Biometrika (2011), **98**, 2, *pp.* 371–380 doi: 10.1093/biomet/asr010

Printed in Great Britain

Sure independence screening and compressed random sensing

BY LINGZHOU XUE AND HUI ZOU

School of Statistics, University of Minnesota, Minneapolis, Minnesota 55455, U.S.A.
lzxue@stat.umn.edu hzou@stat.umn.edu

SUMMARY

Compressed sensing is a very powerful and popular tool for sparse recovery of high dimensional signals. Random sensing matrices are often employed in compressed sensing. In this paper we introduce a new method named aggressive betting using sure independence screening for sparse noiseless signal recovery. The proposal exploits the randomness structure of random sensing matrices to greatly boost computation speed. When using sub-Gaussian sensing matrices, which include the Gaussian and Bernoulli sensing matrices as special cases, our proposal has the exact recovery property with overwhelming probability. We also consider sparse recovery with noise and explicitly reveal the impact of noise-to-signal ratio on the probability of sure screening.

Some key words: Compressed sensing; Random matrix; Sparse recovery; Sub-Gaussian random variable; Sure independence screening.

1. INTRODUCTION

A system of underdetermined linear equations has fewer observations than unknowns. In general, solving an underdetermined linear system is an ill-posed problem, as there are infinitely many solutions. However, underdetermined linear systems have frequently appeared in important engineering problems such as signal/image processing and linear decoding. These applications require sparse solutions of such systems because only a few unknown elements are important. Consider an underdetermined linear equations system $y_{n\times 1} = \Phi_{n\times p} x_{p\times 1}$ where $p \gg n$ and Φ is referred to as the sensing matrix. The sparsest solution is given by

$$P_0: \quad \min \|x\|_{\ell_0} \quad \text{subject to} \quad y = \Phi x,$$

where $\|x\|_0$ is the ℓ_0 norm of x and equals $\sum_{j=1}^{p} I(x_j \neq 0)$. Unfortunately, solving P_0 is computationally infeasible. To overcome the computational difficulty, Donoho (2006) and Candès & Tao (2006) introduced compressed sensing that solves the following ℓ_1 minimization problem

$$P_1: \quad \min \|x\|_{\ell_1} \quad \text{subject to} \quad y = \Phi x,$$

where $\|x\|_{\ell_1} = \sum_{j=1}^{p} |x_j|$ is the ℓ_1 norm of x. Since its introduction, compressed sensing has been attracting growing attention owing to its excellent properties. First, compressed sensing naturally produces a sparse solution via linear programming and thus is computationally manageable. The second, and perhaps the most important, property of compressed sensing is that the solution to P_1 is identical to the solution to P_0, if the sensing matrix Φ satisfies certain conditions (Donoho & Huo, 2001; Candès & Tao, 2005, 2006; Donoho, 2006; Donoho & Elad, 2006).

A popular theoretical framework for studying the exact recovery property of compressed sensing is the restricted isometry property introduced by Candès & Tao (2005). Given an index subset $\mathcal{M} \subseteq \{1, \ldots, p\}$, write $\Phi_{\mathcal{M}} = [\cdots \phi_j \cdots]$, $j \in \mathcal{M}$. The S-restricted isometry constant of Φ, denoted by δ_S, is the smallest quantity such that $(1 - \delta_S)\|c\|_2^2 \leqslant \|\Phi_{\mathcal{M}} c\|_2^2 \leqslant (1 + \delta_S)\|c\|_2^2$ holds for all index subsets $\mathcal{M}$ of cardinality at most S and all real vectors c. Candès & Tao (2005) proved that any s-sparse signal can be recovered exactly by compressed sensing, if the sensing matrix Φ satisfies $\delta_{3s} + 3\delta_{4s} < 2$. Several papers have discussed weaker restricted isometry conditions for exact recovery. See, for example, Candès (2008), Foucart & Lai (2009), Cai et al. (2010) and references therein. However, it is still very hard to verify the restricted isometry property for a given sensing matrix Φ. Typically, certain random sensing matrices are used in compressed sensing such that they satisfy the restricted isometry property with overwhelming probability according to random matrix theory (Candès & Tao, 2005). Widely used examples include Gaussian and Bernoulli sensing matrices; Φ is a Gaussian sensing matrix if the ϕ_{ij} are independent normal variables with zero mean and variance n^{-1}; and Φ is a Bernoulli sensing matrix if the ϕ_{ij} are independent Bernoulli variables such that $\text{pr}(\phi_{ij} = n^{-1/2}) = \text{pr}(\phi_{ij} = -n^{-1/2}) = 0{\cdot}5$.

Since random sensing matrices are often employed in compressed sensing, one may ask whether there is an alternative sparse recovery algorithm specifically designed to work well with random sensing matrices. Current implementations of compressed sensing such as ℓ_1-MAGIC (Candès & Romberg, 2005) could be very inefficient with a large random sensing matrix; see § 2. In such situations, the new sparse recovery algorithm should fully exploit the randomness to boost computation speed.

