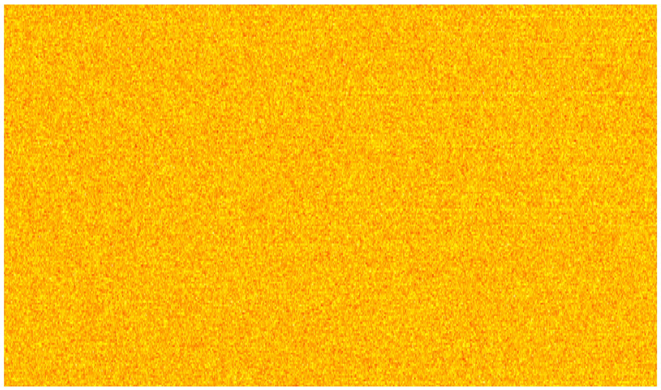
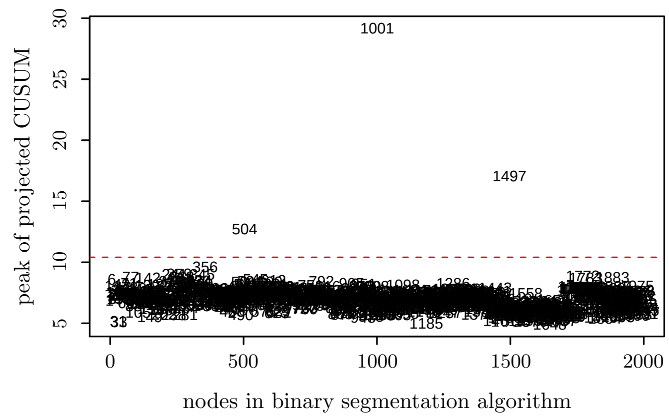
HIGH-DIMENSIONAL CHANGEPOINT DETECTION VIA SPARSE PROJECTION

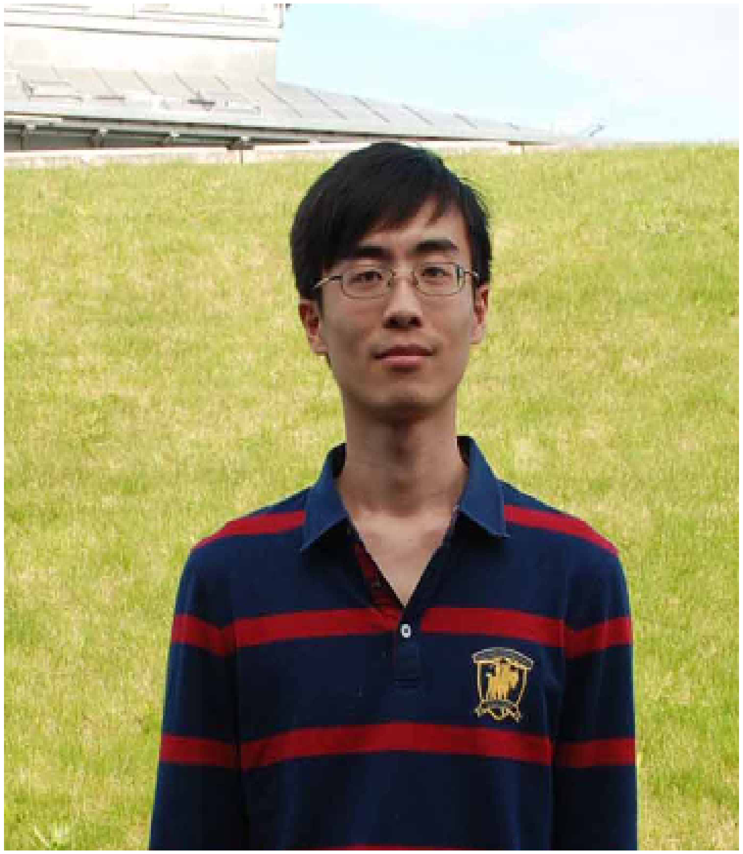




Richard Samworth, University of Cambridge

Joint work with Tengyao Wang

Tengyao



Heterogeneity in Big Data

One of the most commonly-encountered issues with Big Data is heterogeneity.

Departures from traditional, stylised i.i.d. models can take many forms, e.g. missing data, correlated errors, data combined from multiple sources,...

In data streams, heterogeneity is manifested through non-stationarity. Perhaps the simplest model assumes population changes occur at a finite set of time points.

Changepoint estimation

Changepoint problems have a rich history (Page, 1955).

State-of-the-art univariate methods include PELT (Killick, Fearnhead and Eckley, 2012)**, Wild Binary Segmentation** (Fryzlewicz, 2014) **and SMUCE** (Frick, Munk and Sieling, 2014).

Some ideas extend to multivariate settings (Horváth, Kokoszka and Steinebach,1999; Ombao, Von Sachs and Guo, 2005; Aue et al., 2009; Kirch, Mushal and Ombao, 2014).

