Structured-signal recovery from single-bit measurements

Yaniv Plan
Department of Mathematics
University of Michigan, Ann Arbor, Michigan 48104
Email: yplan@umich.edu

***Abstract*—1-bit compressed sensing was introduced by Boufounos and Baraniuk in 2008 as a model of extreme quantization; only the sign of each measurement is retained. Recent theoretical and algorithmic advances, combined with the ease of hardware implementation, show that it is an effective method of signal acquisition. Surprisingly, in the high-noise regime there is almost no information loss from 1-bit quantization. We review and revise recent results, and compare to closely related statistical problems: sparse binary regression and binary matrix completion.**

I. Introduction

Discrete measurements arise both in signal processing and statistical inference, but for different reasons. In some cases, they are inherent to the data—consider a statistical experiment in which the response is a binary variable indicating the presence or absence of a certain disease. In other cases the level of discretization is chosen—consider quantization in analog-to-digital conversion. We focus on the extreme case in which all measurements are binary. For further signal-processing motivation, see [1].

It turns out that the abstract statistical models and signal-processing models nearly match, but with subtle differences that have strong influence on the methods of signal reconstruction and the theoretical challenges. We point out these differences and how the ideas from 1-bit compressed sensing allow new methods and results in binary regression.

In Section II, we describe recent results in *1-bit compressed sensing* and give connections to standard *compressed sensing*. These methods allow for a new semi-parametric approach to *sparse binary regression*, described in Section III. In Section IV we describe modern theoretical results in binary PCA with missing entries, or *binary matrix completion*.

II. 1-bit compressed sensing

Unquantized compressed sensing [7] concerns the reconstruction of sparse signals from linear measurements. Let $\|\boldsymbol{x}\|_0$ give the number of nonzero entries of $\boldsymbol{x}$. We assume that $\|\boldsymbol{x}\|_0 \leq s$ i.e., $\boldsymbol{x}$ is sparse. One observes data of the form

$$y_i = \langle \boldsymbol{a}_i, \boldsymbol{x} \rangle \qquad i = 1, \ldots, m$$

and would like to reconstruct $\boldsymbol{x} \in \mathbb{R}^n$ from $\{y_i, \boldsymbol{a}_i\}$.

Supported by NSF Postdoctoral Research Fellowship under award No. 1103909.

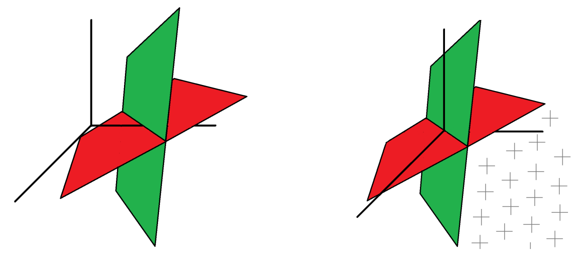

Fig. 1: On left: Linear measurements $y_1 = \langle \boldsymbol{a}_1, \boldsymbol{x} \rangle$ and $y_2 = \langle \boldsymbol{a}_2, \boldsymbol{x} \rangle$ determine that $\boldsymbol{x}$ must lie in the intersection of the two hyperplanes. On right: Single bit measurements $y_1 = \text{sign}(\langle \boldsymbol{a}_1, \boldsymbol{x} \rangle)$ and $y_2 = \text{sign}(\langle \boldsymbol{a}_2, \boldsymbol{x} \rangle)$ determine that $\boldsymbol{x}$ must lie in the region denoted by $+$ signs.

In 1-bit compressed sensing [4], only the sign of each measurement is retained:

$$y_i = \text{sign}(\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle) \qquad i = 1, \ldots, m.$$

Above $\text{sign}(t) = 1$ if $t \geq 0$ and $\text{sign}(t) = -1$ it $t < 0$. In matrix form,

$$\boldsymbol{y} = \text{sign}(\boldsymbol{A}\boldsymbol{x})$$

where $\boldsymbol{A} \in \mathbb{R}^{m \times n}$ is a matrix whose i-th row is equal to $\boldsymbol{a}_i$, and we allow the sign function to act on a vector by acting on each individual entry.

There is a stark geometric difference between these two observation models. In unquantized compressed sensing, each measurement determines a hyperplane in which $\boldsymbol{x}$ must reside. In 1-bit compressed sensing, each measurement determines a hyperplane, but now we are only told which side of the hyperplane $\boldsymbol{x}$ resides on (see Figure 1).

