Recovery of Short, Complex Linear Combinations via ℓ_1 Minimization

Joel A. Tropp, *Member, IEEE*

***Abstract*— This note provides a condition under which ℓ_1 minimization (also known as Basis Pursuit) can recover short linear combinations of complex vectors chosen from fixed, overcomplete collection. This condition has already been established in the real setting by J.-J. Fuchs, who used convex analysis. The proof given here is more direct.**

***Index Terms*— Algorithms, approximation, Basis Pursuit, linear program, redundant dictionaries, sparse representations.**

I. Introduction

The (complex) sparse approximation problem is set in the Hilbert space $\mathbb{C}^d$. For practical reasons, we work in a finite-dimensional space, but the theory can be extended to the infinite-dimensional setting. A *dictionary* for $\mathbb{C}^d$ is a finite collection of unit-norm vectors that spans the whole space. The elements of the dictionary are called *atoms*, and they are denoted by $\boldsymbol{\varphi}_\omega$, where the parameter ω is drawn from an index set Ω. The letter N will indicate the number of atoms in the dictionary. Now, form the *dictionary synthesis matrix*, whose columns are atoms:

$$\boldsymbol{\Phi} \stackrel{\text{def}}{=} \begin{bmatrix} \boldsymbol{\varphi}_{\omega_1} & \boldsymbol{\varphi}_{\omega_2} & \dots & \boldsymbol{\varphi}_{\omega_N} \end{bmatrix}.$$

The order of the atoms does not matter, so long as it is fixed.

Given a signal $\boldsymbol{s}$ from $\mathbb{C}^d$, the problem is to determine the shortest linear combination of atoms that equals the signal. If we define $\|\boldsymbol{b}\|_0$ to be the number of nonzero components of the vector $\boldsymbol{b}$, then we may write this *sparse approximation problem* as

$$\min_{\boldsymbol{b} \in \mathbb{C}^d} \|\boldsymbol{b}\|_0 \qquad \text{subject to} \qquad \boldsymbol{\Phi}\,\boldsymbol{b} = \boldsymbol{s}. \tag{P_0}$$

This problem is somewhat academic since the signals that have a sparse representation using fewer than d atoms form a set of Lebesgue measure zero in $\mathbb{C}^d$ [1, Prop. 4.1]. Nevertheless, the question has value for the insight it can provide on more difficult sparse approximation problems.

One approach to solving (P_0) is to replace the horribly nonlinear function $\|\cdot\|_0$ with the norm $\|\cdot\|_1$ and hope that the solutions coincide. That is,

$$\min_{\boldsymbol{b} \in \mathbb{C}^d} \|\boldsymbol{b}\|_1 \qquad \text{subject to} \qquad \boldsymbol{\Phi}\,\boldsymbol{b} = \boldsymbol{s}. \tag{P_1}$$

This convex minimization problem can be solved efficiently with standard mathematical programming software. Chen, Donoho, and Saunders introduce this method in [2], where they call it *Basis Pursuit*. They provide copious empirical evidence that the method of ℓ_1 minimization can indeed solve (P_0).

Several years ago, Donoho and Huo established that the Basis Pursuit method provably recovers short linear combinations of vectors from *incoherent* dictionaries [3]. Roughly speaking, an incoherent dictionary has small inner products between its atoms. This basic result was sharpened and extended by Elad–Bruckstein [4], Donoho–Elad [5], Gribonval–Nielsen [6], and Tropp [1]. The strongest result in this direction, which we will soon explore, is due to Fuchs [7]. This article provides a completely different method of reaching Fuchs' result.

J. A. Tropp was with ICES, The University of Texas at Austin, Austin, TX 78712. He is presently with the Dept. of Mathematics of The University of Michigan at Ann Arbor. This research was supported by an NSF Graduate Fellowship. E-mail: `jtropp@umich.edu`.

