Recovery of Short, Complex Linear Combinations Via ℓ_1 Minimization

Joel A. Tropp, *Student 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 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 φ_ω, 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

$$\mathbf{\Phi} \stackrel{\text{def}}{=} [\, \varphi_{\omega_1} \quad \varphi_{\omega_2} \quad \ldots \quad \varphi_{\omega_N} \,].$$

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

$$\min_{\boldsymbol{b}\in\mathbb{C}^d} \|\boldsymbol{b}\|_0 \quad \text{subject to} \quad \mathbf{\Phi}\boldsymbol{b} = \boldsymbol{s}. \tag{1}$$

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, Proposition 4.1]. Nevertheless, the question has value for the insight it can provide on more difficult sparse approximation problems.

One approach to solving (1) is to replace the horribly nonlinear function $\|\cdot\|_0$ with the norm $\|\cdot\|_1$ and hope that the solutions coincide. That is,

$$\min_{\boldsymbol{b}\in\mathbb{C}^d} \|\boldsymbol{b}\|_1 \quad \text{subject to} \quad \mathbf{\Phi}\boldsymbol{b} = \boldsymbol{s}. \tag{2}$$

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 (1).

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 correspondence provides a completely different method of reaching Fuchs' result.

Manuscript received June 9, 2004; revised January 12, 2005. This work was supported by a National Science Foundation Graduate Fellowship.

The author was with was with ICES, University of Texas at Austin, Austin, TX 78712 USA. He is now with the Department of Mathematics, University of Michigan, Ann Arbor, MI 48109 USA (e-mail: jtropp@umich.edu).

Communicated by V. A. Vaishampayan, Associate Editor At Large.
Digital Object Identifier 10.1109/TIT.2005.844057

II. Fuchs' Condition

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

$$\boldsymbol{s} = \sum_{\Lambda_{\text{opt}}} b_\lambda \varphi_\lambda$$

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 $\mathbf{\Phi}_{\text{opt}}$ whose columns are the atoms listed in Λ_{opt}

$$\mathbf{\Phi}_{\text{opt}} \stackrel{\text{def}}{=} [\, \varphi_{\lambda_1} \quad \varphi_{\lambda_2} \quad \ldots \quad \varphi_{\lambda_m} \,]$$

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

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

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 $\mathbf{\Phi}_{\text{opt}}\boldsymbol{b}_{\text{opt}}$. If there exists a vector $\boldsymbol{h}$ in $\mathbb{R}^d$ at which

1) $\mathbf{\Phi}_{\text{opt}}^T \boldsymbol{h} = \operatorname{sgn}(\boldsymbol{b}_{\text{opt}})$ and
2) $|\langle \boldsymbol{h}, \varphi_\omega \rangle| < 1$ for each $\omega \notin \Lambda_{\text{opt}}$

then the (unique) solution to the ℓ_1 minimization problem (2) is $\boldsymbol{b}'_{\text{opt}}$, which coincides with the (unique) solution to the sparse approximation problem (1).

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 $\boldsymbol{h} = (\mathbf{\Phi}_{\text{opt}}^\dagger)^T (\operatorname{sgn} \boldsymbol{b}_{\text{opt}})$. We have used the dagger to represent the

Moore–Penrose pseudoinverse, which is defined for full-column-rank matrices by the formula $\mathbf{\Phi}_{\rm opt}^{\dagger} = (\mathbf{\Phi}_{\rm opt}^{*}\mathbf{\Phi}_{\rm opt})^{-1}\mathbf{\Phi}_{\rm opt}^{T}$.

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

$$\left|\left\langle (\mathbf{\Phi}_{\rm opt}^{\dagger})^{T}(\operatorname{sgn} \boldsymbol{b}_{\rm opt}), \varphi_{\omega}\right\rangle\right| < 1, \text{ for every } \omega \notin \Lambda_{\rm opt} \tag{3}$$

then the unique solution to the ℓ_1 minimization problem (2) is $\boldsymbol{b}'_{\rm opt}$.