Do the 1-bit measurements contain sufficient information to reconstruct $\boldsymbol{x}$? Clearly, exact reconstruction is impossible because the measurements only give a finite number of bits of information and the signal lies in an infinite set. Furthermore, the measurements retrieve *no information* about the norm of $\boldsymbol{x}$. Thus, we may only hope to approximate the direction of $\boldsymbol{x}$. Equivalently, we assume that $\boldsymbol{x} \in S^{n-1}$ and endeavor to approximate $\boldsymbol{x}$ itself.

A natural method to reconstruct $\boldsymbol{x}$ is to find a vector that matches the data and has the required structure:

$$\text{Find } \boldsymbol{x}' \quad \text{such that} \quad \|\boldsymbol{x}'\|_0 \le s, \|\boldsymbol{x}'\|_2 = 1 \quad \text{and} \quad \operatorname{sign}(\boldsymbol{A}\boldsymbol{x}') = y. \tag{1}$$

This program has recently been shown to give nearly optimal accuracy. If $\boldsymbol{A}$ is a Gaussian matrix, Jacques et al. [10] show that $O(\delta^{-1} s \log(n/\delta))$ measurements are sufficient to reconstruct $\boldsymbol{x}$ with ℓ_2 error at most δ. Aside from the logarithmic factor, Theorem 1 in [10] shows that this error bound in nearly minimax. It is further shown that a variation on this program provides stability to adversarial noise. Yet there still remain important challenges because the above program contains two nonconvex constraints: $\|\boldsymbol{x}\|_0 \le s$ and $\|\boldsymbol{x}\|_2 = 1$. Thus, there is no known algorithm that is guaranteed to return the solution to the above program in polynomial time.

In order to give a polynomial-time solver, Plan and Vershynin [17] propose a convex programming approach:

$$\min_{\boldsymbol{x}'} \|\boldsymbol{x}'\|_1 \quad \text{such that} \quad \|\boldsymbol{A}\boldsymbol{x}'\|_1 = 1 \quad \text{and} \quad \operatorname{sign}(\boldsymbol{A}\boldsymbol{x}') = y. \tag{2}$$

Above, $\|\boldsymbol{A}\boldsymbol{x}\|_1 = \sum_i |\langle \boldsymbol{a}_i, \boldsymbol{x}\rangle| = \sum_i y_i \langle \boldsymbol{a}_i, \boldsymbol{x}\rangle$ is a linear constraint; in fact, the program can be recast as a linear program. Let $\hat{\boldsymbol{x}}$ be the solution to the above program. Theorem 1.1 in [17] shows that

$$\left\| \frac{\hat{\boldsymbol{x}}}{\|\hat{\boldsymbol{x}}\|_2} - \boldsymbol{x} \right\|_2 \le \delta$$

with high probability provided that $m \ge O(\delta^{-5} s \log^2(n/s))$. We leverage recent results on discrete embeddings [16] to give a slight refinement of this result.

Theorem 1. *Let $s \le m \le n$. Let $\boldsymbol{A}$ have i.i.d. standard normal entries. Suppose that*

$$m \ge C\delta^{-4} s \log^2(n/s).$$

Then, with probability at least $1 - C_1 \exp(-c\delta m)$ the following holds uniformly over all signals $\boldsymbol{x} \in \mathbb{R}^n$ satisfying $\|\boldsymbol{x}\|_1 \le \sqrt{s}$, $\|\boldsymbol{x}\|_2 = 1$. Let $\boldsymbol{y} = \operatorname{sign}(\boldsymbol{A}\boldsymbol{x})$. Then the solution $\hat{\boldsymbol{x}}$ to the linear program (2) *satisfies*

$$\left\| \frac{\hat{\boldsymbol{x}}}{\|\hat{\boldsymbol{x}}\|_2} - \boldsymbol{x} \right\|_2 \le \delta.$$

Above, and in what follows, C and c are absolute numeric constants.