Increasing interest in high-dimensional setting, possibly with a sparsity condition on coordinates of change (Aston and Kirch, 2014; Enikeeva and Harchaoui, 2014; Jirak, 2015; Cho and Fryzlewicz, 2015; Cho, 2016).

Basic model

Let $X = (X_1, \ldots, X_n) \in \mathbb{R}^{p \times n}$ **have independent columns** $X_t \sim N_p(\mu_t, \sigma^2 I_p)$**. Assume there exist changepoints** $1 \leq z_1 < z_2 < \cdots < z_\nu \leq n-1$ **such that**

$$\mu_{z_i+1} = \cdots = \mu_{z_{i+1}} =: \mu^{(i)}, \qquad \forall\, 0 \leq i \leq \nu,$$

where $z_0 := 0$ **and** $z_{\nu+1} := n$**. Writing**

$$\theta^{(i)} := \mu^{(i)} - \mu^{(i-1)}, \qquad 1 \leq i \leq \nu,$$

we assume $\exists k \in \{1, \ldots, p\}$ **s.t.** $\|\theta^{(i)}\|_0 \leq k$ **for** $1 \leq i \leq \nu$**.**

Further model assumptions

Assume stationary run lengths satisfy

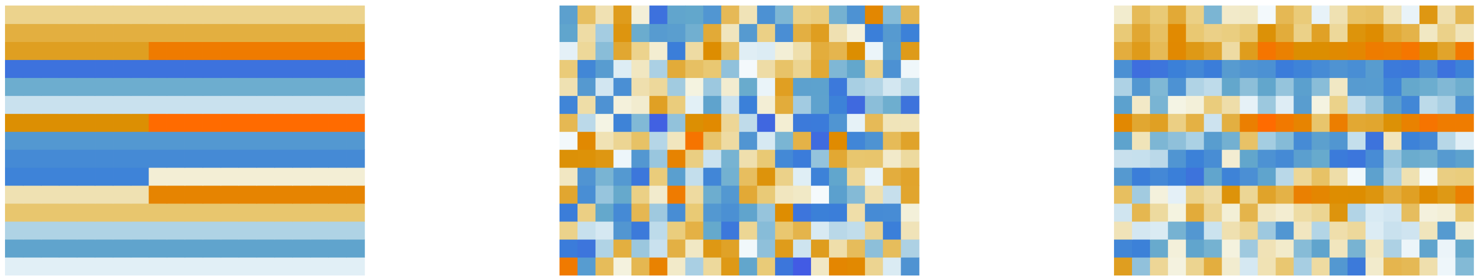
$$\frac{1}{n}\min\{z_{i+1} - z_i : 0 \leq i \leq \nu\} \geq \tau,$$

and the magnitudes of mean changes are such that

$$\|\theta^{(i)}\|_2 \geq \vartheta, \qquad \forall\, 1 \leq i \leq \nu.$$

Let $\mathcal{P}(n, p, k, \nu, \vartheta, \tau, \sigma^2)$ **be the set of distributions of such** X**.**

Projection-based single changepoint estimation



$$\boldsymbol{\mu} \quad + \quad W \quad = \quad X$$

Let $\nu = 1$**, write** $z := z_1$**,** $\theta := \theta^{(1)}$ **and** $\tau := n^{-1}\min\{z, n-z\}$**. For any** $a \in \mathbb{S}^{p-1}$**,**

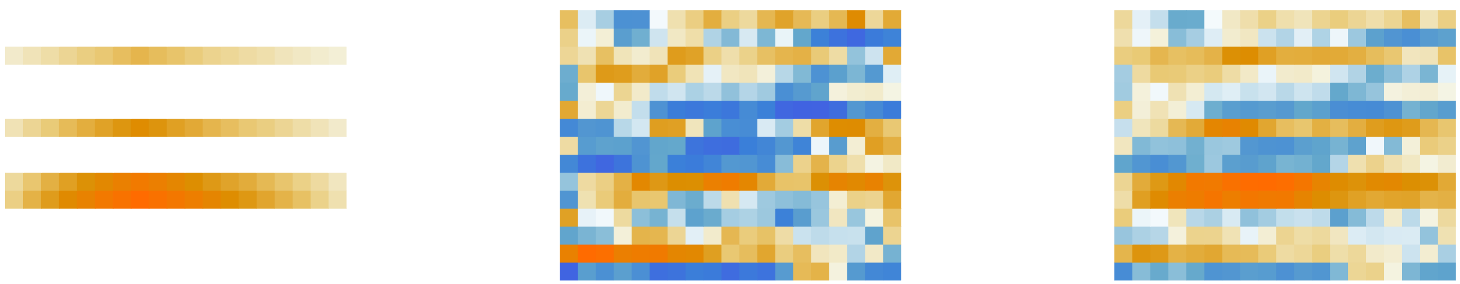
$$a^\top X_t \sim N(a^\top \mu_t, \sigma^2).$$

