Restricted Eigenvalue Properties for Correlated Gaussian Designs

Garvesh Raskutti, Martin Wainwright, Bin Yu

March 24, 2014

Outline

- Introduction
- Various conditions on design matrix
  - Restricted Nullspace condition
  - Restricted Isometry Property
  - Restricted Eigenvalue Condition
- Main results
- Proof of result
- Application examples

Problem Overview

- High-dimensional sparse models

$$y = X\beta^* + w, \qquad y \in \mathbb{R}^n, X \in \mathbb{R}^{n \times p}, w \sim (0, \sigma^2 I_{n\times n}), p >> n$$

- Assumption of exact sparsity

$$S(\beta^*) := \{j \in \{1, ..., p\} | \beta_j^* \neq 0\}, \qquad |S| \leq s$$

- Problem reduces to: Find $\hat{\beta}$ close to β^* such that $\|\beta\|_0 \leq s$
- Convex relaxation: Use ℓ_1-norm

$$\text{Basis pursuit: } \hat{\beta} \in arg \min_{\beta \in \mathbb{R}^p} \|\beta\|_1 \quad \text{such that} \quad X\beta = y$$

$$\text{Lasso: } \hat{\beta} \in arg \min_{\beta \in \mathbb{R}^p} \{\|y - X\beta\|_2^2 + \lambda \|\beta\|_1\}$$

- Under what conditions on matrix X can we recover $\hat{\beta}$?

Restricted Nullspace condition

- Define any set $S \subset \{1, ..., p\}$
- Notations: **n** - number of observations, **p** - number of covariates, ,**k** - sparsity level
- For some constant $\alpha \geq 1$, define the set

$$C(S;\alpha) := \{\theta \in \mathbb{R}^p \mid \|\theta_{S^c}\|_1 \leq \alpha \|\theta_S\|_1\}$$

Restricted Nullspace condition

For a given sparsity index $k \leq p$, the matrix X satisfies the restricted nullspace (RN) condition of order k if $null(X) \cap C(S;1) = \{0\}$ for all subsets of cardinality k

- A sufficient and necessary condition for exact recovery in the noisless setting

Restricted Isometry Property

- For a matrix X define for every integer $1 \leq s \leq |S|$, where $S \subset \{1, ..., p\}$, define the s-restricted isometry constants δ_s to be the smallest quantity such that X_S obeys

$$(1 - \delta_s)\|\beta\|_2^2 \leq \|X_S\beta\|_2^2 \leq (1 + \delta_s)\|\beta\|_2^2$$

  for all subsets $S \subset \{1, ..., p\}$ of cardinality at most s, and all real coefficients $(\beta_j)_{j \in S}$

- RIP requires $\frac{1+\delta}{1-\delta} = \frac{\lambda_{max}(X_S)}{\lambda_{min}(X_S)} = \kappa$ to be close to 1
- $X^T X/n$ should be close to identity matrix $\rightarrow$ covariates cannot be strongly correlated
- Random matrices with i.i.d sub-Gaussian entries satisfy this property w.h.p with n almost linearly scaling with k

Restricted Eigenvalue condition

Restricted Eigenvalue Condition

A $p \times p$ sample covariance matrix $X^T X/n$ satisfies the restricted eigenvalue (RE) condition over S with parameters $(\alpha, \gamma) \in [1, \infty) \times (0, \infty)$ if

$$\frac{1}{n}\theta^T X^T X\theta = \frac{1}{n}\|X\theta\|_2^2 \geq \gamma^2\|\theta\|_2^2 \quad \forall\theta \in C(S;\alpha)$$

- Weaker than the RIP condition
- $X^T X/n$ satisfies RE condition of order k if above condition is satisfied for all subsets $S, |S| = k$
- If X satisfies RE condition then $\|\hat{\beta} - \beta^*\|_2 = O(\sqrt{k \log p/n})$
- Does $X \in \mathbb{R}^{n\times p}, X_i \sim N(0, \Sigma)$ satisfy the RE condition for any Σ?

