Available at
www.ElsevierComputerScience.com
POWERED BY SCIENCE DIRECT
Signal Processing 86 (2006) 511–532
SIGNAL PROCESSING
www.elsevier.com/locate/sigpro

On the stability of the basis pursuit in the presence of noise

David L. Donoho[a], Michael Elad[b,*]

[a]Department of Statistics, Stanford University, Stanford 94306 CA, USA
[b]Department of Computer Science, The Technion-Israel Institute of Technology, Haifa 32000, Israel

Received 27 October 2004; received in revised form 19 April 2005; accepted 30 May 2005
Available online 3 August 2005

Abstract

Given a signal $\mathbf{S} \in \mathscr{R}^N$ and a full-rank matrix $\mathbf{D} \in \mathscr{R}^{\mathbf{N} \times \mathbf{L}}$ with $N < L$, we define the signal's over-complete representation as $\alpha \in \mathscr{R}^L$ satisfying $\mathbf{S} = \mathbf{D}\alpha$. Among the infinitely many solutions of this under-determined linear system of equations, we have special interest in the sparsest representation, i.e., the one minimizing $\|\alpha\|_0$. This problem has a combinatorial flavor to it, and its direct solution is impossible even for moderate L. Approximation algorithms are thus required, and one such appealing technique is the basis pursuit (BP) algorithm. This algorithm has been the focus of recent theoretical research effort. It was found that if indeed the representation is sparse enough, BP finds it accurately.

When an error is permitted in the composition of the signal, we no longer require exact equality $\mathbf{S} = \mathbf{D}\alpha$. The BP has been extended to treat this case, leading to a denoizing algorithm. The natural question to pose is how the above-mentioned theoretical results generalize to this more practical mode of operation. In this paper we propose such a generalization. The behavior of the basis pursuit in the presence of noise has been the subject of two independent very wide contributions released for publication very recently. This paper is another contribution in this direction, but as opposed to the others mentioned, this paper aims to present a somewhat simplified picture of the topic, and thus could be referred to as a primer to this field. Specifically, we establish here the stability of the BP in the presence of noise for sparse enough representations. We study both the case of a general dictionary $\mathbf{D}$, and a special case where $\mathbf{D}$ is built as a union of orthonormal bases. This work is a direct generalization of noiseless BP study, and indeed, when the noise power is reduced to zero, we obtain the known results of the noiseless BP.

Keywords: Basis pursuit denoizing; Sparse representation; Union of ortho-bases; Bound on sparsity; Stability

*Corresponding author. Tel.: +972 4 8294169; fax: +972 4 8293900.
E-mail addresses: donoho@stat.stanford.edu (D.L. Donoho), elad@cs.technion.ac.il (M. Elad).

0165-1684/$ - see front matter
doi:10.1016/j.sigpro.2005.05.027

1. Introduction

1.1. General– Sparse representations

In signal processing, we are often interested in a replacement of the representation, seeking some simplification for an obvious gain. This is the rational behind the so many transforms proposed over the past several centuries, such as the Fourier, Cosine, Wavelets, and many others. The basic idea is to "change language", and describe the signal differently, in the hope that the new description is better for the application in mind. A natural justification for the use of a transform is that signals already use an imposed representation when described as samples as a function of time/space. Generally, there is no reason to believe that this representation is the most appropriate one for our needs.

The ease with which linear transforms are operated and analyzed keeps those as the first priority candidates in defining alternative representations. It is therefore not surprising to find that linear transforms are the more popular ones in theory and practice in signal processing. A linear transform is defined through the use of a full-rank matrix $\mathbf{D} \in \mathscr{R}^{\mathbf{N}\times\mathbf{L}}$, where $L \geqslant N$. Given the signal $\mathbf{S} \in \mathscr{R}^N$, its representation is defined by

$$\mathbf{S} = \mathbf{D}\alpha, \tag{1}$$

where $\alpha \in \mathscr{R}^L$. For the case of $L = N$ (and a non-singular matrix $\mathbf{D}$ due to the full-rank property), the above relationship implies a linear operation both for the forward transform (from $\mathbf{S}$ to α) and its inverse. Many of the practical transforms are of this type, and many of them go further and simplify the matrix $\mathbf{D}$ to be structured and unitary, so that its inverse is easier to operate and both directions can be computed with nearly $O(N)$ operations. Such is the case with the DFT, DCT, the Hadamard, orthonormal wavelet and other transforms.

In this paper we are interested in the case of $L > N$, referred to as the over-complete transforms. When $L > N$, the relationship in (1) poses an under-determined linear set of equations, and thus in general it leads to an infinite number of possible solutions. Further information is therefore needed in order to uniquely define the transform, and this is typically achieved by using regularization, defining the representation as the solution of

$$(P_p)\ \min_{\alpha} \|\alpha\|_p \quad \text{subject to } \mathbf{S} = \mathbf{D}\alpha. \tag{2}$$

For $p = 2$, it is easy to show that again we obtain linearity in both directions (forward and inverse transforms). This case, typically referred to as "Frame Theory", has drawn a lot of attention because of this obvious simplicity. However, it is clear that linearity poses a hard restriction on the space of possibilities, and may cost in performance.

A different and far more complicated approach advocated strongly in recent years is to consider $p = 0$. The ℓ^0 notation is an abused ℓ^p-norm with $p \to 0$, effectively counting the number of non-zeros in the vector α. In such an approach we seek among all feasible representations (satisfying the constraint in (1)) the one with the fewest non-zero entries, this way achieving an ultimate simplicity in representation. Referring to the matrix $\mathbf{D}$ as a dictionary of signal–prototypes as its columns, we build $\mathbf{S}$ as a linear combination of only few of these columns, typically referred to as atoms. Thus, we can think of our signal as a molecule, and the forward transform decomposes it to its building atoms, where we try to use the fewest in this construction [1].

From the numerical standpoint, the forward transform, defined as (P_0), is a non-convex and highly non-smooth optimization problem, with many possible local minimum points. Prior work has established that this problem is an NP-hard one, implying that its complexity grows exponentially with the number of columns in the dictionary [2,3]. Recent study of this problem and methods to approximate its solution give promising new results, indicating that even though complicated, means exist to solve it at least in some cases using either greedy [4–13] or convex programming approaches [1,10,14–20].

The Basis Pursuit (BP) belongs to the second family of methods, using convexization of the original problem in order to get a numerically traceable algorithm. Instead of the original problem

$$(P_0)\ \min_{\alpha} \|\alpha\|_0 \quad \text{subject to } \mathbf{S} = \mathbf{D}\alpha, \tag{3}$$

the BP proposes the solution of

$$(P_1)\ \min_{\alpha} \|\alpha\|_1 \quad \text{subject to } \mathbf{S} = \mathbf{D}\alpha. \tag{4}$$

This constrained optimization problem has a linear programming (LP) structure, for which there are stable and reliable numerical solvers. In the original work that introduced this option [1], it was observed empirically that this algorithm performs very well, implying that if indeed a sparse solution exists, the solution of (P_1) leads to it. Several very recent works studied theoretically this phenomenon and found that indeed under some conditions (P_1) and (P_0) lead to the same solution. In the following we briefly describe these results.

1.2. Known results on BP

In this section we briefly mention several results related to the performance of the BP. The main reason for mentioning these results is the desire to show later that the noisy case results converge to the noiseless ones mentioned here, when the noise power is set to zero.

One common theme to all the works that analyze the approximating algorithms is their use of the *Mutual Incoherence* as a way to characterize the dictionary $\mathbf{D}$. We start by assuming that the columns of $\mathbf{D}$ are all normalized, i.e. $\|\mathbf{d}_k\|_2 = 1$. The mutual incoherence, denoted as M, is defined as the maximal inner product between the dictionary columns assumed to be normalized,

$$M = \max_{1 \leqslant k,j \leqslant L,\ k \neq j} |\mathbf{d}_k^{\mathrm{T}} \mathbf{d}_j|.$$

For a square and unitary matrix $\mathbf{D}$ the mutual incoherence is zero, indicating a total independence between the dictionary's atoms. For general over-complete dictionaries with $L > N$, M is necessarily non-zero, and we desire the smallest possible value so as to get close to the ideal independence exhibited in the unitary setup. In [21] it has been shown that for full-rank dictionaries of size $N \times L$ we have

$$M \geqslant \sqrt{\frac{L-N}{N(L-1)}},$$

and equality is obtained for a family of dictionaries named *Grassmanian Frames*.

We now turn to describe the known results for the BP, starting with the general dictionary case:

Theorem 1. (*see Donoho and Elad* [18]; *Gribonval and Nielson* [19]; *Fuchs* [20]).*If the solution of* (P_0) *satisfies*

$$\|\hat{\alpha}\|_0 < \frac{1+M}{2M}, \tag{5}$$

then the solution of (P_1) *coincides with it exactly, implying the success of the BP.*

This result suggests that for a class of signals having a sparse enough representation, solving (P_1) is just as solving (P_0), and this way we obtain the solution of an NP-hard problem using an LP solver.

The next result refers to a special case of interest when the dictionary has a specific structure. We assume that $L = JN$, and the dictionary is a union of J orthonormal matrices

$$\mathbf{D} = [\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_J] \quad \text{where } \forall\ 1 \leqslant k \leqslant J,\ \ \mathbf{B}_k^{\mathrm{T}}\mathbf{B}_k = \mathbf{I}.$$

Theorem 2. (*see Elad and Bruckstein* [17]; *Gribonval and Nielsen* [19]) *If the solution of* (P_0) *satisfies*

$$\|\hat{\alpha}\|_0 < \left(\sqrt{2} - 1 + \frac{1}{2(J-1)}\right)\frac{1}{M}, \tag{6}$$

then the solution of (P_1) *coincides with it exactly*, *implying the success of the BP*.

For $J = 2$ it is easy to see that the last result is stronger than the more general one, as expected. It is less obvious (but possible) to see the gain for $J > 2$—we refer the reader to a discussion on this matter in [19].

