Reweighted ℓ_1-norm minimization with guarantees: An incremental measurement approach to sparse reconstruction

João F. C. Mota*, Lior Weizman†, Nikos Deligiannis‡, Yonina C. Eldar†, and Miguel R. D. Rodrigues§
* Institute of Sensors, Signals and Systems, Heriot-Watt University, Edinburgh, UK
Email: j.mota@hw.ac.uk
† Department of Electrical Engineering, Technion — Israel Institute of Technology, Israel
Email: weizmanl@tx.technion.ac.il, yonina@ee.technion.ac.il
‡ Department of Electronics and Informatics, Vrije Universiteit Brussel — imec, Belgium
Email: ndeligia@etrovub.be
§ Department of Electronic and Electrical Engineering, University College London, UK
Email: m.rodrigues@ucl.ac.uk

I. Introduction and Problem Statement

Sparse reconstruction consists of finding a sparse vector $x^\star \in \mathbb{R}^n$ from linear measurements $b = Ax^\star$, where $A \in \mathbb{R}^{m\times n}$ with $m \ll n$. A popular approach to this problem is to solve *basis pursuit* (BP) [1]:

$$\underset{x}{\text{minimize}} \ \|x\|_1 \quad \text{subject to} \ \ Ax = b\,, \tag{1}$$

where $\|\cdot\|_1$ is the ℓ_1-norm of x. Because BP is a convex problem, one can not only find its solutions in polynomial time, but also characterize its sample complexity, i.e., the number of measurements m that BP requires to reconstruct $x^\star$. For example, [2], [3] showed that if A has i.i.d. Gaussian entries and $x^\star$ has sparsity s, then $m \gtrsim 2s\log(n/s)$ guarantees that, with high probability, $x^\star$ is the unique solution of BP (1). They also showed that this bound is tight.

Another approach to sparse reconstruction is the *iteratively reweighted ℓ_1-norm* (IRL1) minimization [4] which, starting at an arbitrary $x^0 \in \mathbb{R}^n$, solves a sequence of weighted BP problems:

$$x^{k+1} \in \arg\min_x \ \sum_{i=1}^n \frac{|x_i|}{|x_i^k| + \epsilon} \quad \text{subject to} \ \ Ax = b\,, \tag{2}$$

where $x^k = (x_1^k, \ldots, x_n^k)$ is the estimate of $x^\star$ at iteration k, and $\epsilon > 0$. Experiments show that IRL1 (2) reconstructs $x^\star$ using fewer measurements than BP (1) and several justifications have been given: one is that RIP constants associated to IRL1 are smaller than the ones associated to BP [5]; another is that (2) is a heuristic to solve

$$\underset{x}{\text{minimize}} \ \sum_{i=1}^n \log\left(|x_i| + \epsilon\right) \quad \text{subject to} \ \ Ax = b\,, \tag{3}$$

a nonconvex problem whose objective approximates the cardinality of a vector better than the ℓ_1-norm does [4], [6]; see also [7]–[10].

Both BP and IRL1 require the number of measurements m to be selected in advance. Whereas the minimal m for BP has been well characterized and depends only on the sparsity s of $x^\star$ [2], [3], the minimal m for IRL1 has never been fully characterized and it will necessarily depend on the quality of the initialization x^0, a quantity that seems impossible to estimate without knowing $x^\star$.

Contributions. We design an algorithm based on IRL1 that, independently of the initialization x^0, *provably* reconstructs $x^\star$; in addition, it automatically selects the number of measurements m, via a feedback mechanism between the encoder and the decoder. Experiments show that our algorithm adds no significant computation with respect to IRL1, but the number of measurements that it selects is often smaller than the number of measurements that IRL1 requires, even when we know the exact phase transition of IRL1.

II. Proposed Algorithm

Motivated by medical and astronomical imaging [11]–[14], where sensing is expensive or time-consuming, we propose to reduce the number of measurements by acquiring them incrementally, in small blocks. This departs from the standard paradigm where measurements are taken all at once, but it enables us to select them automatically.

