $M_1 := \frac{K}{\mathrm{tr}\left(\Sigma_V^{1/2}\right)} \Sigma_V^{1/2}$.
Define $M_2 := \begin{pmatrix} M_1 & O_{n\times(K-n)} \\ O_{(K-n)\times n} & I_{K-n} \end{pmatrix}$, $A := \begin{pmatrix} I_n & O_{n\times(K-n)} \\ O_{(K-n)\times n} & O_{(K-n)\times(K-n)} \end{pmatrix}$
Compute $C \in \mathbb{R}^{K\times K}$ such that $M_2 = CC^\top$.
Define $\Lambda \in \mathbb{R}^{K\times K}$ by $\Lambda := C^\top AC$, and apply Algorithm 1 to get a collection of
    vectors $\{x_i\}_{i=1}^K \subset \mathbb{R}^K$.
Define the collection $\{v_i\}_{i=1}^K \subset \mathbb{R}^K$ by $v_i := Cx_i$ for $i = 1, \ldots, K$.
Define the $\ell_2$-optimal dictionary $\{d_i^*\}_{i=1}^K \subset \mathbb{R}^n$ such that the $j^{th}$ component of $d_i^*$ is
    given by $d_i^*(j) := v_i(j)$ for $j = 1, \ldots, n$ and for every $i = 1, \ldots, K$.
Define the optimal scheme $\mathbb{R}^n \ni v \longmapsto f^*(v) := \begin{pmatrix} d_1^* & d_2^* & \cdots & d_K^* \end{pmatrix}^+ v$.
```

4.2 Proof of Lemma 2

Proof We argue by contradiction. Suppose that the assertion of the Lemma is false. If we denote by x_i the orthogonal projection of d_i on X_V and by y_i the orthogonal projection of

d_i on the orthogonal complement of X_V, we must have $\|x_i\| < 1$ for at least one value of i. If f is an optimal scheme of representation, feasibility of f gives, for any $v \in R_V$,

$$\begin{aligned} v = \sum_{i=1}^{K} d_i f_i(v) &= \left(\sum_{i=1}^{K} x_i f_i(v)\right) + \left(\sum_{i=1}^{K} y_i f_i(v)\right) \\ &= \sum_{\substack{i=1, \\ \|x_i\| \neq 0}}^{K} x_i f_i(v) + 0. \end{aligned} \tag{32}$$

Fix a unit vector $x \in X_V$, and define a dictionary $\{d_i^*\}_{k=1}^{K}$ by

$$d_i^* := \begin{cases} \dfrac{x_i}{\|x_i\|} & \text{if } \|x_i\| \neq 0, \\ x & \text{otherwise.} \end{cases}$$

Then clearly

$$\operatorname{span}\{d_i^*\}_{i=1}^{K} \supset \operatorname{span}\{x_i\}_{i=1}^{K} \supset R_V \quad \text{and} \quad \|d_i^*\| = 1 \text{ for all } i = 1, \ldots, K.$$

In other words, the dictionary of vectors $\{d_i^*\}_{i=1}^{K}$ is feasible for the problem (6). Let us now define a scheme f^* by

$$\mathbb{R}^n \ni v \longmapsto f^*(v) := \operatorname{diag}\{\|x_1\|, \|x_2\|, \ldots, \|x_K\|\} f(v) \in \mathbb{R}^K.$$

For any $v \in R_V$, using the dictionary consisting of vectors $\{d_i^*\}_{i=1}^{K}$ we get

$$\sum_{i=1}^{K} d_i^* f_i^*(v) = \sum_{i=1}^{K} d_i^* \|x_i\| f_i(v) = \sum_{\substack{i=1, \\ \|x_i\| \neq 0}}^{K} \frac{x_i}{\|x_i\|} \|x_i\| f_i(v) = v, \tag{33}$$

where the last equality follows from (32). Thus, $f^*(\cdot)$ along with the dictionary of vectors $\{d_i^*\}_{i=1}^{K}$ is feasible for problem (6). But for any $v \in R_V$ we have

$$\|f^*(v)\|^2 = \sum_{i=1}^{K} \big(f_i^*(v)\big)^2 = \sum_{i=1}^{K} \|x_i\|^2 \big(f_i(v)\big)^2 < \sum_{i=1}^{K} \big(f_i(v)\big)^2 = \|f(v)\|^2,$$