1.3. Presence of noise and stability

The above two Theorems assume that a signal has been composed by $\mathbf{S}_0 = \mathbf{D}\alpha_0$, where $\|\alpha_0\|_0$ is known to be sufficiently small. Then, the BP is required to recover this original representation successfully based on the knowledge of $\mathbf{S}_0$. We now generalize this scenario and assume that the signal $\mathbf{S}$ is given to us corrupted by additive noise

$$\mathbf{S} = \mathbf{S}_0 + \mathbf{Z} = \mathbf{D}\alpha_0 + \mathbf{Z}, \tag{7}$$

where $|\mathbf{Z}| \leqslant \mathbf{1} \cdot \varepsilon$. This assumption implies that the deviation in each of the entries of the signal are in the range $[-\varepsilon, \varepsilon]$. The notation $\mathbf{1}$ stands for a vector of ones. A different noise description such as $\|\mathbf{Z}\|_2 \leqslant \varepsilon$ could have been posed. However, as we shall see later, the entry-wise description better fits the analysis method adopted in our proof.

Can we still expect the BP to succeed in finding α_0? Before answering this we should address a more fundamental question of how to operate the BP now that we know that the signal is noisy. Solving (4) disregarding the noise is possible, although we can do far better by solving instead:

$$(P_1(\varepsilon))\ \min_{\alpha} \|\alpha\|_1 \quad \text{subject to } |\mathbf{S} - \mathbf{D}\alpha| \leqslant \mathbf{1} \cdot \varepsilon. \tag{8}$$

One may argue that ε is not known, and then we should solve $(P_1(\delta))$ for an arbitrarily chosen δ. We will refer to this option as BP as well in-spite the evident change in formulation.

As we shall see, since the BP is fed with a corrupted signal, we cannot expect such exact recovery. However, we show that the BP result is very close to α_0 if this representation is sparse enough. Such a result implies that BP is globally stable and robust to the noise. Our study essentially leads to a bound on the deviation between α_0 and the BP outcome. We show that this bound is proportional to ε, and thus when $\varepsilon = 0$ we obtain an exact recovery.

1.4. This paper's structure

In the next section we show the main result of this work, building the stability result for the general dictionary case. In Section 3 we address the same question, but this time refer to a specific choice of dictionary, built as an amalgam of several orthonormal matrices. Due to the specific structure imposed the results are tighter. In Section 4 we refer to two very recent works that also studied the noisy BP case. We will briefly mention these works' contributions and relate them to our results. We summarize and conclude in Section 5.

2. New stability result

We start by stating the Theorem that we will later prove in detail.

Theorem 3. *We assume a signal* $\mathbf{S} = \mathbf{D}\alpha_0 + \mathbf{Z}$ *constructed as a sparse combination of columns of the dictionary* $\mathbf{D}$ *of size* $N \times L$ *with mutual incoherence* M. *We assume a bounded noise*, $|\mathbf{Z}| \leqslant \mathbf{1} \cdot \varepsilon$. *Then, if* α_0 *is sparse enough, satisfying*

$$\|\alpha_0\|_0 < \frac{1+M}{2M + 2\sqrt{N}(\varepsilon+\delta)/T}, \tag{9}$$

then the solution $\hat{\alpha}$ *of*

$$(P_1(\delta)) \ \min_{\alpha} \|\alpha\|_1 \quad \text{subject to } |\mathbf{S} - \mathbf{D}\alpha| \leqslant \mathbf{1} \cdot \delta \tag{10}$$

with $\delta \geqslant \varepsilon$ *exhibits stability*

$$\|\hat{\alpha} - \alpha_0\|_1 \leqslant T. \tag{11}$$

Before turning to prove this result, let us first discuss its meaning. Starting with the case of $\varepsilon = \delta = 0$, this Theorem suggests that if

$$\|\alpha_0\|_0 < \frac{1+M}{2M}, \tag{12}$$

then $\hat{\alpha}$ and α_0 can be arbitrarily close, since any T—even zero—can be used here. In effect, this is exactly the result stated in Theorem 1.

When there is either noise in the signal (i.e. $\varepsilon > 0$) and/or noise is assumed in the BP solver (i.e. $\delta > \varepsilon = 0$) we obtain that guaranteed exact recovery of the original representation, $\hat{\alpha} = \alpha_0$, is impossible for any cardinality of α_0. Still, if we are willing to absorb a slight deviation in the outcome, then guaranteed performance can be claimed. We measure here the deviation between the original and the computed representations using ℓ^1 norm—the reason for this choice will become evident as we will turn to the proof of this result. If we allow a deviation of size T, then for sparse enough representations, the theorem guarantees success of the BP.

It is interesting to see that the critical cardinality to allow these results is a function of the signal-to-noise-ratio (SNR), $T/(\varepsilon + \delta)$. The smaller this value is, the stronger the effects of the noise, and then the requirement on the cardinality become more strict. As the SNR increases, we get near the noiseless case result.

Although not said explicitly, there is a hard restriction on the cardinality that may lead to successful BP behavior. The cardinality of the original representation, $\|\alpha_0\|_0$, must be always smaller than $(1+M)/2M$, no matter what the noise power is. This is an immediate outcome of the inequality presented in (9).

The above theorem could be given a different interpretation if we warp the inequality in (9) to be

$$T > \frac{2\sqrt{N}(\varepsilon+\delta)\|\alpha_0\|_0}{1 + M - 2M\|\alpha_0\|_0}.$$

This inequality suggests that if the cardinality of the representation is known, we can conclude the amount of deviation between the original representation and the BP's result.

As a final comment in this discussion we draw attention to the following limitation in the above result. At this stage we can claim stability only if $\varepsilon \leqslant \delta$, meaning that the noise assumed in the BP should be at least as big as the actual noise contaminating the signal.

We now turn to prove this Theorem, and the proof is a direct extension of the methodology used in [17] and later in [18] to study the noiseless case.

Proof. Consider a candidate solution of (10), β. Since it is indeed a candidate solution, it must satisfy the constraint $|\mathbf{S}-\mathbf{D}\beta|\leqslant \mathbf{1}\cdot\delta$. We compare this possible outcome with the desired outcome being α_0. Since we have assumed that $\varepsilon\leqslant\delta$, α_0 is also satisfying the constraint $|\mathbf{S}-\mathbf{D}\alpha|\leqslant\mathbf{1}\cdot\varepsilon\leqslant\mathbf{1}\cdot\delta$. In order for the BP as in (10) to favor α_0 as its solution, we should require

$$\|\beta\|_1-\|\alpha_0\|_1>0.$$

This should hold true for any candidate solution β. We can pose this requirement as a constrained optimization problem of the form

$$\min_{\beta}\quad \|\beta\|_1-\|\alpha_0\|_1 \text{ subject to }\quad |\mathbf{S}-\mathbf{D}\beta|\leqslant\mathbf{1}\cdot\delta \text{ and } |\mathbf{S}-\mathbf{D}\alpha_0|\leqslant\mathbf{1}\cdot\varepsilon\leqslant\mathbf{1}\cdot\delta.$$

If the minimum of this problem is negative, then BP can fail and produce a representation being different from the desired one.

The strategy employed to solve this problem and see whether it can get negative is to replace the penalty with lower bounds on it, and replace the constraints with wider ones. Both changes give more freedom to the problem to cross the zero towards negative values, and this way we study the worst-case behavior.

We start with the penalty and replace it with a lower bound. This bound was originally proposed in [15] and later used in [17–19]. The bound replaces the use of β by $\beta=\alpha_0+\mathbf{X}$. Defining $\mathscr{S}$ as the indices in the support of α_0, we have

$$\|\beta\|_1-\|\alpha_0\|_1=\|\alpha_0+\mathbf{X}\|_1-\|\alpha_0\|_1\geqslant\|\mathbf{X}\|_1-2\sum_{k\in\mathscr{S}}|x_k|. \tag{13}$$

Turning to the constraints, using our defined vector $\mathbf{X}$ we have

$$-\delta\cdot\mathbf{1}\leqslant\mathbf{D}\beta-\mathbf{S}=\mathbf{D}\mathbf{X}+\mathbf{D}\alpha_0-\mathbf{S}\leqslant\delta\cdot\mathbf{1}.$$

Since $-\varepsilon\cdot\mathbf{1}\leqslant\mathbf{D}\alpha_0-\mathbf{S}\leqslant\varepsilon\cdot\mathbf{1}$, the above relation can be brought to

$$-(\delta+\varepsilon)\cdot\mathbf{1}\leqslant-\delta\cdot\mathbf{1}-(\mathbf{D}\alpha_0-\mathbf{S})\leqslant\mathbf{D}\mathbf{X}\leqslant\delta\cdot\mathbf{1}-(\mathbf{D}\alpha_0-\mathbf{S})\leqslant(\delta+\varepsilon)\cdot\mathbf{1}.$$

Thus, we obtain an alternative, wider, constraint on the difference between the two representations, of the form

$$|\mathbf{D}\mathbf{X}|\leqslant(\delta+\varepsilon)\cdot\mathbf{1}. \tag{14}$$

Combining the new penalty term in (13) with the new constraint in (14) we obtain the following alternative simpler optimization task

$$\min_{\mathbf{X}}\|\mathbf{X}\|_1-2\sum_{k\in\mathscr{S}}|x_k|\quad\text{subject to }|\mathbf{D}\mathbf{X}|\leqslant(\delta+\varepsilon)\cdot\mathbf{1}. \tag{15}$$

If a solution $\mathbf{X}$ can be found such that the penalty value is negative, this indicates the BP's failure to choose the original representation. Unfortunately, though, this problem is hard to solve since it is posed both in terms of the entries of $\mathbf{X}$ and their absolute values. Thus, we further relax the constraint by using the mutual incoherence.