Proof: Proceed as in the proof of Theorem 1.1 in [17], but replace Theorem 2.1 in [17] with Theorem 3.1 in [16]. ■

Remark 1 (Soft sparsity). The assumption that $\|\boldsymbol{x}\|_1 \le \sqrt{s}$ is a relaxation of the exact sparsity constraint $\|\boldsymbol{x}\|_0 \le s$. Indeed, suppose that $\|\boldsymbol{x}\|_0 \le s$. Then by the Cauchy-Schwarz inequality,

$$\|\boldsymbol{x}\|_1 \le \sqrt{\|\boldsymbol{x}\|_0} \cdot \|\boldsymbol{x}\|_2 = \sqrt{\|\boldsymbol{x}\|_0} \le \sqrt{s}.$$

However, the constraint $\|\boldsymbol{x}\|_1 \le \sqrt{s}$ allows for $\boldsymbol{x}$ to be *compressible* instead of exactly sparse—it only requires a fast decay rate of the entries of $\boldsymbol{x}$.

Remark 2 (Optimality and δ dependence). Up to the power of 2 on the logarithm, the number of measurements required for a *fixed* level of accuracy matches what is needed for *unquantized* compressed sensing, and also matches the error bound achieved by the non-convex program (1). Let us also consider the dependence of m on δ and compare to the solution to the non-convex program. If $\hat{\boldsymbol{x}}$ is the solution to (1) the number of measurements required is essentially proportional to δ^{-1}. If $\hat{\boldsymbol{x}}$ is the solution to the convex program (2), Theorem 1 requires m to be proportional to δ^{-4}. On one hand, the former requires exact sparsity while the latter softens this requirement. Further, as shown in [15], in the noisy problem and with *soft sparsity* the δ^{-4} dependence is sometimes optimal. Nevertheless, in the noiseless problem it is an open problem whether the δ dependence can be improved for an efficient solver.

For adaptive approaches to 1-bit compressed sensing with impressive reconstruction guarantees, see [8], [9].

III. Noisy 1-bit compressed sensing

In noisy 1-bit compressed sensing, the data takes the form

$$\boldsymbol{y} = \operatorname{sign}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{z}) \tag{3}$$

where $\boldsymbol{z}$ is a noise term with i.i.d. entries.

In order to reconstruct $\boldsymbol{x}$, one would like to soften the constraint, $\operatorname{sign}(\boldsymbol{A}\boldsymbol{x}) = y$ used in the noiseless problem. A natural way to do this would be to bound the Hamming distance between $\operatorname{sign}(\boldsymbol{A}\boldsymbol{x})$ and $\boldsymbol{y}$. Unfortunately, this would give a non-convex constraint. Thus, Plan and Vershynin [15] suggest a different convex program to estimate $\boldsymbol{x}$:

$$\max_{\boldsymbol{x}'} \sum_i y_i \langle \boldsymbol{a}_i, \boldsymbol{x}'\rangle \quad \text{such that} \quad \|\boldsymbol{x}'\|_2 \le 1, \|\boldsymbol{x}'\|_1 \le \sqrt{s}. \tag{4}$$

The solution enjoys a high level of accuracy.

Theorem 2 ([15], Corollary 3.1)**.** *Fix $\boldsymbol{x} \in S^{n-1}$ satisfying $\|\boldsymbol{x}\|_1 \le \sqrt{s}$. Let $\boldsymbol{A}$ have independent standard normal entries. Let $\boldsymbol{y} = \operatorname{sign}(\boldsymbol{A}\boldsymbol{x} + \boldsymbol{z})$ and suppose that $\boldsymbol{z}$ is a Gaussian noise vector with independent $N(0, \sigma^2)$ entries. Let $\delta > 0$ and suppose that*

$$m \ge C\delta^{-4}(\sigma^2 + 1) s \log(2n/s).$$

Then, with probability at least $1 - 8\exp(-c\delta^4 m)$, the solution $\hat{\boldsymbol{x}}$ to the convex program (4) *satisfies*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \le \delta.$$

In contrast to Theorem 1, this is a non-uniform result—it holds for one fixed $\boldsymbol{x}$ with a random draw of $\boldsymbol{A}$. See Theorem 1.3 in [15] for a uniform result which also considers adversarial noise and for the treatment of much more general signal structures outside of sparsity. We also note that when $\sigma > 1$ this error bound nearly matches the minimax error bound achievable by any estimator from *unquantized measurements*. See [18, Theorem 1] and [15, Section 3]. This has the following implication: *When the signal-to-noise ratio is low, 1-bit measurements contain almost as much information* *as unquantized measurements.* The preceding theoretical conclusion is backed up with numerical evidence in [12].