The method, detailed in Alg. 1, starts with a small number of measurements m^0, e.g., twice the estimated sparsity of $x^\star$ or a fixed fraction of n. At each iteration t, Alg. 1 runs IRL1 (steps 5-9) until the stopping criterion in step 9 is met. We show, using [15], that this takes a finite number of iterations and that the stopping point is a local minimizer of (3). Prior work showed that (2) has sparse limit points [5], [6], not finite convergence to local minimizers. We also establish conditions guaranteeing that a local minimizer of (3) is equal to $x^\star$. These are specified in step 11 and require a grid search in the interval $(0, \overline{\delta}]$, where $\overline{\delta} > 0$. If Alg. 1 finds $\delta \in (0, \overline{\delta}]$ satisfying those conditions, it halts; otherwise, it takes an additional block of γ measurements, in steps 15-17. Rescaling the old entries of A and b in step 17 is necessary to keep all entries of A with the same variance.

Proposition. *Alg. 1 stops in a finite number of iterations $T < \lceil (n - m^0)/\gamma \rceil$, with probability 1. Furthermore, if $\overline{\delta} < |x_i^\star|$, for all i such that $x_i^\star \neq 0$, and $x^\star$ satisfies the equations in step 11 then, with high probability, the algorithm finds $x^\star$, i.e., $x^T = x^\star$.*

This does not specify the number of measurements that Alg. 1 selects. But, in practice, it often selects less measurements than IRL1.

Experiments. We generated $x^\star$ with $(n, s) = (1000, 70)$. Varying m from 1 to 401, we then generated, for each m, 100 realizations of $A \in \mathbb{R}^{m\times n}$ with i.i.d. $\mathcal{N}(0, 1/m)$ entries, and set $b = Ax^\star$. For each realization, we ran BP (1), and IRL1 (2) with $\epsilon = 0.2$ and the stopping criterion of step 11; we also ran 100 independent instances of Alg. 1 with $(m^0, \gamma, \overline{\delta}) = (100, 10, 10^{-2})$. Fig. 1(a) plots the phase transitions, i.e., the success rate versus m. Fig. 1(b) plots the number of times each algorithm solved a weighted BP. As Alg. 1 computes m automatically, it has no phase transition. The vertical lines in Fig. 1(a), which indicate the minimum, median, and maximum number of measurements over the independent 100 runs, show that the number of measurements selected by Alg. 1 coincides with the phase transition of IRL1. Fig. 1(b) shows that this was achieved without adding significant computation.

In conclusion, our algorithm automatically selects a number of measurements smaller than the number of measurements that IRL1 requires, has comparable complexity, and provably reconstructs $x^\star$.

Algorithm 1 Incremental measurement IRL1

Parameters: Initial number of measurements m^0, initialization vector $x^0 \in \mathbb{R}^n$, increment $\gamma \in \mathbb{N}$, and parameters $\epsilon, \overline{\delta} > 0$

Initialization:

1: Generate matrix $A : m^0 \times n$ with entries $\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/m^0)$, and take measurements $b = Ax^\star$

2: Set $m = m^0$ and $t = 0$

Algorithm:

3: **for** $t = 1, 2, \ldots$ **do**

4: $\quad k = 0$; $z^0 = x^t$

5: $\quad$ **repeat**

6: $\quad\quad k \leftarrow k + 1$

7: $\quad\quad$ Set $w_i^k = 1/(|z_i^{k-1}| + \epsilon)$, for $i = 1, \ldots, n$

8: $\quad\quad$ Find a solution z^k of

$$\begin{array}{ll} \underset{z}{\text{minimize}} & \sum_{i=1}^n w_i^k |z_i| \\ \text{subject to} & Az = b \end{array}$$

9: $\quad$ **until** $\sum_{i=1}^n w_i^k |z_i^{k+1}| = \sum_{i=1}^n w_i^k |z_i^k|$