Defining $\mathbf{V}=\mathbf{D}\mathbf{X}$ we know that this vector has bounded entries, $|v_k|\leqslant(\delta+\varepsilon)$. Multiplying this vector by $\mathbf{D}^{\mathrm{T}}$ we get that the outcome is also bounded entry-wise by

$$|\mathbf{D}^{\mathrm{T}}\mathbf{V}|=|\mathbf{D}^{\mathrm{T}}\mathbf{D}\mathbf{X}|\leqslant(\delta+\varepsilon)\sqrt{N}\cdot\mathbf{1}. \tag{16}$$

Here we have exploited that fact that the dictionary columns are all ℓ^2-normalized. On the other hand, we have the trivial relation

$$\mathbf{D}^{\mathrm{T}}\mathbf{D}\mathbf{X}=\mathbf{X}+(\mathbf{D}^{\mathrm{T}}\mathbf{D}-\mathbf{I})\mathbf{X}. \tag{17}$$

The matrix $(\mathbf{D}^{\mathrm{T}}\mathbf{D}-\mathbf{I})$ contains exact zeros on its main diagonal, and all of its off-diagonal entries are smaller or equal to M in magnitude, being inner products of pairs of columns from $\mathbf{D}$. Combining (16) and (17) we have

$$\begin{aligned}|\mathbf{X}| &= |\mathbf{D}^{\mathrm{T}}\mathbf{D}\mathbf{X}-(\mathbf{D}^{\mathrm{T}}\mathbf{D}-\mathbf{I})\mathbf{X}|\\ &\leqslant |\mathbf{D}^{\mathrm{T}}\mathbf{D}\mathbf{X}|+|(\mathbf{D}^{\mathrm{T}}\mathbf{D}-\mathbf{I})\mathbf{X}|\\ &\leqslant (\delta+\varepsilon)\sqrt{N}\cdot\mathbf{1}+|(\mathbf{D}^{\mathrm{T}}\mathbf{D}-\mathbf{I})|\cdot|\mathbf{X}|\\ &\leqslant (\delta+\varepsilon)\sqrt{N}\cdot\mathbf{1}+M(\mathbf{1}-\mathbf{I})\cdot|\mathbf{X}|\\ &= (\delta+\varepsilon)\sqrt{N}\cdot\mathbf{1}+M\|\mathbf{X}\|_1\cdot\mathbf{1}-M|\mathbf{X}|.\end{aligned}$$

In the above we have used the fact that the off-diagonal entries of the Gram matrix $\mathbf{D}^{\mathrm{T}}\mathbf{D}$ are bounded by M in absolute value. The notation $\mathbf{1}$ stands for a matrix with all entries being '1'-es. The above leads to a different constraint resulting with the following optimization problem

$$\begin{aligned}&\min_{\mathbf{X}} \quad \|\mathbf{X}\|_1-2\sum_{k\in\mathcal{S}}|x_k|\\ &\text{subject to} \quad |\mathbf{X}|\leqslant\left(\frac{(\delta+\varepsilon)\sqrt{N}}{1+M}+\frac{M}{1+M}\|\mathbf{X}\|_1\right)\cdot\mathbf{1}.\end{aligned}$$

Since this problem is formulated in terms of the absolute entries of $\mathbf{X}$ we can simplify its description by defining $\mathbf{Y}=|\mathbf{X}|$. Furthermore, in its current formulation, the location of the non-zeros in $\mathcal{S}$ play no role. Thus, we simplify the problem by assuming that the first $|\mathcal{S}|$ entries in α_0 are non-zeros. These changes lead to the problem

$$\begin{aligned}&\min_{\mathbf{Y}} \quad (\mathbf{1}-2\cdot\mathbf{1}_{|\mathcal{S}|})^{\mathrm{T}}\mathbf{Y}\\ &\text{subject to} \left(\mathbf{I}-\frac{M}{1+M}\mathbf{1}\right)\quad\mathbf{Y}\leqslant\frac{(\delta+\varepsilon)\sqrt{N}}{1+M}\cdot\mathbf{1}\ \text{ and }\ \mathbf{Y}\geqslant 0.\end{aligned} \tag{18}$$

The vector $\mathbf{1}_{|\mathcal{S}|}$ has $|\mathcal{S}|$ '1'-es as its first entries and zeros elsewhere.

Instead of solving this LP problem, we turn to its dual. This idea, used successfully in [17] and later in [19], exploits the zero duality gap between the primal and the dual in feasible LP problems [22]. Thus, the sign of the penalty at the extreme (minimum or maximum, depending on the problem) is the same. Thus, if the dual problem is easier to solve, we can use this to draw a conclusion on the primal outcome.

For the primal problem defined in (18), the dual form is given by

$$\begin{aligned}&\max_{\mathbf{U}} \quad -\frac{(\delta+\varepsilon)\sqrt{N}}{1+M}\cdot\mathbf{1}^{\mathrm{T}}\mathbf{U}\\ &\text{subject to} \quad -\left(\mathbf{I}-\frac{M}{1+M}\mathbf{1}\right)\mathbf{U}\leqslant\mathbf{1}-2\cdot\mathbf{1}_{|\mathcal{S}|}\ \text{ and }\ \mathbf{U}\geqslant 0.\end{aligned} \tag{19}$$

It is trivial to see that no matter what $|\mathcal{S}|,\delta,\varepsilon,M$, and N are, the result of this problem is always non-positive and zero at its best. Thus, we conclude that finding α_0 as the recovered representation is impossible. This, however, refers to the case where we desire a perfect recovery. We now impose an additional constraint in (18), restricting $\mathbf{Y}$ to satisfy

$$\mathbf{1}^{\mathrm{T}}\mathbf{Y}=\|\mathbf{X}\|_1=\|\beta-\alpha_0\|_1\geqslant T.$$

This constraint focus the search for the competitive solution outside a T ball (in ℓ^1 for convenience) of the desired solution α_0. The basic rational here is that if a candidate better solution is indeed found but in a proximity to α_0, we consider this as a success as well. Adding this constraint to (18), the new problem become

$$\begin{aligned} &\min_{\mathbf{Y}} \quad \left(\mathbf{1} - 2 \cdot \mathbf{1}_{|\mathscr{S}|}\right)^{\mathrm{T}} \mathbf{Y} \\ &\text{subject to} \quad \left(\mathbf{I} - \frac{M}{1+M}\mathbf{1}\right)\mathbf{Y} \leqslant \frac{(\delta+\varepsilon)\sqrt{N}}{1+M} \cdot \mathbf{1}, \mathbf{1}^{\mathrm{T}}\mathbf{Y} \geqslant T \text{ and } \mathbf{Y} \geqslant 0. \end{aligned} \tag{20}$$

Note that if we have used ℓ^2-norm to measure proximity between solutions we would have lost the convexity. In the chosen format we stay convex and even preserve the LP structure. Thus, as before we turn to the dual and obtain

$$\begin{aligned} &\max_{\mathbf{U}} \quad -\frac{(\delta+\varepsilon)\sqrt{N}}{1+M} \cdot \mathbf{1}^{\mathrm{T}}\mathbf{U} + T \cdot u_0 \\ &\text{subject to} \quad \mathbf{1} \cdot u_0 - \left(\mathbf{I} - \frac{M}{1+M}\mathbf{1}\right)\mathbf{U} \leqslant \mathbf{1} - 2 \cdot \mathbf{1}_{|\mathscr{S}|} \text{ and } \mathbf{U} \geqslant 0,\ u_0 \geqslant 0. \end{aligned} \tag{21}$$

The additional constraint in the primal problem brought an additional scalar unknown to the dual one, u_0.

We are interested in a symbolic solution of the primal (or dual—the outcome should be the same) problem in order to find the relation between the various parameters involved and the condition they must satisfy in order to get a positive penalty value. Such solution is hard to obtain, and thus we have turned to the dual problem. By wisely guessing a solution for it, if this solution leads to a positive penalty, it must imply a positive maximum value for the dual and necessarily a positive value for the primal as well, leading to a guaranteed success of the BP. This is the rational also exercised in [17,19].

How shall we wisely choose a solution? While numerical solution of (21) cannot be used to obtain the desired relation between the parameters, it can certainly shed light on the structure of the solution. Given this structure we can pose it using few parameters, and solve for those parameters to guarantee positive penalty outcome. We have done so (see Appendix A for the details) and based on the results we propose the following structure for the solution of (21):

$$\mathbf{U} = A \cdot \mathbf{1}_{|\mathscr{S}|}.$$

Naturally we must force $A, u_0 \geqslant 0$. The other constraint translates into

$$\mathbf{1} \cdot (u_0 - 1) \leqslant \left(\mathbf{I} - \frac{M}{1+M}\mathbf{1}\right) A \cdot \mathbf{1}_{|\mathscr{S}|} - 2 \cdot \mathbf{1}_{|\mathscr{S}|}.$$

Using the fact that $\mathbf{1}\mathbf{1}_{|\mathscr{S}|} = \mathbf{1} \cdot |\mathscr{S}|$ we obtain

$$\left(u_0 - 1 + \frac{M|\mathscr{S}|A}{1+M}\right)\mathbf{1} \leqslant (A-2) \cdot \mathbf{1}_{|\mathscr{S}|}.$$

We can take these L inequalities and replace them with canonic two, one referring to the on-support and the other to the off-support in $\mathbf{1}_{|\mathscr{S}|}$. This leads to

$$\text{on-support} : u_0 + \frac{M|\mathscr{S}|A}{1+M} \leqslant A - 1 \tag{22}$$

$$\text{off-support} : u_0 + \frac{M|\mathscr{S}|A}{1+M} \leqslant 1. \tag{23}$$

The off-support inequality is stronger than the on-support one if $A \geqslant 2$. Let us suppose that this is the case, and discard of the on-support inequality. For a complete solution we should return to this junction and try the assumption $A \leqslant 2$ as well. Still, even without doing so, as long as we use the off-support constraint while satisfying $A \geqslant 2$, the solution obtained is sufficient for our needs, since we are not interested in maximization of the dual problem.