Hence $a = \theta/\|\theta\|_2 =: v$ **maximises the magnitude of the difference in means between the two segments.**

CUSUM transformation

Define CUSUM transformation $\mathcal{T}_{p,n} : \mathbb{R}^{p\times n} \to \mathbb{R}^{p\times(n-1)}$ **by**

$$[\mathcal{T}(M)]_{j,t} = [\mathcal{T}_{p,n}(M)]_{j,t} := \sqrt{\frac{t(n-t)}{n}}\left(\sum_{r=t+1}^{n} \frac{M_{j,r}}{n-t} - \sum_{r=1}^{t} \frac{M_{j,r}}{t}\right).$$



$$\mathcal{T}(\boldsymbol{\mu}) \quad + \quad \mathcal{T}(W) \quad = \quad \mathcal{T}(X)$$

$$A \quad + \quad E \quad = \quad T$$

SVD of CUSUM transformation

When $\nu = 1$, we can compute A explicitly:

$$A_{j,t} = \begin{cases} \sqrt{\frac{t}{n(n-t)}}(n-z)\theta_j, & \textbf{if } t \leq z \\ \sqrt{\frac{n-t}{nt}} z\theta_j, & \textbf{if } t > z \end{cases} =: (\theta\gamma^\top)_{j,t},$$

so the oracle projection direction is the leading left singular vector of the rank 1 matrix A.

We could therefore consider estimating v by $\hat{v}_{\max,k} \in \mathrm{argmax}_{\tilde{v}\in\mathbb{S}^{p-1}(k)} \|T^\top \tilde{v}\|_2$, and indeed when $n \geq 6$, with probability at least $1 - 4(p\log n)^{-1/2}$,

$$\sin\angle(\hat{v}_{\max,k}, v) \leq \frac{16\sqrt{2}\sigma}{\vartheta}\sqrt{\frac{k\log(p\log n)}{n\tau}}.$$

A computationally efficient projection

Computing the k-sparse leading left singular vector of a matrix is NP-hard (Tillmann and Pfetsch, 2014). However,

$$\max_{u \in \mathbb{S}^{p-1}(k)} \|u^\top T\|_2 = \max_{u \in \mathbb{S}^{p-1}(k), w \in \mathbb{S}^{n-2}} u^\top T w$$
$$= \max_{u \in \mathbb{S}^{p-1}, w \in \mathbb{S}^{n-2}, \|u\|_0 \le k} \langle u w^\top, T \rangle = \max_{M \in \mathcal{M}} \langle M, T \rangle,$$

where $\mathcal{M} := \{M : \|M\|_* = 1, \mathrm{rk}(M) = 1, \mathrm{nnzr}(M) \le k\}$. For $\lambda > 0$, we therefore consider computing

$$\hat{M} \in \operatorname*{argmax}_{M \in \mathcal{S}_1} \bigl\{ \langle T, M \rangle - \lambda \|M\|_1 \bigr\},$$

where $\mathcal{S}_1 := \{M \in \mathbb{R}^{p \times (n-1)} : \|M\|_* \le 1\}$, using ADMM. We can then let $\hat{v}$ be a leading left singular vector of $\hat{M}$.

Alternative relaxation

Let $\mathcal{S}_2 := \{M \in \mathbb{R}^{p\times(n-1)} : \|M\|_2 \leq 1\}$**. Then the simple dual formulation leads to**

$$\tilde{M} := \frac{\textbf{soft}(T,\lambda)}{\|\textbf{soft}(T,\lambda)\|_2} = \operatorname*{argmax}_{M\in\mathcal{S}_2}\bigl\{\langle T, M\rangle - \lambda\|M\|_1\bigr\}.$$

Suppose $\hat{M} \in \mathrm{argmax}_{M\in\mathcal{S}}\bigl\{\langle T, M\rangle - \lambda\|M\|_1\bigr\}$ **for** $\mathcal{S} = \mathcal{S}_1$ **or** $\mathcal{S} = \mathcal{S}_2$ **and let** $\hat{v} \in \mathrm{argmax}_{\tilde{v}\in\mathbb{S}^{p-1}} \|\hat{M}^\top \tilde{v}\|_2$**. If** $n \geq 6$ **and** $\lambda \geq 2\sigma\sqrt{\log(p\log n)}$**, then w.p. at least** $1 - 4(p\log n)^{-1/2}$**,**

$$\sin\angle(\hat{v}, v) \leq \frac{32\lambda\sqrt{k}}{\tau\vartheta\sqrt{n}}.$$

Changepoint estimation after projection

Input: $X \in \mathbb{R}^{p \times n}$, $\lambda > 0$.