In the present paper, we propose a new computationally efficient method for finding the sparsest solution defined in P_0. Our proposal is based on the following fact. Suppose that x is the sparse signal generating y. Define the support set T as $T = \{j : x_j \neq 0\}$. An index set $\mathcal{M}$ of size n is called a secure bet if $T \subset \mathcal{M}$. Observe that if $\mathcal{M}$ is a secure bet, then $x_j = 0$ for all $j \in \mathcal{M}^c$ where $\Phi_{\mathcal{M}} = [\cdots \phi_j \cdots]$, $j \in \mathcal{M}$ and $x_{\mathcal{M}} = (\ldots, x_j, \ldots)^{\mathrm{T}}$, $j \in \mathcal{M}$. Thus, we can write $y = \Phi_{\mathcal{M}} x_{\mathcal{M}}$. Furthermore, $\Phi_{\mathcal{M}}$ is an $n \times n$ matrix and is invertible almost surely if Φ is a random sensing matrix. If $Z = \Phi_{\mathcal{M}}^{-1} y$, then Z is exactly equal to $x_{\mathcal{M}}$. The nonzero components of Z directly reveal x_T and T. Therefore, the sparse recovery problem is equivalent to finding a secure bet.

Finding a secure bet is very difficult in general. Fortunately, for a wide class of random sensing matrices, there is an algorithm that can find a secure bet with overwhelming probability. Specifically, let $\omega = (\omega_1, \ldots, \omega_p)^{\mathrm{T}}$ and $\omega_i = y^{\mathrm{T}} \phi_i$, where ϕ_i is the ith column of Φ. The aggressive betting index set is defined as

$$\mathcal{M}_* = \{1 \leqslant i \leqslant p : |\omega_i| \text{ is among the first } n \text{ largest of all elements of } |\omega|\}.$$

To summarize, we solve underdetermined linear equations in two steps:

Step 1: Find the index set $\mathcal{M}_*$ defined in (2).

Step 2: Compute $Z = \Phi_{\mathcal{M}_*}^{-1} y$ and then set $\tilde{x}_{\mathcal{M}_*} = Z$ and $\tilde{x}_{\mathcal{M}_*^c} = 0$.

We name the above algorithm aggressive betting using sure independence screening, because the marginal correlation learning idea has been studied and named sure independence screening by Fan & Lv (2008), Fan et al. (2008) and Fan & Song (2010).

2. Theoretical and numerical properties of aggressive betting

2·1. *Theoretical results*

In this section we study the exact recovery property of aggressive betting using sure independence screening with a class of random sensing matrices. In what follows, let x be the sparse signal and s be the cardinality of the support set T. Let $\tilde{x}$ be the aggressive betting solution to the underdetermined linear equations. Define the recovered support set as $\tilde{T} = \{j : \tilde{x}_j \neq 0\}$. Exact recovery means that $\tilde{T} = T$ and $\tilde{x}_T = x_T$.

Our theory covers a wide class of random sensing matrices whose entries are independent sub-Gaussian random variables. A random variable z is sub-Gaussian if $E\{\exp(tz)\} \leqslant \exp(ct^2)$ for some $c > 0$ (Buldygin & Kozachenko, 1980).

Definition 1. *The $n \times p$ matrix Φ is a sub-Gaussian sensing matrix with a scale factor σ if its entries are independent random variables with zero mean and variance v and $E(e^{t\phi_{ij}}) \leqslant e^{t^2\sigma^2/2}$.*

It is easy to verify that both Gaussian and Bernoulli sensing matrices are sub-Gaussian sensing matrices. Without loss of generality, assume $v = n^{-1}$. If not, simply consider the transformation $y \leftarrow y(nv)^{-1/2}$ and $\phi_{ij} \leftarrow \phi_{ij}(nv)^{-1/2}$. The linear equations stay unchanged and the variance is n^{-1}. To see the intuition behind aggressive betting, observe that $E(\omega_j) = E(y^{\mathrm{T}}\phi_j) = x_j$, which is clearly a strong motivation for us to consider using the marginal correlation ranking idea to find a secure bet.

Theorem 1. *Let Φ be a sub-Gaussian sensing matrix with a scale factor σ and $E(\phi_{ij}^2) = n^{-1}$. Take any constant ϵ in the range $(0, 0{\cdot}5)$ and assume that $\sigma^2 = dn^{-1}$ and $d > 2^{-7/2}$. Then aggressive betting using sure independence screening recovers T and x_T with probability at least $1 - se^{-c_1 n} - 2pe^{-c_2\rho n/s}$, where $\rho = (\min_{j\in T} x_j^2)(\sum_{j\in T} x_j^2/s)^{-1}$ and*

$$c_1 = 2^{-9/2}d^{-2}(\epsilon - 0{\cdot}5)^2, \quad c_2 = 0{\cdot}25\{(1 + 4d^{-2}\epsilon^2)^{1/2} - 1\}.$$