The penalty term to be maximized is supposed to be positive so as to guarantee the BP's success. This penalty is given by

$$0 < -\frac{(\delta+\varepsilon)\sqrt{N}}{1+M}\cdot \mathbf{1}^{\mathrm{T}}\mathbf{U} + T\cdot u_0 = -\frac{(\delta+\varepsilon)\sqrt{N}A|\mathscr{S}|}{1+M} + T\cdot u_0.$$

Plugging in the off-support constraint we obtain

$$\begin{aligned}0 &< -\frac{(\delta+\varepsilon)\sqrt{N}A|\mathscr{S}|}{1+M} + T\cdot u_0 \leqslant -\frac{(\delta+\varepsilon)\sqrt{N}A|\mathscr{S}|}{1+M} + T\cdot\left(1-\frac{M|\mathscr{S}|A}{1+M}\right)\\ &= \frac{T(1+M) - A\big((\delta+\varepsilon)\sqrt{N}+MT\big)|\mathscr{S}|}{1+M}.\end{aligned}$$

The value of A in the permissible range ($A \geqslant 2$ due to our assumption) that maximizes this value is $A = 2$, and this leads to the requirement

$$|\mathscr{S}| = \|\alpha_0\|_0 < \frac{1+M}{2M + 2\sqrt{N}(\varepsilon+\delta)/T},$$

as claimed by the theorem.

If we alternatively choose as active the on-support constraint in (22) one can easily verify that the analysis leads to the same solution. This implies that under the structure assumed for the solution of the dual LP problem, we have managed to locate the maximum value, thus getting tight result. □

What about the case of $\varepsilon > \delta$? Can we prove stability here just as well? In this case α_0 cannot be proposed as a possible competing solution since it does not satisfy the constraint in (10). While $|\mathbf{D}\alpha_0 - \mathbf{S}| < \mathbf{1}\cdot\varepsilon$ is true, it does not imply necessarily $|\mathbf{D}\alpha_0 - \mathbf{S}| < \mathbf{1}\cdot\delta$. Thus, an alternative solution should be proposed. One option is to choose

$$\hat{\alpha} = \mu\alpha_0 + (1-\mu)\mathbf{D}^{+}\mathbf{S},$$

with $0 \leqslant \mu \leqslant 1$. Put into the constraint we get

$$|\mathbf{D}\hat{\alpha} - \mathbf{S}| = |\mu\mathbf{D}\,\alpha_0 - \mu\mathbf{S}| \leqslant \mathbf{1}\cdot\varepsilon\mu.$$

Thus, by choosing $\mu = \delta/\varepsilon$ we bring the proposed solution to be a feasible solution to (10). Alternatively, we can propose $\hat{\alpha}$ to be the closest feasible solution to α_0, defined as

$$\hat{\alpha} = Arg\min_{\alpha}\|\alpha-\alpha_0\|_1 \quad \text{subject to } |\mathbf{D}\alpha - \mathbf{S}| \leqslant \delta\cdot\mathbf{1}.$$

This LP problem defines our candidate solution and for it we can check the behavior of the BP. We will not proceed with this issue here as it deviates from the main path we have undertaken. We leave this as an open problem for future work.

3. Special case of interest: union of ortho-bases

3.1. Stability result

In this section we treat the same stability problem, but concentrate on a special case where the dictionary is built as a union of J orthonormal matrices of size $N \times N$ each (thus, $L = JN$),

$$\mathbf{D} = [\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_J] \quad \text{where } \forall\ 1 \leqslant k \leqslant J,\ \mathbf{B}_k^{\mathrm{T}}\mathbf{B}_k = \mathbf{I}. \tag{24}$$

The analysis here is very similar to the one presented in [19], and generalizes it for the noisy case. We start by presenting the main result.

Theorem 4. *We assume a signal* $\mathbf{S} = \mathbf{D}\alpha_0 + \mathbf{Z}$ *constructed as a sparse combination of columns of the dictionary* $\mathbf{D}$ *of size* $N \times NJ$, *where* $\mathbf{D}$ *is constructed as a union of* J *orthonormal bases with mutual incoherence* M. *We assume a bounded noise,* $|\mathbf{Z}| \leqslant \mathbf{1} \cdot \varepsilon$. *Then, denoting the number of atoms taken from the* kth *ortho-basis by* $|\mathcal{S}_k|$, *and assuming* $|\mathcal{S}_1| \leqslant |\mathcal{S}_2| \leqslant \cdots \leqslant |\mathcal{S}_J|$, *the solution* $\hat{\alpha}$ *of*

$$(P_1(\delta)) \min_{\alpha} \|\alpha\|_1 \quad \text{subject to } |\mathbf{S} - \mathbf{D}\alpha| \leqslant \mathbf{1} \cdot \delta \tag{25}$$

with $\delta \geqslant \varepsilon$ *exhibits stability, namely*

$$\|\hat{\alpha} - \alpha_0\|_1 \leqslant T, \tag{26}$$

where

$$T > \frac{2(\delta+\varepsilon)\sqrt{N}(1 + M|\mathcal{S}_1|)\sum_{j=1}^{J} |\mathcal{S}_j|/(1 + M|\mathcal{S}_j|)}{1 + 2M|\mathcal{S}_1| - 2M(1 + M|\mathcal{S}_1|)\sum_{j=1}^{J} |\mathcal{S}_j|/(1 + M|\mathcal{S}_j|)}. \tag{27}$$

Furthermore, this result is true if the cardinalities $|\mathcal{S}_k|$ *satisfy*

$$\sum_{j=1}^{J} \frac{|\mathcal{S}_j|}{1 + M|\mathcal{S}_j|} < \frac{1 + 2M|\mathcal{S}_1|}{2M(1 + M|\mathcal{S}_1|)}. \tag{28}$$

Clearly, the requirements posed here are far more complicated and hard to interpret intuitively. Instead of getting a clear requirement on the cardinality of $\|\alpha_0\|_0$, the condition is with respect to the number of non-zeros in each of the ortho-bases. Still, at the heart of this result lays the same concept of obtained stability with the BP for sparse enough representations. We will prove first the result as presented here, and then turn to create a weaker but simpler version of it.

Proof. In fact, large portions of this proof are identical to the proof presented in the previous section. The general and the special dictionary cases both lead to the desire to solve (parallel to (15) in the previous proof, with the additional error-permitting constraint) the problem

$$\min_{\mathbf{X}} \quad \|\mathbf{X}\|_1 - 2\sum_{k \in \mathcal{S}} |x_k| \text{ subject to } \quad |\mathbf{D}\mathbf{X}| \leqslant (\delta+\varepsilon) \cdot \mathbf{1} \quad \text{and } \|\mathbf{X}\|_1 \geqslant T, \tag{29}$$

and obtain a positive value for the penalty, so as to guarantee BP's success. The two proofs depart here because of the different treatment we can now give to the constraint $|\mathbf{D}\mathbf{X}| \leqslant (\delta+\varepsilon) \cdot \mathbf{1}$, exploiting the structure of $\mathbf{D}$ to obtain a tighter inequality. The constraint can be written differently as

$$|[\mathbf{B}_1, \mathbf{B}_2, \ldots, \mathbf{B}_J]\mathbf{X}| = |\mathbf{B}_1 \cdot [\mathbf{I}, \mathbf{B}_1^{\mathrm{T}}\mathbf{B}_2, \ldots, \mathbf{B}_1^{\mathrm{T}}\mathbf{B}_J]\mathbf{X}| \leqslant (\delta+\varepsilon) \cdot \mathbf{1},$$

We use the property that for an orthonormal matrix $\mathbf{B}$ of size $N \times N$

$$|\mathbf{B}\mathbf{V}| \geqslant \frac{1}{\sqrt{N}}|\mathbf{V}|.$$

Using this inequality and breaking the vector $\mathbf{X}$ into J parts that match the J blocks in the dictionary, we obtain a weaker requirement of the form

$$|[\mathbf{I}, \mathbf{B}_1^{\mathrm{T}}\mathbf{B}_2, \ldots, \mathbf{B}_1^{\mathrm{T}}\mathbf{B}_J]\mathbf{X}| = \left|\mathbf{X}_1 + \sum_{k=2}^{J} \mathbf{B}_1^{\mathrm{T}}\mathbf{B}_k\mathbf{X}_k\right| \leqslant (\delta+\varepsilon)\sqrt{N} \cdot \mathbf{1}.$$

This leads to the inequality

$$|\mathbf{X}_1| \leqslant (\delta+\varepsilon)\sqrt{N}\cdot\mathbf{1} + \sum_{k=2}^{J} |\mathbf{B}_1^{\mathrm{T}}\mathbf{B}_k|\cdot|\mathbf{X}_k| \leqslant \left((\delta+\varepsilon)\sqrt{N} + M\sum_{k=2}^{J}\|\mathbf{X}_k\|_1\right)\cdot\mathbf{1}.$$

This inequality should be posed J times, for each of the J parts of the vector $\mathbf{X}$. For convenience we denote

$$\mathbf{P} = \begin{bmatrix} \mathbf{I} & -M\cdot\mathbf{1} & -M\cdot\mathbf{1} & \cdots & -M\cdot\mathbf{1} \\ -M\cdot\mathbf{1} & \mathbf{I} & -M\cdot\mathbf{1} & \cdots & -M\cdot\mathbf{1} \\ -M\cdot\mathbf{1} & -M\cdot\mathbf{1} & \mathbf{I} & \cdots & -M\cdot\mathbf{1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -M\cdot\mathbf{1} & -M\cdot\mathbf{1} & -M\cdot\mathbf{1} & \dots & \mathbf{I} \end{bmatrix}. \tag{30}$$

Using this matrix and putting the new constraint to the problem in (29) we get

$$\begin{aligned} &\min_{\mathbf{Y}} \quad (\mathbf{1}-2\cdot\mathbf{1}_{|\mathscr{S}|})^{\mathrm{T}}\mathbf{Y} \\ &\text{subject to} \quad \mathbf{P}\mathbf{Y} \leqslant (\delta+\varepsilon)\sqrt{N}\cdot\mathbf{1} \quad \text{and} \quad -\mathbf{1}^{\mathrm{T}}\mathbf{Y} \leqslant -T. \end{aligned} \tag{31}$$

This problem is similar to the problem posed in (20) with two major difference: First, the matrix $(\mathbf{I}-(M/(1+M))\mathbf{1})$ in the old-version constraint is replaced with $\mathbf{P}$. This change exploits the knowledge about some of the entries in the Gram matrix being identically zero. The second difference is the lack of division by $M+1$ in the constraint. Both differences eventually imply that the feasible set of $\mathbf{Y}$ is smaller than before, and thus we expect to get weaker requirements for the BP success.