Main Results

- Linear model $y_i = X_i^T \beta + w_i, X_i \sim N(0, \Sigma)$
- Define: $\rho^2(\Sigma) = \max_{j=1,..,p} \Sigma_{jj}$

Theorem 1

For any Gaussian random design $X \in \mathbb{R}^{n \times p}$ with i.i.d. $N(0, \Sigma)$ rows, there are universal positive constants c, c' such that

$$\frac{\|Xv\|_2}{\sqrt{n}} \geq \frac{1}{4}\|\Sigma^{1/2} v\|_2 - 9\rho(\Sigma)\sqrt{\frac{\log p}{n}}\|v\|_1, \text{ for all } v \in \mathbb{R}^p$$

with probability atleast $1 - c' exp(-cn)$

- Insight into eigenstructure of sample covariance matrix $\hat{\Sigma} = X^T X / n$

Main results

Corollary 1 (Restricted eigenvalue property)

Suppose that Σ satisfies the RE condition of order k with parameters (α, γ). Then for universal positive constants c, c', c'', if the sample size satisfies

$$n > c'' \frac{\rho^2(\Sigma)(1+\alpha)^2}{\gamma^2} k \log p$$

then the matrix $\hat{\Sigma} = X^T X / n$ satisfies the RE condition with parameters $(\alpha, \frac{\gamma}{8})$ with probability at least $1 - c' exp(-cn)$.

- Proof: Use $\|v\|_1 = \|v_S\|_1 + \|v_{S^c}\|_1 \leq (1+\alpha)\sqrt{k}\|v\|_2$ and $\|\Sigma^{1/2} v\|_2 \geq \gamma \|v\|_2$ and substitute in Theorem 1, we get $9(1+\alpha)\rho(\Sigma)\sqrt{\frac{k \log p}{n}} \leq \gamma/8$
- The sample size scales as $\Omega(k \log p)$ as long as $\rho(\Sigma)$ is bounded

Proof outline

- The result bounds $\|Xv\|_2$ in terms of $\|\Sigma^{1/2}v\|$ and $\|v\|_1$ for all v w.h.p
- Step 1: Consider set: $V(r) := \{v \in \mathbb{R}^p \mid \|\Sigma^{1/2}v\|_2 = 1, \|v\|_1 \leq r\}$
  - Condition holds trivially when $\Sigma^{1/2}v = 0$
  - For any vector $v \in \mathbb{R}^p$ consider $\tilde{v} = v/\|\Sigma^{1/2}v\|$. Condition is scale invariant. Hence holds for v if it holds for $\tilde{v}$.
- Step 2: Define random variable:

$$M(r, X) := 1 - \inf_{v \in V(r)} \frac{\|Xv\|_2}{\sqrt{n}} = \sup_{v \in V(r)} \left\{1 - \frac{\|Xv\|_2}{\sqrt{n}}\right\}$$

  - Step 2a: Upper bound $\mathbb{E}[M(r, X)]$
  - Step 2b: Establish concertration around the mean
- Step 3: Peeling argument to show that analysis holds with high probability and uniformly for all r

Step 2a: Bounding the expectation

Lemma 1

For any radius $r > 0$ such that $V(r)$ is non-empty, we have

$$\mathbb{E}[M(r, X)] \leq \frac{1}{4} + 3\rho(\Sigma)\sqrt{\frac{\log p}{n}}\, r$$

- Define the Gaussian random variable $Y_{u,v} := u^T X v$
- $-\inf_{v \in V(r)} \|Xv\|_2 = -\inf_{v \in V(r)} \sup_{u \in S^{n-1}} u^T X v = \sup_{v \in V(r)} \inf_{u \in S^{n-1}} u^T X v$
- Upper bound $1 + \mathbb{E}[\sup_{v \in V(r)} \inf_{u \in S^{n-1}} Y_{u,v}]$

Step 2a: Bounding the expectation

Gordon's inequality

Suppose that $\{Y_{u,v}, (u,v) \in U \times V\}$ *and* $\{Z_{u,v}, (u,v) \in U \times V\}$ *are two zero-mean Gaussian processes on* $U \times V$*. Let* $\sigma(.)$ *denote the standard deviation of its argument. Suppose these two processes satisfy the inequality*

$$\sigma(Y_{u,v} - Y_{u',v'}) \leq \sigma(Z_{u,v} - Z_{u',v'}), \text{ for all pairs } (u,v) \text{ and } (u',v') \in U \times V$$

where equality holds when $v = v'$*. Then we are guaranteed that*

$$\mathbb{E}[\sup_{v \in V} \inf_{u \in U} Y_{u,v}] \leq \mathbb{E}[\sup_{v \in V} \inf_{u \in U} Z_{u,v}]$$

- Find a $Z_{u,v}$ such that the above condition is satisfied and computing $\mathbb{E}[\sup_{v \in V} \inf_{u \in U} Z_{u,v}]$ is easy

Step 2a: Bounding the expectation

- X can be expressed as $X = W\Sigma^{1/2}$, where $W \in \mathbb{R}^{n\times p}$ is a matrix with i.i.d. $N(0,1)$ entries. Therefore $Y_{u,v} = u^T W\Sigma^{1/2} v = u^T W \tilde{v}$
- Define $\tilde{v} = \Sigma^{1/2} v$
- Compute $\sigma^2(Y_{u,v} - Y_{u',\tilde{v}'})$

$$\sigma^2(Y_{u,\tilde{v}} - Y_{u',\tilde{v}'}) := \mathbb{E}(\sum_{i=1}^{n}\sum_{j=1}^{p} W_{i,j}(u_i\tilde{v}_j - u'_i\tilde{v}'_j))^2 = |||u\tilde{v}^T - (u')(\tilde{v}')^T|||_F^2$$

