On the Identifiability of Overcomplete Dictionaries via the Minimisation Principle Underlying K-SVD

Karin Schnass

Abstract

This article gives theoretical insights into the performance of K-SVD, a dictionary learning algorithm that has gained significant popularity in practical applications. The particular question studied here is when a dictionary $\Phi \in \mathbb{R}^{d\times K}$ can be recovered as local minimum of the minimisation criterion underlying K-SVD from a set of N training signals $y_n = \Phi x_n$. A theoretical analysis of the problem leads to two types of identifiability results assuming the training signals are generated from a tight frame with coefficients drawn from a random symmetric distribution. First, asymptotic results showing, that in expectation the generating dictionary can be recovered exactly as a local minimum of the K-SVD criterion if the coefficient distribution exhibits sufficient decay. Second, based on the asymptotic results it is demonstrated that given a finite number of training samples N, such that $N/\log N = O(K^3 d)$, except with probability $O(N^{-Kd})$ there is a local minimum of the K-SVD criterion within distance $O(KN^{-1/4})$ to the generating dictionary.

Index Terms

dictionary learning, sparse coding, sparse component analysis, K-SVD, finite sample size, sampling complexity, dictionary identification, minimisation criterion, sparse representation

1 Introduction

As the universe expands so does the information we are collecting about and in it. New and diverse sources such as the internet, astronomic observations, medical diagnostics, etc., confront

Karin Schnass is with the Computer Vision Laboratory, University of Sassari, Porto Conte Ricerche, 07041 Alghero, Italy, email: kschnass@uniss.it

us with a flood of data in ever increasing dimensions and while we have a lot of technology at our disposal to acquire these data, we are already facing difficulties in storing and even more importantly interpreting them. Thus in the last decades high-dimensional data processing has become a very challenging and interdisciplinary field, requiring the collaboration of researchers capturing the data on one hand and researchers from computer science, information theory, electric engineering and applied mathematics, developing the tools to deal with the data on the other hand. One of the most promising approaches to dealing with high-dimensional data so far has proven to be through the concept of sparsity.

A signal is called sparse if it has a representation or good approximation in a dictionary, i.e. a representation system like an orthonormal basis or frame, [10], such that the number of dictionary elements, also called atoms, with non-zero coefficients is small compared to the dimension of the space. Modelling the signals as vectors $y \in \mathbb{R}^d$ and the dictionary accordingly as a matrix collecting K normalised atom-vectors as its columns, i.e. $\Phi = (\phi_1, \dots \phi_K), \phi_i \in \mathbb{R}^d, \|\phi_i\|_2 = 1$, we have

$$y \approx \sum_{i \in I} x(i)\phi_i,$$

for a set I of size S, i.e. $|I| = S$, which is small compared to the ambient dimension, i.e. $S \ll d \leq K$.

The above characterisation already shows why sparsity provides such an elegant way of dealing with high-dimensional data. No matter the size of the original signal, given the right dictionary, its size effectively reduces to a small number of non-zero coefficients. For instance the sparsity of natural images in wavelet bases is the fundamental principle underlying the compression standard JPEG 2000.

Classical sparsity research studies two types of problems. The first line of research investigates how to perform the dimensionality reduction algorithmically, i.e. how to find the sparse approximations of a signal given the sparsity inducing dictionary. By now there exists a substantial amount of theory including a vast choice of algorithms, e.g. [13], [9], [29], [6], [12], together with analysis about their worst case or average case performance, [38], [39], [35], [20]. The second line of research investigates how sparsity can be exploited for efficient data processing. So it has been shown that sparse signals are very robust to noise or corruption and can therefore easily be denoised, [15], or restored from incomplete information. This second effect is being exploited in the very active research field of *compressed sensing*, see [14], [8], [31].

However, while sparsity based methods have proven very efficient for high-dimensional data processing, they suffer from one common drawback. They all rely on the existence of a dictionary providing sparse representations for the data at hand.

The traditional approach to finding efficient dictionaries is through the careful analysis of the given data class, which for instance has led to the development of wavelets, [11], and curvelets, [7], for natural images. However when faced with a (possibly exotic) new signal class this analytic approach has the disadvantage of requiring too much time and effort. Therefore, more recently, researchers have started to investigate the possibilities of learning the appropriate dictionary directly from the new data class, i.e. given N signals $y_n \in \mathbb{R}^d$, stored as columns in a matrix $Y = (y_1, \dots, y_N)$ find a decomposition

$$Y \approx \Phi X$$

into a $d \times K$ dictionary matrix Φ with unit norm columns and a $K \times N$ coefficient matrix with sparse columns. Looking at the matrix decomposition we can immediately see that, on top of being the key to sparse data processing schemes, dictionary learning is actually a powerful data-analysis tool. Indeed within the blind source separation community dictionary learning is known as sparse component analysis (the dictionary atoms are the sparse components) and this data-analysis point of view has been a parallel driving force for the development of dictionary learning.

So far the research focus in dictionary learning has been on algorithmic development rather than theoretic analysis. This means that by now there are several dictionary learning algorithms, which are efficient in practice and therefore popular in applications, see [16], [23], [3], [26], [42], [24], [36] or [32] for a more complete survey, but only comparatively little theory. Some theoretical insights come from the blind source separation community, [43], [18], and more recently from a set of generalisation bounds for learned dictionaries, [27], [40], [28], [19], which predict the quality of a learned dictionary for future data, but unfortunately do not directly imply uniqueness of the 'true' dictionary nor guarantee recoverability by an efficient algorithm, However, especially to justify the use of dictionary learning as data analysis tool, we need theoretical identification results quantifying the conditions on the dictionary, the coefficient model generating the sparse signals and the number of training signals under which a scheme will be successful.

While it is true that for most schemes we do not yet understand their behaviour, there exists a

handful of exceptions to this rule, [4], [21], [17], [22], [37][1]. For these schemes there are known conditions under which a dictionary can be recovered from a given signal class, but unfortunately they all have certain drawbacks limiting their practical applicability. In [4] the authors themselves state that the algorithm is only of theoretical interest because of its computational complexity and also for the ℓ_1-minimisation principle, suggested in [43], [30] and studied in [21], [17], [22], finding a local minimum is computational sufficiently challenging to prohibit the learning of very high-dimensional dictionaries. Finally, the ER-SpUD algorithm, [37], has the disadvantage that it can only learn a basis, but not an overcomplete dictionary.

In this paper we will start bridging the gap between practically efficient and provably efficient dictionary learning schemes, by providing identification results for the minimisation principle underlying K-SVD (K-Singular Value Decomposition), one of the most widely applied dictionary algorithms.

K-SVD was introduced by Aharon, Elad and Bruckstein in [3] as a generalisation of the K-means clustering process. The starting point for the algorithm is the following minimisation criterion. Given some signals $Y = (y_1, \dots, y_N)$, $y_n \in \mathbb{R}^d$, find

$$\min_{\Phi \in \mathcal{D}, X \in \mathcal{X}_S} \|Y - \Phi X\|_F^2 \tag{1}$$

for $\mathcal{D} := \{\Phi = (\phi_1, \dots, \phi_K), \phi_i \in \mathbb{R}^d, \|\phi_i\|_2 = 1\}$ and $\mathcal{X}_S := \{X = (x_1, \dots, x_N), x_n \in \mathbb{R}^K, \|x_n\|_0 \leq S\}$, where $\|x\|_0$ counts the number of non-zero entries of x, and $\|\cdot\|_F$ denotes the Frobenius norm. In other words we are looking for the dictionary that provides on average the best S-term approximation to the signals in Y.

K-SVD aims to find the minimum of (1) by alternating two procedures, a) fixing the dictionary Φ and finding a new close to optimal coefficient matrix X^{new} column-wise, using a sparse approximation algorithm such as (Orthogonal) Matching Pursuit, [38], or Basis Pursuit, [9], and b) updating the dictionary atom-wise, choosing the updated atom ϕ_i^{new} to be the left singular vector to the maximal singular value of the matrix having as its columns the residuals $y_n - \sum_{k \neq i} \phi_k x_n(k)$ of all signals y_n to which the current atom ϕ_i contributes, i.e. $X_{ni} = x_n(i) \neq 0$. If in every step for every signal the best sparse approximation is found the K-SVD algorithm is guaranteed to find a local minimiser of (1). However because of the non-optimal sparse approximation procedure it can in general not be guaranteed to converge to a local minimiser

1. For the sake of completeness we also mention (without discussion) some very recent results, developed while this work has been under review, [5], [2], [1] .

of (1) unless $S = 1$ and a greedy algorithm is used, see also the discussion in Section 5. We will not go further into algorithmic details, but refer the reader to the original paper [3] as well as [4]. Instead we concentrate on the theoretical aspects of the posed minimisation problem.

First it will be convenient to rewrite the objective function using the fact that for any signal y_n the best S-term approximation using Φ is given by the largest projection onto a set of S atoms $\Phi_I = (\phi_{i_1} \dots \phi_{i_S})$, i.e.,

$$\begin{aligned}\min_{\Phi\in\mathcal{D}, X\in\mathcal{X}_S} \|Y - \Phi X\|_F^2 &= \min_{\Phi\in\mathcal{D}} \sum_n \min_{\|x_n\|_0\leq S} \|y_n - \Phi x_n\|_2^2 \\ &= \min_{\Phi\in\mathcal{D}} \sum_n \min_{|I|=S} \|y_n - \Phi_I \Phi_I^\dagger y_n\|_2^2 \\ &= \|Y\|_F - \max_{\Phi\in\mathcal{D}} \sum_n \max_{|I|=S} \|\Phi_I \Phi_I^\dagger y_n\|_2^2,\end{aligned}$$

where $\Phi_I^\dagger$ denotes the Moore-Penrose pseudo inverse of Φ_I. Abbreviating the projection onto the span of $(\phi_i)_{i\in I}$ by $P_I(\Phi) = \Phi_I \Phi_I^\dagger$, we can thus replace the minimisation problem in (1) with the following maximisation problem,

$$\max_{\Phi\in\mathcal{D}} \sum_n \max_{|I|=S} \|P_I(\Phi) y_n\|_2^2. \tag{2}$$

From the above formulation it is quite easy to see the motivation for the proposed learning criterion. Indeed assume that the training signals are all $\bar{S}$-sparse in an admissible dictionary $\bar{\Phi} \in \mathcal{D}$, i.e. $Y = \bar{\Phi}\bar{X}$ and $\|\bar{x}_i\|_0 \leq \bar{S}$, then clearly there is a global maximum[2] of (2) at $\bar{\Phi}$, respectively a global minimum of (1) at $(\bar{\Phi}, \bar{X})$, as long as $\bar{S} \leq S$. However in practice we will be facing something like,

$$y_n = \bar{\Phi}\bar{x}_n + r_n \qquad \text{or} \qquad Y = \bar{\Phi}\bar{X} + R, \tag{3}$$

where the coefficient vectors $\bar{x}_n$ in $\bar{X}$ are only approximately S-sparse or rapidly decaying and the pure signals $\bar{\Phi}\bar{x}_n$ are corrupted with noise $R = (r_1, \dots, r_K)$. In this case it is no longer trivial or obvious that $\bar{\Phi}$ is a local maximum of (2), but we can hope for a result of the following type.

Goal 1.1. *Assume that the signals y_n are generated as in* (3)*, with x_n drawn from a distribution of approximately sparse or decaying vectors and r_n random noise. As soon as the number of signals N is*

2. $\bar{\Phi}$ is a global maximiser together with all $2^K K!$ dictionaries consisting of a permutation of the atoms in $\bar{\Phi}$ provided with a ± 1 sign. For a more detailed discussion on the uniqueness of the maximiser/minimiser see eg. [21].

large enough $N \geq C$, with high probability $p \approx 1$ there will be a local maximum of (2) *within distance ε from $\bar{\Phi}$.*

The rest of this paper is organised as follows. After introducing some notation in Section 2, we first give conditions on the dictionary and the coefficients which allow for asymptotic identifiability by studying when $\bar{\Phi}$ is exactly at a local maximum in the limiting case, where we replace the sum in (2) with the expectation,

$$\max_{\Phi \in \mathcal{D}} \mathbb{E}_y \left(\max_{|I|=S} \|P_I(\Phi) y\|_2^2 \right). \tag{4}$$

Thus in Section 3 we will prove identification results for (4) assuming first a simple (discrete, noise-free) signal model and then progressing to a noisy, continuous signal model. In Section 4 we will go from asymptotic results to results for finite sample sizes and prove versions of Theorem 1.1 that under the same assumptions as the asymptotic results quantify the sizes of the parameters ε, p in terms of the number of training signals N and the size of C in terms of the number of atoms K. In the last section we will discuss the implications of our results for practical applications, compare them to existing identification results and point out some directions for future research.

2 Notations and Conventions

Before we jump into the fray, we collect some definitions and lose a few words on notations; usually subscripted letters will denote vectors with the exception of c and ε where they are numbers, eg. $(x_1, \dots, x_K) = X \in \mathbb{R}^{d \times K}$ vs. $c = (c_1, \dots, c_K) \in \mathbb{R}^K$, however, it should always be clear from the context what we are dealing with.

For a matrix M, we denote its (conjugate) transpose by $M^\star$ and its Moore-Penrose pseudo inverse by $M^\dagger$. We denote its operator norm by $\|M\|_{2,2} = \max_{\|x\|_2=1} \|Mx\|_2$ and its Frobenius norm by $\|M\|_F = \operatorname{tr}(M^\star M)^{1/2}$, remember that we have $\|M\|_{2,2} \leq \|M\|_F$.

We consider a **dictionary** Φ a collection of K unit norm vectors $\phi_i \in \mathbb{R}^d$, $\|\phi_i\|_2 = 1$. By abuse of notation we will also refer to the $d \times K$ matrix collecting the atoms as its columns as the dictionary, i.e. $\Phi = (\phi_i, \dots \phi_K)$. The maximal absolute inner product between two different atoms is called the **coherence** μ of a dictionary, $\mu = \max_{i \neq j} |\langle \phi_i, \phi_j \rangle|$.

By Φ_I we denote the restriction of the dictionary to the atoms indexed by I, i.e. $\Phi_I = (\phi_{i_1} \dots \phi_{i_S})$, $i_j \in I$, and by $P_I(\Phi)$ the orthogonal projection onto the span of the atoms indexed by I, i.e. $P_I(\Phi) = \Phi_I \Phi_I^\dagger$. Note that in case the atoms indexed by I are linearly independent we have

$\Phi_I^\dagger = (\Phi_I^\star \Phi_I)^{-1}\Phi_I^\star$.

(Ab)using the language of compressed sensing we denote the minimal eigenvalue of $\Phi_I^\star\Phi_I$ by $1-\delta_I(\Phi)$ and define the **lower isometry constant** $\delta_S(\Phi)$ of the dictionary as $\delta_S(\Phi) := \max_{|I|\leq S}\delta_I(\Phi) \leq 1$. If any set of S atoms is linearly independent we have $\delta_S(\Phi) < 1$ and in general we have the bound $\delta_S(\Phi) \leq \mu(S-1)$. When clear from the context we will usually omit the reference to the dictionary. For more details on isometry constants, see for instance [8]. For two dictionaries Φ, Ψ we define the distance between each other as the maximal distance between two corresponding atoms, i.e.

$$d(\Phi,\Psi) := \max_i \|\phi_i - \psi_i\|_2. \tag{5}$$

We consider a **frame** F a collection of $K \geq d$ vectors $f_i \in \mathbb{R}^d$ for which there exist two positive constants A, B such that for all $v \in \mathbb{R}^d$ we have

$$A\|v\|_2^2 \leq \sum_{i=1}^K |\langle f_i, v\rangle|^2 \leq B\|v\|_2^2. \tag{6}$$

If B can be chosen equal to A, i.e. $B = A$, the frame is called tight and if all elements of a tight frame have unit norm we have $A = K/d$. The operator $FF^\star$ is called frame operator and by (6) its spectrum is bounded by A, B. For more details on frames, see e.g. [10].

