Electronic Journal of Statistics
Vol. 8 (2014) 3031–3061
ISSN: 1935-7524
DOI: 10.1214/15-EJS983

On higher order isotropy conditions and lower bounds for sparse quadratic forms*

Sara van de Geer and Alan Muro

Seminar for Statistics
ETH Zürich
Rämistrasse 101
8092 Zürich, Switzerland
e-mail: geer@stat.math.ethz.ch; muro@stat.math.ethz.ch

Abstract: This study aims at contributing to lower bounds for empirical compatibility constants or empirical restricted eigenvalues. This is of importance in compressed sensing and theory for ℓ_1-regularized estimators. Let X be an $n \times p$ data matrix with rows being independent copies of a p-dimensional random variable. Let $\hat{\Sigma} := X^T X/n$ be the inner product matrix. We show that the quadratic forms $u^T \hat{\Sigma} u$ are lower bounded by a value converging to one, uniformly over the set of vectors u with $u^T \Sigma_0 u$ equal to one and ℓ_1-norm at most M. Here $\Sigma_0 := \mathbb{E}\hat{\Sigma}$ is the theoretical inner product matrix, which we assume to exist. The constant M is required to be of small order $\sqrt{n/\log p}$. We assume moreover m-th order isotropy for some $m > 2$ and sub-exponential tails or moments up to order $\log p$ for the entries in X. As a consequence, we obtain convergence of the empirical compatibility constant to its theoretical counterpart, and similarly for the empirical restricted eigenvalue. If the data matrix X is first normalized so that its columns all have equal length we obtain lower bounds assuming only isotropy and no further moment conditions on its entries. The isotropy condition is shown to hold for certain martingale situations.

MSC 2010 subject classifications: 62J07.
Keywords and phrases: Compatibility, isotropy, quadratic forms, restricted eigenvalue, sparsity, transfer principle.

Received June 2014.

1. Introduction

Let X be an $n \times p$ data matrix with rows being i.i.d. copies of a random vector $X_0^T \in \mathbb{R}^p$. We consider the empirical inner product matrix $\hat{\Sigma} = X^T X/n$. For a vector $u \in \mathbb{R}^p$, let $\|u\|_q$ be its ℓ_q-norm ($1 \le q \le \infty$). We examine sparse quadratic forms $u^T \hat{\Sigma} u$, where u is sparse in the sense that $\|u\|_1 \le M$ for some constant $M \ge 1$. We will provide lower bounds for $\min\{u^T \hat{\Sigma} u : \ u^T \Sigma_0 u = 1, \ \|u\|_1 \le M\}$ with $\Sigma_0 := \mathbb{E}\hat{\Sigma}$ being the theoretical inner product matrix which we assume to exist. The constant M will be required to be of small order $\sqrt{n/\log p}$.

*The authors gratefully acknowledge financial support of the Swiss National Science Foundation, grant nr. 20PA20E-134495. We are moreover very grateful to Guillaume Lecué for his helpful comments.

A motivation to study lower bounds for quadratic forms comes from theory for ℓ_1-penalized estimation methods. Here, the so-called (empirical) compatibility constant plays an important role. It is defined as follows. For S being a subset of $\{1, \ldots, p\}$, write $u_{j,S} = u_j \mathrm{l}\{j \in S\}$ $(j = 1, \ldots, p)$ and $u_{-S} := u - u_S$. The compatibility constant ([17]) is

$$\hat{\phi}^2(L, S) := \min\{|S| u^T \hat{\Sigma} u : \ \|u_S\|_1 = 1, \ \|u_{-S}\|_1 \le L\}.$$

The condition $\hat{\phi}^2(L, S) > 0$ for suitable values of L and S allows one to establish oracle inequalities for the Lasso. Indeed, let u^0 be the sparse vector we want to recover and let $S := \{j : \ u_j^0 \neq 0\}$ be its active set. Let $\xi \in \mathbb{R}^n$ be a "noise" vector. Consider the Lasso

$$\hat{u} := \arg\min_{u \in \mathbb{R}^p} \left\{ \|\xi + Xu^0 - Xu\|_{2,n}^2 + 2\lambda \|u\|_1 \right\}$$

where $\lambda > 0$ is a tuning parameter and where we use the notation $\|v\|_{2,n}^2 := v^T v / n$, $v \in \mathbb{R}^n$. For $\lambda > \lambda_0 =: \|\xi^T X\|_\infty / n$ it holds that

$$\|X(\hat{u} - u^0)\|_{2,n}^2 \le (\lambda + \lambda_0)^2 |S| / \hat{\phi}^2(L, S) \tag{1.1}$$

where $L := (\lambda + \lambda_0)/(\lambda - \lambda_0)$. We refer to [4] and the references therein. In the literature result (1.1) is considered to be an "oracle inequality" if – in a suitable asymptotic formulation – the constant L remains bounded (i.e. λ is of the same order as λ_0) and $\hat{\phi}^2(L, S)$ stays away from zero. In the present paper, this case may serve as benchmark case. We give non-asymptotic results and some asymptotic consequences showing that under certain conditions $\hat{\phi}^2(L, S)$ indeed stays away from zero.

Closely related is the so-called null space property[1] (see e.g. [6]) used in exact recovery. One says that X has the null space property relative to S if for all $u \in \mathbb{R}^p$ with $Xu = 0$ it holds that $\|u_S\|_1 < \|u_{-S}\|_1$. The null space property is the same as the condition $\hat{\phi}(1, S) > 0$ and implies in the noiseless case exact recovery of a sparse signal u^0 with active set S using basis pursuit ([5]):

$$\arg\min\{\|u\|_1 : \ Xu = Xu^0\} = u^0.$$

The compatibility constant is also a close relative of the (empirical) restricted eigenvalue defined in [2] as

$$\hat{\kappa}^2(L, S) := \min\{u^T \hat{\Sigma} u : \ \|u\|_2 = 1, \ \|u_{-S}\|_1 \le L \|u_S\|_1\}.$$

It is easy to see that $\hat{\kappa}^2(L, S) \le \hat{\phi}^2(L, S)$. An example where $\hat{\kappa}^2(L, S)$ will be much smaller than $\hat{\phi}^2(L, S)$ is given in [19]. In that example $\hat{\kappa}^2(1, S)$ is about $1/|S|$ whereas $\hat{\phi}^2(1, S)$ is about $1/2$. One sees that for large values of $|S|$ the difference is substantial.

[1] We thank Emmanuel Candés for pointing this out.

In many cases (e.g. when applying the Lasso) the data are first normalized: for $\hat{\sigma}_j^2 := \hat{\Sigma}_{j,j}$ one replaces the j-th column X_j of X by $\tilde{X}_j / \hat{\sigma}_j$, $j = 1, \ldots, p$. Therefore we study in Section 5 the compatibility constant for normalized design

$$\tilde{\phi}^2(L, S) := \min\{|S| u^T \hat{R} u : \ \|u_S\|_1 = 1, \ \|u_{-S}\|_1 \leq L\}$$

and restricted eigenvalue for normalized design

$$\tilde{\kappa}^2(L, S) := \min\{u^T \hat{R} u : \ \|u\|_2 = 1, \ \|u_{-S}\|_1 \leq L \|u_S\|_1\}.$$

where $\hat{R} := \tilde{X}^T \tilde{X}/n$ with $\tilde{X} := (\tilde{X}_1, \ldots, \tilde{X}_p)$ being the normalized Gram matrix.

1.1. Organization of the paper

After some notations and definitions in the next section, we present in Section 3 a bound for sparse quadratic forms. The lower bounds for the empirical compatibility constant and empirical restricted eigenvalue follow from this. The upper bounds depend on fourth moments. We will show that $\hat{\phi}(L, S)$ converges to its theoretical counterpart, and similarly for $\hat{\kappa}(L, S)$ (see Theorem 4.2). For this we need $(L+1)\sqrt{s}$ to be of small order $\sqrt{n/\log p}$ (for the lower bound). This is detailed in Section 4. In Section 5 we consider the transfer principle from [13] which allows one to show that for the case where the data are normalized very weak moment conditions suffice. Section 6 is devoted to a discussion with related work. There, we summarize the comparison of results in Table 1. In Section 7 we make a brief comparison of the results when we drop the isotropy assumption. We show convergence of $|u^T \hat{\Sigma} u - u^T \Sigma_0 u|$ uniformly over $\|u\|_1 \leq M$ assuming sub-exponential entries in X_0. In Section 8 we examine the higher order isotropy condition. Finally, Section 9 contains the proofs.

2. Notation and definitions

We let $\Sigma_0 := \mathbb{E} X_0 X_0^T = \mathbb{E}\hat{\Sigma}$ be the theoretical inner product matrix. Its smallest eigenvalue is denoted by ψ_0^2. We do not assume $\psi_0 > 0$. For $m \geq 1$, and Z a real-valued random variable, we introduce the notation

$$\|Z\|_m^m := \mathbb{E}|Z|^m.$$

Thus $u^T \Sigma_0 u = \|\langle X_0 u \rangle\|_2^2$ where $\langle X_0 u \rangle$ is the inner product $X_0^T u$, $u \in \mathbb{R}^p$.

Let $X_{i,\cdot}^T$ be the i-th row of X ($i = 1, \ldots, n$). We write for a function $f : \mathbb{R}^p \to \mathbb{R}$,

$$\|f\|_{2,n}^2 := \|f(X)\|_{2,n}^2 := \frac{1}{n} \sum_{i=1}^n f^2(X_{i,\cdot})$$

and

$$\|f\|_2^2 := \|f(X_0)\|_2^2 := \mathbb{E}\|f(X)\|_{2,n}^2,$$

so that $\|Xu\|_{2,n}^2 = u^T \hat{\Sigma} u$.

Definition 2.1. We say that a random variable Z is Bernstein with constants σ and K if for all $k \in \{2, 3, \ldots\}$

$$\mathbb{E}|Z|^k \leq \frac{k!}{2} K^{k-2} \sigma^2.$$

Definition 2.2. We say that a random variable $Z \in \mathbb{R}$ is sub-Gaussian with constant C if for all $\lambda > 0$

$$\mathbb{E} \exp[\lambda |Z|] \leq 2 \exp[\lambda^2 C^2 / 2].$$

Let us denote for $k = 1, 2$ the Orlicz norm by

$$\|Z\|_{\Psi_k} := \inf\{c > 0 : \ \mathbb{E} \exp[|Z/c|^k] - 1 \leq 1\}.$$

Then being a Bernstein random variable is equivalent to having finite $\| \cdot \|_{\Psi_1}$-norm (i.e., being sub-exponential) and sub-Gaussianity is equivalent to a finite $\| \cdot \|_{\Psi_2}$-norm. We have chosen for the Definitions 2.1 and 2.2 in order to have simple explicit dependence on the constants later on.

Note that if a random variable is sub-Gaussian with constant C it is also Bernstein with constants $\sigma = 2$ and $K = \sqrt{2} C$. Moreover, a Bernstein random variable Z with constants σ and K always has $\sigma \leq 3K$ and so $\|Z\|_m \leq mK$ for all $m \in \{3, 4, \ldots\}$.

We use the definition of [12] or [11] of a sub-Gaussian vector (in a slightly alternative formulation).

Definition 2.3. A random vector $X_0 \in \mathbb{R}^p$ is sub-Gaussian with constant C if for all $u \in \mathbb{R}^p$ with $\|\langle X_0 u \rangle\|_2 = 1$ the random variable $\langle X_0 u \rangle$ is sub-Gaussian with constant C.

The main concept we will use in this paper is weak isotropy, for which we now present the definition.

Definition 2.4. Let $m \geq 2$. The random vector $X_0 \in \mathbb{R}^p$ is weakly m-th order isotropic with constant C_m if for all $u \in \mathbb{R}^p$ with $\|\langle X_0 u \rangle\|_2 = 1$ it holds that

$$P(|\langle X_0 u \rangle| > t) \leq (C_m / t)^m \ \forall \ t > 0.$$

A Gaussian vector is sub-Gaussian with constant 1 and is strongly m-th order isotropic (defined in Definition 6.1) with constant $\sqrt{2} \Gamma^{\frac{1}{m}}((1+m)/2) \pi^{-\frac{1}{m}}$, $m \geq 2$.

