Sparse change detection in high-dimensional linear regression

Tengyao Wang

London School of Economics and Political Science

Statistics Seminar, University of Geneva

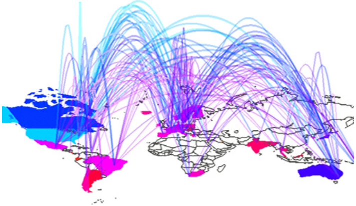
18 Nov 2022

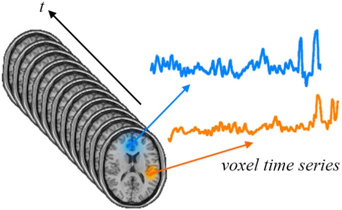
Collaborator

Fengnan Gao
Fudan University

High-dimensional changepoint models

- The evolution of technology enables the collection of vast amounts of time-ordered data:
  - Healthcare devices
  - Covid case numbers
  - Network traffic data
  - Trading data of financial instruments





- Changes in the dynamics of the data streams are frequently of interest, leading to a renaissance of research on changepoint analysis.

Changepoint in regression coefficients

- When data consist of covariate-response pairs, we are often interested in changes in the regression function.
- Observations $(X_t, Y_t) \in \mathbb{R}^p \times \mathbb{R}$ for $t = 1, \ldots, n$ generated from

$$Y_t = X_t^\top \beta_t + \epsilon_t,$$

  where $\epsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$.
- Coefficients $\beta_1, \ldots, \beta_n$ piecewise constant with changepoints at $z_1, \ldots, z_\nu$

$$\beta_t = \beta^{(r)} \quad \text{for } z_{r-1} < t \leq z_r, 1 \leq r \leq \nu + 1.$$

  (Convention: $z_0 = 0$, $z_{\nu+1} = n$)
- **Goal**: estimate the changepoint locations $z_1, \ldots, z_\nu$.

Classically …

- When $p \ll n$, least squares estimators work well (Bai, 1997; Bai and Perron 1998, Julious, 2001)
- For a fixed ν, find the optimal partition of $\{1, \ldots, n\}$ into $\nu + 1$ segments such that the sum of RSS of least squares fit within each segment is minimised:

$$(\hat{z}_1, \ldots, \hat{z}_\nu) = \operatorname*{argmin}_{\tilde{z}_1 < \tilde{z}_2 < \cdots < \tilde{z}_\nu} \sum_{r=1}^{\nu+1} \min_{\tilde{\beta}} \sum_{t=\tilde{z}_{r-1}+1}^{\tilde{z}_r} (Y_t - X_t^\top \tilde{\beta})^2.$$

- If ν is unknown, compare goodness-of-fit from different choices of ν, e.g. using BIC.

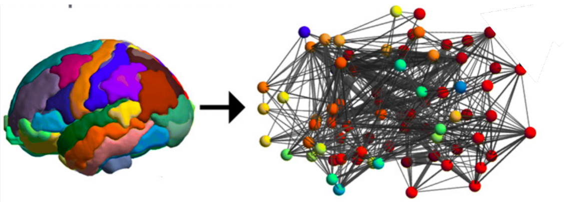
Challenges in high dimensions

- When $p \asymp n$, the above least squares approach no longer works.
- Several approaches were proposed to analyse changepoints in high-dimensional regression problems (Lee et al., 2016; Kaul et al., 2019; Rinaldo et al., 2021; Wang et al., 2021).
  - These works impose the additional assumption that all regression coefficients $\beta^{(1)}, \ldots, \beta^{(\nu+1)}$ are sparse.
  - This allows reasonable estimation of $\beta^{(r)}$, $1 \leq r \leq \nu + 1$ given a candidate set of changepoints
  - Choose the best candidate set using goodness-of-fit statistics
- In contrast, we will only assume that the **changes are sparse**:

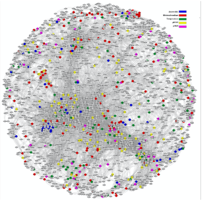
$$\|\beta^{(r+1)} - \beta^{(r)}\|_0 \leq k.$$

Real data examples

- Differential networks: find changepoints in the dynamics of Gaussian graphical models over time.
  - Brain connectivity network
  - Gene-gene interaction network
  - Financial network model between countries
- Central players in the network may have dense connection to other nodes, but their changes may still be sparse.



Bansal et al. (*Sci. Adv.* 2019)



Chen et al. (*PLOS ONE*, 2015)

Single changepoint setup

- We focus first on the single changepoint problem, i.e. $\nu = 1$, we write $z = z_1$.
- Observations $(X_t, Y_t) \in \mathbb{R}^p \times \mathbb{R}$ for $t = 1, \ldots, n$ generated from

$$Y_t = X_t^\top(\beta^{(1)}\mathbb{1}_{\{t\le z\}} + \beta^{(2)}\mathbb{1}_{\{t>z\}}) + \epsilon_t,$$

  where $\epsilon_t \overset{\text{iid}}{\sim} N(0, \sigma^2)$.
- Assume $\|\beta^{(2)} - \beta^{(1)}\|_0 \le k$ and $p < n$.