Finally we introduce the Landau symbols O, o to characterise the growth of a function. We write $f(\varepsilon) = O(g(\varepsilon))$ if $\lim_{\varepsilon\to 0} f(\varepsilon)/g(\varepsilon) = C < \infty$ and $f(\varepsilon) = o(g(\varepsilon))$ if $\lim_{\varepsilon\to 0} f(\varepsilon)/g(\varepsilon) = 0$.

3 ASYMPTOTIC IDENTIFICATION RESULTS

As mentioned in the introduction if the signals y are all S-sparse in a dictionary $\bar\Phi$ then clearly there is a global minimum of (1) or global maximum of (4) with parameter S at $\bar\Phi$. However what happens if we do not have perfect S-sparsity? Let us start with a very simple negative example of a coefficient distribution for which the original generating dictionary is not at a local maximum for the case $S = 1$.

Example 3.1. Let U be an orthonormal basis and let the signals be generated as $y = Ux$, where x is a randomly 2-sparse, 'flat' coefficients sequence, i.e. we pick an index set $I = \{i, j\}$ and two signs $\sigma_{i/j} = \pm 1$ uniformly at random and set $x(k) = \sigma_k$ for $k \in I$ and zero else. Then there is no local maximum of (4) with $S = 1$ at U. Indeed since the signals are all 2-sparse the maximal inner product with all atoms in U is the same as the maximal inner product with

only $d-1$ atoms. This degree of freedom we can use to construct an ascent direction. Choose $U_\varepsilon = (u_1, \dots, u_{d-1}, (u_d + \varepsilon u_1)/\sqrt{1+\varepsilon^2})$. Using the identity $\max_i \|P_i(\Phi)y\|_2^2 = \|\Phi^\star y\|_\infty^2$ we get,

$$\begin{aligned}
\mathbb{E}_y\left(\|U_\varepsilon^\star y\|_\infty^2\right) &= \mathbb{E}_x\left(\|U_\varepsilon^\star U x\|_\infty^2\right) \\
&= \mathbb{E}_x\left(\|(x(1), \dots, x(d-1), \tfrac{x(d)+\varepsilon x(1)}{\sqrt{1+\varepsilon^2}})\|_\infty^2\right) \\
&= 1 \cdot (1 - \mathbb{P}(I = \{1,d\} \cap \sigma_1 = \sigma_d)) + \frac{(1+\varepsilon)^2}{1+\varepsilon^2} \cdot \mathbb{P}(I = \{1,d\} \cap \sigma_1 = \sigma_d) \\
&= 1 + \frac{\varepsilon}{1+\varepsilon^2} \cdot \frac{1}{d(d-1)},
\end{aligned}$$

which is larger than $\mathbb{E}_y\left(\|U^\star y\|_\infty^2\right) = 1$.

From the above example we see that in order to have a local maximum at the original dictionary we need a signal/coefficient model where the coefficients show some type of decay.

3.1 A simple model of decaying coefficients

To get started we consider a very simple coefficient model, constructed from a non-negative, non-increasing sequence $c \in \mathbb{R}^K$ with $\|c\|_2 = 1$, which we permute uniformly at random and provide with random $\pm$ signs. To be precise for a permutation $p : \{1, ..., K\} \to \{1, ..., K\}$ and a sign sequence σ, $\sigma_i = \pm 1$, we define the sequence $c_{p,\sigma}$ component-wise as $c_{p,\sigma}(i) := \sigma_i c_{p(i)}$, and set $y = \Phi x$ where $x = c_{p,\sigma}$ with probability $(2^K K!)^{-1}$.

The normalisation $\|c\|_2 = 1$ has the advantage that for dictionaries, which are an orthonormal basis, the resulting signals also have unit norm and for general dictionaries the signals have unit square norm in expectation, i.e. $\mathbb{E}(\|y\|_2^2) = 1$. This reflects the situation in practical applications, where we would normalise the signals in order to equally weight their importance.

Armed with this model we can now prove a first dictionary identification result for (4).

Theorem 3.1. *Let Φ be a unit norm tight frame with frame constant $A = K/d$ and lower isometry constant δ_S. Let x be a random permutation of a positive, nonincreasing sequence c, where $c_1 \geq c_2 \geq c_3 \dots \geq c_K \geq 0$ and $\|c\|_2 = 1$, provided with random $\pm$ signs, i.e. $x = c_{p,\sigma}$ with probability $\mathbb{P}(p,\sigma) = (2^K K!)^{-1}$. Assume that the signals are generated as $y = \Phi x$. If there exists $\kappa > 0$ such that for $I_p := p^{-1}(\{1, \dots S\})$ we have*

$$\|P_{I_p}(\Phi)\Phi c_{p,\sigma}\|_2 - \max_{|I|=S, I \neq I_p} \|P_I(\Phi)\Phi c_{p,\sigma}\|_2 \geq 2\kappa, \quad \forall \sigma, p, \tag{7}$$

then there is a local maximum of (4) *at* Φ.

Moreover for $\Psi \neq \Phi$ *we have* $\mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Psi)y\|_2^2\right) < \mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Phi)y\|_2^2\right)$ *as soon as*

$$d(\Phi,\Psi) \leq \frac{\kappa\sqrt{1-\delta_S}}{\sqrt{\frac{9S}{2}}\left(1+4\sqrt{\log\left(\frac{60A\binom{K}{S}^2}{\kappa\lambda_S(1-\delta_S)}\right)}\right)}, \tag{8}$$

where $\lambda_S = \frac{c_1^2+\ldots+c_S^2}{S} - \frac{1-c_1^2-\ldots-c_S^2}{K-S}$ *and* $\delta_S < 1$ *because of* (7).

Proof: The basic idea of the proof is that for the original dictionary the maximal response is always attained for the set I_p and that for most signals, i.e. most sign sequences, also for a perturbed dictionary the maximal response is still at I_p. Since the average loss of a perturbed dictionary over most sign sequences,

$$\mathbb{E}_p\mathbb{E}_\sigma\left(\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2\right) < \mathbb{E}_p\mathbb{E}_\sigma\left(\|P_{I_p}(\Phi)\Phi c_{p,\sigma}\|_2^2\right), \tag{9}$$

is larger than the possible gain on exceptional sign sequences we have a maximum at Φ. More detailed sketches and a version of the proof for $S = 1$ can be found in [34], [33].

Following the proof idea we first calculate the expectation using the original dictionary Φ. Condition (7) quite obviously (and artlessly) guarantees that the maximum is always attained for the set I_p, so setting $\gamma_S^2 := c_1^2 + \ldots + c_S^2$ we get from Lemma A.1 in the appendix,

$$\begin{aligned}\mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Phi)y\|_2^2\right) &= \mathbb{E}_p\mathbb{E}_\sigma\left(\|P_{I_p}(\Phi)\Phi c_{p,\sigma}\|_2^2\right)\\ &= \frac{A(1-\gamma_S^2)S}{(K-S)} + \left(\frac{\gamma_S^2}{S} - \frac{1-\gamma_S^2}{K-S}\right)\binom{K}{S}^{-1}\sum_{I:|I|=S}\|\Phi_I\|_F^2. \end{aligned}\tag{10}$$

To compute the expectation for a perturbation of the original dictionary we first note that we can parametrise all ε-perturbations Ψ of the original dictionary Φ, i.e. $d(\Phi,\Psi) = \varepsilon$, as

$$\psi_i = (1-\varepsilon_i^2/2)\phi_i + (\varepsilon_i^2 - \varepsilon_i^4/4)^{\frac{1}{2}} z_i,$$

for some z_i with $\langle\phi_i, z_i\rangle = 0, \|z_i\|_2 = 1$ and some ε_i with $\max_i \varepsilon_i = \varepsilon$. For conciseness of the following presentation we define $\alpha_i := 1-\varepsilon_i^2/2$, $\omega_i := (\varepsilon_i^2 - \varepsilon_i^4/4)^{\frac{1}{2}}$ and $b_i := \omega_i/\alpha_i z_i$. Further we define $A_I = \mathrm{diag}(\alpha_i)_{i\in I}$ and $W_I = \mathrm{diag}(\omega_i)_{i\in I}$ to get $\Psi_I = \Phi_I A_I + Z_I W_I$ and $B_I = Z_I W_I A_I^{-1}$. Note that some perturbations, e.g. small rotations, will be also unit norm tight frames but in general the perturbed dictionaries will not be tight.

As pointed out in the proof idea our strategy will be to show that for a fixed permutation p

with high probability (over σ) the maximal projection is still onto the atoms indexed by I_p. For any index set I of size S we can bound the projection onto a perturbed dictionary as,

$$\begin{aligned}\|P_I(\Psi)y\|_2^2 &= \|P_I(\Phi)y\|_2^2 + \langle P_I(\Phi)y, \big(P_I(\Psi) - P_I(\Phi)\big)y\rangle + \langle P_I(\Psi)y, \big(P_I(\Psi) - P_I(\Phi)\big)y\rangle \\ &\leq \|P_I(\Phi)y\|_2^2 + 2\|y\|_2\|\big(P_I(\Psi) - P_I(\Phi)\big)y\|_2 \\ &\leq \|P_I(\Phi)y\|_2^2 + 2\|y\|_2^2\|\|P_I(\Psi) - P_I(\Phi)\|_{2,2} \\ &\leq \|P_I(\Phi)y\|_2^2 + 2A\max_{|I|=S}\|P_I(\Psi) - P_I(\Phi)\|_F, \end{aligned} \tag{11}$$

leading to

$$\max_{|I|=S}\|P_I(\Psi)y\|_2^2 \leq \|P_{I_p}(\Phi)y\|_2^2 + 2A\max_{|I|=S}\|P_I(\Psi) - P_I(\Phi)\|_F. \tag{12}$$

However (12) is a quite pessimistic estimate since for most $y = \Phi c_{p,\sigma}$, meaning for most σ, the expression $\|\big(P_I(\Psi) - P_I(\Phi)\big)y\|_2$ will be much smaller than $\|P_I(\Psi) - P_I(\Phi)\|_{2,2}\|y\|_2$. Indeed we can estimate its typical size via the following convenient if not optimal concentration inequality for Rademacher series from [25], Chapter 4.

Corollary 3.2 (of Theorem 4.7 in [25]). *For a vector-valued Rademacher series $V = \sum_i \sigma_i v_i$, i.e. for σ_i independent Bernoulli variables with $\mathbb{P}(\sigma_i = \pm 1) = 1/2$ and $v_i \in \mathbb{R}^n$, and $t > 0$ we have,*

$$\mathbb{P}(\|V\|_2 > t) \leq 2\exp\left(\frac{-t^2}{32\mathbb{E}(\|V\|_2^2)}\right). \tag{13}$$

Applied to $v_i = c_{p(i)}\big(P_I(\Psi) - P_I(\Phi)\big)\phi_i$ this leads to the following estimate,

$$\begin{aligned}\mathbb{P}\left(\|\big(P_I(\Psi) - P_I(\Phi)\big)\Phi c_{p,\sigma}\|_2 > t\right) &\leq 2\exp\left(\frac{-t^2}{32\sum_i c_{p(i)}^2\|\big(P_I(\Psi) - P_I(\Phi)\big)\phi_i\|_2^2}\right) \\ &\leq 2\exp\left(\frac{-t^2}{32\sum_i c_{p(i)}^2\|P_I(\Psi) - P_I(\Phi)\|_{2,2}^2}\right) \\ &\leq 2\exp\left(\frac{-t^2}{32\|P_I(\Psi) - P_I(\Phi)\|_F^2}\right), \end{aligned} \tag{14}$$

whenever $P_I(\Psi) \neq P_I(\Phi)$ - otherwise we trivially have $\mathbb{P}\left(\|\big(P_I(\Psi) - P_I(\Phi)\big)\Phi c_{p,\sigma}\|_2 > t\right) = 0$. We now define the set Σ_p,

$$\Sigma_p := \bigcup_{I:|I|=S}\{\sigma : \|\big(P_I(\Psi) - P_I(\Phi)\big)\Phi c_{p,\sigma}\|_2 > \kappa\}, \tag{15}$$

whose size we can estimate using (14) with $t = \kappa$ and a union bound,

$$\mathbb{P}(\Sigma_p) \leq 2\sum_{I:P_I(\Psi)\neq P_I(\Phi)}\exp\left(\frac{-\kappa^2}{32\|P_I(\Psi) - P_I(\Phi)\|_F^2}\right) := \eta_S. \tag{16}$$

Note that whenever $\sigma \notin \Sigma_p$ we have $\max_I \|P_I(\Psi)\Phi c_{p,\sigma}\|_2 = \|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2$, since using the (reversed) triangular inequality we have

$$\begin{aligned}
\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2 &\geq \|P_{I_p}(\Phi)\Phi c_{p,\sigma}\|_2 - \|\big(P_I(\Psi) - P_I(\Phi)\big)\Phi c_{p,\sigma}\|_2 \\
&\geq \|P_{I_p}(\Phi)\Phi c_{p,\sigma}\|_2 - \kappa \\
&\geq \max_{I:I\neq I_p} \|P_I(\Phi)\Phi c_{p,\sigma}\|_2 + \kappa \\
&\geq \max_{I:I\neq I_p} \left(\|P_I(\Phi)\Phi c_{p,\sigma}\|_2 + \|\big(P_I(\Psi) - P_I(\Phi)\big)\Phi c_{p,\sigma}\|_2\right) \\
&\geq \max_{I:I\neq I_p} \|P_I(\Psi)\Phi c_{p,\sigma}\|_2. \qquad (17)
\end{aligned}$$

To finally calculate the expectation over σ for a perturbed dictionary we split it into a sum over the sign sequences contained in Σ_p and its complement. We can estimate,

$$\begin{aligned}
&\mathbb{E}_\sigma\left(\max_{|I|=S} \|P_I(\Psi)c_{p,\sigma}\|_2^2\right) \\
&= \sum_{\sigma\in\Sigma_p} \max_{|I|=S} \|P_I(\Psi)\Phi c_{p,\sigma}\|_2^2 + \sum_{\sigma\notin\Sigma_p} \max_{|I|=S} \|P_I(\Psi)\Phi c_{p,\sigma}\|_2^2 \\
&\leq \sum_{\sigma\in\Sigma_p} \left(\|P_{I_p}(\Phi)c_{p,\sigma}\|_2^2 + 2A \max_{|I|=S} \|P_I(\Psi) - P_I(\Phi)\|_F\right) + \sum_{\sigma\notin\Sigma_p} \max_{|I|=S} \|P_I(\Psi)\Phi c_{p,\sigma}\|_2^2 \qquad (18) \\
&\leq \sum_{\sigma\in\Sigma_p} \left(\|P_{I_p}(\Psi)c_{p,\sigma}\|_2^2 + 4A \max_{|I|=S} \|P_I(\Psi) - P_I(\Phi)\|_F\right) + \sum_{\sigma\notin\Sigma_p} \|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2 \qquad (19) \\
&\leq 4\eta_S A \max_{|I|=S} \|P_I(\Psi) - P_I(\Phi)\|_F + \mathbb{E}_\sigma\left(\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2\right),
\end{aligned}$$

where we have used (11), reversing the roles of Φ and Ψ and choosing $I = I_p$, to go from (18) to (19). Using the expression for $\mathbb{E}_p\mathbb{E}_\sigma\left(\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2\right)$ derived in Lemma A.1 in the appendix we get the following bound for the expectation of the maximal projection using a perturbed dictionary,

$$\begin{aligned}
\mathbb{E}_p\mathbb{E}_\sigma\left(\max_{|I|=S} \|P_I(\Psi)\Phi c_{p,\sigma}\|_2^2\right) \leq \frac{A(1-\gamma_S^2)S}{(K-S)} + \left(\frac{\gamma_S^2}{S} - \frac{1-\gamma_S^2}{K-S}\right)\binom{K}{S}^{-1} \sum_{I:|I|=S} \|P_I(\Psi)\Phi_I\|_F^2 \\
+ 4\eta_S A \max_{|I|=S} \|P_I(\Psi) - P_I(\Phi)\|_F. \qquad (20)
\end{aligned}$$

