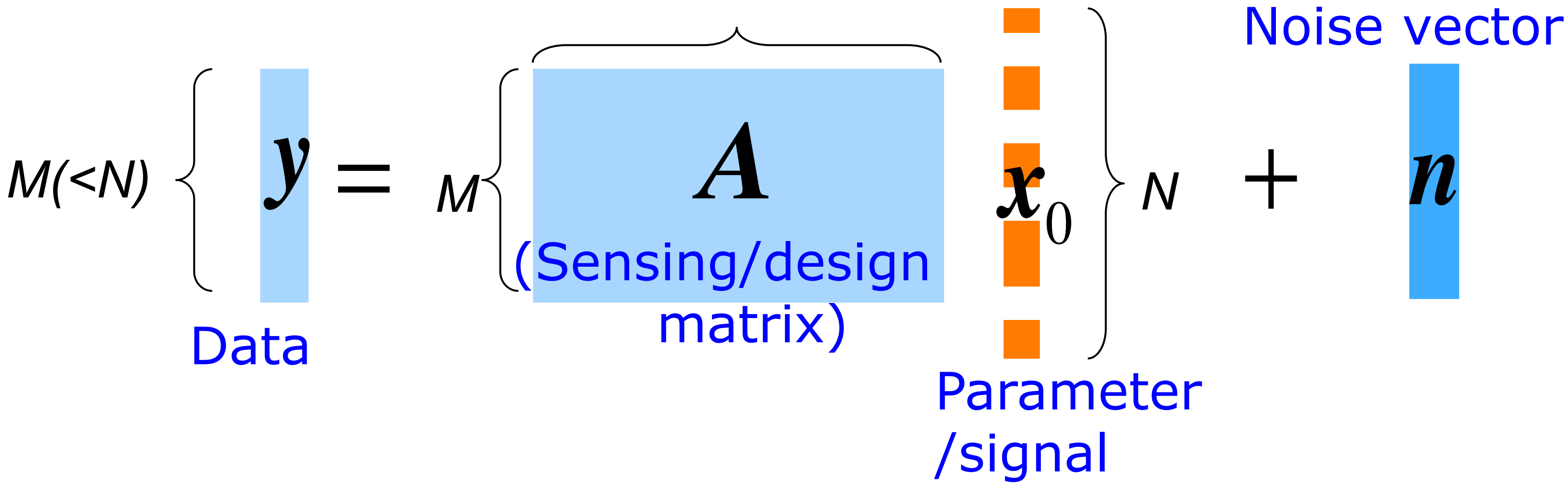
A statistical mechanics approach to de-biasing and uncertainty estimation in LASSO

Takashi Takahashi and YK
Tokyo Institute of Technology,
Japan

Outline

- Motivation and setup
- Statistical mechanics approach
  - Replica analysis
  - Cavity/TAP approach
- Experimental validation
- Summary

Underdetermined sparse regression



- Basic example of high-dimensional statistics
- Underdetermined as $M<N$
  - Unique estimator can not be determined by minimizing $\mathcal{L} = \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|_2^2$ squared loss.
  - There are uncountably many solutions...

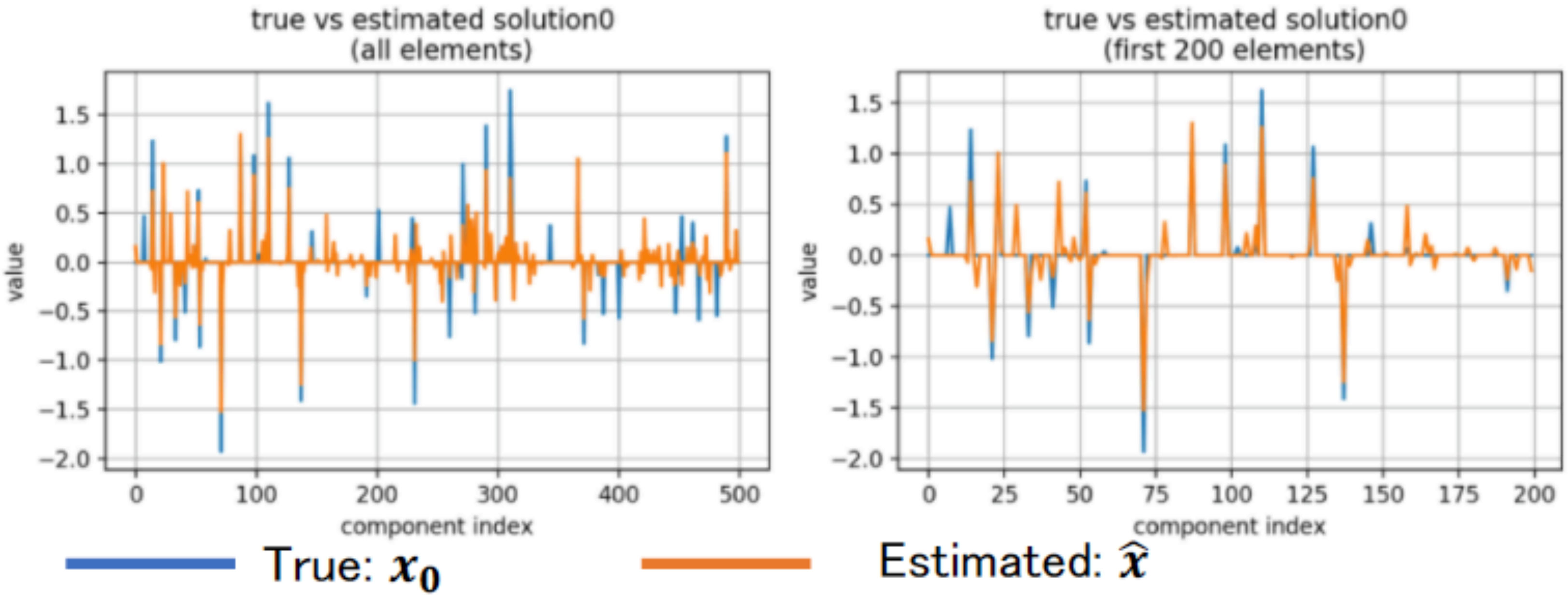
LASSO/L1-norm regularized estimator

- A popular solution:
  Least absolute shrinkage and selection operator (LASSO)

$$\hat{\boldsymbol{x}}^{\mathrm{LASSO}}(\boldsymbol{y}, A; \lambda) = \underset{\boldsymbol{x}}{\arg\min} \left\{ \underbrace{\frac{1}{2} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2}_{\text{loss}} + \lambda \underbrace{\|\boldsymbol{x}\|_1}_{\text{regularizer}} \right\}$$

- Formulated as a convex optimization problem
- An estimator is given as a numerical solution
- Obtained estimator is sparse
  - LASSO performs on
    regression + *variable selection*