10: $\quad$ Set $x^t = z^k$

11: $\quad$ Defining $\mathcal{I}_\delta^t := \left\{ i : |x_i^t| \geq \delta \right\}$ and $s_\delta^t := \left|\mathcal{I}_\delta^t\right|$, find $0 < \delta \leq \overline{\delta}$ such that $0 < s_\delta^t < n$, and also

$$m \geq 2 \left[\sum_{i \in \mathcal{I}_\delta^t} \left(\frac{\epsilon}{|x_i^t| + \epsilon} \right)^{2-2p} \right] \log \frac{n}{s_\delta^t} + \frac{7}{5} s_\delta^t + 1$$

$$s_\delta^t \leq \frac{10}{3} \left[\sum_{i \in \mathcal{I}_\delta^t} \left(\frac{\epsilon}{|x_i^t| + \epsilon} \right)^{2-2p} \right] \log \frac{n}{s_\delta^t}$$

12: $\quad$ **if** such δ exists **then**

13: $\quad\quad$ Terminate algorithm and return $\widehat{x}$, where

$$\widehat{x}_i = \begin{cases} x_i^t & , i \in \mathcal{I}_\delta^t \\ 0 & , i \notin \mathcal{I}_\delta^t \end{cases} .$$

14: $\quad$ **else**

15: $\quad\quad$ Set $m \leftarrow m + \gamma$

16: $\quad\quad$ Generate $\overline{A}^t : \gamma \times n$ with entries $\overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 1/m)$

17: $\quad\quad$ Take measurements $\overline{b}^t = \overline{A}^t x^\star$ and update

$$A \leftarrow \begin{bmatrix} \sqrt{\frac{m-\gamma}{m}} A \\ \overline{A}^t \end{bmatrix} \qquad b \leftarrow \begin{bmatrix} \sqrt{\frac{m-\gamma}{m}} b \\ \overline{b}^t \end{bmatrix}$$

18: $\quad$ **end if**

19: **end for**

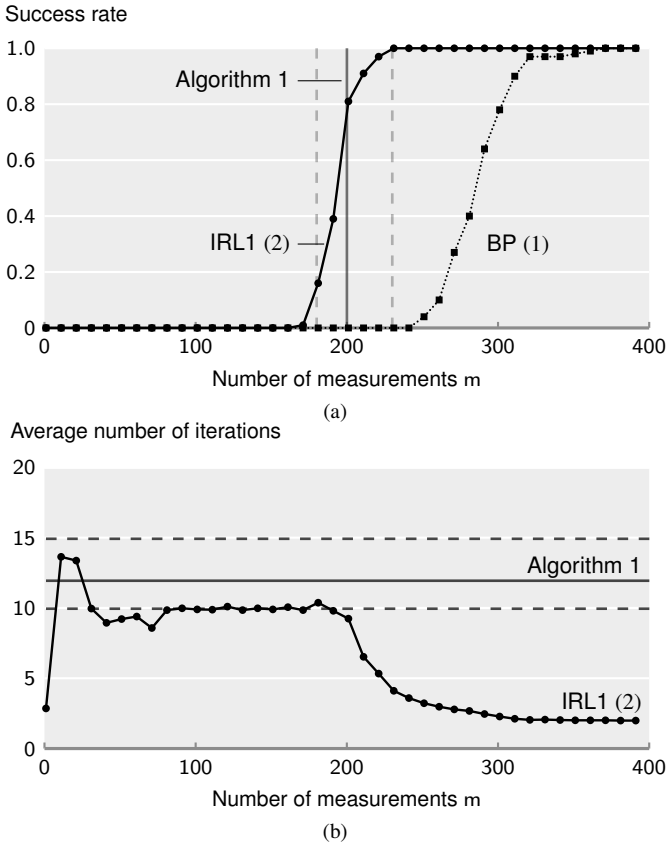

Figure 1. (a) Phase transitions, over 100 random trials, of BP and IRL1 with $\epsilon = 0.2$. Alg. 1, also executed over 100 random trials, computes the number of measurements automatically: the median was 200 (solid line), the minimum and maximum 180 and 230 (dashed lines). (b) Average number of iterations (on k) of (2). The horizontal lines indicate the minimum, median, and maximum number of times Alg. 1 solved the weighted BP problem in step 8, that is, the cumulative number of its inner iterations.