We are now ready to compare the above expression to the corresponding one for the original dictionary. Abbreviating $\lambda_S = \frac{\gamma_S^2}{S} - \frac{1-\gamma_S^2}{K-S}$ and using the estimates for $\|P_I(\Psi) - P_I(\Phi)\|_F$ and

$\|P_I(\Psi)\Phi_I\|_F^2 - \|\Phi_I\|_F^2$ from Lemma A.2 in the appendix, we get

$$\begin{aligned}
&\mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Psi)y\|_2^2\right) - \mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Phi)y\|_2^2\right) \\
&\quad\le 4A\max_{|I|=S}\|P_I(\Psi)-P_I(\Phi)\|_F \sum_{P_I(\Psi)\neq P_I(\Phi)} \exp\left(\frac{-\kappa^2}{32\|P_I(\Psi)-P_I(\Phi)\|_F^2}\right) \\
&\qquad\qquad + \lambda_S\binom{K}{S}^{-1}\sum_{I:|I|=S}\left(\|P_I(\Psi)\Phi_I\|_F^2 - \|\Phi_I\|_F^2\right) \\
&\quad\le \frac{4AC_1}{\sqrt{1-\delta_S}}\max_{|I|=S}\|Q_I(\Phi)B_I\|_F \sum_{I:Q_I(\Phi)B_I\neq 0} \exp\left(\frac{-\kappa^2(1-\delta_S)}{32C_1^2\|Q_I(\Phi)B_I\|_F^2}\right) \\
&\qquad\qquad - \lambda_S\binom{K}{S}^{-1}\sum_{I:|I|=S} C_2\|Q_I(\Phi)B_I\|_F^2, \qquad (21)
\end{aligned}$$

with $C_1 = 1.487$ and $C_2 = 0.897$ and where we have used that (8) implies $\varepsilon \le \frac{\sqrt{1-\delta_S}}{21\sqrt{S}}$. Denote by $\bar{I}$ the set for which $\|Q_I(\Phi)B_I\|_F$ is maximal. We can further estimate,

$$\begin{aligned}
&\mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Psi)y\|_2^2\right) - \mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Phi)y\|_2^2\right) \\
&\quad\le \frac{4AC_1}{\sqrt{1-\delta_S}}\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F\binom{K}{S}\exp\left(\frac{-\kappa^2(1-\delta_S)}{32C_1^2\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F^2}\right) - \lambda_S\binom{K}{S}^{-1}C_2\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F^2.
\end{aligned}$$

Thus to have a local maximum at Φ we need to show that for $\varepsilon \neq 0$ small enough we have

$$\frac{4AC_1}{\sqrt{1-\delta_S}}\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F\binom{K}{S}\exp\left(\frac{-\kappa^2(1-\delta_S)}{32C_1^2\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F^2}\right) < \lambda_S\binom{K}{S}^{-1}C_2\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F^2,$$

or equivalently that

$$\frac{4AC_1}{\lambda_S C_2\sqrt{1-\delta_S}}\binom{K}{S}^2\exp\left(\frac{-\kappa^2(1-\delta_S)}{32C_1^2\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F^2}\right) < \|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F. \qquad (22)$$

Applying Lemma A.3 we get that for $\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F > 0$ the inequality above is satisfied if we have

$$\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F \le \frac{4\kappa\sqrt{1-\delta_S}}{C_1\sqrt{32}\left(1+\sqrt{1+16\log\left(\frac{4\sqrt{32}C_1^2A\binom{K}{S}^2}{C_2\kappa\lambda_S(1-\delta_S)}\right)}\right)}. \qquad (23)$$

Employing the bound $\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F^2 \le \|B_{\bar{I}}\|_F^2 \le S\varepsilon^2/(1-\varepsilon^2)$ this is further implied by

$$\frac{\varepsilon}{\sqrt{1-\varepsilon^2}} \le \frac{\kappa\sqrt{1-\delta_S}}{C_1\sqrt{2S}\left(1+4\sqrt{\log\left(\frac{4\sqrt{32}e^{1/16}C_1^2A\binom{K}{S}^2}{C_2\kappa\lambda_S(1-\delta_S)}\right)}\right)}, \qquad (24)$$

which is in turn implied by (8).

Finally all that remains to show is that for $\varepsilon > 0$ we have $\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F > 0$. Assume conversely

that for $\varepsilon > 0$ we have $\|Q_{\bar{I}}(\Phi)B_{\bar{I}}\|_F = 0$ meaning that $\|Q_I(\Phi)B_I\|_F = 0$ for all I of size S. We can then find an index ι for which we have $\psi_\iota = (1-\varepsilon^2/2)\phi_i + (\varepsilon^2 - \varepsilon^4/4)^{\frac{1}{2}} z_\iota$ for some z_i with $\langle \phi_\iota, z_\iota \rangle = 0$ and $\|z_\iota\|_2$=1. For all I of size S containing ι we have $Q_I(\Phi)b_\iota = 0$ and therefore $Q_I(\Phi)z_\iota = 0$. Choose J to be any set of size $S-1$ containing ι. For all $j \notin J$ we have $Q_{J\cup j}(\Phi)z_\iota = 0$ or $P_{J\cup j}(\Phi)z_\iota = z_\iota$, which means that either z_ι is in the span of Φ_J and therefore $Q_J(\Phi)z_\iota = 0$ or that ϕ_j is in the span of (Φ_J, z_i) for all $j \notin J$. However this would mean that Φ has rank $S < d$ which is a contradiction to Φ being a frame and we can conclude that $Q_I(\Phi)z_\iota = 0$ for all I of size $S-1$ containing ι. Iterating the argument we get that z_ι has to be in the span of ϕ_ι which is a contradiction to $\langle \phi_\iota, z_\iota \rangle = 0$ and $\|z_\iota\|_2$=1. □

Remark 3.2. (a) To make the theorem more applicable it would be nice to have a concrete condition in terms of the coherence of the dictionary rather than the abstract condition in (7). Indeed it can be shown, see [34] Appendix C, that we can find a $\kappa > 0$ if we have $S\mu < 1/2$ and

$$c_S > \frac{1-S\mu}{1-2S\mu} c_{S+1} + \frac{4\mu}{1-2S\mu} \sum_{i>S+1} |c_i|. \tag{25}$$

In some cases we can also easily derive estimates for κ.

If Φ is an orthonormal basis we have

$$\kappa \geq \frac{c_S^2 - c_{S+1}^2}{2\sqrt{c_1^2 + \ldots + c_S^2}}, \tag{26}$$

and if $S = 1$ we have

$$\kappa \geq (c_1 - c_2)(1-\mu) - 2\mu \sum_{i=3}^{K} c_i. \tag{27}$$

(b) Next note that in some sense the theorem is sharp. Assume that Φ is an orthonormal basis. Then we simply have $\|P_I(\Phi)\Phi c_{p,\sigma}\|_2^2 = \sum_{i\in I} c_{p(i)}^2$ and the condition to be a local minimum reduces to $c_S > c_{S+1}$. However similar to Example 3.1 if $c_S = c_{S+1}$ we can again construct an ascent direction and so Φ is not a local maximum.

(c) Finally before extending Theorem 3.1 to more general coefficient models we want to motivate why we used the condition that Φ is a tight frame.

Assume the same conditions as in Theorem 3.1 but that Φ is not tight, i.e. $A\|v\|_2^2 \leq \sum_i |\langle v, \phi_i \rangle|^2 \leq B\|v\|_2^2$, with $A < B$. Going through the proof we see that using (74) instead of (75) from

Lemma A.1 we get

$$\mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Phi)y\|_2^2\right) = \mathbb{E}_{p,\sigma}\left(\max_{|I|=S}\|P_{I_p}(\Phi)\Phi c_{p,\sigma}\|_2^2\right)$$
$$= \binom{K}{S}^{-1}\left(\frac{1-\gamma_S^2}{K-S}\sum_I\|P_I(\Phi)\Phi\|_F^2 + \left(\frac{\gamma_S^2}{S}-\frac{1-\gamma_S^2}{K-S}\right)\sum_I\|P_I(\Phi)\Phi_I\|_F^2\right),$$

and

$$\mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Psi)y\|_2^2\right) \geq \mathbb{E}_{p,\sigma}\left(\max_{|I|=S}\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2\right)$$
$$= \binom{K}{S}^{-1}\left(\frac{1-\gamma_S^2}{K-S}\sum_I\|P_I(\Psi)\Phi\|_F^2 + \left(\frac{\gamma_S^2}{S}-\frac{1-\gamma_S^2}{K-S}\right)\sum_I\|P_I(\Psi)\Phi_I\|_F^2\right).$$

Moreover by replacing A with B in (11) and (12) we get the new upper bound,

$$\mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Psi)y\|_2^2\right) \leq \mathbb{E}_{p,\sigma}\left(\max_{|I|=S}\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2\right) + 4B\eta_S\max_{|I|=S}\|P_I(\Psi)-P_I(\Phi)\|_F. \tag{28}$$

Since $B\eta_S$ is still of order $o(\varepsilon^2)$ to prove that Φ is a local maximum it suffices to show that up to second order $\mathbb{E}_{p,\sigma}\left(\max_{|I|=S}\|P_{I_p}(\Phi)\Phi c_{p,\sigma}\|_2^2\right) - \mathbb{E}_{p,\sigma}\left(\max_{|I|=S}\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2\right) > 0$. Conversely if we can find perturbation directions z_i such that the reversed inequality holds, Φ is not a local maximum. Using (81) from the appendix, we get

$$\begin{aligned}
&\binom{K}{S}\left(\mathbb{E}_{p,\sigma}\left(\max_{|I|=S}\|P_{I_p}(\Phi)\Phi c_{p,\sigma}\|_2^2\right) - \mathbb{E}_{p,\sigma}\left(\max_{|I|=S}\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2\right)\right)\\
&\quad= \frac{1-\gamma_S^2}{K-S}\sum_I\left(\|P_I(\Phi)\Phi\|_F^2 - \|P_I(\Psi)\Phi\|_F^2\right) + \lambda_S\sum_I\left(\|P_I(\Phi)\Phi_I\|_F^2 - \|P_I(\Psi)\Phi_I\|_F^2\right)\\
&\quad= \frac{1-\gamma_S^2}{K-S}\sum_I\operatorname{tr}\left(\Phi^\star\left(P_I(\Phi)-P_I(\Psi)\right)\Phi\right) + \lambda_S\sum_I\|Q_I(\Phi)B_I\|_F^2 + O(\varepsilon^3)\\
&\quad= \frac{1-\gamma_S^2}{K-S}\sum_I 2\operatorname{tr}\left(\Phi^\star Q_I(\Phi)B_I\Phi_I^\dagger\Phi\right) + O(\varepsilon^2) + \lambda_S\sum_I\|Q_I(\Phi)B_I\|_F^2 + O(\varepsilon^3)
\end{aligned} \tag{29}$$

The term $\sum_I \operatorname{tr}\left(\Phi^\star Q_I(\Phi)B_I\Phi_I^\dagger\Phi\right)$ is linear in B and thus can be negative. Since it is also of order $O(\varepsilon)$ whenever $\gamma_S < 1$ a necessary condition to have a local maximum exactly at Φ is that for all $B_I = Z_I W_I A_I^{-1}$,

$$\sum_I \operatorname{tr}\left(\Phi^\star Q_I(\Phi)B_I\Phi_I^\dagger\Phi\right) = 0. \tag{30}$$

In case $S = 1$ we have $Q_i(\Phi)b_i = b_i$ since $b_i \perp \phi_i$ and the condition above reduces to

$$\sum_i \operatorname{tr}\left(\Phi^\star b_i\phi_i^\star\Phi\right) = 0 \quad\Leftrightarrow\quad \sum_i \phi_i^\star\Phi\Phi^\star b_i = 0 \tag{31}$$

Choosing in turn $\omega_k = 0$ except for $k = i$ this means that for all i and $z_i \perp \phi_i$ we need to have

$$\frac{\omega_i}{\alpha_i}\langle z_i, \Phi\Phi^\star\phi_i\rangle = 0, \tag{32}$$

which is equivalent to every atom ϕ_i being an eigenvector of the frame operator, i.e. $\Phi\Phi^\star\phi_i = \lambda_i\phi_i$, $\forall i$. While this condition is certainly fulfilled when Φ is a tight frame (corresponding to $\lambda_i = A$), it is sufficient for Φ to be a collection of m tight frames for m orthogonal subspaces of $\mathbb{R}^d$ - corresponding to the case $\Phi = (\Phi_{\lambda_1}, \ldots, \Phi_{\lambda_m})$ with $\Phi\Phi^\star\Phi_{\lambda_i} = \lambda_i\Phi_{\lambda_i}$. Going through the same analysis as in the proof of Theorem 3.1 we see that in this second case Φ is again a local maximum under the additional condition that $c_1^2 > \frac{B-A+1}{B-A+K}$, where $A = \min_i \lambda_i$ and $B = \max_i \lambda_i$. In case $S > 1$, Condition (30) is again implied by tightness of the dictionary but it is an open question whether conversely it implies tightness of the dictionary. However, for simplicity we will henceforth restrict our analysis to the situation where Φ is a tight frame.

3.2 A continuous model of decaying coefficients

After proving a recovery result for the simple coefficient model of the last section we would like to extend it to a wider range of coefficient distributions, especially continuous ones.
Looking back at the proof of Theorem 3.1 we see that apart from the condition ensuring optimality of the projection P_{I_p} it also relied heavily on the equal probability of all sign sequences and permutations changing our base coefficient sequence. We therefore make the following definition.

Definition 3.1. *A probability measure ν on the unit sphere $S^{K-1} \subset \mathbb{R}^K$ is called symmetric if for all measurable sets $\mathcal{X} \subseteq S^{K-1}$, for all sign sequences $\sigma \in \{-1, 1\}^K$ and all permutations p we have*

$$\nu(\sigma\mathcal{X}) = \nu(\mathcal{X}), \quad \textit{where} \quad \sigma\mathcal{X} := \{(\sigma_1 x_1, \ldots, \sigma_K x_K) : x \in \mathcal{X}\}, \quad \textit{and} \tag{33}$$

$$\nu(p(\mathcal{X})) = \nu(\mathcal{X}), \quad \textit{where} \quad p(\mathcal{X}) := \{(x_{p(1)}, \ldots, x_{p(K)}) : x \in \mathcal{X}\}. \tag{34}$$

We are now ready to state a version of Theorem 3.1 for more general coefficient distributions.

Theorem 3.3. *Let Φ be a unit norm tight frame with frame constant $A = K/d$ and lower isometry constant δ_S. Let x be drawn from a symmetric probability distribution ν on the unit sphere and assume that the signals are generated as $y = \Phi x$. If there exists $\kappa > 0$ such that for $c(x)$ a non-increasing rearrangement of the absolute values of x and $I_p := p^{-1}(\{1, \ldots S\})$ we have,*

$$\nu\left(\min_{p,\sigma}\left(\|P_{I_p}(\Phi)\Phi c_{p,\sigma}(x)\|_2 - \max_{|I|=S, I\neq I_p}\|P_I(\Phi)\Phi c_{p,\sigma}(x)\|_2\right) \geq 2\kappa\right) = 1 \tag{35}$$

then there is a local maximum of (4) *at* Φ.