Example of LASSO solution



■ <u>Parameters</u>

$$A_{\mu i} \sim \text{i.i.d. } \mathcal{N}\left(0, N^{-1/2}\right),\ x_{0,i} \sim \text{i.i.d. } 0.9\delta(x) + 0.1\mathcal{N}(0,1),\ n \sim \text{i.i.d. } \mathcal{N}(0, 0.05)$$

$$N = 500,\ \alpha = M/N = 0.5$$

Good properties of LASSO

1. Easy to perform

$$Q_1 : \min_{\boldsymbol{x}} \left\{ \frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 \right\}$$

is a convex optimization problem

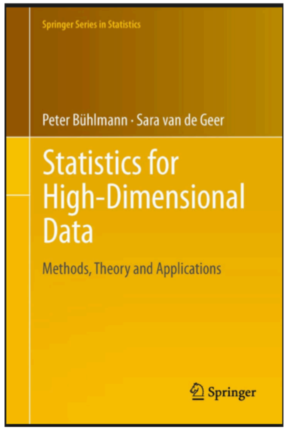
2. Consistent in ideal settings

$$\boxed{\|\boldsymbol{x}_0 - \hat{\boldsymbol{x}}\|_2 \leq \frac{C s_0 \sigma^2}{M} \log N}$$

with high probability

when

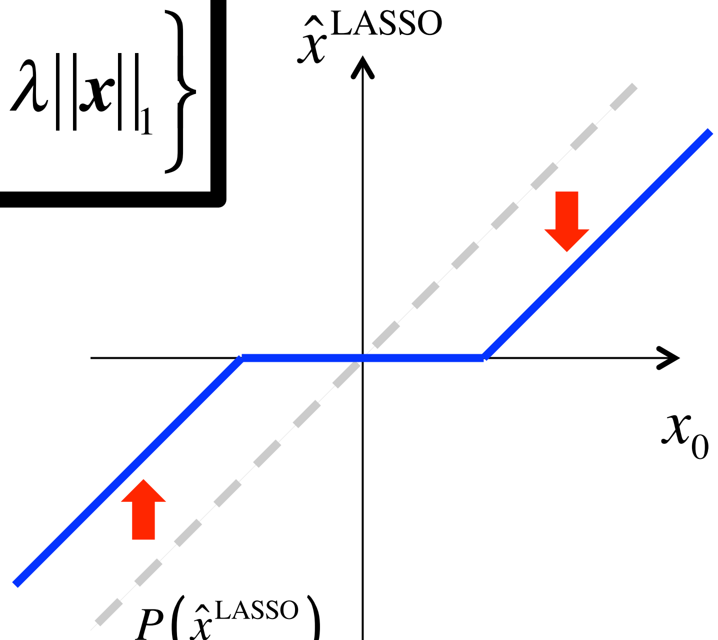
- $\boldsymbol{x}_0$ is s_0-sparse ($\|\boldsymbol{x}_0\|_0 = s_0$)
- A satisfies certain “good” properties



See “Statistics for High-Dimensional Data” (Bulmann and van de Geer) for details

Unsatisfactory properties of LASSO

$$\boxed{\hat{\boldsymbol{x}}^{\mathrm{LASSO}}(\boldsymbol{y}, A; \lambda) = \arg\min_{\boldsymbol{x}} \left\{ \frac{1}{2} \|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda \|\boldsymbol{x}\|_1 \right\}}$$



1. <u>Biased</u>

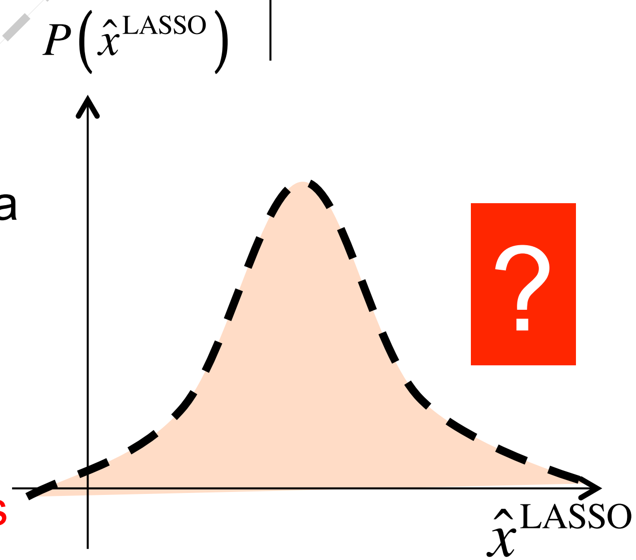
$$\left| \mathbb{E}\left[\hat{x}_i^{\mathrm{LASSO}} \right]_{A,\xi} - x_{o,i} \right| > 0 \text{ for } \lambda > 0$$

2. <u>Dist. of $\hat{x}^{\mathrm{LASSO}}$ is unknown</u>

$\hat{\boldsymbol{x}}^{\mathrm{LASSO}}(\boldsymbol{y}, A; \lambda)$: No analytical expression
Complicated dependence on data

We cannot evaluate its confidence interval

We cannot quantify the reliability of obtained results



Aim of this talk: Post-processing of LASSO

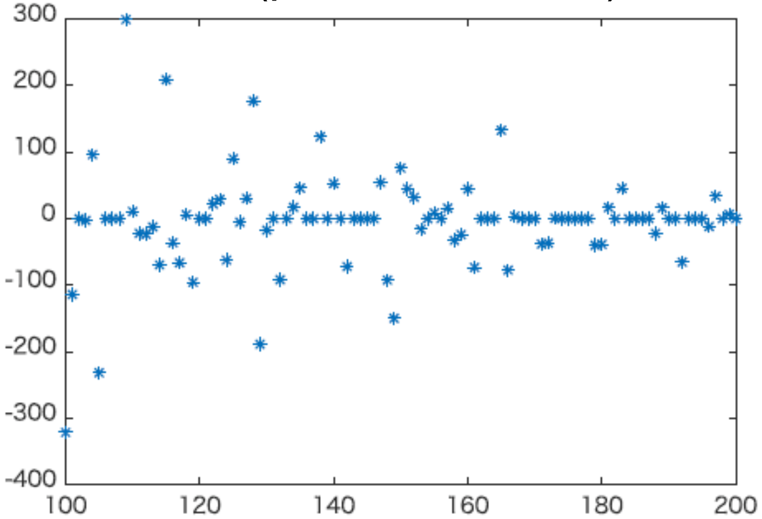
We aim to construct an estimator $\hat{x}^{\mathrm{debiased}}$ that has the following properties from the LASSO solution.