Dense nuisance parameters

- This problem is an example of high-dimensional inference in the presence of dense nuisance parameters.
- True parameter of interest is $\beta^{(2)} - \beta^{(1)}$, which is sparse. The dense nuisance parameter $\beta^{(1)} + \beta^{(2)}$ interferes with the inference.
- Relation to the literature
  - The Neyman–Scott paradox (Neyman and Scott, 1948)
  - High-dimensional change-point problems (e.g. Cho and Fryzlewicz, 2015; Jirak, 2015; W. and Samworth, 2018; Enikeeva and Harchaoui, 2019)
  - Matched-pair survival analysis (Battey and Cox, 2020)

Our method: complementary sketching

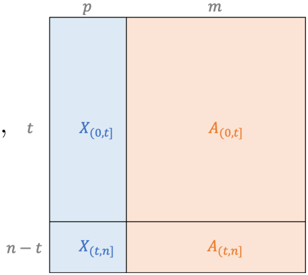
A complimentary sketching

Complementary sketching

- Define $m := n - p$, $X := (X_1^\top, \ldots, X_n^\top)^\top$ and write $X_{(s,e]}$ for the submatrix of X using rows $s+1, \ldots, e$.
- **Procedure:** Given data $X \in \mathbb{R}^{n \times p}$ and $Y \in \mathbb{R}^n$,
  1. Construct $A \in \mathbb{R}^{n \times m}$ such that A has orthonormal columns orthogonal to the column space of X.
  2. For each $t \in \{1, \ldots, n-1\}$, compute

$$W_t := \begin{pmatrix} A_{(0,t]}^\top & -A_{(t,n]}^\top \end{pmatrix} \begin{pmatrix} X_{(0,t]} \\ X_{(t,n]} \end{pmatrix} \in \mathbb{R}^{m \times p},$$

$$Z := \begin{pmatrix} A_{(0,t]}^\top & A_{(t,n]}^\top \end{pmatrix} \begin{pmatrix} Y_{(0,t]} \\ Y_{(t,n]} \end{pmatrix} \in \mathbb{R}^m.$$



- Similar to orthogonal sketching, but sketches the covariate matrix and the response vector in opposite ways in the second block.

Elimination of nuisance parameters

- Why does complementary sketching work?
- Write $\theta := (\beta^{(1)} - \beta^{(2)})/2$ and $\zeta := (\beta^{(1)} + \beta^{(2)})/2$.

$$
\begin{aligned}
Z &= A_{(0,z]}^{\top} Y_{(0,z]} + A_{(z,n]}^{\top} Y_{(z,n]} \\
&= A_{(0,z]}^{\top} (X_{(0,z]} \beta^{(1)} + \epsilon_{(0,z]}) + A_{(z,n]}^{\top} (X_{(z,n]} \beta^{(2)} + \epsilon_{(z,n]}) \\
&= A_{(0,z]}^{\top} X_{(0,z]} \theta + \cancel{A_{(0,z]}^{\top} X_{(0,z]} \zeta} - A_{(z,n]}^{\top} X_{(z,n]} \theta + \cancel{A_{(z,n]}^{\top} X_{(z,n]} \zeta} \\
&\qquad + A_{(0,z]} \epsilon_{(0,z]} + A_{(z,n]} \epsilon_{(z,n]} \\
&= W_z \theta + \xi,
\end{aligned}
$$

- We have eliminated the contribution of the nuisance parameter ζ in Z.
- This idea of complementary sketching was first used in a two-sample testing problem (Gao and W. 2022).
- The changepoint problem is reduced to finding t such that W_t forms a 'best sparse linear approximation' to Z.

The charcoal algorithms

- Several different approaches are possible once we have eliminated the nuisrance parameter, which we collectively call the **charcoal** (changepoint in regression via a complementary-sketching algorithm) methdology.
- **charcoal$_{\mathbf{corr}}$**: $Q_t := \{\mathrm{diag}(W_t^\top W_t)\}^{-1/2} W_t^\top Z$,

$$\hat{z}^{\mathrm{corr}} := \operatorname*{argmax}_t \|\mathbf{soft}(Q_t, \lambda)\|_2^2.$$

Algorithm 1: Pseudocode for change-point estimation

Input: $X \in \mathbb{R}^{n\times p}, Y \in \mathbb{R}^n$ satisfying $n > p$, $\lambda \geq 0$, $\alpha > 0$

Set $m \leftarrow n - p$;

Form $A \in \mathbb{O}^{n\times m}$ with columns orthogonal to the column space of X;

Compute $Z \leftarrow A^\top Y$;

Set $W_0 = \mathbf{0}_{m\times p}$;

for $1 \leq t \leq n-1$ **do**

  Compute $W_t \leftarrow W_{t-1} + 2a_t x_t^\top$;

  Compute $Q_t = \{\mathrm{diag}(W_t^\top W_t)\}^{-1/2} W_t^\top Z$;

  Compute $H_t \leftarrow \|\mathrm{soft}(Q_t, \lambda)\|_2$;

Output: $\hat{z} := \arg\max_{\alpha n < t < (1-\alpha)n} H_t$.