As before we turn to the dual LP problem and choose a parametric solution. The dual problem is given by

$$\begin{aligned} &\max_{\mathbf{U}} \quad -(\delta+\varepsilon)\sqrt{N}\cdot\mathbf{1}^{\mathrm{T}}\mathbf{U} + T\cdot u_0 \\ &\text{subject to} \quad \mathbf{1}\cdot u_0 - \mathbf{P}\mathbf{U} \leqslant \mathbf{1}-2\cdot\mathbf{1}_{|\mathscr{S}|} \\ &\qquad\qquad \text{and} \quad \mathbf{U}\geqslant 0,\ u_0\geqslant 0. \end{aligned} \tag{32}$$

We break the vector $\mathbf{U}$ into $J=L/N$ equal parts, each referring to a different orthonormal matrix in our dictionary. Then we propose the solution (see the Appendix for an explanation for this choice of solution)

$$k=1,2,\ldots,J, \quad \mathbf{U}_k = A_k\cdot\mathbf{1}_{|\mathscr{S}_k|}. \tag{33}$$

Here we use the notation $|\mathscr{S}_k|$ to designate the number of non-zero entries referring to the kth ortho-basis. The positivity constraint implies $\forall k,\ A_k\geqslant 0$. Plugging this solution into the main constraint we obtain

$$\left(u_0 - 1 - MA_k|\mathscr{S}_k| + M\sum_{j=1}^{J}A_j|\mathscr{S}_j|\right)\mathbf{1} \leqslant (A_k-2)\mathbf{1}_{|\mathscr{S}_k|}, \quad k=1,2,3,\ldots,J.$$

As before, these inequalities encompass two scalar relations, referring to the on-and the off-support of $\mathbf{1}_{|\mathscr{S}_k|}$, being

$$\text{on-support:} \quad u_0 + M\sum_{j=1}^{J}A_j|\mathscr{S}_j| \leqslant A_k(1+M|\mathscr{S}_k|)-1,\ k=1,2,3,\ldots,J,$$

$$\text{off-support:} \quad u_0 + M\sum_{j=1}^{J}A_j|\mathscr{S}_j| \leqslant 1 + MA_k|\mathscr{S}_k|,\ k=1,2,3,\ldots,J.$$

As in the previous section, we can choose one type of constraint to be stronger than the other and work with it, while being consistent (verifying that this assumption is necessarily true). Based on the numerical results (see Appendix) we assume hereafter that the on-support constraints are stronger. This assumption is true if

$$\forall k = 1, 2, 3, \ldots, J, \quad A_k \leqslant 2.$$

Thus, we must enforce these inequalities as part of our solution. We further assume that all the on-support inequalities are met with exact equality, implying that all are active constraint. Thus

$$A_1(1 + M|\mathscr{S}_1|) = A_2(1 + M|\mathscr{S}_2|) = \cdots = A_J(1 + M|\mathscr{S}_J|) = C \tag{34}$$

and also

$$u_0 = C - 1 - M \sum_{j=1}^{J} A_j|\mathscr{S}_j|. \tag{35}$$

Returning to the dual LP penalty term, we should enforce $0 < -(\delta + \varepsilon)\sqrt{N} \cdot \mathbf{1}^{\mathrm{T}}\mathbf{U} + T \cdot u_0$. Plugging the relations in (34) and (35) we obtain

$$\begin{aligned}
0 &< -(\delta + \varepsilon)\sqrt{N} \cdot \sum_{j=1}^{J} A_j|\mathscr{S}_j| + T \cdot u_0 \\
&= -(\delta + \varepsilon)\sqrt{N} \cdot \sum_{j=1}^{J} A_j|\mathscr{S}_j| + T \cdot \left(C - 1 - M \sum_{j=1}^{J} A_j|\mathscr{S}_j|\right) \\
&= TC - T - \left(\sum_{j=1}^{J} \frac{|\mathscr{S}_j|C}{1 + M|\mathscr{S}_j|}\right) \cdot ((\delta + \varepsilon)\sqrt{N} + MT).
\end{aligned}$$

In order to maximize this expression, we should choose the maximal possible value for C. Assuming an ordered sequence $|\mathscr{S}_1| \leqslant |\mathscr{S}_2| \leqslant \cdots \leqslant |\mathscr{S}_J|$, we can choose $A_1 = 2$, and all other values being $A_j \leqslant 2$ while satisfying (34), and maximizing C, leading to $C = 2(1 + M|\mathscr{S}_1|)$. This leads to the requirement

$$0 < 2T(1 + M|\mathscr{S}_1|) - T - 2\left(\sum_{j=1}^{J} \frac{|\mathscr{S}_j|(1 + M|\mathscr{S}_1|)}{1 + M|\mathscr{S}_j|}\right) \cdot ((\delta + \varepsilon)\sqrt{N} + MT).$$

Written differently as a requirement on the ball radius around the true representation, T, we write

$$T > \frac{2(\delta + \varepsilon)\sqrt{N}(1 + M|\mathscr{S}_1|)\sum_{j=1}^{J} |\mathscr{S}_j|/(1 + M|\mathscr{S}_j|)}{1 + 2M|\mathscr{S}_1| - 2M(1 + M|\mathscr{S}_1|)\sum_{j=1}^{J} |\mathscr{S}_j|/(1 + M|\mathscr{S}_j|)} \tag{36}$$

obtaining the claimed result. This result is feasible only if the denominator is positive, thus adding a second requirement of the form

$$1 + 2M|\mathscr{S}_1| - 2M(1 + M|\mathscr{S}_1|) \sum_{j=1}^{J} \frac{|\mathscr{S}_j|}{1 + M|\mathscr{S}_j|} > 0.$$

This requirement can be rearranged to a simpler form, similar to the result obtained in [19], being

$$\sum_{j=1}^{J} \frac{|\mathscr{S}_j|}{1 + M|\mathscr{S}_j|} < \frac{1 + 2M|\mathscr{S}_1|}{2M(1 + M|\mathscr{S}_1|)}.$$

and this concludes the proof. □

The last condition appears (in slightly different structure but completely equivalent to the one here) in [19] as the condition on the cardinalities that leads to the success of the BP. This result parallels our sparsity requirement posed in (12) for the general dictionary case.

3.2. Two ortho-bases case revisited

When dealing with two ortho-bases ($J=2$), the above conditions could be further simplified. The requirement (27) after few algebraic steps becomes

$$\mathrm{SNR} = \frac{T}{\delta+\varepsilon} > 2\sqrt{N}\cdot\frac{|\mathcal{S}_1|+|\mathcal{S}_2|+2M|\mathcal{S}_1||\mathcal{S}_2|}{1-M|\mathcal{S}_2|-2M^2|\mathcal{S}_1||\mathcal{S}_2|}. \tag{37}$$

The condition for this bound to hold true is the positivity of the denominator:

$$1-M|\mathcal{S}_2|-2M^2|\mathcal{S}_1||\mathcal{S}_2|>0. \tag{38}$$

This requirement is exactly the one posed in [17] for the success of the BP in the noiseless case with a dictionary built by two ortho-bases.

Assuming that the denominator is positive, the above expression suggests a lower bound on the SNR that can be dealt with successfully with the BP. This bound increases with both cardinalities $|\mathcal{S}_1|$ and $|\mathcal{S}_2|$, exploding to infinity on the boundary of the condition in (38).

The above result could be read differently, leading to a different interpretation. Given the desired SNR (dictated by our application), we can ask what are the cardinalities $|\mathcal{S}_1|$ and $|\mathcal{S}_2|$ that will enable it. Using (37) we obtain the quadratic inequality that the cardinalities are to satisfy

$$2M\left(1+M\frac{\mathrm{SNR}}{2\sqrt{N}}\right)|\mathcal{S}_1||\mathcal{S}_2|+\left(1+M\frac{\mathrm{SNR}}{2\sqrt{N}}\right)|\mathcal{S}_2|+|\mathcal{S}_1|-\frac{\mathrm{SNR}}{2\sqrt{N}}<0. \tag{39}$$

Recall that we assume throughout these inequalities that $|\mathcal{S}_1|\leqslant|\mathcal{S}_2|$. Thus, we can sweep through $|\mathcal{S}_1|=1,2,3,\ldots$, and per each solve for the permissible $|\mathcal{S}_2|$ values, this way mapping the boundary of possible cardinalities that allow stability with the pre-specified SNR. Fig. 1 was obtained this way for varying values of SNR. As can be seen, as the SNR goes to infinity the cardinalities $|\mathcal{S}_1|$ and $|\mathcal{S}_2|$ are approaching those of the noiseless case as described in [17]. For convenience we plot the bound (12) of the general noiseless case, the bound $|\mathcal{S}_1|+|\mathcal{S}_2|<1/M$, and the bound $|\mathcal{S}_1|+|\mathcal{S}_2|<(\sqrt{2}-0.5)/M$, which is the simplified requirement for the success of the BP in the noiseless case with a dictionary built as two ortho-bases.

In fact, in this case we can obtain an analytic bound of the form $\|\alpha_0\|_0=|\mathcal{S}_1|+|\mathcal{S}_2|<f(\mathrm{SNR},M,N)$, depicting a worst case simplified bound. We achieve this by rearranging (39) to be

$$|\mathcal{S}_1|<\frac{(\mathrm{SNR}/2\sqrt{N})-(1+M(\mathrm{SNR}/2\sqrt{N}))|\mathcal{S}_2|}{2M(1+M(\mathrm{SNR}/2\sqrt{N}))|\mathcal{S}_2|+1}, \tag{40}$$

and then adding $|\mathcal{S}_2|$ to both sides, getting

$$|\mathcal{S}_1|+|\mathcal{S}_2|<\frac{(\mathrm{SNR}/2\sqrt{N})-(1+M(\mathrm{SNR}/2\sqrt{N}))|\mathcal{S}_2|}{2M(1+M(\mathrm{SNR}/2\sqrt{N}))|\mathcal{S}_2|+1}+|\mathcal{S}_2|. \tag{41}$$

The right-hand side expression bounding $|\mathcal{S}_1|+|\mathcal{S}_2|$ is a function of $|\mathcal{S}_2|$. By minimizing this function with respect to $|\mathcal{S}_2|$ we can replace this expression with a function of the form $f(\mathrm{SNR},M,N)$, and obtain the desired result.