1. Unbiased

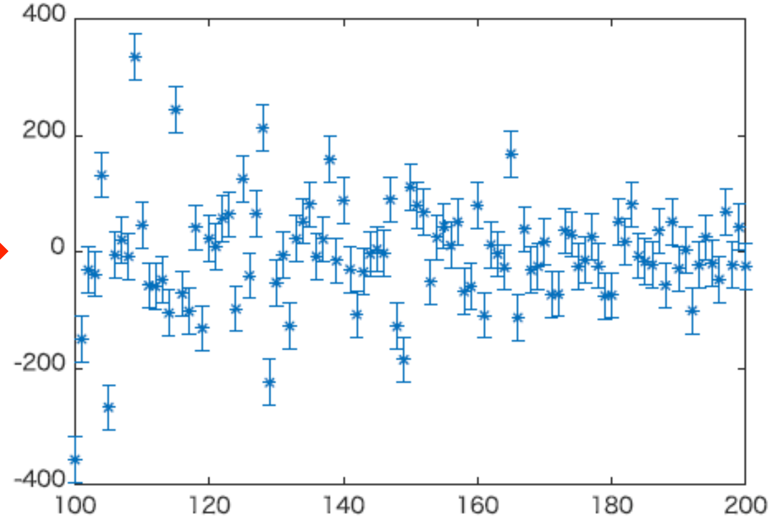
$$\left| \mathbb{E}\left[\hat{x}_i^{\mathrm{debiased}} \right]_{A,\boldsymbol{n}} - x_{o,i} \right| = 0 \text{ for } \forall \lambda > 0$$

2. Confidence intervals are available

$\hat{x}^{\mathrm{LASSO}}$ (point estimation)



$\hat{x}^{\mathrm{debiased}}$ (interval estimation)

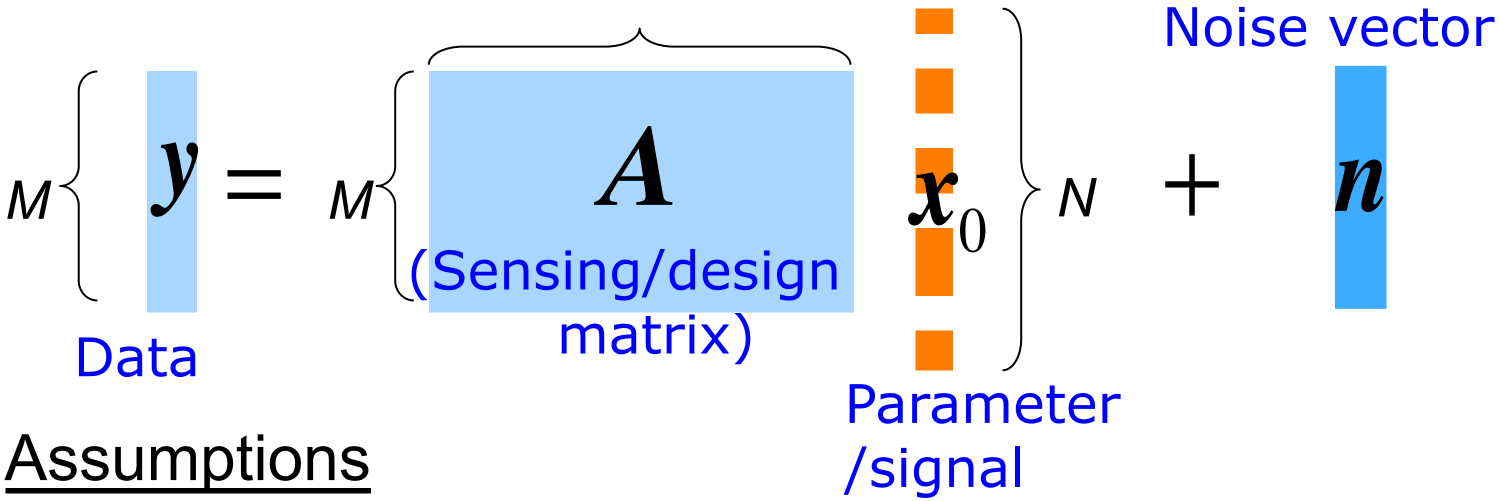


Related work

- Javanmard and Montanari (2014)
- van de Geer et al (2014)
- Zhang and Zhang (2014)
  - Fixed $\boldsymbol{x}_0$
  - Fixed *A*
  - Very sparse signals $\rho = \|\boldsymbol{x}_0\|_0 / N \to 0$
  - Mathematically rigorous

- Ours
  - Fixed $\boldsymbol{x}_0$
  - Random *A*
  - Finite sparsity $\rho = \|\boldsymbol{x}_0\|_0 / N = O(1)$
  - Careful review of the existing stat. mech. results

Setup



- **Assumptions**
  - $\boldsymbol{n} \sim \mathcal{N}\left(0_M, \sigma^2 I_M\right)$
  - $\boldsymbol{x}_0 : \rho N - \text{sparse}\ \left(\|\boldsymbol{x}_0\|_0 = N\rho,\ \rho = O(1);\ \|\boldsymbol{x}_0\|_2^2 = N\rho\sigma_X^2\right)$
  - A : random matrix satisfying the following property

    $A^\top A = ODO^\top$: eigenvalue decomposition

    $\begin{cases} O \sim \text{uniform dist. of } O(N) \\ D_{ii} \sim \text{a fixed eigenvalue dist.} \quad \mu_{A^\top A}(s) \end{cases}$ ← Can be relaxed slightly

Statistical mechanics approach

- LASSO cost = Hamiltonian

$$\frac{1}{2}\|\boldsymbol{y}-A\boldsymbol{x}\|_2^2+\lambda\|\boldsymbol{x}\|_1=H\left(\boldsymbol{x}\middle|\boldsymbol{y},A;\lambda\right)$$

- Gibbs-Boltzmann distribution

$$p_\beta\left(\boldsymbol{x}\middle|\boldsymbol{y},A;\lambda\right)=\frac{e^{-\beta H\left(\boldsymbol{x}\middle|\boldsymbol{y},A;\lambda\right)}}{Z_\beta\left(\boldsymbol{y},A;\lambda\right)}$$

- LASSO solution = Mean in $\beta\to\infty$

$$\begin{aligned}\hat{\boldsymbol{x}}^{\mathrm{LASSO}}\left(\boldsymbol{y},A;\lambda\right)&=\arg\min_{\boldsymbol{x}}\left\{\frac{1}{2}\|\boldsymbol{y}-A\boldsymbol{x}\|_2^2+\lambda\|\boldsymbol{x}\|_1\right\}\\&=\lim_{\beta\to+\infty}\int\boldsymbol{x}p_\beta\left(\boldsymbol{x}\middle|\boldsymbol{y},A;\lambda\right)d\boldsymbol{x}\end{aligned}$$

One can analyze $\hat{\boldsymbol{x}}^{\mathrm{LASSO}}$ using standard methods of stat. mech.