At first sight, (3) 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 pseudoinverse 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 (3) 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 $\mathbf{\Phi}_{\rm opt}\boldsymbol{b}_{\rm opt}$. The condition

$$\left\|\mathbf{\Phi}_{\rm opt}^{\dagger}\boldsymbol{\varphi}_{\omega}\right\|_{1} < 1, \quad \text{for every } \omega \notin \Lambda_{\rm opt} \tag{4}$$

implies that the unique solution to the ℓ_1 minimization problem (2) is $\boldsymbol{b}'_{\rm opt}$.

In fact, the proof of [1] establishes this condition in the complex setting. The same article demonstrates that (4) 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 (4) 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 $\mathbf{\Phi}_{\rm opt}\boldsymbol{b}_{\rm opt}$. If there exists a vector $\boldsymbol{h}$ in $\mathbb{C}^d$ at which

1) $\mathbf{\Phi}_{\rm opt}^{*}\boldsymbol{h} = \operatorname{sgn}(\boldsymbol{b}_{\rm opt})$ and
2) $|\langle \boldsymbol{h}, \varphi_{\omega}\rangle| < 1$ for each $\omega \notin \Lambda_{\rm opt}$

then the (unique) solution to the ℓ_1 minimization problem (2) is $\boldsymbol{b}'_{\rm opt}$, which coincides with the (unique) solution to the sparse approximation problem (1).

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 $\mathbf{\Phi}_{\rm opt}\boldsymbol{b}_{\rm opt}$. Say that the vector $\boldsymbol{b}_{\rm opt}$ has m components (all nonzero), and let $\Lambda_{\rm opt}$ index these components. Assume too that there exists a vector $\boldsymbol{h}$ in $\mathbb{C}^d$ at which

1) $\mathbf{\Phi}_{\rm opt}^{*}\boldsymbol{h} = \operatorname{sgn}(\boldsymbol{b}_{\rm opt})$ and
2) $|\langle \boldsymbol{h}, \varphi_{\omega}\rangle| < 1$ for each $\omega \notin \Lambda_{\rm opt}$.

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

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

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

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

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

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

$$\begin{aligned}
\|\boldsymbol{b}'_{\rm opt}\|_1 &\le \|\boldsymbol{b}_{\rm alt}\|_1 \; \|\mathbf{\Phi}_{\rm alt}^{*}\boldsymbol{h}\|_\infty \\
&< \|\boldsymbol{b}_{\rm alt}\|_1 \\
&= \|\boldsymbol{b}'_{\rm alt}\|_1.
\end{aligned}$$

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 (1) 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 $\mathbf{\Phi}_{\rm opt}\boldsymbol{b}_{\rm opt}$. If it happens that

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

then the unique solution to the ℓ_1 minimization problem (2) is $\boldsymbol{b}'_{\rm opt}$.

Remark 8: One of the anonymous referees outlined another proof of Theorem 5 via classical duality theory. The dual of (2) is

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

If a coefficient vector $\boldsymbol{b}$ is feasible for (2), then $\operatorname{Re}\langle \boldsymbol{s}, \boldsymbol{u}\rangle \le \|\boldsymbol{b}\|_1$ for every dual-feasible $\boldsymbol{u}$. Strong duality implies that $\boldsymbol{b}'_{\rm opt}$ is a minimizer of (2) if and only if we can identify a dual-feasible $\boldsymbol{u}$ for which $\operatorname{Re}\langle \boldsymbol{s}, \boldsymbol{u}\rangle =$

$\|\boldsymbol{b}'_{\rm opt}\|_1$. 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

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

To see that $\boldsymbol{b}'_{\rm opt}$ uniquely solves (2), observe that the third equality can hold only if the support of $\boldsymbol{b}_{\rm opt}$ equals $\Lambda_{\rm opt}$.

ACKNOWLEDGMENT

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. Inf. 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. Comput.*, vol. 20, no. 1, pp. 33–61, 1999.