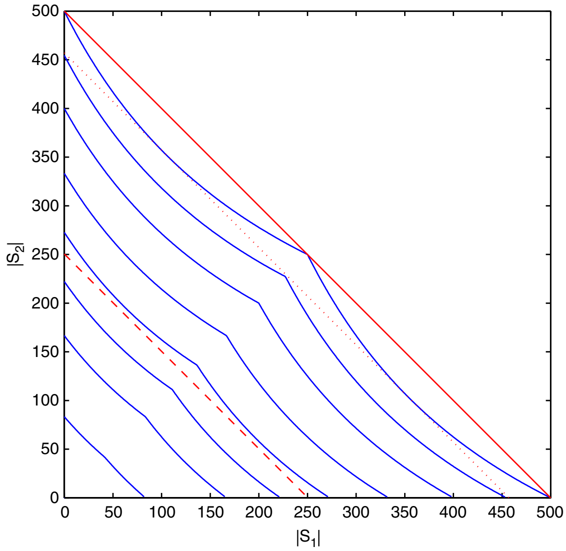

Fig. 1. The requirement on the cardinalities $|\mathcal{S}_1|$ and $|\mathcal{S}_2|$ as a function of the SNR (its values are in the range $1E+5$ to $1E+10$). The parameters used here are $N = 500^2$ and $M = \frac{1}{500}$. In red we plot the lines $|\mathcal{S}_1| + |\mathcal{S}_2| < (M+1)/2M$, $|\mathcal{S}_1| + |\mathcal{S}_2| < 1/M$, and $|\mathcal{S}_1| + |\mathcal{S}_2| < (\sqrt{2} - 0.5)/M$.

Taking a derivative of the right-hand side and zeroing it leads to the value $|\mathcal{S}_2|$ that minimizes this expression (the second derivative confirm this claim but we omit this development here). It is easy to verify that this leads to

$$|\mathcal{S}_2|_{\mathrm{opt}} = \frac{\sqrt{(1 + M(\mathrm{SNR}/2\sqrt{N})) + 2M\,(\mathrm{SNR}/2\sqrt{N}) \cdot (1 + M(\mathrm{SNR}/2\sqrt{N}))} - 1}{2M(1 + M(\mathrm{SNR}/2\sqrt{N}))}. \tag{42}$$

Plugging this into (41) we obtain a lower bound on the right-hand side. Instead of simplifying what seems like a very complex expression, we simply plug $|\mathcal{S}_2|_{\mathrm{opt}}$ to (41), getting

$$|\mathcal{S}_1| + |\mathcal{S}_2| < \frac{(\mathrm{SNR}/2\sqrt{N}) - (1 + M(\mathrm{SNR}/2\sqrt{N}))|\mathcal{S}_2|_{\mathrm{opt}}}{2M(1 + M(\mathrm{SNR}/2\sqrt{N}))|\mathcal{S}_2|_{\mathrm{opt}} + 1} + |\mathcal{S}_2|_{\mathrm{opt}}. \tag{43}$$

Fig. 2 presents the original bound curves and the simplified requirement developed above. As can be seen, every curve bound is matched with a diagonal line of slope -1 that bounds the curve from below, thus guaranteeing worst-case requirement.

We have special interest in the case where $\mathrm{SNR} \to \infty$, since we are already familiar with results for the noiseless case. We have

$$\lim_{\mathrm{SNR}\to\infty} |\mathcal{S}_2|_{\mathrm{opt}} = \lim_{\mathrm{SNR}\to\infty} \frac{\sqrt{\left(1 + M(\mathrm{SNR}/2\sqrt{N})\right) + 2M\,(\mathrm{SNR}/2\sqrt{N}) \cdot \left(1 + M(\mathrm{SNR}/2\sqrt{N})\right)} - 1}{2M(1 + M(\mathrm{SNR}/2\sqrt{N}))} = \frac{1}{\sqrt{2}M}.$$

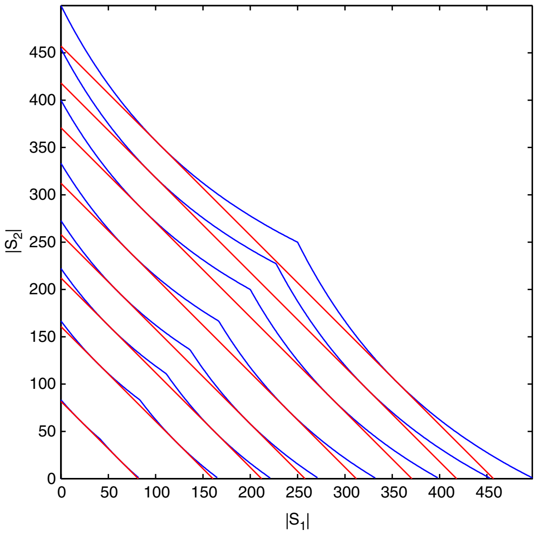

Fig. 2. The requirement on the cardinalities $|\mathscr{S}_1|$ and $|\mathscr{S}_2|$—original (Eq. (37)) in blue, and simplified (Eq. (43)) forms in red. The parameters are as in Fig. 1.

Plugged into (43) we obtain that the bound becomes

$$\begin{aligned} |\mathscr{S}_1|+|\mathscr{S}_2| &< \lim_{\mathrm{SNR}\to\infty} \frac{(\mathrm{SNR}/2\sqrt{N})-(1+M(\mathrm{SNR}/2\sqrt{N}))|\mathscr{S}_2|_{\mathrm{opt}}}{2M(1+M(\mathrm{SNR}/2\sqrt{N}))|\mathscr{S}_2|_{\mathrm{opt}}+1}+|\mathscr{S}_2|_{\mathrm{opt}} \\ &= \lim_{\mathrm{SNR}\to\infty} \frac{1-M|\mathscr{S}_2|_{\mathrm{opt}}}{2M^2|\mathscr{S}_2|_{\mathrm{opt}}}+|\mathscr{S}_2|_{\mathrm{opt}} = \frac{\sqrt{2}-0.5}{M}, \end{aligned} \tag{44}$$

which is the result given in [17].

3.3. The general case—simplified bound

An important shortcoming of the obtained result in Theorem 4 is the lack of a simple condition in the form of (9) that relates the radius of error, the noise power, and the cardinality of the original representation. We seek a condition similar to (43) that was developed for the two ortho-bases case. Note that since the requirement in (43) does not have a simple form, we will be satisfied with a computable expression being of the form $f(\mathrm{SNR}, M, N, J)$.

The following theorem walks through similar scenery as in [19] to propose such a simpler structured requirement. As we will show shortly, the result in [19] is correct, but not tight, and thus can be further improved. We shall propose such an improvement, and generalize both bounds to the noisy case.

Theorem 5. *We assume a signal* $\mathbf{S}=\mathbf{D}\alpha_0+\mathbf{Z}$ *constructed as a sparse combination of columns of the dictionary* $\mathbf{D}$ *of size* $N\times NJ$, *where* $\mathbf{D}$ *is constructed as a union of* J *orthonormal bases with mutual incoherence* M. *We assume a bounded noise*, $|\mathbf{Z}|\leqslant\mathbf{1}\cdot\varepsilon$. *Then, stability of the BP with* ℓ^1 *error smaller than* T *is* *guaranteed if*

$$\|\alpha_0\|_0 < \begin{cases} \dfrac{1}{M} \cdot \left[\dfrac{(J-1+\sqrt{(MR+1)/(2MR+1)})^2}{J-1+1/2MR+1} - J \right] & J \in \mathscr{C}(M,R) \\ \dfrac{1}{M} \cdot \dfrac{(J-1)MR/(2MR+1)}{J-1-(MR/(2MR+1))} & \textit{Else}, \end{cases} \tag{45}$$

where

$$R = \frac{T}{2\sqrt{N}(\varepsilon+\delta)}, \tag{46}$$

and the condition needed above, $J \in \mathscr{C}(M,R)$, is all J satisfying

$$\frac{(J-1+\sqrt{(MR+1)/(2MR+1)})^2}{J-1+1/2MR+1} - J \geqslant \frac{(J-1)(1-\sqrt{(MR+1)/(2MR+1)})}{\sqrt{(MR+1)/(2MR+1)}}. \tag{47}$$

Before proving this theorem, a natural question that should be asked is how this bound matches the more general form given in Section 3. Fig. 3 presents a comparison of the bounds in (9) and the upper part of (45), assuming $J \in \mathscr{C}(M,R)$. We have fixed $M = 1E-2$ and we show the bound curves as a function of J for varying R. As can be seen, for small enough J the new bound is "more generous" due to its ability to exploit the dictionary structure. The crossing between the bounds is an artifact that will be cleared at a later stage. This behavior implies that the obtained bound is wrong and should be fixed—this is indeed done by the condition posed and the alternative bound supplied. We will present more about these two bounds as we proceed into the proof.

Proof. In the previous theorem we got condition (36) in order to guarantee stability of the BP,

$$\frac{\mathrm{SNR}}{2\sqrt{N}} > \frac{(1+M|\mathscr{S}_1|)\sum_{j=1}^{J}|\mathscr{S}_j|/(1+M|\mathscr{S}_j|)}{1+2M|\mathscr{S}_1|-2M(1+M|\mathscr{S}_1|)\sum_{j=1}^{J}|\mathscr{S}_j|/(1+M|\mathscr{S}_j|)}.$$

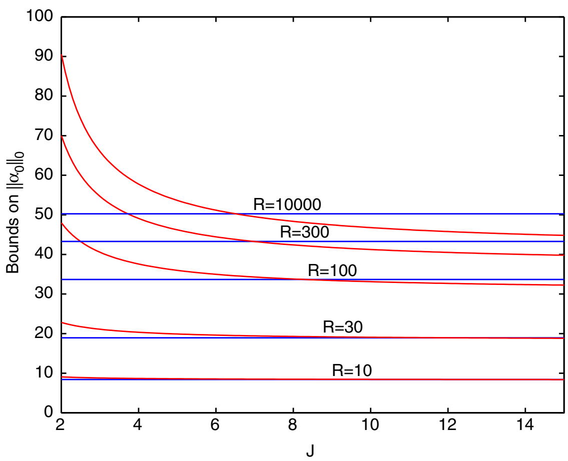

Fig. 3. Comparison between the general bound in (9) and the special one in (45) developed for the union of ortho-bases. Per each value of R the blue curve describes the general bound and the red one shows the new bound. M is assumed to be 0.01, thus giving cardinalities smaller than $1/M = 100$. The case of $R \to \infty$ corresponds to the noiseless case, where these results match those in [19].