II. Fuchs' Condition

Imagine that the sparsest representation of a given signal $\boldsymbol{s}$ requires m atoms, say

$$\boldsymbol{s} = \sum_{\Lambda_{\text{opt}}} b_\lambda\, \boldsymbol{\varphi}_\lambda,$$

where $\Lambda_{\text{opt}} \subset \Omega$ is an index set of size m. Without loss of generality, assume that the atoms in Λ_{opt} are linearly independent and that the coefficients b_λ are nonzero. Otherwise, the signal has an exact representation using fewer than m atoms.

From the dictionary synthesis matrix, extract the $d \times m$ matrix $\boldsymbol{\Phi}_{\text{opt}}$ whose columns are the atoms listed in Λ_{opt}:

$$\boldsymbol{\Phi}_{\text{opt}} \stackrel{\text{def}}{=} \begin{bmatrix} \boldsymbol{\varphi}_{\lambda_1} & \boldsymbol{\varphi}_{\lambda_2} & \dots & \boldsymbol{\varphi}_{\lambda_m} \end{bmatrix},$$

where λ_k ranges over Λ_{opt}. Note that $\boldsymbol{\Phi}_{\text{opt}}$ is nonsingular because its columns form a linearly independent set. The signal can now be expressed as

$$\boldsymbol{s} = \boldsymbol{\Phi}_{\text{opt}}\, \boldsymbol{b}_{\text{opt}}$$

for a vector $\boldsymbol{b}_{\text{opt}}$ of m complex coefficients, which vector formally belongs to $\mathbb{C}^{\Lambda_{\text{opt}}}$.

A few other preliminaries remain. It is sometimes necessary to extend a short coefficient vector with zeros so that it lies in $\mathbb{C}^\Omega$. We indicate this operation with a prime mark ($'$). For example, we might extend the m-dimensional vector $\boldsymbol{b}_{\text{opt}}$ to the N-dimensional vector $\boldsymbol{b}'_{\text{opt}}$ whose nonzero entries all lie at coordinates indexed by Λ_{opt}. Finally, we require a precise definition of the signum function.

Definition 1 (Signum Function): Applied to a complex number, the signum function $\operatorname{sgn}(\cdot)$ returns the unimodular part of that number, i.e.,

$$\operatorname{sgn}(r\,\mathrm{e}^{\mathrm{i}\,\theta}) = \begin{cases} \mathrm{e}^{\mathrm{i}\,\theta} & \text{when } r > 0 \text{ and} \\ 0 & \text{when } r = 0. \end{cases}$$

We extend the signum function to vectors by applying it to each component separately.

For the case of a real dictionary in a real vector space, Fuchs has developed a condition under which the unique solution to the Basis Pursuit problem is $\boldsymbol{b}'_{\text{opt}}$.

Theorem 2 (Fuchs [7]): Suppose that the sparsest representation of a real vector is $\boldsymbol{\Phi}_{\text{opt}}\,\boldsymbol{b}_{\text{opt}}$. If there exists a vector $\boldsymbol{h}$ in $\mathbb{R}^d$ at which

1) $\boldsymbol{\Phi}_{\text{opt}}^T\,\boldsymbol{h} = \operatorname{sgn}(\boldsymbol{b}_{\text{opt}})$ and
2) $|\langle \boldsymbol{h}, \boldsymbol{\varphi}_\omega \rangle| < 1$ for each ω not listed in Λ_{opt},

then the (unique) solution to the ℓ_1 minimization problem (P_1) is $\boldsymbol{b}'_{\text{opt}}$, which coincides with the (unique) solution to the sparse approximation problem (P_0).