[3] D. L. Donoho and X. Huo, "Uncertainty principles and ideal atomic decomposition," *IEEE Trans. Inf. Theory*, vol. 47, no. 7, 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. Inf. Theory*, vol. 48, no. 9, pp. 2558–2567, Sep. 2002.

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

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

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

[8] R. Gribonval and M. Nielsen, "On the Exponential Convergence of Matching Pursuits in Quasi-Incoherent Dictionaries," Université de Rennes I, Rennes, France, IRISA Rep. 1619, 2004.

Sum Power Iterative Water-Filling for Multi-Antenna Gaussian Broadcast Channels

Nihar Jindal, *Member, IEEE*, Wonjong Rhee, *Member, IEEE*, Sriram Vishwanath, *Member, IEEE*, Syed Ali Jafar, *Member, IEEE*, and Andrea Goldsmith, *Fellow, IEEE*

***Abstract*—In this correspondence, we consider the problem of maximizing sum rate of a multiple-antenna Gaussian broadcast channel (BC). It was recently found that dirty-paper coding is capacity achieving for this channel. In order to achieve capacity, the optimal transmission policy (i.e., the optimal transmit covariance structure) given the channel conditions and power constraint must be found. However, obtaining the optimal transmission policy when employing dirty-paper coding is a computationally complex nonconvex problem. We use duality to transform this problem into a well-structured convex multiple-access channel (MAC) problem. We exploit the structure of this problem and derive simple and fast iterative algorithms that provide the optimum transmission policies for the MAC, which can easily be mapped to the optimal BC policies.**

***Index Terms*—Broadcast channel, dirty-paper coding, duality, multiple-access channel (MAC), multiple-input multiple-output (MIMO), systems.**

I. INTRODUCTION

In recent years, there has been great interest in characterizing and computing the capacity region of multiple-antenna broadcast (downlink) channels. An achievable region for the multiple-antenna downlink channel was found in [3], and this achievable region was shown to achieve the sum rate capacity in [3], [10], [12], [16], and was more recently shown to achieve the full capacity region in [14]. Though these results show that the general dirty-paper coding strategy is optimal, one must still optimize over the transmit covariance structure (i.e., how transmissions over different antennas should be correlated) in order to determine the optimal transmission policy and the corresponding sum rate capacity. Unlike the single-antenna broadcast channel (BC), sum capacity is not in general achieved by transmitting to a single user. Thus, the problem cannot be reduced to a point-to-point multiple-input multiple-output (MIMO) problem, for which simple expressions are known. Furthermore, the direct optimization for sum rate capacity is a computationally complex

Manuscript received July 21, 2004; revised December 15, 2004. The work of some of the authors was supported by the Stanford Networking Research Center. The material in this correspondence was presented in part at the International Symposium on Information Theory, Yokohama, Japan, June/July 2003, and at the Asilomar Conference on Signals, Systems, and Computers, Asilomar, CA , Nov. 2002. This work was initiated while all the authors were at Stanford University.

N. Jindal is with the Department of Electrical and Computer Engineering, University of Minnesota, Minneapolis, MN 55455 USA (e-mail: nihar@ece.umn.edu).

W. Rhee is with the ASSIA, Inc., Redwood City, CA 94065 USA (e-mail: wonjong@dsl.stanford.edu).

S. Vishwanath is with the Department of Electrical and Computer Engineering, University of Texas, Austin, Austin, TX 78712 USA (e-mail: sriram@ece.utexas.edu).

S. A. Jafar is with Electronic Engineering and Computer Science, University of California, Irvine, Irvine, CA 92697-2625 USA (e-mail: syed@ece.uci.edu)

A. Goldsmith is with the Department of Electrical Engineering, Stanford University, Stanford, CA 94305-9515 USA (e-mail: andrea@systems.stanford.edu).

Communicated by M. Medard, Associate Editor for Communications.

Digital Object Identifier 10.1109/TIT.2005.844082