Reorganizing this expression and using the definitions $R = \mathrm{SNR}/2\sqrt{N}$ and $y_k = M|\mathscr{S}_k|,\ k = 1, 2, \ldots, J$, we obtain the condition

$$g(y_1, y_2, \ldots, y_J) = \sum_{j=2}^{J} \frac{y_j}{1 + y_j} - \frac{MR - y_1}{(2MR + 1)(1 + y_1)} < 0.$$

Among all the support combinations $\{y_j\}_{j=1}^{J}$ that sum to a constant C we are interested in those that bring the requirement to its extreme. Thus we should solve the optimization problem

$$\max_{y_1, y_2, \ldots, y_J} \quad g(y_1, y_2, \ldots, y_J) \quad \text{subject to} \quad \sum_{j=1}^{J} y_j = C \ \text{ and } \ \{0 \leqslant y_1 \leqslant y_j\}_{j \geqslant 2}. \tag{48}$$

Using Lagrange multipliers method we obtain

$$\mathscr{L}\{y_1, y_2, \ldots, y_J\} = \sum_{j=2}^{J} \frac{y_j}{1 + y_j} - \frac{MR - y_1}{(2MR + 1)(1 + y_1)} - \lambda \left(\sum_{j=1}^{J} y_j - C \right). \tag{49}$$

We deliberately disregard the inequality constraints $\{0 \leqslant y_1 \leqslant y_j\}_{j \geqslant 2}$ assuming that the solution will satisfy these nevertheless. This, of course, must be verified when a solution is obtained. Taking the derivatives with respect to y_j we obtain

$$\frac{\partial \mathscr{L}}{\partial y_1} = 0 \rightarrow y_1 = \sqrt{\frac{1 + MR}{1 + 2MR}} \cdot \frac{1}{\sqrt{\lambda}} - 1 \quad \text{and for } j \geqslant 2: \ \frac{\partial \mathscr{L}}{\partial y_j} = 0 \ \rightarrow \ y_j = \frac{1}{\sqrt{\lambda}} - 1. \tag{50}$$

It is easily verified that the constraints $y_1 \leqslant y_j$ for $j \geqslant 2$ are indeed satisfied.

Having found the expressions for y_j we first put them into the sum constraint to obtain a relation between C and the Lagrange multiplier λ. This gives

$$\sum_{j=1}^{J} y_j = C = (J - 1)\left(\frac{1}{\sqrt{\lambda}} - 1 \right) + \sqrt{\frac{1 + MR}{1 + 2MR}} \cdot \frac{1}{\sqrt{\lambda}} - 1,$$

leading to

$$\frac{1}{\sqrt{\lambda}} = \frac{C + J}{J - 1 + \sqrt{(1 + MR)/(1 + 2MR)}}. \tag{51}$$

Similarly, we can describe $g(y_1, y_2, \ldots, y_J)$ as a function of λ, obtaining

$$0 > g(y_1, y_2, \ldots, y_J) = (J - 1)(1 - \sqrt{\lambda}) - \frac{\sqrt{(MR + 1)(2MR + 1)}\sqrt{\lambda} - 1}{1 + 2MR}$$

The above, together with (51), give a requirement on λ being

$$\sqrt{\lambda} = \frac{J - 1 + \sqrt{(MR + 1)/(2MR + 1)}}{C + J} > \frac{J - 1 + 1/(2MR + 1)}{J - 1 + \sqrt{(MR + 1)/(2MR + 1)}}. \tag{52}$$

The relation in (52) leads to the desired bound on the cardinality, by

$$\|\alpha_0\|_0 = \sum_{j=1}^{J} |\mathscr{S}_j| = \frac{1}{M} \cdot \sum_{j=1}^{J} y_j = \frac{C}{M} < \frac{1}{M} \cdot \left[\frac{(J - 1 + \sqrt{(MR + 1)/(2MR + 1)})^2}{J - 1 + 1/(2MR + 1)} - J \right].$$

This is the claim of the theorem, although we still have to establish that it is conditioned. For that, we should now add another ingredient that is missing in [19]. The above result is correct only if the solution of (48) satisfies $y_1 \geqslant 0$. We have already argued that if this is true, then all other y_j are non-negative as well,

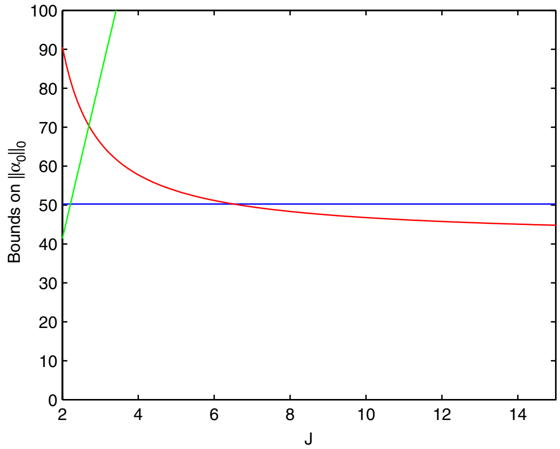

Fig. 4. For the same comparison as in Fig. 3 and with $R = 1E4$, the green line presents the condition for the correctness of the simple bound obtained. If the bound (red curve) is above the green one, the bound is correct. In this figure this happens to be true for $J = 2$ only.

thus removing the need to test their non-negativity. Thus, we must require

$$y_1 \geqslant 0 \quad \rightarrow \quad \sqrt{\frac{(MR+1)}{(2MR+1)}} \geqslant \sqrt{\lambda}. \tag{53}$$

Using the expression for λ as in (51) we get yet another condition on C being

$$C \geqslant \frac{(J-1)\left(1-\sqrt{(MR+1)/(2MR+1)}\right)}{\sqrt{(MR+1)/(2MR+1)}}. \tag{54}$$

Using the expression for C given above, this is equivalent to the requirement

$$\frac{(J-1+\sqrt{(MR+1)/(2MR+1)})^2}{J-1+1/(2MR+1)} - J \geqslant \frac{(J-1)\left(1-\sqrt{(MR+1)/(2MR+1)}\right)}{\sqrt{(MR+1)/(2MR+1)}}. \tag{55}$$

This requirement essentially leads to a quadratic equation posing a condition on J. We can pose this condition on the bound of cardinality by dividing by M. Fig. 4 present the two bounds as in Fig. 3 for $R = 1E4$ (representing nearly a noiseless case), and along these curves we present the restriction on C/M as in (54). As we can see, for small enough J values the condition is met (the red curve is beneath the condition). In fact, even $J = 3$ already is not satisfying this condition, and thus the bound obtained is wrong.

The reason for this phenomenon is that as J grows, the cardinality should be divided between the J sections of the representation vector. For small cardinalities, the optimal solution of (48) should lead to $y_1 = 0$ and equal division of the support between the other $J-1$ parts. In fact, even this is not exact since we have to propose an integer solution, but we will disregard this effect. This explains why our "tighter" bound is not truly tighter as J crosses some critical value, as manifested in Fig. 5.

Considering the case where (54) is not satisfied, we necessarily have to assume $y_1 = 0$. Then the optimization posed in (48) leads to $y_2 = y_3 = \cdots = y_J = x$. The sum constraint in (48) leads to

$$(J-1)x = C \quad \rightarrow \quad x = \frac{C}{J-1},$$

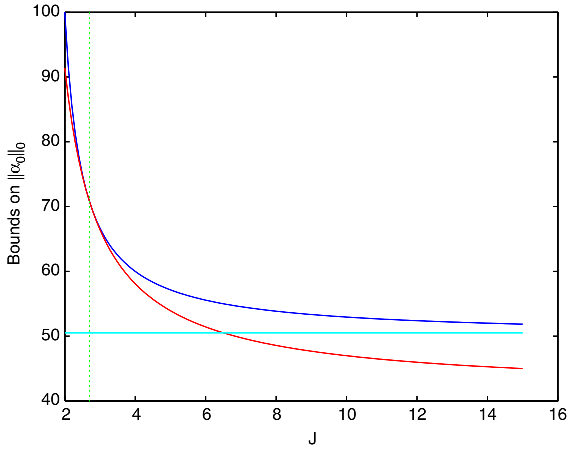

Fig. 5. Comparing the bounds for the noiseless case: (i) the general bound $(0.5(1+1/M))$ in cyan, (ii) the simplified bound in [19] $(1/M(\sqrt{2}-1+(1/2(J-1))))$ in red, and (iii) the new bound $(1/M(0.5+1/(4J-6)))$ in blue. We also show the cross-over point $J \approx 2.7$ (green).

and the value of the penalty in (48) becomes

$$g(y_1, y_2, \ldots, y_J) = \frac{(J-1)C}{J-1+C} - \frac{MR}{2MR+1}.$$

Requiring negativity of this function leads to the condition

$$g(y_1, y_2, \ldots, y_J) < 0 \quad \rightarrow \quad C < \frac{((J-1)MR)/(2MR+1)}{J-1-(MR/(2MR+1))},$$

and this leads to the final cardinality requirement

$$\|\alpha_0\|_0 = \sum_{j=1}^{J} |\mathscr{S}_j| = \frac{1}{M} \cdot \sum_{j=1}^{J} y_j = \frac{C}{M} < \frac{1}{M} \cdot \frac{((J-1)MR)/(2MR+1)}{J-1-(MR/(2MR+1))}, \tag{56}$$

which is the requirement for values of J for which the previous results are wrong (due to $y_1 < 0$). This concludes the proof as we have shown the two bounds and the decision rule between them. □