A. Sparse binary regression

Sparse binary regression, and in particular sparse logistic regression, are often used to explain statistical data in which the response variable is binary. It is common to assume that the data is generated according to the *generalized linear model:* $y_i \in \{+1, -1\}$ is a Bernoulli random variable satisfying

$$\mathbb{E}\, y_i = \theta(\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle) \tag{5}$$

for some function $\theta : \mathbb{R} \to [0, 1]$. Note that this model implies that

$$P(y_i = 1) = \frac{\theta(\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle) + 1}{2} =: f(\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle).$$

Thus, the noisy 1-bit compressed sensing model 3 can always be recast using the generalized linear model by taking $f(t) := P(z_i \geq -t)$. The two are equivalent as long as $1 - f$ is a distribution function.

There are a number of theoretical results in sparse binary regression, focusing on sparse logistic regression [3], [5], [11], [13], [14], [19], [20]. A main message is that $O(\delta^{-2} s \log(n))$ measurements are sufficient to reconstruct $\boldsymbol{x}$ up to error δ by using ℓ_1-penalized maximum likelihood estimation [14]. Interestingly, these results allow for the reconstruction of both the direction, and norm of $\boldsymbol{x}$. However, there are two limitations to this maximum-likelihood-based approach: 1) knowledge of the function θ defining the generalized linear model is imperative, and 2) as the norm of $\boldsymbol{x}$ increases, the negative log-likelihood loses the strong convexity needed in the theoretical treatment.

The ideas from 1-bit compressed sensing allow us to overcome these two limitations. Indeed, the solution to (4) remains accurate for nearly any generalized linear model, but knowledge of the function θ is unnecessary in the reconstruction of $\boldsymbol{x}$. To make this precise, define

$$\lambda := \mathbb{E}\, g\, \theta(g)$$

where $g \sim N(0, 1)$. λ gives a notion of how correlated the response y is with the linear functionals $\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle$. Higher correlation improves reconstruction. For example, if f is the logistic function, then $\lambda \approx 0.41$.

Theorem 3 ([15], Corollary 3.1)**.** *Fix $\boldsymbol{x} \in S^{n-1}$ satisfying $\|\boldsymbol{x}\|_1 \leq \sqrt{s}$. Let $\boldsymbol{A}$ have independent standard normal entries and suppose that $\boldsymbol{y}$ follows the generalized linear model* (5)*. Let $\delta > 0$ and suppose that*

$$m \geq C\delta^{-4}\lambda^{-2} s \log(2n/s).$$

Then, with probability at least $1 - 8\exp(-c\delta^4 m)$, the solution $\hat{\boldsymbol{x}}$ to the convex program (4) *satisfies*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \delta.$$

Remark 3. The assumption that $\|\boldsymbol{x}\|_2 = 1$ in the theorem is really no assumption at all, since the norm of $\boldsymbol{x}$ may be absorbed into the definition of θ simply by rescaling the function. Further, suppose the following two mild assumptions on the model: 1) θ is monotonically increasing and 2) $\theta(0) = 0$. Then for a positive scalar t and standard normal random variable g, $\mathbb{E}\, \theta(tg)g$ is an increasing function of t. The implication is that rescaling θ, to absorb the norm of $\boldsymbol{x}$ causes λ to *increase* as long as $\|\boldsymbol{x}\|_2 \geq 1$. Thus, the reconstruction of the direction of $\boldsymbol{x}$ only improves as the magnitude of $\boldsymbol{x}$ increases. This contrasts with the maximum-likelihood approach discussed above.

B. sub-gaussian measurements

Up until now, we have considered Gaussian measurement vectors. One may ask whether 1-bit compressed sensing is possible with other random measurement schemes.

Let us consider Bernoulli measurement vectors in which each entry of $\boldsymbol{a}_i$ is an independent Bernoulli random variable, so that $\boldsymbol{a}_i \in \{+1, -1\}^n$. It is well known [7] that measurements of this form lead to near-optimal results in unquantized compressed sensing. Does the same hold true for 1-bit compressed sensing?

Consider two candidate signals $\boldsymbol{x} = (1, 0, 0, \ldots, 0)$ and $\bar{\boldsymbol{x}} = (1, 0.9, 0, 0, \ldots, 0)$. Then one has $\mathrm{sign}(\langle \boldsymbol{a}_i, \boldsymbol{x} \rangle) = \mathrm{sign}(\langle \boldsymbol{a}_i, \bar{\boldsymbol{x}} \rangle)$ *deterministically*. Thus, when the measurement vectors are Bernoulli, $\boldsymbol{x}$ and $\bar{\boldsymbol{x}}$ are indistinguishable, and reconstruction of either is ill-posed.

