Inferring change points in high-dimensional linear regression via Approximate Message Passing

Xiaoqi (Shirley) Liu
University of Cambridge

Joint work with
Gabriel Arpino, Ramji Venkataramanan

International Conference on Machine Learning (ICML) 2024, July 2024

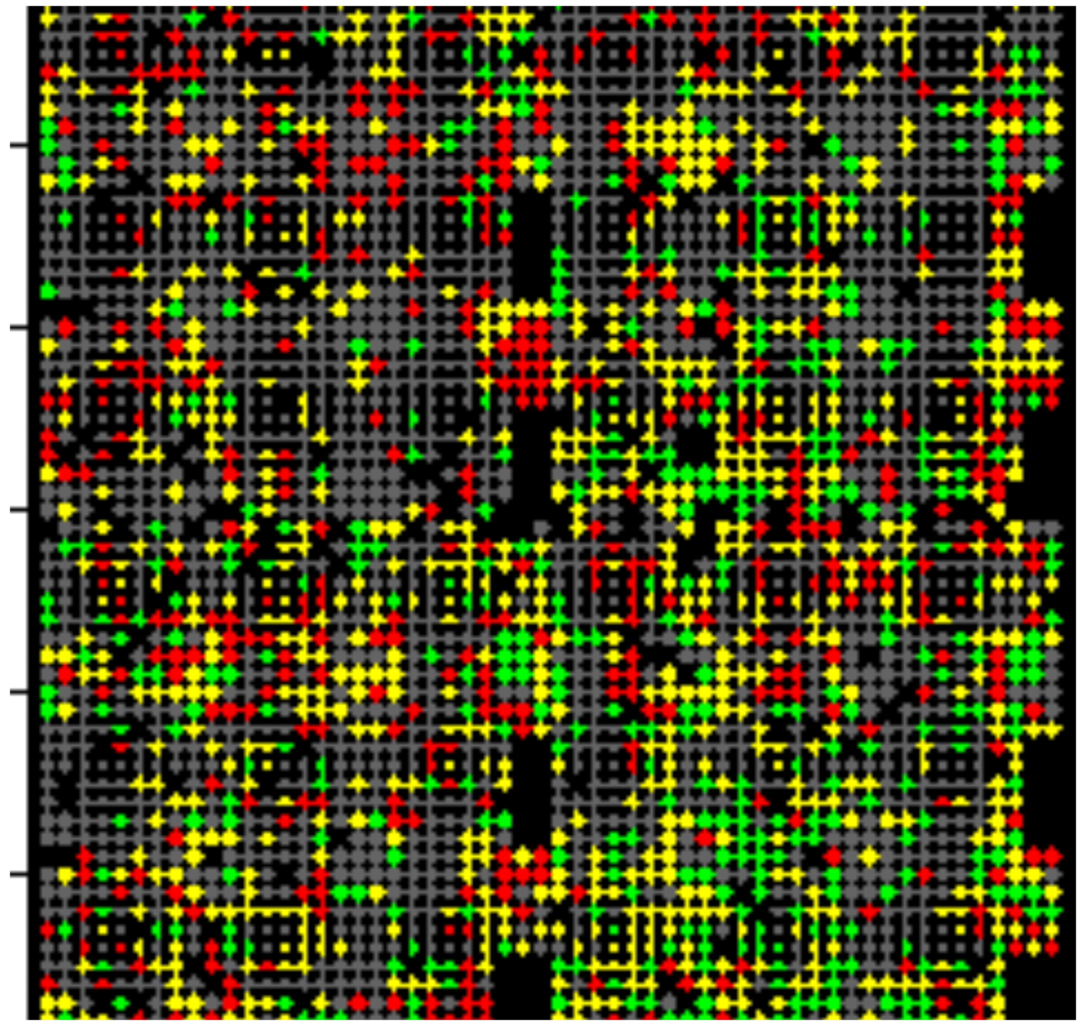
Vast amounts of **time-ordered, non-stationary** data

Vast amounts of **time-ordered, non-stationary** data

Stock prices

Vast amounts of **time-ordered, non-stationary** data

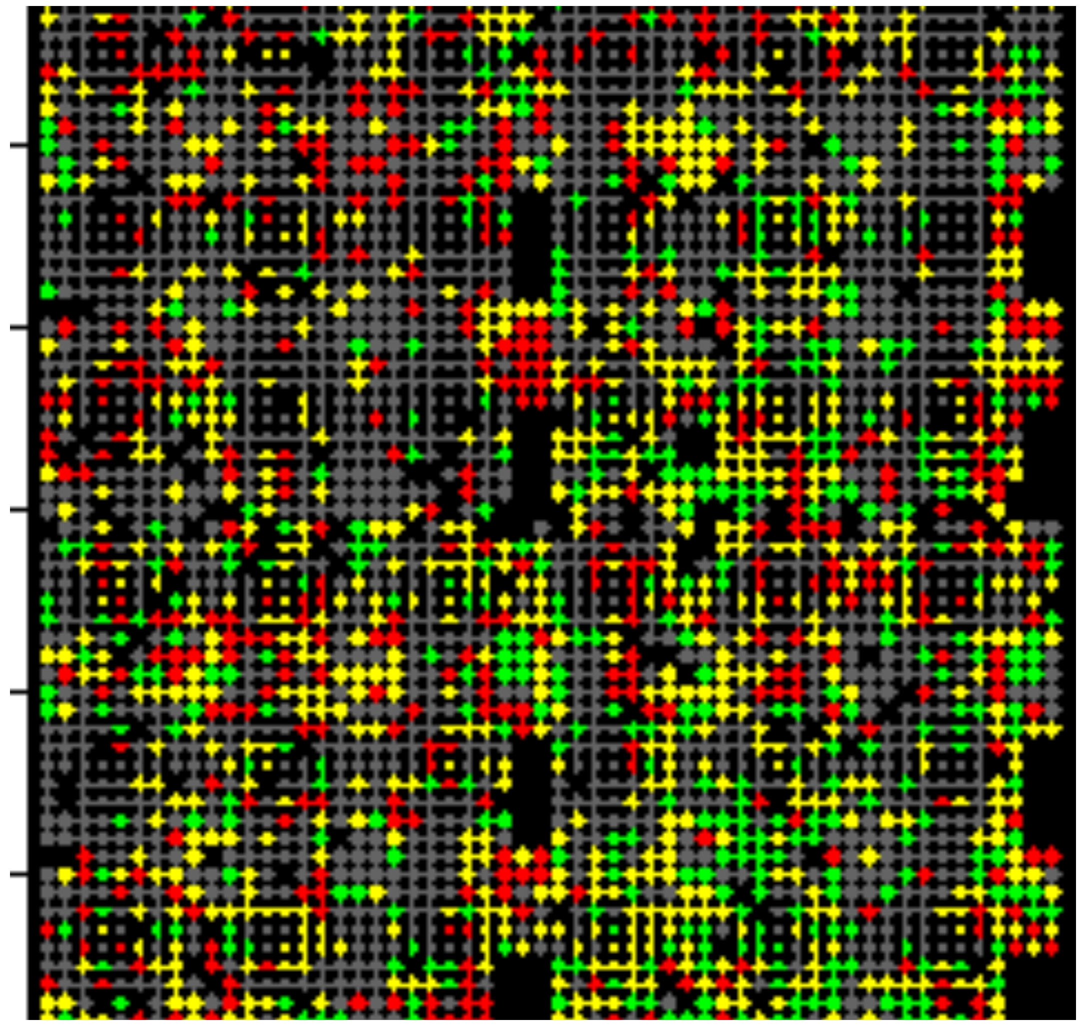
Stock prices



Gene expression

Vast amounts of **time-ordered, non-stationary** data

Stock prices



Gene expression

Want to understand how underlying **data generating mechanisms** change over time

