

A Compressed Introduction to Compressed Sensing

Benjamin Peterson

April 4, 2015

Abstract

We attempt to convey a sense of compressed sensing. Specifically, we discuss how ℓ_1 minimization and the restricted isometry property for matrices can be used for sparse recovery of underdetermined linear systems even in the presence of noise.

1 Introduction

Suppose $A \in \mathcal{M}_{m \times n}(\mathbb{R})$ is a matrix and we obtain a measurement $y \in \mathbb{R}^m$, which we know to be the image of some $x \in \mathbb{R}^n$ under A . If the system $Ax = y$ is underdetermined (i.e., $m < n$), then x is not unique. Suppose however that we want a single solution and we know x is “small”. We might try to pick an x which solves the equation with the least ℓ_2 norm by solving the convex program

$$\min_{x \in \mathbb{R}^n} \|x\|_{\ell_2} \quad Ax = y. \quad (\text{L})$$

This program has the advantage that a unique solution is guaranteed. Unfortunately, the solution to (L) is often not the best one for applications.

Many real world signals are known to be sparse (i.e. having few nonzero entries) in some basis. Thus, we might alternatively seek a sparse solution for $Ax = y$. The sparsest solution to the linear equation is given by the program

$$\min_{x \in \mathbb{R}^n} \|x\|_{\ell_0} \quad Ax = y. \quad (\text{S})$$

Here the “norm” $\|x\|_{\ell_0}$ counts the number of nonzero entries in a vector x . Solving (S) is a difficult NP-hard problem—the subset-sum problem may be reduced to it—in combinatorial optimization [16]. We will replace (S) with the convex program

$$\min_{x \in \mathbb{R}^n} \|x\|_{\ell_1} \quad Ax = y. \quad (\text{P})$$

Solving this program is often called “basis pursuit” and can be efficiently done with linear programming or specialized algorithms [13, ch. 15]. Relating the solutions of (S) and (P) is major goal in compressed sensing which we will explore in this paper.

There is an intuitive geometrical reason that we might expect the ℓ_1 norm to be better at finding sparse solutions than the ℓ_2 norm. The program (L) finds smallest ball around the origin that intersects the solution subspace of $Ax = y$. The level sets of the ℓ_1 norm are polyhedra, which emphasize the axes. Therefore, it seems “likely” that the (P) finds a sparse solution in the solution subspace.

We now introduce a restriction on matrices that allows sparse solutions to be easily recovered. Recall that a vector is said to be k -sparse if it has at most k nonzero entries.

Definition 1.1 ([7]). A matrix $A \in \mathcal{M}_{m \times n}(\mathbb{R})$ satisfies the *restricted isometry property* (RIP) of order k if there exists a $\delta_k \geq 0$ such that for all k -sparse vectors $x \in \mathbb{R}^n$,

$$(1 - \delta_k)\|x\|_{\ell_2}^2 \leq \|Ax\|_{\ell_2}^2 \leq (1 + \delta_k)\|x\|_{\ell_2}^2.$$

The smallest such δ_k is called A 's *restricted isometry constant* of order k .

Notice that $\delta_1 \leq \delta_2 \leq \dots$. The RIP is related to vector space frames [8] and the Johnson-Lindenstrauss lemma [1]. Unfortunately, checking whether a matrix has the restricted isometry property is NP-Hard in general [19]. On the other hand, many families of random matrices (e.g. having Gaussian or Bernoulli entries) satisfy the RIP with high probability [6, §1.3].

