CS 754 Project Report

Adithya Bhaskar(190050005) and Jash Ashish Shah (190050052)

May 2021

1 Referred publication

The stem paper for this work is "Computable Performance Bounds on Sparse Recovery" [1] by Gongguo Tang and Arye Nehorai, which appeared in IEEE Transactions on Signal Processing in the issue of January 2015. Our work is a reproduction of theirs, with the experiments repeated and fresh results shown.

2 Introduction and Existing Work

It is an ongoing research topic to find sufficient conditions for sparse recovery that are both tight and easily verifiable. That is, one typically wishes to find sufficient constraints on the the matrix $A \in \mathbb{R}^{m \times n}$, where $m \leq n$, such that for "sufficiently" sparse x, the value of x is efficiently recoverable from the compressive measurement

$$y = Ax + \eta$$

where η is the noise vector, assumed to be small in magnitude. The three most common methods of recovering x from the measurements of y, and those that are our focus, are

- *Basis Pursuit Denoising (BP).* The reconstruction x^* is the solution to

$$\min_{x^* \in \mathbb{R}^n} ||x^*||_1 \text{ s.t. } ||y - Ax^*||_\diamond \leq \varepsilon$$

  Above, $\diamond \in \{0, 1, 2, \infty\}$ is a norm, and ε symbolizes the error level (roughly around the expected magnitude of η).

- *The Dantzig Selector(DS).* [2] The reconstruction x^* is the solution to

$$\min_{x^* \in \mathbb{R}^n} ||x^*||_1 \text{ s.t. } ||A^T(y - Ax^*)||_\infty \leq \mu$$

  Above μ signifies the error level.

- *LASSO.* [3] The optimization problem for LASSO is

$$\min_{x^* \in \mathbb{R}^n} \frac{1}{2}||y - Ax^*||_2^2 + \mu||x^*||_1$$

The three most well known sufficient conditions for sparse recovery are the Restricted Isometry Property [2], Nullspace Property [4] and Mutual Coherence [5] based bounds. We briefly describe each in turn.

Definition 1. Given a matrix $A \in \mathbb{R}^{m \times n}$, the *Restricted Isometry Constant of order k* of A is the least value of δ_k, such that for all $x \in \mathbb{R}^n$ that are k-sparse (have at most k nonzero components), we have

$$1 - \delta_k \leq \frac{||Ax||_2^2}{||x||_2^2} \leq 1 + \delta_k$$

if $\delta_k < 1$, the matrix A is said to satisfy the Restricted Isometry Property (RIP) of order k.
The bounds afforded by RIP are

- For BP, if x is assumed to be k-sparse, then assuming $\delta_{2k}(A) < \sqrt{2} - 1$, we have

$$||x^* - x||_2 \leq \frac{4\sqrt{1 + \delta_{2k}(A)}}{1 - (1 + \sqrt{2})\delta_{2k}(A)}$$

- For DS, if x is assumed to be k-sparse, then if $\delta_{2k}(A) + \delta_{3k}(A) < 1$, then

$$||x^* - x||_2 \leq \frac{4\sqrt{k}}{1 - \delta_{2k}(A) - \delta_{3k}(A)}$$

While the bounds due to RIP are sufficiently tight for most applications, they are incredibly hard to verify. In fact, certifying RIP is NP-hard, and the only positive results known so far is that RIP holds with high probability for certain classes of random matrices.
The Nullspace Property (NSP) concerns itself with the problem (we assume that x is k-sparse):

$$\min_{x^* \in \mathbb{R}^n} ||x^*||_1 \text{ s.t. } Ax^* = Ax$$

NSP claims that the solution to this problem is exact, i.e. $x^* = x$, if and only if

$$\sum_{i \in S} |z_i| < \sum_{i \notin S} |z_i| \quad \forall z \in \ker(A),\ |S| \leq k$$

where $\ker(A) := \{z : Az = 0\}$ is the kernel of A. Another version of stating the above result is via

$$\alpha_k = \max_z ||z||_{k,1} \text{ s.t. } Az = 0,\ ||z|| \leq 1$$

Where $\alpha_k < \frac{1}{2}$ is both necessary and sufficient. Unfortunately, the above optimization is incredibly hard to solve, and it is known that the verification of NSP is NP-hard.
Finally, coming to mutual coherence, we define
Definition 2. The *mutual coherence* (MC) $\mu(A)$ of a matrix $A \in \mathbb{R}^{m \times n}$ with unit normalized columns a_i is defined to be

$$\max_{i \neq j} |a_i^T a_j|$$

The mutual coherence bound for BP, for example, is

$$||x^* - x||_1 \leq \frac{2\varepsilon}{1 - \mu(A)(4k - 1)}$$

While the MC of a matrix is easily found, the bounds it leads to are often quite loose. With the above serving as motivation for needing better bounds on sampling matrices that are also easily verifiable, we introduce the definition of $\omega_\diamond$, and comment on its verification and computation.