Consider the asymptotic set-up where s and n vary with p such that $s < n$ and $p \to \infty$. Then Theorem 1 says that $n \gg c_2^{-1}\rho^{-1}s\log(p)$ is sufficient for consistent recovery. As pointed out by a referee, this asymptotic lower bound is comparable with that obtained in Fletcher et al. (2009) for compressed sensing using a Gaussian sensing matrix. Fletcher et al. (2009) considered $\mathcal{MC} = \{1 \leqslant i \leqslant p : |\omega_i|$ is among the first s largest of all elements of $|\omega|\}$ and showed that $\mathrm{pr}(\mathcal{MC} = T) \to 1$ if $n \gg 8\rho^{-1}s\log(p - s)$. If Φ is Gaussian we can let $d = 1$ and $c_2 = 0{\cdot}1$ in Theorem 1, so our asymptotic lower bound for n becomes $10\rho^{-1}s\log(p)$. There are two fundamental differences between our work and that of Fletcher et al. (2009). Firstly, the definition of $\mathcal{MC}$ requires knowing the value of s, which is unrealistic in practice. In other words, $\mathcal{MC}$ is not an estimator of T. We do not have such a problem in our approach. Secondly, Fletcher et al. (2009) considered only compressed sensing with a Gaussian sensing matrix and derived their asymptotic bound by using special properties of the normal distribution extensively. In contrast, our asymptotic lower bound for n holds for not only the Gaussian sensing matrix but also a much broader class of sub-Gaussian sensing matrices.

Compressed sensing has a nice universal property. If the sensing matrix satisfies a restricted isometry condition such as $\delta_s < 0{\cdot}307$ (Cai et al., 2010), then compressed sensing recovers all s sparse signals. Many stochastic sparse recovery algorithms, on the other hand, need conditions that depend on the sparse signal and hence aim to recover a fixed signal. For example, a key

Table 1. *A comparison of the computation speed of aggressive betting and compressed sensing using ℓ_1-MAGIC. We report the total time (in seconds) of three runs.*

Sensing matrix	(p, n, s)	Compressed sensing	Aggressive betting
Gaussian	$(2^{10}, 200, 2^3)$	5·27	0·19
Gaussian	$(10 \times 2^{11}, 800, 2^4)$	Not completed	4·83
Bernoulli	$(2^{10}, 200, 2^3)$	4·45	0·26
Bernoulli	$(10 \times 2^{11}, 800, 2^4)$	Not completed	7·17

assumption made in Wainwright (2009a) is that the design matrix satisfies an incoherence condition whose definition depends on the true support of the signal. Lv & Fan (2009) assume that the smallest nonzero component of the signal is greater than a certain threshold. In Fletcher et al. (2009) and Theorem 1, ρ plays a critical role. Asymptotically we should require $\rho \gg s \log(p)/n$. Although it is interesting theoretically to use a universal sensing matrix for all sparse recovery problems, in practice one should carefully choose Φ according to the signal to be recovered. Mairal et al. (2010) proposed ways to learn Φ in order to adapt it to specific data.

2·2. *Numerical results*

We conduct simulation experiments to examine the performance of aggressive betting using sure independence screening. The experimental procedure is similar to that in Candès & Tao (2005) and is described as follows.

(i) Select the length of input signal p, the sample size n and the size of support set s, $p > n > s$.
(ii) Generate the $n \times p$ sensing matrix Φ, either Gaussian or Bernoulli.
(iii) Pick a support set T of size s uniformly at random and generate a standard Gaussian vector of length s. Let x_T be the sign of the random Gaussian vector and $x_{T^c} = 0$.
(iv) Create the transformed signal $y = \Phi x$.
(v) Recover the signal $\tilde{x}$ through compressed sensing and aggressive betting.
(vi) Repeat steps (ii)–(v) 100 times.

To perform compressed sensing we used the collection of MATLAB routines ℓ_1-MAGIC (Candès & Romberg, 2005). In Step (ii) of aggressive betting, we used linsolve to solve the reduced linear equation system. We used two different combinations of (p, n, s). The results are recorded in Table 1. Both algorithms achieved perfect recovery in our experiments, so we focused on comparing the computation speed. All timings were carried out on a laptop with Intel Core2 Duo CPU 1·6 GHz. Table 1 shows that aggressive betting is at least 20 times faster than compressed sensing using ℓ_1-MAGIC. When $p = 10 \times 2^{11}$ and $n = 800$, we ran out of memory for using the ℓ_1-MAGIC package to perform compressed sensing and hence could not get the results. But it only took about two seconds to finish aggressive betting.

3. SURE INDEPENDENCE SCREENING AND ROBUST COMPRESSED SENSING

Candès et al. (2006) considered stable signal recovery from incomplete and inaccurate measurements and named it robust compressed sensing. Consider a contaminated linear system $y_{n\times 1} = \Phi_{n\times p} x_{p\times 1} + e_{n\times 1}$, where $p \gg n$ and $e = (e_1, \ldots, e_n)$ denotes the measurement error. Robust compressed sensing (Candès et al., 2006) finds the sparsest solution to the following ℓ_1

minimization problem,

$$P_2: \quad \min \|x\|_{\ell_1} \quad \text{subject to} \quad \|y - \Phi x\|_{\ell_2} \leqslant \nu,$$

where ν denotes the size of the error term e. Some other papers have also considered sparse recovery with noise data (Wainwright, 2009a,b; Fletcher et al., 2009; Lv & Fan, 2009).