It is somewhat difficult to interpret the hypotheses of this theorem, and there is no known method for checking them directly. We may obtain a more intuitive corollary by choosing a natural value for the auxiliary vector $\boldsymbol{h}$. From the subspace of vectors that satisfy Condition (1) of Theorem 2, select the one with minimal ℓ_2 norm, namely $\boldsymbol{h} = (\boldsymbol{\Phi}_{\mathrm{opt}}^{\dagger})^T (\operatorname{sgn} \boldsymbol{b}_{\mathrm{opt}})$. We have used the dagger to represent the Moore–Penrose pseudo-inverse, which is defined for full-column-rank matrices by the formula $\boldsymbol{\Phi}_{\mathrm{opt}}^{\dagger} = (\boldsymbol{\Phi}_{\mathrm{opt}}^{*} \boldsymbol{\Phi}_{\mathrm{opt}})^{-1} \boldsymbol{\Phi}_{\mathrm{opt}}^{T}$.

Corollary 3 (Fuchs [7]): Suppose that the sparsest representation of a real vector is $\boldsymbol{\Phi}_{\mathrm{opt}} \boldsymbol{b}_{\mathrm{opt}}$. If it happens that

$$\left| \left\langle (\boldsymbol{\Phi}_{\mathrm{opt}}^{\dagger})^T (\operatorname{sgn} \boldsymbol{b}_{\mathrm{opt}}), \boldsymbol{\varphi}_{\omega} \right\rangle \right| < 1 \quad \text{for every } \omega \text{ not contained in } \Lambda_{\mathrm{opt}}, \tag{1}$$

then the solution to the ℓ_1 minimization problem (P_1) is $\boldsymbol{b}'_{\mathrm{opt}}$.

At first sight, Condition (1) may look just as confusing as Conditions (1) and (2) of Theorem 2. It will become more clear, perhaps, upon inspection. The presence of the pseudo-inverse shows that the conditioning of the optimal synthesis matrix plays a major role in how well ℓ_1 minimization can recover the synthesis coefficients: Basis Pursuit works best when the set of optimal atoms is more or less orthogonal. It is also important that the nonoptimal atoms are significantly different from the optimal atoms. Condition (1) also shows that the signs of the coefficients significantly affect the performance of the method. If we choose the worst possible disbursement of signs, then we obtain a third condition.

Corollary 4 (Tropp [1]): Suppose that the sparsest representation of a real vector is $\boldsymbol{\Phi}_{\mathrm{opt}} \boldsymbol{b}_{\mathrm{opt}}$. The condition

$$\left\| \boldsymbol{\Phi}_{\mathrm{opt}}^{\dagger} \boldsymbol{\varphi}_{\omega} \right\|_1 < 1 \quad \text{for every } \omega \text{ not contained in } \Lambda_{\mathrm{opt}} \tag{2}$$

implies that the unique solution to the ℓ_1 minimization problem (P_1) is $\boldsymbol{b}'_{\mathrm{opt}}$.

In fact, the proof of [1] establishes this condition in the complex setting. The same article demonstrates that (2) can guarantee the success of another algorithm, Orthogonal Matching Pursuit. Moreover, it offers techniques for checking the condition. Recently, Gribonval and Vandergheynst have proven that a third algorithm, Matching Pursuit, also performs well when Condition (2) is in force [8].

III. Generalization of Fuchs' Theorem

We may reach a complex version of Theorem 2 by modifying the proof of Corollary 4 that appears in [1].

Theorem 5: Suppose that the sparsest representation of a complex vector is $\boldsymbol{\Phi}_{\mathrm{opt}} \boldsymbol{b}_{\mathrm{opt}}$. If there exists a vector $\boldsymbol{h}$ in $\mathbb{C}^d$ at which

1) $\boldsymbol{\Phi}_{\mathrm{opt}}^{*} \boldsymbol{h} = \operatorname{sgn}(\boldsymbol{b}_{\mathrm{opt}})$ and
2) $|\langle \boldsymbol{h}, \boldsymbol{\varphi}_{\omega} \rangle| < 1$ for each ω not listed in Λ_{opt},

then the (unique) solution to the ℓ_1 minimization problem (P_1) is $\boldsymbol{b}'_{\mathrm{opt}}$, which coincides with the (unique) solution to the sparse approximation problem (P_0).