Replica + RMT analysis

- Replica symmetric free energy keeping $\boldsymbol{x}_0$ fixed
  - Parisi and Potters (1995), Takeda et al (2006), ⋯

$$\left(\alpha \triangleq \frac{M}{N},\ Dz \triangleq \frac{e^{-\frac{z^2}{2}}}{\sqrt{2\pi}} \right)$$

$$f = -\lim_{\beta\to\infty}\lim_{n\to 0}\lim_{N\to\infty}\frac{1}{\beta}\frac{\partial}{\partial n}\ln \mathbb{E}\left[\left(Z_\beta(\boldsymbol{y},A;\lambda)\right)^n\right]_{A,\boldsymbol{n}}$$

$$= \mathop{\mathrm{extr}}_{\chi,\hat{\chi},Q,\hat{Q},m,\hat{m}} \left\{ \begin{array}{l} -G'\left(-\chi; A^\top A\right)\left(Q - 2m + \rho\sigma_X^2 - \chi\sigma^2\right) - \frac{\alpha}{2}\sigma^2 + \frac{\hat{Q}Q}{2} - \frac{\hat{\chi}\chi}{2} - \hat{m}m \\ + \lim_{N\to\infty}\sum_{i\sim=1}^{N}\int \boxed{\min_{x_i}\left\{\frac{\hat{Q}}{2}x_i^2 - \left(\hat{m}x_{o,i} + \sqrt{\hat{\chi}}z_i\right)x_i + \lambda|x_i|\right\}} Dz_i \end{array} \right\}$$

$$\Rightarrow \quad \mathbb{E}\left[\left(\hat{x}_i^{\mathrm{LASSO}}\right)^p\right]_{A,\boldsymbol{n}} = \int \left(\mathop{\arg\min}_{x_i}\left\{\frac{\hat{Q}}{2}x_i^2 - \left(\hat{m}x_{o,i} + \sqrt{\hat{\chi}}z_i\right)x_i + \lambda|x_i|\right\}\right)^p Dz_i$$

$$\boxed{G\left(x; A^\top A\right) \triangleq \mathop{\mathrm{extr}}_{z}\left\{-\frac{1}{2}\int \rho_{A^\top A}(s)\ln(z-s)\,ds + \frac{zx}{2}\right\} - \frac{1}{2}\ln|x| - \frac{1}{2}}$$

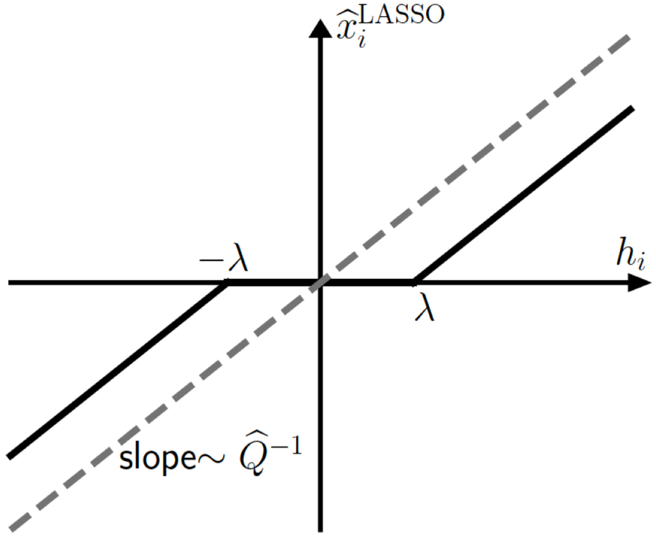
Characterizes the property of the matrix ensemble

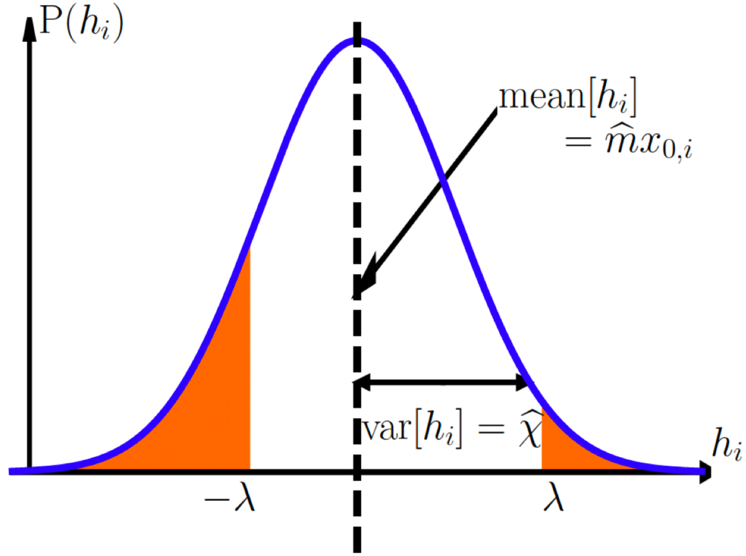
Decoupling principle

- Estimator of each component: Soft-thresholding of an statistically independent Gaussian field

$$\mathbb{E}\left[\left(\hat{x}_i^{\text{LASSO}}\right)^p\right]_{A,\boldsymbol{n}} = \int \left(\underset{x_i}{\arg\min} \left\{ \frac{\hat{Q}}{2} x_i^2 - \underbrace{\left(\hat{m} x_{o,i} + \sqrt{\hat{\chi}} z_i \right)}_{\equiv h_i \sim \mathcal{N}\left(\hat{m}x_{0,i}, \hat{\chi}\right)} x_i + \lambda |x_i| \right\} \right)^p Dz_i$$

$$\boxed{\hat{x}_i^{\text{LASSO}} = \hat{Q}^{-1}\left(h_i - \lambda\,\text{sgn}(h_i)\right)\Theta\left(|h_i| - \lambda\right)}$$





Consideration

- Statistical properties of the estimator are completely determined by 3 macroscopic variables $\hat{Q}, \hat{m}\left(=\hat{Q}\right), \hat{\chi}$.
- If these variables and the Gaussian field h_i are available, one can construct a debiased estimator $\hat{x}_i^{\mathrm{debiased}}$ as follows:

$$
\begin{cases}
\bullet\ h_i \left(= \hat{m} x_{0,i} + \sqrt{\hat{\chi}} z_i\right) \\
\bullet\ \hat{x}_i^{\mathrm{debiased}} = \dfrac{h_i}{\hat{Q}} = \dfrac{\hat{m}}{\hat{Q}} x_{0,i} + \dfrac{\sqrt{\hat{\chi}}}{\hat{Q}} z_i = x_{0,i} + \dfrac{\sqrt{\hat{\chi}}}{\hat{Q}} z_i \sim \mathcal{N}\left(x_{0,i}, \dfrac{\hat{\chi}}{\hat{Q}^2}\right) \\
\bullet\ 100\left(1-\alpha_{\mathrm{sig}}\right)\%\ \mathrm{CI}: \left[\hat{x}_i^{\mathrm{debiased}} - \hat{Q}^{-1}\sqrt{\hat{\chi}}\, u\left(1-\dfrac{\alpha_{\mathrm{sig}}}{2}\right), \hat{x}_i^{\mathrm{debiased}} + \hat{Q}^{-1}\sqrt{\hat{\chi}}\, u\left(1-\dfrac{\alpha_{\mathrm{sig}}}{2}\right)\right]
\end{cases}
$$

How can we know $\hat{Q}, \hat{m}\left(=\hat{Q}\right), \hat{\chi}$, and $\{h_i\}$ from a single sample of data? → ***Cavity/TAP approach***

Cavity/TAP approach

- Stat. mech. evaluation of the LASSO estimator for a single sample of $\boldsymbol{y}, A$