The aggressive betting method does not work for the robust compressed sensing problem because $Z = \Phi^{-1}_{\mathcal{M}_*} y = x_{\mathcal{M}_*} + \Phi^{-1}_{\mathcal{M}_*} e$, which means we can no longer recover T and x_T simply by looking at Z. Following the spirit of sure independence screening (Fan and Lv, 2008), it is natural to perform sure independence screening to reduce the dimension and then do robust compressed sensing. The key argument is to establish the sure screening property, namely that $\mathcal{M}_*$ contains T with overwhelming probability.

THEOREM 2. *Consider the sub-Gaussian sensing matrix in Theorem* 1. *Let e be a vector of sub-Gaussian measurement errors with a common scale factor σ_e and $\sigma_e^2 = d_e n^{-1}$. Define $\kappa = d_e d^{-1} (\sum_{j \in T} x_j^2)^{-1}$. Take any constant ϵ in the range* $(0, 0{\cdot}5)$. *Then the sure screening property holds with probability at least* $1 - s e^{-c_1 n} - 2 p e^{-c_3 \rho n / s}$, *where* $c_3 = 0{\cdot}25[\{1 + 4\epsilon^2 d^{-2} (1 + \kappa)^{-1}\}^{1/2} - 1]$ *and* $c_1 = 2^{-9/2} d^{-2} (\epsilon - 0{\cdot}5)^2$.

Here we comment on some important technical differences between our analysis and that of Fan & Lv (2008). First, the two papers study different Φ matrices. In Fan & Lv (2008), Φ is the design matrix in which the rows of Φ are independent and identically distributed random vectors. Moreover, the theory of Fan & Lv (2008) requires that Φ satisfies a crucial concentration property which, as commented therein, is not easy to verify in general. Fan & Lv (2008) verified the concentration property when the rows of Φ are multivariate normal by using deep random matrix theory results. Fan & Lv's (2008) theory covers correlated normal sensing matrices. In our analysis the entries of Φ are independent but not necessarily identically distributed, and Φ may be non-Gaussian. Moreover, the proof of Theorem 2 only uses Chernoff bounds, without citing any advanced random matrix theory. Secondly, our result explicitly reveals the impact of the noise-to-signal ratio on sure screening. To be more specific, suppose Φ is a Gaussian sensing matrix and the error is also Gaussian. Then by definition κ is the noise-to-signal ratio. Obviously, increasing κ reduces c_3 and hence the probability lower bound in Theorem 2. By simple calculations, we see that $n \gg (8\kappa + 10)\rho^{-1} s \log(p)$ is sufficient to ensure that the sure screening property holds with high probability. In their discussion of Fan & Lv (2008), Levina & Zhu (2008) conducted some simulation to show how the noise-to-signal ratio influences the performance of sure independence screening. They wondered if the noise-to-signal ratio could be incorporated into the theory of sure independence screening explicitly. Our analysis provides some insight into their simulation findings.

We further present some numerical examples to support the use of sure independence screening together with robust compressed sensing. The simulation procedure is very similar to that in § 2·2 except that we used new mixed sensing matrices in which the odd columns are sampled independently from $N(0, n^{-1})$ and the even number columns are sampled independently from $\mathrm{Ber}(\pm n^{-1/2}, 0{\cdot}5)$. In our simulation the transformed signal y was generated by $y = \Phi x + e$ and $\sigma_e = 0{\cdot}005$. Following Candès et al. (2006), we let the positive constant ν in P_2 be $\nu^2 = n\sigma_e^2 + (8n)^{1/2}\sigma_e^2$ with the intuition that $\|e\|_{\ell_2}^2$ follows a chi-square distribution with mean $n\sigma_e^2$ and variance $2n\sigma_e^4$. Robust compressed sensing has been implemented by using the log-barrier method in the collection of MATLAB routines ℓ_1-MAGIC (Candès & Romberg, 2005), which we used in our experiments.

Table 2. *Comparison of robust compressed sensing and the new algorithm, which first conducts sure independence screening and then solves robust compressed sensing with the reduced dimension. The reported time (in seconds) is the total time of three runs.*

	$(2^{10}, 200, 2^3)$		$(10 \times 2^{11}, 800, 2^4)$	
	Time	Frequency of exact recovery	Time	Frequency of exact recovery
New algorithm	9·78	81·8	179·01	96·1
Robust compressed sensing	105·53	63·0	Not completed	Not completed

Table 2 shows that sure independence screening can greatly shorten the computing time and help improve the quality of the solution. The frequency of exact recovery is increased from 63 per cent to 81·8 per cent after using sure independence screening. For larger systems with $(p, n, s) = (10 \times 2^{11}, 800, 2^4)$ we could not do robust compressed sensing using ℓ_1-MAGIC. On the other hand, with the help of sure independence screening, the computation can be done rather quickly and the frequency of exact recovery is 96 percent.

ACKNOWLEDGEMENT

The authors thank the editor and referees for their helpful comments. This work was supported by a grant from the National Science Foundation, U.S.A.

