

Restricted nullspace and eigenvalues for random designs

Kaveh S. Nobari

Lectures in High-Dimensional Statistics

Department of Mathematics and Statistics
Lancaster University

Contents

- 1 Motivation
- 2 Preliminaries: Gaussian comparison inequalities
 - A general comparison result
 - Sudakov-Fernique inequality & Gordon's inequality
- 3 Theorem 7.16
 - Restricted eigenvalue condition
 - Proof

1 Motivation

2 Preliminaries: Gaussian comparison inequalities

- A general comparison result
- Sudakov-Fernique inequality & Gordon's inequality

3 Theorem 7.16

- Restricted eigenvalue condition
- Proof

Motivation

- Recall Theorem 7.13 of Wainwright (2019), which provides the bounds $\|\hat{\theta} - \theta^*\|_2$ for the Lagrangian Lasso, Constrained Lasso and the Relaxed Basis Pursuit Program, on the premise that the design matrix X satisfies the restricted eigenvalue (RE) condition.
- In practice it is difficult to verify that a given design matrix X satisfies this condition.
- It is, however, possible to give **high-probability results** in the case of **random design matrices**.
- **Pairwise incoherence** and **RIP** conditions are **sufficient** conditions to certify the restricted nullspace and eigenvalue properties, and are suitable for **isotropic** designs (i.e. Identity population covariance matrix for the rows X_i of the matrix).
- The following Theorem provides results for random design matrices that do not necessarily have isotropic structure.

- 1 Motivation
- 2 Preliminaries: Gaussian comparison inequalities
 - A general comparison result
 - Sudakov-Fernique inequality & Gordon's inequality
- 3 Theorem 7.16
 - Restricted eigenvalue condition
 - Proof

A general comparison result

We are given two Gaussian processes, say $\{Y_\theta, \theta \in \mathbb{T}\}$ and $\{Z_\theta, \theta \in \mathbb{T}\}$, both indexed by the same set \mathbb{T} . Suppose we wish to compare the two processes in some sense. Let

$$F(X) := \sup_{\theta \in \mathbb{T}} X_\theta$$

and suppose we would like to assess the conditions under which $\mathbb{E}[F(X)] > \mathbb{E}[F(Y)]$ or vice-versa. The results which permit such comparisons are known as **Gaussian comparison inequalities**.

In general, it is useful to have a uniform control (or have a bound) on a random process, $(X_\theta)_{\theta \in \mathbb{T}}$

$$\mathbb{E}[\sup_{\theta \in \mathbb{T}} X_\theta].$$

A general comparison result

The suprema of Gaussian processes are defined by taking limits of maxima over finite subsets. Hence, it is sufficient to consider a **finite** \mathbb{T} , say $\mathbb{T} = \{1, \dots, N\}$ for some integer N , where for convenience, we denote

$$[N] = \{1, \dots, N\}.$$

In what follows, we first consider general comparison inequalities and subsequently Slepian's comparison inequality for Gaussian processes. It essentially suggests that **the faster the process grows, the farther it gets** [see Vershynin (2018)].

Theorem (A general comparison result)

Let (X_1, \dots, X_N) and (Y_1, \dots, Y_N) be a pair of centered Gaussian random vectors, and suppose there exist disjoint subsets A and B of $[N] \times [N]$, such that

$$\begin{aligned}\mathbb{E}[X_i X_j] &\leq \mathbb{E}[Y_i Y_j], & \forall (i, j) \in A, \\ \mathbb{E}[X_i X_j] &\geq \mathbb{E}[Y_i Y_j], & \forall (i, j) \in B, \\ \mathbb{E}[X_i X_j] &= \mathbb{E}[Y_i Y_j], & \forall (i, j) \notin A \cup B.\end{aligned}$$

Let $F : \mathbb{R}^N \rightarrow \mathbb{R}$ be a twice-differentiable function, and suppose that

$$\begin{aligned}\frac{\partial^2 F}{\partial u_i \partial u_j}(u) &\geq 0 & \forall (i, j) \in A \\ \frac{\partial^2 F}{\partial u_i \partial u_j}(u) &\leq 0 & \forall (i, j) \in B.\end{aligned}$$

Then, we are guaranteed that $\mathbb{E}[F(X)] \leq \mathbb{E}[F(Y)]$.

An important Corollary of the above Theorem is **Slepian's inequality**.