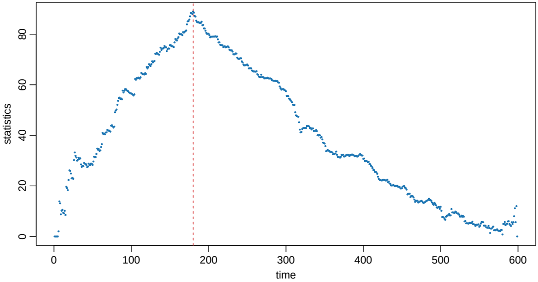
The charcoal algorithms

- Several different approaches are possible once we have eliminated the nuisrance parameter, which we collectively call the **charcoal** (changepoint in regression via a complementary-sketching algorithm) methdology.
- **charcoal$_{\text{corr}}$**: $Q_t := \{\mathrm{diag}(W_t^\top W_t)\}^{-1/2} W_t^\top Z$,

$$\hat{z}^{\mathrm{corr}} := \operatorname*{argmax}_t \|\mathbf{soft}(Q_t, \lambda)\|_2^2.$$

- Several different approaches are possible once we have eliminated the nuisrance parameter, which we collectively call the **charcoal** (changepoint in regression via a complementary-sketching algorithm) methdology.
- **charcoal**$_{\mathbf{corr}}$: $Q_t := \{\mathrm{diag}(W_t^\top W_t)\}^{-1/2} W_t^\top Z$,

$$\hat{z}^{\mathrm{corr}} := \operatorname*{argmax}_t \|\mathbf{soft}(Q_t, \lambda)\|_2^2.$$



The charcoal algorithms

- Several different approaches are possible once we have eliminated the nuisrance parameter, which we collectively call the **charcoal** (changepoint in regression via a complementary-sketching algorithm) methdology.
- **charcoal$_{\textbf{corr}}$**: $Q_t := \{\mathrm{diag}(W_t^\top W_t)\}^{-1/2} W_t^\top Z$,
$$\hat{z}^{\mathrm{corr}} := \operatorname*{argmax}_t \|\mathbf{soft}(Q_t, \lambda)\|_2^2.$$
- **charcoal$_{\textbf{proj}}$**: let $\hat{v}$ be the leading left singular vector of $\mathbf{soft}(Q, \lambda)$, estimate
$$\hat{z}^{\mathrm{proj}} := \operatorname*{argmax}_t (\hat{v}^\top Q_t).$$
- **charcoal$_{\textbf{lasso}}$**: simply run Lasso on (W_t, Z) to find the best fit
$$\hat{\theta}_t(\lambda_t) := \operatorname*{argmin}_\theta \biggl\{ \frac{1}{2m} \|Z - W_t\theta\|_2^2 + \lambda_t \|\theta\|_1 \biggr\}$$
$$\hat{z}^{\mathrm{lasso}} := \operatorname*{argmin}_t \|Z - W_t \hat{\theta}_t(\lambda_t)\|_2^2,$$

Multiple changepoints

- The **charcoal** algorithms can be combined with any of the top-down methods to recursively identify multilple changepoints.
- We use the narrowest-over-threshold method (Baranowski et al., 2019)

Algorithm 4: Pseudocode for multiple changepoint estimation

Input: $X \in \mathbb{R}^{n\times p}, Y \in \mathbb{R}^n$ satisfying $n - p > 0$, a soft threshold level $\lambda \geq 0$, burn-in parameter $\alpha \geq 0$, number of intervals M, testing threshold $T > 0$

Set $\hat{Z} \leftarrow \emptyset$ and generate M pairs of integers $(s_1, e_1), \ldots, (s_M, e_M)$ uniformly from $\{(a, b] : a, b \in \mathbb{N} \cup \{0\}, b - a > p\}$.

Run $\mathtt{NOT}(0, n)$ where $\mathtt{NOT}$ is defined below.

Let $\hat{\nu} \leftarrow |\hat{Z}|$ and sort elements of $\hat{Z}$ in increasing order to yield $\hat{z}_1 < \cdots < \hat{z}_{\hat{\nu}}$.

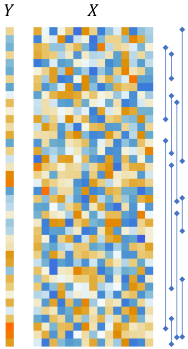
Output: $\hat{z}_1, \ldots, \hat{z}_{\hat{\nu}}$