APPENDIX

This section contains the proofs of Theorems 1 and 2. We first introduce a useful lemma.

LEMMA 1. *If η is a sub-Gaussian random variable with $E(\eta) = 0$ and $E\{\exp(t\eta)\} \leqslant \exp(t^2\tau^2/2)$, then*

$$E\{\exp(t\eta^2)\} \leqslant \begin{cases} \exp\{tE(\eta^2) + 2^{9/2}\tau^4 t^2\}, & t \in [-1/4\tau^2, 0), \\ (1 - 2t\tau^2)^{-1/2}, & t \in (0, 1/4\tau^2]. \end{cases}$$

Proof. Let $Z \sim N(0, 1)$. For any $t : t \in (0, 1/4\tau^2]$,

$$E\{\exp(t\eta^2)\} = E_\eta[E_Z\{\exp(2^{1/2}t^{1/2}Z\eta)\}] = E_Z[E_\eta\{\exp(2^{1/2}t^{1/2}Z\eta)\}].$$

Using the properties of sub-Gaussian distributions, we have

$$E_Z[E_\eta\{\exp(2^{1/2}t^{1/2}Z\eta)\}] \leqslant E_Z\{\exp(t\tau^2Z^2)\} = (1 - 2t\tau^2)^{-1/2}. \quad \text{(A1)}$$

For any $t \in [-1/4\tau^2, 0)$, we consider

$$\begin{aligned} E\{\exp(t\eta^2)\} &= \sum_{m=0}^{\infty} \frac{t^m E\{(\eta^2)^m\}}{m!} \\ &= 1 + tE(\eta^2) + \sum_{m=2}^{\infty} \frac{(4\tau^2 t)^m E\{(\eta^2/4\tau^2)^m\}}{m!} \\ &\leqslant 1 + tE(\eta^2) + (4\tau^2|t|)^2 E\{\exp(\eta^2/4\tau^2)\}. \quad \text{(A2)} \end{aligned}$$

By (A1) we have $E\{\exp(\eta^2/4\tau^2)\} \leqslant 2^{1/2}$, so (A2) yields

$$E\{\exp(t\eta^2)\} \leqslant 1 + tE(\eta^2) + 2^{9/2}\tau^4t^2 \leqslant \exp\{tE(\eta^2) + 2^{9/2}\tau^4t^2\}. \qquad \square$$

Proof of Theorem 1. Take a real number $0 < \epsilon < 0{\cdot}5$ and let $\delta = \epsilon \min_{j\in T} |x_j|$. Note that

$$\mathrm{pr}(\min_{i\in T} |\omega_i| > \delta) = 1 - \mathrm{pr}(\min_{i\in T} |\omega_i| \leqslant \delta) \geqslant 1 - \sum_{i\in T} \mathrm{pr}(|\omega_i| \leqslant \delta), \tag{A3}$$

$$\mathrm{pr}(|\omega_i| \leqslant \delta) \leqslant \mathrm{pr}(\omega_i \leqslant \delta) I(x_i > 0) + \mathrm{pr}(\omega_i \geqslant -\delta) I(x_i < 0). \tag{A4}$$

For each $i \in T$, $\omega_i = \sum_{j\in T} x_j \phi_j^{\mathrm{T}} \phi_i = x_i \phi_i^{\mathrm{T}} \phi_i + \sum_{j\in T\setminus\{i\}} x_j \phi_j^{\mathrm{T}} \phi_i$. Therefore

$$\mathrm{pr}(\omega_i \leqslant \delta) \leqslant \mathrm{pr}(x_i \phi_i^{\mathrm{T}} \phi_i \leqslant 2\delta) + \mathrm{pr}\left(\sum_{j\in T\setminus\{i\}} x_j \phi_j^{\mathrm{T}} \phi_i \leqslant -\delta \right), \tag{A5}$$

$$\mathrm{pr}(\omega_i \geqslant -\delta) \leqslant \mathrm{pr}(x_i \phi_i^{\mathrm{T}} \phi_i \geqslant -2\delta) + \mathrm{pr}\left(\sum_{j\in T\setminus\{i\}} x_j \phi_j^{\mathrm{T}} \phi_i \geqslant \delta \right). \tag{A6}$$

Combining (A3)–(A6), it follows that

$$\begin{aligned} \mathrm{pr}(\min_{i\in T} |\omega_i| > \delta) \geqslant 1 &- \sum_{i\in T} \mathrm{pr}(x_i \phi_i^{\mathrm{T}} \phi_i \leqslant 2\delta) I(x_i > 0) - \sum_{i\in T} \mathrm{pr}(x_i \phi_i^{\mathrm{T}} \phi_i \geqslant -2\delta) I(x_i < 0) \\ &- \sum_{i\in T} \mathrm{pr}\left(\left| \sum_{j\in T\setminus\{i\}} x_j \phi_j^{\mathrm{T}} \phi_i \right| \geqslant \delta \right). \end{aligned} \tag{A7}$$

We also have

$$\mathrm{pr}(\max_{i\in T^c} |\omega_i| < \delta) \geqslant 1 - \sum_{i\in T^c} \mathrm{pr}\left(\left| \sum_{j\in T} x_j \phi_j^{\mathrm{T}} \phi_i \right| \geqslant \delta \right). \tag{A8}$$

We bound each probability in (A7) and (A8) by repeatedly using Lemma 1 and Chernoff bounds.