3 Definition of $\omega_\diamond$ and related bounds

First, we give the definition of $\omega_\diamond$:
Definition 3. For any real number $s \in [1, n]$ and matrix $A \in \mathbb{R}^{m \times n}$, we define

$$\omega_\diamond(Q, s) := \min_{z: ||z||_1/||z||_\infty \leq s} \frac{||Qz||_\diamond}{||z||_\infty}$$

where $\diamond$ is a norm, and usually Q is one of A and $A^T A$.
We next present some error bounds in terms of $\omega_\diamond$.
Theorem 1. Suppose x is k-sparse, and the error η satisfies $||\eta||_\diamond \leq \varepsilon$, $||A^T\eta||_\infty \leq \mu$ and $||A^T\eta||_\infty \leq \kappa\mu$, $\kappa \in (0, 1)$ for the BP, DS and LASSO problems respectively. Then the solution to the problem BP above, satisfies

$$||x^* - x||_\infty \leq \frac{2\varepsilon}{\omega_\diamond(A, 2k)}$$

Similarly, the solution to DS satisfies

$$||x^* - x||_\infty \leq \frac{2\mu}{\omega_\infty(A^TA, 2k)}$$

and the solution to LASSO satisfies

$$||x^* - x||_\infty \leq \frac{(1+\kappa)\mu}{\omega_\infty(A^TA, \frac{2k}{1-\kappa})}$$

Corollary 1. Note that since both x and x^* are k sparse, their difference is $2k$-sparse, and thus $||x^* - x||_2 \leq \sqrt{2k}||x^* - x||_\infty$. So

$$||x^* - x||_2 \leq \frac{2\sqrt{2k}\varepsilon}{\omega_\diamond(A, 2k)}$$

for BP,

$$||x^* - x||_2 \leq \frac{2\mu\sqrt{2k}}{\omega_\infty(A^TA, 2k)}$$

for the DS, and

$$||x^* - x||_2 \leq \frac{(1+\kappa)\sqrt{2k}\mu}{\omega_\infty(A^TA, \frac{2k}{1-\kappa})}$$

for the LASSO estimator.
Proof. Define $h = x^* - x$. Then, we have

$$c||h||_{k,1} \geq ||h||_1$$

Where $c = 2$ for BP and DS, and $c = \frac{2}{1-\kappa}$ for the LASSO. To see this, for example for BP, note that if S is the support set of x, then we have $||x||_1 \geq ||x+h||_1 \geq ||x||_1 - ||h_S||_1 + ||h_{\bar{S}}||_1$. This implies $||h||_1 \leq 2||h_S||_1 \leq 2||h||_{k,1}$. Similarly the other results too can be derived. Using this, we can derive the following bounds:

$$||h||_1 \leq ck||h||_\infty$$

$$||h||_2 \leq \sqrt{ck}||h||_\infty$$

Now, note that for BP

$$||A(x^* - x)||_\diamond \leq ||y - Ax^*||_\diamond + ||y - Ax||_\diamond \leq \varepsilon + ||\eta||_\diamond \leq 2\varepsilon$$

Similarly,

$$||A^TA(x^* - x)||_\infty \leq 2\mu$$

for the Dantzig selector, and

$$||A^TA(x^* - x)||_\infty \leq (1+\kappa)\mu$$

for the LASSO estimator. Theorem 1 follows from the definition of $\omega_\diamond$.