Step 1: **Perform CUSUM transformation** $T \leftarrow \mathcal{T}(X)$

Step 2: **Find** $\hat{M} \in \mathrm{argmax}_{M \in \mathcal{S}} \bigl\{ \langle T, M \rangle - \lambda \|M\|_1 \bigr\}$ **for** $\mathcal{S} = \mathcal{S}_1$ **or** $\mathcal{S}_2$

Step 3: **Find** $\hat{v} \in \mathrm{argmax}_{\tilde{v} \in \mathbb{S}^{p-1}} \|\hat{M}^\top \tilde{v}\|_2$.

Step 4: **Let** $\hat{z} \in \mathrm{argmax}_{1 \le t \le n-1} |\hat{v}^\top T_t|$**, where** T_t **is the** t**th column of** T**, and set** $\bar{T}_{\max} \leftarrow |\hat{v}^\top T_{\hat{z}}|$

Output: $\hat{z}$, $\bar{T}_{\max}$

Sample-splitting version performance

Suppose $\sigma > 0$ **is known and** $X \sim P \in \mathcal{P}(n, p, k, 1, \vartheta, \tau, \sigma^2)$**. Let** $\hat{z}$ **be the output of sample-splitting algorithm with input** X**,** σ **and** $\lambda := 2\sigma\sqrt{\log(p \log n)}$**. If** $n \geq 6$ **is even and**

$$\frac{\sigma}{\vartheta\tau}\sqrt{\frac{k \log(p \log n)}{n}} \leq \frac{\sqrt{3}}{128},$$

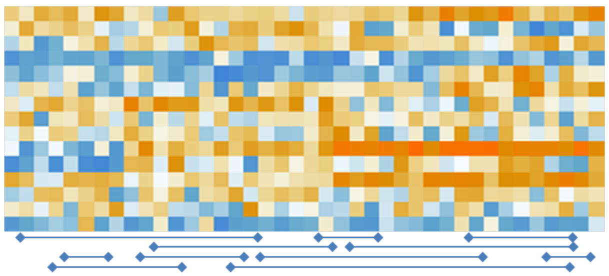
then with probability at least $1 - 4\{p \log(n/2)\}^{-1/2} - 2/n$**,**

$$\frac{1}{n}|\hat{z} - z| \leq \frac{32\sigma}{\vartheta}\sqrt{\frac{\log n}{n\tau}}.$$

If $\log p = O(\log n)$**,** $\vartheta \asymp n^{-a}$**,** $\tau \asymp n^{-b}$**,** $k \asymp n^c$ **and** $a + b + c/2 < 1/2$**, then rate of convergence is** $o(n^{-\frac{1-2a-b}{2}+\delta})$ **for all** $\delta > 0$**.**

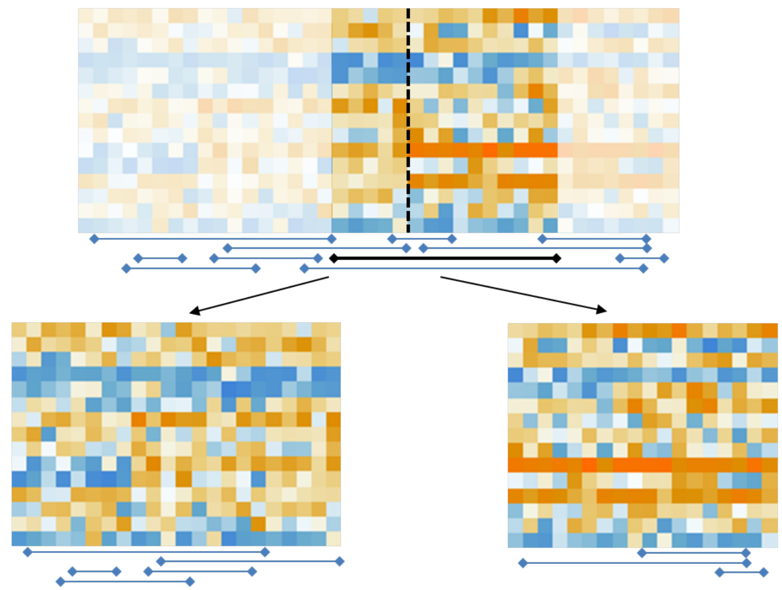
Multiple changepoint estimation – `inspect`

Wild binary segmentation scheme (Fryzlewicz, 2014)