3.4. The noiseless case—updated results

We now check the obtained simplified results for $R \rightarrow \infty$ and compare to the result in [19]. The Theorem's first bound becomes

$$\lim_{R\rightarrow\infty} \frac{1}{M} \cdot \left[\frac{\left(J-1+\sqrt{(MR+1)/(2MR+1)}\right)^2}{J-1+1/(2MR+1)} - J \right] = \frac{1}{M} \cdot \left[\sqrt{2} - 1 + \frac{1}{2(J-1)} \right]$$

and this is the result shown given in [19]. However, based on (54), the above result is true only for small enough J, satisfying

$$C \geqslant \lim_{R\rightarrow\infty} \frac{(J-1)\left(1-\sqrt{(MR+1)/(2MR+1)}\right)}{\sqrt{(MR+1)/(2MR+1)}} = (J-1)(\sqrt{2}-1). \tag{57}$$

Thus, using the above result $C = \sqrt{2} - 1 + (1/2(J-1))$ and using the above inequality we get

$$C = \sqrt{2} - 1 + \frac{1}{2(J-1)} \geqslant (J-1)(\sqrt{2}-1) \quad \rightarrow J \leqslant \frac{2\sqrt{2}}{4(\sqrt{2}-1)} + 1 = 2.7071. \tag{58}$$

For $J \geqslant 3$ we thus use the alternative bound as in (56), being

$$\|\alpha_0\|_0 < \lim_{R\to\infty} \frac{1}{M} \cdot \frac{(J-1)MR/(2MR+1)}{J-1-MR/(2MR+1)} = \frac{1}{M} \cdot \left(\frac{1}{2} + \frac{1}{4J-6}\right). \tag{59}$$

The various bounds are described for the noiseless case in Fig. 5. We see that after the critical value of J as in (58) we get a more optimistic bound that preserves the improvement over the general case. As a side note we should add that this gain is lost again for high enough J (above 50, where the solution again needs to be corrected. We believe that this effect should be attributed to the limit on M as J grows (specifically—it can no longer stay $1/\sqrt{N}$), but we will not pursue this matter here.

4. Relation to existing work

Analysis of the BP is a new topic and contributions to this field were made in the past 3–4 years. Among these, the treatment of the noisy case was addressed only recently in two parallel major efforts [10,11]. To the best of the authors knowledge, these manuscripts, along with this one are the only attempts to study how the BP behaves in the presence of noise.

The analysis in [10] addresses uniqueness of sparse representations in the presence of noise, stability of the BP, recovery of the correct support with BP, and similar analysis (stability and support recovery) for several versions of the greedy algorithm. The results proposed here could be regarded as direct extension of the work in [10]. Here we have concentrated on the stability of the BP, but the outcome is different in several ways:

1. The result obtained here approaches that of the noiseless case for zero noise. This is not true for the results reported in [10].
2. The analysis is based on a different noise model and error measurements—while this is not crucial, we believe that it is these changes that made it possible to draw the tighter results.
3. In this paper we have added the treatment of the union of orthobases as a special case of interest.

The work by Tropp in [11] parallels the one in [10], but offers a different point of view and thus somewhat different results. Rather than discussing the stability property, Tropp concentrates on showing equivalence between the ℓ^0 and the ℓ^1 problem formations. Several modes of operations are discussed and studied, with sparsity and representation-accuracy forced either as constraints or penalties. As in the comparison made above, the results given here are tighter than those in Tropps work as well.

Beyond the bound-tightening that takes place in this work, and the treatment of the amalgam of orthobases as a special kind of dictionary, this paper offers another more important benefit. The description of the problem, the Theorems proposed, and especially the proof to back them up are all far simpler than those discussed in [10,11]. In fact, we suggest that if anyone is to study the analysis of the BP under noise results, this paper should necessarily be the starting point, as it offers a simple picture, and one that gives smooth continuation of the study of the noiseless BP case to the noisy one. This work is a direct extension of [18] and the work that preceded it in [17]. We believe that this is the true and major contribution of this work.

As a final note we should mention that as a by product of this work, in the study of the union of ortho-bases as a dictionary, this paper proposes a new and corrected result for the equivalence bound, compared to the work in [19].

5. Conclusions

In this paper we addressed the BP behavior in the presence of noise. We have proven two major stability results, one that addresses the general dictionary case and the other for the special case where the dictionary is built as a union of ortho-base. The bottom line result could be summarized by the the following phrases: *The BP successfully recovers sparse representations of signals, even if those signals are contaminated by noise*. In this work we give the exact conditions for this claim to be true.

Acknowledgements

The author would like to thank Prof. Irad Yavneh from the Computer Science department at the Technion, Israel Institute of Technology, for his help in deciphering some confusing inequalities that turned important for this work.

Appendix A. —Matlab Programs to Solve (21) and (32)

The following Matlab program builds the problem as posed in (21) and obtains a solution for it.

```
L=60; N=15;
M=sqrt((L-N)/(L-1)/N);
s=2; c=0.001; T=0.1;
f=[-ones(L,1)*c; T];
A=[-eye(L)+M/(1+M)*ones(L), ones(L,1)];
B=ones(L,1); B(1:s)=-1;
LB=zeros(L+1,1); UB=1000*ones(L+1,1);
U=linprog(-f,A,B,[],[],LB,UB);
X=linprog(B,-A',-f,[],[],LB);
disp([f'*U,B'*X]); plot(U)
```

Similar to the above, the following program simulates (32) and builds a solution for it.

```
L=60; N=15;
M=sqrt((L-N)/(L-1)/N);
s=3; c=0.001; T=0.1;
f=[-ones(L,1)*c; T];
A=[M*ones(L), ones(L,1)];
for k=1:1:L/N,
  A((k-1)*N+1:k*N,(k-1)*N+1:k*N)=-eye(N);
end;
B=ones(L,1); B(1:s)=-1;
LB=zeros(L+1,1); UB=1000*ones(L+1,1);
```

```
U=linprog(-f,A,B,[],[],LB,UB);
X=linprog(B,-A',-f,[],[],LB);
disp([f'*U,B'*X]); plot(U)
```

References

[1] S.S. Chen, D.L. Donoho, M.A. Saunders, Atomic decomposition by basis pursuit, SIAM Rev. 43 (1) (2001) 129–159.
[2] B.K. Natarajan, Sparse approximate solutions to linear systems, SIAM J. Comput. 24 (1995) 227–234.
[3] G. Davis, S. Mallat, M. Avellaneda, Greedy adaptive approximation, J. Constr. Approx. 13 (1997) 57–98.
[4] S. Mallat, Z. Zhang, Matching pursuit in a time-frequency dictionary, IEEE Trans. Signal Process. 41 (1993) 3397–3415.
[5] Y.C. Pati, R. Rezaiifar, P.S. Krishnaprasad, Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition, Proceedings of the 27th Annual Asilomar Conference on Signals, Systems, and Computers, 1993.
[6] C. Couvreur, Y. Bresler, On the optimality of the backward greedy algorithm for the subset selection problem, SIAM J. Matrix Anal. Appl. 21 (3) (2000) 797–808.
[7] A.C. Gilbert, S. Muthukrishnan, M.J. Strauss, Approximation of functions over redundant dictionaries using coherence, 14th Annual ACM–SIAM Symposium Discrete Algorithms, 2003.
[8] J.A. Tropp, A.C. Gilbert, S. Muthukrishnan, M.J. Strauss, Improved sparse approximation over quasi-incoherent dictionaries, IEEE International Conference on Image Processing, Barcelona, September, 2003.
[9] J.A. Tropp, Greed is good: algorithmic results for sparse approximation, IEEE Trans. Inform. Theory 50 (10) (2004) 2231–2242.
[10] D.L. Donoho, M. Elad, V. Temlyakov, Stable recovery of sparse overcomplete representations in the presence of noise, IEEE Trans. Inform. Theory, to appear.
[11] J.A. Tropp, Just relax: convex programming methods for subset selection and sparse approximation, IEEE Trans. Inform. Theory, to appear.
[12] V.N. Temlyakov, Greedy algorithms and *m*-term approximation, J. Approx. Theory 98 (1999) 117–145.
[13] V.N. Temlyakov, Weak greedy algorithms, Adv. Comput. Math. 5 (2000) 173–187.
[14] I.F. Gorodnitsky, B.D. Rao, Sparse signal reconstruction from limited data using FOCUSS: a re-weighted norm minimization algorithm, IEEE Trans. Signal Process. 45 (3) (March 1997) 600–616.
[15] D.L. Donoho, X. Huo, Uncertainty principles and ideal atomic decomposition, IEEE Trans. Inform. Theory 47 (7) (2001) 2845–2862.
[16] X. Huo, Sparse image representation via combined transforms, Ph.D. Thesis, Stanford, 1999.
[17] M. Elad, A.M. Bruckstein, A generalized uncertainty principle and sparse representation in pairs of $\Re^N$ bases, IEEE Trans. Inform. Theory 48 (2002) 2558–2567.
[18] D.L. Donoho, M. Elad, Optimally sparse representation in general (non-orthogonal) dictionaries via ℓ^1 minimization, Proc. Natl. Acad. Sci. 100 (2002) 2197–2202.
[19] R. Gribonval, M. Nielsen, Sparse representations in unions of bases, IEEE Trans. Inform. Theory 49 (12) (2003) 3320–3325.
[20] J.-J. Fuchs, On sparse representations in arbitrary redundant bases, IRISA Technical Report, Univ. de Rennes I, Dec. IEEE Trans. Inform. Theory 50 (6) (2004) 1341–1344.
[21] T. Strohmer, R. Heath Jr., Grassmannian frames with applications to coding and communcations, Appl. Comp. Harm. Anal. 14 (3) 257–275.
[22] D.P. Bertsekas, Nonlinear Programming, Athena Scientific, Belmont, MA, 1995.