Theorem 1.2 ([7, lemma 1.2]). *Suppose $A \in \mathcal{M}_{m \times n}$ satisfies RIP with $\delta_{2s} < 1$. Then the equation $Ax = y$ has an unique s -sparse solution given by (S).*

Proof. Suppose $x, x' \in \mathbb{R}^m$ are s -sparse and $Ax = Ax'$. By the RIP,

$$0 \leq (1 - \delta_{2s})\|x - x'\|_{\ell_2} \leq \|A(x - x')\|_{\ell_2} = 0.$$

Necessarily, $x = x'$. □

Before stating the main theorem, we introduce the notion of error. In reality, we do not know Ax exactly. Instead, we measure $y = Ax + z$, where $z \in \mathbb{R}^m$ is a small error vector satisfying $\|z\|_{\ell_2} \leq \epsilon$. We would like our recovery of x to be robust against error. To deal with this, we generalize (P) to the (still convex) program

$$\min_{x \in \mathbb{R}^n} \|x\|_{\ell_1} \quad \|Ax - y\|_{\ell_2} \leq \epsilon. \quad (\text{P}')$$

Finally, we can state the main theorem of this lecture. We denote the vector containing only the s largest entries of x by x_s .

Theorem 1.3 ([5, 1.3]). *Suppose $A \in \mathcal{M}_{m \times n}(\mathbb{R})$ satisfies RIP with $\delta_{2s} < \sqrt{2} - 1$. Let x^* denote the solution to (P'). Then, there are constants C_0, C_1 depending only on δ_{2s} such that*

$$\|x^* - x\|_{\ell_2} \leq C_0 s^{-1/2} \|x - x_s\|_{\ell_1} + C_1 \epsilon.$$

Theorem 1.3 tells us that when RIP holds, basis pursuit recovers solutions very close to sparse solutions of the equation.

2 Proof of Theorem 1.3

We closely follow [5].

Set $x = x^* + h$. Our goal is to show $\|h\|_{\ell_2}$ is small. The RIP for A only gives us control over sparse vectors, so we start by breaking h up into s -sparse vectors. Let T_0 be the indexes of the s largest entries of x . Let T_1 be the indexes of the s largest entries of $h_{T_0^c}$. Let T_2 be the indexes of the s largest entries of $h_{(T_0 \cup T_1)^c}$ and so on. We observe that h can be written as the sum of s -sparse vectors $h_{T_0} + h_{T_1} + h_{T_2} + \dots$. Using the triangle inequality,

$$\|x - x^*\|_{\ell_2} = \|h\|_{\ell_2} \leq \|h_{T_0 \cup T_1}\|_{\ell_2} + \|h_{(T_0 \cup T_1)^c}\|_{\ell_2}. \quad (1)$$

We will estimate the two terms of (1) separately and then combine them to prove the theorem. We start by showing the $\|h_{(T_0 \cup T_1)^c}\|_{\ell_2}$ term can be bounded in terms of the first term $\|h_{T_0 \cup T_1}\|_{\ell_2}$. Several useful intermediate inequalities are obtained along the way.

Lemma 2.1 (Tail estimates).

$$\|h_{T_0^c}\|_{\ell_1} \leq \|h_{T_0}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1} \quad (2)$$

$$\sum_{j \geq 2} \|h_{T_j}\|_{\ell_2} \leq s^{-1/2} \|h_{T_0^c}\|_{\ell_1} \quad (3)$$

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} \leq \|h_{T_0}\|_{\ell_2} + 2s^{-1/2} \|x - x_s\|_{\ell_1} \quad (4)$$

Proof. Since x is feasible for (P') and x^* is a minimum,

$$\|x\|_{\ell_1} \geq \|x + h\|_{\ell_1} = \sum_{i \in T_0} |x_i + h_i| + \sum_{i \in T_0^c} |x_i + h_i| \geq \|x_{T_0}\|_{\ell_1} - \|h_{T_0}\|_{\ell_1} + \|h_{T_0^c}\|_{\ell_1} - \|x_{T_0^c}\|_{\ell_1}.$$

The last step uses the triangle inequality twice. Rewriting and applying the reverse triangle inequality proves (2).

$$\|h_{T_0^c}\|_{\ell_1} \leq \|x\|_{\ell_1} - \|x_{T_0}\|_{\ell_1} + \|x_{T_0^c}\|_{\ell_1} + \|h_{T_0}\|_{\ell_1} \leq \|h_{T_0}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1}$$

If $j \geq 2$,

$$\|h_{T_j}\|_{\ell_2} \leq s^{1/2} \|h_{T_j}\|_{\ell_\infty} \leq s^{-1/2} \|h_{T_{j-1}}\|_{\ell_1}.$$