where the inequality is due to the fact that $\|x_i\| < 1$ for at least one i. This contradicts the assumption that the pair $\{d_i\}_{i=1}^{K}$ along with the scheme f is optimal for (6). ■

4.3 Proof of Theorem 3

Proof The problem (9) is similar to problem (7) except for the first constraint. In (7) we optimize over vectors taking values on the surface of the unit sphere, whereas in (9) we optimize over vectors taking values on the surface of the ellipsoid $\big\{x \in \mathbb{R}^m \;\big|\; x^\top (B^\top B) x = 1\big\}$. Following the arguments in the proof of Theorem 1 till (22), one can conclude that the optimal value, if it exists, of problem (9) is bounded below by the optimal value, if it exists, of the problem

$$
\begin{aligned}
&\underset{\{\delta_i\}_{i=1}^K}{\text{minimize}} && \operatorname{tr}\Bigg(\Sigma_{V_X}\bigg(\sum_{i=1}^{K}\delta_i\delta_i^\top\bigg)^{-1}\Bigg)\\
&\text{subject to} && \begin{cases}\delta_i^\top(B^\top B)\delta_i = 1 \text{ for all } i = 1, 2, \ldots, K,\\ \operatorname{span}\{\delta_i\}_{i=1}^K = \mathbb{R}^m,\end{cases}
\end{aligned}
\tag{34}
$$

where $\Sigma_{V_X} := \mathsf{E}_\mu\big[V_X V_X^\top\big] = \big((B^\top B)^{-1}B^\top\big)\,\mathsf{E}_\mu\big[VV^\top\big]\big((B^\top B)^{-1}B^\top\big)^\top$.

Let us define:

○ S to be the feasible region of the problem (34),

○ $R := \big\{H \in \mathbb{S}_{++}^{m\times m} \;\big|\; \operatorname{tr}\big(H(B^\top B)\big) = K\big\}$, and

○ the map $\big(\mathbb{R}^m\big)^K \ni (\delta_1, \delta_2, \ldots, \delta_K) \longmapsto h(\delta_1, \delta_2, \ldots, \delta_K) := \sum_{i=1}^K \delta_i\delta_i^\top \in \mathbb{S}_+^{m\times m}$.

From Corollary 16 we see that for every $H \in R$ there exists a collection of vectors $\{\delta_i\}_{i=1}^K$ such that

$$
\sum_{i=1}^{K}\delta_i\delta_i^\top = H \quad \text{and} \quad \delta_i^\top(B^\top B)\delta_i = \frac{\operatorname{tr}\big(H(B^\top B)\big)}{K} = 1,
$$

which, along with the fact that $\operatorname{rank}(H) = m \Rightarrow \operatorname{span}\{\delta_i\}_{i=1}^K = \mathbb{R}^m$, imply that

$$
R \subset h(S). \tag{35}
$$

Moreover, for any collection $\{\delta_i\}_{i=1}^K \in S$, we have

$$
\operatorname{tr}\big(h(\delta_1, \delta_2, \ldots, \delta_K)(B^\top B)\big) = \sum_{i=1}^{K}\delta_i^\top(B^\top B)\delta_i = K \quad \text{and} \quad h(\delta_1, \delta_2, \ldots, \delta_K) \in \mathbb{S}_{++}^{m\times m},
$$

which implies that

$$
h(S) \subset R. \tag{36}
$$

From (35) and (36) we conclude that $R = h(S)$. In other words, instead of optimizing over the feasible collection of vectors in S in (34), one can equivalently optimize over the set of symmetric positive definite matrices in R. This consideration leads us to the problem:

$$
\begin{aligned}
&\underset{H \in \mathbb{S}_{++}^{m\times m}}{\text{minimize}} && \operatorname{tr}\big(\Sigma_{V_X}H^{-1}\big)\\
&\text{subject to} && \operatorname{tr}\big(H(B^\top B)\big) - K = 0.