Corollary (Slepian's inequality)

Let $X \in \mathbb{R}^N$ and $Y \in \mathbb{R}^N$ be centered Gaussian random vectors, such that

$$\begin{aligned}\mathbb{E}[X_i X_j] &\geq \mathbb{E}[Y_i Y_j], \quad \forall i \neq j, \\ \mathbb{E}[X_i^2] &= \mathbb{E}[Y_i^2], \quad \forall i = 1, \dots, N.\end{aligned}$$

Then for every $\tau \in \mathbb{R}$, we have

$$P \left[\max_{i=1, \dots, N} X_i \geq \tau \right] \leq P \left[\max_{i=1, \dots, N} Y_i \geq \tau \right].$$

Then we are guaranteed

$$\mathbb{E} \left[\max_{i=1, \dots, N} X_i \right] \leq \mathbb{E} \left[\max_{i=1, \dots, N} Y_i \right].$$

In what follows, we show an extension of Slepian's inequality - namely, Gordon's inequality.

Corollary (Gordon's inequality)

Let $(X_{ui})_{u \in U, i \in N}$ and $(Y_{ui})_{u \in U, i \in N}$ be two centered Gaussian processes indexed by pair of points (u, i) in a product set $U \times N$. Assume that we have

$$\begin{aligned} \mathbb{E}[X_{ui}^2] &= \mathbb{E}[Y_{ui}^2], & \mathbb{E}[(X_{ui} - X_{uj})^2] &\leq \mathbb{E}[(Y_{ui} - Y_{uj})^2], & \forall u, i, j \\ \mathbb{E}[(X_{ui} - X_{vj})^2] &\geq \mathbb{E}[(Y_{ui} - Y_{vj})^2], & \forall u \neq v, \forall i, j \end{aligned}$$

Then fore every $\tau \geq 0$, we have

$$P \left[\inf_{u \in U} \sup_{i \in [N]} X_{ui} \geq \tau \right] \leq P \left[\inf_{u \in U} \sup_{i \in [N]} Y_{ui} \geq \tau \right].$$

Consequently,

$$\mathbb{E} \left[\inf_{u \in U} \sup_{i \in [N]} X_{ui} \right] \leq \mathbb{E} \left[\inf_{u \in U} \sup_{i \in [N]} Y_{ui} \right].$$

1 Motivation

2 Preliminaries: Gaussian comparison inequalities

- A general comparison result
- Sudakov-Fernique inequality & Gordon's inequality

3 Theorem 7.16

- Restricted eigenvalue condition
- Proof

Sudakov-Fernique inequality

Note that the earlier results are expressed in terms of variances and covariances. In many cases, it is more convenient to compare two Gaussian processes in terms of their associated **pseudomatrix**

$$\rho_X^{(i,j)} = \mathbb{E} [(X_i - X_j)^2], \quad \text{and} \quad \rho_Y^{(i,j)} = \mathbb{E} [(Y_i - Y_j)^2],$$

which leads to the Sudakov-Fernique Theorem.

Theorem (Sudakov-Fernique inequality)

Given a pair of centered N -dimensional Gaussian vectors (X_1, \dots, X_n) and (Y_1, \dots, Y_N) , suppose that

$$\mathbb{E}[(X_i - X_j)^2] \leq \mathbb{E}[(Y_i - Y_j)^2], \quad \forall (i, j) \in [N] \times [N].$$

Then $\mathbb{E}[\max_{j=1, \dots, N} X_j] \leq \mathbb{E}[\max_{j=1, \dots, N} Y_j]$

- 1 Motivation
- 2 Preliminaries: Gaussian comparison inequalities
 - A general comparison result
 - Sudakov-Fernique inequality & Gordon's inequality
- 3 Theorem 7.16
 - Restricted eigenvalue condition
 - Proof

Theorem 7.16

Theorem (7.16)

Consider a random matrix $X \in \mathbb{R}^{n \times d}$, in which each row $x_i \in \mathbb{R}^d$ is drawn i.i.d. from $N(0, \Sigma)$. Moreover, let $\rho^2(\Sigma)$ be the maximum diagonal entry of a covariance matrix Σ . Then there are universal positive constants $c_1 < 1 < c_2$, such that

$$\frac{\|X\theta\|_2^2}{n} \geq c_1 \|\sqrt{\Sigma}\theta\|_2^2 - c_2 \rho^2(\Sigma) \frac{\log d}{n} \|\theta\|_1^2, \quad \forall \theta \in \mathbb{R}^d \quad (1)$$

with probability at least $1 - \frac{\exp(-n/32)}{1 - \exp(-n/32)}$.