- Define $Z_{u,v} = \vec{g}^T u + \vec{h}^T \Sigma^{1/2} v = \vec{g}^T u + \vec{h}^T \tilde{v}$, where $\vec{g} \sim N(0, I_{n\times n})$, $\vec{h} \sim N(0, I_{p\times p})$
- Compute $\sigma^2(Z_{u,v} - Z_{u',v'})$

$$\sigma^2(Z_{u,v} - Z_{u',v'}) = \|u - u'\|_2^2 + \|v - v'\|_2^2$$

- Condition in Gordon's inequality is satisfied

Step 2a: Bounding the expectation

- Applying Gordon's inequality

$$\mathbb{E}[\sup_{v\in V(r)} \inf_{u\in S^{n-1}} u^T X v] \leq \mathbb{E}[\inf_{u\in S^{n-1}} \vec{g}^T u] + \mathbb{E}[\sup_{v\in V(r)} \vec{h}^T \Sigma^{1/2} v]$$

$$= -\mathbb{E}[\|\vec{g}\|_2] + \mathbb{E}[\sup_{v\in V(r)} \vec{h}^T \Sigma^{1/2} v]$$

- By definition of V(r)

$$\sup_{v\in V(r)} |\vec{h}^T \Sigma^{1/2} v| \leq \sup_{v\in V(r)} \|v\|_1 \|\Sigma^{1/2}\vec{h}\|_\infty \leq r\|\Sigma^{1/2}\vec{h}\|_\infty$$

- Each element $(\Sigma^{1/2}\vec{h})_j$ is zero-mean Gaussian with variance Σ_{jj}. According to known results on Gaussian maxima

$$\mathbb{E}[\|\Sigma^{1/2}\vec{h}\|_\infty] \leq 3\sqrt{\rho^2(\Sigma)\log p}, \text{ where } \rho^2(\Sigma) = max_j \Sigma_{jj}$$

- $\mathbb{E}[\|\vec{g}\|_2] \geq \frac{3}{4}\sqrt{n}$ for all $n \geq 10$ by standard χ^2 tail bounds
- Putting together the pieces gives us the required result

Step 2b: Concentration around the mean

Lemma 2

For any r such that V(r) is non-empty, we have

$$\mathbb{P}\left[M(r,X) \geq \frac{3t(r)}{2}\right] \leq 2exp(-nt^2(r)/8)$$

where

$$t(r) := \frac{1}{4} + 3r\rho(\Sigma)\sqrt{\frac{\log p}{n}}$$

- Following from previous result suffices to show that

$$\mathbb{P}[|M(r,X) - \mathbb{E}[M(r,X)]| \geq t(r)/2] \leq 2exp(-nt^2(r)/8)$$

Step 2b: Concentration around the mean

- A function $F : \mathbb{R}^m \to \mathbb{R}$ is Lipschitz with constant L if $|F(x) - F(y)| \le L\|x - y\|_2 \ \forall x, y \in \mathbb{R}^m$

Theorem

Let $w \sim N(0, I_{m \times m})$ be an m-dimensional Gaussian random variable. Then for any L-Lipschitz function F, we have

$$\mathbb{P}\left[|F(w) - \mathbb{E}[F(w)]| \ge t\right] \le 2exp(-\frac{t^2}{2L^2}), \ \forall t \ge 0$$

- The tail bound above will follow if we show the Lipschitz constant L is less than $\frac{1}{\sqrt{n}}$

Step 2b: Concentration around the mean

- Define $h(W) = \sup_{v \in V(r)} (1 - \|W\Sigma^{1/2}v\|_2/\sqrt{n})$
- Proof:

$$
\begin{aligned}
\sqrt{n}[h(W) - h(W')] &= \sup_{v \in V(r)} -\|W\Sigma^{1/2}v\|_2 - \sup_{v \in V(r)} \|W'\Sigma^{1/2}v\|_2 \\
&= -\|W\Sigma^{1/2}\hat{v}\|_2 - \sup_{v \in V(r)} (-\|W'\Sigma^{1/2}v\|) \\
&\leq \|W'\Sigma^{1/2}\hat{v}\|_2 - \|W\Sigma^{1/2}\hat{v}\|_2 \\
&\leq \sup_{v \in V(r)} (\|(W - W')\Sigma^{1/2}v\|_2) \\
&\leq \| \sup_{v \in V(r)} (\|\Sigma^{1/2}v\|_2)\} |||(W - W'|||_2 \\
&\leq \| \sup_{v \in V(r)} (\|\Sigma^{1/2}v\|_2)\} |||(W - W'|||_F \\
&= |||W - W'|||_F
\end{aligned}
$$