However, it turns out that the above negative example is atypical. Signal reconstruction with Bernoulli measurements is possible—as long as the signal is not *too sparse*. This will be quantified by a bound on the maximum entry of $\boldsymbol{x}$ in Theorem 4 below.

Recall that a random variable η is called sub-gaussian if it has a sub-gaussian tail: $\mathbb{P}\left(\eta > t\right) \leq Ce^{-ct^2}$. Recall also that Bernoulli random variables are sub-gaussian.

Theorem 4 ([2], Theorem 1.3)**.** *Fix $\boldsymbol{x} \in S^{n-1}$ satisfying $\|\boldsymbol{x}\|_1 \leq \sqrt{s}$. Let a be a symmetric, sub-gaussian random variable with unit variance. Let $\boldsymbol{A}$ be generated with coordinates that are independent copies of a. Assume that $\boldsymbol{y}$ follows the generalized linear model* (5) *and the first three derivatives of θ are bounded. Suppose*

$$m \geq C\delta^{-4}\lambda^{-2} s \log(n/s)$$

and let $\hat{\boldsymbol{x}}$ be the solution to the convex program (4)*. Then with probability at least $1 - 4\exp(-c\delta^4 m)$*

$$\|\hat{\boldsymbol{x}} - \boldsymbol{x}\|_2 \leq \delta + C\left(\frac{\|\boldsymbol{x}\|_\infty}{\lambda^3}\right)^{1/4}.$$

For a more precise treatment which also allows for θ to be the discontinuous sign function, see [2]. We note that this theorem allows no correlation between the entries of $\boldsymbol{A}$. For a treatment of the case when each row of $\boldsymbol{A}$ is Gaussian with correlations between entries see Section 3.4 in [15].

C. General signal structures

While sparse signal structures are intrinsic in the sparse binary regression model and in some compressed sensing problems, it is often of interest to consider more general signal structures. As a common example, $\boldsymbol{x}$ may not be sparse itself, but it may be sparse in a known dictionary, so that $\boldsymbol{x} = \boldsymbol{D}\boldsymbol{v}$ for a sparse vector v. Alternatively, $\boldsymbol{x}$ itself could be a matrix with low rank.

In general, the signal structure of $\boldsymbol{x}$ is defined by knowledge of a set K to which $\boldsymbol{x}$ belongs. Since we assume that $\|\boldsymbol{x}\|_2 = 1$, we may also assume that $K \subset B_2$ where B_2 is the unit ball. In this case, Vershynin and Plan [15] suggest to take the estimate of $\boldsymbol{x}$ to be the solution to the following program.

$$\max_{\boldsymbol{x}'} \sum_{i=1}^{m} y_i \langle \boldsymbol{a}_i, \boldsymbol{x}' \rangle \quad \text{such that} \quad \boldsymbol{x}' \in K. \tag{6}$$

For example, we may take $K = B_2 \cap \sqrt{s} B_1^n$ where B_1^n is the ℓ_1 ball. This recovers the convex program (4).

In this general case, reconstruction of $\boldsymbol{x}$ to accuracy δ requires $O(\delta^{-4} w(K)^2)$ binary measurements, where $w(K)$ is the *Gaussian mean width* of K. See [15] for details.

IV. Binary matrix completion

In a complementary line of research, Davenport et al. [6] analyze the following problem. Suppose that you see a subset of entries of a binary matrix, i.e., a matrix filled with ± 1 entries. From the observed entries, what information can be determined about unobserved entries? Problems of this nature arise in various applications. Consider for example the voting history of US senators on a number of bills, but with missing votes when a senator is out of town; or consider binary recommendation systems such as Pandora, in which one wishes to recommend unrated songs based on observed user ratings. For more applications see [6]. Davenport et al. assume that the data follows from the generalized linear model, but with three large differences from the considerations of the previous sections: 1) the measurements only give information about single entries of the matrix, 2) a low-rank structure is assumed in place of a sparse structure, and 3) θ (from Equation (5)) is assumed to be known. Under these assumptions, the authors show that nuclear-norm constrained maximum likelihood estimation gives minimax optimal reconstruction of the probability distribution of the unseen entries.