Definition 2.4 (and 2.3) are invariant under rotations: if $\psi_0 > 0$ one may without loss of generality assume $\Sigma_0 = I$ here. We however explicitly do not assume $\Sigma_0 = I$ because conditions on the ℓ_1-norm are not invariant under rotation. In contrast to the literature where the "isotropic" case is sometimes defined as the case $\Sigma_0 = I$ our definition of isotropy is rather to be understood as uniformity in all one-dimensional directions (very much like isotropy of functions in Besov spaces).

3. Lower bounds for sparse quadratic forms under higher order isotropy

The first result of Theorem 3.1 below is as in [16] and is given for completeness. It is only of interest when p is smaller than n. The result is improved in [7]. We refer to Section 6 for a discussion. The second result of Theorem 3.1 extends the situation to the case where p can be larger than n but ℓ_1-restrictions are invoked. Here we need bounds on Rademacher averages. A Rademacher sequence is a sequence of independent random variables $\epsilon_1, \ldots, \epsilon_n$, where each ϵ_i takes values ± 1 with probability 1/2. We assume that $\epsilon := (\epsilon_1, \ldots, \epsilon_n)^T$ is independent of X. Consider the Rademacher averages

$$W^T := (W_1, \ldots, W_p)^T := \epsilon^T X/n.$$

and let $\|W\|_\infty := \max_{1 \le j \le p} |W_j|$. We will need bounds for $\mathbb{E}\|W\|_\infty$. If the entries in X_0 are Bernstein with constants σ_X and K_X, then applying Lemma 14.12 in [4] gives

$$\mathbb{E}\|W\|_\infty \le \sigma_X \sqrt{\frac{2\log(2p)}{n}} + K_X \frac{\log(2p)}{n}. \tag{3.1}$$

and it is this bound that is invoked in the second result of Theorem 3.1. Further bounds for $\mathbb{E}\|W\|_\infty$ are discussed in Subsection 3.1.

Theorem 3.1. *Suppose that for some $m > 2$ the random vector X_0 is weakly m-th order isotropic with constant C_m and define*

$$D_m := [2C_m]^{\frac{m}{m-1}} (m-1)/(m-2). \tag{3.2}$$

Then for all $t > 0$ with probability at least $1 - \exp[-t]$

$$\inf_{\|\langle X_0 u\rangle\|_2 = 1} \|Xu\|_{2,n}^2 - 1 \ge -\left[D_m \left(16\sqrt{\frac{p}{n}} + \sqrt{\frac{2t}{n}} \right)^{\frac{m-2}{m-1}} + \frac{8D_m^2}{3} \left(\frac{t}{n} \right)^{\frac{m-2}{m-1}} \right]. \tag{3.3}$$

If in addition the entries in X_0 are Bernstein with constants σ_X and K_X, then for all $t > 0$ with probability at least $1 - \exp[-t]$

$$\inf_{\|\langle X_0 u\rangle\|_2 = 1,\ \|u\|_1 \le M} \|Xu\|_{2,n}^2 - 1 \ge -\Delta_n^{\mathrm{L}}(M, t) \tag{3.4}$$

where

$$\Delta_n^{\mathrm{L}}(M, t) := D_m \left(16 M \delta_n + \sqrt{\frac{2t}{n}} \right)^{\frac{m-2}{m-1}} + \frac{8D_m^2}{3} \left(\frac{t}{n} \right)^{\frac{m-2}{m-1}} \tag{3.5}$$

with $\delta_n := \sigma_X \sqrt{\frac{2\log(2p)}{n}} + K_X \frac{\log(2p)}{n}$.

Asymptotics In an asymptotic formulation suppose that C_m, K_X and σ_X remain fixed and that $(\log p)/n = o(1)$. Then the second result (3.4) of Theorem 3.1 says that for $M = o(\sqrt{n/\log p})$ one has

$$\inf_{\|\langle X_0 u\rangle\|_2 = 1,\ \|u\|_1 \le M} \|Xu\|_{2,n}^2 \ge 1 - o_{\mathbb{P}}(1).$$

Remark 3.1. The constant t in the formulation of Theorem 3.1 allows one to choose the confidence level of the result. If t is large (for example p large and $t = \log p$) the bounds will be true with large probability. Of course for very large t the bounds will become void.

Remark 3.2. We have not attempted to obtain small constants in the bound of Theorem 3.1. In fact, the last "smaller order" term in the expression (3.5) for $\Delta_n^{\mathrm{L}}(M,t)$ can be refined but this will make the expressions more involved.

Remark 3.3. The technique to prove Theorem 3.1 does not rely on the fact that we consider squared functions $|(Xu)_i|^2$, $i = 1, \ldots, n$. For example, one may use it for bounding

$$\inf_{\|\langle X_0 u\rangle\|_2 = 1, \|u\|_1 \le M} \|Xu\|_{q,n}^q,$$

where for $q \ge 1$

$$\|Xu\|_{q,n}^q := \frac{1}{n}\sum_{i=1}^n |(Xu)_i|^q.$$

Then one could e.g. use weak isotropy conditions of order $m > q$. However, a motivation for having such results is perhaps lacking.

Theorem 3.1 is based on a truncation argument. For the case of a sub-Gaussian vector X_0 the truncation level can be taken rather small, leading to an improved bound. We present this for completeness in the next lemma.

Lemma 3.1. *If the random vector X_0 is sub-Gaussian with constant C we find that for all $t > 0$ with probability at least $1 - \exp[-t]$*

$$\inf_{\|\langle X_0 u\rangle\|_2 = 1,\ \|u\|_1 \le M} \|Xu\|_{2,n}^2 - 1$$

$$\ge -\sqrt{2}Cb(1 + 2\sqrt{2\log(C/b)}) - 16C^2 t \cdot \frac{\log(C/b)}{3n}$$

where

$$b := 16 \min\left\{ M\delta_n', \sqrt{\frac{p}{n}} \right\} + \sqrt{\frac{2t}{n}},$$

with $\delta_n' := C\sqrt{\frac{2\log(2p)}{n}}$.

3.1. Bounds for $\mathbb{E}\|\boldsymbol{W}\|_\infty$

Inequality (3.1) presents a bound for $\mathbb{E}\|W\|_\infty$ assuming Bernstein conditions. This bound is then invoked in Theorem 3.1. One may derive alternative bounds for $\mathbb{E}\|W\|_\infty$ and adjust the definion of δ_n in Theorem 3.1 accordingly. For example one may impose existence of k-th moments of the entries of X_0 where k is of order $\log p$. The paper [9] presents refined results, which we cite in the next lemma.

Lemma 3.2. *Let $Z_1, \ldots, Z_n$ be i.i.d. copies of a mean-zero random variable $Z \in \mathbb{R}$ and $\bar{Z} := \sum_{i=1}^n Z_i/n$. Suppose that for some constants κ_1 and $\alpha \geq 1/2$ one has*

$$\|Z\|_k \leq \kappa_1 k^{\alpha},\ 2 \leq k \leq k_0.$$

Then for $n \geq k_0^{\max\{2\alpha-1,1\}}$ and for all $k \leq k_0$

$$\|\bar{Z}\|_k \leq c_0 \exp[2\alpha - 1]\kappa_1 \sqrt{k/n},$$

where c_0 is a universal constant.

Corollary 3.1. *Suppose that for some constants κ_1, $\eta \geq 2/\log p$ and $\alpha \geq 1/2$ one has*

$$\max_{1 \leq j \leq p} \|X_{0,j}\|_k \leq \kappa_1 k^{\alpha},\ 2 \leq k \leq k_0 := \eta \log p. \tag{3.6}$$

Then for $n \geq k_0^{\max\{2\alpha-1,1\}}$

$$\max_{1 \leq j \leq p} \|W_j\|_{k_0} \leq c_0 \exp[2\alpha - 1]\kappa_1 \sqrt{k_0/n},$$

where c_0 is a universal constant. But then

$$\mathbb{E}\|W\|_\infty \leq c_1 \sqrt{\log p/n}$$

where $c_1 = c_0 \kappa_1 \sqrt{\eta} \exp[2\alpha - 1 + 1/\eta]$.

4. Convergence of the compatibility constant and restricted eigenvalue

An "almost isometric" (in a terminology from [7]) lower bound for the empirical compatibility constant and empirical restricted eigenvalue follows easily from Theorem 3.1 as is shown in the next theorem.

Recall that $S \subset \{1, \ldots, p\}$ is an arbitrary subset. Let for $s := |S|$

$$\phi_0^2(L, S) := \min\{s\|\langle X_0 u\rangle\|_2^2 :\ \|u_S\|_1 = 1,\ \|u_{-S}\|_1 \leq L\}.$$

be the theoretical compatibility constant and

$$\kappa_0^2(L, S) := \min\{\|\langle X_0 u\rangle\|_2^2 :\ \|u_S\|_2 = 1,\ \|u_{-S}\|_1 \leq L\|u_S\|_1\}$$

be the theoretical restricted eigenvalue.

Theorem 4.1. *Under the conditions of Theorem 3.1 and using its notation we find that for all $t > 0$, with probability at least $1 - \exp[-t]$*

$$\frac{\hat{\phi}^2(L, S)}{\phi_0^2(L, S)} - 1 \geq -\Delta_n^{\mathrm{L}}((L+1)\sqrt{s}/\phi_0(L, S), t)$$

as well as

$$\frac{\hat{\kappa}^2(L, S)}{\kappa_0^2(L, S)} - 1 \geq -\Delta_n^{\mathrm{L}}((L+1)\sqrt{s}/\kappa_0(L, S), t).$$

Note that Theorem 4.1 does not depend on the smallest eigenvalue ψ_0 of Σ_0 nor on its maximal eigenvalue. If $\psi_0 > 0$ one may however want to insert the bounds $\phi_0(L,S) \ge \kappa_0(L,S) \ge \psi_0$. We refer to the "Asymptotics" paragraph at the end of this section for a further discussion.

The next issue is whether $\hat{\phi}(L,S)$ actually converges to $\phi_0(L,S)$ and $\hat{\kappa}(L,S)$ to $\kappa_0(L,S)$. This part follows easily from the lower bounds of Theorem 4.1 and convergence of $\|Xu\|_{2,n}^2 - \|\langle X_0 u\rangle\|_2^2$ for fixed values of u, for which in turn we e.g. would like to have fourth moments. If $m > 4$, this 4-th order moment condition follows from m-th order weak isotropy. If however X_0 is only m-th order weakly isotropic for $m \le 4$ we need some other means to check 4-th moments. The next lemma can be invoked.

Lemma 4.1. *Suppose that the entries in X_0 are Bernstein with constants σ_X and $K_X \ge \sigma_X$ and that for some constant $c_0 \ge 1$ and for $c_1 = 2(1+c_0)(K_X+\sigma_X)$ we have*

$$c_1 M \log(2p) \le p^{c_0/2}.$$

Then for all u with $\|\langle X_0 u\rangle\|_2 \le 1$ and $\|u\|_1 \le M$ we have

$$\|\langle X_0 u\rangle\|_4^2 \le \sqrt{2} c_1 M \log(2p).$$

Combining Theorem 4.1 with Lemma 4.1 gives the upper and lower bounds shown in the next theorem.

Theorem 4.2. *Suppose that X_0 is weakly m-th order isotropic with constant C_m and that the entries in X_0 are Bernstein with constants σ_X and $K_X > \sigma_X$. For the case $m \le 4$ we assume in addition that for some constant $c_0 \ge 1$ and for $c_1 := 2(1+c_0)(K_X+\sigma_X)$*

$$c_1 (L+1)\sqrt{s} \log(2p) \le p^{c_0/2}. \tag{4.1}$$

Define D_m as in (3.2) and $\Delta_n^{\mathrm{L}}(M,t)$ as in (3.5). For all $t > 0$, with probability at least $1 - \exp[-t] - 1/t$

$$-\Delta_n^{\mathrm{L}}((L+1)\sqrt{s}/\phi_0(L,S), t) \le \frac{\hat{\phi}^2(L,S)}{\phi_0^2(L,S)} - 1 \le \Delta_n^{\mathrm{U}}((L+1)\sqrt{s}, t)$$

and

$$-\Delta_n^{\mathrm{L}}((L+1)\sqrt{s}/\kappa_0(L,S), t) \le \frac{\hat{\kappa}^2(L,S)}{\kappa_0^2(L,S)} - 1 \le \Delta_n^{\mathrm{U}}((L+1)\sqrt{s}, t),$$

where

$$\Delta_n^{\mathrm{U}}((L+1)\sqrt{s}, t) = \begin{cases} c_1(L+1)\sqrt{s}\log(2p)\sqrt{\frac{2t}{n}} & m \le 4 \\ \min\left\{ c_1(L+1)\sqrt{s}\log(2p), C_m^2 \sqrt{\frac{m}{2(m-4)}} \right\} \sqrt{\frac{2t}{n}} & m > 4 \end{cases}.$$

Asymptotics In an asymptotic formulation we assume that the constants $1/\phi_0(L,S)$, C_m, σ_X and K_X remain bounded. Then it follows from Theorem 4.2 that under its conditions, as long as $(L+1)\sqrt{s} = o(\sqrt{n/\log p})$

$$\hat{\phi}^2(L,S) \geq \phi_0^2(L,S) - o_{\mathbb{P}}(1).$$

If $m \leq 4$, c_0 is fixed and $(L+1)\sqrt{s} = o(\sqrt{n}/\log p)$ we find

$$\hat{\phi}^2(L,S) \leq \phi_0^2(L,S) + o_{\mathbb{P}}(1).$$

Similar results hold for the restricted eigenvalue. (Note that for c_0 fixed and $p > n$ condition (4.1) follows from the already imposed condition $(L+1)\sqrt{s} = o(\sqrt{n/\log p})$.) Thus, in the upper bound an additional $\sqrt{\log p}$ appears in the requirement on M. This term can be omitted if $m > 4$ or if we assume the entries in X_0 are sub-Gaussian instead of Bernstein.