Let $t_1 = -2^{-11/2} n^{-1} \sigma^{-4} x_i^{-2} (2\delta - x_i) > 0$ when $x_i > 0$. By $d > 2^{-9/2}$ and $\sigma^2 = dn^{-1}$ we check $-t_1 x_i > -1/4\sigma^2$. Thus, by Lemma 1 we have

$$\begin{aligned} \mathrm{pr}(x_i \phi_i^{\mathrm{T}} \phi_i \leqslant 2\delta) &\leqslant e^{2t_1\delta} \prod_{k=1}^{n} E\{\exp(-t_1 x_i \phi_{ik}^2)\} \leqslant \exp\{(2\delta - x_i) t_1 + n 2^{9/2} \sigma^4 x_i^2 t_1^2\} \\ &\leqslant \exp\left\{ -\frac{n}{2^{9/2} d^2} \left(\frac{1}{2} - \epsilon \right)^2 \right\}. \end{aligned} \tag{A9}$$

Similarly, when $x_i < 0$ we have

$$\mathrm{pr}(x_i \phi_i^{\mathrm{T}} \phi_i \geqslant -2\delta) \leqslant \exp\left\{ -\frac{n}{2^{9/2} d^2} \left(\frac{1}{2} - \epsilon \right)^2 \right\}. \tag{A10}$$

Let $\lambda = (\min_{j\in T} x_j^2)(\sum_{j\in T} x_j^2)^{-1}$. Define a nonnegative function $f(t) = (1+t)^{1/2} - \log\{1 + (1+t)^{1/2}\} + \log(2) - 1$ for $t > 0$. Denote $\tilde{Z} = \sum_{j\in T\setminus\{i\}} x_j \phi_j$, and $Z = \sum_{j\in T} x_j \phi_j$. Note that $\tilde{z}_k$ and z_k are sub-Gaussian random variables with $\tau = \sigma(\sum_{j\in T\setminus\{i\}} x_j^2)^{1/2}$ and $\tau = \sigma(\sum_{j\in T} x_j^2)^{1/2}$, respectively. Let

$t_2 = n\{-1 + 1 + 4\delta^2 d^{-2}/(\sum_{j \in T\setminus\{i\}} x_j^2)\}^{1/2}/2\delta$ and note that

$$
\begin{aligned}
\mathrm{pr}\left(\sum_{j\in T\setminus\{i\}} x_j \phi_j^{\mathrm{T}}\phi_i \geqslant \delta\right) &\leqslant e^{-t_2\delta}\prod_{k=1}^{n} E_{\tilde{z}_k}[E_{\phi_{ik}}\{\exp(t_2\tilde{z}_k\phi_{ik})\}] \\
&\leqslant e^{-t_2\delta}\prod_{k=1}^{n} E_{\tilde{z}_k}\left\{\exp\left(\frac{1}{2}t_2^2\sigma^2\tilde{z}_k^2\right)\right\} \\
&\leqslant e^{-t_2\delta}\left(1 - t_2^2\sigma^4 \sum_{j\in T\setminus\{i\}} x_j^2\right)^{-n/2} \\
&= \exp\left\{-\frac{n}{2} f\left(\frac{4\delta^2}{d^2\sum_{j\in T\setminus\{i\}} x_j^2}\right)\right\},
\end{aligned}
$$

where we have set $\sigma^2 = dn^{-1}$ in the last step. Likewise, we can show

$$
\mathrm{pr}\left(\sum_{j\in T\setminus\{i\}} x_j \phi_j^{\mathrm{T}}\phi_i \leqslant -\delta\right) \leqslant \exp\left\{-\frac{n}{2} f\left(\frac{4\delta^2}{d^2\sum_{j\in T\setminus\{i\}} x_j^2}\right)\right\}. \tag{A11}
$$

Hence, we have

$$
\mathrm{pr}\left(\left|\sum_{j\in T\setminus\{i\}} x_j \phi_j^{\mathrm{T}}\phi_i\right| \geqslant \delta\right) \leqslant 2\exp\left\{-\frac{n}{2} f\left(\frac{4\delta^2}{d^2\sum_{j\in T\setminus\{i\}} x_j^2}\right)\right\} \leqslant 2\exp\left\{-\frac{n}{2} f\left(\frac{4\epsilon^2\lambda}{d^2}\right)\right\}. \tag{A12}
$$

By the same arguments, we also obtain

$$
\mathrm{pr}\left(\left|\sum_{j\in T} x_j \phi_j^{\mathrm{T}}\phi_i\right| \geqslant \delta\right) \leqslant 2\exp\left\{-\frac{n}{2} f\left(\frac{4\epsilon^2\lambda}{d^2}\right)\right\}, \quad \text{for } i \in T^c. \tag{A13}
$$