4 Verification and computation of $\omega_\diamond$

4.1 Verification of $\omega_\diamond > 0$

Clearly, for the above bounds to hold any meaning, we need $\omega_\diamond(Q, s) > 0$ to hold for appropriate Q, s. We will now see how to verify this property. The verification is equivalent to verifying that $\frac{||z||_1}{||z||_\infty} > s$ for all z such that $Qz = 0$. To this end, we can compute

$$s_* := \min\left\{\frac{||z||_1}{||z||_\infty} : Qz = 0\right\}$$

And then verify that $s < s_*$ to check the positiveness of $\omega_\diamond(Q, s)$. We can rewrite the above as

$$\frac{1}{s_*} = \max_z ||z||_\infty \text{ s.t. } Qz = 0, ||z||_1 \le 1$$

This in turn can be solved by solving n linear programs,

$$\max_i \left\{ \max_z z_i \text{ s.t. } Qz = 0, ||z||_1 \le 1 \right\}$$

It is noteworthy that in the case of a row subsampled submatrix of a Hadamard matrix, all of the above optimizations yield the same value.
This implies that we can re-state Proposition 4 of [6] as thus:
Proposition 1. For any $m \times n$ matrix A with $n \ge 32m$, one has

$$s_* = \min \left\{ \frac{||z||_1}{||z||_\infty} : Qz = 0 \right\} < 2\sqrt{2m}$$

4.2 Computation of $\omega_\diamond$

We would like to solve

$$\omega_\diamond(Q, s) = \min_z \frac{||Qz||_\diamond}{||z||_\infty} \text{ s.t. } \frac{||z||_1}{||z||_\infty} \le s$$

Consider the similar but different optimization problem(s),

$$\omega_\diamond^i(Q, s) = \min_z \frac{||Qz||_\diamond}{|z_i|} \text{ s.t. } \frac{||z||_1}{|z_i|} \le s$$

Then, we have
Proposition 2.

$$\min_i \omega_\diamond^i(Q, s) = \omega_\diamond(Q, s)$$

Proof. Firstly note that since $||z||_\infty \ge |z_i|$ for all i, we have

$$\left\{ z : \frac{||z||_1}{|z_i|} \le s \right\} \subseteq \left\{ z : \frac{||z||_1}{||z||_\infty} \le s \right\}$$

And also, $\frac{||Qz||_\diamond}{||z||_\infty} \le \frac{||Qz||_\diamond}{|z_i|}$. Thus, since $\omega_\diamond$ is the minimization of a smaller function over a superset of the domain of the other optimization, we have $\omega_\diamond(Q, s) \le \omega_\diamond^i(Q, s)$ for all i, and hence

$$\min_i \omega_\diamond^i(Q, s) \ge \omega_\diamond(Q, s)$$

But also, consider a vector z^* such that it achieves the minimum possible value in $\omega_\delta(Q, s)$, and let index i^* be such that $|z^*_{i^*}| = ||z^*||_\infty$. Then,

$$\omega_\diamond(Q, s) = \frac{||Qz^*||_1}{||z^*||_\infty} = \frac{||Qz^*||_1}{|z^*_{i^*}|} \ge \omega_\diamond^{i^*}(Q, s) \ge \min_i \omega_\diamond^i(Q, s)$$

Combining this with the previously derived inequality we get the result.
Using Proposition 2, we can write

$$\omega_\diamond(Q, s) = \min_i \left(\min_z ||Qz||_\diamond \text{ s.t. } z_i = 1, ||z||_1 \le s \right)$$

We can rephrase this as:
Theorem 2. The value of $\omega_\diamond(Q, s)$ equals the minimum objective value of the n optimization problems:

$$\min_{\lambda \in \mathbb{R}^{n-1}} ||Q_i - Q^{-i}\lambda||_\diamond \text{ s.t. } ||\lambda||_1 \le s - 1$$

Above, Q_i is the i-th column of Q, and $Q^{-i} \equiv Q(:, -i)$ is the matrix Q with the i-th column removed.
Looking at theorem 2, we see that each of its optimizations measure how well a column can be approximated

	m	51	77	102	128	154	179	205
	s_*	4.795	6.376	7.860	9.558	11.805	14.911	19.186
k	k_*	2	3	3	4	5	7	9
1	ω bd	4.707	3.987	3.708	3.699	3.489	3.452	3.384
	RIC bd			46.588	18.662	14.620	14.526	12.787
2	ω bd	29.684	12.532	9.0694	8.296	7.0436	6.656	6.199
	RIC bd							57.915
3	ω bd		109.619	25.352	17.609	12.408	11.121	9.791
	RIC bd							
4	ω bd				49.088	22.952	18.347	14.798
	RIC bd							
5	ω bd					58.942	31.431	22.182
	RIC bd							
6	ω bd						62.265	32.804
	RIC bd							
7	ω bd						235.536	50.233
	RIC bd							
8	ω bd							87.823
	RIC bd							
9	ω bd							262.907
	RIC bd							

Table 1: Comparison of the ω_2 based bounds and the RIC based bounds on the $l2$ norm of the error for BP in case of an $m \times 256$ sub-matrix of a 256×256 Bernoulli random matrix

by a linear combination of the other columns. In that sense, it can be viewed as a generalization of the Mutual Coherence (MC) property.
Again, we have
Corollary 2. For Hadamard matrices, the objective values of the n optimization problems of Theorem 2 are equal.