Moreover for $\Psi \neq \Phi$ *we have* $\mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Psi)y\|_2^2\right) < \mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Phi)y\|_2^2\right)$ *as soon as*

$$d(\Phi,\Psi) \leq \frac{\kappa\sqrt{1-\delta_S}}{\sqrt{\frac{9S}{2}}\left(1+4\sqrt{\log\left(\frac{60A\binom{\kappa}{S}^2}{\kappa\bar{\lambda}_S(1-\delta_S)}\right)}\right)}, \tag{36}$$

where $\bar{\lambda}_S = \frac{E_x(c_1^2(x)+\ldots+c_S^2(x))}{S} - \frac{1-E_x(c_1^2(x)+\ldots+c_S^2(x))}{K-S}$ *and* $\delta_S < 1$ *because of* (35).

Proof: Let c denote the mapping that assigns to each $x \in S^{K-1}$ the non increasing rearrangement of the absolute values of its components, i.e. $c_i(x) = |x_{p(i)}|$ for a permutation p such that $c_1(x) \geq c_2(x) \geq \ldots \geq c_K(x) \geq 0$. Then the mapping c together with the probability measure ν on S^{K-1} induces a pull-back probability measure ν_c on $c(S^{K-1})$, by $\nu_c(\Omega) := \nu(c^{-1}(\Omega))$ for any measurable set $\Omega \subseteq c(S^{K-1})$. With the help of this new measure we can rewrite the expectations we need to calculate as,

$$\begin{aligned}\mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Phi)y\|_2^2\right) &= \mathbb{E}_x\left(\max_{|I|=S}\|P_I(\Phi)\Phi x\|_2^2\right)\\ &= \int_x \max_{|I|=S}\|P_I(\Phi)\Phi x\|_2^2 d\nu(x)\\ &= \int_{c(x)} \mathbb{E}_p\mathbb{E}_\sigma \max_{|I|=S}\|P_I(\Phi)c_{p,\sigma}(x)\|_2^2 d\nu_c(x).\end{aligned} \tag{37}$$

The expectation inside the integral should seem familiar. Indeed we have calculated it already in the proof of Theorem 3.1 for $c(x)$ a fixed decaying sequence satisfying

$$\|P_{I_p}(\Phi)\Phi c_{p,\sigma}\|_2 - \max_{|I|=S, I\neq I_p}\|P_I(\Phi)\Phi c_{p,\sigma}\|_2 \geq 2\kappa, \quad \forall\sigma, p. \tag{38}$$

By (35) this property is satisfied almost surely and so by applying Lemma A.1 we get,

$$\begin{aligned}\mathbb{E}_x\left(\max_{|I|=S}\|P_I(\Phi)\Phi x\|_2^2\right) &= \int_{c(x)} \mathbb{E}_p\mathbb{E}_\sigma\|P_{I_p}(\Phi)c_{p,\sigma}(x)\|_2^2 d\nu_c(x)\\ &= \int_{c(x)} \frac{A(1-\bar{\gamma}_S^2(x))S}{(K-S)} + \left(\frac{\bar{\gamma}_S^2(x)}{S} - \frac{1-\bar{\gamma}_S^2(x)}{K-S}\right)\binom{K}{S}^{-1}\sum_{I:|I|=S}\|\Phi_I\|_F^2 d\nu_c(x),\end{aligned}$$

where $\gamma_S^2(x) := c_1^2(x) + \ldots + c_S^2(x)$. Since for the integral term we simply have

$$\int_{c(x)} \gamma_S^2(x) d\nu_c(x) = \mathbb{E}_x\left(\max_{|I|=S}\|x_I\|_2^2\right) = \bar{\gamma_S}^2, \tag{39}$$

we arrive at the following estimate analogue to (10)

$$\mathbb{E}_x\left(\max_{|I|=S}\|P_I(\Phi)\Phi x\|_2^2\right) = \frac{A(1-\bar{\gamma}_S^2)S}{(K-S)} + \left(\frac{\bar{\gamma}_S^2}{S} - \frac{1-\bar{\gamma}_S^2}{K-S}\right)\binom{K}{S}^{-1}\sum_{I:|I|=S}\|\Phi_I\|_F^2. \tag{40}$$

Using the same argument we also get an estimate for the expectation of a perturbed dictionary analogue to (41), i.e.

$$\mathbb{E}_x\left(\max_{|I|=S}\|P_I(\Psi)\Phi x\|_2^2\right) \leq \frac{A(1-\bar{\gamma}_S^2)S}{(K-S)} + \left(\frac{\bar{\gamma}_S^2}{S} - \frac{1-\bar{\gamma}_S^2}{K-S}\right)\binom{K}{S}^{-1}\sum_{I:|I|=S}\|P_I(\Psi)\Phi_I\|_F^2 + 4\eta_S A \max_{|I|=S}\|P_I(\Psi)-P_I(\Phi)\|_F. \tag{41}$$

where

$$\eta_S = 2\sum_{I:P_I(\Psi)\neq P_I(\Phi)}\exp\left(\frac{-\kappa^2}{32\|P_I(\Psi)-P_I(\Phi)\|_F^2}\right). \tag{42}$$

The rest of the proof simply consists of replacing γ_S with $\bar{\gamma}_S$ in the proof of Theorem 3.1. □

Remark 3.3. (a) Again the abstract condition in (35) can be satisfied, i.e. we can find $\kappa > 0$, if we have $S\mu < 1/2$ and

$$\nu\left(c_S(x) > \frac{1-S\mu}{1-2S\mu}c_{S+1}(x) + \frac{4\mu}{1-2S\mu}\sum_{i>S+1}|c_i(x)|\right) = 1. \tag{43}$$

(b) Note that with the available tools it is also be possible to extend Theorem 3.3 to signal models with coefficient distributions approaching the limit in (35), i.e. $\kappa = 0$. However to keep the presentation concise we will not go into further details here but refer the interested reader to [33] or [34] for the proof idea and some simple example distributions approaching the limit in the case of an orthonormal basis.

3.3 Bounded white noise

With the tools used to prove the two noiseless identification results in the last two subsections it is also possible to analyse the case of (very small) bounded white noise.

Theorem 3.4. *Let Φ be a unit norm tight frame with frame constant $A = K/d$ and lower isometry constant δ_S. Assume that the signals y are generated as $y = \Phi x + r$, where x is drawn from a symmetric decaying probability distribution ν on the unit sphere S^{K-1} and r is a bounded random white noise vector, i.e. there exist two constants $\rho, \rho_{\max}$ such that $\|r\|_2 \leq \rho_{\max}$ almost surely, $\mathbb{E}(r) = 0$ and $\mathbb{E}(rr^\star) = \rho^2 I$. If there exists $\kappa > 0$ such that for $c(x)$ a non-increasing rearrangement of the absolute values of x and $I_p := p^{-1}(\{1,\dots S\})$ we have,*

$$\nu\left(\min_{p,\sigma}\left(\|P_{I_p}(\Phi)\Phi c_{p,\sigma}(x)\|_2 - \max_{|I|=S, I\neq I_p}\|P_I(\Phi)\Phi c_{p,\sigma}(x)\|_2\right) \geq 2\kappa + 2\rho_{\max}\right) = 1, \tag{44}$$

then there is a local maximum of (4) *at* Φ.

Moreover for $\Psi \neq \Phi$ *we have* $\mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Psi)y\|_2^2\right) < \mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Phi)y\|_2^2\right)$ *as soon as*

$$d(\Phi,\Psi) \leq \frac{\kappa\sqrt{1-\delta_S}}{\sqrt{\frac{9S}{2}}\left(1+4\sqrt{\log\left(\frac{60A_r\binom{\kappa}{S}^2}{\kappa\bar{\lambda}_S(1-\delta_S)}\right)}\right)}, \tag{45}$$

where $\bar{\lambda}_S = \frac{E_x(c_1^2(x)+\ldots+c_S^2(x))}{S} - \frac{1-E_x(c_1^2(x)+\ldots+c_S^2(x))}{K-S}$ *and* $A_r = (\sqrt{A}+\rho_{\max})^2$. *Again* $\delta_S < 1$ *is implied by* (44).

Proof: We streamline the proof, since it relies on the same ideas as those of Theorem 3.1 and Theorem 3.3. For a noisy signal $y = \Phi x + r = \Phi c_{p,\sigma}(x) + r$ the condition in (44) guarantees that the maximal response for the original dictionary Φ is taken at I_p , since we have

$$\|P_{I_p}(\Phi)y\|_2 \geq \|P_{I_p}(\Phi)\Phi c_{p,\sigma}(x)\|_2 + \|P_I(\Phi)r\|_2 \geq \|P_{I_p}(\Phi)\Phi c_{p,\sigma}(x)\|_2 - \rho_{\max}, \tag{46}$$

$$\|P_I(\Phi)y\|_2 \leq \|P_I(\Phi)\Phi c_{p,\sigma}(x)\|_2 + \|P_I(\Phi)r\|_2 \leq \|P_I(\Phi)\Phi c_{p,\sigma}(x)\|_2 + \rho_{\max}. \tag{47}$$

Thus we get,

$$\begin{aligned}
\mathbb{E}_y\left(\max_{|I|=S}\|P_I(\Phi)y\|_2^2\right) &= \mathbb{E}_{r,x}\left(\max_{|I|=S}\|P_I(\Phi)y\|_2^2\right)\\
&= \mathbb{E}_r\left(\int_{c(x)}\mathbb{E}_p\mathbb{E}_\sigma\max_{|I|=S}\|P_I(\Phi)\left((\Phi)c_{p,\sigma}(x)+r\right)\|_2^2 d\nu_c(x)\right)\\
&= \mathbb{E}_r\left(\int_{c(x)}\mathbb{E}_p\mathbb{E}_\sigma\|P_{I_p}\left((\Phi)c_{p,\sigma}(x)+r\right)\|_2^2 d\nu_c(x)\right)\\
&= \int_{c(x)}\mathbb{E}_p\mathbb{E}_\sigma\mathbb{E}_r\left(\|P_{I_p}(\Phi)\left((\Phi)c_{p,\sigma}(x)+r\right)\|_2^2\right)d\nu_c(x)\\
&= \int_{c(x)}\mathbb{E}_p\mathbb{E}_\sigma\left(\|P_{I_p}(\Phi)c_{p,\sigma}(x)\| + \mathbb{E}_r\|P_{I_p}(\Phi)r\|_2^2\right)d\nu_c(x)\\
&= \frac{A(1-\bar{\gamma}_S^2)S}{(K-S)} + \left(\frac{\bar{\gamma}_S^2}{S} - \frac{1-\bar{\gamma}_S^2}{K-S}\right)\binom{K}{S}^{-1}\sum_{I:|I|=S}\|\Phi_I\|_F^2 + S\rho^2.
\end{aligned} \tag{48}$$

For a perturbed dictionary and a noisy signal y we can bound the response using the set I analogue to (11),

$$\begin{aligned}
\|P_I(\Psi)y\|_2^2 &\leq \|P_I(\Phi)y\|_2^2 + 2\|y\|_2^2\|P_I(\Psi)-P_I(\Phi)\|_{2,2}\\
&\leq \|P_I(\Phi)y\|_2^2 + 2(\sqrt{A}+\rho_{\max})^2\max_{|I|=S}\|P_I(\Psi)-P_I(\Phi)\|_F\\
&\leq \|P_{I_p}(\Phi)y\|_2^2 + 2(\sqrt{A}+\rho_{\max})^2\max_{|I|=S}\|P_I(\Psi)-P_I(\Phi)\|_F.
\end{aligned} \tag{49}$$

Reversing the roles of Ψ and Φ and setting $I = I_p$ in the inequality above then leads to

$$\|P_I(\Psi)y\|_2^2 \leq \|P_{I_p}(\Psi)y\|_2^2 + 4(\sqrt{A} + \rho_{\max})^2 \max_{|I|=S} \|P_I(\Psi) - P_I(\Phi)\|_F. \tag{50}$$

Using the sets Σ_p as defined in (15) we get the following estimate for $y = \Phi c_{p,\sigma}(x) + r$ with $\sigma \notin \Sigma_p$ and all $I \neq I_p$,

$$\begin{aligned}
\|P_I(\Psi)y\|_2 &\leq \|P_I(\Psi)c_{p,\sigma}(x)\|_2 + \|P_I(\Phi)r\|_2 \\
&\leq \|P_I(\Phi)c_{p,\sigma}(x)\|_2 + \| \left(P_I(\Psi) - P_I(\Phi)\right) c_{p,\sigma}(x)\|_2 + \rho_{\max} \\
&\leq \|P_I(\Phi)c_{p,\sigma}(x)\|_2 + \kappa + \rho_{\max} \\
&\leq \|P_{I_p}(\Phi)c_{p,\sigma}(x)\|_2 - \kappa - \rho_{\max} \\
&\leq \|P_{I_p}(\Phi)c_{p,\sigma}(x)\|_2 - \| \left(P_{I_p}(\Psi) - P_{I_p}(\Phi)\right) c_{p,\sigma}(x)\|_2 - \rho_{\max} \\
&\leq \|P_{I_p}(\Psi)c_{p,\sigma}(x)\|_2 - \|P_{I_p}(\Phi)r\|_2 \leq \|P_{I_p}(\Psi)y\|_2.
\end{aligned} \tag{51}$$

Thus we can estimate the expectation for a perturbed dictionary as

$$\begin{aligned}
\mathbb{E}_y\left(\max_{|I|=S} \|P_I(\Psi)y\|_2^2\right) &= \mathbb{E}_{r,x}\left(\max_{|I|=S} \|P_I(\Psi)y\|_2^2\right) \\
&= \mathbb{E}_r\left(\int_{c(x)} \mathbb{E}_p\mathbb{E}_\sigma\left(\max_{|I|=S} \|P_I(\Psi)\left(\Phi c_{p,\sigma}(x) + r\right)\|_2^2\right) d\nu_c(x)\right) \\
&= \mathbb{E}_r\left(\int_{c(x)} \mathbb{E}_p\left(\sum_{\sigma\in\Sigma_p} \max_{|I|=S} \|\dots\|_2^2 + \sum_{\sigma\notin\Sigma_p} \max_{|I|=S} \|\dots\|_2^2\right) d\nu_c(x)\right) \\
&\leq \mathbb{E}_r\left(\int_{c(x)} \mathbb{E}_p\mathbb{E}_\sigma\left(\|P_{I_p}(\Psi)\left(\Phi c_{p,\sigma}(x) + r\right)\|_2\right) d\nu_c(x)\right) \\
&\qquad + 4\eta_S(\sqrt{A} + \rho_{\max})^2 \max_{|I|=S} \|P_I(\Psi) - P_I(\Phi)\|_F \\
&\leq \frac{A(1-\bar{\gamma}_S^2)S}{(K-S)} + \left(\frac{\bar{\gamma}_S^2}{S} - \frac{1-\bar{\gamma}_S^2}{K-S}\right)\binom{K}{S}^{-1} \sum_{I:|I|=S} \|P_I(\Psi)\Phi_I\|_F^2 + S\rho^2 \\
&\qquad + 4\eta_S(\sqrt{A} + \rho_{\max})^2 \max_{|I|=S} \|P_I(\Psi) - P_I(\Phi)\|_F.
\end{aligned} \tag{52}$$

The rest of the proof simply consists of replacing γ_S with $\bar{\gamma}_S$ and A with $(\sqrt{A} + \rho_{\max})^2$ in the proof of Theorem 3.1. □

4 FINITE SAMPLE SIZE RESULTS

Finally we make the step from the asymptotic identification results derived in the last section to an identification result for a finite number of training samples. We consider the maximisation

problem,

$$\max_{\Psi\in\mathcal{D}} \frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\Psi)y_n\|_2^2. \tag{53}$$

The main idea is that whenever Ψ is near to Φ we have

$$\frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\Psi)y_n\|_2^2 \approx \mathbb{E}_y\left(\max_{|I|=S} \|P_I(\Psi)y\|_2^2\right) < \mathbb{E}_y\left(\max_{|I|=S} \|P_I(\Phi)y\|_2^2\right) \approx \frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\Phi)y_n\|_2^2.$$

Concretising the sharpness of $\approx$ quantitatively and making sure that it is valid for all possible ε-perturbations at the same time, leads to the following theorem.