\end{aligned}
\tag{37}
$$

Letting $M := (B^\top B)^{1/2}H(B^\top B)^{1/2}$, we write the optimization problem (37) with M as the variable instead of H. Due to this change of variables, the constraint and the objective function become

$$
\operatorname{tr}\big(H(B^\top B)\big) = \operatorname{tr}\big((B^\top B)^{1/2}H(B^\top B)^{1/2}\big) = \operatorname{tr}(M),
$$

and

$$
\begin{aligned}
\operatorname{tr}\big(\Sigma_{V_X}H^{-1}\big) &= \operatorname{tr}\big(\Sigma_{V_X}(B^\top B)^{1/2}M^{-1}(B^\top B)^{1/2}\big)\\
&= \operatorname{tr}\big((B^\top B)^{1/2}\Sigma_{V_X}(B^\top B)^{1/2}M^{-1}\big)\\
&= \operatorname{tr}\big(\Sigma M^{-1}\big),
\end{aligned}
\tag{38}
$$

where
$$\begin{aligned}\Sigma &:= (B^\top B)^{1/2}\Sigma_{V_X}(B^\top B)^{1/2}\\ &= (B^\top B)^{1/2}\big((B^\top B)^{-1}B^\top\big)\,\mathsf{E}_\mu\big[VV^\top\big]\big((B^\top B)^{-1}B^\top\big)^\top(B^\top B)^{1/2}\\ &= (B^\top B)^{-1/2}\big(B^\top\Sigma_V B\big)(B^\top B)^{-1/2}.\end{aligned} \tag{39}$$

Using (38) we write the problem (37) equivalently as:

$$\begin{aligned}&\underset{M\in\mathbb{S}^{n\times n}_{++}}{\text{minimize}} && \operatorname{tr}\big(\Sigma M^{-1}\big)\\ &\text{subject to} && \operatorname{tr}(M)-K=0.\end{aligned} \tag{40}$$

The problem (40) is identical to (26), which implies that the problem (40) is solvable, and an optimizer is

$$M^* := \frac{K}{\operatorname{tr}\big(\Sigma^{1/2}\big)}\Sigma^{1/2}.$$

Therefore, the problem (37) is solvable, and an optimizer is

$$\begin{aligned}H^* &:= (B^\top B)^{-1/2}M^*(B^\top B)^{-1/2}\\ &= \frac{K}{\operatorname{tr}\big(\Sigma^{1/2}\big)}\big((B^\top B)^{-1/2}\Sigma^{1/2}(B^\top B)^{-1/2}\big).\end{aligned} \tag{41}$$

From Corollary 16 it follows that there exists a collection $\{\delta_i^*\}_{i=1}^K$ of vectors such that

$$\sum_{i=1}^K \delta_i^*\delta_i^{*\top} = H^* \quad\text{and}\quad \delta_i^{*\top}\big(B^\top B\big)\delta_i^* = \frac{\operatorname{tr}\big(H^*(B^\top B)\big)}{K} = 1.$$

Employing arguments similar to those given in the proof of Theorem 1, we now conclude that the pair

○ the collection of vectors $\{\delta_i^*\}_{i=1}^K$, and

○ the scheme $f_X^*(u) = \begin{pmatrix}\delta_1^* & \delta_2^* & \cdots & \delta_K^*\end{pmatrix}^+ u$,

is optimal for the problem (9). Using the optimal solution of (9), we define a dictionary-scheme pair as:

$$\begin{cases} d_i^* := B\delta_i^* \quad \text{for } i=1,\ldots,K,\\ \mathbb{R}^n \ni v \longmapsto f^*(v) := f_X^*\Big(\big((B^\top B)^{-1}B^\top\big)v\Big) = \Big(\delta_1^* \quad \delta_2^* \quad \cdots \quad \delta_K^*\Big)^+\big((B^\top B)^{-1}B^\top\big)v. \end{cases} \tag{42}$$

It is clear that the pair in (42) is feasible for the problem (6), and that the corresponding objective function evaluates to the optimal value of the problem (34). Therefore, along with the assertion of Lemma 2 we can conclude that the problem (6) is solvable, and in fact an optimal solution is given by (42) with the optimal value of $\dfrac{\big(\operatorname{tr}(\Sigma^{1/2})\big)^2}{K}$. This completes the proof. ■

As in the case $X_V = \mathbb{R}^n$, we now provide the Algorithm 3 to obtain an optimal dictionary-scheme pair for the general ℓ_2-optimal dictionary problem (6). The algorithm takes the matrix Σ_V and the size of the dictionary K as its inputs. From Σ_V we extract a matrix $B \in \mathbb{R}^{n\times m}$ containing a set of basis vectors for image(Σ_V) in its columns, these vectors form a basis for X_V.