1 Motivation

2 Preliminaries: Gaussian comparison inequalities

- A general comparison result
- Sudakov-Fernique inequality & Gordon's inequality

3 Theorem 7.16

- Restricted eigenvalue condition
- Proof

Proof

Let us restrict our attention to vectors belonging in the ellipse $\mathbb{S}^{d-1}(\Sigma) = \{\theta \in \mathbb{R}^d \mid \|\sqrt{\Sigma}\theta\|_2 = 1\}$. In which case, the results of the Theorem simplifies to

$$\frac{\|X\theta\|_2^2}{n} \geq c_1 - c_2 \rho^2(\Sigma) \frac{\log d}{n} \|\theta\|_1^2, \quad \forall \theta \in \mathbb{R}^d$$

Furthermore, we define the function $g(t) := 2\rho(\Sigma)\sqrt{\frac{\log d}{n}}t$, and the associated **bad** event

$$\xi := \left\{ X \in \mathbb{R}^{n \times d} \mid \inf_{\theta \in \mathbb{S}^{d-1}(\Sigma)} \frac{\|X\theta\|_2}{\sqrt{n}} \leq \frac{1}{4} - 2g(\|\theta\|_1) \right\}. \quad (2)$$

Note that squaring both sides of the bound (2) yields

$$\inf_{\theta \in \mathbb{S}^{d-1}(\Sigma)} \frac{\|X\theta\|_2^2}{n} \leq \frac{1}{16} - g(\|\theta\|_1) + 4g^2(\|\theta\|_1).$$

Proof

It is first claimed that on ξ^c , the lower bound (16) holds. Defining $a = \frac{1}{4}$, $b = 2g(\|\theta\|_1)$ and $c = \frac{\|X\theta\|_2}{\sqrt{n}}$, we have

$$c \geq \max\{a - b, 0\}$$

on the event ξ^c . Further, this lower bound implies

$$c^2 \geq (1 - \delta)^2 a^2 - \frac{b^2}{\delta^2} \quad \text{for any } \delta \in (0, 1).$$

If $\frac{b}{\delta} \geq a$, then the lower bound is trivial. Otherwise, we may assume that $b \leq \delta a$, in which case the bound $c \geq a - b$ implies that $c \geq (1 - \delta)a$, and hence that $c^2 \geq (1 - \delta)^2 a^2$. Settings $(1 - \delta)^2 = \frac{1}{2}$ then yields the claim. Hence, the remainder of the proof is devoted to upper bounding $P[\xi]$

Proof

For radii (r_ℓ, r_u) such that $0 \leq r_\ell \leq r_u$, let us define the sets

$$\mathbb{K}(r_\ell, r_u) := \{\theta \in \mathbb{S}^{d-1}(\Sigma) \mid g(\|\theta\|_1) \in [r_\ell, r_u]\} \quad (3)$$

along with the events

$$\mathcal{A}(r_\ell, r_u) := \left\{ \inf_{\theta \in \mathbb{K}(r_\ell, r_u)} \frac{\|X\theta\|_2}{\sqrt{n}} \leq \frac{1}{2} - 2r_u \right\}. \quad (4)$$

Proof

Lemma

For any pair of radii $0 \leq r_\ell \leq r_u$, we have

$$P[\mathcal{A}(r_\ell, r_u)] \leq \exp\left(-\frac{n}{32}\right) \exp\left(-\frac{n}{2}r_u^2\right). \quad (5)$$

Moreover, for $\mu = 1/4$, we have

$$\xi \subseteq \mathcal{A}(0, \mu) \cup \left(\bigcup_{\ell=1}^{\infty} \mathcal{A}(2^{\ell-1}\mu, 2^\ell\mu) \right). \quad (6)$$

Before completing the proof, first, recall Boole's inequality:

Proof

Definition (Boole's inequality)

For a countable set of events E_1, E_2, \dots , we have

$$P\left(\bigcup_i E_i\right) \leq \sum_i P(E_i)$$

For more details see Grimmett and Stirzaker (2020).