5. Bounds for the compatibility constant and restricted eigenvalue using the transfer principle

In this section, we assume for simplicity that Σ_0 has ones on the diagonal. We let $\hat{\sigma}_j^2 := \hat{\Sigma}_{j,j} = \|X_j\|_{2,n}^2$, $j = 1, \ldots, p$ where X_j denotes the j-th column of X.

5.1. The transfer principle

The transfer principle given in the next theorem is from [13]. As shown in the latter paper it can be used to move from the case $p \leq n$ to $p > n$ assuming ℓ_1-conditions. We will apply this technique here as well, for non-normalized design in Theorem 5.2 and for normalized design in Theorem 5.3. The results are compared with [13] in Section 6.

Theorem 5.1. *Let A be a symmetric $p \times p$ matrix with $A_{j,j} \geq 0$ for all $j \in \{1, \ldots, p\}$. Let $d \in \{2, \ldots, p\}$ and suppose that for all $J \subset \{1, \ldots, p\}$ with cardinality $|J| = d$ and all $u \in \mathbb{R}^p$ one has*

$$u_J^T A u_J \geq 0.$$

Then for all $u \in \mathbb{R}^p$

$$u^T A u \geq -\max_j A_{j,j} \|u\|_1^2/(d-1).$$

We will invoke the transfer principle via the following corollary (as well as directly in the proof of Theorem 5.3). The corollary is as in [13] and we state it here in our notation for ease of reference.

Corollary 5.1. *Let $M^2 \in \{2, \ldots, p\}$ and $0 \leq \Delta < 1$. Consider the events*

$$\mathcal{A} := \left\{ \frac{\|X u_J\|_{2,n}^2}{\|\langle X_0 u\rangle_J\|_2^2} \geq 1 - \Delta, \ \forall\, u \in \mathbb{R}^p, \ \forall\, J \subset \{1, \ldots, p\} \text{ with } \ |J| \leq M^2 \right\}$$

and, for some $\epsilon > 0$, the event

$$\mathcal{B} := \Big\{ \max_{1 \le j \le p} \hat{\sigma}_j^2 \le 1 + \epsilon \Big\}.$$

Then on $\mathcal{A} \cap \mathcal{B}$ for all $\|u\|_1 \le M \|\langle X_0 u \rangle\|_2$

$$\|Xu\|_{2,n}^2 \ge (1 - 3\Delta - 2\epsilon) \|\langle X_0 u \rangle\|_2^2.$$

To put this corollary to work we insert the first result of Theorem 3.1.

Theorem 5.2. *Suppose that for some $m > 2$ the random vector X_0 is weakly m-th order isotropic with constant C_m. Define D_m as in Theorem 3.1. Let for $M^2 \in \{2, \ldots, p\}$*

$$\bar{\Delta}_n^{\mathrm{L}}(M,t) := D_m \left(16 M \frac{(1 + \sqrt{2 \log p})}{\sqrt{n}} + \sqrt{\frac{2t}{n}} \right)^{\frac{m-2}{m-1}} + \frac{8 D_m^2}{3} \left(\frac{t + M^2 \log p}{n} \right)^{\frac{m-2}{m-1}}$$

and let, for some $\epsilon > 0$, $\mathcal{B}$ be the event

$$\mathcal{B} := \Big\{ \max_{1 \le j \le p} \hat{\sigma}_j^2 \le 1 + \epsilon \Big\}.$$

Then with probability at least $1 - \exp[-t] - \mathbb{P}(\mathcal{B}^c)$, we have that uniformly in $\|u\|_1 \le M \|\langle X_0 u \rangle\|_2$,

$$\|Xu\|_{2,n}^2 \ge \left(1 - 3 \bar{\Delta}_n^{\mathrm{L}}(M,t) - 2\epsilon \right) \|\langle X_0 u \rangle\|_2^2.$$

The above theorem invokes Theorem 3.1 for handling the event $\mathcal{A}$. One may also use the results in [13] for the case of m-th order strong isotropy (defined in Definition 6.1) with $m \ge 4$ or those which can be deduced from [7] for the case m-th order weak isotropy with $m > 2$ (the latter paper does not explicitly treat an event of the form $\mathcal{A}$). For the case $m < 4$ for instance the arguments in [7] would allow to replace $\bar{\Delta}_n(M,t)$ in Theorem 5.2 (which is of order $[M\sqrt{\log p / n}]^{\frac{m-2}{m-1}}$ by a term of order $[M\sqrt{\log p/n} \log(1/(\sqrt{\log p/n} M))]^{2(m-2)/m}$.

Clearly, one can again apply the results to the compatibility constant and restricted eigenvalue as in Theorem 4.1. This gives the following corollary.

Corollary 5.2. *Assume the conditions of Theorem 5.2 and let $(L+1)^2 s \in \{2, \ldots, p\}$. Then with probability at least $1 - \exp[-t] - \mathbb{P}(\mathcal{B}^c)$*

$$\frac{\hat{\phi}^2(L,S)}{\phi_0^2(L,S)} \ge 1 - 3 \bar{\Delta}_n^{\mathrm{L}}((L+1)\sqrt{s}/\phi_0(L,S), t) - 2\epsilon$$

as well as

$$\frac{\hat{\kappa}^2(L,S)}{\kappa_0^2(L,S)} \ge 1 - 3 \bar{\Delta}_n^{\mathrm{L}}((L+1)\sqrt{s}/\kappa_0(L,S), t) - 2\epsilon.$$

5.2. The behaviour of $\max_j \hat{\sigma}_j^2$

Recall that in Theorem 3.1 the lower bound for $\inf\{\|Xu\|_{2,n}: \ \|\langle X_0 u\rangle\|_2 = 1, \ \|u\|_1 \leq M\}$ depends on the bound δ_n for $\mathbb{E}\|W\|_\infty$. Bounding $\mathbb{E}\|W\|_\infty$ leads to moment conditions on the entries in X_0. The transfer principle now leads to requiring a bound for $\max_j \hat{\sigma}_j^2$ where $\hat{\sigma}_j^2 = \sum_{i=1}^n X_{i,j}^2/n$. The latter is clearly a more difficult task than the former. In the non-normalized case this appears to be the price to pay for application of the elegant transfer principle.

We first assume sub-Gaussian tail behaviour in Lemma 5.1 and then moments up to order $\log p$ in Lemma 5.2.

Lemma 5.1. *Suppose that the entries of X_0 are sub-Gaussian with constant C. Then for all $t > 0$*

$$\mathbb{P}\biggl(\max_{1\leq j\leq p} |\hat{\sigma}_j^2 - 1|/(8C^2) \geq \sqrt{\frac{2\log(2p)}{n}} + \sqrt{\frac{2t}{n}} + \frac{\log(2p)}{n} + \frac{t}{n} \biggr) \leq \exp[-t].$$

Lemma 5.2. *Suppose the conditions of Corollary 3.1 with constants $\kappa_1 \geq 1$, $\eta \geq 2/\log p$ and $\alpha \geq 1/4$:*

$$\max_{1\leq j\leq p} \|X_{0,j}\|_k \leq \kappa_1 k^\alpha, \ \ 2 \leq k \leq k_0 := \eta \log p.$$

Then for $n \geq (k_0/2)^{\max\{4\alpha-1,1\}}$ and all $t > 0$ with probability at least $1 - 1/t$

$$\max_{1\leq j\leq p} \hat{\sigma}_j^2 \leq 1 + c_1 t^{2/k_0} \sqrt{\log p/n},$$

where $c_1 := c_0 \exp[4\alpha - 1 + 2/\eta]\kappa_1^2 2^{2\alpha+1}\sqrt{\eta/2}$ with c_0 a universal constant.

For example, when $\alpha = 1$ the direct approach of Theorem 3.1 requires $n \geq \log p$ (see Corollary 3.1) whereas the transfer principle leads to requiring $n \geq \log^3 p$.

5.3. Normalized design

Define $\tilde{X}_j := X_j/\hat{\sigma}_j, \ j = 1, \ldots, p, \ \tilde{X} := (\tilde{X}_1, \ldots, \tilde{X}_p)$ and

$$\hat{R} := \tilde{X}^T \tilde{X}/n.$$

Thus, $\hat{R} = \hat{D}^{-1/2}\hat{\Sigma}\hat{D}^{-1/2}$ where $\hat{D} = \text{diag}(\hat{\sigma}_1^2, \ldots, \hat{\sigma}_p^2)$. Define for $S \subset \{1, \ldots, p\}$ a set with cardinality $s := |S|$ the (empirical) compatibility constant for normalized design

$$\tilde{\phi}^2(L, S) := \min\{su^T\hat{R}u: \ \|u_S\|_1 = 1, \ \|u_{-S}\|_1 \leq L\}.$$

Similarly, the (empirical) restricted eigenvalue for normalized design is

$$\tilde{\kappa}^2(L, S) := \min\{u^T\hat{R}u: \ \|u_S\|_2 = 1, \ \|u_{-S}\|_1 \leq L\|u_S\|_1\}.$$

In [4] the (theoretical) adaptive restricted eigenvalue is defined as

$$\kappa_*^2(L, S) := \min\{\|\langle X_0 u\rangle\|_2^2: \ \|u_S\|_2 = 1, \ \|u_{-S}\|_1 \leq L\sqrt{s}\}.$$

Clearly $\kappa_*^2(L,S) \le \kappa_0^2(L,S)$. We prove in Theorem 5.3 that the empirical compatibility constant $\tilde{\phi}^2(L,S)$ can be bounded from below by the theoretical adaptive restricted eigenvalue. The theorem establishes that compatibility needs no further moment conditions on the entries in X_0. If we do assume such moment conditions on the entries $X_{0,j}$ with $j \in S$, the results can be extended to restricted eigenvalues, as shown in [13] for the case of 4-th order strong isotropy (defined in Definition 6.1), and as shown in the next theorem.

Theorem 5.3. *Suppose that for some $m > 2$ the random vector X_0 is weakly m-th order isotropic with constant C_m. Define D_m as in Theorem 3.1, and let $\bar{\Delta}_n^{\rm L}(M,t)$ be defined as in Theorem 5.2. Let, for some $\epsilon > 0$ sufficiently small, $\mathcal{B}_S$ and $\mathcal{C}_S$ be the events*

$$\mathcal{B}_S := \left\{ \max_{j\in S} \hat{\sigma}_j^2 \le 1+\epsilon \right\},\ \mathcal{C}_S := \left\{ \sum_{j\in S} \hat{\sigma}_j^2/s \le 1+\epsilon \right\}. \tag{5.1}$$

Let $0 < \Delta < 1$ be arbitrary and define $M^2(\Delta)$ as the largest value of $M^2 \in \{2,\ldots,p\}$ such that $\bar{\Delta}_n^{\rm L}(M(\Delta),t) \le \Delta$ (assuming such a value exists). Define $L(\Delta,\epsilon) := L\sqrt{1+\epsilon}/(1-\sqrt{\Delta})$. Then with probability at least $1-\exp[-t]-\mathbb{P}(\mathcal{C}_S^{\rm c})$ we have

$$\tilde{\phi}^2(L,S) \ge \frac{\kappa_*^2(L(\Delta,\epsilon),S)}{(1+\epsilon)} - \frac{(L+1)^2 s}{M^2(\Delta)-1}.$$

Moreover, with probability at least $1-\exp[-t]-\mathbb{P}(\mathcal{B}_S)$

$$\tilde{\phi}^2(L,S) \ge \frac{\phi_0^2(L(\Delta,\epsilon),S)}{(1+\epsilon)} - \frac{(L+1)^2 s}{M^2(\Delta)-1}$$

and

$$\tilde{\kappa}^2(L,S) \ge \frac{\kappa_0^2(L(\Delta,\epsilon),S)}{(1+\epsilon)} - \frac{(L+1)^2 s}{M^2(\Delta)-1}.$$

Asymptotics The above theorem shows that when $1/\kappa_*(L,S) = \mathcal{O}(1)$, $C_m = \mathcal{O}(1)$ and $(L+1)\sqrt{s} = o(\sqrt{n/\log p})$, then also $1/\tilde{\phi}(L,S) = \mathcal{O}_{\mathbb{P}}(1)$ and in fact $\liminf_{n\to\infty} \tilde{\phi}(L,S)/\kappa_*(L,S) \ge 1$ since $\kappa_*(L,S)$ is continuous in L and $\sum_{j\in S}\hat{\sigma}_j^2/s = 1 + o_{\mathbb{P}}(n^{-\frac{m-2}{2}})$.