Theorem 4.1. *Let Φ be a unit norm tight frame with frame constant $A = K/d$ and lower isometry constant $\delta_S < 1 - \frac{S}{d}$. Assume that the signals y_n are generated as $y_n = \Phi x_n$, where x_n is drawn from a symmetric decaying probability distribution ν on the unit sphere S^{K-1} and r is a bounded random white noise vector with $\|r\|_2 \leq \rho_{\max}$ almost surely, $\mathbb{E}(r) = 0$ and $\mathbb{E}(rr^\star) = \rho^2$. Further assume that there exists $\kappa > 0$ such that for $c(x)$ a non-increasing rearrangement of the absolute values of x and $I_p := p^{-1}(\{1,\dots S\})$ we have,*

$$\nu\left(\min_{p,\sigma}\left(\|P_{I_p}(\Phi)\Phi c_{p,\sigma}(x)\|_2 - \max_{|I|=S, I\neq I_p} \|P_I(\Phi)\Phi c_{p,\sigma}(x)\|_2\right) \geq 2\kappa + 2\rho_{\max}\right) = 1. \tag{54}$$

Abbreviate $\bar{\lambda}_S = \frac{E_x(c_1^2(x)+\dots+c_S^2(x))}{S} - \frac{1-E_x(c_1^2(x)+\dots+c_S^2(x))}{K-S}$, $A_r = (\sqrt{A}+\rho_{\max})^2$ and $C_S := 1 - \frac{S}{d(1-\delta_S)}$. If for some $0 < q < 1/4$ the number of samples N satisfies

$$2N^{-q} + N^{-2q} \leq \frac{\kappa\sqrt{1-\delta_S}}{\sqrt{\frac{9S}{2}}\left(1 + 4\sqrt{\log\left(\frac{135A_rK\binom{K}{S}}{\kappa\bar{\lambda}_S C_S S(1-\delta_S)}\right)}\right)}, \tag{55}$$

then except with probability

$$\exp\left(-\frac{N^{1-4q}\bar{\lambda}_S^2 S^2 C_S^2}{4K^2A_r^2} + Kd\log\left(\frac{NKA_r}{2\bar{\lambda}_S S C_S}\right)\right), \tag{56}$$

there is a local maximum of (53) *resp. local minimum of* (1) *within distance at most $2N^{-q}$ to Φ, i.e. for the local maximum $\tilde{\Psi}$ we have $\dot{(}\tilde{\Psi}, \Phi) \leq 2N^{-q}$.*

Proof: Conceptually we need to show that for some $\varepsilon_{\min}(N) < \varepsilon_{\max}(N)$ and with probability $p(N)$ for all perturbations Ψ with $\varepsilon_{\min}(N) \leq d(\Psi,\Phi) \leq \varepsilon_{\max}(N)$ we have

$$\frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\Psi)y_n\|_2^2 < \frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\Phi)y_n\|_2^2 \tag{57}$$

To do this we need to add three ingredients to the asymptotic results of Theorem 3.4, 1) that with high probability for a fixed dictionary Ψ the sum of signal responses concentrates around

its expectation, 2) a dense enough net for the space of all perturbations and 3) and a Lipschitz-type bound for the mapping $\Psi \longrightarrow \max_{|I|=S} \|P_I(\Psi)y_n\|_2^2$. Then we can argue that an arbitrary perturbation will be close to a perturbation in the net, for which the sum concentrates around its expectation. This expectation is in turn is smaller than the expectation of the generating dictionary, around which the sum for the generating dictionary concentrates.

We start with the Lipschitz-type bound for the mapping $\Psi \longrightarrow \max_{|I|=S} \|P_I(\Psi)y_n\|_2^2$ on the set of perturbations with $d(\Psi, \Phi) \leq \varepsilon_{\max}$. Analogue to (11) we have for any index set I of size S,

$$\begin{aligned}\|P_I(\Psi)y\|_2^2 &\leq \|P_I(\bar{\Psi})y\|_2^2 + 2A_r \max_{|I|=S} \|P_I(\Psi) - P_I(\Phi)\|_F \\ &\leq \max_{|I|=S} \|P_I(\bar{\Psi})y\|_2^2 + 2A_r \max_{|I|=S} \|P_I(\Psi) - P_I(\Phi)\|_F. \end{aligned} \tag{58}$$

Since this is true for all I we further get that

$$\max_{|I|=S} \|P_I(\Psi)y\|_2^2 \leq \max_{|I|=S} \|P_I(\bar{\Psi})y\|_2^2 + 2A_r \max_{|I|=S} \|P_I(\Psi) - P_I(\Phi)\|_F,$$

and reversing the roles of Ψ and $\bar{\Psi}$ leads to

$$\left| \max_{|I|=S} \|P_I(\Psi)y\|_2^2 - \max_{|I|=S} \|P_I(\bar{\Psi})y\|_2^2 \right| \leq 2A_r \max_{|I|=S} \|P_I(\Psi) - P_I(\Phi)\|_F. \tag{59}$$

From Lemma A.2 we know that

$$\|P_I(\Psi) - P_I(\bar{\Psi})\|_F^2 \leq \frac{2S\frac{d(\Psi,\bar{\Psi})^2}{1-d(\Psi,\bar{\Psi})^2}}{\|\Psi_I^\dagger\|_{2,2}^{-1}\left(\|\Psi_I^\dagger\|_{2,2}^{-1} - 2\sqrt{S}\frac{d(\Psi,\bar{\Psi})}{\sqrt{1-d(\Psi,\bar{\Psi})^2}}\right)}.$$

Now note that $\|\Psi_I^\dagger\|_{2,2}^{-1}$ is simply the minimal singular value of Ψ_I. Since we have $\delta_S < 1 - S/d$ we get,

$$\begin{aligned}\|\Psi_I^\dagger\|_{2,2}^{-1} = \sigma_{\min}(\Psi_I) = \sigma_{\min}(\Phi_I A_I + Z_I W_I) &\geq \sigma_{\min}(\Phi_I)\sigma_{\min}(A_I) - \sigma_{\max}(Z_I W_I) \\ &\geq \sqrt{1-\delta_S}(1-\varepsilon^2/2) - \sqrt{S}\varepsilon. \end{aligned} \tag{60}$$

The combination of the last three estimates, together with some simplifications, using the fact that both ε and $d(\Psi, \bar{\Psi})$ will be smaller than $\varepsilon_{\max} \leq \frac{\sqrt{1-\delta_S}}{21\sqrt{S}}$, leads to the final bound,

$$\left| \max_{|I|=S} \|P_I(\Psi)y_n\|_2^2 - \max_{|I|=S} \|P_I(\bar{\Psi})y_n\|_2^2 \right| \leq d(\Psi, \bar{\Psi}) \cdot \frac{C_L A_r \sqrt{S}}{\sqrt{1-\delta_S}}, \tag{61}$$

with $C_L = 3.139$. Next for $Y_n = \max_{|I|=S} \|P_I(\Psi)y_n\|_2^2$ we have $Y_n \in [0, A_r]$ and therefore by Hoeffding's inequality,

$$\mathbb{P}\left(\left| \frac{1}{N} \sum_{n=1}^{N} \max_{|I|=S} \|P_I(\Psi)y_n\|_2^2 - \mathbb{E}(\max_{|I|=S} \|P_I(\Psi)y_1\|_2^2) \right| \geq t \right) \leq e^{-Nt^2/A_r^2}.$$

The last ingredient is a δ-net for all perturbations Ψ with $d(\Psi, \Psi) \leq \varepsilon_{\max}$, i.e. a finite set of perturbations $\mathcal{N}$ such that for every Ψ we can find $\bar{\Psi} \in \mathcal{N}$ with $d(\Psi, \bar{\Psi}) < \delta$. Remembering the parametrisation of all ε-perturbations from the proof of Theorem 3.1 we see that the space we need to cover is the product of K balls with radius $\varepsilon_{\max}$ in $\mathbb{R}^d$. Following for example the argument in Lemma 2 of [41] we know that for the m-dimensional ball of radius $\varepsilon_{\max}$ we can find a δ net $\mathcal{N}_d$ with

$$\sharp\mathcal{N}_d \leq \left(\varepsilon_{\max} + \frac{2\varepsilon_{\max}}{\delta}\right)^d.$$

Thus for the product of K balls in $\mathbb{R}^d$ we can construct a δ-net $\mathcal{N}$ as the product of K δ-nets $\mathcal{N}_d$. Assuming that $\delta < 1$ we then have,

$$\sharp\mathcal{N} \leq \left(\varepsilon_{\max} + \frac{2\varepsilon_{\max}}{\delta}\right)^{Kd} \leq \left(\frac{3\varepsilon_{\max}}{\delta}\right)^{Kd}.$$

Using a union bound we can now estimate the probability that for all perturbations in the net the sum of responses concentrates around its expectation, as

$$\mathbb{P}\left(\exists\bar{\Psi} \in \mathcal{N} : \left|\frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\Psi)y_n\|_2^2 - \mathbb{E}(\max_{|I|=S} \|P_I(\Psi)y_1\|_2^2)\right| \geq t\right) \leq \left(\frac{3\varepsilon_{\max}}{\delta}\right)^{Kd} e^{-Nt^2/A_r^2}.$$

We can now turn to the triangle inequality argument. For a perturbation Ψ with $d(\Psi, \Phi) = \varepsilon \leq \varepsilon_{\max}$ we can find $\bar{\Psi} \in \mathcal{N}$ with $d(\Psi, \bar{\Psi}) \leq \delta$ and $d(\bar{\Psi}, \Phi) = \bar{\varepsilon}$. We then have

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\Psi)y_n\|_2^2 - \frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\Phi)y_n\|_2^2 \\
&\quad= \frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\Psi)y_n\|_2^2 - \frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\bar{\Psi})y_n\|_2^2 \\
&\qquad\quad+ \frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\bar{\Psi})y_n\|_2^2 - \mathbb{E}\left(\max_{|I|=S} \|P_I(\bar{\Psi})y_n\|_2^2\right) \\
&\qquad\qquad+ \mathbb{E}\left(\max_{|I|=S} \|P_I(\bar{\Psi})y_n\|_2^2\right) - \mathbb{E}\left(\max_{|I|=S} \|P_I(\Phi)y_n\|_2^2\right) \\
&\qquad\qquad\quad+ \mathbb{E}\left(\max_{|I|=S} \|P_I(\Phi)y_n\|_2^2\right) - \frac{1}{N}\sum_{n=1}^{N} \max_{|I|=S} \|P_I(\Phi)y_n\|_2^2 \\
&\leq \mathbb{E}\left(\max_{|I|=S} \|P_I(\bar{\Psi})y_n\|_2^2\right) - \mathbb{E}\left(\max_{|I|=S} \|P_I(\Phi)y_n\|_2^2\right) + 2t + \delta\frac{C_L A_r\sqrt{S}}{\sqrt{1-\delta_S}}. \qquad (62)
\end{aligned}$$

Using the expression for the respective expectations for a noisy signal from (48) and (52) and

the abbreviation $\bar{\lambda}_S = \frac{\bar{\gamma}_S^2}{S} - \frac{1-\bar{\gamma}_S^2}{K-S}$ we get,

$$\begin{aligned}
&\frac{1}{N}\sum_{n=1}^{N}\max_{|I|=S}\|P_I(\Psi)y_n\|_2^2 - \frac{1}{N}\sum_{n=1}^{N}\max_{|I|=S}\|P_I(\Phi)y_n\|_2^2 \\
&\qquad \leq 4A_r \max_{|I|=S}\|P_I(\bar{\Psi}) - P_I(\Phi)\|_F \sum_{I:\ldots}\exp\left(\frac{-\kappa^2}{32\|P_I(\bar{\Psi}) - P_I(\Phi)\|_F^2}\right) \\
&\qquad\qquad + \bar{\lambda}_S\binom{K}{S}^{-1}\sum_{|I|=S}\left(\|P_I(\bar{\Psi})\Phi_I\|_F^2 - \|\Phi_I\|_F^2\right) + 2t + \delta\frac{C_L A_r\sqrt{S}}{\sqrt{1-\delta_S}} \\
&\qquad \leq \frac{4A_rC_1}{\sqrt{1-\delta_S}}\max_{|I|=S}\|\bar{B}_I\|_F\binom{K}{S}\exp\left(\frac{-\kappa^2(1-\delta_S)}{32C_1^2\max_{|I|=S}\|\bar{B}_I\|_F^2}\right) \\
&\qquad\qquad - C_2\bar{\lambda}_S\binom{K}{S}^{-1}\sum_{|I|=S}\|Q_I(\Phi)\bar{B}_I\|_F^2 + 2t + \delta\frac{C_L A_r\sqrt{S}}{\sqrt{1-\delta_S}},
\end{aligned} \tag{63}$$

where we have used Lemma A.2 and that $\|Q_I(\Phi)\bar{B}_I\|_F \leq \|\bar{B}_I\|_F$. Using the condition on the isometry constant we now derive a (for all practical purposes) sharper lower bound than simply $\max_I \|Q_I(\Phi)\bar{B}_I\|_F^2$ for the sum in the equation above,

$$\begin{aligned}
\binom{K}{S}^{-1}\sum_{I}\|Q_I(\Phi)\bar{B}_I\|_F^2 &= \binom{K}{S}^{-1}\sum_{I}\left(\|\bar{B}_I\|_F^2 - \|P_I(\Phi)\bar{B}_I\|_F^2\right) \\
&= \binom{K}{S}^{-1}\binom{K-1}{S-1}\|\bar{B}\|_F^2 - \binom{K}{S}^{-1}\sum_{I}\|(\Phi_I^\dagger)^\star\Phi_I^\star\bar{B}_I\|_F^2 \\
&\geq \frac{S}{K}\|\bar{B}\|_F^2 - \binom{K}{S}^{-1}\sum_{I}\|\Phi_I^\dagger\|_{2,2}^2\|\Phi_I^\star\bar{B}_I\|_F^2 \\
&\geq \frac{S}{K}\|\bar{B}\|_F^2 - \frac{1}{1-\delta_S}\binom{K}{S}^{-1}\sum_{I}\|\Phi_I^\star\bar{B}_I\|_F^2 \\
&\geq \frac{S}{K}\|\bar{B}\|_F^2 - \frac{1}{1-\delta_S}\binom{K}{S}^{-1}\binom{K-2}{S-2}\|\Phi^\star\bar{B}\|_F^2 \\
&\geq \frac{S}{K}\left(1 - \frac{A}{1-\delta_S}\frac{S-1}{K-1}\right)\|\bar{B}\|_F^2.
\end{aligned} \tag{64}$$