5 Numerical Experiments

In this section, we compare the performance of the Omega bounds to those obtained using RIP. To solve the optimization problems of Theorem 2, we use `spgl1` for $\diamond = 2$, and `CVX` for $\diamond = 1, \infty$. Since the RIC of a matrix of any order is difficult to calculate, we use Monte Carlo simulations: for example, to calculate $\delta_{2k}(A)$ we take 5000 random sub-matrices of size $m \times 2k$ from A, calculate their minimal and maximal singular values σ_m and σ_M, and calculate the estimate of δ_{2k} as $\max(1-\sigma_m^2, \sigma_M^2-1)$. We take the maximum across all the random samples as the final estimate for the RIC. We utilize the RIC bounds stated earlier. Note that we *underestimate* the RIC, and hence our approximated RIC bounds will be *tighter* than the actual RIC based bounds. We then compare the bounds for both BP and DS, for the following choices: an $m \times n$ submatrix of a {Bernoulli, Guassian, Hadmard} matrix, with $m \in \{51, 77, 102, 128, 154, 179, 205\}$, n $= 256$ and the columns unit normalized. We also display the value of s_* and $k_* = \lfloor \frac{s_*}{2} \rfloor$, which tells us the maximum value of k for which the Omega bounds hold. Empty entries denote that the bounds were either not applicable (i.e. were negative) or infinite.

We note that the $\omega_\diamond(\cdot)$ not only gives bounds for a wider range of m, but is also always better, with the exception of $k = 11, 12$ for the Dantzig selector in case of $m = 205$ for the Hadamard submatrix. We conclude that the $\omega \diamond (\cdot)$ based bounds are not only much more easily calculable but also tighter than the corresponding RIC bounds.
We then compare the ω bounds to the Mutual Coherence bounds shown above. We compare them for a random $m \times n$ submatrix of a 2048×2048 Hadamard submatrix where n is held fixed at 2048 and the ration m/n is varied to produce different curves, each plotted for values of s from 1 to 40. Certain iterations hung inside a function of spgl1 which we could not fix, and hence we have used NaN values for the ω bounds in those places which explains a bit of the erratic behaviour observed for large s. The reason why the ω bounds extend further is that the MC bounds quickly become inapplicable to the matrix. We

	m	51	77	102	128	154	179	205
	s_*	4.681	6.106	7.717	9.873	11.964	15.049	19.435
k	k_*	2	3	3	4	5	7	9
1	ω bd	4.429	4.118	3.750	3.610	3.522	3.442	3.356
	RIC bd		160.750	38.605	19.409	22.422	12.966	13.573
2	ω bd	29.038	13.155	9.553	7.733	7.289	6.664	6.187
	RIC bd						2416.142	5483.134
3	ω bd		365.042	27.597	15.840	12.878	11.172	15.044
	RIC bd							
4	ω bd				41.884	23.335	18.430	22.654
	RIC bd							
5	ω bd					59.762	31.273	35.034
	RIC bd							
6	ω bd						61.932	57.046
	RIC bd							
7	ω bd						212.015	50.233
	RIC bd							
8	ω bd							107.179
	RIC bd							
9	ω bd							293.531
	RIC bd							

Table 2: Comparison of the ω_2 based bounds and the RIC based bounds on the $l2$ norm of the error for BP in case of an $m \times 256$ sub-matrix of a 256×256 Guassian random matrix

	m	51	77	102	128	154	179	205
	s_*	5.452	7.242	9.428	11.468	14.212	19.231	25.039
k	k_*	2	3	4	5	7	9	12
1	ω bd	3.850	3.548	3.316	3.243	3.165	3.053	3.000
	RIC bd	23.654	21.298	9.202	7.964	6.527	5.656	6.005
2	ω bd	13.466	8.528	6.624	6.020	5.491	4.961	4.698
	RIC bd			129.779	27.257	11.868	8.491	8.013
3	ω bd		27.905	13.159	10.251	8.483	7.047	6.396
	RIC bd				102.229	29.078	12.246	10.992
4	ω bd			37.049	18.869	13.083	9.631	8.287
	RIC bd					65.391	17.585	17.434
5	ω bd				50.241	21.731	13.151	10.532
	RIC bd						20.196	17.777
6	ω bd					45.552	18.454	13.348
	RIC bd						41.097	27.534
7	ω bd					512.231	27.631	17.070
	RIC bd						561.943	41.851
8	ω bd						47.890	22.330
	RIC bd							42.762
9	ω bd						133.479	30.465
	RIC bd							
10	ω bd							44.926
	RIC bd							
11	ω bd							78.218
	RIC bd							
12	ω bd							239.012
	RIC bd							