Note that we have started using the conjugate transpose symbol * instead of the transpose symbol ^{T} because we have moved to the complex setting. Our proof requires a simple lemma.

Lemma 6: Suppose that $\boldsymbol{z}$ is a vector whose components are all nonzero and that $\boldsymbol{v}$ is a vector whose entries do not have identical moduli. Then $|\langle \boldsymbol{z}, \boldsymbol{v} \rangle| < \|\boldsymbol{z}\|_1 \, \|\boldsymbol{v}\|_{\infty}$.

The lemma is straightforward to establish, so we continue with the proof of the theorem.

Proof: Suppose that $\boldsymbol{s}$ is a signal whose sparsest representation is $\boldsymbol{\Phi}_{\mathrm{opt}} \boldsymbol{b}_{\mathrm{opt}}$. Say that the vector $\boldsymbol{b}_{\mathrm{opt}}$ has m components (all nonzero), and let Λ_{opt} index these components. Assume too that there exists a vector $\boldsymbol{h}$ in $\mathbb{C}^d$ at which

1) $\boldsymbol{\Phi}_{\mathrm{opt}}^{*} \boldsymbol{h} = \operatorname{sgn}(\boldsymbol{b}_{\mathrm{opt}})$ and
2) $|\langle \boldsymbol{h}, \boldsymbol{\varphi}_{\omega} \rangle| < 1$ for each ω not listed in Λ_{opt}.

Let $\boldsymbol{s} = \boldsymbol{\Phi}_{\mathrm{alt}} \boldsymbol{b}_{\mathrm{alt}}$ be a different representation of the signal. We may suppose that its components are all nonzero and that they are indexed by Λ_{alt}. It must be shown that the ℓ_1 norm of the extended coefficient vector $\boldsymbol{b}'_{\mathrm{opt}}$ is strictly less than the ℓ_1 norm of the extended coefficient vector $\boldsymbol{b}'_{\mathrm{alt}}$. We begin with a calculation that should explain itself.

$$\begin{aligned}
\left\| \boldsymbol{b}'_{\mathrm{opt}} \right\|_1 &= \|\boldsymbol{b}_{\mathrm{opt}}\|_1 \\
&= |(\operatorname{sgn} \boldsymbol{b}_{\mathrm{opt}})^* \, \boldsymbol{b}_{\mathrm{opt}}| \\
&= |(\boldsymbol{h}^* \, \boldsymbol{\Phi}_{\mathrm{opt}}) \, \boldsymbol{b}_{\mathrm{opt}}| \\
&= |\boldsymbol{h}^* \, \boldsymbol{s}| \\
&= |\boldsymbol{h}^* \, (\boldsymbol{\Phi}_{\mathrm{alt}} \, \boldsymbol{b}_{\mathrm{alt}})| \\
&= |\langle \boldsymbol{b}_{\mathrm{alt}}, \boldsymbol{\Phi}_{\mathrm{alt}}^{*} \, \boldsymbol{h} \rangle| \,.
\end{aligned}$$

Now assume that the vector $\boldsymbol{\Phi}_{\mathrm{alt}}^{*} \boldsymbol{h}$ has components whose moduli are not identical. By assumption, $\boldsymbol{b}_{\mathrm{alt}}$ has no zero entries, so we may apply the lemma. Hence

$$\begin{aligned}
\left\| \boldsymbol{b}'_{\mathrm{opt}} \right\|_1 &< \|\boldsymbol{b}_{\mathrm{alt}}\|_1 \, \|\boldsymbol{\Phi}_{\mathrm{alt}}^{*} \boldsymbol{h}\|_{\infty} \\
&= \|\boldsymbol{b}_{\mathrm{alt}}\|_1 \, \max_{\lambda \in \Lambda_{\mathrm{alt}}} |\langle \boldsymbol{h}, \boldsymbol{\varphi}_{\lambda} \rangle| \\
&\le \|\boldsymbol{b}_{\mathrm{alt}}\|_1 \\
&= \left\| \boldsymbol{b}'_{\mathrm{alt}} \right\|_1 .
\end{aligned}$$

The second inequality holds because the conditions we have placed on $\boldsymbol{h}$ imply that $|\langle \boldsymbol{h}, \boldsymbol{\varphi}_{\omega} \rangle| \le 1$ for every ω in Ω.