Using $A = K/d$ and denoting the index set for which $\|\bar{B}_I\|_F$ is maximal by $\bar{I}$ then leads to the bound

$$\binom{K}{S}^{-1}\sum_{I}\|Q_I(\Phi)\bar{B}_I\|_F^2 \geq \frac{S}{K}\left(1 - \frac{S}{d(1-\delta_S)}\right)\max_{|I|=S}\|\bar{B}_I\|_F^2. \tag{65}$$

Substituting the estimate above into (63) we further get

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}\max_{|I|=S}\|P_I(\Psi)y_n\|_2^2-\frac{1}{N}\sum_{n=1}^{N}\max_{|I|=S}\|P_I(\Phi)y_n\|_2^2\\ \leq -\frac{C_2\bar{\lambda}_S S}{K}\left(1-\frac{S}{d(1-\delta_S)}\right)\|\bar{B}_{\bar{I}}\|_F^2\\ +\frac{4A_rC_1}{\sqrt{1-\delta_S}}\binom{K}{S}\|\bar{B}_{\bar{I}}\|_F\exp\left(\frac{-\kappa^2(1-\delta_S)}{32C_1^2\|\bar{B}_{\bar{I}}\|_F^2}\right)+2t+\delta\frac{C_LA_r\sqrt{S}}{\sqrt{1-\delta_S}},\end{aligned}$$

with $C_1=1.487$ and $C_2=0.897$. Abbreviating $C_S=1-\frac{S}{d(1-\delta_S)}$ by Lemma A.3 we have

$$\frac{4A_rC_1}{\sqrt{1-\delta_S}}\binom{K}{S}\|\bar{B}_{\bar{I}}\|_F\exp\left(\frac{-\kappa^2(1-\delta_S)}{32C_1^2\|\bar{B}_{\bar{I}}\|_F^2}\right)\leq\frac{(C_2-0.5)\bar{\lambda}_SC_SS}{K}\|\bar{B}_{\bar{I}}\|_F^2 \tag{66}$$

as soon as

$$\|B_{\bar{I}}\|_F\leq\frac{4\kappa\sqrt{1-\delta_S}}{C_1\sqrt{32}\left(1+4\sqrt{\log\left(\frac{4\sqrt{32}C_1^2e^{1/16}AK\binom{K}{S}}{(C_2-0.5)\kappa\bar{\lambda}_SC_SS(1-\delta_S)}\right)}\right)}, \tag{67}$$

which is satisfied if

$$\bar{\varepsilon}\leq\frac{\kappa\sqrt{1-\delta_S}}{\sqrt{\frac{9S}{2}}\left(1+4\sqrt{\log\left(\frac{135A_rK\binom{K}{S}}{\kappa\bar{\lambda}_SC_SS(1-\delta_S)}\right)}\right)}:=\varepsilon_{\max}+\delta. \tag{68}$$

Under the condition above, which defines $\varepsilon_{\max}$ up to δ, we further have

$$\begin{aligned}\frac{1}{N}\sum_{n=1}^{N}\max_{|I|=S}\|P_I(\Psi)y_n\|_2^2-\frac{1}{N}\sum_{n=1}^{N}\max_{|I|=S}\|P_I(\Phi)y_n\|_2^2\\ \leq-\frac{\bar{\lambda}_SC_SS}{2K}\bar{\varepsilon}^2+2t+\delta\frac{C_LA_r\sqrt{S}}{\sqrt{1-\delta_S}}\\ \leq-\frac{\bar{\lambda}_SC_SS}{2K}(\varepsilon-\delta)^2+2t+\delta\frac{C_LA_r\sqrt{S}}{\sqrt{1-\delta_S}}\\ \leq-\frac{\bar{\lambda}_SC_SS}{2K}\varepsilon^2+2t+\delta(\frac{(C_L+\varepsilon_{\max})A_r\sqrt{S}}{\sqrt{1-\delta_S}}.\end{aligned} \tag{69}$$

We now choose $t=N^{-2q}\frac{\bar{\lambda}_SC_SS}{2K}$ and $\delta=N^{-2q}\frac{\bar{\lambda}_SC_S\sqrt{S(1-\delta_S)}}{(C_L+\varepsilon_{\max})A_rK}$ to get, that except with probability,

$$\exp\left(-\frac{N^{1-4q}\bar{\lambda}_S^2C_S^2S^2}{4K^2A_r^2}+Kd\log\left(\frac{3\varepsilon_{\max}(C_L+\varepsilon_{\max})A_rKN^{2q}}{\bar{\lambda}_SC_S\sqrt{S(1-\delta_S)}}\right)\right), \tag{70}$$

we have

$$\frac{1}{N}\sum_{n=1}^{N}\max_{|I|=S}\|P_I(\Psi)y_n\|_2^2-\frac{1}{N}\sum_{n=1}^{N}\max_{|I|=S}\|P_I(\Phi)y_n\|_2^2\leq-\frac{\bar{\lambda}_SC_SS}{2K}(\varepsilon^2-4N^{-2q}), \tag{71}$$

which is smaller than zero as long as $\varepsilon > 2N^{-q} := \varepsilon_{\min}$. The statement follows from the bound $3\varepsilon_{\max}(C_L + \varepsilon_{\max}) \leq \frac{\sqrt{1-\delta_S}}{2\sqrt{S}}$, and ensuring that $\varepsilon_{\min} < \varepsilon_{\max}$ using the crude bound $\delta \leq N^{-2q}$. $\square$

Remark 4.1. Note that in case $S = 1$ the above theorem is not only a result for the K-SVD minimisation principle but actually for K-SVD. While for $S > 1$ the decay-condition is not strong enough to ensure that the sparse approximation algorithm used for K-SVD always finds the best approximation as soon as we are close enough to the generating dictionary, in the case $S = 1$ any simple greedy algorithm, e.g. thresholding, will always find the best 1-term approximation to any signal given any dictionary. Thus given the right initialisation and sufficiently many training samples K-SVD can recover the generating dictionary up to the prescribed precision with high probability. To make the theorem more applicable we quickly concretise how the distance between the generating dictionary Φ and the local minimum output by K-SVD $\tilde{\Psi}$ decreases with the sample size. If we want the success probability to be of the order $1 - N^{-Kd}$ we need

$$\frac{-N^{1-4q}\lambda_S^2}{4K^2C_L^2} + Kd\log(NKC_L/\lambda_S) \approx -Kd\log N,$$

or $N^{1-4q} \approx K^3 d\log N$ meaning that $-q \approx -\frac{1}{4} + \frac{\log K}{\log N}$. Thus we have

$$\log\left(d(\Phi,\tilde{\Psi})\right) = -q\log N \approx -\frac{\log N}{4} + \log K$$

or

$$d(\Phi,\tilde{\Psi}) \approx KN^{-1/4}. \tag{72}$$

Let us now turn to a discussion of our results.

5 Discussion

We have shown that the minimisation principle underlying K-SVD (1) can identify a tight frame with arbitrary precision from signals generated from a wide class of decaying coefficients distributions, provided that the training sample size is large enough. For the case $S = 1$ in particular this means that K-SVD in combination with a greedy algorithm can recover the generating dictionary up to prescribed precision. To illustrate our results we conducted two experiments.

5.1 Experiments

The first experiment demonstrates that the requirement on the dictionary to be tight in order to be identifiable translates to the case of finitely many training samples. For simplicity and to allow for a visual representation of the outcome it was conducted in $\mathbb{R}^2$. We generated 1000 coefficients by drawing c_2 uniformly at random from the interval $[0, 0.6]$, setting $c_1 = \sqrt{1 - c_2^2}$, randomly permuting the resulting vector and providing it with random $\pm$ signs. We then generated four sets of signals, using four bases with increasing coherence and the same coefficients, and for each set of signals found the minimiser of the K-SVD criterion (1) with $S = 1$. Figure 1 shows the objective function for the case of an orthonormal basis, while Figure 2 shows the four signal sets, the generating bases and the recovered bases. As predicted by our theoretical results when the generating basis is orthogonal it is also the minimiser of the K-SVD criterion, while for an oblique generating basis the minimiser is distorted towards the maximal eigenvector of the basis. Since for a 2-dimensional basis in combination with our coefficient distribution the abstract condition in (35) is always fulfilled, this effect can only be due to the violation of the tightness-condition.

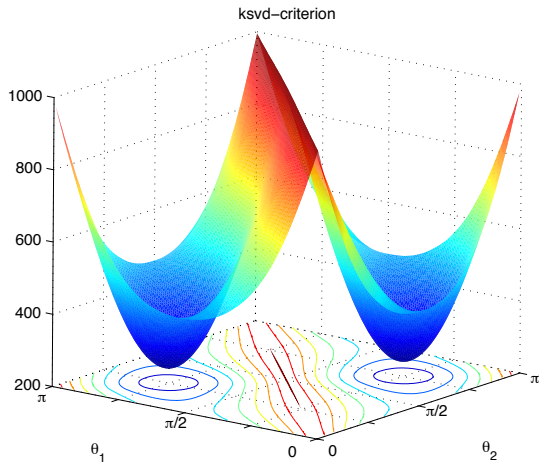


Fig. 1. The K-SVD-criterion for the signals created from the decaying coefficients and an orthonormal basis, the admissible dictionaries are parametrised by two angles (θ_1, θ_2), i.e. $\phi_i = (\cos\theta_i, \sin\theta_i)$.

The second experiment illustrates how the local minimum near the generating dictionary

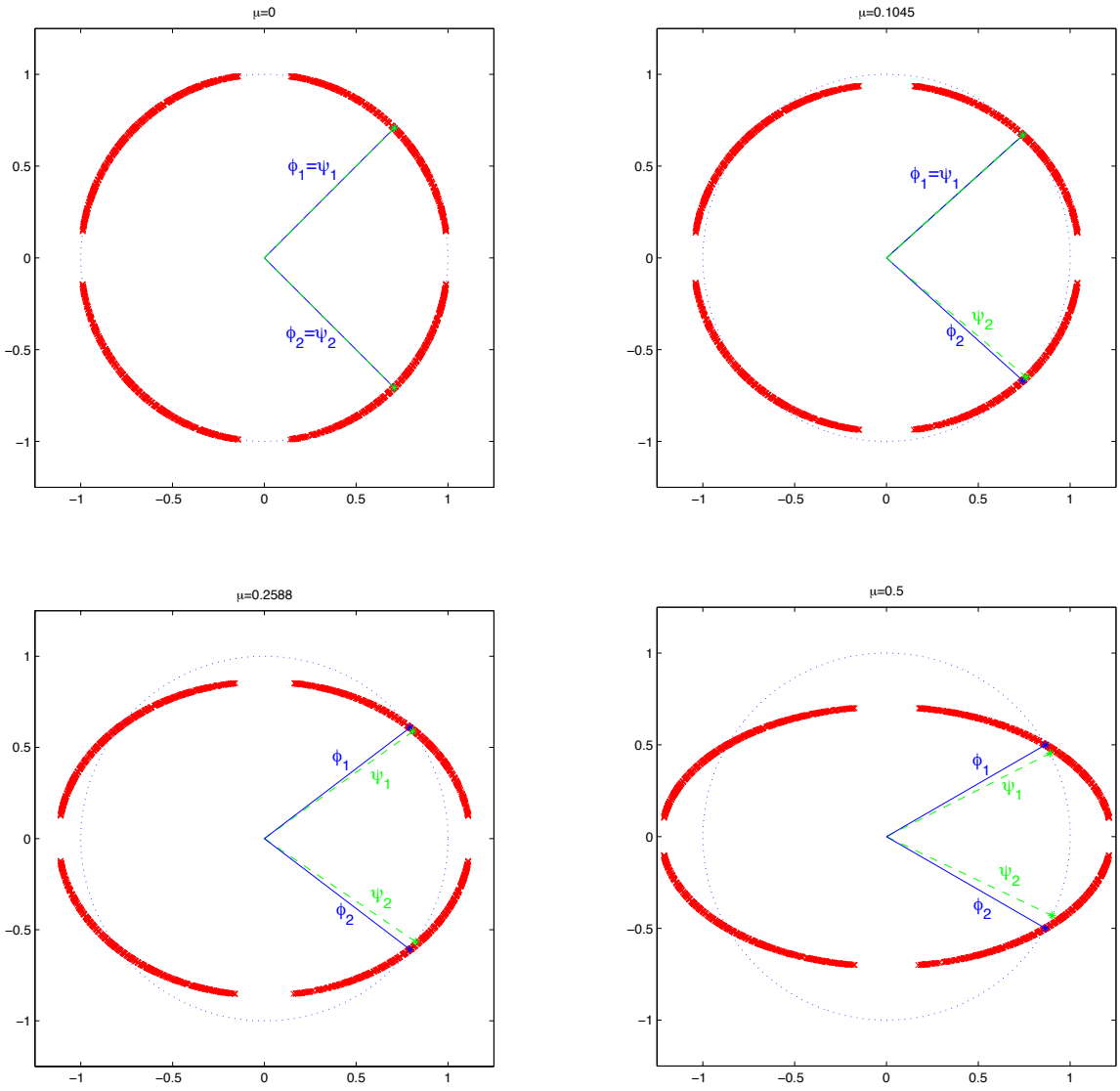
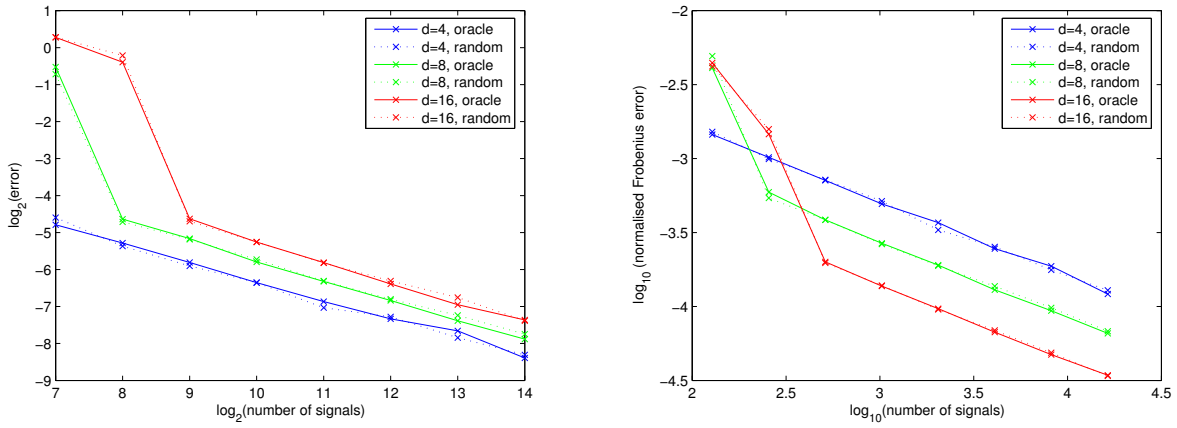


Fig. 2. Signals created from various bases $\Phi = (\phi_1, \phi_2)$ with increasing coherence μ, together with the corresponding minimiser $\Psi = (\psi_1, \psi_2)$ of the K-SVD-criterion for $S = 1$.

approaches the generating dictionary as the number of signals increases. As generating dictionary we choose the union of two orthonormal bases, the Hadamard and the Dirac basis, in dimension $d = 4, 8, 16$, i.e. $K = 2d$. We then generated 2-sparse signals by first drawing c_1 uniformly at random from the interval $[0.99, 1]$, setting $c_2 = \sqrt{1 - c_1^2}$, meaning $c_2 \in [0, 0.1]$, and $c_i = 0$ for $i \geq 3$ and then setting $y = \Phi c_{\sigma,p}$ for a uniformly at random chosen sign sequence σ and permutation p. We then run the original K-SVD algorithm as described in [3], with a

greedy algorithm, and sparsity parameter $S = 1$, using both an oracle initialisation (i.e. the generating dictionary) and a random initialisation, on training sets containing $128 \cdot 2^n$ signals for n increasing from 0 to 7. Figure 3 (a) plots the maximal distance between two corresponding atoms of the generating and the learned dictionary, $d(\Phi, \tilde{\Psi}) = \max_i \|\phi_i - \psi_i\|_2$, averaged over 10 runs. Figure 3 (b) is designed to be comparable to the experiment conducted for the noisy ℓ_1-criterion in [22] and plots the normalised Frobenius norm between the generating and the learned dictionary, $\|\Phi - \tilde{\Psi}\|_F / \sqrt{dK^3}$, averaged over 10 runs.