Gibbs free energy

$$\Phi\left(\boldsymbol{m}\middle|\,\boldsymbol{y},A;\lambda\right)=\max_{\boldsymbol{h}}\left[\boldsymbol{h}\cdot\boldsymbol{m}-\lim_{\beta\to+\infty}\frac{1}{\beta}\ln\int e^{-\beta H\left(\boldsymbol{x}|\boldsymbol{y},A;\lambda\right)+\beta\boldsymbol{h}\cdot\boldsymbol{x}}\,d\boldsymbol{x}\right]$$

$$=\max_{\boldsymbol{h}}\left[\boldsymbol{h}\cdot\boldsymbol{m}+\min_{\boldsymbol{x}}\left[\frac{1}{2}\left\|\boldsymbol{y}-A\boldsymbol{x}\right\|_2^2+\lambda\left\|\boldsymbol{x}\right\|_1-\boldsymbol{h}\cdot\boldsymbol{x}\right]\right]$$

$$\left(=\frac{1}{2}\left\|\boldsymbol{y}-A\boldsymbol{m}\right\|_2^2+\lambda\left\|\boldsymbol{m}\right\|_1\right)$$

$$\hat{x}^{\mathrm{LASSO}}\left(\boldsymbol{y},A;\lambda\right)=\arg\min_{\boldsymbol{m}}\left\{\Phi\left(\boldsymbol{m}\middle|\,\boldsymbol{y},A;\lambda\right)\right\}$$

Cavity/TAP approach

- Adaptive TAP/EC approach
  - Opper and Winther (2001,2006), YK and Vehkapera (2014), ···

Generalized free energy

$$\tilde{\Phi}(\boldsymbol{m}|\boldsymbol{y},A;\lambda,l)=\max_{\boldsymbol{h}}\left[\boldsymbol{h}\cdot\boldsymbol{m}+\min_{\boldsymbol{x}}\left[\frac{l}{2}\|\boldsymbol{y}-A\boldsymbol{x}\|_2^2+\lambda\|\boldsymbol{x}\|_1-\boldsymbol{h}\cdot\boldsymbol{x}\right]\right]$$

$$\Phi=\tilde{\Phi}(l=1)=\int_0^1\frac{\partial\tilde{\Phi}(l)}{\partial l}dl+\tilde{\Phi}(l=0)$$

$$\simeq\max_{\boldsymbol{h}}\left[\frac{1}{2}\|\boldsymbol{y}-A\boldsymbol{m}\|_2^2-\frac{\Lambda\|\boldsymbol{m}\|_2^2}{2}+\boldsymbol{h}\cdot\boldsymbol{m}-\frac{1}{2\Lambda}\sum_{i=1}^N(|h_i|-\lambda)^2\Theta(|h_i|-\lambda)\right]$$

$$\left(\Lambda=2G'(-\chi;A^\top A),\chi=\frac{1}{N\Lambda}\sum_{i=1}^N\Theta(|h_i|-\lambda)\right)$$

Macroscopic moment matching
+ Gaussian approximation

Statistical property of matrix ensemble is summarized in *G*-func.

Adaptive TAP equation

- Extremum condition of the free energy offers

$$\begin{cases} \boldsymbol{h} = \Lambda \boldsymbol{m} + A^\top (\boldsymbol{y} - A\boldsymbol{m}) \\ m_i = \dfrac{h_i - \lambda \operatorname{sgn}(h_i)}{\Lambda} \Theta(|h_i| - \lambda) \\ \Lambda = 2G'(-\chi; A^\top A) \\ \chi = \dfrac{1}{N\Lambda} \displaystyle\sum_{i=1}^{N} \Theta(|h_i| - \lambda) \end{cases}$$

Onsager reaction coeff

- We can evaluate the Gaussian fields $\{h_i\}$ from these equations.

But, how can we evaluate macroscopic variables $\hat{Q}, \hat{m}, \hat{\chi}$, which are averaged quantities w.r.t. $A, \boldsymbol{y}$ from a single sample?

Self-averaging property

- For $N \gg 1$, the self-averaging property indicates that the macroscopic variables $\hat{Q}, \hat{m}, \hat{\chi}$ can be evaluated from a <u>single sample of $A, \boldsymbol{y}$</u>.
- Concretely, the replica/cavity equivalence for *N>>1* provides the following correspondence:

Replica (averaged)	Cavity/TAP (single sample)
$\hat{Q} = \hat{m}$	$\Lambda\left(=2G'(-\chi)\right)$
$\hat{\chi}$	$\frac{G''(-\chi)\left\|\boldsymbol{y}-A\hat{\boldsymbol{x}}^{\mathrm{LASSO}}\right\|_2^2}{M\left(G'(-\chi)-\chi G''(-\chi)\right)}+\frac{-G''(-\chi)+2G'(-\chi)}{G'(-\chi)-\chi G''(-\chi)}\sigma^2$

  - We need to estimate the noise variance σ^2 from data.

Main result

1. $\hat{\boldsymbol{x}}^{\mathrm{LASSO}} = \underset{\boldsymbol{x}}{\arg\min}\left[\frac{1}{2}\|\boldsymbol{y} - A\boldsymbol{x}\|_2^2 + \lambda\|\boldsymbol{x}\|\right]$

Any algorithm:
VAMP, LARS, CD, ⋯

2. Solve $\Lambda = 2G'\left(-\chi; A^\top A\right),\ \chi = \frac{1}{\Lambda N}\#\left\{i \middle| x_i^{\mathrm{LASSO}} \neq 0\right\}$

3. $\boldsymbol{h} = \Lambda\hat{\boldsymbol{x}}^{\mathrm{LASSO}} + A^\top\left(\boldsymbol{y} - A\hat{\boldsymbol{x}}^{\mathrm{LASSO}}\right)$

4. $\hat{\chi} = \frac{G''(-\chi)\|\boldsymbol{y} - A\hat{\boldsymbol{x}}^{\mathrm{LASSO}}\|_2^2}{M\left(G'(-\chi) - \chi G''(-\chi)\right)} + \frac{-G''(-\chi) + 2G'(-\chi)}{G'(-\chi) - \chi G''(-\chi)}\sigma^2$

5. $\hat{\boldsymbol{x}}^{\mathrm{debiased}} = \frac{\boldsymbol{h}}{\Lambda} \sim \mathcal{N}\left(\boldsymbol{x}_0, \frac{\hat{\chi}}{\Lambda^2}\right)$

Very fast!

As for the noise variance σ^2, a cross-validation-based estimator (Reid et al (2013)) empirically exhibits good performance.

Experimental set up

We examined the utility of the developed debiasing method by application to the following two matrix ensembles of $M = \alpha N \ (\alpha = 0.5)$.