6. Related work

Before discussing related work we present the definitions of the concepts used. Recall that in this paper we require weak isotropy (see Definition 2.4).

Definition 6.1. Let $m \ge 2$. The random vector $X_0 \in \mathbb{R}^p$ is strongly m-th order isotropic with constant C_m if for all $u \in \mathbb{R}^p$ with $\|\langle X_0 u\rangle\|_2 = 1$ it holds that

$$\|\langle X_0 u\rangle\|_m \le C_m.$$

Definition 6.2. The random vector $X_0 \in \mathbb{R}^p$ satisfies the L_1-L_2 property with constant C if for all $u \in \mathbb{R}^p$ with $\|\langle X_0 u\rangle\|_2 = 1$ it holds that

$$\|\langle X_0 u\rangle\|_1 \ge 1/C.$$

Definition 6.3. The random vector $X_0 \in \mathbb{R}^p$ satisfies the small ball property with constants $C_1 > 0$ and $C_2 > 0$ if for all $u \in \mathbb{R}^p$ with $\|\langle X_0 u\rangle\|_2 = 1$ it holds that

$$P\big(|\langle X_0 u\rangle| \ge 1/C_1\big) \ge 1/C_2.$$

It can be shown that for appropriate constants one has (for $m > 2$)

$$\text{strong } m\text{-th order isotropy} \Rightarrow \text{weak } m\text{-th order isotropy}$$

$$\Rightarrow L_1\text{-}L_2 \text{ property} \Rightarrow \text{small ball property.}$$

E.g. for the last implication see [7].

6.1. Relation of this work with [7] and [13]

The paper [7] obtains lower bounds for the smallest eigenvalue of $\hat\Sigma$ for the case $p \le n$. Their approach allows one to show that for $p \ll n$ it holds that $u^T\hat\Sigma u \ge (1-\Delta)u^T\Sigma_0 u$, uniformly in $u \in \mathbb{R}^p$, with large probability, for some small Δ. Such a result is not stated explicitly but it is easy to infer. The bounds in [7] are better than the first result (3.3) of Theorem 3.1. The paper employs a type of "peeling device" and the fact that for all $0 < a < b$

$$\{\{x:\ |xu| > K\}:\ \|\langle X_0 u\rangle\|_2 = 1,\ K \in (a, b]\}$$

is a VC-class with dimension at most p. If we have "good" bounds for the entropy for $\|\cdot\|_{2,n}$ of the classes

$$\{\{x:\ |xu| > K\}:\ \|\langle X_0 u\rangle\|_2 = 1,\ \|u\|_1 \le M,\ K \in (a, b]\}$$

their argument can be extended to the case $p > n$ with ℓ_1-restrictions. However, how to derive "good" entropy bounds for such classes is as yet not clear to us.

Both papers [7] and [13] assume m-th order isotropy (defined here in Definitions 2.4 and 6.1). The paper [7] has results with weak isotropy for any $m > 2$, whereas [13] assumes strong isotropy with $m = 4$. The paper [13] shows that by a transfer principle (described here in Theorem 5.1), a result for $p \le n$ can be invoked to derive that also for the case $p \gg n$ one has $u^T\hat\Sigma u \ge (1-\Delta')u^T\Sigma_0 u$, uniformly in $\|u\|_1 \le M u^T\Sigma_0 u$, with large probability, for some small Δ' and not too large M (generally of small order $\sqrt{n/\log p}$). In the present paper we consider weak isotropy with $m > 2$ as in [7] and we show by a direct method that $u^T\hat\Sigma u \ge (1-\Delta)u^T\Sigma_0 u$ uniformly in $\|u\|_1 \le M u^T\Sigma_0 u$ with large probability for some small Δ. Here, we assume sub-exponential tails for the entries in X_0, or, inserting results from [9], existence of moments up to order $\log p$ for these entries. We compared the result with the one using the transfer principle of [13]. Our finding is that the transfer principle needs slightly stronger moment conditions. In fact, our direct approach requires a bound for the maximum of the p Rademacher averages of the columns of X, whereas the approach using the transfer principle makes it necessary to have a bound for the maximal length of the p columns of X. Both can be dealt with by assuming higher order moments, but clearly the Rademacher averages need less moments than the lengths.

The paper [13] shows that when the columns of X are normalized to have all equal length, then the transfer principle leads to lower bounds for the (empirical) compatibility constant and (empirical) restricted eigenvalues assuming only 4-th order strong isotropy and moments of order bigger than 4 for the entries in X_0. We presented this result in Section 5 relaxing 4-th order strong isotropy to m-th order weak isotropy with $m > 2$. Moreover, we derive that the compatibility constant $\hat{\phi}^2(L, S)$ is positive with large probability assuming only isotropy but no additional moment assumptions on the entries in X_0. Thus, using normalized design one obtains exact recovery under isotropy only.

6.2. *Further related work*

In [15] a result of [14] concerning a lower bound for restricted eigenvalues is extended from the Gaussian case to the sub-Gaussian case. The paper [1] considers the case of log-concave distributions, which is related to sub-exponentiality of the vector X_0 (the sub-exponential variant of Definition 2.3). The papers [16] and [7] provide lower bounds for the empirical smallest eigenvalue $\hat{\psi}^2 := \min\{u^T \hat{\Sigma} u : \ \|u\|_2 = 1\}$ for the case where p is at most n. The paper [16] uses higher order isotropy conditions (defined in Definitions 6.1 and 2.4) and the paper [7] uses these too, but they in addition explore small ball properties (defined in Definition 6.3). The paper [9] considers the null space property, restricted eigenvalues $\hat{\kappa}^2(L, S)$ and results uniformly in S with fixed cardinality s, invoking small ball properties. Indeed, they show that small ball properties are very natural requirements when one aims at lower bounds. With the small ball property one obtains an "isomorphic" bound (we call this a result of type II in Table 1), that is, in a standard asymptotic framework the lower bound remains strictly smaller than the theoretical counterpart. For the case $p > n$, [9] has general moment conditions. It requires the stronger ("sub-Gaussian") conditions of Lemma 5.2 instead of the ("sub-exponential") condition (3.6) of Corollary 3.1 to arrive at the growth condition $s = o(n/\log p)$. The paper shows that moment conditions are necessary for exact recovery (see also [8]), but for the case $p \ll n$ it proves lower bounds for empirical smallest eigenvalues assuming only the small ball property and no further moment conditions (the covariance matrix is not required to exist here).

In Table 1 we present a summary of the results in the cited papers in comparison with the present paper. Of course it is not possible to make a simple comparison doing all aspects of the cited papers justice. The summary should be seen as focussing on what are in our view the relevant differences.

7. The case of (almost) bounded random variables

The bounded case is considered [15] and a reformulation is in [18]. It is shown there that when $\|X_0\|_\infty \leq K_X$ then for a universal constant c_1 and for all $t > 0$,

Table 1

The entry "pp" stands for the present paper. With "isotropy" we mean weak or strong isotropy. The "sub-Gaussian" results concern the lower tails for quadratic forms. With "conditions on p" we mean conditions stronger than the asymptotic one $\log p/n \to 0$. The entries "normalized" stand for normalized design and "non-normalized" for non-normalized design. The symbols κ^2, ψ^2 and ϕ^2 are shorthand for restricted eigenvalue, smallest eigenvalue and compatibility constant respectively. With results of "type I" we mean results in terms of theoretical counterparts. Results of "type II" are in terms of the constants occurring e.g. in the small ball property. The "moment conditions" are apart from isotropy (or small ball properties) those on the entries of X_0

	[13]	[7]	[9]	pp
isotropy	$m \geq 4$	$m > 2$	no	$m > 2$
small ball	no	yes	yes	no
conditions on p	no	$p < n$	no	no
sub-Gaussian	yes	yes	no	no
normalized	yes	no	no	yes
results	κ^2	ψ^2	ϕ^2	κ^2 and ϕ^2
type of result	I	I & II	II	I
moment conditions:				
non-normalized	-	-	sub-Gaussian type	sub-exponential type
normalized	weak for κ^2	-	-	none for ϕ^2

with probability at least $1 - \exp[-t]$

$$\sup_{\|\langle X_0 u\rangle\|_2 \leq 1,\ \|u\|_1 \leq M} \left| \|Xu\|_{2,n} - \|\langle X_0 u\rangle\|_2^2 \right| / c_1 \leq M K_X \sqrt{\frac{\log p \log^3 n + t}{n}}$$

$$+ M^2 K_X^2 \frac{\log p \log^3 n + t}{n}.$$

Observe this inequality goes both ways, and it does not require higher order isotropy conditions. On the other hand, the bounds involve an additional $\log^3 n$-factor. If we replace the assumption of bounded random variables by (say) a sub-Gaussian assumption but *do* assume strong (say) isotropy we can again use a truncation argument and obtain an inequality that goes both ways. Admittedly, the number of $\log p$- and $\log n$-terms increases.

We first present an auxiliary truncation lemma.

Lemma 7.1. *Suppose X_0 is strongly m-th order isotropic with constant $\tilde{C}_m$ and that its components are sub-Gaussian with constant C. Let $t > 0$ be arbitrary and let*

$$A(t) := \{x \in \mathbb{R}^p : \max_{1 \leq j \leq p} |x_j| \leq C(\sqrt{2t + 2\log(2p) + 2m(\log n)/(m-2)})\}.$$

Then for all u with $\|\langle X_0 u\rangle\|_2 = 1$

$$\|(\langle X_0 u\rangle) \mathrm{l}\{X_0 \notin A(t)\}\|_2^2 \leq \tilde{C}_m^2 \exp[-t(m-2)/m]/n.$$

Theorem 7.1. *Suppose X_0 is strongly m-th order isotropic with constant $\tilde{C}_m$ and that its components are sub-Gaussian with constant C. Then for a universal* *constant* c_1 *and for all* $t > 0$ *with probability at least* $1 - (1 + n^{-m/(m-2)}) \exp[-t]$

$$\sup_{\|\langle X_0 u\rangle\|_2 \le 1,\ \|u\|_1 \le M} \left| \|Xu\|_{2,n}^2 - \|\langle X_0 u\rangle\|_2^2 \right| / c_1$$

$$\le MC \sqrt{\frac{(2t + 2\log(2p) + 2m(\log n)/(m-2))(\log p \log^3 n + t)}{n}}$$

$$+ M^2 C^2 \frac{(2t + 2\log(2p) + 2m(\log n)/(m-2))(\log p \log^3 n + t)}{n}$$

$$+ \tilde{C}_m^2 \exp[-t(m-2)/m]/n.$$

8. Higher order isotropy

If X_0 is (strongly or weakly) m-th order isotropic with constant C_m and A is a $q \times p$ matrix, then clearly AX_0 is also (strongly or weakly) m-th order isotropic with constant C_m. In other words, the property is invariant under linear transformations. The same is true for sub-Gaussianity. In particular, we have invariance under any permutation of the $X_{0,j}$.