(a) (b)

Fig. 3. Error between the generating Hadamard-Dirac dictionary Φ in $\mathbb{R}^d$ and the output $\tilde{\Psi}$ of the K-SVD algorithm with parameter $S = 1$; the error is measured as $d(\Phi, \tilde{\Psi}) = \max_i \|\phi_i - \psi_i\|_2$) in (a) and as $\|\Phi - \tilde{\Psi}\|_F / \sqrt{dK^3}$ in (b).

As expected we have a log-linear relation between the number of samples and the reconstruction error. However our predictions seem to be too pessimistic. So rather than an inclination of $-\frac{1}{4}$ we see one of $-\frac{1}{2}$ indicating that $d(\Phi, \tilde{\Psi}) \approx N^{-\frac{1}{2}}$. We also see that both the oracle and the random initialisation lead to the same results, raising the question of uniqueness of the equivalent local minima, compare also [22].

5.2 Future work

Finally let us point out further research directions based on a comparison of our results for the K-SVD-minimisation principle to existing identification results. Compared to the available

identification results for the ℓ_1-minimisation principle,

$$\min_{\Phi\in\mathcal{D},X:Y=\Phi X}\sum_{ij}|X_{ij}|, \tag{73}$$

it seems at first glance that the K-SVD-criterion requires a larger sample size than the ℓ_1-criterion, i.e. $N^{1-4q}/\log N = O(K^3 d)$ as opposed to $O(d^2\log d)$ reported in [21] for a basis and $O(K^3)$ reported in [17] for an overcomplete dictionary. Also it does not allow for exact identification with high probability but only guarantees stability. However, this effect may be due to the more general signal model which assumes decay rather than exact sparsity. Indeed it is very interesting to compare our results to a recent result for a noisy version of the ℓ_1-minimisation principle, [22], which provides stability results under unbounded white noise and, omitting log factors, also derives a sampling complexity of $O(K^3 d)$.

Another difference, apparently intrinsic to the two minimisation criteria is that probably the K-SVD criterion can only identify tight dictionary frames exactly, while the ℓ_1-criterion allows identification of arbitrary dictionaries. Thus to support the use of the K-SVD criterion for the learning of non-tight dictionaries also theoretically, we plan to study the stability of the K-SVD criterion under non-tightness by analysing the maximal distance between an original, non tight dictionary with condition number $\sqrt{B/A} > 1$ and the closest local maximum, cp. also Figure 2. Compared to identification results for the ER-SpUD algorithm, [37], our results have the advantage of being valid also for overcomplete dictionaries and not exactly sparse signals. The disadvantage is that our results are valid only locally and in case $S > 1$ only for a criterion, not an algorithm. An important research direction therefore is to analyse how close the output of K-SVD is to the local minimum of the K-SVD criterion given the same initialisation in the general case.

The last research direction we want to point out is how much decay of the coefficients is actually necessary. For the case $S = 1$, it is quite easy to see, compare also [33], [34], that a condition of the type $c_1 > c_2 + 2\mu\|c\|_1$ ensures that the maximal inner product is always attained at $i_p = p^{-1}(1)$. However, typically we have $|\langle\phi_i, \Phi c_{p,\sigma}\rangle| \approx c_{p(i)} \pm \mu$. Therefore a condition such as $c_1 > c_2 + O(\mu)$, which allows for outliers, i.e. signals for which the maximal projection is not at i_p, might be sufficient to prove - if not exact identifiability - at least stability. Together with the inspiring techniques from [22], we expect the tools developed in the course of such an analysis to allow us also to deal with unbounded white noise.

ACKNOWLEDGMENTS

This work was supported by the Austrian Science Fund (FWF) under Grant no. Y432 and J3335 and improved thanks to the reviewers' detailed and pointed comments.

Also I would like to thank Massimo Fornasier for SUPPORT (in capital letters), Maria Mateescu for proof-reading the proposal leading to grant J3335 and helping me with the shoeshine, Remi Gribonval for pointing out the connection between an early version of (4) with $S = 1$ and K-SVD and Jan Vybiral for reading several ugly draft versions.

APPENDIX A

TECHNICAL DETAILS FOR THE PROOF OF THEOREM 3.1

Lemma A.1. *For two frames Φ, Ψ we have*

$$\begin{aligned}&\mathbb{E}_p\mathbb{E}_\sigma\left(\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2\right)\\&\quad=\binom{K}{S}^{-1}\left(\frac{1-\gamma_S^2}{K-S}\sum_I\|P_I(\Psi)\Phi\|_F^2+\left(\frac{\gamma_S^2}{S}-\frac{1-\gamma_S^2}{K-S}\right)\sum_I\|P_I(\Psi)\Phi_I\|_F^2\right),\end{aligned}\tag{74}$$

where $\gamma_S^2 := c_1^2 + \ldots + c_S^2$.

In case Φ is a tight frame with frame constant A and $\delta_S(\Psi) < 1$ this reduces to

$$\mathbb{E}_p\mathbb{E}_\sigma\left(\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2\right)=\frac{A(1-\gamma_S^2)S}{K-S}+\left(\frac{\gamma_S^2}{S}-\frac{1-\gamma_S^2}{K-S}\right)\binom{K}{S}^{-1}\sum_I\|P_I(\Psi)\Phi_I\|_F^2.\tag{75}$$

Proof: We have

$$\mathbb{E}_p\mathbb{E}_\sigma\left(\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2\right)=\sum_i\mathbb{E}_p\left(c_{p(i)}^2\|P_{I_p}(\Psi)\phi_i\|_2^2\right)\tag{76}$$

For each i we now split the set of all permutations $\mathcal{P}$ into disjoint sets $\mathcal{P}_{Ik}^i$, defined as

$$\mathcal{P}_{Ik}^i := \{p : p(I) = \{1,\ldots,S\}, p(i) = k\},$$

where I is a subset of $\{1,\ldots,K\}$ with $|I| = S$ and $k = 1\ldots K$. We then have $\mathcal{P} = \cup_{I,k}\mathcal{P}_{jk}^i$ and

$$|\mathcal{P}_{Ik}^i| = \begin{cases}(K-S-1)!S! & \text{if } i\notin I \text{ and } k\geq S+1\\(K-S)!(S-1)! & \text{if } i=j\in I \text{ and } k=p(j)\\0 & \text{else}\end{cases}.$$

Using these sets we can compute the expectations in (76) as follows

$$\begin{aligned}
\mathbb{E}_p\left(c_{p(i)}^2\|P_{I_p}(\Psi)\phi_i\|_2^2\right) &= \frac{1}{K!}\sum_I\sum_k\sum_{p\in\mathcal{P}_{Ik}^i} c_k^2\|P_I(\Psi)\phi_i\|_2^2 \\
&= \binom{K}{S}^{-1}\frac{1}{K-S}\sum_{I:i\notin I}\sum_{k\geq S+1} c_k^2\|P_I(\Psi)\phi_i\|_2^2 + \binom{K}{S}^{-1}\frac{1}{S}\sum_{I:i\in I}\sum_{k\leq S} c_k^2\|P_I(\Psi)\phi_i\|_2^2 \\
&= \binom{K}{S}^{-1}\left(\frac{1-c_1^2-\ldots-c_S^2}{K-S}\sum_{I:i\notin I}\|P_I(\Psi)\phi_i\|_2^2 + \frac{c_1^2+\ldots+c_S^2}{S}\sum_{I:i\in I}\|P_I(\Psi)\phi_i\|_2^2\right).
\end{aligned}$$

Abbreviating $\gamma_S^2 := c_1^2+\ldots+c_S^2$ and re-substituting the above expression into (76) leads to,

$$\begin{aligned}
\binom{K}{S}\mathbb{E}_p\mathbb{E}_\sigma\left(\|P_{I_p}(\Psi)\Phi c_{p,\sigma}\|_2^2\right) &= \frac{1-\gamma_S^2}{K-S}\sum_i\sum_{I:i\notin I}\|P_I(\Psi)\phi_i\|_2^2 + \frac{\gamma_S^2}{S}\sum_i\sum_{I:i\in I}\|P_I(\Psi)\phi_i\|_2^2 \\
&= \frac{1-\gamma_S^2}{K-S}\sum_i\sum_I\|P_I(\Psi)\phi_i\|_2^2 + \left(\frac{\gamma_S^2}{S}-\frac{1-\gamma_S^2}{K-S}\right)\sum_i\sum_{I:i\in I}\|P_I(\Psi)\phi_i\|_2^2 \\
&= \frac{1-\gamma_S^2}{K-S}\sum_I\sum_i\|P_I(\Psi)\phi_i\|_2^2 + \left(\frac{\gamma_S^2}{S}-\frac{1-\gamma_S^2}{K-S}\right)\sum_I\sum_{i\in I}\|P_I(\Psi)\phi_i\|_2^2 \\
&= \frac{1-\gamma_S^2}{K-S}\sum_I\|P_I(\Psi)\Phi\|_F^2 + \left(\frac{\gamma_S^2}{S}-\frac{1-\gamma_S^2}{K-S}\right)\sum_I\|P_I(\Psi)\Phi_I\|_F^2.
\end{aligned}$$

If Φ is a tight frame and $\delta_S(\Psi)<1$, meaning Ψ_I always has full rank, we have $\|P_I(\Psi)\Phi\|_F^2 = \mathrm{tr}(\Phi^\star P_I(\Psi)^\star P_I(\Psi)\Phi) = \mathrm{tr}(P_I(\Psi)\Phi\Phi^\star) = AS$, which leads to the second statement. □

Lemma A.2. *Let Φ be a dictionary with isometry constant $\delta_S<1$ and Ψ be an ε perturbation of Φ, i.e. $d(\Phi,\Psi)=\varepsilon$. So we can write $\Psi=\Phi A+ZW$, where $A=\mathrm{diag}((1-\varepsilon_i^2/2)_i)$, $W=\mathrm{diag}((\varepsilon_i^2-\varepsilon_i^4/4)^{1/2})_i)$ for $\max_i\varepsilon_i=\varepsilon$ and $Z=(z_1,\ldots z_K)$ where $\langle z_i,\phi_i\rangle=0$. Abbreviating $Q_I(\Phi)=\mathbb{I}_d-P_I(\Phi)$, where $\mathbb{I}_d$ is the identity matrix in $\mathbb{R}^{d\times d}$, and $B_I=Z_IW_IA_I^{-1}$, where $A_I=\mathrm{diag}((1-\varepsilon_i^2/2)_{i\in I})$ and $W_I=\mathrm{diag}((\varepsilon_i^2-\varepsilon_i^4/4)^{1/2})_{i\in I})$, we have*

$$\|P_I(\Phi)-P_I(\Psi)\|_F^2 \leq \frac{2\|Q_I(\Phi)B_I\|_F^2}{\sqrt{1-\delta_S}\left(\sqrt{1-\delta_S}-2\|B_I\|_F\right)},$$

and

$$\|P_I(\Psi)\Phi_I\|_F^2 \leq \|\Phi_I\|_F^2 - \|Q_I(\Phi)B_I\|_F^2 + \frac{2\|Q_I(\Phi)B_I\|_F^2\|B_I\|_F}{\sqrt{1-\delta_S}-\|B_I\|_F} + \frac{\|Q_I(\Phi)B_I\|_F^4}{\left(\sqrt{1-\delta_S}-2\|B_I\|_F\right)^2},$$

whenever ε is small enough.

In particular when $\varepsilon\leq\frac{\sqrt{1-\delta_S}}{21\sqrt{S}}$ we have

$$\|P_I(\Phi)-P_I(\Psi)\|_F \leq 1.487\cdot\frac{\|Q_I(\Phi)B_I\|_F}{\sqrt{1-\delta_S}} \qquad \textit{and} \qquad \|P_I(\Psi)\Phi_I\|_F^2 \leq \|\Phi_I\|_F^2 - 0.897\cdot\|Q_I(\Phi)B_I\|_F^2.$$

Proof: We first compute the projection $P_I(\Psi) = \Psi_I(\Psi_I^\star\Psi_I)^{-1}\Psi_I^\star$ in terms of Φ_I and B_I. Since $\delta_S < 1$ the matrix $\Phi_I^\star\Phi_I$ is invertible and we can write $\Phi_I^\dagger = (\Phi_I^\star\Phi_I)^{-1}\Phi_I^\star$. We now split Ψ_I into the part contained in the span of Φ_I and the rest,

$$\begin{aligned}\Psi_I &= P_I(\Phi)\Psi_I + Q_I(\Phi)\Psi_I \\ &= \Phi_I A_I + P_I(\Phi)Z_I W_I + Q_I(\Phi)Z_I W_I \\ &= \left(\Phi_I(\mathbb{I}_S + \Phi_I^\dagger B_I) + Q_I(\Phi)B_I\right) A_I. \end{aligned} \tag{77}$$

Next we calculate $(\Psi_I^\star\Psi_I)^{-1}$. Using the expression in (77) we have

$$\Psi_I^\star\Psi_I = A_I\left((\mathbb{I}_S + \Phi_I^\dagger B_I)^\star\Phi_I^\star\Phi_I(\mathbb{I}_S + \Phi_I^\dagger B_I) + B_I^\star Q_I(\Phi)B_I\right)A_I.$$

Using the fact that $\|\Phi_I^\dagger\|_{2,2}^2 = \|(\Phi_I^\star\Phi_I)^{-1}\|_{2,2} \leq (1-\delta_S)^{-1}$ we can estimate

$$\|\Phi_I^\dagger B_I\|_{2,2} \leq \|\Phi_I^\dagger\|_{2,2}\|B_I\|_F \leq \sqrt{1-\delta_S}\frac{\varepsilon\sqrt{S}}{\sqrt{1-\varepsilon^2}}. \tag{78}$$

Since this is smaller than 1 for ε small enough, we can calculate the inverse of $(\mathbb{I}_S + \Phi_I^\dagger B_I)$ using a Neumann series, i.e.

$$(\mathbb{I}_S + \Phi_I^\dagger B_I)^{-1} = \mathbb{I}_S + \sum_{i=1}^{\infty}(-\Phi_I^\dagger B_I)^i,$$

with $\|(\mathbb{I}_S + \Phi_I^\dagger B_I)^{-1}\|_{2,2} \leq (1 - \|\Phi_I^\dagger B_I\|_{2,2})^{-1}$. This allows us to rewrite $\Psi_I^\star\Psi_I$ as,

$$\begin{aligned}\Psi_I^\star\Psi_I = A_I(\mathbb{I}_S + \Phi_I^\dagger B_I)^\star\Phi_I^\star\Phi_I\left(\mathbb{I}_S + R_I\right)(\mathbb{I}_S + \Phi_I^\dagger B_I)A_I, \\ \text{for } R_I = (\Phi_I^\star\Phi_I)^{-1}(\mathbb{I}_S + \Phi_I^\dagger B_I)^{\star-1}B_I^\star Q_I(\Phi)B_I(\mathbb{I}_S + \Phi_I^\dagger B_I)^{-1},\end{aligned} \tag{79}$$

and we can estimate

$$\begin{aligned}\|R_I\|_{2,2} &\leq \|(\Phi_I^\star\Phi_I)^{-1}\|_{2,2}\|(\mathbb{I}_S + \Phi_I^\dagger B_I)^{-1}\|_{2,2}^2\|Q_I(\Phi)B_I\|_{2,2}^2 \\ &\leq \frac{\|Q_I(\Phi)B_I\|_F^2}{\left(\|\Phi_I^\dagger\|_{2,2}^{-1} - \|B_I\|_F\right)^2} \\ &\leq \frac{\frac{S\varepsilon^2}{1-\varepsilon^2}}{1-\delta_S - 2\frac{S\varepsilon^2}{1-\varepsilon^2}}.\end{aligned} \tag{80}$$