Therefore, from (6) and the union bound provided by Boole's inequality, we have

$$\begin{aligned} P[\xi] &\leq P[\mathcal{A}(0, \mu)] + \sum_{l=1}^{\infty} P[\mathcal{A}(2^{l-1}\mu, 2^l\mu)] \\ &\leq \exp\left(-\frac{n}{32}\right) \left\{ \sum_{\ell=0}^{\infty} \exp\left(-\frac{n}{2}\right) 2^{2\ell} \mu^2 \right\}. \end{aligned}$$

Proof

Since $\mu = 1/4$ and $2^{2\ell} \geq 2\ell$, we have

$$\begin{aligned} P[\xi] &\leq \exp\left(-\frac{n}{32}\right) \sum_{\ell=0}^{\infty} \exp\left(-\frac{n}{2} 2^{2\ell} \mu^2\right) \leq \exp\left(-\frac{n}{32}\right) \sum_{\ell=0}^{\infty} \exp(-n\mu^2)^\ell \\ &\leq \frac{\exp\left(-\frac{n}{32}\right)}{1 - \exp\left(-\frac{n}{32}\right)} \end{aligned}$$

Proof of the Lemma:

Let us begin with (6). Let $\theta \in \mathbb{S}^{d-1}(\Sigma)$ validate the event ξ ; then it must belong either to $\mathbb{K}(0, \mu)$ or to a set $\mathbb{K}(2^{\ell-1}\mu, 2^\ell\mu)$, for some $\ell = 1, 2, \dots$.

Case 1: First suppose $\theta \in \mathbb{K}(0, \mu)$, so that $g(\|\theta\|_1) \leq \mu = 1/4$. Since θ validates ξ , we have

$$\frac{\|X\theta\|_2}{\sqrt{n}} \leq \frac{1}{4} - 2g(\|\theta\|_1) \leq \frac{1}{4} = \frac{1}{2} - \mu,$$

Proof

showing that event $\mathcal{A}(0, \mu)$ must happen.

Case 2: Otherwise, we must have $\theta \in \mathbb{K}(2^{\ell-1}\mu, 2^\ell\mu)$ for some $\ell = 1, 2, \dots$, and moreover

$$\frac{\|X\theta\|_2}{\sqrt{n}} \leq \frac{1}{4} - 2g(\|\theta\|_1) \leq \frac{1}{2} - 2(2^\ell\mu) \leq \frac{1}{2} - 2^\ell\mu.$$

which shows that the event $\mathcal{A}(2^{\ell-1}\mu, 2^\ell\mu)$ must happen.

Let us now establish the tail bound (5). It is equivalent to upper bound the random variable $T(r_\ell, r_u) := -\inf_{\theta \in \mathbb{K}(r_\ell, r_u)} \frac{\|X\theta\|_2}{\sqrt{n}}$. By the dual representation of the l_2 -norm [see Muscat (2014)], we have

$$T(r_\ell, r_u) = -\inf_{\theta \in \mathbb{K}(r_\ell, r_u)} \sup_{u \in \mathbb{S}^{n-1}} \frac{\langle u, X\theta \rangle}{\sqrt{n}} = \sup_{\theta \in \mathbb{K}(r_\ell, r_u)} \inf_{u \in \mathbb{S}^{n-1}} \frac{\langle u, X\theta \rangle}{\sqrt{n}}.$$

Proof

We now may write $X = W\sqrt{\Sigma}$, where $W \in \mathbb{R}^{n \times d}$ is standard Gaussian matrix and define the vector $v = \sqrt{\Sigma}\theta$, then

$$- \inf_{\theta \in \mathbb{K}(r_\ell, r_u)} \frac{\|X\theta\|_2}{\sqrt{n}} = \sup_{v \in \tilde{\mathbb{K}}(r_\ell, r_u)} \inf_{u \in \mathbb{S}^{n-1}} \underbrace{\frac{\langle u, Wv \rangle}{\sqrt{n}}}_{Z_{u,v}} \quad (7)$$

where $\tilde{\mathbb{K}}(r_\ell, r_u) = \{v \in \mathbb{R}^d \mid \|v\|_2 = 1, g(\Sigma^{-\frac{1}{2}}v) \in [r_\ell, r_u]\}$.