For $i = 1, 2, \ldots, n$,

$$\boldsymbol{x}_i, \boldsymbol{\beta} \in \mathbb{R}^p$$

For $i = 1, 2, \ldots, n,$

$\boldsymbol{x}_i, \boldsymbol{\beta} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta} + \varepsilon_i$$

For $i = 1, 2, \ldots, n$,

$\boldsymbol{x}_i, \boldsymbol{\beta}^{(l)} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(1)} + \varepsilon_i$$

For $i = 1, 2, \ldots, n$,

$\boldsymbol{x}_i, \boldsymbol{\beta}^{(l)} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(1)} + \varepsilon_i$$

For $i = 1, 2, \ldots, n,$

$\boldsymbol{x}_i, \boldsymbol{\beta}^{(l)} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(1)} + \varepsilon_i$$

$i = \eta_1$

For $i = 1, 2, \ldots, n$,

$\boldsymbol{x}_i, \boldsymbol{\beta}^{(l)} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(1)} + \varepsilon_i$$

$i = \eta_1$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(2)} + \varepsilon_i$$

For $i = 1, 2, \ldots, n,$

$\boldsymbol{x}_i, \boldsymbol{\beta}^{(l)} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(1)} + \varepsilon_i$$

$i = \eta_1$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(2)} + \varepsilon_i$$

For $i = 1, 2, \ldots, n$,

$\boldsymbol{x}_i, \boldsymbol{\beta}^{(l)} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(1)} + \varepsilon_i$$

$$i = \eta_1$$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(2)} + \varepsilon_i$$

$$\vdots$$

For $i = 1, 2, \ldots, n$,

$\boldsymbol{x}_i, \boldsymbol{\beta}^{(l)} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(1)} + \varepsilon_i$$

$i = \eta_1$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(2)} + \varepsilon_i$$

$\vdots$

$i = \eta_{L^*-1}$

For $i = 1, 2, \ldots, n$,

$\boldsymbol{x}_i, \boldsymbol{\beta}^{(l)} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(1)} + \varepsilon_i$$

$i = \eta_1$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(2)} + \varepsilon_i$$

$\vdots$

$i = \eta_{L^*-1}$

L^* unknown constant,
given $L^* < L$

For $i = 1, 2, \ldots, n$,

$\boldsymbol{x}_i, \boldsymbol{\beta}^{(l)} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(1)} + \varepsilon_i$$

$i = \eta_1$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(2)} + \varepsilon_i$$

$\vdots$

$i = \eta_{L^*-1}$

L^* unknown constant,

given $L^* < L$

$\boldsymbol{\eta} := [\eta_1, \ldots, \eta_{L^*-1}]$

For $i = 1, 2, \ldots, n$,

$\boldsymbol{x}_i, \boldsymbol{\beta}^{(l)} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(1)} + \varepsilon_i$$

$$i = \eta_1$$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(2)} + \varepsilon_i$$

$$\vdots$$

$$i = \eta_{L^*-1}$$

L^* unknown constant, given $L^* < L$

$\boldsymbol{\eta} := [\eta_1, \ldots, \eta_{L^*-1}]$

Goal: recover change point locations $\{\eta_l\}_{l=1}^{L^*-1}$ from $\{\boldsymbol{x}_i, y_i\}_{i=1}^{n}$, and estimate signals $\{\boldsymbol{\beta}^{(l)}\}_{l=1}^{L^*}$

For $i = 1, 2, \ldots, n$,

$\boldsymbol{x}_i, \boldsymbol{\beta}^{(l)} \in \mathbb{R}^p$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(1)} + \varepsilon_i$$

$i = \eta_1$

$$y_i = \boldsymbol{x}_i^\top \boldsymbol{\beta}^{(2)} + \varepsilon_i$$

$\vdots$

$i = \eta_{L^*-1}$

L^* unknown constant, given $L^* < L$

$\boldsymbol{\eta} := [\eta_1, \ldots, \eta_{L^*-1}]$

Goal: recover change point locations $\{\eta_l\}_{l=1}^{L^*-1}$ from $\{\boldsymbol{x}_i, y_i\}_{i=1}^n$, and estimate signals $\{\boldsymbol{\beta}^{(l)}\}_{l=1}^{L^*}$

Scaling regime: $n, p \to \infty$, $n/p \to$ fixed constant δ

$$\hat{\eta}_1, \ldots, \hat{\eta}_{L^*-1} = \underbrace{\arg\min_{\tilde{\eta}_1 \leq \tilde{\eta}_2 \leq \ldots \leq \tilde{\eta}_{L^*-1}} \sum_{l=1}^{L^*}}_{\text{search through different partitioning}} \underbrace{\min_{\tilde{\boldsymbol{\beta}} \in \mathbb{R}^p} \sum_{i=\tilde{\eta}_{l-1}+1}^{\tilde{\eta}_l} \left(y_i - \boldsymbol{x}_i^\top \tilde{\boldsymbol{\beta}}\right)^2}_{\text{Residual Sum of Squares}}$$

$p \ll n$

$$\hat{\eta}_1, \ldots, \hat{\eta}_{L^*-1} = \underbrace{\arg\min_{\tilde{\eta}_1 \le \tilde{\eta}_2 \le \ldots \le \tilde{\eta}_{L^*-1}} \sum_{l=1}^{L^*}}_{\text{search through different partitioning}} \underbrace{\min_{\tilde{\boldsymbol{\beta}} \in \mathbb{R}^p} \sum_{i=\tilde{\eta}_{l-1}+1}^{\tilde{\eta}_l} \left(y_i - \boldsymbol{x}_i^\top \tilde{\boldsymbol{\beta}}\right)^2}_{\text{Residual Sum of Squares}}$$

~~Residual Sum of Squares~~ ~~$p \ll n$~~

penalised maximum likelihood e.g., LASSO

$n, p \to \infty$

[Zhang et al. 2015, Leonardi and Bühlmann 2016, Lee et al. 2016, Rinaldo et al. 2021]

Sparse prior: penalised maximum likelihood + partitioning

[Zhang et al. 2015, Leonardi and Bühlmann 2016, Lee et al. 2016, Rinaldo et al. 2021]

Sparse prior: penalised maximum likelihood + partitioning

[Zhang et al. 2015, Leonardi and Bühlmann 2016, Lee et al. 2016, Rinaldo et al. 2021]

Sparse difference prior: complementary sketching

[Wang et al. 2021, Gao and Wang 2022]

Sparse prior: penalised maximum likelihood + partitioning

[Zhang et al. 2015, Leonardi and Bühlmann 2016, Lee et al. 2016, Rinaldo et al. 2021]

Sparse difference prior: complementary sketching

[Wang et al. 2021, Gao and Wang 2022]

Limitations:

Sparse prior: penalised maximum likelihood + partitioning

[Zhang et al. 2015, Leonardi and Bühlmann 2016, Lee et al. 2016, Rinaldo et al. 2021]

Sparse difference prior: complementary sketching

[Wang et al. 2021, Gao and Wang 2022]

Limitations:

- Restricted to certain signal priors
- Provide point estimates, without uncertainty quantification

Sparse prior: penalised maximum likelihood + partitioning

[Zhang et al. 2015, Leonardi and Bühlmann 2016, Lee et al. 2016, Rinaldo et al. 2021]

Sparse difference prior: complementary sketching

[Wang et al. 2021, Gao and Wang 2022]

Limitations:

- Restricted to certain signal priors
- Provide point estimates, without uncertainty quantification

We address these limitations via:

Sparse prior: penalised maximum likelihood + partitioning

[Zhang et al. 2015, Leonardi and Bühlmann 2016, Lee et al. 2016, Rinaldo et al. 2021]

Sparse difference prior: complementary sketching

[Wang et al. 2021, Gao and Wang 2022]

Limitations:

- Restricted to certain signal priors
- Provide point estimates, without uncertainty quantification

We address these limitations via:

Approximate Message Passing (AMP)

[Donoho et al. 2009, Bayati and Montanari 2011, Rangan 2011, Berthier et al. 2019, Celentano and Montanari 2022, Feng et al. 2022, Gerbelot and Berthier 2023]

Aside: AMP for sparse recovery

Aside: AMP for sparse recovery

$$\boxed{\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \qquad \begin{array}{l} \text{entries of } \boldsymbol{\beta} \overset{\text{iid}}{\sim} p_{\bar{\beta}} \\ \text{entries of } \boldsymbol{X} \overset{\text{iid}}{\sim} N(0,1/n) \end{array}}$$

Aside: AMP for sparse recovery

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

entries of $\boldsymbol{\beta} \overset{\text{iid}}{\sim} p_{\bar{\beta}}$

entries of $\boldsymbol{X} \overset{\text{iid}}{\sim} \mathcal{N}(0,1/n)$

modified residual: $\boldsymbol{r}^t = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^t + b_t \boldsymbol{r}^{t-1}$

signal estimate: $\boldsymbol{\beta}^{t+1} = \eta_{\substack{\text{soft}\\\text{thres}}}\left(\boldsymbol{\beta}^t + \boldsymbol{X}^\top \boldsymbol{r}^t; \theta_t\right)$

Aside: AMP for sparse recovery

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

entries of $\boldsymbol{\beta} \overset{\text{iid}}{\sim} p_{\bar{\beta}}$

entries of $\boldsymbol{X} \overset{\text{iid}}{\sim} N(0,1/n)$

modified residual: $\boldsymbol{r}^t = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^t \boxed{+ b_t \boldsymbol{r}^{t-1}}$ debias term

signal estimate: $\boldsymbol{\beta}^{t+1} = \eta_{\substack{\text{soft}\\\text{thres}}}\left(\boldsymbol{\beta}^t + \boldsymbol{X}^\top \boldsymbol{r}^t; \theta_t\right)$

Aside: AMP for sparse recovery

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

entries of $\boldsymbol{\beta} \overset{\text{iid}}{\sim} p_{\bar{\beta}}$

entries of $\boldsymbol{X} \overset{\text{iid}}{\sim} N(0,1/n)$

modified residual: $\boldsymbol{r}^t = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^t \boxed{+ b_t \boldsymbol{r}^{t-1}}$ debias term

signal estimate: $\boldsymbol{\beta}^{t+1} = \eta_{\substack{\text{soft}\\\text{thres}}}\left(\boldsymbol{\beta}^t + \boldsymbol{X}^\top \boldsymbol{r}^t; \theta_t\right)$

distribution of entries of $\boldsymbol{\beta}^t + \boldsymbol{X}^\top \boldsymbol{r}^t - \boldsymbol{\beta} \quad \to N(0, \kappa^t)$

Aside: AMP for sparse recovery

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

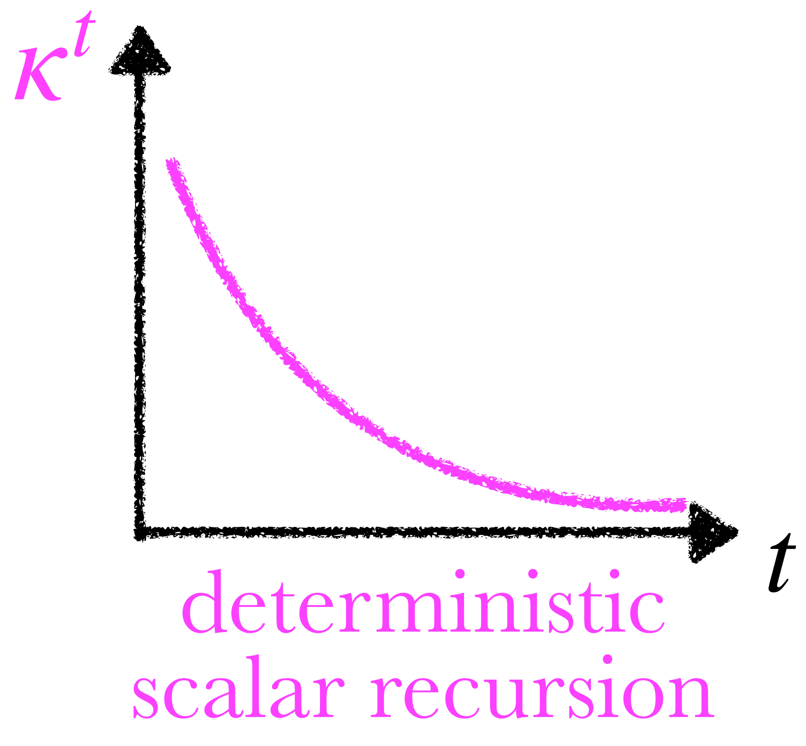
entries of $\boldsymbol{\beta} \overset{\text{iid}}{\sim} p_{\bar{\beta}}$

entries of $\boldsymbol{X} \overset{\text{iid}}{\sim} N(0, 1/n)$

modified residual: $\boldsymbol{r}^t = \boldsymbol{y} - \boldsymbol{X}\boldsymbol{\beta}^t \boxed{+ b_t \boldsymbol{r}^{t-1}}$ debias term

signal estimate: $\boldsymbol{\beta}^{t+1} = \eta_{\substack{\text{soft}\\\text{thres}}}\left(\boldsymbol{\beta}^t + \boldsymbol{X}^\top \boldsymbol{r}^t; \theta_t\right)$

distribution of entries of $\boldsymbol{\beta}^t + \boldsymbol{X}^\top \boldsymbol{r}^t - \boldsymbol{\beta} \rightarrow N(0, \kappa^t)$



AMP for change point inference

AMP for change point inference

Signal is $\boldsymbol{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \boldsymbol{\beta}^{(1)} & \dots & \boldsymbol{\beta}^{(L)} \\ \downarrow & & \downarrow \end{bmatrix} \sim p_{\bar{\boldsymbol{B}}}$ and define $\boldsymbol{\Theta} := \boldsymbol{X}\boldsymbol{B}$.

AMP for change point inference

Signal is $\boldsymbol{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \beta^{(1)} & \dots & \beta^{(L)} \\ \downarrow & & \downarrow \end{bmatrix} \sim p_{\bar{\boldsymbol{B}}}$ and define $\boldsymbol{\Theta} := X\boldsymbol{B}$.

$$\boldsymbol{\Theta}^t = X\hat{\boldsymbol{B}}^t - \boldsymbol{R}^{t-1}\left(\boldsymbol{F}^t\right)^\top$$

$$\boldsymbol{R}^t = g^t\left(\boldsymbol{\Theta}^t, \boldsymbol{y}\right)$$

produce residual $\boldsymbol{R}^t$

infer change points

$$\boldsymbol{B}^{t+1} = X^\top \boldsymbol{R}^t - \hat{\boldsymbol{B}}^t\left(\boldsymbol{C}^t\right)^\top$$

$$\hat{\boldsymbol{B}}^t = f^t\left(\boldsymbol{B}^t\right)$$

produce signal estimate $\hat{\boldsymbol{B}}^t$

AMP for change point inference

Signal is $\boldsymbol{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \beta^{(1)} & \dots & \beta^{(L)} \\ \downarrow & & \downarrow \end{bmatrix} \sim p_{\bar{B}}$ and define $\boldsymbol{\Theta} := X\boldsymbol{B}$.