1. <u>i.i.d. Gaussian ensemble</u>
   - $A_{ij} \sim \mathcal{N}\left(0, N^{-1}\right)$
   - AED: Marchenko-Pastur distribution

$$\mu_{A^\top A}(s) = (1-\alpha)\delta(s) + \frac{\alpha}{2\pi}\frac{\sqrt{\left[\left(1+\sqrt{\alpha}\right)^2 - s\right]\left[s-\left(1-\sqrt{\alpha}\right)^2\right]}}{s}$$

2. <u>Random DCT ensemble</u>
   - Random sampling of $M=\alpha N$ rows from N x N DCT matrix
   - AED: $\mu_{A^\top A}(s) = (1-\alpha)\delta(s) + \alpha\delta(s-1)$
   - <u>Not rotationally invariant!!!</u> But, the developed technique is applicable (Cakmak and Opper (2018))

Experimental set up

Other settings:

- Sparse signal: Bernoulli-Gaussian dist.

$$x_{0,i} \sim_{\text{i.i.d.}} (1-\rho)\delta(x) + \rho\mathcal{N}(0,1)$$

- Noise estimator: CV-based (Reid et al (2013))

$$\hat{\sigma}^2 \equiv \frac{1}{M - \#\left\{i \middle| \hat{x}_i^{\text{LASSO}} \neq 0\right\}} \left\|\boldsymbol{y} - A\hat{\boldsymbol{x}}^{\text{LASSO}}\left(\boldsymbol{y}, A; \hat{\lambda}\right)\right\|_2^2$$

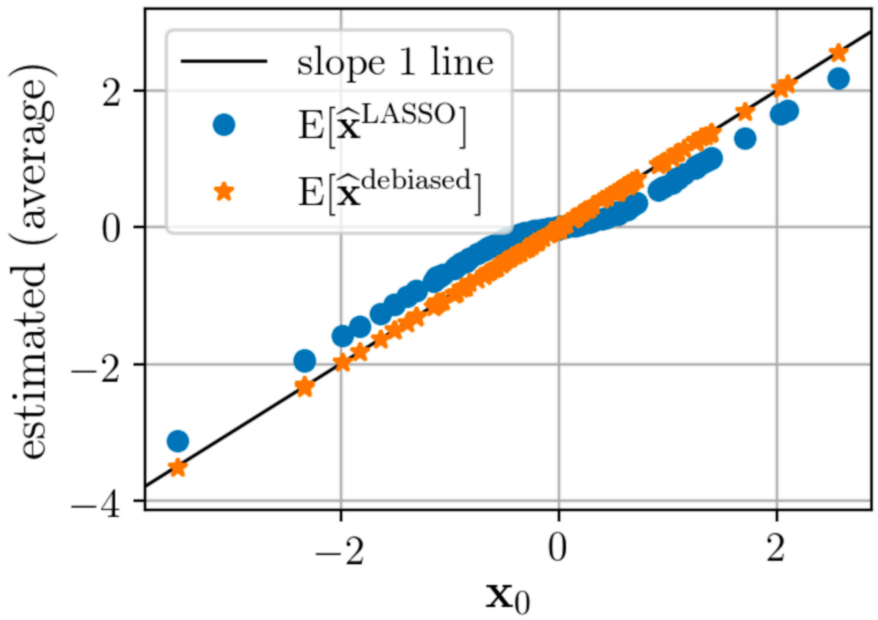
$\hat{\lambda}$: Determined so that CV error is minimized
Approximate CV formula (Obuchi and YK (2016))

- System parameters:

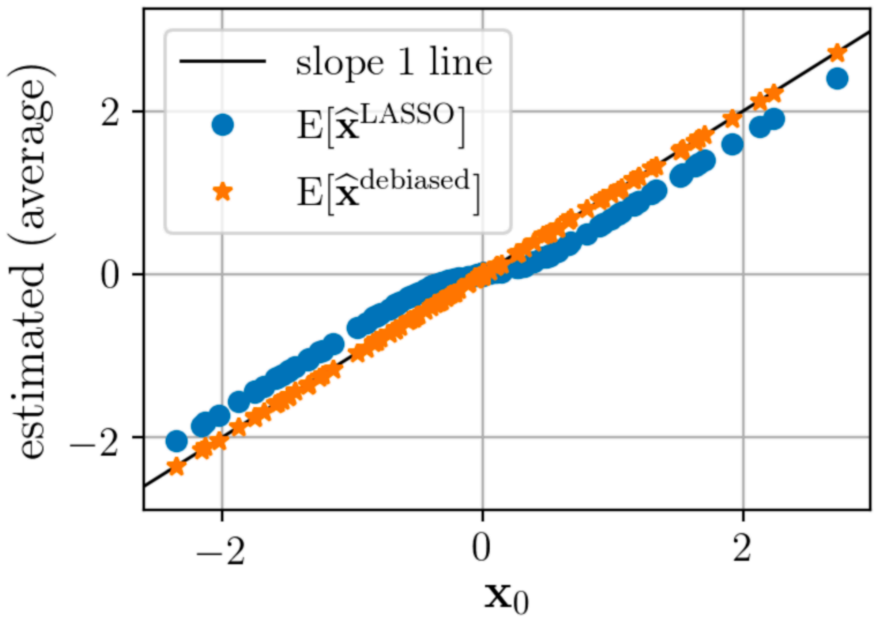
$$N = 1000,\ \alpha = M/N = 0.5,\ \sigma^2 = 0.2,\ \rho = 0.1$$