The pair (u, v) range over a subset of $\mathbb{S}^{n-1} \times \mathbb{S}^{d-1}$, hence each $Z_{u,v}$ is a centered Gaussian variable with variance n^{-1} . Furthermore, Gordon's **Gaussian comparison principle** introduced earlier may be applied - i.e. to compare the Gaussian process $\{Z_{u,v}\}$ to the centered Gaussian process with elements

$$Y_{u,v} := \frac{\langle g, u \rangle}{\sqrt{n}} + \frac{\langle h, v \rangle}{\sqrt{n}},$$

Proof

where $g \in \mathbb{R}^n$, $h \in \mathbb{R}^d$ have i.i.d. $N(0, 1)$ entries. Applying Gordon's inequality we find that

$$\begin{aligned} \mathbb{E}[T(r_\ell, r_u)] &= \mathbb{E} \left[\sup_{v \in \tilde{\mathbb{K}}(r_\ell, r_u)} \inf_{u \in \mathbb{S}^{n-1}} Z_{u,v} \right] \leq \mathbb{E} \left[\sup_{v \in \tilde{\mathbb{K}}(r_\ell, r_u)} \inf_{u \in \mathbb{S}^{n-1}} Y_{u,v} \right] \\ &= \mathbb{E} \left[\sup_{v \in \tilde{\mathbb{K}}(r_\ell, r_u)} \frac{\langle h, v \rangle}{\sqrt{n}} \right] \\ &\quad + \mathbb{E} \left[\inf_{u \in \mathbb{S}^{n-1}} \frac{\langle g, u \rangle}{\sqrt{n}} \right] \\ &= \mathbb{E} \left[\sup_{\theta \in \mathbb{K}(r_\ell, r_u)} \frac{\langle \sqrt{\Sigma} h, \theta \rangle}{\sqrt{n}} \right] - \mathbb{E} \left[\frac{\|g\|_2}{\sqrt{n}} \right]. \end{aligned}$$

Proof

On one hand, we have $\mathbb{E}[\|g\|_2] \geq \sqrt{n}\sqrt{\frac{2}{\pi}}$. On the other hand, applying Holder's inequality yields

$$\mathbb{E} \left[\sup_{\theta \in \mathbb{K}(r_\ell, r_u)} \frac{\langle \sqrt{\Sigma} h, \theta \rangle}{\sqrt{n}} \right] \leq \mathbb{E} \left[\sup_{\theta \in \mathbb{K}(r_\ell, r_u)} \|\theta\|_1 \frac{\|\sqrt{\Sigma} h\|_\infty}{\sqrt{n}} \right] \stackrel{(i)}{\leq} r_u$$

where step (i) follows since $\mathbb{E} \left[\frac{\|\sqrt{\Sigma} h\|_\infty}{\sqrt{n}} \right] \leq 2\rho(\Sigma) \sqrt{\frac{\log d}{n}}$ and $\sup_{\theta \in \mathbb{K}(r_\ell, r_u)} \|\theta\|_1 \leq \frac{r_u}{2\rho(\Sigma) \sqrt{(\log d)/n}}$ from (3) of \mathbb{K} . Putting it all together, we have shown that

$$\mathbb{E}[T(r_\ell, r_u)] \leq -\sqrt{\frac{2}{\pi}} + r_u. \quad (8)$$

From the dual representation (7), we see that the random variable $\sqrt{n}T(r_\ell, r_u)$ is a 1-Lipschitz function of standard Gaussian matrix W , so that it implies the upper tail bound $P[T(r_\ell, r_u) \geq \mathbb{E}[T(r_\ell, r_u)] + \delta] \leq e^{-\delta^2/2}$ for all $\delta > 0$. Define

Proof

the constant $C = \sqrt{\frac{2}{\pi}} - \frac{1}{2} \geq \frac{1}{4}$. Setting $\delta = C + r_u$ and using our upper bound on the mean (8) yields

$$\begin{aligned} P[T(r_\ell, r_u) \geq -\frac{1}{2} + 2r_u] &\leq \exp\left(-\frac{n}{2}C^2\right) \exp\left(-\frac{n}{2}Cr_u^2\right) \\ &\leq \exp\left(-\frac{n}{32}\right) \exp\left(-\frac{n}{2}r_u^2\right) \end{aligned}$$

as claimed.

References

- Grimmett, G. and Stirzaker, D. (2020). *Probability and random processes*. Oxford university press.
- Muscat, J. (2014). *Functional analysis: an introduction to metric spaces, Hilbert spaces, and Banach algebras*. Springer.
- Vershynin, R. (2018). *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge university press.
- Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, volume 48. Cambridge University Press.