Table 3: Comparison of the ω_2 based bounds and the RIC based bounds on the $l2$ norm of the error for BP in case of an $m \times 256$ sub-matrix of a 256×256 Hadamard matrix

	m	51	77	102	128	154	179	205
	s_*	4.739	6.229	7.582	9.689	12.439	15.016	19.720
k	k_*	2	3	3	4	6	7	9
1	ω bd	7.212	5.185	4.808	4.642	4.109	4.018	4.142
	RIC bd		121.069	18.859	22.108	15.742	10.395	8.732
2	ω bd	93.841	35.373	27.812	22.015	14.166	12.773	11.525
	RIC bd					139.744	135.698	43.885
3	ω bd		1005.203	219.572	97.353	55.487	42.949	32.359
	RIC bd							
4	ω bd				463.773	169.756	135.901	88.275
	RIC bd							
5	ω bd					555.060	366.644	203.168
	RIC bd					hspace1em		
6	ω bd					4625.112	1125.134	430.752
	RIC bd							
7	ω bd						4914.432	980.651
	RIC bd							
8	ω bd							2806.531
	RIC bd							
9	ω bd							10926.015
	RIC bd							

Table 4: Comparison of the ω_2 based bounds and the RIC based bounds on the $l2$ norm of the error for DS in case of an $m \times 256$ sub-matrix of a 256×256 Bernoulli random matrix

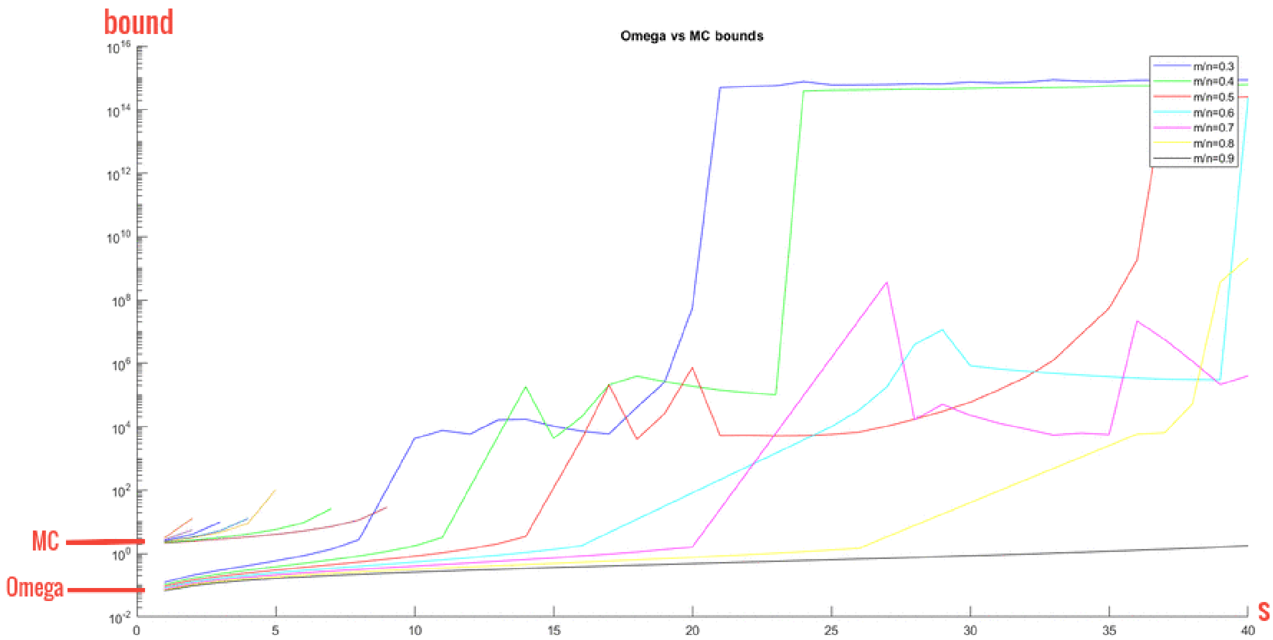
	m	51	77	102	128	154	179	205
	s_*	4.657	6.207	8.067	9.951	12.171	14.977	19.949
k	k_*	2	3	4	4	6	7	9
1	ω bd	6.364	4.946	4.597	4.341	4.211	4.005	4.151
	RIC bd		46.605	18.001	12.802	10.725	10.608	9.835
2	ω bd	100.422	36.580	23.091	17.610	14.442	12.885	11.917
	RIC bd						92.568	43.525
3	ω bd		1044.297	129.800	77.987	54.106	42.068	33.092
	RIC bd							
4	ω bd			6814.667	362.321	232.627	125.853	89.656
	RIC bd							
5	ω bd					760.609	321.814	193.400
	RIC bd							
6	ω bd					15076.555	852.697	408.262
	RIC bd							
7	ω bd						4590.848	1000.523
	RIC bd							
8	ω bd							2155.104
	RIC bd							
9	ω bd							6137.915
	RIC bd							