Summing yields (3).

$$\sum_{j \geq 2} \|h_{T_j}\|_{\ell_2} \leq s^{-1/2} \sum_{j \geq 1} \|h_{T_j}\|_{\ell_1} = s^{-1/2} \|h_{T_0^c}\|_{\ell_1}$$

From this, we immediately obtain

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} = \left\| \sum_{j \geq 2} h_{T_j} \right\|_{\ell_2} \leq \sum_{j \geq 2} \|h_{T_j}\|_{\ell_2} \leq s^{-1/2} \|h_{T_0^c}\|_{\ell_1}. \quad (5)$$

By the Cauchy-Schwarz inequality,

$$\|h_{T_0}\|_{\ell_1} \leq s^{1/2} \|h_{T_0}\|_{\ell_2}. \quad (6)$$

Applying (5), (2), and (6) proves (4).

$$\|h_{(T_0 \cup T_1)^c}\|_{\ell_2} \leq s^{-1/2} \|h_{T_0^c}\|_{\ell_1} \leq s^{-1/2} (\|h_{T_0}\|_{\ell_1} + 2\|x_{T_0^c}\|_{\ell_1}) \leq \|h_{T_0}\|_{\ell_2} + 2s^{-1/2} \|x_{T_0^c}\|_{\ell_1}$$

□

Next, we want to bound the main part of the error term $\|h_{T_0 \cup T_1}\|_{\ell_2}$. We begin with a lemma.

Lemma 2.2. *Suppose x and x' are s -sparse and s' -sparse respectively and supported on disjoint sets. Then*

$$|\langle Ax, Ax' \rangle| \leq \delta_{s+s'} \|x\|_{\ell_2} \|x'\|_{\ell_2}$$

Proof. We may assume without a loss of generality that $\|x\|_{\ell_2} = \|x'\|_{\ell_2} = 1$. The RIP tells us that

$$2(1 - \delta_{s+s'}) = (1 - \delta_{s+s'}) \|x + x'\|_{\ell_2}^2 \leq \|Ax \pm Ax'\|_{\ell_2}^2 \leq (1 + \delta_{s+s'}) \|x + x'\|_{\ell_2}^2 = 2(1 + \delta_{s+s'}).$$

By the polarization identity,

$$|\langle Ax, Ax' \rangle| \leq \frac{1}{4} \left| \|Ax + Ax'\|_{\ell_2}^2 - \|Ax - Ax'\|_{\ell_2}^2 \right| \leq \delta_{s+s'}.$$

□

Lemma 2.3 (Main term estimate).

$$\|h_{T_0 \cup T_1}\|_{\ell_2} \leq (1 - \rho)^{-1} (\alpha \epsilon + 2\rho s^{-1/2} \|x - x_s\|_{\ell_1}) \quad (7)$$

where

$$\alpha \equiv \frac{2\sqrt{1 - \delta_{2s}}}{1 - \delta_{2s}}, \quad \rho \equiv \frac{\sqrt{2}\delta_{2s}}{1 - \delta_{2s}}.$$

Proof. RIP allows us to control the size of $\|h_{T_0 \cup T_1}\|_{\ell_2}$ with $\|Ah_{T_0 \cup T_1}\|_{\ell_2}$, and so we start by bounding the latter. We break $\|Ah_{T_0 \cup T_1}\|_{\ell_2}$ into parts using properties of the inner product.

$$\|Ah_{T_0 \cup T_1}\|_{\ell_2}^2 = \langle Ah_{T_0 \cup T_1}, Ah \rangle - \sum_{j \geq 2} (\langle Ah_{T_0}, Ah_{T_j} \rangle + \langle Ah_{T_1}, Ah_{T_j} \rangle) \quad (8)$$

From the triangle inequality and hypothesis,

$$\|Ah\|_{\ell_2} = \|A(x - x^*)\|_{\ell_2} \leq \|Ax^* - y\|_{\ell_2} + \|y - Ax\|_{\ell_2} \leq 2\epsilon. \quad (9)$$