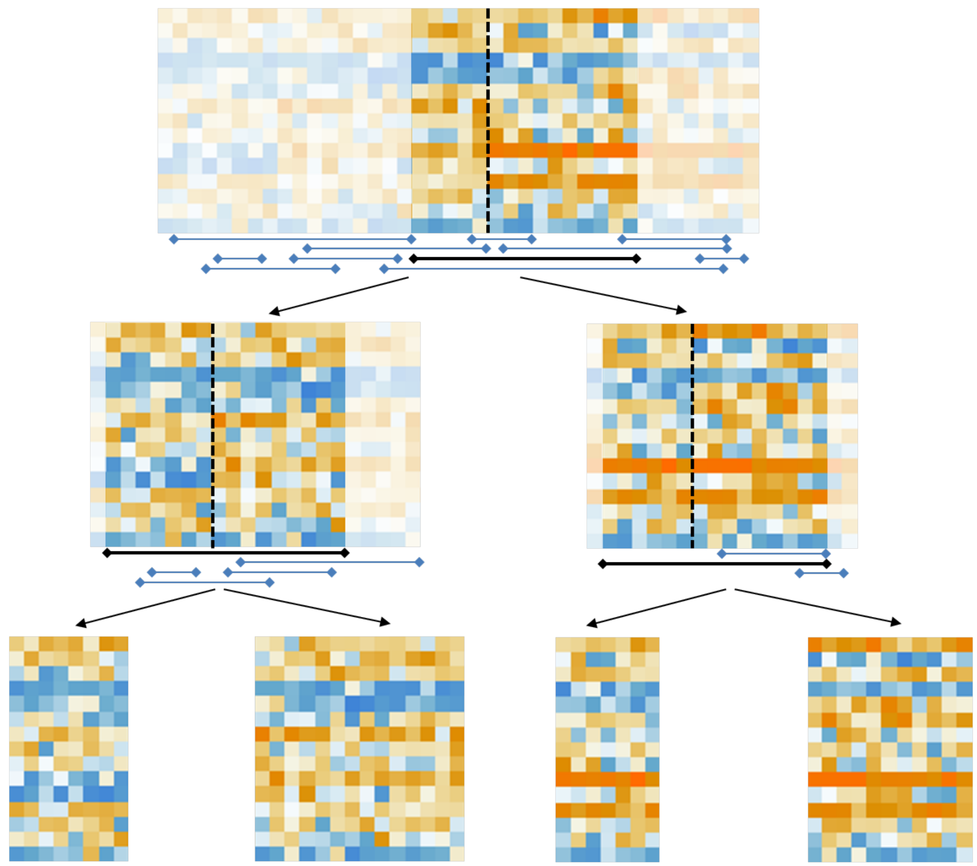
Multiple changepoint estimation – `inspect`

Wild binary segmentation scheme (Fryzlewicz, 2014)

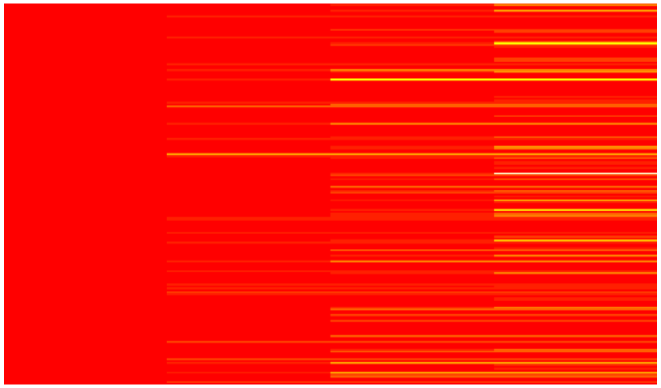


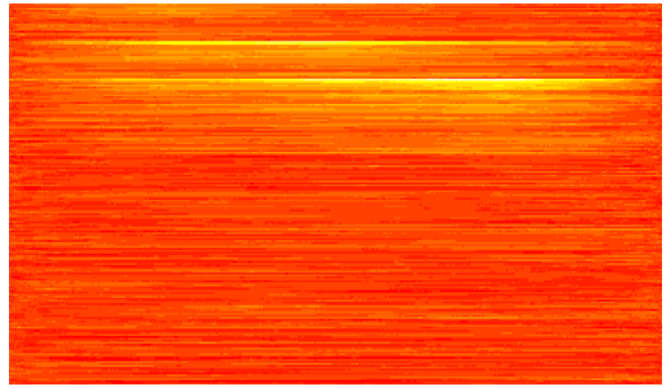
Multiple changepoint estimation – `inspect`

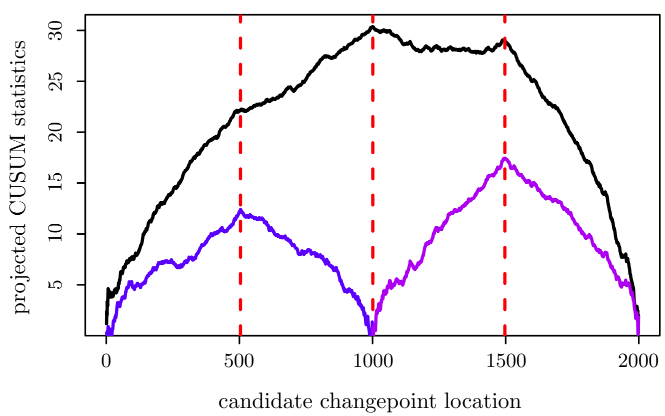
Wild binary segmentation scheme (Fryzlewicz, 2014)