```
Algorithm 3: A procedure to obtain $\ell_2$-optimal dictionary.
  Input: A matrix $\Sigma_V \in \mathbb{S}_+^{n\times n}$ and a number $K \geqslant m := \dim(X_V) = \mathrm{rank}(\Sigma_V)$.
  Output: An $\ell_2$-optimal dictionary-scheme pair $\big(\{y_i^*\}_{i=1}^K, f^*\big)$.
  Compute a basis $\{b_i\}_{i=1}^m$ for image$(\Sigma_V)$ and define $B := \begin{pmatrix} b_1 & b_2 & \cdots & b_m \end{pmatrix}$.
  Define $\Sigma := (B^\top B)^{-1/2}\big(B^\top \Sigma_V B\big)(B^\top B)^{-1/2}$.
  Compute $H := \frac{K}{\mathrm{tr}\left(\Sigma^{1/2}\right)}\big((B^\top B)^{-1/2}\Sigma^{1/2}(B^\top B)^{-1/2}\big)$.
  Define $M := \begin{pmatrix} H & O_{m\times(K-m)} \\ O_{(K-m)\times m} & I_{K-m} \end{pmatrix}$, $A := \begin{pmatrix} B^\top B & O_{m\times(K-m)} \\ O_{(K-m)\times m} & O_{(K-m)\times(K-m)} \end{pmatrix}$
  Compute $C \in \mathbb{R}^{K\times K}$ such that $M = CC^\top$.
  Define $\Lambda \in \mathbb{R}^{K\times K}$ by $\Lambda := C^\top AC$, and apply Algorithm 1 to get a collection of
    vectors $\{x_i\}_{i=1}^K \subset \mathbb{R}^K$.
  Define the collection $\{v_i\}_{i=1}^K \subset \mathbb{R}^K$ as $v_i := Cx_i$ for $i = 1, \ldots, K$.
  Define the collection $\{\delta_i^*\}_{i=1}^K \subset \mathbb{R}^m$ such that the $j^{th}$ component of $\delta_i^*$ is given by
    $\delta_i^*(j) := v_i(j)$ for $j = 1, \ldots, m$ and for every $i = 1, \ldots, K$.
  Define the $\ell_2$-optimal dictionary $\{d_i^*\}_{i=1}^K \subset \mathbb{R}^n$ as $d_i^* := B\delta_i^*$ for $i = 1, \ldots, K$.
  Define the optimal scheme $\mathbb{R}^n \ni v \longmapsto f^*(v) := \begin{pmatrix} d_1^* & d_2^* & \cdots & d_K^* \end{pmatrix}^+ v \in \mathbb{R}^K$.
```

5. Conclusion and future directions

In this article we have provided an explicit solution of the ℓ_2-optimal dictionary problem in the form of a *rank*-1 *decomposition* of a specific positive definite matrix derived from given data, together with algorithms to compute the corresponding ℓ_2-optimal dictionaries.

The analysis in this article assumes that the second moment of the random vector whose samples are to be represented is known. An online algorithm which estimates the second moment of the random vector and computes the dictionary vectors in parallel is being developed, and will be reported in subsequent articles.

Acknowledgments

We sincerely thank Prof. V. S. Borkar for his valuable pointers on convexity of the ℓ_2-optimal dictionary problem, Prof. K. S. Mallikarjuna Rao for pointing us to the literature on rank-1 decompositions of matrices, and Prof. N. Khaneja for helpful discussions.

References

B. D. O. Anderson and J. B. Moore. *Optimal Control: Linear Quadratic Methods*. Courier Corporation, 2007.

J. J. Benedetto and M. Fickus. Finite Normalized Tight Frames. *Advances in Computational Mathematics*, 18(2-4):357–385, 2003.