The first term of (8) can be bounded using Cauchy-Schwarz, the RIP, and (9).

$$|\langle Ah_{T_0 \cup T_1}, Ah \rangle| \leq \|Ah_{T_0 \cup T_1}\|_{\ell_2} \|Ah\|_{\ell_2} \leq 2\epsilon \sqrt{1 + \delta_{2s}} \|h_{T_0 \cup T_1}\|_{\ell_2} \quad (10)$$

Since T_0 and T_1 are disjoint Cauchy-Schwarz gives

$$\|h_{T_0}\|_{\ell_2} + \|h_{T_1}\|_{\ell_2} \leq \sqrt{2} \|h_{T_0 \cup T_1}\|_{\ell_2}.$$

To estimate the sum term of (9), we apply lemma 2.2 and (3).

$$\begin{aligned} \sum_{j \geq 2} |\langle Ah_{T_0}, Ah_{T_j} \rangle| + |\langle Ah_{T_1}, Ah_{T_j} \rangle| &\leq \delta_{2s} (\|h_{T_0}\|_{\ell_2} + \|h_{T_1}\|_{\ell_2}) \sum_{j \geq 2} \|h_{T_j}\|_{\ell_2} \\ &\leq \delta_{2s} \sqrt{2} s^{-1/2} \|h_{T_0 \cup T_1}\|_{\ell_2} \|h_{T_0^c}\|_{\ell_1} \end{aligned} \quad (11)$$

We now have

$$(1 - \delta_{2s}) \|h_{T_0 \cup T_1}\|_{\ell_2}^2 \leq \|Ah_{T_0 \cup T_1}\|_{\ell_2}^2 \leq \|h_{T_0 \cup T_1}\|_{\ell_2} (2\epsilon \sqrt{1 + \delta_{2s}} + \sqrt{2} \delta_{2s} s^{-1/2} \|h_{T_0^c}\|_{\ell_1})$$

from RIP, (10), and (11). We divide by $(1 - \delta_{2s}) \|h_{T_0 \cup T_1}\|_{\ell_2}$.

$$\|h_{T_0 \cup T_1}\|_{\ell_2} \leq \alpha\epsilon + \rho s^{-1/2} \|h_{T_0^c}\|_{\ell_1}$$

From (2) and (6), we obtain

$$\|h_{T_0 \cup T_1}\|_{\ell_2} \leq \alpha\epsilon + \rho s^{-1/2} (\|h_{T_0}\|_{\ell_1} + 2\|x - x_s\|_{\ell_1}) \leq \alpha\epsilon + \rho \|h_{T_0 \cup T_1}\|_{\ell_2} + 2\rho s^{-1/2} \|x - x_s\|_{\ell_1}.$$

By the hypothesis of theorem 1.3, $\rho < 1$, and we get

$$\|h_{T_0 \cup T_1}\|_{\ell_2} \leq (1 - \rho)^{-1} (\alpha\epsilon + 2\rho s^{-1/2} \|x - x_s\|_{\ell_1}),$$

which completes the lemma's proof. \square

Applying our estimates (4) and lemma 2.3 to the two terms of (1), we have

$$\|h\|_{\ell_2} \leq 2\|h_{(T_0 \cup T_1)}\|_{\ell_2} + 2s^{-1/2} \|x - x_s\|_{\ell_1} \leq 2(1 - \rho)^{-1} (\alpha\epsilon + (1 + \rho)s^{-1/2} \|x - x_s\|_{\ell_1}).$$

This completes the proof of theorem 1.3.

3 Remarks

A huge amount of research in compressed sensing has appeared since the publication of the original papers [6, 10] around 2004-2006. Among other things, researchers have investigated techniques for sparse recovery besides basis pursuit e.g. matching pursuit [20]. There is also replacement condition for RIP called the nullspace property, which is a necessary and sufficient condition for (S) and (P) to have the same solutions [9].