For ε small enough this is again smaller than 1 and so we can again use a Neumann series to calculate the inverse,

$$(\Psi_I^\star\Psi_I)^{-1} = A_I^{-1}(\mathbb{I}_S + \Phi_I^\dagger B_I)^{-1}\left(\mathbb{I}_S + \sum_{i=1}^{\infty}(-R_I)^i\right)(\Phi_I^\star\Phi_I)^{-1}(\mathbb{I}_S + \Phi_I^\dagger B_I)^{-1\star}A_I^{-1}.$$

Thus we finally get for the projection onto the perturbed atoms indexed by I,

$$P_I(\Psi) = \left(\Phi_I + Q_I(\Phi)B_I(\mathbb{I}_S + \Phi_I^\dagger B_I)^{-1}\right) \cdot \left(\mathbb{I}_S + \sum_{i=1}^{\infty}(-R_I)^i\right)(\Phi_I^\star\Phi_I)^{-1}\left(\Phi_I + Q_I(\Phi)B_I(\mathbb{I}_S + \Phi_I^\dagger B_I)^{-1}\right)^\star. \quad (81)$$

To calculate $\|P_I(\Phi) - P_I(\Psi)\|_F^2$ up to order $O(\varepsilon^2)$ we need to keep track of all terms involving B_I up to second order. We have,

$$\begin{aligned}\|P_I(\Phi) - P_I(\Psi)\|_F^2 &= \operatorname{tr}(P_I(\Phi)) - \operatorname{tr}(P_I(\Phi)P_I(\Psi)) + \operatorname{tr}(P_I(\Psi)) \\ &= 2S - 2\operatorname{tr}((\Phi_I^\star\Phi_I)^{-1}\Phi_I^\star\Psi_I(\Psi_I^\star\Psi_I)^{-1}\Psi_I^\star\Phi_I) \\ &= 2S - 2\operatorname{tr}\left(\mathbb{I}_S + \sum_{i=1}^{\infty}(-R_I)^i\right) \le 2\sum_{i=1}^{\infty}\|R_I\|_F^i, \end{aligned} \quad (82)$$

and employing the bound for $\|R_I\|_F$ from (80) leads us to,

$$\begin{aligned}\|P_I(\Phi) - P_I(\Psi)\|_F^2 &\le \frac{2\|Q_I(\Phi)B_I\|_F^2}{\left(\|\Phi_I^\dagger\|_{2,2}^{-1} - \|B_I\|_F\right)^2 - \|Q_I(\Phi)B_I\|_F^2} \\ &\le \frac{2\|Q_I(\Phi)B_I\|_F^2}{\|\Phi_I^\dagger\|_{2,2}^{-1}\left(\|\Phi_I^\dagger\|_{2,2}^{-1} - 2\|B_I\|_F\right)} \le \frac{2\|Q_I(\Phi)B_I\|_F^2}{\sqrt{1-\delta_S}\left(\sqrt{1-\delta_S} - 2\|B_I\|_F\right)}. \end{aligned} \quad (83)$$

Similarily we get for $\|P_I(\Psi)\Phi_I\|_F^2$,

$$\begin{aligned}\|P_I(\Psi)\Phi_I\|_F^2 &= \operatorname{tr}(\Phi_I^\star\Psi_I(\Psi_I^\star\Psi_I)^{-1}\Psi_I^\star\Phi_I) \\ &= \operatorname{tr}\left(\Phi_I^\star\Phi_I\left(\mathbb{I}_S + \sum_{i=1}^{\infty}(-R_I)^i\right)\right) \\ &= \operatorname{tr}(\Phi_I^\star\Phi_I) - \operatorname{tr}\left(\left(\mathbb{I}_S + \sum_{i=1}^{\infty}(-\Phi_I^\dagger B_I)^i\right)^\star B_I^\star Q_I(\Phi)B_I\left(\mathbb{I}_S + \sum_{i=1}^{\infty}(-\Phi_I^\dagger B_I)^i\right)\right) \\ &\qquad + \operatorname{tr}\left(\Phi_I^\star\Phi_I\sum_{i=2}^{\infty}(-R_I)^i\right) \\ &= \operatorname{tr}(\Phi_I^\star\Phi_I) - \operatorname{tr}(B_I^\star Q_I(\Phi)B_I) - 2\operatorname{tr}\left(B_I^\star Q_I(\Phi)B_I\sum_{i=1}^{\infty}(-\Phi_I^\dagger B_I)^i\right) \\ &\qquad - \operatorname{tr}\left(\left(\sum_{i=1}^{\infty}(-\Phi_I^\dagger B_I)^i\right)^\star B_I^\star Q_I(\Phi)B_I\sum_{i=1}^{\infty}(-\Phi_I^\dagger B_I)^i\right) + \operatorname{tr}\left(\Phi_I^\star\Phi_I\sum_{i=2}^{\infty}(-R_I)^i\right). \end{aligned}$$

Taking into account that the fourth term in the above equation is always smaller than zero we

finally get the bound,

$$\begin{aligned}\|P_I(\Psi)\Phi_I\|_F^2 &\leq \|\Phi_I\|_F^2 - \|Q_I(\Phi)B_I\|_F^2 + 2\|Q_I(\Phi)B_I\|_F^2 \sum_{i=1}^{\infty} \|\Phi_I^\dagger B_I\|_F^i + \|\Phi_I^\star \Phi_I R_I\|_F \sum_{i=1}^{\infty} \|R_I\|_F^i \\ &\leq \|\Phi_I\|_F^2 - \|Q_I(\Phi)B_I\|_F^2 + \frac{2\|Q_I(\Phi)B_I\|_F^2\|B_I\|_F}{\|\Phi_I^\dagger\|_{2,2}^{-1} - \|B_I\|_F} + \frac{\|Q_I(\Phi)B_I\|_F^4}{\left(\|\Phi_I^\dagger\|_{2,2}^{-1} - 2\|B_I\|_F\right)^2} \\ &\leq \|\Phi_I\|_F^2 - \|Q_I(\Phi)B_I\|_F^2 + \frac{2\|Q_I(\Phi)B_I\|_F^2\|B_I\|_F}{\sqrt{1-\delta_S} - \|B_I\|_F} + \frac{\|Q_I(\Phi)B_I\|_F^4}{\left(\sqrt{1-\delta_S} - 2\|B_I\|_F\right)^2}. \end{aligned} \tag{84}$$

□

Lemma A.3. *For $a, b, \xi > 0$,*

$$\xi \leq \frac{4b}{1+\sqrt{1+16\log(\frac{a}{b})}} \qquad \textit{implies that} \qquad a\exp\left(\frac{-b^2}{\xi^2}\right) < \xi. \tag{85}$$

Proof: We have

$$a\exp\left(\frac{-b^2}{\xi^2}\right) < \xi \quad \Leftrightarrow \quad \frac{a}{b}\exp\left(-\frac{b^2}{\xi^2}\right) < \left(\frac{b}{\xi}\right)^{-1} \quad \Leftrightarrow \quad \log\left(\frac{a}{b}\right) - \frac{b^2}{\xi^2} < -\log\left(\frac{b}{\xi}\right).$$

Since $\log x < x/2$ for $x \geq 0$ the last inequality is implied by

$$\frac{b^2}{\xi^2} - \frac{b}{2\xi} \geq \log\left(\frac{a}{b}\right),$$

which is satisfied as soon as

$$\frac{b}{\xi} \geq \frac{1}{4}\left(1+\sqrt{1+16\log\left(\frac{a}{b}\right)}\right) \qquad \Leftrightarrow \qquad \xi \leq \frac{4b}{1+\sqrt{1+16\log(\frac{a}{b})}}.$$

□

REFERENCES

[1] A. Agarwal, A. Anandkumar, P. Jain, P. Netrapalli, and R. Tandon. Learning sparsely used overcomplete dictionaries via alternating minimization. *arXiv:1310.7991*, 2013.

[2] A. Agarwal, A. Anandkumar, and P. Netrapalli. Exact recovery of sparsely used overcomplete dictionaries. *arXiv:1309.1952*, 2013.

[3] M. Aharon, M. Elad, and A.M. Bruckstein. K-SVD: An algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Transactions on Signal Processing.*, 54(11):4311–4322, November 2006.

[4] M. Aharon, M. Elad, and A.M. Bruckstein. On the uniqueness of overcomplete dictionaries, and a practical way to retrieve them. *Journal of Linear Algebra and Applications*, 416:48–67, July 2006.

[5] S. Arora, R. Ge, and A. Moitra. New algorithms for learning incoherent and overcomplete dictionaries. *arXiv:1308.6273*, 2013.

[6] T. Blumensath and M.E. Davies. Iterative thresholding for sparse approximations. *Journal of Fourier Analysis and Applications*, 14(5-6):629–654, 2008.

[7] E. Candès, L. Demanet, D.L. Donoho, and L. Ying. Fast discrete curvelet transforms. *Multiscale Modeling & Simulation*, 5(3):861–899, 2006.

[8] E. Candès, J. Romberg, and T. Tao. Robust uncertainty principles: exact signal reconstruction from highly incomplete frequency information. *IEEE Transactions on Information Theory*, 52(2):489–509, 2006.

[9] S.S. Chen, D.L. Donoho, and M.A. Saunders. Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 1998.

[10] O. Christensen. *An Introduction to Frames and Riesz Bases*. Birkhäuser, 2003.

[11] I. Daubechies. *Ten Lectures on Wavelets*. CBMS-NSF Lecture Notes. SIAM, 1992.

[12] I. Daubechies, R.A. DeVore, M. Fornasier, and S. Güntürk. Iteratively reweighted least squares minimization for sparse recovery. *Communications on Pure and Applied Mathematics*, 63(1):1–38, January 2010.

[13] G. Davis, S. Mallat, and M. Avellaneda. Adaptive greedy approximations. *Constructive Approximation*, 13:57–98, 1997. Springer-Verlag New York Inc.

[14] D.L. Donoho. Compressed sensing. *IEEE Transactions on Information Theory*, 52(4):1289–1306, 2006.

[15] D.L. Donoho, M. Elad, and V.N. Temlyakov. Stable recovery of sparse overcomplete representations in the presence of noise. *IEEE Transactions on Information Theory*, 52(1):6–18, January 2006.

[16] D.J. Field and B.A. Olshausen. Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, 381:607–609, 1996.

[17] Q. Geng, H. Wang, and J. Wright. On the local correctness of ℓ^1-minimization for dictionary learning. *arXiv:1101.5672*, 2011.

[18] P. Georgiev, F.J. Theis, and A. Cichocki. Sparse component analysis and blind source separation of underdetermined mixtures. *IEEE Transactions on Neural Networks*, 16(4):992–996, 2005.

[19] R. Gribonval, R. Jenatton, F. Bach, M. Kleinsteuber, and M. Seibert. Sample complexity of dictionary learning and other matrix factorizations. *arXiv:1312.3790*, 2013.

[20] R. Gribonval, H. Rauhut, K. Schnass, and P. Vandergheynst. Atoms of all channels, unite! Average case analysis of multi-channel sparse recovery using greedy algorithms. *Journal of Fourier Analysis and Applications*, 14(5):655–687, 2008.

[21] R. Gribonval and K. Schnass. Dictionary identifiability - sparse matrix-factorisation via l_1-minimisation. *IEEE Transactions on Information Theory*, 56(7):3523–3539, July 2010.

[22] R. Jenatton, F. Bach, and R. Gribonval. Local stability and robustness of sparse dictionary learning in the presence of noise. *preprint*, 2012.

[23] K. Kreutz-Delgado, J.F. Murray, B.D. Rao, K. Engan, T. Lee, and T.J. Sejnowski. Dictionary learning algorithms for sparse representation. *Neural Computations*, 15(2):349–396, 2003.

[24] K. Kreutz-Delgado and B.D. Rao. FOCUSS-based dictionary learning algorithms. In *SPIE 4119*, 2000.

[25] M. Ledoux and M. Talagrand. *Probability in Banach spaces. Isoperimetry and processes.* Springer-Verlag, Berlin, Heidelberg, NewYork, 1991.

[26] J. Mairal, F. Bach, J. Ponce, and G. Sapiro. Online learning for matrix factorization and sparse coding. *Journal of Machine Learning Research*, 11:19–60, 2010.

[27] A. Maurer and M. Pontil. K-dimensional coding schemes in Hilbert spaces. *IEEE Transactions on Information Theory*, 56(11):5839–5846, 2010.

[28] N.A. Mehta and A.G. Gray. On the sample complexity of predictive sparse coding. *arXiv:1202.4050*, 2012.

[29] D. Needell and J.A. Tropp. CoSaMP: Iterative signal recovery from incomplete and inaccurate samples. *Applied Computational Harmonic Analysis*, 26(3):301–321, 2009.

[30] M.D. Plumbley. Dictionary learning for ℓ_1-exact sparse coding. In M.E. Davies, C.J. James, and S.A. Abdallah, editors, *International Conference on Independent Component Analysis and Signal Separation*, volume 4666, pages 406–413. Springer, 2007.

[31] DSP Rice University. Compressive sensing resources. *http://www.compressedsensing.com/*.

[32] R. Rubinstein, A. Bruckstein, and M. Elad. Dictionaries for sparse representation modeling. *Proceedings of the IEEE*, 98(6):1045–1057, 2010.

[33] K. Schnass. Dictionary identification results for K-SVD with sparsity parameter 1. In *SampTA13*, 2013.

[34] K. Schnass. On the identifiability of overcomplete dictionaries via the minimisation principle underlying K-SVD. *arXiv:1301.3375*, 2013.

[35] K. Schnass and P. Vandergheynst. Average performance analysis for thresholding. *IEEE Signal Processing Letters*, 14(11):828–831, 2007.

[36] K. Skretting and K. Engan. Recursive least squares dictionary learning algorithm. *IEEE Transactions on Signal Processing*, 58(4):2121–2130, April 2010.

[37] D. Spielman, H. Wang, and J. Wright. Exact recovery of sparsely-used dictionaries. In *Conference on Learning Theory (arXiv:1206.5882)*, 2012.

[38] J.A. Tropp. Greed is good: Algorithmic results for sparse approximation. *IEEE Transactions on Information Theory*, 50(10):2231–2242, October 2004.

[39] J.A. Tropp. On the conditioning of random subdictionaries. *Applied Computational Harmonic Analysis*, 25(1-24), 2008.

[40] D. Vainsencher, S. Mannor, and A.M. Bruckstein. The sample complexity of dictionary learning. *Journal of Machine Learning Research*, 12(3259-3281), 2011.

[41] R. Vershynin. Introduction to the non-asymptotic analysis of random matrices. In Y. Eldar and G. Kutyniok, editors, *Compressed Sensing, Theory and Applications*, chapter 5. Cambridge University Press, 2012.

[42] 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, June 2009.

[43] M. Zibulevsky and B.A. Pearlmutter. Blind source separation by sparse decomposition in a signal dictionary. *Neural Computations*, 13(4):863–882, 2001.