On the contrary, suppose that each component of the vector $\boldsymbol{\Phi}_{\mathrm{alt}}^{*} \boldsymbol{h}$ has the same modulus. As noted in Section II, the matrix $\boldsymbol{\Phi}_{\mathrm{opt}}$ is nonsingular, so $\boldsymbol{\Phi}_{\mathrm{opt}} \boldsymbol{b}_{\mathrm{opt}}$ is the unique representation of $\boldsymbol{s}$ using the vectors in Λ_{opt}. Moreover, Λ_{opt} is the smallest possible index set whose atoms can represent $\boldsymbol{s}$. Thus Λ_{alt} contains at least one index, say λ_0, that is not contained in Λ_{opt}. By assumption, the number $|\langle \boldsymbol{h}, \boldsymbol{\varphi}_{\lambda_0} \rangle|$ is strictly less than one. We may identify this number as a component of $\boldsymbol{\Phi}_{\mathrm{alt}}^{*} \boldsymbol{h}$. In consequence, *every* component of the vector $\boldsymbol{\Phi}_{\mathrm{alt}}^{*} \boldsymbol{h}$ has modulus less than one. Therefore, we may calculate that

$$\begin{aligned}
\left\| \boldsymbol{b}'_{\mathrm{opt}} \right\|_1 &\le \|\boldsymbol{b}_{\mathrm{alt}}\|_1 \, \|\boldsymbol{\Phi}_{\mathrm{alt}}^{*} \boldsymbol{h}\|_{\infty} \\
&< \|\boldsymbol{b}_{\mathrm{alt}}\|_1 \\
&= \left\| \boldsymbol{b}'_{\mathrm{alt}} \right\|_1 .
\end{aligned}$$

In words, any set of nonoptimal coefficients for representing the signal has strictly larger ℓ_1 norm than the optimal coefficients. We conclude that Basis Pursuit must recover these optimal coefficients. Finally, suppose that the hypotheses of the theorem hold, while the sparse approximation problem (P_0) has two distinct solutions. The preceding argument shows that each one would have a strictly smaller ℓ_1 norm than the other, a *reductio ad absurdum*. ■

A complex version of Corollary 3 follows immediately.

Corollary 7: Suppose that the sparsest representation of a complex vector is $\boldsymbol{\Phi}_{\text{opt}}\, \boldsymbol{b}_{\text{opt}}$. If it happens that

$$\left| \left\langle (\boldsymbol{\Phi}_{\text{opt}}^{\dagger})^{*} \, (\operatorname{sgn} \boldsymbol{b}_{\text{opt}}), \boldsymbol{\varphi}_{\omega} \right\rangle \right| < 1 \qquad \text{for every } \omega \text{ not listed in } \Lambda_{\text{opt}},$$

then the unique solution to the ℓ_1 minimization problem (P_1) is $\boldsymbol{b}'_{\text{opt}}$.