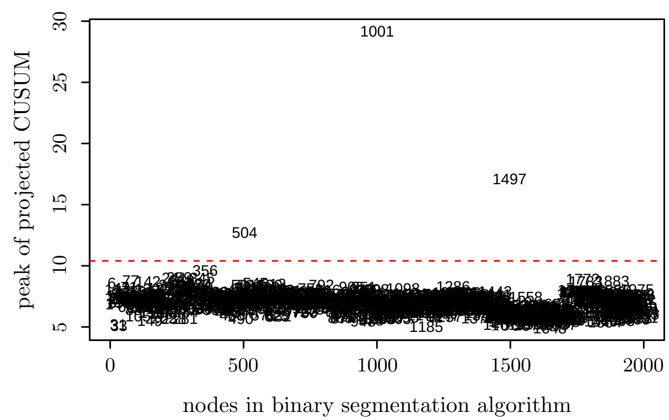


Example






projected CUSUM statistics
5
10
15
20
25
30
0
500
1000
1500
2000
candidate changepoint location


peak of projected CUSUM
5
10
15
20
25
30
1001
1497
504
356
1185
0
500
1000
1500
2000
nodes in binary segmentation algorithm

$\mathcal{S}_1$ or $\mathcal{S}_2$?

Angles (in degrees) between oracle projection direction v and estimated projection directions $\hat{v}_{\mathcal{S}_1}$ (using $\mathcal{S}_1$) and $\hat{v}_{\mathcal{S}_2}$ (using $\mathcal{S}_2$), for different choices of ϑ.

ϑ	0.1	0.2	0.3	0.4	0.5
$\angle(\hat{v}_{\mathcal{S}_1}, v)$	80.3	63.1	51.6	39.4	28.6
$\angle(\hat{v}_{\mathcal{S}_2}, v)$	79.5	63.9	52.9	40.6	30.2
ϑ	0.6	0.7	0.8	0.9	1
$\angle(\hat{v}_{\mathcal{S}_1}, v)$	25.8	21.7	19.0	16.7	14.4
$\angle(\hat{v}_{\mathcal{S}_2}, v)$	27.3	23.4	20.4	18.0	15.6

Single changepoint simulations – RMSE

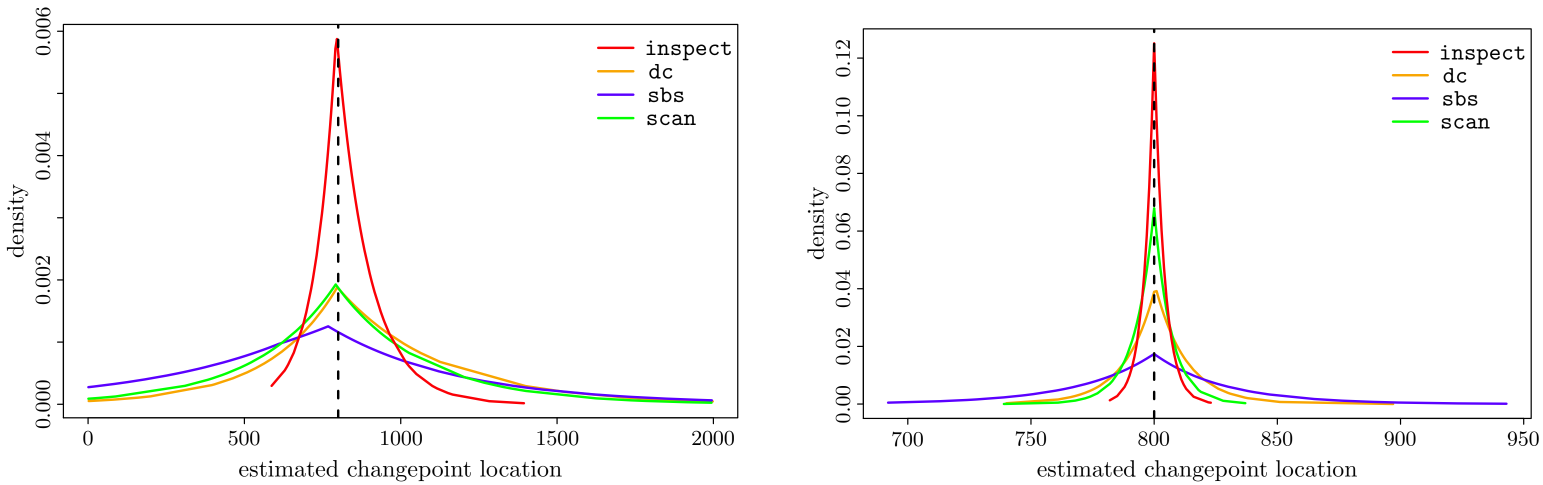
$\theta = (1, 2^{-1/2}, \ldots, k^{-1/2}, 0, \ldots, 0)^\top \in \mathbb{R}^p$.