As far as applications are concerned, compressed sensing is being used in areas as diverse as tomography, astronomy, machine learning, linear coding, and experiment design [2, 3]. It is particularly useful in situations where minimizing the work done in sensors is important such as space probes. Gimmicks like single-pixel cameras [11] have captured the public imagination.

Finally, for readers still curious, there are a lot of other (possibly more palatable) introductions to compressed sensing [4, 12, 13, 14, 15, 17, 18].

References

- [1] Richard Baraniuk, Mark Davenport, Ronald DeVore, and Michael Wakin, *A simple proof of the restricted isometry property for random matrices*, Constructive Approximation **28** (2008), no. 3, 253–263.
- [2] Richard G Baraniuk, *More is less: signal processing and the data deluge*, Science **331** (2011), no. 6018, 717–719.
- [3] Emmanuel Candes and Terence Tao, *The dantzig selector: statistical estimation when p is much larger than n* , The Annals of Statistics (2007), 2313–2351.
- [4] Emmanuel J Candès and Michael B Wakin, *An introduction to compressive sampling*, Signal Processing Magazine, IEEE **25** (2008), no. 2, 21–30.
- [5] Emmanuel J Candès, *The restricted isometry property and its implications for compressed sensing*, Comptes Rendus Mathématique **346** (2008), no. 9, 589–592.
- [6] Emmanuel J Candès, Justin K Romberg, and Terence Tao, *Stable signal recovery from incomplete and inaccurate measurements*, Communications on pure and applied mathematics **59** (2006), no. 8, 1207–1223.
- [7] Emmanuel J Candès and Terence Tao, *Decoding by linear programming*, IEEE Transactions on Information Theory **51** (2005), no. 12, 4203–4215.
- [8] Ole Christensen, *An introduction to frames and riesz bases*, Springer Science & Business Media, 2003.
- [9] Albert Cohen, Wolfgang Dahmen, and Ronald DeVore, *Compressed sensing and best k -term approximation*, Journal of the American mathematical society **22** (2009), no. 1, 211–231.
- [10] David L Donoho, *Compressed sensing*, IEEE Transactions on Information Theory **52** (2006), no. 4, 1289–1306.
- [11] Marco F Duarte, Mark A Davenport, Dharmpal Takhar, Jason N Laska, Ting Sun, Kevin E Kelly, and Richard G Baraniuk, *Single-pixel imaging via compressive sampling*, IEEE Signal Processing Magazine **25** (2008), no. 2, 83.
- [12] Yonina C Eldar and Gitta Kutyniok, *Compressed sensing: theory and applications*, Cambridge University Press, 2012.
- [13] Simon Foucart and Holger Rauhut, *A mathematical introduction to compressive sensing*, Springer, 2013.
- [14] Kazunori Hayashi, Masaaki Nagahara, and Toshiyuki Tanaka, *A user’s guide to compressed sensing for communications systems*, IEICE transactions on communications **96** (2013), no. 3, 685–712.
- [15] Gitta Kutyniok, *Theory and applications of compressed sensing*, GAMM-Mitteilungen **36** (2013), no. 1, 79–101.
- [16] Balas Kausik Natarajan, *Sparse approximate solutions to linear systems*, SIAM journal on computing **24** (1995), no. 2, 227–234.
- [17] Saad Qaisar, Rana Muhammad Bilal, Wafa Iqbal, Muqaddas Naureen, and Sungyoung Lee, *Compressive sensing: From theory to applications, a survey*, Journal of Communications and Networks **15** (2013), no. 5, 443–456.
- [18] Justin Romberg, *Imaging via compressive sampling [introduction to compressive sampling and recovery via convex programming]*, IEEE Signal Processing Magazine **25** (2008), no. 2, 14–20.

- [19] Andreas M Tillmann and Marc E Pfetsch, *The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing*, IEEE Transactions on Information Theory **60** (2014), no. 2, 1248–1259.
- [20] Joel A Tropp and Anna C Gilbert, *Signal recovery from random measurements via orthogonal matching pursuit*, IEEE Transactions on Information Theory **53** (2007), no. 12, 4655–4666.