$$\boxed{\begin{aligned} \boldsymbol{\Theta}^t &= X\hat{\boldsymbol{B}}^t - \boldsymbol{R}^{t-1}\left(\boldsymbol{F}^t\right)^\top \\ \boldsymbol{R}^t &= g^t\left(\boldsymbol{\Theta}^t, \boldsymbol{y}\right) \end{aligned}} \qquad \boxed{\begin{aligned} \boldsymbol{B}^{t+1} &= X^\top \boldsymbol{R}^t - \hat{\boldsymbol{B}}^t\left(\boldsymbol{C}^t\right)^\top \\ \hat{\boldsymbol{B}}^t &= f^t\left(\boldsymbol{B}^t\right) \end{aligned}}$$

produce residual $\boldsymbol{R}^t$

infer change points

produce signal estimate $\hat{\boldsymbol{B}}^t$

Bayes-optimal

$$f^t\left(\boldsymbol{B}^t\right) = \mathbb{E}\left[\bar{\boldsymbol{B}} \mid \bar{\boldsymbol{B}} + \boldsymbol{G}_B^t = \boldsymbol{B}^t\right]$$

AMP for change point inference

Signal is $\boldsymbol{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \beta^{(1)} & \dots & \beta^{(L)} \\ \downarrow & & \downarrow \end{bmatrix} \sim p_{\bar{B}}$ and define $\boldsymbol{\Theta} := X\boldsymbol{B}$.

g^t explicitly uses prior info about change points $\boldsymbol{\eta} \sim p_{\bar{\eta}}$

$$\boxed{\begin{aligned} \boldsymbol{\Theta}^t &= X\hat{\boldsymbol{B}}^t - \boldsymbol{R}^{t-1}\left(\boldsymbol{F}^t\right)^\top \\ \boldsymbol{R}^t &= g^t\left(\boldsymbol{\Theta}^t, \boldsymbol{y}\right) \end{aligned}}$$

produce residual $\boldsymbol{R}^t$

infer change points

$$\boxed{\begin{aligned} \boldsymbol{B}^{t+1} &= X^\top \boldsymbol{R}^t - \hat{\boldsymbol{B}}^t\left(\boldsymbol{C}^t\right)^\top \\ \hat{\boldsymbol{B}}^t &= f^t\left(\boldsymbol{B}^t\right) \end{aligned}}$$

produce signal estimate $\hat{\boldsymbol{B}}^t$

Bayes-optimal

$$f^t\left(\boldsymbol{B}^t\right) = \mathbb{E}\left[\bar{\boldsymbol{B}} \mid \bar{\boldsymbol{B}} + \boldsymbol{G}_B^t = \boldsymbol{B}^t\right]$$

AMP for change point inference

Signal is $\boldsymbol{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \boldsymbol{\beta}^{(1)} & \dots & \boldsymbol{\beta}^{(L)} \\ \downarrow & & \downarrow \end{bmatrix} \sim p_{\bar{\boldsymbol{B}}}$ and define $\boldsymbol{\Theta} := \boldsymbol{X}\boldsymbol{B}$.

g^t explicitly uses prior info about change points $\boldsymbol{\eta} \sim p_{\bar{\eta}}$

$$\boxed{\begin{aligned} \boldsymbol{\Theta}^t &= \boldsymbol{X}\hat{\boldsymbol{B}}^t - \boldsymbol{R}^{t-1}\left(\boldsymbol{F}^t\right)^\top \\ \boldsymbol{R}^t &= g^t\left(\boldsymbol{\Theta}^t, \boldsymbol{y}\right) \end{aligned}}$$

produce residual $\boldsymbol{R}^t$

infer change points

$$\boxed{\begin{aligned} \boldsymbol{B}^{t+1} &= \boldsymbol{X}^\top \boldsymbol{R}^t - \hat{\boldsymbol{B}}^t\left(\boldsymbol{C}^t\right)^\top \\ \hat{\boldsymbol{B}}^t &= f^t\left(\boldsymbol{B}^t\right) \end{aligned}}$$

produce signal estimate $\hat{\boldsymbol{B}}^t$

Bayes-optimal

$f^t\left(\boldsymbol{B}^t\right) = \mathbb{E}\left[\bar{\boldsymbol{B}} \mid \bar{\boldsymbol{B}} + \boldsymbol{G}_B^t = \boldsymbol{B}^t\right]$

Theorem (informal)

AMP for change point inference

Signal is $\boldsymbol{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \boldsymbol{\beta}^{(1)} & \dots & \boldsymbol{\beta}^{(L)} \\ \downarrow & & \downarrow \end{bmatrix} \sim p_{\bar{\boldsymbol{B}}}$ and define $\boldsymbol{\Theta} := \boldsymbol{X}\boldsymbol{B}$.

g^t explicitly uses prior info about change points $\boldsymbol{\eta} \sim p_{\bar{\boldsymbol{\eta}}}$

$$\boldsymbol{\Theta}^t = \boldsymbol{X}\hat{\boldsymbol{B}}^t - \boldsymbol{R}^{t-1}\left(\boldsymbol{F}^t\right)^\top$$

$$\boldsymbol{R}^t = g^t\left(\boldsymbol{\Theta}^t, \boldsymbol{y}\right)$$

produce residual $\boldsymbol{R}^t$

infer change points

$$\boldsymbol{B}^{t+1} = \boldsymbol{X}^\top\boldsymbol{R}^t - \hat{\boldsymbol{B}}^t\left(\boldsymbol{C}^t\right)^\top$$

$$\hat{\boldsymbol{B}}^t = f^t\left(\boldsymbol{B}^t\right)$$

produce signal estimate $\hat{\boldsymbol{B}}^t$

Bayes-optimal

$$f^t\left(\boldsymbol{B}^t\right) = \mathbb{E}\left[\bar{\boldsymbol{B}} \mid \bar{\boldsymbol{B}} + \boldsymbol{G}_{\boldsymbol{B}}^t = \boldsymbol{B}^t\right]$$

Theorem (informal)

$$\boldsymbol{B}^t \overset{\mathbb{P}}{\simeq} \boldsymbol{B} + \boldsymbol{G}_{\boldsymbol{B}}^t \qquad \text{rows of } \boldsymbol{G}_{\boldsymbol{B}}^t \overset{\text{iid}}{\sim} N(\boldsymbol{0}, \boldsymbol{\kappa}_{\boldsymbol{B}}^t)$$

$$\boldsymbol{\Theta}^t \overset{\mathbb{P}}{\simeq} \boldsymbol{\Theta} + \boldsymbol{G}_{\boldsymbol{\Theta}}^t \qquad \text{rows of } \boldsymbol{G}_{\boldsymbol{\Theta}}^t \overset{\text{iid}}{\sim} N(\boldsymbol{0}, \boldsymbol{\kappa}_{\boldsymbol{\Theta}}^t)$$

AMP for change point inference

Signal is $\boldsymbol{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \boldsymbol{\beta}^{(1)} & \dots & \boldsymbol{\beta}^{(L)} \\ \downarrow & & \downarrow \end{bmatrix} \sim p_{\bar{\boldsymbol{B}}}$ and define $\boldsymbol{\Theta} := \boldsymbol{X}\boldsymbol{B}$.