Function $\mathtt{NOT}(s,\, e)$
- Set $\mathcal{M}_{s,e} \leftarrow \{m : (s_m, e_m] \subseteq (s, e]\}$
- **for** $m \in \mathcal{M}_{s,e}$ **do**
  - Run Algorithm 2 with input $X_{(s_m,e_m]}, Y_{(s_m,e_m]}, \lambda$ and α, and let $\hat{z}^{(m)}$ and $H^{(m)}_{\max}$ be the output.
- **end**
- $\mathcal{M}^*_{s,e} \leftarrow \{m \in \mathcal{M}_{s,e} : H^{(m)}_{\max} > T\}$
- **if** $\mathcal{M}^*_{s,e} \neq \emptyset$ **then**
  - $m_0 \leftarrow \arg\min_{m \in \mathcal{M}} (e_m - s_m)$
  - $b \leftarrow \hat{s}_{m_0} + \hat{z}^{(m_0)}$
  - $\hat{Z} \leftarrow \hat{Z} \cup \{b\}$
  - $\mathtt{NOT}(s,\, b)$
  - $\mathtt{NOT}(b,\, e)$
- **end**

end

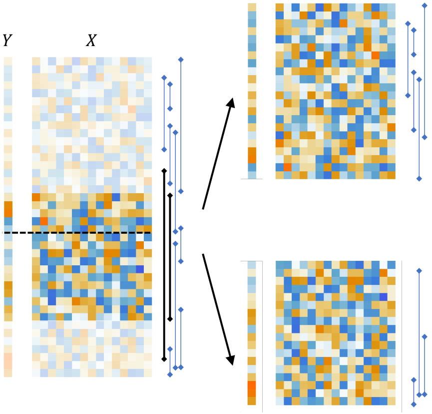
Multiple changepoints

- The **charcoal** algorithms can be combined with any of the top-down methods to recursively identify multilple changepoints.
- We use the narrowest-over-threshold method (Baranowski et al., 2019)



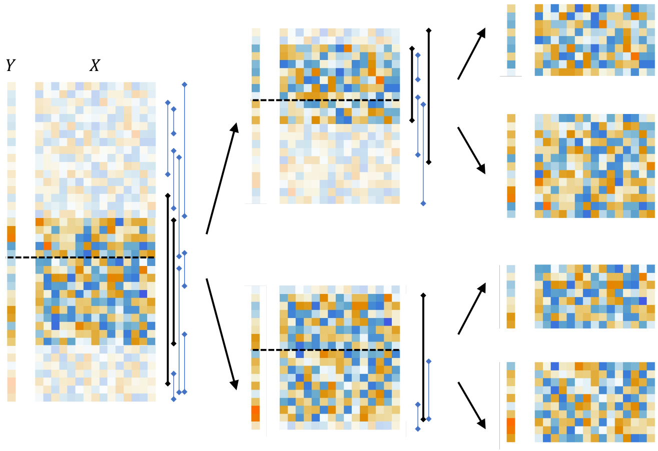
Multiple changepoints

- The **charcoal** algorithms can be combined with any of the top-down methods to recursively identify multilple changepoints.
- We use the narrowest-over-threshold method (Baranowski et al., 2019)



Multiple changepoints

- The **charcoal** algorithms can be combined with any of the top-down methods to recursively identify multilple changepoints.
- We use the narrowest-over-threshold method (Baranowski et al., 2019)



Theoretical results

- Test statistics are formed from

$$Q_t = \{\mathrm{diag}(W_t^\top W_t)\}^{-1/2}\bigl(W_t^\top W_z\theta + W_t^\top \xi\bigr)$$

- **Key step**: show that $W_t^\top W_z$ is close to $4t(n-z)(n-p)n^{-2}I_p$ in k-operator norm uniformly over t.
- Difficult to control $\{\mathrm{diag}(W_t^\top W_t)\}^{-1/2}$ uniformly over t. For theoretical analysis, we look at a slight variant where

$$Q_t = \sqrt{\frac{n}{t(n-t)}}W_t^\top Z = \sqrt{\frac{n}{t(n-t)}}\bigl(W_t^\top W_z\theta + W_t^\top \xi\bigr).$$

Theoretical analysis of the variant

- Test statistics are formed from

$$Q_t = \sqrt{\frac{n}{t(n-t)}}\left(W_t^\top W_z \theta + W_t^\top \xi\right)$$

- **Key step**: show that $W_t^\top W_z$ is close to $4t(n-z)(n-p)n^{-2} I_p$ in k-operator norm uniformly over t for $t \le z$.

Theoretical analysis of the variant

- Test statistics are formed from

$$Q_t = \sqrt{\frac{n}{t(n-t)}}\left(W_t^\top W_z \theta + W_t^\top \xi\right)$$

- **Key step**: show that $W_t^\top W_z$ is close to $4t(n-z)(n-p)n^{-2} I_p$ in k-operator norm uniformly over t for $t \leq z$.
- Hence $H_t := \|\mathbf{soft}(Q_t, \lambda)\|_2$ is close to $\tilde{H}_t := \sqrt{\frac{n}{t(n-t)}}\|(W_t^\top W_z \theta)_S\|_2$