n	p	k	z	ϑ	inspect	dc	sbs	scan
1000	200	10	400	0.57	**32.3**	82.2	99.6	46.2
1000	200	14	400	0.41	**97.2**	274.5	215.7	218.1
1000	200	200	400	0.57	**65.5**	262.3	180.1	156.4
1000	500	10	400	0.57	**48.2**	125.7	181.4	106.1
1000	500	22	400	0.52	**86.9**	240.5	235.5	190.3
1000	500	500	400	0.89	24.5	106.4	96.8	**22.5**
1000	1000	10	400	0.57	**48.6**	118.6	185.4	149.4
1000	1000	32	400	0.62	**58.7**	143.9	171.4	151.3
1000	1000	1000	400	1.26	**10.1**	28.1	42.7	15.1
2000	200	10	800	0.35	**126.3**	327.5	293.9	221.1
2000	200	14	800	0.41	**88.1**	213.7	155.2	121.0
2000	200	200	800	0.57	**57.6**	221.3	155.1	60.9
2000	500	10	800	0.35	**169.9**	348.1	456.0	305.5
2000	500	22	800	0.33	**195.2**	578.4	511.8	535.9
2000	500	500	800	0.89	**21.3**	45.0	62.4	27.0
2000	1000	10	800	0.35	**131.5**	416.4	460.5	397.7
2000	1000	32	800	0.40	**138.4**	441.0	448.6	401.6
2000	1000	1000	800	1.26	**6.7**	30.8	33.7	13.8

Changepoint density estimates

Left: $(n, p, k, z, \vartheta) = (2000, 1000, 32, 800, 0.40)$.

Right: $(n, p, k, z, \vartheta) = (2000, 1000, 32, 800, 1.02)$.



Misspecified settings

$(n, p, k, z, \vartheta) = (2000, 1000, 32, 800, 1.7)$.

Model	**inspect**	**dc**	**sbs**	**scan**
$\mathrm{M}_{\mathrm{unif}}$	**3.0**	13.8	17.6	3.8
$\mathrm{M}_{\mathrm{exp}}$	**2.8**	11.9	47.7	5.5
$\mathrm{M}_{\mathrm{cs,loc}}(0.2)$	**3.4**	8.4	17.5	6.8
$\mathrm{M}_{\mathrm{cs,loc}}(0.5)$	**5.6**	10.8	23.7	8.4
$\mathrm{M}_{\mathrm{cs}}(0.5)$	**1.5**	7.5	14.2	3.5
$\mathrm{M}_{\mathrm{cs}}(0.9)$	**2.5**	6.5	10.2	2.9
$\mathrm{M}_{\mathrm{temp}}(0.1)$	**4.0**	16.9	96.2	10.1
$\mathrm{M}_{\mathrm{temp}}(0.3)$	**14.5**	24.9	226.4	14.7