In the next subsection, we assume that the $\{X_{0,j}\}$ form a directed acyclic graph (possibly after some linear transformation) where the noise terms are a martingale difference array with fixed sub-Gaussian tail behaviour. Then we extend in Subsections 8.2 and 8.3 the situation where the conditional tail behaviour is sub-Gaussian or Bernstein, with constants depending on predictable random variables. We consider there a filtration $\{\mathcal{F}_j\}_{j=1}^p$ and predictable random variables $\{V_j\}_{j=1}^p$ that satisfy for some constants $m > 2$ and μ_m

$$\max_{1 \le j \le p} \|V_j\|_m \le \mu_m.$$

We investigate strong m-th order isotropy. In fact we give explicit expressions for $\|\langle X_0 u\rangle\|_m$ in terms of $\|u\|_2$. This implies strong isotropy if we assume the smallest eigenvalue ψ_0^2 of Σ_0 is positive. Obviously this also implies a bound for the largest eigenvalue $\psi_{\max}^2$ of Σ_0:

$$\psi_{\max}^2 \le \max\{\|\langle X_0 u\rangle\|_m^2 :\ \|u\|_2 = 1\}.$$

8.1. Directed acyclic graphs

Let X_0 be a vector of random variables with mean zero and covariance matrix $\Sigma_0 := \mathbb{E} X_0 X_0^T$. We want to find conditions such that for all $u \in \mathbb{R}^p$, with $\|\langle X_0 u\rangle\|_2 = 1$ the random variable $\langle X_0 u\rangle$ is sub-Gaussian with constant C. We will examine this here for the situation where the graph of X_0 has a directed acyclic graph (DAG) structure that is, satisfying (after an appropriate permutation of the indexes) the structural equations model

$$X_{0,1} = \epsilon_{0,1},\ X_{0,j} = \sum_{k=1}^{j-1} X_{0,k}\beta_{k,j} + \epsilon_{0,j},\ j = 2, \ldots, p \tag{8.1}$$

where $\{\epsilon_{0,j}\}_{j=1}^p$ is a martingale difference array for the filtration $\{\mathcal{F}_j\}_{j=0}^{p-1}$. We assume $X_{0,j}$ is $\mathcal{F}_j$-measurable, $j = 1, \ldots, p$. We moreover assume that $\omega_j^2 := \mathrm{var}(\epsilon_{0,j}) = \mathbb{E}\mathrm{var}(X_j|\mathcal{F}_{j-1})$ exists for all j. Note that model (8.1) holds when X_0 is Gaussian for example. More generally, the standard linear structural equations model is a special case. The latter model assumes that for $j \geq 2$, the noise $\epsilon_{0,j}$ is independent of $\{X_{0,k}\}_{k=1}^{j-1}$, and that $\epsilon_{0,1}, \ldots \epsilon_{0,p}$ are independent mean-zero random variables.

Lemma 8.1. *Assume the structural equations model (8.1). Assume in addition that for some constant C and for all $\lambda \in \mathbb{R}$*

$$\mathbb{E}(\exp[\lambda \epsilon_{0,j}/\omega_j]|\mathcal{F}_{j-1}) \leq \exp[\lambda^2 C^2/2], j = 1, \ldots, p.$$

Then X_0 is sub-Gaussian with constant C.

The above lemma follows from the fact that its condition implies that the vector $\epsilon_0 := (\epsilon_{0,1}, \ldots, \epsilon_{0,p})^T$ is sub-Gaussian with constant C. If ϵ_0 is (strongly or weakly) m-th order isotropic with constant C_m, then under the structural equations model (8.1) the vector X_0 is also (strongly or weakly) m-th order isotropic with constant C_m. This follows from the fact that X_0 is a linear transformation of ϵ_0. One may use the results of the next two subsections to check isotropy of ϵ_0.

8.2. The conditionally sub-Gaussian case

Let $\{\mathcal{F}_j\}_{j=0}^p$ be a filtration and for $j = 1, \ldots, p$, let $X_{0,j}$ be $\mathcal{F}_j$-measurable and V_j be $\mathcal{F}_{j-1}$-measurable. We assume that for some $m > 2$,

$$\max_{1 \leq j \leq p} \|V_j\|_m := \mu_m < \infty.$$

Lemma 8.2. *Suppose that for all j*

$$\mathbb{E}(X_{0,j}|\mathcal{F}_{j-1}) = 0, \ \mathbb{E}(\exp[\lambda X_{0,j}]|\mathcal{F}_{j-1}) \leq \exp[\lambda^2 V_j^2/2] \ \forall \ \lambda \in \mathbb{R}.$$

If $\{V_j\}_{j=1}^p$ is $\mathcal{F}_0$-measurable then for all $\|u\|_2 = 1$

$$\|\langle X_0 u \rangle\|_m \leq \sqrt{2m}\mu_m.$$

For general predictable $\{V_j\}$ we have for $2 < m_0 < m$ and all $\|u\|_2 = 1$

$$\|\langle X_0 u \rangle\|_m \leq \sqrt{\frac{2m}{m - m_0}} \left(\frac{3m\Gamma(m_0/2 + 1)}{m - m_0} \right)^{1/m_0} \mu_m.$$

8.3. The conditionally Bernstein (or sub-exponential) case

Let as in the previous sub-section $\{\mathcal{F}_j\}_{j=0}^p$ be a filtration and for $j = 1, \ldots, p$, let $X_{0,j}$ be $\mathcal{F}_j$-measurable and V_j be $\mathcal{F}_{j-1}$-measurable and satisfying for some $m > 2$,

$$\max_{1 \leq j \leq p} \|V_j\|_m := \mu_m < \infty.$$

As in the previous section, we prove strong isotropy but now under a different condition.

Lemma 8.3. *Suppose that for some constant K and all j*

$$\mathbb{E}(X_{0,j}|\mathcal{F}_{j-1}) = 0,\ \mathbb{E}(|X_{0,j}|^k|\mathcal{F}_{j-1}) \le \frac{k!}{2}K^{k-2}V_j^2,\ k = 2, 3, \ldots.$$

If the $\{V_j\}_{j=1}^p$ are non-random, then for all $\|u\|_2 = 1$

$$\|\langle X_0 u\rangle\|_m \le \sqrt{2m}\mu_m + mK.$$

If $\{V_j\}_{j=1}^p$ is $\mathcal{F}_0$-measurable we get for all $\|u\|_2 = 1$

$$\|\langle X_0 u\rangle\|_m \le 2^{1-1/m_0}\sqrt{2m}\mu_m + 2^{1-1/m_0} mK.$$

For general predictable $\{V_j\}_{j=1}^p$ we have for all $2 < m_0 < m$ and all $\|u\|_2 = 1$

$$\|\langle X_0 u\rangle\|_{m_0} \le \sqrt{\frac{2m}{m-m_0}}\biggl(\frac{3m\Gamma(m_0/2+1)}{m-m_0}\biggr)^{m_0/2+1}\mu_m + \biggl(3\Gamma(m_0+1)\biggr)^{1/m_0} K.$$

Note that the conditions of the above lemma imply that the entries in X_0 are Bernstein with constants μ_2 and K, where $\mu_2 := \max_{1\le j\le p}\|V_j\|_2 \le \mu_m$. In other words, the conditions of the lemma imply the bound of Theorem 3.1 with $\delta_n = \mu_2\sqrt{2\log(2p)/n} + K\log(2p)/n$ and with m replaced by any $m_0 < m$.

9. Proofs

9.1. Proofs for Section 3

Recall that Theorem 3.1 presents lower bounds for sparse quadratic forms.

Proof of Theorem 3.1. For $Z \in \mathbb{R}$, and $K > 0$, we introduce the truncated version

$$[Z]_K := \begin{cases} -K, & Z < -K \\ Z, & |Z| \le K \\ +K, & Z > K \end{cases}.$$

We obviously have for any $K > 0$ and $u \in \mathbb{R}^p$

$$\|Xu\|_{2,n}^2 \ge \|[Xu]_K\|_{2,n}^2 \tag{9.1}$$

where $[Xu]_K$ is the vector $\{[(Xu)_i]_K :\ i = 1, \ldots, n\}$ with, for $i \in \{1, \ldots, n\}$, $(Xu)_i$ be the i-th component of the vector Xu. Moreover, whenever $\|\langle X_0 u\rangle\|_2 = 1$ by the weak isotropy

$$1 - \|[\langle X_0 u\rangle]_K\|_2^2 \le 2C_m^m K^{-(m-2)}/(m-2).$$

Here, we used the formula

$$1 - \|[\langle X_0 u\rangle]_K\|_2^2 = \int_0^\infty \mathbb{P}(|\langle X_0 u\rangle| > \sqrt{K^2+t})dt.$$

We note that

$$\mathbb{E} \sup_{\|\langle X_0 u\rangle\|_2 = 1,\ \|u\|_1 \le M} \Big| \|[Xu]_K\|_{2,n}^2 - \|[\langle X_0 u\rangle]_K\|_2^2 \Big|$$

$$= \frac{1}{K^2}\mathbb{E} \sup_{\|\langle X_0 u\rangle\|_2 = 1/K,\ \|u\|_1 \le M/K} \Big| \|[Xu]_1\|_{2,n}^2 - \|[\langle X_0 u\rangle]_1\|_2^2 \Big|.$$

Let

$$\mathbf{Z} := \sup_{\|\langle X_0 u\rangle\|_2 = 1/K,\ \|u\|_1 \le M/K} \Big| \|[Xu]_1\|_{2,n}^2 - \|[\langle X_0 u\rangle]_1\|_2^2 \Big|.$$

By symmetrization (see e.g. [20], p. 108) and contraction ([10], p. 112),

$$\mathbb{E}\mathbf{Z} \le 2\mathbb{E} \sup_{\|\langle X_0 u\rangle\|_2 = 1/K,\ \|u\|_1 \le M/K} \left| \frac{1}{n}\sum_{i=1}^n \epsilon_i [(Xu)_i]_1^2 \right|$$

$$\le 8\mathbb{E} \sup_{\|\langle X_0 u\rangle\|_2 = 1/K,\ \|u\|_1 \le M/K} \left| \frac{1}{n}\sum_{i=1}^n \epsilon_i (Xu)_i \right|$$

since the mapping $Z \mapsto [Z]_1^2$ is 2-Lipschitz. Continuing with the last bound, we will apply

$$\mathbb{E} \sup_{\|\langle X_0 u\rangle\|_2 = 1/K} \left| \frac{1}{n}\sum_{i=1}^n \epsilon_i (Xu)_i \right| = \frac{1}{K}\sqrt{\frac{p}{n}}$$

for deriving (3.3) and

$$\mathbb{E} \sup_{\|u\|_1 \le M/K} \left| \frac{1}{n}\sum_{i=1}^n \epsilon_i (Xu)_i \right| = \frac{M}{K}\mathbb{E}\|W\|_\infty \le \frac{M}{K}\delta_n$$

for deriving (3.4). In other words

$$\mathbb{E}\mathbf{Z} \le \frac{8}{K}\min\left\{ M\delta_n, \sqrt{\frac{p}{n}} \right\}.$$

Next we apply the concentration inequality of [3] to $\mathbf{Z}$. We get for all $t > 0$

$$\mathbb{P}\left(\mathbf{Z} \ge \mathbb{E}\mathbf{Z} + \frac{2t}{3n} + \sqrt{2t/n}\sqrt{1/K^2 + 4\mathbb{E}\mathbf{Z}} \right) \le \exp[-t]$$

where we used for $\|\langle X_0 u\rangle\|_2 \le 1/K$ the bound

$$\mathrm{var}([\langle X_0 u\rangle]_1^2) \le \mathbb{E}[\langle X_0 u\rangle]_1^2 \le \|\langle X_0 u\rangle\|_2^2 \le 1/K^2.$$

We invoke that

$$\sqrt{2t/n}\sqrt{1/K^2 + 4\mathbb{E}\mathbf{Z}} \le \sqrt{2t/n}(1/K + 2\sqrt{\mathbb{E}\mathbf{Z}})$$

$$\le \frac{\sqrt{2t/n}}{K} + \frac{2t}{n} + \mathbb{E}\mathbf{Z}.$$

This gives for all $t > 0$

$$\mathbb{P}\biggl(\mathbf{Z} \ge 2\mathbb{E}\mathbf{Z} + \frac{\sqrt{2t/n}}{K} + \frac{8t}{3n}\biggr) \le \exp[-t]$$

and hence

$$\mathbb{P}\biggl(\mathbf{Z} \ge \frac{16}{K}\min\biggl\{M\delta_n, \sqrt{\frac{p}{n}}\biggr\} + \frac{\sqrt{2t/n}}{K} + \frac{8t}{3n}\biggr) \le \exp[-t].$$