- Test statistics are formed from

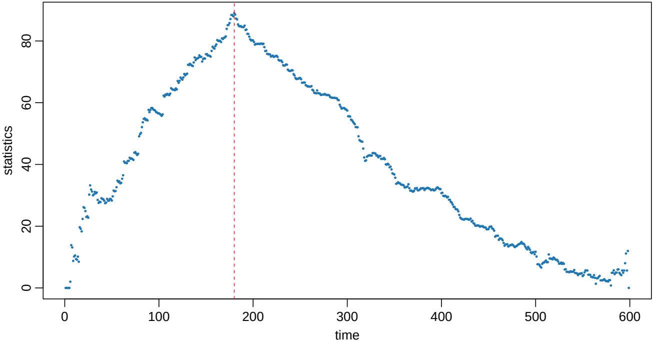
$$Q_t = \sqrt{\frac{n}{t(n-t)}}\big(W_t^\top W_z\theta + W_t^\top \xi\big)$$

- **Key step**: show that $W_t^\top W_z$ is close to $4t(n-z)(n-p)n^{-2}I_p$ in k-operator norm uniformly over t for $t \le z$.
- Hence $H_t := \|\mathbf{soft}(Q_t, \lambda)\|_2$ is close to $\tilde{H}_t := \sqrt{\frac{n}{t(n-t)}}\|(W_t^\top W_z\theta)_S\|_2$
- This is in turn approximately

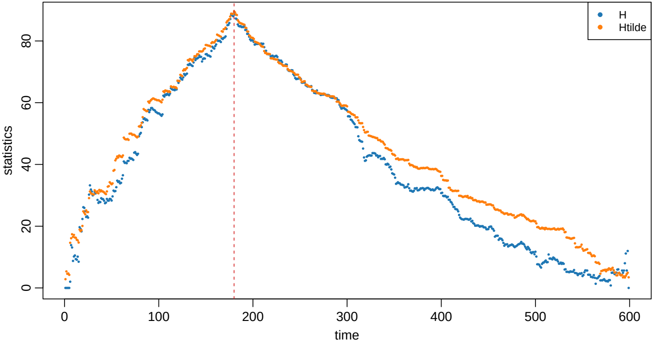
$$h_t := \frac{4(n-p)\|\theta\|_2}{n}\left\{\sqrt{\frac{t}{n(n-t)}}(n-z)\mathbb{1}_{\{t\le z\}} + \sqrt{\frac{n-t}{nt}}z\mathbb{1}_{\{t>z\}}\right\}.$$

Theoretical analysis of the variant

- Graphical illustration of the proof sketch:

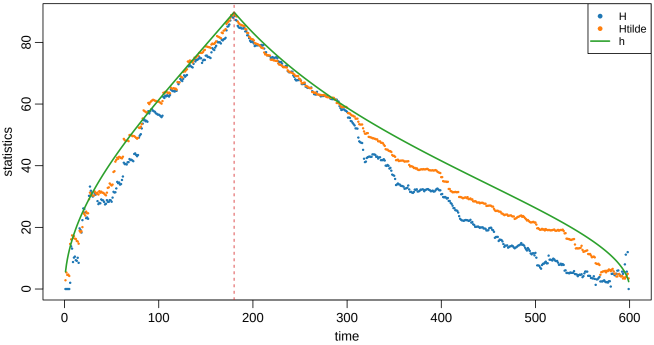


- Graphical illustration of the proof sketch:



Theoretical analysis of the variant

- Graphical illustration of the proof sketch:



- To prove estimation accuracy:
  1. Understand the sharpness of peak of $(h_t : 1 \leq t \leq n-1)$
     — this turns out to be the same as the univariate CUSUM curve
  2. Control $|H_t - \tilde{H}_t|$ and $|\tilde{H}_t - h_t|$ uniformly over t.

Theoretical guarantees

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Assumptions

(A1) Random design: $x_t \sim N_p(0, I_p)$ independently for $t = 1, \ldots, n$

(A2) Asymptotic regime: n, z, p satisfies $p < n$ and $z/n \to \tau \in (0,1)$ and $(n-p)/n \to \eta \in (0,1)$ as $n \to \infty$.

Theorem. Assume Conditions (A1) and (A2). Suppose that $\|\theta\|_2 \leq 1$, $k \leq p/2$. There exists $c, C > 0$, depending only on τ, η, such that if $\lambda > c\sigma \log p$, then asymptotically with probability 1, for all but finitely many n's, we have

$$\frac{|\hat{z}^{\mathrm{corr}} - z|}{n} \leq \frac{C\lambda\sqrt{k}}{\sqrt{n}\|\theta\|_2}.$$

Theoretical guarantees

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Theorem. Under the same condition as above, There exists $c, C > 0$, depending only on τ, η, such that if $\lambda > c\sigma \log p$, then asymptotically with probability 1, for all but finitely many n's, we have

$$\sin \angle(\hat{v}^{\mathrm{proj}}, \theta) \leq \frac{C\lambda\sqrt{k}}{\sqrt{n}\|\theta\|_2}.$$

Hence, $\hat{z}^{\mathrm{proj}}$ satisfies

$$\frac{|\hat{z}^{\mathrm{proj}} - z|}{n} \leq \frac{C\lambda^2\sqrt{k}\log p}{\sqrt{n}\|\theta\|_2^2}.$$