g^t explicitly uses prior info about change points $\boldsymbol{\eta} \sim p_{\bar{\boldsymbol{\eta}}}$

$$\boxed{\begin{aligned} \boldsymbol{\Theta}^t &= \boldsymbol{X}\hat{\boldsymbol{B}}^t - \boldsymbol{R}^{t-1}\left(\boldsymbol{F}^t\right)^\top \\ \boldsymbol{R}^t &= g^t\left(\boldsymbol{\Theta}^t, \boldsymbol{y}\right) \end{aligned}} \qquad \boxed{\begin{aligned} \boldsymbol{B}^{t+1} &= \boldsymbol{X}^\top \boldsymbol{R}^t - \hat{\boldsymbol{B}}^t\left(\boldsymbol{C}^t\right)^\top \\ \hat{\boldsymbol{B}}^t &= f^t\left(\boldsymbol{B}^t\right) \end{aligned}}$$

produce residual $\boldsymbol{R}^t$

infer change points

produce signal estimate $\hat{\boldsymbol{B}}^t$

Proposition (informal) (Hausdorff distance d_H)

$$d_H\left(\boldsymbol{\eta}, \hat{\boldsymbol{\eta}}(\boldsymbol{\Theta}^t, \boldsymbol{y})\right)/n \overset{\mathbb{P}}{\simeq} \mathbb{E}_{\boldsymbol{V}^t_{\boldsymbol{\Theta}}, \bar{\boldsymbol{\Theta}}}\left[d_H\left(\boldsymbol{\eta}, \hat{\boldsymbol{\eta}}(\bar{\boldsymbol{\Theta}}\boldsymbol{\nu}^t_{\boldsymbol{\Theta}} + \boldsymbol{G}^t_{\boldsymbol{\Theta}}, \bar{\boldsymbol{y}}(\bar{\boldsymbol{\Theta}}, \boldsymbol{\eta}))\right)/n\right]$$

AMP for change point inference

Signal is $\boldsymbol{B} = \begin{bmatrix} \uparrow & & \uparrow \\ \beta^{(1)} & \dots & \beta^{(L)} \\ \downarrow & & \downarrow \end{bmatrix} \sim p_{\bar{B}}$ and define $\boldsymbol{\Theta} := X\boldsymbol{B}$.

g^t explicitly uses prior info about change points $\boldsymbol{\eta} \sim p_{\bar{\eta}}$

$$\boxed{\begin{aligned} \boldsymbol{\Theta}^t &= X\hat{\boldsymbol{B}}^t - \boldsymbol{R}^{t-1}\left(\boldsymbol{F}^t\right)^\top \\ \boldsymbol{R}^t &= g^t\left(\boldsymbol{\Theta}^t, \boldsymbol{y}\right) \end{aligned}}$$

produce residual $\boldsymbol{R}^t$

infer change points

$$\boxed{\begin{aligned} \boldsymbol{B}^{t+1} &= X^\top \boldsymbol{R}^t - \hat{\boldsymbol{B}}^t\left(\boldsymbol{C}^t\right)^\top \\ \hat{\boldsymbol{B}}^t &= f^t\left(\boldsymbol{B}^t\right) \end{aligned}}$$

produce signal estimate $\hat{\boldsymbol{B}}^t$

Proposition (informal) (Hausdorff distance d_H)

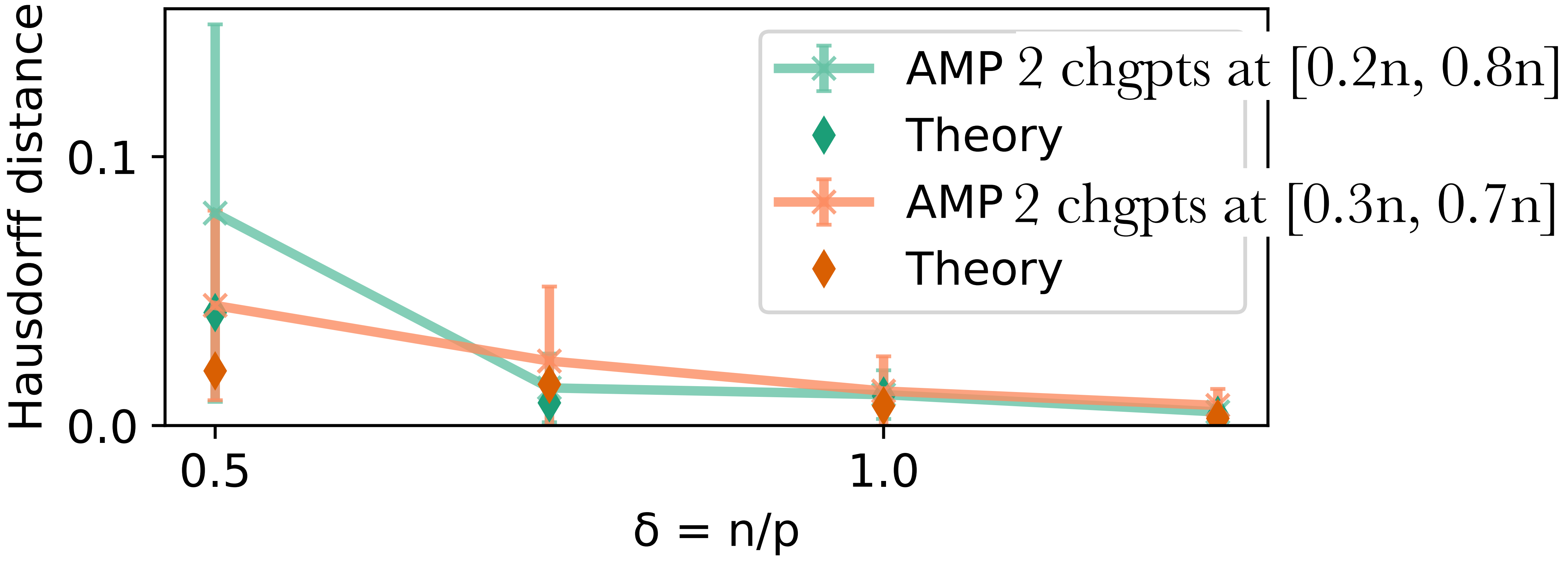
$$d_H\left(\boldsymbol{\eta}, \hat{\boldsymbol{\eta}}(\boldsymbol{\Theta}^t, \boldsymbol{y})\right)/n \overset{\mathbb{P}}{\simeq} \mathbb{E}_{\boldsymbol{V}^t_{\boldsymbol{\Theta}}, \bar{\boldsymbol{\Theta}}}\left[d_H\left(\boldsymbol{\eta}, \hat{\boldsymbol{\eta}}(\bar{\boldsymbol{\Theta}}\boldsymbol{\nu}^t_{\boldsymbol{\Theta}} + \boldsymbol{G}^t_{\boldsymbol{\Theta}}, \bar{\boldsymbol{y}}(\bar{\boldsymbol{\Theta}}, \boldsymbol{\eta}))\right)/n\right]$$

Similar result in terms of posterior $p(\boldsymbol{\eta} \,|\, \boldsymbol{\Theta}^t, \boldsymbol{y})$