So with probability at least $1 - \exp[-t]$

$$\inf_{\|\langle X_0 u\rangle\|_2 = 1,\ \|u\|_1 \le M} \|Xu\|_{2,n}^2 - 1 \ge -\frac{2C_m^m}{(m-2)K^{m-2}}$$

$$- 16K\min\biggl\{M\delta_n, \sqrt{\frac{p}{n}}\biggr\} - K\sqrt{\frac{2t}{n}} - \frac{8K^2 t}{3n}.$$

We now let

$$K := [2C_m^m]^{\frac{1}{m-1}} b^{-\frac{1}{m-1}},$$

where

$$b := 16\min\biggl\{M\delta_n, \sqrt{\frac{p}{n}}\biggr\} + \sqrt{\frac{2t}{n}}.$$

Then

$$\frac{2C_m^m}{(m-2)K^{m-2}} + Kb = D_m b^{\frac{m-2}{m-1}},$$

and

$$\frac{8K^2 t}{3n} \le \frac{8D_m^2}{3}\biggl(\frac{t}{n}\biggr)^{\frac{m-2}{m-1}}.$$

It follows that with probability at least $1 - \exp[-t]$

$$\inf_{\|\langle X_0 u\rangle\|_2 = 1,\ \|u\|_1 \le M} \|Xu\|_{2,n}^2 - 1$$

$$\ge -D_m\biggl(16\min\{M\delta_n, \sqrt{\frac{p}{n}}\biggr\} + \sqrt{\frac{2t}{n}}\biggr)^{\frac{m-2}{m-1}} - \frac{8D_m^2}{3}\biggl(\frac{t}{n}\biggr)^{\frac{m-2}{m-1}}. \qquad \square$$

Remark 9.1. With assumptions weaker than the weak isotropy assumption used in the present paper, for example with the L_1-L_2 property, one can prove lower bounds along the same lines as for Theorem 3.1. One applies instead of the truncation inequality (9.1) in the proof of Theorem 3.1 the inequality

$$\|Xu\|_{2,n} \geq \|Xu\|_{1,n},$$

where

$$\|Xu\|_{1,n} = \frac{1}{n}\sum_{i=1}^{n} |(Xu)_i|$$

One can then proceed using the arguments following (9.1) in the proof of Theorem 3.1 using the Lipschitz property of the absolute value function $Z \mapsto |Z|$. For results assuming only the small ball property, we refer to [9].

We now provide a proof for the sub-Gaussian case along the same lines as the proof of Theorem 3.1.

Proof of Lemma 3.1. We use the same notation as in the proof of Theorem 3.1 for truncation at a value K. Whenever $\|\langle X_0 u\rangle\|_2 = 1$,

$$1 - \|[\langle X_0 u\rangle]_K\|_2^2 \int_0^\infty = \mathbb{P}(|\langle X_0 u\rangle| > \sqrt{K^2 + t})dt$$

$$\leq 2\int_0^\infty \exp[-(K^2 + t)/(2C^2)] = 4C^2 \exp[-K^2/(2C^2)].$$

We choose

$$K = C\sqrt{2\log(C/b)}$$

where

$$b := 16 \min\left\{M\delta_n', \sqrt{\frac{p}{n}}\right\} + \sqrt{\frac{2t}{n}}.$$

The result then follows by the same arguments as those used for Theorem 3.1 and inserting that in the sub-Gaussian case one has $\mathbb{E}\|W\|_\infty \leq \delta_n'$. ☐

9.2. Proofs for Section 4

We first proof the "almost isometric" bound for the compatibility constant and restricted eigenvalue.

Proof of Theorem 4.1. By Theorem 3.1 we know that uniformly in u with $\|u\|_1 \leq M\|\langle X_0 u\rangle\|_2$ with probability at least $1 - \exp[-t]$

$$\|Xu\|_{2,n}^2 \geq (1 - \Delta_n^{\mathrm{L}}(M,t))\|\langle X_0 u\rangle\|_2^2.$$

If $\|u_S\|_1 = 1$ and $\|u_{-S}\|_1 \leq L$ we clearly have

$$\|u\|_1 \leq (L+1) = (L+1)\|u_S\|_1 \leq (L+1)\sqrt{s}\|\langle X_0 u\rangle\|_2/\phi_0(L,S).$$

This implies the lower bound for the compatibility constant. If $\|u_S\|_2 = 1$ and $\|u_{-S}\|_1 \leq L\|u_S\|_1$ we again have $\|u\|_1 \leq (L+1)\sqrt{s}\|u\|_2 \leq (L+1)\sqrt{s}\|\langle X_0 u\rangle\|_2/\kappa_0(L,S)$ which implies the result for the restricted eigenvalue. ☐

We now check the fourth moments, i.e. the second moments of quadratic forms.

Proof of Lemma 4.1. One readily sees that each $X_{0,j}$ has Orlizc norm $\| \cdot \|_{\Psi_1}$ bounded by $K_X + \sigma_X$:

$$\mathbb{E} \exp[|X_{0,j}|/(K_X + \sigma_X)] - 1 \leq 1, \ \forall \ j.$$

Hence for all $t > 0$ and all j

$$\mathbb{P}(|X_{0,j}| > t(K_X + \sigma_X)) \leq 2 \exp[-t].$$

It follows that for all $t > 0$

$$\mathbb{P}\left(\max_j |X_{0,j}| > [t + (1 + c_0) \log(2p)](K_X + \sigma_X) \right) \leq \exp[-t]/(2p)^{c_0}.$$

Clearly

$$\|\langle X_0 u \rangle\|_4^4 = \underbrace{\mathbb{E}|\langle X_0 u \rangle|^4 \mathrm{l}\{\max_j |X_{0,j}| \leq 2(1 + c_0) \log(2p)(K_X + \sigma_X)\}}_{:=i}$$

$$+ \underbrace{\mathbb{E}|\langle X_0 u \rangle|^4 \mathrm{l}\{\max_j |X_{0,j}| > 2(1 + c_0) \log(2p)(K_X + \sigma_X)\}}_{:=ii}.$$

We have for $\|u\|_1 \leq M$ and $\|\langle X_0 u \rangle\|_2 \leq 1$

$$\begin{aligned} i &\leq (2(1 + c_0)M)^2 (K_X + \sigma_X)^2 \log^2(2p) \mathbb{E} \|\langle X_0 u \rangle\|_2^2 \\ &\leq (2(1 + c_0)M)^2 (K_X + \sigma_X)^2 \log^2(2p). \end{aligned}$$

Now for a random variable Z satisfying for all $t > 0$ $\mathbb{P}(|Z| > bt + K/2) \leq c \exp[-t]$ for certain constants b, c and K

$$\begin{aligned} \mathbb{E} Z^4 \mathrm{l}\{|Z| > K\} &= \mathbb{E}(Z^4 - K^4) \mathrm{l}\{|Z| > K\} + K^4 \mathbb{P}(|Z| > K) \\ &= \int_0^\infty \mathbb{P}(Z^4 > t + K^4) dt + K^4 \mathbb{P}(|Z| > K) \\ &\leq \int_0^\infty \mathbb{P}(Z > (t/8)^{1/4} + K/2) dt + K^4 \mathbb{P}(|Z| > K). \end{aligned}$$

Here we used that $(t^{1/4} + K)^4 \leq 8(t + K^4)$ and $8^{1/4} \leq 2$. So we get

$$\mathbb{E} Z^4 \mathrm{l}\{|Z| > K\} \leq 8b^4 \int_0^\infty \mathbb{P}(Z > bs + K/2) ds^4 + K^4 \mathbb{P}(|Z| > K)$$

$$\leq 8b^4 c \int_0^\infty \exp[-s] ds^4 + cK^4 = (8 \times 4!) b^4 c + cK^4 \exp[-K/(2b)]$$

$$\leq (4b)^4 c + cK^4.$$

Apply this to $|Z| := \max_{1\le j\le p}|X_{0,j}|$. Then we can take $b = (K_X + \sigma_X)$, $c = 1/(2p)^{c_0}$ and $K = 2(1+c_0)\log(2p)(K_X+\sigma_X)$. We find

$$\mathbb{E}\max_{1\le j\le p}|X_{0,j}|^4 \mathrm{l}\{\max_{1\le j\le p}|X_{0,j}| > 2(1+c_0)\log(2p)(K_X+\sigma_X)\}$$

$$\le [4(K_X+\sigma_X)]^4/(2p)^{c_0} + \biggl(2(1+c_0)\log(2p)(K_X+\sigma_X)\biggr)^4/(2p)^{c_0}$$

$$\le \biggl(2(1+c_0)\log(2p)(K_X+\sigma_X)\biggr)^4/p^{c_0}$$

since $\log(2p)\ge 1$ and hence $2(1+c_0)\log(2p)\ge 4$. But then

$$\begin{aligned} ii &\le M^4\mathbb{E}\biggl(\max_{1\le j\le p}|X_{0,j}|^4\{\max_{1\le j\le p}|X_{0,j}| > 2(1+c_0)\log(2p)(K_X+\sigma_X)\}\biggr) \\ &\le [2(1+c_0)M]^4(K_X+\sigma_X)^4\log^4(2p)/p^{c_0} \\ &\le [2(1+c_0)M]^2(K_X+\sigma_X)^2\log^2(2p) \end{aligned}$$

where in the last step we invoked the assumption of the lemma. We conclude

$$\|\langle X_0u\rangle\|_4^4 \le i + ii \le 2[2(1+c_0)M]^2(K_X+\sigma_X)^2\log^2(2p). \qquad \square$$

As a result, we can now obtain lower and upper bounds for the compatibility constant and restricted eigenvalue.

Proof of Theorem 4.2. We only need to prove the upper bounds as the lower bounds are from Theorem 4.1. Let $s := |S|$ and let u^* be defined by

$$\phi_0^2(L,S) := s\|\langle X_0u^*\rangle\|_2^2.$$

Then

$$\hat{\phi}^2(L,S) \le s\|Xu^*\|_{2,n}^2 = \phi_0^2(L,S) + s\biggl(\|Xu^*\|_{2,n}^2 - \|\langle X_0u^*\rangle\|_2^2\biggr).$$

But by Chebyshev's inequality, for all $t>0$

$$\mathbb{P}\biggl(\|Xu^*\|_{2,n}^2 - \|\langle X_0u^*\rangle\|_2^2 > \sqrt{\frac{t}{n}}\|\langle X_0u^*\rangle\|_4^2\biggr) \le 1/t.$$

Insert the bound of Lemma 4.1 for $\|\langle X_0u^*\rangle\|_4/\|\langle X_0u^*\rangle\|_2^4$ or, in the case $m>4$, the bound

$$\begin{aligned} \|\langle X_0u^*\rangle\|_4^4 &\le C_m^4 + \int_{C_m^4}^\infty \mathbb{P}(|\langle X_0u^*\rangle| \ge t^{1/4})dt \\ &\le C_m^4 + C_m^m\int_{C_m^4}^\infty t^{-m/4}dt = C_m^4 m/(m-4). \end{aligned}$$

This gives that with probability at least $1-1/t$

$$\begin{aligned} s\biggl(\|Xu^*\|_{2,n}^2 - \|\langle X_0u^*\rangle\|_2^2\biggr) &\le s\|\langle X_0u\rangle^*\|_2^2\Delta_n^{\mathrm{U}}((L+1)\sqrt{s},t) \\ &= \phi_0^2(L,S)\Delta_n^{\mathrm{U}}((L+1)\sqrt{s},t). \end{aligned}$$

The result for the restricted eigenvalue follows in the same way. $\square$

9.3. Proofs for Section 5

We use the transfer principle to obtain lower bounds for sparse quadratic forms.

Proof of Theorem 5.2. We apply result (3.3) of Theorem 3.1 to

$$\inf\{\|Xu_J\|_{2,n} - 1 : \ \|\langle X_0 u_J\rangle\|_2 = 1\}.$$

where J is a fixed subset of $\{1, \ldots, p\}$ with $|J| \le M^2$. There are at most p^{M^2} such subsets. Hence, by the union bound and replacing in the expression (3.3) of Theorem 3.1 the value p by M^2 and t by $t + M^2 \log p$ we have that with probability at least $1 - \exp[-t]$

$$\|Xu_J\|_{2,n}^2 \ge \left(1 - \bar{\Delta}_n^{\mathrm{L}}(M, t)\right)\|\langle X_0 u_J\rangle\|_2^2 \ \forall \ u \in \mathbb{R}^p, \ \forall \ |J| \le M^2.$$

The result follows now from Corollary 5.1. □

To handle the event $\mathcal{B} = \{\max_j \hat{\sigma}_j^2 \le 1 + \epsilon\}$ we gave two lemmas. Here are their proofs.