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

Theorem. Under the same condition as above, There exists $c, C > 0$, depending only on τ, η, such that if $\lambda > c\sigma \log p$, then asymptotically with probability 1, for all but finitely many n's, we have

$$\sin \angle(\hat{v}^{\mathrm{proj}}, \theta) \leq \frac{C\lambda\sqrt{k}}{\sqrt{n}\|\theta\|_2}.$$

Hence, a sample-splitting variant of $\hat{z}^{\mathrm{proj}}$ satisfies

$$\frac{|\hat{z}^{\mathrm{proj}} - z|}{n} \leq \frac{C\lambda\sqrt{k}\log p}{\sqrt{n}\|\theta\|_2}.$$

Optimality of the estimator

- Consistent estimation is possible when $\|\theta\|_2/\sigma \gg \sqrt{\frac{k\log^2 p}{n}}$.
- This is essentially the SNR required to **test for a change** even if the location of changepoint z is known. Let $P^X_{z,\beta^{(1)},\beta^{(2)}}$ be the distribution of Y conditional on X, changepoint z and parameters $\beta^{(1)}$ and $\beta^{(2)}$. We test

$$H_0 : \theta = 0 \quad \text{vs} \quad H_1 : \theta \in \Theta_{p,k}(\rho) := \{\theta : \|\theta\|_2/\sigma \geq \rho, \|\theta\|_0 \leq k\}$$

- Define the **minimax risk** of testing

$$\mathcal{M}_X(k,\rho) := \inf_{\psi}\biggl\{\sup_{\beta\in\mathbb{R}^p} P^X_{z,\beta,\beta}(\psi \neq 0) + \sup_{\substack{\beta_1,\beta_2\in\mathbb{R}^p \\ (\beta_1-\beta_2)/2\in\Theta_{p,k}(\rho)}} P^X_{z,\beta_1,\beta_2}(\psi \neq 1)\biggr\},$$

Theorem. Assume (A1), (A2), and $k \leq p^\alpha$ for some $\alpha < 1/2$. There exists a universal constant $c > 0$ such that if $\rho \leq \sqrt{\frac{c(1-2\alpha)k\log p}{n}}$, then

$$\mathcal{M}_X(k,\rho) \xrightarrow{\text{a.s.}} 1.$$

Numerical studies

Comparison of variants

- Gaussian Orthogonal Ensemble design matrices with a single changepoint at $z = 0.3n$
- $\theta^{(1)}$ sampled as a Gaussian vector, $\theta^{(2)} - \theta^{(1)}$ randomly generated k-sparse vector with ℓ_2 norm ρ.
- $\textbf{charcoal}_{\textbf{corr}}$ and $\textbf{charcoal}'_{\textbf{corr}}$ uses a burn-in parameter of 0.1.

n	p	k	ρ	corr	corr′	proj	proj′	lasso
600	200	3	1	7.16	8.67	7.17	11.05	12.95
			2	2.04	3.22	1.95	2.81	3.04
			4	0.93	2.35	1.24	2.16	1.47
		14	1	16.75	18.14	19.69	34.44	82.36
			2	3.22	3.76	3.19	4.03	6.94
			4	1.62	2.29	2.20	2.65	2.00
1200	400	3	1	6.61	7.13	6.20	7.63	12.14
			2	1.64	1.86	1.96	2.40	3.39
			4	1.11	2.06	0.94	2.06	1.43
		20	1	16.70	19.51	11.01	14.94	101.81
			2	2.90	2.98	3.92	4.11	10.12
			4	1.86	2.50	1.64	1.91	3.20

Table: $\mathbb{E}|\hat{z} - z|$ estimated over 100 Monte Carlo repetitions.

Comparisons with other methods

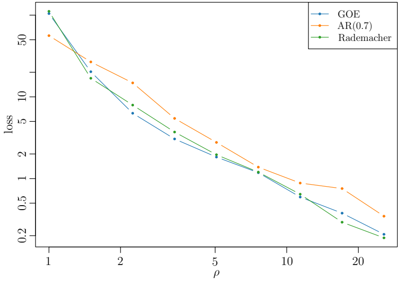
- Existing methods in literature require sparsity of $\theta^{(r)}$ for all r.
- We compare with
  - The VPBS algorithm of Rinaldo et al., 2021
  - A two-sided Lasso-based approach of Lee et al. (2016) (LSS) and Leonardi and Bühlmann (2016) (LB)
  - a two-stage refinement approach of Kaul et al. (2019) (KJF)
- We compare the performance of various methods in a single changepoint estimation task with $n = 1200$, $z = 360$.