Table 5: Comparison of the ω_2 based bounds and the RIC based bounds on the $l2$ norm of the error for DS in case of an $m \times 256$ sub-matrix of a 256×256 Guassian random matrix

	m	51	77	102	128	154	179	205
	s_*	5.408	7.125	8.687	11.803	14.179	18.570	24.127
k	k_*	2	3	4	5	7	9	12
1	ω bd	4.808	4.537	3.796	3.415	3.351	3.245	3.151
	RIC bd	37.952	14.005	10.827	8.372	7.149	6.331	5.350
2	ω bd	31.602	17.720	11.128	7.690	7.031	6.196	5.582
	RIC bd			308.921	20.534	14.458	11.843	9.679
3	ω bd		96.346	43.409	17.420	14.391	10.053	8.736
	RIC bd				70.645	24.697	18.474	13.520
4	ω bd			282.160	47.496	30.709	16.662	13.169
	RIC bd					54.258	30.297	18.348
5	ω bd				148.008	71.281	28.894	19.876
	RIC bd					93.302	64.843	22.314
6	ω bd					189.678	52.254	30.739
	RIC bd						68.445	29.524
7	ω bd					2955.929	99.330	49.836
	RIC bd						139.971	35.990
8	ω bd						217.323	79.874
	RIC bd						5582.074	46.747
9	ω bd						1284.652	129.004
	RIC bd							64.014
10	ω bd							226.250
	RIC bd							120.338
11	ω bd							503.928
	RIC bd							155.265
12	ω bd							9477.510
	RIC bd							153.968

Table 6: Comparison of the ω_2 based bounds and the RIC based bounds on the $l2$ norm of the error for DS in case of an $m \times 256$ sub-matrix of a 256×256 Hadamard matrix

observe that not only are the ω bounds applicable for a larger range of s values, but are also always tighter than the corresponding MC bounds.



6 Conclusion

In this reproduction of the stem paper, we considered the $l1$ norm minimization based recovery of sparse signal, and asked whether there existed simultaneously verifiable and tight sufficient conditions. We showed that the answer was yes, giving as proof the $\omega_\diamond(\cdot)$ function. We showed that the value of ω was efficiently calculable *and* that it gave significantly tighter bounds than it's RIC as well as MC counterpart *and* was simultaneously applicable to a larger set of matrices. It seems likely that the tradeoff between tightness and calculability is not fundamental after all.

References

[1] Gongguo Tang and Arye Nehorai. Computable performance bounds on sparse recovery. *IEEE Transactions on Signal Processing*, 63(1):132–141, 2015.

[2] Emmanuel J. Candès, Justin K. Romberg, and Terence Tao. Stable signal recovery from incomplete and inaccurate measurements. *Communications on Pure and Applied Mathematics*, 59(8):1207–1223, 2006.

[3] Robert Tibshirani. Regression shrinkage and selection via the lasso. *JOURNAL OF THE ROYAL STATISTICAL SOCIETY, SERIES B*, 58:267–288, 1994.

[4] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore. Compressed sensing and best k-term approximation. *American Mathematical Society*, 22:211–231, 01 2009.

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

[6] Gongguo Tang and Arye Nehorai. Performance analysis of sparse recovery based on constrained minimal singular values. *IEEE Transactions on Signal Processing*, 59(12):5734–5745, Dec 2011.