Proof of Lemma 5.1. Recall we assumed in the beginning of Section 5 that $\|X_{0,j}\|_2 = 1$ for all j. The assumption that the $X_{0,j}$ are sub-Gaussian implies

$$\|X_{0,j}\|_{\Psi_2} \le 2C.$$

Hence

$$\mathbb{E}|X_{0,j}|^{2k} \le k!(4C^2)^k$$

and so

$$\mathbb{E}|X_{0,j}^2 - \mathbb{E}X_{0,j}^2|^k \le 2^{k-1}k!(4C^2)^k = \frac{k!}{2}(8C^2)^k.$$

By Lemma 14.13 in [4] we find

$$\mathbb{P}\left(\max_{1\le j\le p} |\hat{\sigma}_j^2 - 1|/(8C^2) \ge \sqrt{\frac{2\log(2p)}{n}} + \sqrt{\frac{2t}{n}} + \frac{\log(2p)}{n} + \frac{t}{n}\right) \le \exp[-t]. \quad \square$$

Proof of Lemma 5.2. The moment conditions imply that

$$\max_{1\le j\le p} \|X_{0,j}^2\|_k = \max_{1\le j\le p} \|X_{0,j}\|_{2k}^2 \le \kappa_1^2(2k)^{2\alpha} = \kappa_1^2 2^{2\alpha}k^{2\alpha}, \ 1 \le k \le k_0/2.$$

But then

$$\max_{1\le j\le p} \|X_{0,j}^2 - 1\|_k \le \kappa_1^2 2^{2\alpha}k^{2\alpha} + 1 \le \kappa_1^2 2^{2\alpha+1}k^{2\alpha}, \ 1 \le k \le k_0/2.$$

We therefore have by Lemma 3.2

$$\max_{1\le j\le p} \|\hat{\sigma}_j^2 - 1\|_{k_0/2} \le c_0 \exp[4\alpha - 1]\kappa_1^2 2^{2\alpha+1}\sqrt{k_0/(2n)}.$$

It follows that

$$\left(\mathbb{E}\max_{1\le j\le p}|\hat\sigma_j^2-1|^{k_0/2}\right)^{\frac{2}{k_0}} \le p^{\frac{2}{k_0}} c_0 \exp[4\alpha-1]\kappa_1^2 2^{2\alpha+1}\sqrt{k_0/(2n)}$$

$$= c_0\exp[4\alpha-1+2/\eta]\kappa_1^2 2^{2\alpha+1}\sqrt{\eta\log p/(2n)} := c_1\sqrt{\log p/n}.$$

But then by Chebyshev's inequality, for all $t>0$

$$\mathbb{P}\left(\max_{1\le j\le p}\hat\sigma_j^2 - 1 \ge c_1 t^{2/k_0}\sqrt{\log p/n}\right)\le 1/t. \qquad \square$$

Here is the proof for the case of normalized design.

Proof of Theorem 5.3. Consider the event

$$\mathcal{A} := \left\{\|Xu_J\|_{2,n}^2 \ge (1-\Delta)\|\langle X_0 u_J\rangle\|_2^2,\ \forall\, u\in\mathbb{R}^p,\ J\subset\{1,\ldots,p\},\ |J|\le M^2(\Delta)\right\}$$

Then on $\mathcal{A}$, by the transformation $u\mapsto \hat D^{-1/2}u$,

$$u_J^T\hat R u_J \ge (1-\Delta)u_J\hat D^{-1/2}\Sigma_0\hat D^{-1/2}u_J,$$

for all $u\in\mathbb{R}^p,\ J\subset\{1,\ldots,p\},\ |J|\le M^2(\Delta)$. The diagonal of

$$\hat D^{-1/2}(\hat\Sigma-(1-\Delta)\Sigma_0)\hat D^{-1/2}$$

is non-negative on $\mathcal{A}$ and less than or equal to 1. So on $\mathcal{A}$ by the transfer principle (Theorem 5.1) we know for all $u\in\mathbb{R}^p$ with $\|u\|_1^2\le (L+1)^2$ that

$$u^T\hat R u \ge (1-\Delta)u^T\hat D^{-1/2}\Sigma_0\hat D^{-1/2}u - \frac{(L+1)^2}{M^2(\Delta)-1}.$$

We now note that

$$\inf_{\|u_S\|_1=1,\ \|u_{-S}\|_1\le L} u^T\hat D^{-1/2}\Sigma_0\hat D^{-1/2}u = \inf_{\sum_{j\in S}\hat\sigma_j|u_j|=1,\ \sum_{j\notin S}\hat\sigma_j|u_j|\le L} u^T\Sigma_0 u.$$

But on $\mathcal{C}_S$

$$\sum_{j\in S}\hat\sigma_j|u_j| \le \left(\sum_{j\in S}\hat\sigma_j^2\right)^{1/2}\|u_S\|_2 \le \sqrt{s(1+\epsilon)}\|u\|_2.$$

Moreover on $\mathcal{A}$

$$\sum_{j\notin S}\hat\sigma_j|u_j| \ge (1-\sqrt{\Delta})\|u_{-S}\|_1.$$

Hence

$$\inf_{\|u_S\|_1=1,\ \|u_{-S}\|_1\le L} u^T\hat D^{-1/2}\Sigma_0\hat D^{-1/2}u$$

$$\ge \inf_{\|u_S\|_2\ge 1/\sqrt{s(1+\epsilon)},\ \|u_{-S}\|_1\le L/(1-\sqrt{\Delta})} \|\langle X_0 u\rangle\|_2^2$$

$$= \inf_{\|u_S\|_2 \ge 1/\sqrt{s(1+\epsilon)},\ \|u_{-S}\|_1 \le L/(1-\sqrt{\Delta})} \frac{\|\langle X_0 u\rangle\|_2^2}{\|u_S\|_2^2} \|u_S\|_2^2$$

$$\ge \inf_{\|u_S\|_2 = 1,\ \|u_{-S}\|_1 \le L\sqrt{s(1+\epsilon)}/(1-\sqrt{\Delta})} \frac{\|\langle X_0 u\rangle\|_2^2}{s(1+\epsilon)}$$

$$= \frac{\kappa_*^2 (L\sqrt{1+\epsilon}/(1-\sqrt{\Delta}), S)}{s(1+\epsilon)}.$$

The further bounds on the event $\mathcal{A} \cap \mathcal{B}_S$ follow in the same way. $\square$

9.4. Proofs for Section 7

We show that a vector X_0 which is m-th order strongly isotropic and has p sub-Gaussian entries is up to constants "almost bounded" by $\sqrt{\log(2p)}$.

Proof of Lemma 7.1. The sub-Gaussianity implies that for all j and all $s > 0$

$$P(|X_{0,j}|/C \ge \sqrt{2s}) \le 2\exp[-s].$$

We find

$$P\left(\max_{1 \le j \le p} |X_{0,j}|/C \ge \sqrt{2t + 2\log(2p) + 2m(\log n)/(m-2)} \right)$$

$$\le 2p \exp[-(t + \log(2p) + 2m(\log n)/(m-2))] = \exp[-t] n^{-\frac{m}{m-2}}.$$

The proof is finished by applying the inequality

$$\|(\langle X_0 u\rangle) \mathrm{l}\{X_0 \notin A(t)\}\|_2^2 \le \tilde{C}_m^2 \left(P(X_0 \notin A(t)) \right)^{\frac{m-2}{m}}. \quad \square$$

If we have n independent copies of a vector X_0 which is m-th order strongly isotropic and has p sub-Gaussian entries these $n \times p$ variables are up to constants "almost bounded" by $\sqrt{\log(np)}$. For such bounded random variables, we now prove to have uniform convergence of the empirical norm.

Proof of Theorem 7.1. Let $A := A(t)$ be defined as in Lemma 7.1. Recall the notation

$$\|f(X)\|_{2,n}^2 := \frac{1}{n} \sum_{i=1}^n f^2(X_i)$$

so

$$\|(Xu)\mathrm{l}\{X \in A\}\|_{2,n}^2 := \frac{1}{n} \sum_{i=1}^n ((Xu)_i)^2 \mathrm{l}\{X_i \in A\}.$$

Write

$$\|Xu\|_{2,n}^2 - \|\langle X_0 u\rangle\|_2^2 = \|(Xu)\mathrm{l}\{X \in A\}\|_{2,n}^2 - \|(\langle X_0 u\rangle)\mathrm{l}\{X_0 \in A\}\|_2^2$$

$$+ \|(Xu)\mathrm{l}\{X \notin A\}\|_{2,n}^2 - \|(\langle X_0 u\rangle)\mathrm{l}\{X_0 \notin A\}\|_2^2$$

We have

$$\mathbb{P}\bigg(\|(Xu)\mathrm{l}\{X \notin A\}\|_{2,n}^2 \neq 0\bigg) = \mathbb{P}(\exists\ i:\ X_i \notin A)$$

$$= \mathbb{P}(\max_{1\le i\le n}\max_{1\le j\le p}|X_{i,j}| \ge \sigma_X(\sqrt{2t+2\log(2p)+2m\log(n)/(m-2)})$$

$$\le n^{-\frac{m}{m-2}}\exp[-t].$$

Moreover, by the method in [18], for a universal constant c_1 with probability at least $1-\exp[-t]$

$$\sup_{\|\langle X_0 u\rangle\|_2\le 1,\ \|u\|_1\le M}\bigg|\|(Xu)\mathrm{l}\{X\in A\}\|_{2,n} - \|(\langle X_0 u\rangle)\mathrm{l}\{X_0\in A\}\|_2^2\bigg|/c_1$$

$$\le M\sigma_X\sqrt{\frac{(2t+2\log(2pn))(\log p\log^3 n + t)}{n}}$$

$$+ M^2\sigma_X^2\frac{(2t+2\log(2pn))(\log p\log^3 n+t)}{n}. \qquad \square$$

9.5. Proofs for Section 8

First comes the result for directed acyclic graphs.

Proof of Lemma 8.1. We may write $X_0 = X_0 B + \epsilon_0$ where $B = \{\beta_{k,j}\}$ with $\beta_{k,j} = 0$ for $k \ge j$. Thus $X_0(I-B) = \epsilon_0$ so it suffices to show that ϵ_0 is sub-Gaussian with constant C. Note that for $k \neq j$, say $k < j$

$$\mathbb{E}(\epsilon_{0,j}\epsilon_{0,k}) = \mathbb{E}\bigg(\epsilon_{0,k}\mathbb{E}(\epsilon_{0,j}|\mathcal{F}_{j-1})\bigg) = 0.$$

We let $\Omega^2 := \mathrm{diag}(\omega_1^2,\dots,\omega_p^2)$. Suppose that $\|\Omega u\|_2 = 1$. Then for all $\lambda\in\mathbb{R}$

$$\mathbb{E}\exp[\lambda\epsilon_0 u] = \mathbb{E}\exp[\lambda\sum_{j=1}^p \epsilon_{0,j}u_j] \le \exp[\lambda^2 C^2\sum_{j=1}^p u_j^2\omega_j^2/2]$$

$$= \exp[\lambda^2\|\Omega u\|_2^2 C^2/2] = \exp[\lambda^2 C^2/2].$$

Hence, using $\mathrm{e}^{|x|}\le \mathrm{e}^x + \mathrm{e}^{-x}$, for all $\lambda \ge 0$,

$$\mathbb{E}\exp[\lambda|\epsilon_0 u|] \le 2\exp[\lambda^2 C^2/2]. \qquad \square$$

We now prove isotropy under conditional sub-Gaussian assumptions.