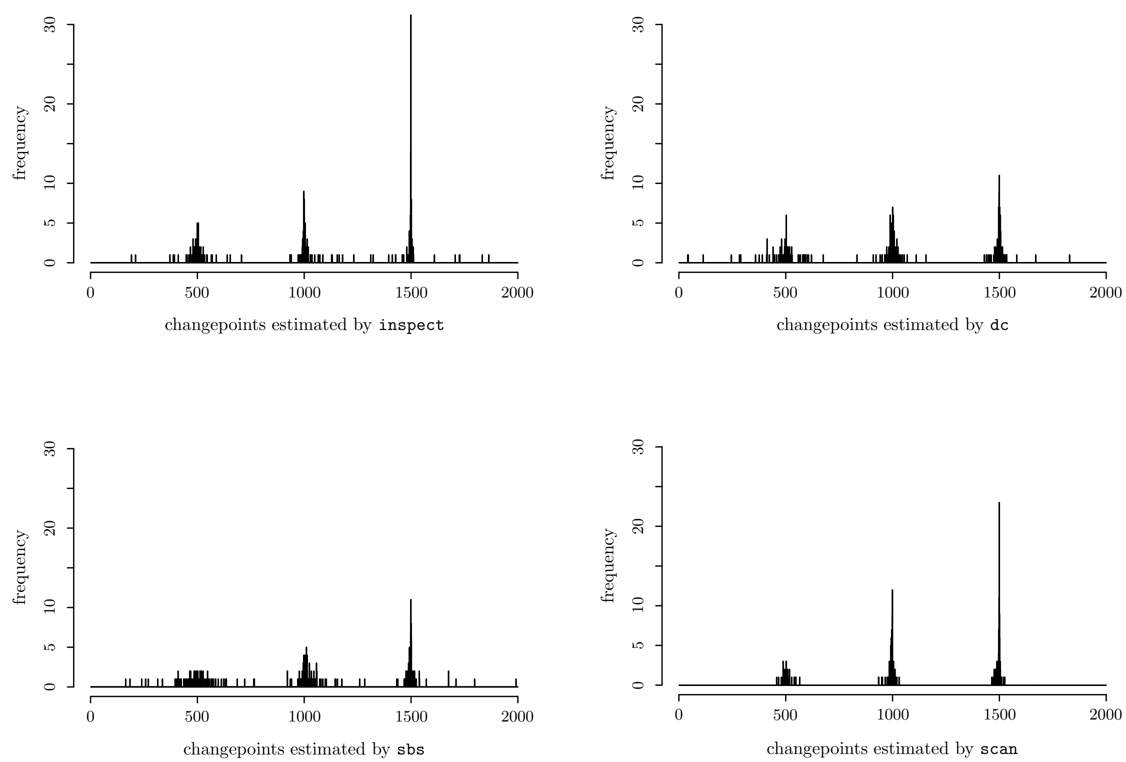
Multiple changepoint simulations

$n = 2000$, $p = 200$, $k = 40$, $z = (500, 1000, 1500)$. **Writing** $\vartheta^{(i)} := \|\theta^{(i)}\|_2$, **set** $(\vartheta^{(1)}, \vartheta^{(2)}, \vartheta^{(3)}) = \vartheta(1, 1.5, 2)$.

ϑ	method	$\hat{\nu}$						Rand	% best
		0	1	2	3	4	5		
0.63	**inspect**	0	0	10	73	14	3	0.91	46
	dc	0	0	23	63	13	1	0.86	14
	sbs	0	0	6	69	22	3	0.85	24
	scan	0	0	65	33	2	0	0.79	16
0.51	**inspect**	0	0	23	50	22	5	0.82	52
	dc	0	0	47	40	12	1	0.76	22
	sbs	0	0	30	48	14	8	0.77	20
	scan	0	0	94	6	0	0	0.71	7
0.38	**inspect**	0	0	48	42	10	0	0.77	55
	dc	0	7	66	23	4	0	0.69	18
	sbs	0	0	58	36	6	0	0.70	14
	scan	0	11	88	1	0	0	0.68	26

Histograms of estimated changepoints

$n = 2000$, $p = 200$, $k = 40$, $z = (500, 1000, 1500)$,
$(\vartheta^{(1)}, \vartheta^{(2)}, \vartheta^{(3)}) = (0.63, 0.95, 1.26)$, $\sigma = 1$.



Summary

- `inspect` **is a new method for high-dimensional changepoint estimation**
- **Convex relaxation used to find projection direction, then CUSUM and WBS to identify multiple changepoints**
- `R` **package** `InspectChangepoint` **available!**

References

- **Aston, J. A. D. and Kirch, C. (2014) Change points in high dimensional settings. `arxiv:1409.1771`.**
- **Aue, A., Hörmann, S., Horváth, L. and Reimherr, M. (2009). Break detection in the covariance structure of multivariate time series models. *Ann. Statist.* 37, 4046–4087.**
- **Cho, H. (2016) Change-point detection in panel data via double CUSUM statistic. *preprint*.**
- **Cho, H. and Fryzlewicz, P. (2015) Multiple changepoint detection for high dimensional time series via sparsified binary segmentation. *J. R. Stat. Soc. Ser. B*, 77, 475–507.**
- **Enikeeva, F. and Harchaoui, Z. (2014) High-dimensional change-point detection with sparse alternatives. `arxiv:1312.1900v2`.**
- **Frick, K., Munk, A. and Sieling, H. (2014) Multiscale change point inference. *J. R. Stat. Soc. Ser. B* 76, 495–580.**
- **Fryzlewicz, P. (2014) Wild binary segmentation for multiple change-point detection. *Ann. Statist.*, 42, 2243–2281.**

- **Horváth, L., Kokoszka, P. and Steinebach, J. (1999) Testing for changes in dependent observations with an application to temperature changes. *J. Multi. Anal.*, 68, 96–199.**
- **Jirak, M. (2015) Uniform change point tests in high dimension. *Ann. Statist.*, 43, 2451–2483.**
- **Killick, R., Fearnhead, P. and Eckley, I. A. (2012) Optimal detection of changepoints with a linear computational cost. *J. Amer. Stat. Assoc.* 107, 1590–1598.**
- **Kirch, C., Mushal, B. and Ombao, H. (2015) Detection of changes in multivariate time series with applications to EEG data. *J. Amer. Statist. Assoc.*, 110, 1197–1216.**
- **Ombao, H., Von Sachs, R. and Guo, W. (2005) SLEX analysis of multivariate nonstationary time series. *J. Amer. Statist. Assoc.*, 100, 519–531.**
- **Page, E. S. (1955) A test for a change in a parameter occurring at an unknown point. *Biometrika*, 42, 523–527.**
- **Tillmann, A. N. and Pfetsch M. E. (2014) The computational complexity of the restricted isometry property, the nullspace property, and related concepts in compressed sensing. *IEEE Trans. Inform. Theory*, 60, 1248–1259.**
- **Wang, T. and Samworth, R. J. (2016) High-dimensional changepoint estimation via sparse projection.** `http://arxiv.org/abs/1606.06246.`