Corrigendum in "Just Relax: Convex Programming Methods for Identifying Sparse Signals in Noise"

Joel A. Tropp, *Member, IEEE*

***Abstract*— This note closes a gap in the proof of the main lemma from the paper "Just Relax: Convex Programming Methods for Identifying Sparse Signals in Noise."**

***Index Terms*— Algorithms, approximation methods, Basis Pursuit, convex program, optimization methods, linear regression, Orthogonal Matching Pursuit, sparse representations**

I. Introduction

The article [1] studies the minimizers of the nonsmooth convex function

$$L(\boldsymbol{b}) = \frac{1}{2}\left\|\boldsymbol{s} - \boldsymbol{\Phi}\boldsymbol{b}\right\|_2^2 + \gamma\left\|\boldsymbol{b}\right\|_1, \tag{L}$$

which plays an important role in sparse approximation and compressive sampling. The key result [1, Lem. 6] is a sufficient condition for the support of the minimizer to be contained within a specified index set. This lemma can then be used to study circumstances in which minimizers of (L) correctly identify the support of a sparse signal $\boldsymbol{s}$ contaminated with noise.

It has come to the author's attention that the proof of [1, Lem. 6] relies on a claim that holds only for differentiable convex functions. Although the argument requires some amplification, the lemma is true as originally stated. This note provides a complete proof that corrects the error.

A. Notation

All notation is recycled from [1], but we repeat the essential pieces for the convenience of the reader. As usual, $\left\|\cdot\right\|_p$ denotes the ℓ_p vector norm with respect to the standard basis. The angle bracket $\langle\cdot,\ \cdot\rangle$ represents the Hermitian inner product, which is linear in the first variable and conjugate-linear in the second variable.

Let Ω be an index set, and consider the linear space $\mathbb{C}^\Omega$ of complex-valued vectors indexed by Ω. The standard basis for $\mathbb{C}^\Omega$ is the family $\{\mathbf{e}_\omega : \omega \in \Omega\}$, where the vector $\mathbf{e}_\omega$ equals one in the component ω and zero in the remaining components.

We study signals that lie in the space $\mathbb{C}^d$. Consider a family of vectors $\{\boldsymbol{\varphi}_\omega : \omega \in \Omega\} \subset \mathbb{C}^d$, and form a matrix $\boldsymbol{\Phi}$ using these vectors as columns. The matrix maps a vector of coefficients in $\mathbb{C}^\Omega$ into a signal by the rule

$$\boldsymbol{\Phi}\boldsymbol{c} = \sum\nolimits_{\omega\in\Omega} c_\omega \boldsymbol{\varphi}_\omega.$$

The adjoint maps a signal into a coefficient vector by the rule

$$(\boldsymbol{\Phi}^*\boldsymbol{s})(\omega) = \langle \boldsymbol{s},\ \boldsymbol{\varphi}_\omega\rangle.$$

Given a subset Λ of Ω, we write $\boldsymbol{\Phi}_\Lambda$ for the submatrix of $\boldsymbol{\Phi}$ whose columns are listed in Λ. When $\boldsymbol{\Phi}_\Lambda$ has full column rank, the pseudoinverse

$$\boldsymbol{\Phi}_\Lambda^\dagger = (\boldsymbol{\Phi}_\Lambda^*\boldsymbol{\Phi}_\Lambda)^{-1}\boldsymbol{\Phi}_\Lambda^*.$$

For a fixed signal $\boldsymbol{s}$, we define a coefficient vector $\boldsymbol{c}_\Lambda = \boldsymbol{\Phi}_\Lambda^\dagger \boldsymbol{s}$ and a signal approximation $\boldsymbol{a}_\Lambda = \boldsymbol{\Phi}_\Lambda \boldsymbol{c}_\Lambda$. This approximation $\boldsymbol{a}_\Lambda$ can be seen as the orthogonal projection of $\boldsymbol{s}$ onto the range of $\boldsymbol{\Phi}_\Lambda$.

Date: 4 September 2008. Typos corrected 6 September 2008.

J. A. Tropp is with Applied and Computational Mathematics, California Institute of Technology, Pasadena, CA 91125-5000 USA. E-mail: jtropp@acm.caltech.edu. Supported in part by ONR Award N00014-08-1-0883.

Finally, for a convex function f we write $\partial f(\boldsymbol{x})$ for the subdifferential of f at the point $\boldsymbol{x}$.

B. The Correlation Condition

To establish [1, Lem. 6], the first step is to study minimizers of the function (L) that are restricted to have fixed support. We state the result without proof, referring the reader to [1, Lem. 5].