Comparisons with other methods

p	k	ρ	$\text{charcoal}_{\text{proj}}$	$\text{charcoal}_{\text{lasso}}$	VPBS	LB	KJF	LSS
400	3	1	**7.2**	13.2	452.4	556.1	238.8	472.2
		2	**2.2**	3.5	476.3	569.2	239.3	364.1
		4	**1.1**	1.5	434.2	532.8	239.1	272.1
		8	**0.7**	0.8	326.3	496.8	239.1	310.8
	20	1	**12.4**	85.4	422.7	528.8	238.9	479.5
		2	**3.0**	9.2	494.9	546.8	238.9	284.5
		4	**2.0**	2.6	431.9	553.1	239.1	268.5
		8	1.9	**0.8**	356.2	513.3	239.3	261.5
	400	1	**162.2**	344.2	477.8	569.8	238.8	429.9
		2	**46.3**	338.4	504.0	583.2	238.8	252.4
		4	25.3	**13.3**	446.3	554.1	238.9	285.6
		8	20.7	**3.0**	355.6	487.6	239.1	250.1
1000	3	1	**60.7**	113.3	241.6	429.5	237.2	227.3
		2	**8.3**	11.8	243.4	441.4	239.0	228.2
		4	**2.9**	4.0	239.5	366.9	243.9	230.6
		8	2.4	**1.4**	235.1	245.1	262.2	230.7
	31	1	300.3	364.9	233.4	440.1	238.8	**227.4**
		2	**71.7**	140.9	242.5	469.5	238.9	228.3
		4	16.0	**12.5**	251.3	358.4	238.9	224.5
		8	13.7	**4.6**	244.5	249.0	238.2	230.1
	1000	1	275.5	359.8	232.6	483.0	239.3	**231.8**
		2	256.9	320.8	238.4	447.4	238.9	**229.2**
		4	224.1	**91.0**	242.7	378.2	239.1	228.0
		8	194.5	**39.6**	246.4	253.5	242.4	226.7

Model misspecification

- We focused on GOE design and Gaussian noise to facilitate theoretical analysis
- Our methodology can be applied in more general settings
- We vary design to have i) $N_p(0, \Sigma)$ rows with $\Sigma = (0.7^{|i-j|})_{1 \le i,j \le p}$, or ii) Rademacher entries
- We vary noise distribution to t_4, t_6, centred $\mathrm{Exp}(1)$ or Rademacher distributions.



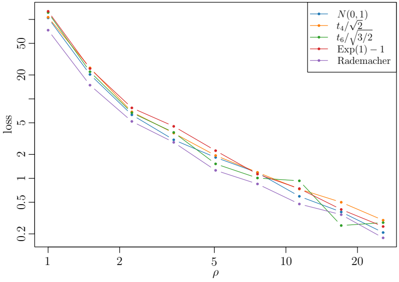


Figure: Robustness to varying design matrices and noise distributions.

- We use **charcoal** in conjunction with **NOT** (Baranowski et al. (2019) for multiple changepoint estimation.
- We consider two simulation settings

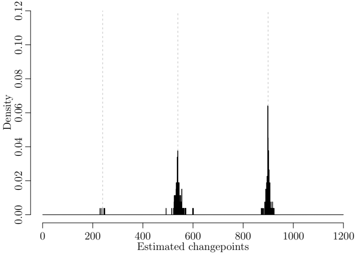
(M1) $n = 1200, p = 200, \nu = 3,$
$(z_1, z_2, z_3)/n = (0.2, 0.55, 0.75),$
$(\|\theta^{(1)}\|_2, |\theta^{(2)}\|_2, |\theta^{(3)}\|_2) = \rho_{\min} \times (1, 1.5, 2),$
$\|\theta^{(1)}\|_0 = \|\theta^{(2)}\|_0 = \|\theta^{(3)}\|_0 = k.$

(M2) $n = 2400, p = 400, \nu = 4,$
$(z_1, z_2, z_3, z_4)/n = (0.3, 0.55, 0.75, 0.9),$
$(\|\theta^{(1)}\|_2, |\theta^{(2)}\|_2, |\theta^{(3)}\|_2, |\theta^{(4)}\|_2) = \rho_{\min} \times (1, 1.15, 1.45, 2.18),$
$\|\theta^{(1)}\|_0 = \|\theta^{(2)}\|_0 = \|\theta^{(3)}\|_0 = \|\theta^{(4)}\|_0 = k.$

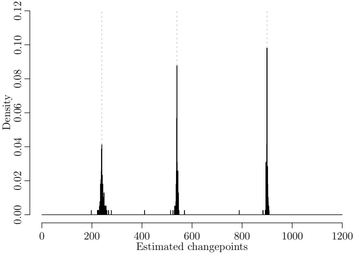
Multiple changepoint

n	p	k	$\rho_{\min}$	$\hat{\nu} - \nu$ value					Haus	ARI
				-3	-2	-1	0	1		
1200	200	3	0.8	0	0	96	4	0	292.8	0.742
			1.2	0	0	22	78	0	75.4	0.918
			1.6	0	0	0	98	2	8.8	0.978
		10	0.8	0	2	97	1	0	304.9	0.71
			1.2	0	0	42	55	3	141.1	0.856
			1.6	0	0	1	96	3	18	0.96
		100	0.8	3	67	30	0	0	591.7	0.303
			1.2	0	4	88	8	0	319.3	0.611
			1.6	0	0	52	46	2	217.1	0.759
2400	400	3	0.8	0	0	25	75	0	155.3	0.881
			1.2	0	0	0	100	0	14.3	0.975
			1.6	0	0	0	100	0	10.1	0.983
		10	0.8	0	15	53	32	0	376.9	0.72
			1.2	0	0	2	98	0	37.3	0.945
			1.6	0	0	1	99	0	21	0.97
		100	0.8	42	57	1	0	0	1154.9	0.184
			1.2	0	32	54	14	0	647	0.457
			1.6	0	0	14	84	2	376.9	0.658