D. P. Bertsekas. *Dynamic Programming and Optimal Control*, volume 1. Athena Scientific Belmont, MA, 1995.

R. Bhatia. *Matrix Analysis*, volume 169. Springer-Verlag, New York, 1997.

R. Bhatia. *Positive Definite Matrices.* Princeton University Press, 2009.

S. Boyd and L. Vandenberghe. *Convex Optimization.* Cambridge University Press, 2004.

O. Christensen. *An Introduction to Frames and Riesz Bases.* Applied and Numerical Harmonic Analysis. Birkhäuser/Springer, [Cham], 2nd edition, 2016.

F. Clarke. *Functional Analysis, Calculus of Variations and Optimal Control*, volume 264. Springer Science & Business Media, 2013.

I. Daubechies, A. Grossmann, and Y. Meyer. Painless Nonorthogonal Expansions. *Journal of Mathematical Physics*, 27(5):1271–1283, 1986.

K. K. Delgado, J. F. Murray, B. D. Rao, K. Engan, T. Lee, and T. J. Sejnowski. Dictionary Learning Algorithms for Sparse Representation. *Neural Computation*, 15(2):349–396, 2003.

D. Liberzon. *Calculus of Variations and Optimal Control Theory: A Concise Introduction.* Princeton University Press, 2012.

J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online Dictionary Learning for Sparse Coding. In *Proceedings of the 26th Annual International Conference on Machine Learning*, pages 689–696. ACM, 2009a.

J. Mairal, J. Ponce, G. Sapiro, A. Zisserman, and F. R. Bach. Supervised Dictionary Learning. In *Advances in Neural Information Processing Systems*, pages 1033–1040, 2009b.

S. G. Mallat and Z. Zhang. Matching Pursuits With Time-Frequency Dictionaries. *IEEE Transactions on Signal Processing*, 41(12):3397–3415, 1993.

P. C. Müller and H. I. Weber. Analysis and optimization of certain qualities of controllability and observability for linear dynamical systems. *Automatica*, 8(3):237–246, 1972.

B. A. Olshausen and D. J. Field. Sparse Coding With an Overcomplete Basis Set: A Strategy Employed by v1? *Vision Research*, 37(23):3311–3325, 1997.

K. R. Parthasarathy. *Probability Measures on Metric Spaces.* AMS Chelsea Publishing, Providence, RI, 2005. Reprint of the 1967 original.

F. Pasqualetti, S. Zampieri, and F. Bullo. Controllability metrics, limitations and algorithms for complex networks. *IEEE Transactions on Control of Network Systems*, 1(1):40–52, 2014.

E. B. Saff and A. Kuijlaars. Distributing Many Points on a Sphere. *The Mathematical Intelligencer*, 19(1):5–11, 1997.

M. R. Sheriff and D. Chatterjee. On a Frame Theoretic Measure of Quality of LTI Systems. *arXiv preprint arXiv:1703.07539*, 2017.

K. Skretting and K. Engan. Recursive Least Squares Dictionary Learning Algorithm. *IEEE Transactions on Signal Processing*, 58(4):2121–2130, 2010.

J. F. Sturm and S. Zhang. On Cones of Nonnegative Quadratic Functions. *Mathematics of Operations Research*, 28(2):246–267, 2003.

I. Tošić and P. Frossard. Dictionary Learning. *Signal Processing Magazine, IEEE*, 28(2): 27–38, 2011.

M. Yaghoobi, T. Blumensath, and M. E. Davies. Dictionary Learning for Sparse Approximations With the Majorization Method. *IEEE Transactions on Signal Processing*, 57(6): 2178–2191, 2009.

M. Yang, L. Zhang, X. Feng, and D. Zhang. Fisher Discrimination Dictionary Learning for Sparse Representation. In *IEEE International Conference on Computer Vision (ICCV), 2011*, pages 543–550. IEEE, 2011.

F. Zhang. *Matrix Theory: Basic Results and Techniques.* Springer Science & Business Media, 2011.

G. Zimmermann. Normalized Tight Frames in Finite Dimensions. In *Recent Progress in Multivariate Approximation*, pages 249–252. Springer, 2001.