From (A7) and (A9)–(A12) we see that

$$
\mathrm{pr}(\min_{i\in T} |\omega_i| > \delta) \geqslant 1 - s\exp\left\{-\frac{n\left(\epsilon - \frac{1}{2}\right)^2}{2^{9/2}d^2}\right\} - 2s\exp\left\{-\frac{n}{2} f\left(\frac{4\epsilon^2\lambda}{d^2}\right)\right\}. \tag{A14}
$$

From (A8) and (A13) we see

$$
\mathrm{pr}(\max_{i\in T^c} |\omega_i| < \delta) \geqslant 1 - 2(p-s)\exp\left\{-\frac{n}{2} f\left(\frac{4\epsilon^2\lambda}{d^2}\right)\right\}. \tag{A15}
$$

Therefore, by combining (A14) and (A15) we conclude that the probability of exact recovery is at least

$$
1 - s\exp\left\{-\frac{n(\epsilon - 1/2)^2}{2^{9/2}d^2}\right\} - 2p\exp\left\{-\frac{n}{2} f\left(\frac{4\epsilon^2\lambda}{d^2}\right)\right\}. \tag{A16}
$$

Finally, we can further simplify the exponent in the third term of (A16). Note that $f(0) = 0$ and $f'(t) = \{2 + 2(1+t)^{1/2}\}^{-1}$, so the mean value theorem gives $f(4\lambda\epsilon^2 d^{-2}) = 2\lambda\epsilon^2 d^{-2}\{1 + (1+t^*)^{1/2}\}^{-1}$ for some $0 < t^* < 4\lambda\epsilon^2 d^{-2} < 4\epsilon^2/d^2$. Thus, we have $2\lambda\epsilon^2 d^{-2}\{1 + (1 + 4\epsilon^2 d^{-2})^{1/2}\}^{-1} < f(4\lambda\epsilon^2 d^{-2}) < \lambda\epsilon^2 d^{-2}$. Note that $\lambda = \rho/s$. □

Proof of Theorem 2. Take $0<\epsilon<0{\cdot}5$ and let $\delta=\epsilon\min_{j\in T}|x_j|$. For each $i\in T$, $\omega_i=x_i\phi_i^{\mathrm{T}}\phi_i+\left(\sum_{j\in T\setminus\{i\}}x_j\phi_j^{\mathrm{T}}+e^{\mathrm{T}}\right)\phi_i$. Therefore, we have

$$\mathrm{pr}(\omega_i\leqslant\delta)\leqslant\mathrm{pr}(x_i\phi_i^{\mathrm{T}}\phi_i\leqslant 2\delta)+\mathrm{pr}\left\{\left(\sum_{j\in T\setminus\{i\}}x_j\phi_j^{\mathrm{T}}+e^{\mathrm{T}}\right)\phi_i\leqslant-\delta\right\},$$

$$\mathrm{pr}(\omega_i\geqslant-\delta)\leqslant\mathrm{pr}(x_i\phi_i^{\mathrm{T}}\phi_i\geqslant-2\delta)+\mathrm{pr}\left\{\left(\sum_{j\in T\setminus\{i\}}x_j\phi_j^{\mathrm{T}}+e^{\mathrm{T}}\right)\phi_i\geqslant\delta\right\}.$$

As in the proof of Theorem 1, we have

$$\mathrm{pr}(\min_{i\in T}|\omega_i|>\delta)\geqslant 1-s\exp\left\{-\frac{n}{2^{9/2}d^2}\left(\frac{1}{2}-\epsilon\right)^2\right\}-\sum_{i\in T}\mathrm{pr}\left\{\left|\left(\sum_{j\in T\setminus\{i\}}x_j\phi_j^{\mathrm{T}}+e^{\mathrm{T}}\right)\phi_i\right|\geqslant\delta\right\},$$

and also

$$\mathrm{pr}(\max_{i\in T^c}|\omega_i|<\delta)\geqslant 1-\sum_{i\in T^c}\mathrm{pr}\left\{\left|\left(\sum_{j\in T}x_j\phi_j^{\mathrm{T}}+e^{\mathrm{T}}\right)\phi_i\right|\geqslant\delta\right\}.$$

Let $\lambda=(\min_{j\in T}x_j^2)(\sum_{j\in T}x_j^2)^{-1}=\rho/s$, $\tilde{Z}=\sum_{j\in T\setminus\{i\}}x_j\phi_j+e$, and $Z=\sum_{j\in T}x_j\phi_j+e$. Note that $\tilde{z}_k$ and z_k are sub-Gaussian random variables with $\tau=(\sum_{j\in T\setminus\{i\}}\sigma^2x_j^2+\sigma_e^2)^{1/2}$ and $\tau=(\sum_{j\in T}\sigma^2x_j^2+\sigma_e^2)^{1/2}$, respectively. Let $t_3=n[-1+\{1+4\delta^2(\sum_{j\in T\setminus\{i\}}d^2x_j^2+dd_e)^{-1}\}^{1/2}]/2\delta$ and we can show