Lemma 1 (Restricted Minimizers): Suppose that $\boldsymbol{\Phi}_\Lambda$ has full column rank, and let $\boldsymbol{b}_\star$ minimize the objective function (L) over all coefficient vectors supported on Λ. A necessary and sufficient condition on such a minimizer is that

$$\boldsymbol{c}_\Lambda - \boldsymbol{b}_\star = \gamma(\boldsymbol{\Phi}_\Lambda^*\boldsymbol{\Phi}_\Lambda)^{-1}\boldsymbol{g} \tag{1}$$

where the vector $\boldsymbol{g}$ is drawn from $\partial\left\|\boldsymbol{b}_\star\right\|_1$. Moreover, the minimizer is unique.

The main result provides a sufficient condition under which the restricted minimizer is also the global minimizer of the objective function.

Lemma 2 (Correlation Condition): Suppose that $\boldsymbol{\Phi}_\Lambda$ has full column rank, and let $\boldsymbol{b}_\star$ minimize the function (L) over all coefficient vectors supported on Λ. Suppose that

$$\left\|\boldsymbol{\Phi}^*(\boldsymbol{s} - \boldsymbol{a}_\Lambda)\right\|_\infty < \gamma\left[1 - \max_{\omega\notin\Lambda}\left|\left\langle \boldsymbol{\Phi}_\Lambda^\dagger\boldsymbol{\varphi}_\omega,\ \boldsymbol{g}\right\rangle\right|\right]$$

where $\boldsymbol{g} \in \partial\left\|\boldsymbol{b}_\star\right\|_1$ is determined by (1). It follows that $\boldsymbol{b}_\star$ is the unique global minimizer of (L).

Together, these two lemmata provide detailed information about the performance of convex programming methods for sparse approximation, as discussed in [1].

II. Proof of Lemma 2

Let $\boldsymbol{b}_\star$ be the unique minimizer of (L) over coefficient vectors supported on Λ. We develop a sufficient condition under which

$$L(\boldsymbol{b}_\star + \boldsymbol{h}) - L(\boldsymbol{b}_\star) > 0$$

whenever the norm of the perturbation $\boldsymbol{h}$ is small enough. Since the objective function (L) is convex, it follows that $\boldsymbol{b}_\star$ is the unique global minimizer.

Each perturbation admits a unique decomposition

$$\boldsymbol{h} = \boldsymbol{u} + \boldsymbol{v}$$

where $\operatorname{supp}(\boldsymbol{u}) \subset \Lambda$ and $\operatorname{supp}(\boldsymbol{v}) \subset \Lambda^c$. Without loss of generality, we may pose some additional constraints. First, we take $\boldsymbol{v} \neq \mathbf{0}$, since Lemma 1 already addresses the complementary case. We also instate the bound $\left\|\boldsymbol{u}\right\|_\infty \le \delta$ for a small, positive number δ, which reflects the requirement that the perturbation is tiny.

To begin the calculation, write the perturbed objective function as

$$L(\boldsymbol{b}_\star + \boldsymbol{h}) = \frac{1}{2}\left\|\boldsymbol{s} - \boldsymbol{\Phi}(\boldsymbol{b}_\star + \boldsymbol{u}) - \boldsymbol{\Phi}\boldsymbol{v}\right\|_2^2 + \gamma\left\|(\boldsymbol{b}_\star + \boldsymbol{u}) + \boldsymbol{v}\right\|_1.$$

Expand the ℓ_2 norm to obtain

$$\begin{aligned}\left\|\boldsymbol{s} - \boldsymbol{\Phi}(\boldsymbol{b}_\star + \boldsymbol{u}) - \boldsymbol{\Phi}\boldsymbol{v}\right\|_2^2 &\\ = \left\|\boldsymbol{s} - \boldsymbol{\Phi}(\boldsymbol{b}_\star + \boldsymbol{u})\right\|_2^2 &+ \left\|\boldsymbol{\Phi}\boldsymbol{v}\right\|_2^2\\ - 2\operatorname{Re}\langle \boldsymbol{s} - \boldsymbol{\Phi}\boldsymbol{b}_\star,\ \boldsymbol{\Phi}\boldsymbol{v}\rangle &+ 2\operatorname{Re}\langle\boldsymbol{\Phi}\boldsymbol{u},\ \boldsymbol{\Phi}\boldsymbol{v}\rangle.\end{aligned}$$

Since the vectors $\boldsymbol{b}_\star + \boldsymbol{u}$ and $\boldsymbol{v}$ have disjoint support,

$$\|(\boldsymbol{b}_\star + \boldsymbol{u}) + \boldsymbol{v}\|_1 = \|\boldsymbol{b}_\star + \boldsymbol{u}\|_1 + \|\boldsymbol{v}\|_1 .$$