Simulation results

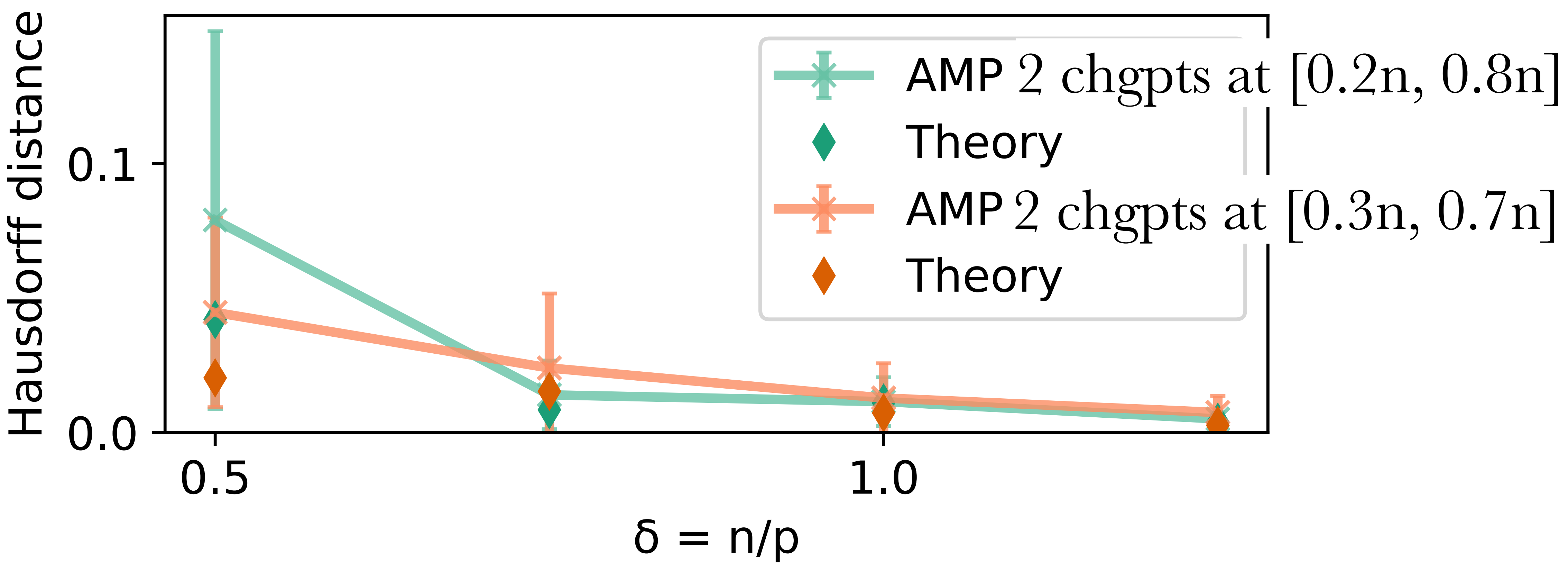
Simulation results

Hausdorff distance

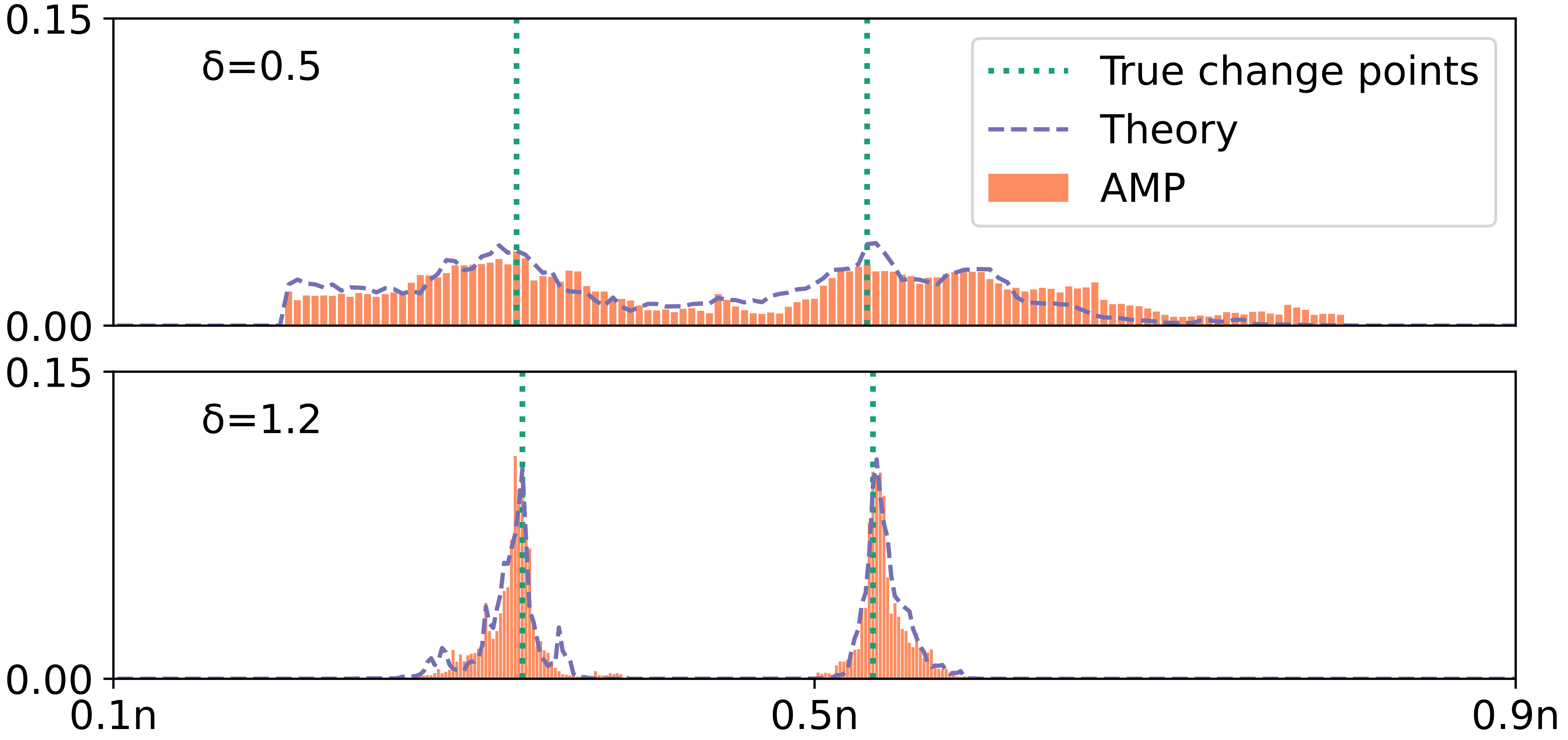


Simulation results

Hausdorff distance

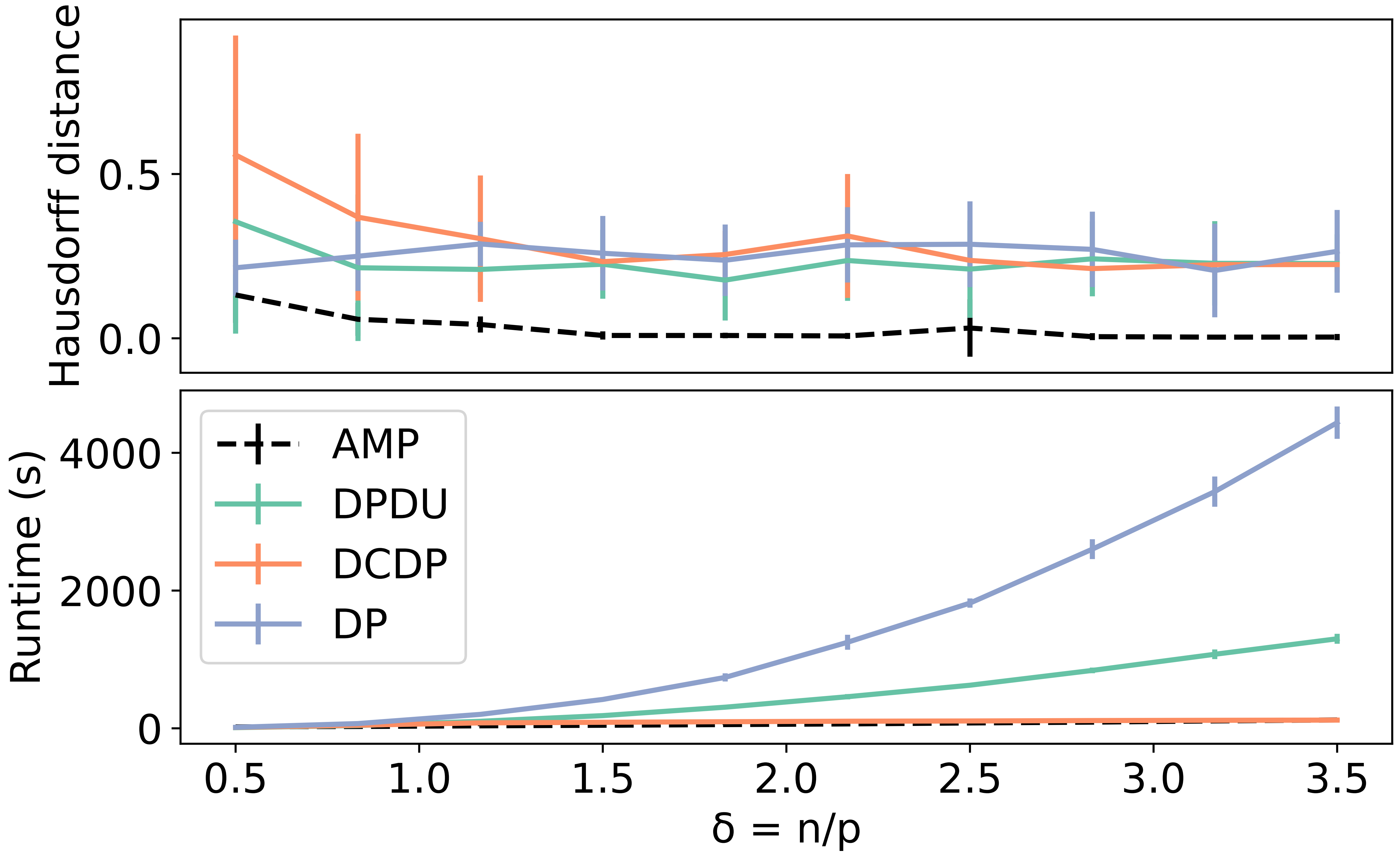


Posterior

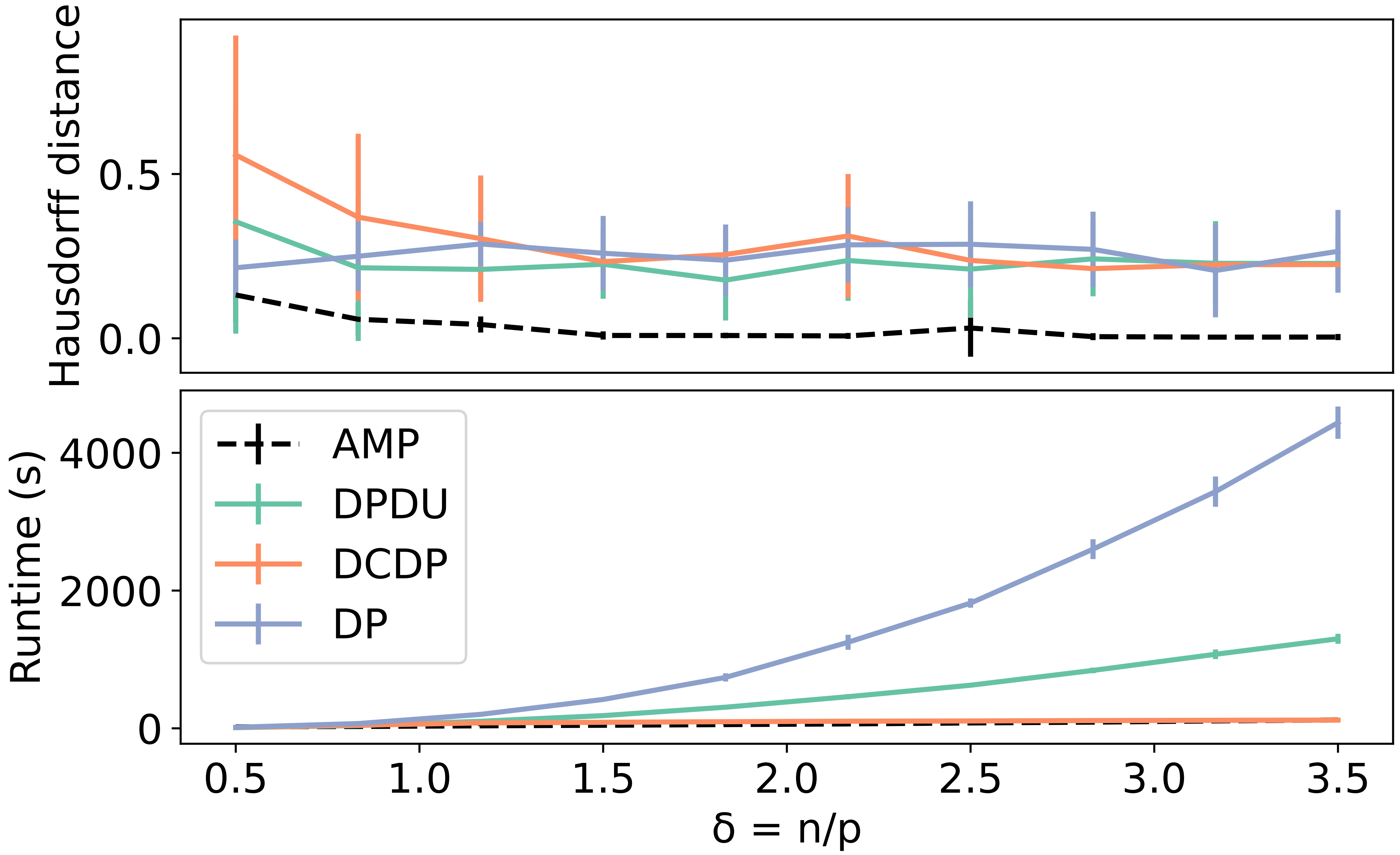


Comparison with other algorithms

Comparison with other algorithms



Comparison with other algorithms



Computational cost: LASSO+partitioning: $O(n^2p^2)$, $O(n^3)$

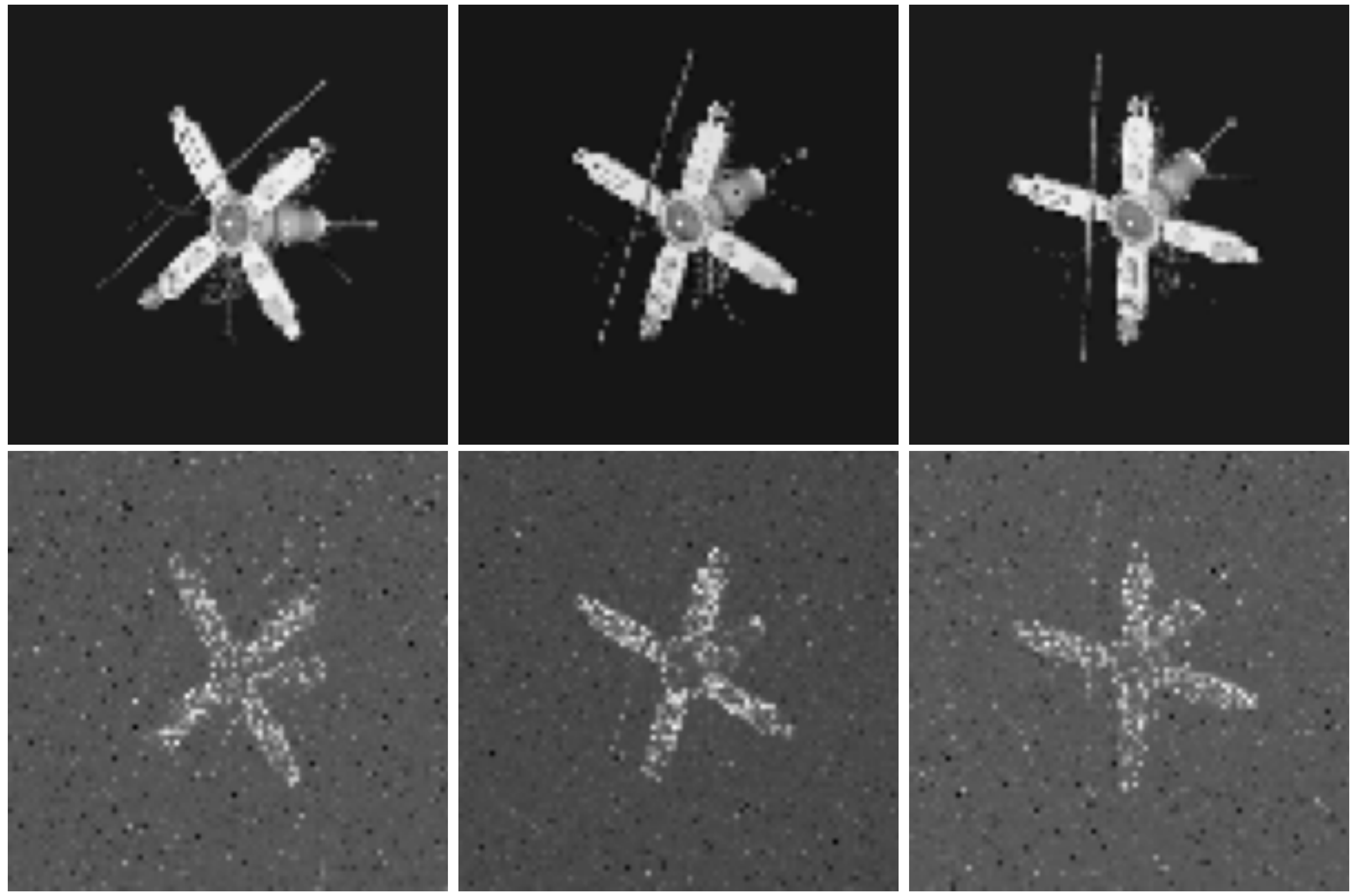
AMP: $O(np)$

Experiments on real images