Step 3: Peeling argument

- $V(r)$ defined such that $\|v\|_1 \leq r$. Need to prove Theorem 1 for all r
- Argument at a high level is as follows
  - Theorem holds for all v in set $V(r)$
  - Consider the event

    $$T := \{\exists v \in \mathbb{R}^p \text{ s.t. } \|\Sigma^{1/2} v\| = 1 \text{ and } (1 - \|Xv\|_2/\sqrt{n}) \geq 3t(\|v\|_1)/2\}$$

  - Bound $\mathbb{P}(T)$ by a union bound over all suitably defined subsets $V(r)$
- Peeling argument yields the bound $\mathbb{P}[T^c] \geq 1 - cexp(-c'n)$ for some constants c, c'

Step 3: Peeling argument

- Define: An objective function $f(v; X)$, $v \in \mathbb{R}^p$, X is a random vector h is any function $h : \mathbb{R}^p \to \mathbb{R}$

Lemma 3

Suppose that $g(r) \geq \mu$ for all $r \geq 0$, and that there exists some constant $c > 0$ such that for all $r > 0$, we have the tail bound

$$\mathbb{P}[\sup_{v \in A, h(v) \leq r} f(v; X) \geq g(r)] \leq 2exp(-ca_n g^2(r))$$

for $a_n > 0$. Define event $E := \{\exists v \in A \text{ such that } f(v; X) \geq 2g(h(v))\}$ Then $\mathbb{P}[E] \leq \frac{2exp(-4ca_n\mu^2)}{1-exp(-4ca_n\mu^2)}$

- In this case: $f(v, X) = 1 - \|Xv\|_2/\sqrt{n}$, $h(v) = \|v\|_1$, $g(r) = 3t(r)/2$, $a_n = n$, $A = \{v \in \mathbb{R}^p \; | \|\Sigma^{1/2}v\|_2 = 1\}$, and $\mu = 3/8$

Applications: Toeplitz matrices

- Toeplitz matrix structure

$$\begin{pmatrix} a & b & c & d & e \\ f & a & b & c & d \\ g & f & a & b & c \\ h & g & f & a & b \\ i & h & g & f & a \end{pmatrix}$$

- Consider Σ has Toeplitz structure with $\Sigma_{ij} = a^{|i-j|}$ for some $a \in [0, 1)$. Common in autoregressive processes
- Minimum eigenvalue $\lambda_{min}(\Sigma) = 1 - a > 0$, independent of p
- Condition number $\kappa = \lambda_{max}(\Sigma_{SS})/\lambda_{min}(\Sigma_{SS})$ grows as parameter a increases towards 1
- RE property satisfied with high probability but RIP violated once $a < 1$ is sufficiently large

Applications: Spiked identity model

- Spiked identity model

  $\Sigma := (1-a)I_{p\times p} + a\vec{1}\vec{1}^T$, $a \in [0,1)$ and $\vec{1} \in \mathbb{R}^p$ is the vector of all ones

- Minimum eigenvalue: $\lambda_{min}(\Sigma) = 1 - a$, $\rho^2(\Sigma) = 1$
- According to Corollary 1: Sample covariance matrix $\hat{\Sigma} = X^T X/n$ will satisfy RE property with high probability when $n = \Omega(k \log p)$
- For any $|S| = k$ consider Σ_{SS}

$$\frac{\lambda_{max}(\Sigma_{SS})}{\lambda_{min}(\Sigma_{SS})} = \frac{1 + a(k-1)}{1-a}$$

- Condition number diverges as k increases

Highly degenerate covariance matrices

- Σ is not full rank
- Generate a degenerate covariance matrix
  - Sample n times from a $N(0, \Sigma)$ distribution
  - Sample covariance matrix $\hat{\Sigma} = X^T X/n, n < p$
  - Therefore $\hat{\Sigma}$ is rank degenerate
  - According to Corollary 1 $\hat{\Sigma}$ satisfies RE property of order k with high probability
  - Now sample n times from $N \sim (0, \hat{\Sigma})$.
- According to Corollary 1 resampled empirical covariance will also have RE property
- Example relevant for a bootstrap-type calculation for assessing errors of the Lasso

Conclusions

- One of the first papers to consider correlated Gaussian matrices
- Result uses Gordon's inequality applicable to only Gaussian design matrices

Thank you