Combine the last three relations, and identify the quantity $L(\boldsymbol{b}_\star + \boldsymbol{u})$ to reach

$$\begin{aligned} L(\boldsymbol{b}_\star + \boldsymbol{h}) - L(\boldsymbol{b}_\star) &= L(\boldsymbol{b}_\star + \boldsymbol{u}) - L(\boldsymbol{b}_\star) \\ &\quad + \frac{1}{2}\|\boldsymbol{\Phi v}\|_2^2 - \operatorname{Re}\langle \boldsymbol{s} - \boldsymbol{\Phi b}_\star,\ \boldsymbol{\Phi v}\rangle \\ &\quad + \operatorname{Re}\langle \boldsymbol{\Phi u},\ \boldsymbol{\Phi v}\rangle + \gamma\|\boldsymbol{v}\|_1 . \qquad (2) \end{aligned}$$

This identity holds for each perturbation $\boldsymbol{h} = \boldsymbol{u} + \boldsymbol{v}$.

The next step is to develop a lower bound on the right-hand side of (2). Lemma 1 states that $\boldsymbol{b}_\star$ minimizes (L) over coefficient vectors supported on Λ. As a result,

$$L(\boldsymbol{b}_\star + \boldsymbol{u}) - L(\boldsymbol{b}_\star) \geq 0.$$

The quadratic term $\|\boldsymbol{\Phi v}\|_2^2$ is also nonnegative, hence

$$\begin{aligned} &L(\boldsymbol{b}_\star + \boldsymbol{h}) - L(\boldsymbol{b}_\star) \\ &\quad \geq \gamma\|\boldsymbol{v}\|_1 - |\langle \boldsymbol{s} - \boldsymbol{\Phi b}_\star,\ \boldsymbol{\Phi v}\rangle| - |\langle \boldsymbol{\Phi u},\ \boldsymbol{\Phi v}\rangle| . \qquad (3) \end{aligned}$$

It is intuitive that the final term, which is quadratic, has smaller order than the other terms, so we will ultimately be able to neglect it.

Let us focus on the second term from the right-hand side of (3). Evidently, we can write

$$\boldsymbol{v} = \left[\sum\nolimits_{\omega\notin\Lambda} \theta_\omega \mathbf{e}_\omega\right] \|\boldsymbol{v}\|_1$$

where $\|\boldsymbol{\theta}\|_1 = 1$. Using this expression, we see that

$$\boldsymbol{\Phi v} = \left[\sum\nolimits_{\omega\notin\Lambda} \theta_\omega \boldsymbol{\varphi}_\omega\right] \|\boldsymbol{v}\|_1 .$$

Invoke the triangle inequality and then Jensen's inequality to obtain

$$\begin{aligned} |\langle \boldsymbol{s} - \boldsymbol{\Phi b}_\star,\ \boldsymbol{\Phi v}\rangle| &\leq \left[\sum\nolimits_{\omega\notin\Lambda} |\theta_\omega|\, |\langle \boldsymbol{s} - \boldsymbol{\Phi b}_\star,\ \boldsymbol{\varphi}_\omega\rangle|\right] \|\boldsymbol{v}\|_1 \\ &\leq \max_{\omega\notin\Lambda} |\langle \boldsymbol{s} - \boldsymbol{\Phi b}_\star,\ \boldsymbol{\varphi}_\omega\rangle| \cdot \|\boldsymbol{v}\|_1 . \end{aligned}$$

To control the third term from the right-hand side of (3), we use standard operator norm bounds. Indeed,

$$\begin{aligned} |\langle \boldsymbol{\Phi u},\ \boldsymbol{\Phi v}\rangle| &= |\langle \boldsymbol{\Phi}^*\boldsymbol{\Phi u},\ \boldsymbol{v}\rangle| \\ &\leq \|\boldsymbol{\Phi}^*\boldsymbol{\Phi u}\|_\infty \|\boldsymbol{v}\|_1 \\ &\leq \delta\|\boldsymbol{\Phi}^*\boldsymbol{\Phi}\|_{\infty,\infty} \|\boldsymbol{v}\|_1 \end{aligned}$$

where we have applied the fact that $\|\boldsymbol{u}\|_\infty \leq \delta$.

Introduce the last two estimates into (3) to discover that

$$\begin{aligned} &L(\boldsymbol{b}_\star + \boldsymbol{h}) - L(\boldsymbol{b}_\star) \\ &\quad \geq \left[\gamma - \max_{\omega\notin\Lambda} |\langle \boldsymbol{s} - \boldsymbol{\Phi b}_\star,\ \boldsymbol{\varphi}_\omega\rangle| - \delta\|\boldsymbol{\Phi}^*\boldsymbol{\Phi}\|_{\infty,\infty}\right] \|\boldsymbol{v}\|_1 . \end{aligned}$$

Since we may select δ as small as we like, the right-hand side is strictly positive for each small perturbation $\boldsymbol{h}$, provided that

$$\gamma - \max_{\omega\notin\Lambda} |\langle \boldsymbol{s} - \boldsymbol{\Phi b}_\star,\ \boldsymbol{\varphi}_\omega\rangle| > 0. \qquad (4)$$

The remaining challenge is to find a more desirable condition which ensures that (4) holds.