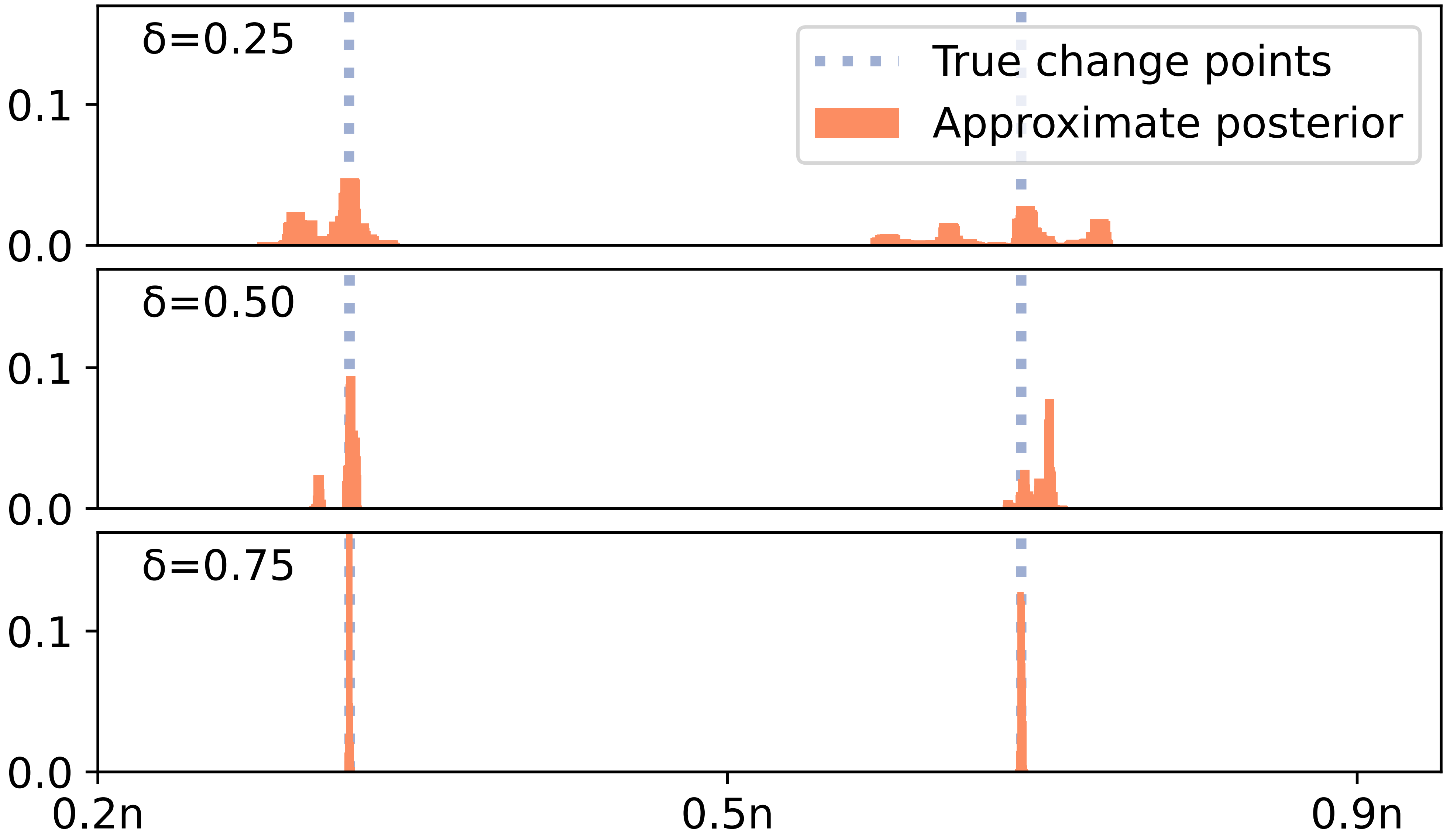
$\boldsymbol{\beta}^{(1)}$	$\boldsymbol{\beta}^{(2)}$	$\boldsymbol{\beta}^{(3)}$
Rotated 0⁰	Rotated 30⁰	Rotated 45⁰

Ground truth signals:

Estimated signals after 10 AMP iterations:



Posterior:



Summary

Summary

Summary

First work to apply AMP algorithms to change point inference:

$(n, p \to \infty,\ n/p \to \delta)$

Summary

First work to apply AMP algorithms to change point inference:

$(n, p \to \infty,\, n/p \to \delta)$

- Bayesian approach to choosing denoisers g^t, f^t
- Near-optimal computational complexity: $O(np)$
- Exact asymptotic performance guarantees (e.g., Hausdorff distance, posterior)

Summary

First work to apply AMP algorithms to change point inference:

$(n, p \to \infty,\ n/p \to \delta)$

- Bayesian approach to choosing denoisers g^t, f^t
- Near-optimal computational complexity: $O(np)$
- Exact asymptotic performance guarantees (e.g., Hausdorff distance, posterior)

Future work:

generalised linear models,

beyond iid Gaussian $\boldsymbol{X}$ matrices,

online change point inference…