References

[1] S. Chen, D. Donoho, and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM J. Sci. Comp.*, vol. 20, no. 1, pp. 33–61, 1998.

[2] V. Chandrasekaran, B. Recht, P. Parrilo, and A. Willsky, "The convex geometry of linear inverse problems," *Found. Computational Mathematics*, vol. 12, pp. 805–849, 2012.

[3] D. Amelunxen, M. Lotz, M. Mccoy, and J. Tropp, "Living on the edge: Phase transitions in convex programs with random data," *Information and Inference: A Journal of the IMA*, pp. 1–71, 2014.

[4] E. Candès, M. Wakin, and S. Boyd, "Enhancing sparsity by reweighted ℓ_1 minimization," *J. Fourier Anal. Appl.*, vol. 14, pp. 877–905, 2008.

[5] D. Needell, "Noisy signal recovery via iterative reweighted ℓ_1-minimization," in *Asilomar Conf. Signals, Systems, and Computers*, 2009, pp. 113–117.

[6] Y.-B. Zhao and D. Li, "Reweighted ℓ_1-minimization for sparse solutions to underdetermined linear systems," *SIAM J. Optim.*, vol. 22, no. 3, pp. 1065–1088, 2012.

[7] S. Foucart and M.-J. Lai, "Sparsest solutions of underdetermined linear systems via l_q-minimization for $0 < q \leq 1$," *Applied and Computational Harmonic Analysis*, vol. 26, pp. 395–407, 2009.

[8] X. Chen and W. Zhou, "Convergence of reweighted ℓ_1 minimization algorithms and unique solution of truncated ℓ_p minimization," The Hong Kong Polytechnique University, Hong Kong, Tech. Rep., 2010.

[9] M.-J. Lai and J. Wang, "An unconstrained l_q minimization with $0 < q \leq 1$ for sparse solution of underdetermined linear systems," *SIAM J. Optim.*, vol. 21, no. 1, pp. 82–101, 2011.

[10] S. Zhou, N. Xiu, Y. Wang, and L. Kong, "A null-space-based weighted ℓ_1 minimization approach to compressed sensing," *Information and Inference: A Journal of the IMA*, vol. 5, pp. 76–102, 2016.

[11] M. Lustig, D. Donoho, and J. Pauly, "Sparse MRI: The application of compressed sensing to rapid MR imaging," *Magnetic Resonance in Medicine*, vol. 58, pp. 1182–1195, 2007.

[12] L. Weizman, Y. C. Eldar, and D. Ben-Bashat, "Compressed sensing for longitudinal MRI: An adaptive-weighted approach," *Med. Phys.*, vol. 42, no. 9, pp. 5195–5207, 2015.

[13] K. Choi, J. Wang, L. Zhu, T.-S. Suh, S. Boyd, and L. Xing, "Compressed sensing based cone-beam computed tomography reconstruction with a first-order method," *Med. Phys.*, vol. 37, no. 9, pp. 5113–5125, 2010.

[14] Y. Wiaux, L. Jacques, G. Puy, A. M. Scaife, and P. Vandergheynst, "Compressed sensing imaging techniques for radio interferometry," *Mon. Not. R. Astron. Soc.*, vol. 395, pp. 1733–1742, 2009.

[15] O. L. Mangasarian, *Applied Mathematics and Parallel Computing*. Physica-Verlag HD, 1996, ch. Machine Learning via Polyhedral Concave Minimization, pp. 175–188.