$$\mathrm{pr}\left\{\left(\sum_{j\in T\setminus\{i\}}x_j\phi_j^{\mathrm{T}}+e^{\mathrm{T}}\right)\phi_i\geqslant\delta\right\}\leqslant\exp\left\{-\frac{n}{2}f\left(\frac{4\delta^2}{\sum_{j\in T\setminus\{i\}}d^2x_j^2+dd_e}\right)\right\},$$

where we used the substitutions $\sigma^2=dn^{-1}$ and $\sigma_e^2=d_en^{-1}$. Likewise,

$$\mathrm{pr}\left\{\left(\sum_{j\in T\setminus\{i\}}x_j\phi_j^{\mathrm{T}}+e^{\mathrm{T}}\right)\phi_i\leqslant-\delta\right\}\leqslant\exp\left\{-\frac{n}{2}f\left(\frac{4\delta^2}{\sum_{j\in T\setminus\{i\}}d^2x_j^2+dd_e}\right)\right\}.$$

Similar to (A13) we can also obtain

$$\mathrm{pr}\left\{\left|\left(\sum_{j\in T}x_j\phi_j^{\mathrm{T}}+e^{\mathrm{T}}\right)\phi_i\right|\geqslant\delta\right\}\leqslant 2\exp\left[-\frac{n}{2}f\left\{\frac{4\epsilon^2\lambda}{d^2(1+\kappa)}\right\}\right],\quad\text{for } i\in T^c.$$

The remaining arguments are identical to those used in the proof of Theorem 1, so we omit them. □

References

Buldygin, V. V. & Kozachenko, Y. V. (1980). Sub-Gaussian random variable. *Ukrainskii Matematicheskii Zhurnal* **32**, 723–30.

Cai, T., Wang, L. & Xu, G. (2010). New bounds for restricted isometry constants. *IEEE Trans. Info. Theory* **56**, 4388–94.

Candès, E. (2008). The restricted isometry property and its implications for compressed sensing. *Compte Rendus de l' Academie des Sciences. Paris, France, ser. I,* **346**, 589–92.

Candès, E. & Romberg, J. (2005). ℓ_1-MAGIC : Recovery of sparse signals via convex programming. *Users Guide of the ℓ_1-MAGIC Toolbox for MATLAB* , http://www.acm.caltech.edu/l1magic/.

Candès, E., Romberg, J. & Tao, T. (2006). Stable signal recovery from incomplete and inaccurate measurements. *Commun. Pure Appl. Math.* **59**, 1207–23.

CANDÈS, E. & TAO, T. (2005). Decoding by linear programming. *IEEE Trans. Info. Theory* **51**, 4203–15.
CANDÈS, E. & TAO, T. (2006). Near-optimal signal recovery from random projections: universal encoding strategies? *IEEE Trans. Info. Theory* **52**, 5406–25.
DONOHO, D. (2006). Compressed sensing. *IEEE Trans. Info. Theory* **52**, 1289–306.
DONOHO, D. & ELAD, M. (2006). For most large underdetermined systems of equations, the minimal ℓ_1-norm solution is the sparsest solution. *Commun. Pure Appl. Math.* **59**, 797–829.
DONOHO, D. & HUO, X. (2001). Uncertainty principles and ideal atomic decomposition. *IEEE Trans. Info. Theory* **47**, 2845–62.
FAN, J. & LV, J. (2008). Sure independence screening for ultrahigh dimensional feature space (with discussion). *J. R. Statist. Soc.* B **70**, 849–911.
FAN, J., SAMWORTH, R. & WU, Y. (2008). Ultrahigh dimensional variable selection: beyond the linear model. *J. Mach. Learn. Res.* **10**, 1829–53.
FAN, J. & SONG, R. (2010). Sure independence screening in generalized linear models with NP-dimensionality. *Ann. Statist.* **38**, 3567–604.
FLETCHER, A., RANGAN, S. & GOYAL, V. (2009). Necessary and sufficient conditions for sparsity pattern recovery. *IEEE Trans. Info. Theory* **55**, 5758–72.
FOUCART, S. & LAI, M.-J. (2009). Sparsest solutions of underdetermined linear systems via ℓ_q-minimization for $0 < q \leqslant 1$. *Appl. Comp. Harm. Anal.* **26**, 395–407.
LEVINA, E. & ZHU, J. (2008). Discussion of "Sure independence screening for ultrahigh dimensional feature space". *J. R. Statist. Soc.* B **70**, 897–8.
LV, J. & FAN, Y. (2009). A unified approach to model selection and sparse recovery using regularized least squares. *Ann. Statist.* **37**, 3498–528.
MAIRAL, J., BACH, F., PONCE, J. & SAPIRO, G. (2010). Online learning for matrix factorization and sparse coding. *J. Mach. Learn. Res.* **11**, 19–60.
WAINWRIGHT, M. (2009a). Sharp thresholds for high-dimensional and noisy sparsity recovery using ℓ_1-constrained quadratic programming (Lasso). *IEEE Trans. Info. Theory* **55**, 2183–202.
WAINWRIGHT, M. (2009b). Information-theoretic limits on sparsity recovery in the high-dimensional and noisy setting. *IEEE Trans. Info. Theory* **55**, 5728–41.

[*Received June* 2010. *Revised January* 2011]