V. Conclusion

Binary data is intrinsic to many naturally arising inverse problems, and also arises in extreme quantization. But the signal or model that is to be reconstructed often comes from an infinite, albeit low-dimensional, set. This blend of continuous and discrete leads to interesting challenges in developing and analyzing accurate methods of signal reconstruction. We reviewed a number of recent results, which show that 1-bit measurements can give comparable information to unquantized measurements. Further, the methods of 1-bit compressed sensing allow for a semi-parametric treatment of sparse binary regression. One question that arises naturally from the above work is whether we can give a semi-parametric treatment of binary matrix completion which does not assume knowledge of θ.

Acknowledgment

The author would like to thank Roman Vershynin, Mark Davenport, Ewout Vandenberg, Mary Wootters, Albert Ai, and Alex Lapanowski for beneficial discussions. Most of the work reviewed in this paper is based on joint work with the above authors.

References

[1] One-bit compressive sampling webpage. http://dsp.rice.edu/1bitCS/.

[2] A. Ai, A. Lapanowski, Y. Plan, and R. Vershynin. One-bit compressed sensing with non-gaussian measurements. Submitted. Available at http://arxiv.org/abs/1208.6279.

[3] F. Bach. Self-concordant analysis for logistic regression. *Electronic Journal of Statistics*, 4:384–414, 2010.

[4] P. Boufounos and R. Baraniuk. 1-Bit compressive sensing. In *42nd Annual Conference on Information Sciences and Systems (CISS)*, March 2008.

[5] F. Bunea. Honest variable selection in linear and logistic regression models via ℓ_1 and $\ell_1 + \ell_2$ penalization. *Electronic Journal of Statistics*, 2:1153–1194, 2008.

[6] M. Davenport, Y. Plan, E. van den Berg, and M. Wootters. One-bit matrix completion. Submitted. Available at http://arxiv.org/abs/1209.3672.

[7] Y. Eldar and G. Kutyniok, editors. *Compressed Sensing*. Cambridge University Press, 2012.

[8] C. Gunturk, M. Lammers, A. Powell, R. Saab, and O. Ylmaz. Sigma delta quantization for compressed sensing. In *44th Annual Conference on Information Sciences and Systems (CISS)*. IEEE, 2010.

[9] A. Gupta, R. Nowak, and B. Recht. Sample complexity for 1-bit compressed sensing and sparse classification. In *International Symposium on Information Theory (ISIT)*. IEEE, 2010.

[10] L. Jacques, J. Laska, P. Boufounos, and R. Baraniuk. Robust 1-bit compressive sensing via binary stable embeddings of sparse vectors. 2011. Preprint. Available at http://arxiv.org/abs/1104.3160.

[11] S.M. Kakade, O. Shamir, K. Sridharan, and A. Tewari. Learning exponential families in high-dimensions: Strong convexity and sparsity. 2009. Preprint. Available at http://arxiv.org/abs/0911.0054.

[12] J. Laska and R. Baraniuk. Regime change: Bit-depth versus measurement-rate in compressive sensing. 2011. Preprint. Available at http://arxiv.org/abs/1110.3450.

[13] L. Meier, S. Van De Geer, and P. Bühlmann. The group lasso for logistic regression. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 70(1):53–71, 2008.

[14] S. Negahban, P. Ravikumar, M.J. Wainwright, and B. Yu. A unified framework for high-dimensional analysis of M-estimators with decomposable regularizers. 2010. Preprint. Available at http://arxiv.org/abs/1010.2731.

[15] Y. Plan and R. Vershynin. Robust 1-bit compressed sensing and sparse logistic regression: A convex programming approach. *IEEE Transactions on Information Theory*. Accepted for publication, 2012.

[16] Y. Plan and R. Vershynin. Dimension reduction by random hyperplane tessellations. 2011. Submitted. Available at http://arxiv.org/abs/1111.4452.

[17] Y. Plan and R. Vershynin. One-bit compressed sensing by linear programming. *Communications on Pure and Applied Math*, 2011. Accepted for publication. Available at http://arxiv.org/abs/1109.4299.

[18] G. Raskutti, M.J. Wainwright, and B. Yu. Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *Information Theory, IEEE Transactions on*, 57(10):6976–6994, 2011.

[19] P. Ravikumar, M.J. Wainwright, and J.D. Lafferty. High-dimensional ising model selection using 1-regularized logistic regression. *The Annals of Statistics*, 38(3):1287–1319, 2010.

[20] S.A. Van De Geer. High-dimensional generalized linear models and the lasso. *The Annals of Statistics*, 36(2):614–645, 2008.