Debiasing

i.i.d. Gaussian

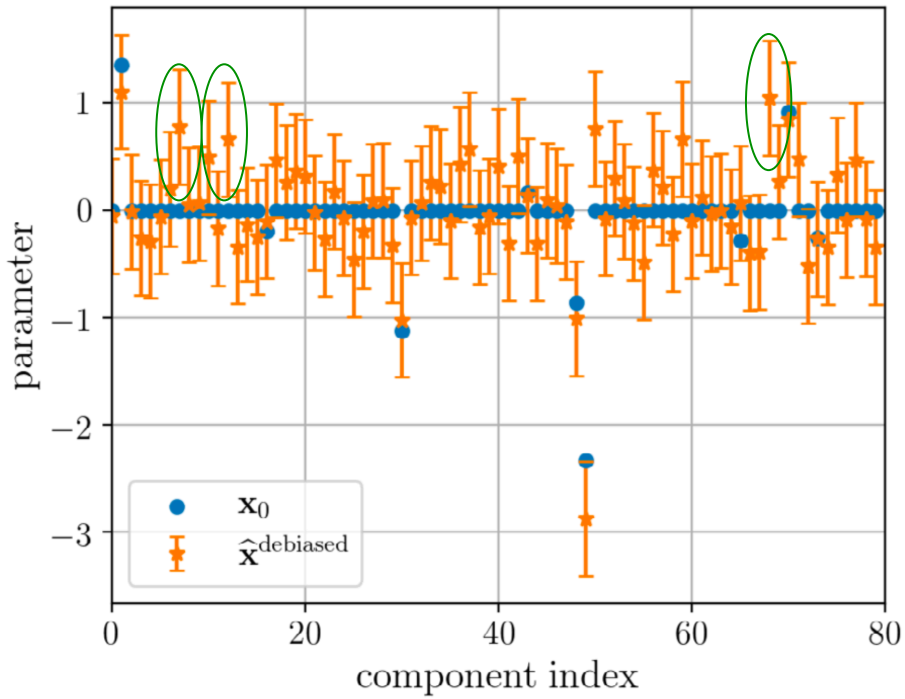


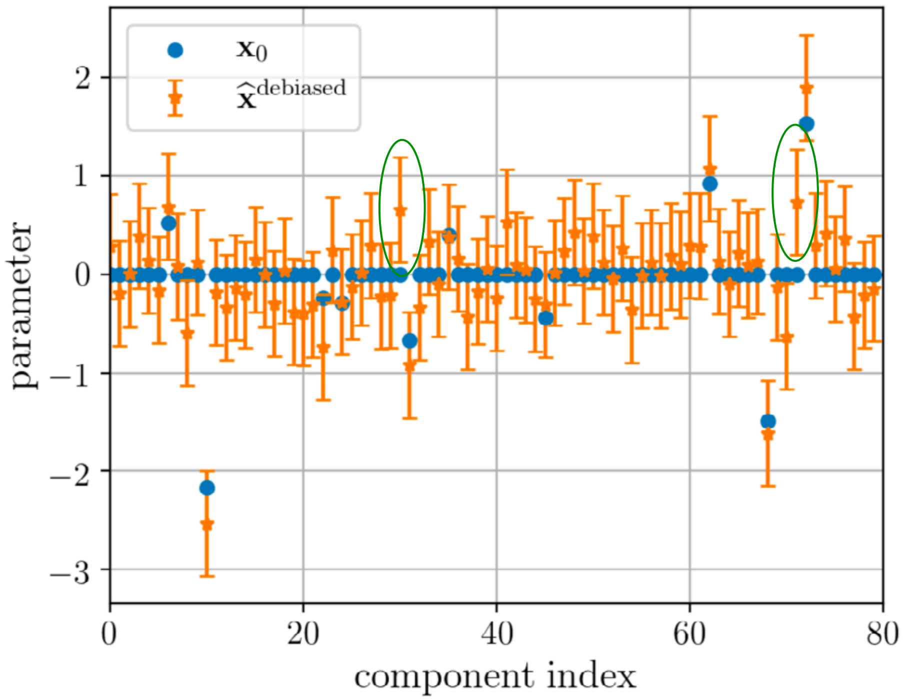
random DCT



Each point represents average over 1000 sets of A and $\boldsymbol{y}$

Confidence interval (CI)





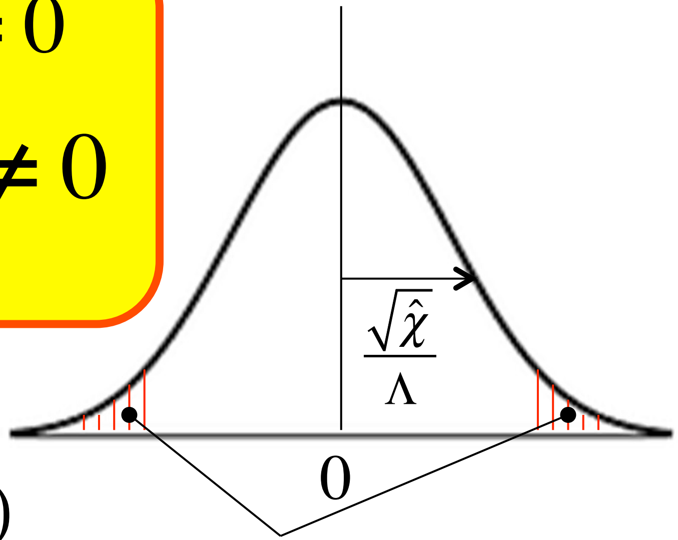
Error bars indicate 95% confidence intervals evaluated from a single data sample $A, \boldsymbol{y}$.
$$\left[\hat{x}_i^{\text{debiased}} - 1.96\frac{\sqrt{\hat{\chi}}}{\Lambda},\ \hat{x}_i^{\text{debiased}} + 1.96\frac{\sqrt{\hat{\chi}}}{\Lambda}\right]$$

The CIs actually cover true parameters $x_{0,i}$ with a probability of about 95%.

Statistical testing

$$\begin{cases} \text{Null hypothesis} & H_{0,i} : x_{0,i} = 0 \\ \text{Alternative hypothesis} & H_{1,i} : x_{0,i} \neq 0 \end{cases}$$

#Relevant for “variable selection”



Rejection region (RR)
Significance level: α_{sig}

Testing procedure

$$\hat{T}_i(\boldsymbol{y}, A; \lambda) = \begin{cases} 1, & \hat{x}_i^{\mathrm{debiased}} \in \mathrm{RR}(\alpha_{\mathrm{sig}}) \ (\text{reject}) \\ 0, & \text{otherwise} \ (\text{accept}) \end{cases}$$

Performance measure

$$\mathrm{FPR} = \frac{\#\{\hat{T}_i = 1 \text{ and } x_{0,i} = 0\}}{\#\{x_{0,i} = 0\}} \equiv \alpha_{\mathrm{sig}}$$

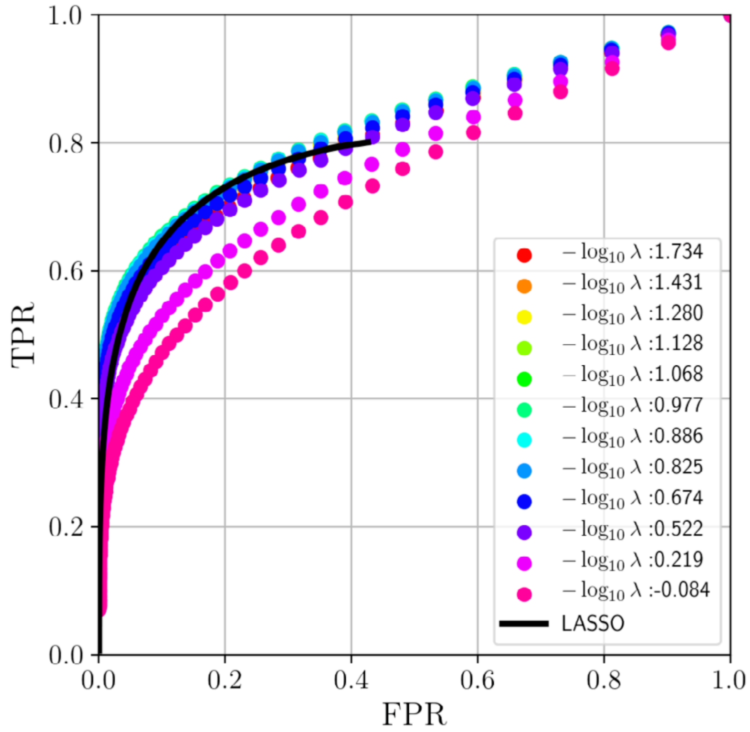
(control parameter)

$$\mathrm{TPR} = \frac{\#\{\hat{T}_i = 1 \text{ and } x_{0,i} = 1\}}{\#\{x_{0,i} = 1\}}$$