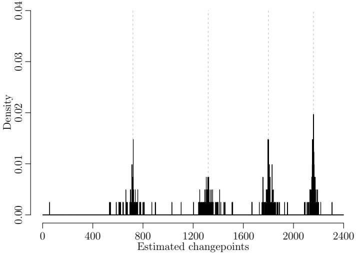
Multiple changepoint



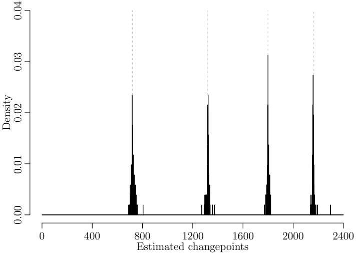
$n = 1200, p = 200, k = 3, \rho_{\min} = 0.8$



$n = 1200, p = 200, k = 3, \rho_{\min} = 1.6$



$n = 2400, p = 400, k = 10, \rho_{\min} = 0.8$



$n = 2400, p = 400, k = 10, \rho_{\min} = 1.6$

Figure: Histogram of estimated changepoint locations in four settings.

Summary

- It is possible to estimate sparse changes in high-dimensional regression coefficients, even if the coefficients themselves are dense.
- Use complementary sketching to eliminate nuisance parameter.
- Implementation available in `github.com/gaofengnan/charcoal/`

Summary

THE LONDON SCHOOL OF ECONOMICS AND POLITICAL SCIENCE

- It is possible to estimate sparse changes in high-dimensional regression coefficients, even if the coefficients themselves are dense.
- Use complementary sketching to eliminate nuisance parameter.
- Implementation available in `github.com/gaofengnan/charcoal/`

- Main references:

Gao, F. and Wang, T. (2022) Two-sample testing of high-dimensional linear regression coefficients via complementary sketching. *Ann. Statist.*, **50**, 2950–2972.

Gao, F. and Wang, T. (2022+) Sparse change detection in high-dimensional linear regression. *Preprint*, arxiv:2208.06326.

Thank you!

References

- Bai, J. (1997) Estimation of a change point in multiple regression models. *Review of Economics and Statistics*, **79**, 551–563.
- Bai, J. and Perron, P. (1998) Estimating and testing linear models with multiple structural changes. *Econometrica*, **66**, 47–78.
- Baranowski, R., Chen, Y. and Fryzlewicz, P. (2019) Narrowest-over-threshold detection of multiple change points and change-point-like features. *J. Roy. Statist. Soc., Ser. B*, **81**, 649–672.
- Battey, H. S. and Cox, D. R. (2020) High dimensional nuisance parameters: an example from parametric survival analysis. *Information Geometry*, **3**, 119–148.
- Cho, H. and Fryzlewicz, P. (2015) Multiple-change-point detection for high dimensional time series via sparsified binary segmentation. *J. Roy. Statist. Soc., Ser. B*, **77**, 475–507.
- Enikeeva, F. and Harchaoui, Z. (2019) High-dimensional change-point detection under sparse alternatives. *Ann. Statist.*, **47**, 2051–2079.
- Jirak, M. (2015) Uniform change point tests in high dimension. *Ann. Statist.*, **43**, 2451–2483.

- Julious, S. A. (2001) Inference and estimation in a changepoint regression problem. *J. Roy. Statist. Soc., Ser. D*, **50**, 51–61.
- Kaul, A., Jandhyala, V. K. and Fotopoulos, S. B. (2019) An efficient two step algorithm for high dimensional change point regression models without grid search. *J. Mach. Learn. Res.*, **20**, (111), 1–40.
- Lee, S., Seo, M. H. and Shin, Y. (2016) The lasso for high dimensional regression with a possible change point. *J. Roy. Statist. Soc., Ser. B*, **78**, 193–210.
- Leonardi, F. and Bühlmann, P. (2016) Computationally efficient change point detection for high-dimensional regression. *arXiv preprint*, arXiv:1601.03704.
- Neyman, J. and Scott, E. L. (1948) Consistent estimates based on partially consistent observations. *Econometrica*, **16**, 1–32.
- Rinaldo, A., Wang, D., Wen, Q., Willett, R. and Yu, Y. (2021) Localizing changes in high-dimensional regression models. *Proc. Mach. Learn. Res.*, **130**.
- Wang, D., Yu, Y. and Rinaldo, A. (2021) Optimal covariance change point localization in high dimensions. *Bernoulli*, **27**, 554–575.
- Wang, T. and Samworth, R. J. (2018) High dimensional change point estimation via sparse projection. *J. Roy. Statist. Soc., Ser., B*, **80**, 57–83.