Proof of Lemma 8.2. We clearly have for all $\lambda\in\mathbb{R}$ and all $u\in\mathbb{R}^p$

$$\mathbb{E}\bigg(\exp\bigg[\lambda\langle X_0 u\rangle - \lambda^2\sum_{j=1}^p u_j^2 V_j^2/2\bigg]\bigg|\mathcal{F}_0\bigg) \le 1.$$

If the $\{V_j\}_{j=1}^p$ are $\mathcal{F}_0$-measurable this gives

$$\mathbb{E}(\exp[\lambda \langle X_0 u\rangle] | \mathcal{F}_0) \le \exp\biggl[\lambda^2 \sum_{j=1}^p u_j^2 V_j^2/2\biggr].$$

We now use that (see e.g. [4], Lemma 14.7)

$$\mathbb{E}(|\langle X_0 u\rangle|^m | \mathcal{F}_0) \le \left(\frac{m}{\lambda} + \frac{\lambda \sum_{j=1}^p u_j^2 V_j^2}{2}\right)^m,$$

and we choose $\lambda = \sqrt{2m}/(\sum_{j=1}^p u_j V_j)^{1/2}$. This gives

$$\mathbb{E}(|\langle X_0 u\rangle|^m | \mathcal{F}_0) \le (2m)^{m/2}\left(\sum_{j=1}^p u_j V_j\right)^{m/2}.$$

But then

$$\mathbb{E}|\langle X_0 u\rangle|^m \le (2m)^{m/2} \|u\|_2^m \mu_m^m.$$

For the case where $\{V_j\}_{j=1}^p$ is predictable, we use that

$$\mathbb{E}\left(\exp\biggl[\lambda\langle X_0 u\rangle - \lambda^2 \sum_{j=1}^p u_j^2 V_j^2/2\biggr]\right) \le 1.$$

and hence by standard arguments, for any positive a and b

$$\mathbb{P}\left(\frac{|\langle X_0 u\rangle|}{\|u\|_2 \mu_m} \ge a,\ \frac{\sum_{j=1}^p u_j^2 V_j^2}{\|u\|_2^2 \mu_m^2} \le b^2\right) \le 2\exp[-a^2/(2b^2)].$$

Choosing $a = b\sqrt{2m\log b}$ and $b = \mathrm{e}^{s/m}$ gives

$$\mathbb{P}\left(\frac{|\langle X_0 u\rangle|}{\|u\|_2 \mu_m} \ge \mathrm{e}^{s/m}\sqrt{2s},\ \frac{\sum_{j=1}^p u_j^2 V_j^2}{\|u\|_2^2 \mu_m^2} \le \mathrm{e}^{2s/m}\right) \le 2\mathrm{e}^{-s}.$$

We thus find

$$\mathbb{P}\left(\frac{|\langle X_0 u\rangle|}{\|u\|_2 \mu_m} \ge \mathrm{e}^{s/m}\sqrt{2s}\right) \le 3\mathrm{e}^{-s}.$$

We have

$$\begin{aligned}
\frac{\mathbb{E}|\langle X_0 u\rangle|^{m_0}}{\|u\|_2^{m_0}\mu_m^{m_0}} &= \int_0^\infty \mathbb{P}\left(\frac{|\langle X_0 u\rangle|}{\|u\|_2 \mu_m} > t^{1/m_0}\right) dt \\
&= \int_0^\infty \mathbb{P}\left(\frac{|\langle X_0 u\rangle|}{\|u\|_2 \mu_m} > \mathrm{e}^{s/m}\sqrt{2s}\right) d(\mathrm{e}^{s/m}\sqrt{2s})^{m_0} \\
&\le 3\int_0^\infty \mathrm{e}^{-s} d(\mathrm{e}^{s/m}\sqrt{2s})^{m_0} \\
&= 3\int_0^\infty (\mathrm{e}^{s/m}\sqrt{2s})^{m_0}\mathrm{e}^{-s} ds \\
&= 3(2^{m_0/2})\left(\frac{m}{m-m_0}\right)^{m_0/2+1}\Gamma(m_0/2+1).
\end{aligned}$$

□

The final proof concerns isotropy under conditional sub-exponential assumptions.

Proof of Lemma 8.3. We invoke the inequality

$$\log \mathbb{E} \exp[Z] \le \mathbb{E}e^{|Z|} - 1 - \mathbb{E}|Z|$$

(see e.g. Lemma 14.1 in [4]) which holds for a random variable $Z \in \mathbb{R}$ with mean zero. Moreover

$$\mathbb{E}e^{|Z|} - 1 - \mathbb{E}|Z| = \sum_{k=2}^{\infty} \mathbb{E}|Z|^k / k!.$$

By the Bernstein condition one readily sees that for all $\lambda \in \mathbb{R}$ and $u \in \mathbb{R}^p$ with $|\lambda| K \|u\|_2 < 1$

$$\mathbb{E}\biggl(\exp\biggl[\lambda \langle X_0 u \rangle - \lambda^2 \sum_{j=1}^p u_j^2 V_j^2 / (2(1 - |\lambda| K \|u\|_2) \biggr] \bigg| \mathcal{F}_0 \biggr) \le 1.$$

If the $\{V_j\}_{j=1}^p$ are $\mathcal{F}_0$-measurable this gives

$$\mathbb{E}(\exp[\lambda |\langle X_0 u \rangle|] | \mathcal{F}_0) \le \exp\biggl[\lambda^2 \sum_{j=1}^p u_j^2 V_j^2 / (2(1 - |\lambda| K \|u\|_2) \biggr].$$

So then (see e.g. [4], Lemma 14.7) for $0 < \lambda < K\|u\|_2$

$$\mathbb{E}(|\langle X_0 u \rangle|^m | \mathcal{F}_0) \le \left(\frac{m}{\lambda} + \frac{\lambda \sum_{j=1}^p u_j^2 V_j^2}{2(1 - \lambda K \|u\|_2)} \right)^m .$$

Now choose

$$\frac{1}{\lambda} = K\|u\|_2 + \left(\frac{\sum_{j=1}^p u_j^2 V_j^2}{2m} \right)^{1/2}.$$

Then we get

$$\mathbb{E}(|\langle X_0 u \rangle|^m | \mathcal{F}_0) \le \left(\sqrt{2m} \biggl(\sum_{j=1}^p u_j^2 V_j^2 \biggr)^{1/2} + m\|u\|_2 K \right)^m .$$

This implies the result for non-random $\{V_j\}_{j=1}^p$. If they are $\mathcal{F}_0$-measurable we find

$$\mathbb{E}|\langle X_0 u \rangle|^m \le 2^{m-1} \Bigl(\sqrt{2m} \|u\|_2 \mu_m \Bigr)^m + 2^{m-1} \Bigl(m \|u\|_2 K \Bigr)^m .$$

If the $\{V_j\}_{j=1}^p$ are only predictable, we use that for all positive a and b and for $\tilde{K} := K/\mu_m$

$$\mathbb{P}\biggl(\frac{|\langle X_0 u \rangle|}{\|u\|_2 \mu_m} \ge b\sqrt{2a} + \tilde{K} a, \ \frac{\sum_{j=1}^p u_j^2 V_j^2}{\|u\|_2^2 \mu_m^2} \le b^2 \biggr) \le 2\exp[-a].$$

Write $a = s$ and $b = \mathrm{e}^{s/m}$ to find that

$$\mathbb{P}\biggl(\frac{|\langle X_0 u\rangle|}{\|u\|_2 \mu_m} \ge \mathrm{e}^{s/m}\sqrt{2s} + \tilde{K}s, \ \frac{\sum_{j=1}^p u_j^2 V_j^2}{\|u\|_2^2 \mu_m^2} \le \mathrm{e}^{2s/m} \biggr) \le 2\exp[-s]$$

and so

$$\mathbb{P}\biggl(\frac{|\langle X_0 u\rangle|}{\|u\|_2 \mu_m} \ge \mathrm{e}^{s/m}\sqrt{2s} + \tilde{K}s \biggr) \le 3\exp[-s].$$

It follows that

$$\mathbb{E}\biggl(\frac{|\langle X_0 u\rangle|}{\|u\|_2 \mu_m} \biggr)^{m_0} = \int_0^\infty \mathbb{P}\biggl(\frac{|\langle X_0 u\rangle|}{\|u\|_2 \mu_m} \ge t^{1/m_0} \biggr) dt$$

$$\le 3 \int_0^\infty \Bigl(\mathrm{e}^{s/m}\sqrt{2s} + \tilde{K}s \Bigr)^{m_0} \mathrm{e}^{-s} ds.$$

But

$$\int_0^\infty \Bigl(\mathrm{e}^{s/m}\sqrt{2s} \Bigr)^{m_0} \mathrm{e}^{-s} ds = (2^{m_0/2}) \biggl(\frac{m}{m - m_0} \biggr)^{m_0/2+1} \Gamma(m_0/2 + 1)$$

and

$$\int_0^\infty (\tilde{K}s)^{m_0} \mathrm{e}^{-s} ds = \tilde{K}^{m_0} \Gamma(m_0 + 1).$$

Hence by the triangle inequality

$$\frac{\|\langle X_0 u\rangle\|_{m_0}}{\|u\|_2 \mu_m} \le \sqrt{\frac{2m}{m - m_0}} \biggl(\frac{3m\Gamma(m_0/2 + 1)}{m - m_0} \biggr)^{m_0/2+1} + \tilde{K} \biggl(3\Gamma(m_0+1) \biggr)^{1/m_0}. \ \square$$

References

[1] Adamczak, R., Litvak, A. E., Pajor, A., and Tomczak-Jaegermann, N., Quantitative estimates of the convergence of the empirical covariance matrix in log-concave ensembles. *J. Amer. Math. Soc*, 23(2):535–561, 2010. MR2601042

[2] Bickel, P., Ritov, Y., and Tsybakov, A., Simultaneous analysis of Lasso and Dantzig selector. *Annals of Statistics*, 37:1705–1732, 2009. MR2533469

[3] Bousquet, O., A Bennet concentration inequality and its application to suprema of empirical processes. *Comptes Rendus de l'Académie des Sciences, Paris*, 334:495–550, 2002. MR1890640

[4] Bühlmann, P. and van de Geer, S., *Statistics for High-Dimensional Data: Methods, Theory and Applications*. Springer, 2011. MR2807761

[5] Chen, S. S., Donoho, D. L., and Saunders, M. A., Atomic decomposition by basis pursuit. *SIAM Journal on Scientific Computing*, 20(1):33–61, 1998. MR1639094

[6] Gribonval, R. and Nielsen, M., Highly sparse representations from dictionaries are unique and independent of the sparseness measure. *Applied and Computational Harmonic Analysis*, 22(3):335–355, 2007. MR2311858

[7] Koltchinskii, V. and Mendelson, S., Bounding the smallest singular value of a random matrix without concentration. *Arxiv preprint arXiv: 1312.3580*, 2013.

[8] Lecué, G. and Mendelson, S., Necessary moment conditions for exact reconstruction via basis pursuit. *Arxiv preprint arXiv: 404.3116*, 2014.

[9] Lecué, G. and Mendelson, S., Compressed sensing under weak moment assumptions. *Arxiv preprint arXiv: 1401.2188*, 2014.

[10] Ledoux, M. and Talagrand, M., *Probability in Banach Spaces: Isoperimetry and Processes*. Springer Verlag, New York, 1991. MR1102015

[11] Loh, P.-L. and Wainwright, M. J., High-dimensional regression with noisy and missing data: Provable guarantees with non-convexity. *Annals of Statistics*, 40:1637–1664, 2012. MR3015038

[12] Mendelson, S., Pajor, A., and Tomczak-Jaegermann, N., Uniform uncertainty principle for Bernoulli and subgaussian ensembles. *Constructive Approximation*, 28(3):277–289, 2008. MR2453368

[13] Oliveira, R. I., The lower tail of random quadratic forms, with applications to ordinary least squares and restricted eigenvalue properties. *Arxiv preprint arXiv: 1312.2903*, 2013.

[14] Raskutti, G., Wainwright, M. J., and Yu, B., Restricted eigenvalue properties for correlated Gaussian designs. *Journal of Machine Learning Research*, 11:2241–2259, 2010. MR2719855

[15] Rudelson, M. and Zhou, S., Reconstruction from anisotropic random measurements. *Journal of Machine Learning Research – Proceedings Track 23*, pages 10.1–10.24, 2012. MR3061256

[16] Srivastava, N. and Vershynin, R., Covariance estimation for distributions with $2+\varepsilon$ moments. *The Annals of Probability*, 41(5):3081–3111, 2013. MR3127875

[17] van de Geer, S., The deterministic Lasso. *The JSM Proceedings*, 2007.

[18] van de Geer, S., On the uniform convergence of empirical norms and inner products, with application to causal inference. *Electronic Journal of Statistics*, 8:543–574, 2014. MR3211024

[19] van de Geer, S. and Bühlmann, P., On the conditions used to prove oracle results for the Lasso. *Electronic Journal of Statistics*, pages 1360–1392, 2009. MR2576316

[20] van der Vaart, A. W. and Wellner, J. A., *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer-Verlag, New York, 1996. ISBN 0-387-94640-3. MR1385671