Remark 8: One of the anonymous referees outlined another proof of Theorem 5 via classical duality theory. The dual of (P_1) is

$$\max_{\boldsymbol{u}} \; \operatorname{Re} \langle \boldsymbol{s}, \boldsymbol{u} \rangle \qquad \text{subject to} \qquad \| \boldsymbol{\Phi}^{*} \, \boldsymbol{u} \|_{\infty} \leq 1.$$

If a coefficient vector $\boldsymbol{b}$ is feasible for (P_1), then $\operatorname{Re} \langle \boldsymbol{s}, \boldsymbol{u} \rangle \leq \| \boldsymbol{b} \|_1$ for every dual-feasible $\boldsymbol{u}$. Strong duality implies that $\boldsymbol{b}'_{\text{opt}}$ is a minimizer of (P_1) if and only if we can identify a dual-feasible $\boldsymbol{u}$ for which $\operatorname{Re} \langle \boldsymbol{s}, \boldsymbol{u} \rangle = \left\| \boldsymbol{b}'_{\text{opt}} \right\|_1$. Suppose that there exists a vector $\boldsymbol{h}$ that meets Conditions (1) and (2) of Theorem 5. It is clear that this vector $\boldsymbol{h}$ is dual feasible, and furthermore

$$\begin{aligned} \operatorname{Re} \langle \boldsymbol{s}, \boldsymbol{h} \rangle &= \operatorname{Re} \left\langle \boldsymbol{\Phi} \, \boldsymbol{b}'_{\text{opt}}, \boldsymbol{h} \right\rangle \\ &= \operatorname{Re} \left\langle \boldsymbol{b}'_{\text{opt}}, \boldsymbol{\Phi}^{*} \, \boldsymbol{h} \right\rangle \\ &= \operatorname{Re} \left\langle \boldsymbol{b}'_{\text{opt}}, \operatorname{sgn} \boldsymbol{b}'_{\text{opt}} \right\rangle \\ &= \left\| \boldsymbol{b}'_{\text{opt}} \right\|_1 . \end{aligned}$$

To see that $\boldsymbol{b}'_{\text{opt}}$ uniquely solves (P_1), observe that the third equality can hold only if the support of $\boldsymbol{b}_{\text{opt}}$ equals Λ_{opt}.

Acknowledgments

The author wishes to thank both anonymous referees for their insightful remarks.

References

[1] J. A. Tropp, "Greed is good: Algorithmic results for sparse approximation," *IEEE Trans. Inform. Theory*, vol. 50, no. 10, pp. 2231–2242, Oct. 2004.

[2] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by Basis Pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1999, electronic.

[3] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inform. Theory*, vol. 47, pp. 2845–2862, Nov. 2001.

[4] M. Elad and A. M. Bruckstein, "A generalized uncertainty principle and sparse representation in pairs of bases," *IEEE Trans. Inform. Theory*, vol. 48, no. 9, pp. 2558–2567, 2002.

[5] D. L. Donoho and M. Elad, "Maximal sparsity representation via ℓ_1 minimization," *Proc. Natl. Acad. Sci.*, vol. 100, pp. 2197–2202, March 2003.

[6] R. Gribonval and M. Nielsen, "Sparse representations in unions of bases," *IEEE Trans. Inform. Theory*, vol. 49, no. 12, pp. 3320–3325, Dec. 2003.

[7] J.-J. Fuchs, "On sparse representations in arbitrary redundant bases," *IEEE Trans. Inform. Th.*, vol. 50, no. 6, pp. 1341–1344, June 2004.

[8] R. Gribonval and M. Nielsen, "On the exponential converge of matching pursuits in quasi-incoherent dictionaries," Université de Rennes I, IRISA Report 1619, 2004.

Joel A. Tropp (S'03) earned the B.A. degree in Plan II Liberal Arts Honors and the B.S. degree in Mathematics from the University of Texas at Austin (UT-Austin) in May 1999, graduating *magna cum laude*. He continued his graduate studies in Computational and Applied Mathematics (CAM) at UT-Austin where he completed the M.S. degree in May 2001 and the Ph.D. degree in August 2004.

Dr. Tropp is presently an Assistant Professor in the Mathematics Department of the University of Michigan at Ann Arbor.

As an undergraduate, Dr. Tropp was a recipient of the Barry M. Goldwater National Science Scholarship and a semi-finalist for the British Marshall. He won a CAM Graduate Fellowship in 1999 and a National Science Foundation Graduate Fellowship in 2000.