To that end, we write

$$\boldsymbol{s} - \boldsymbol{\Phi b}_\star = (\boldsymbol{s} - \boldsymbol{\Phi c}_\Lambda) + \boldsymbol{\Phi}(\boldsymbol{c}_\Lambda - \boldsymbol{b}_\star).$$

By definition, $\boldsymbol{\Phi c}_\Lambda = \boldsymbol{a}_\Lambda$. Invoke the fact that $\boldsymbol{c}_\Lambda - \boldsymbol{b}_\star$ is supported inside Λ along with the characterization from Lemma 1 to see that

$$\boldsymbol{\Phi}(\boldsymbol{c}_\Lambda - \boldsymbol{b}_\star) = \boldsymbol{\Phi}_\Lambda(\boldsymbol{c}_\Lambda - \boldsymbol{b}_\star) = \gamma(\boldsymbol{\Phi}_\Lambda^\dagger)^*\boldsymbol{g}$$

where $\boldsymbol{g} \in \partial\|\boldsymbol{b}_\star\|_1$. Thus, for each index ω,

$$|\langle \boldsymbol{s} - \boldsymbol{\Phi b}_\star,\ \boldsymbol{\varphi}_\omega\rangle| \leq |\langle \boldsymbol{s} - \boldsymbol{a}_\Lambda,\ \boldsymbol{\varphi}_\omega\rangle| + \left|\left\langle \boldsymbol{\Phi}_\Lambda^\dagger\boldsymbol{\varphi}_\omega,\ \boldsymbol{g}\right\rangle\right| .$$

It follows that a sufficient condition for (4) to hold is that

$$\gamma - \max_{\omega\notin\Lambda}\left[\gamma\left|\left\langle \boldsymbol{\Phi}_\Lambda^\dagger\boldsymbol{\varphi}_\omega,\ \boldsymbol{g}\right\rangle\right| + |\langle \boldsymbol{s} - \boldsymbol{a}_\Lambda,\ \boldsymbol{\varphi}_\omega\rangle|\right] > 0.$$

This inequality is in force whenever

$$\max_{\omega\notin\Lambda} |\langle \boldsymbol{s} - \boldsymbol{a}_\Lambda,\ \boldsymbol{\varphi}_\omega\rangle| < \gamma\left[1 - \max_{\omega\notin\Lambda}\left|\left\langle \boldsymbol{\Phi}_\Lambda^\dagger\boldsymbol{\varphi}_\omega,\ \boldsymbol{g}\right\rangle\right|\right] . \qquad (5)$$

To complete the argument, we just need to rewrite the left-hand side of the latter relation. By construction, the vector $\boldsymbol{s} - \boldsymbol{a}_\Lambda$ is orthogonal to $\boldsymbol{\varphi}_\omega$ for each $\omega \in \Lambda$. Therefore, the left-hand side of (5) does not change if we maximize over all $\omega \in \Omega$:

$$\max_{\omega\notin\Lambda} |\langle \boldsymbol{s} - \boldsymbol{a}_\Lambda,\ \boldsymbol{\varphi}_\omega\rangle| = \max_{\omega\in\Omega} |\langle \boldsymbol{s} - \boldsymbol{a}_\Lambda,\ \boldsymbol{\varphi}_\omega\rangle|$$

Finally, note that

$$\max_{\omega\in\Omega} |\langle \boldsymbol{s} - \boldsymbol{a}_\Lambda,\ \boldsymbol{\varphi}_\omega\rangle| = \|\boldsymbol{\Phi}^*(\boldsymbol{s} - \boldsymbol{a}_\Lambda)\|_\infty .$$

We arrive at the sufficient condition stated in Lemma 2.

References

[1] J. A. Tropp, "Just relax: Convex programming methods for identifying sparse signals in noise," *IEEE Trans. Info. Theory*, vol. 51, no. 3, pp. 1030–1051, Mar. 2006.

Joel A. Tropp (S'03, M'04) 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 in 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 2001 and the Ph.D. degree in 2004. His graduate work was supported by an NSF Graduate Fellowship and a CAM Graduate Fellowship.

In 2004, Dr. Tropp joined the Mathematics Department of the University of Michigan at Ann Arbor as Research Assistant Professor. In 2005, he won an NSF Mathematical Sciences Postdoctoral Research Fellowship, and he was appointed T. H. Hildebrandt Research Assistant Professor.

Since August 2007, Dr. Tropp has been Assistant Professor of Computational and Applied Mathematics at the California Institute of Technology.