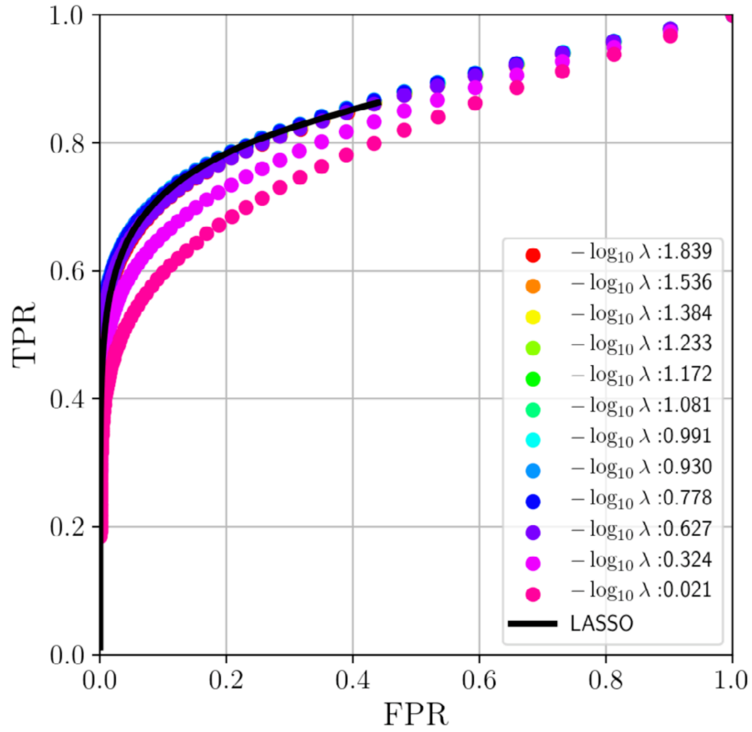
(desired to be maximized)

Statistical testing

i.i.d. Gaussian



random DCT



ROC curve (FPR-TPR curve) is maximized when variance $\dfrac{\hat{\chi}}{\Lambda^2}$ is minimized.

The resulting detection performance is (slightly) better than that achieved by naively employing LASSO varying l_1 strength λ.

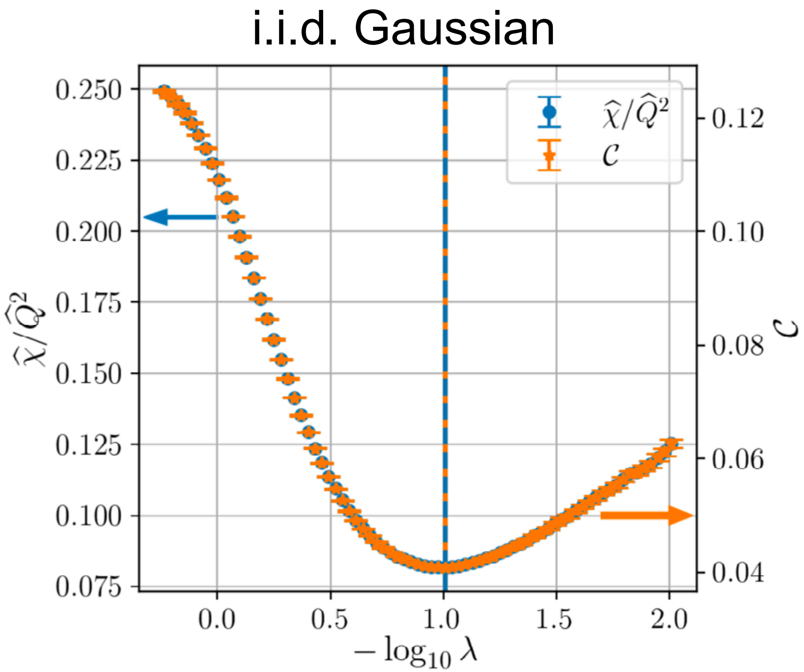
CI size versus CV error

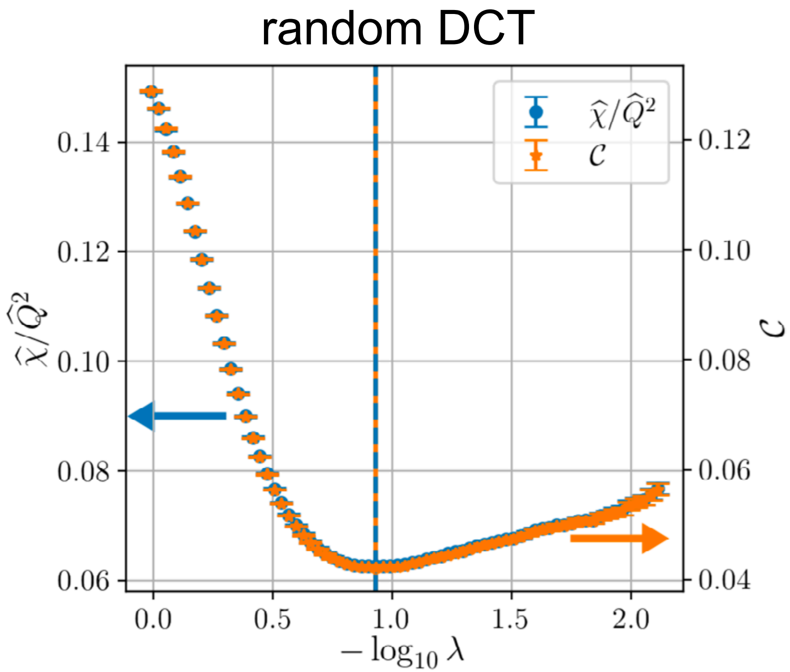
- Minimization of cross-validation (CV) error is one of the standard approaches for determining λ in LASSO.

- On the other hand, maximizing the estimation accuracy by <u>minimizing the size of CI</u> would offer an alternative criterion for selecting λ.

- Which criterion is better?

CI size versus CV error



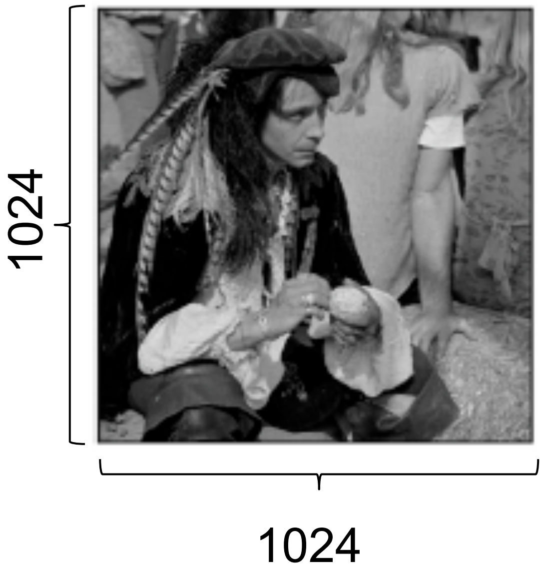


- For i.i.d. Gaussian and random DCT ensembles, CI size and CV error are linearly related, so that the two criteria always lead to the same result!
- The relation between CI size and CV error does not necessarily hold for other ensembles.
- However, as far as we examined, these two criteria are always minimized at the same value of λ when the matrix ensemble is rotationally invariant.

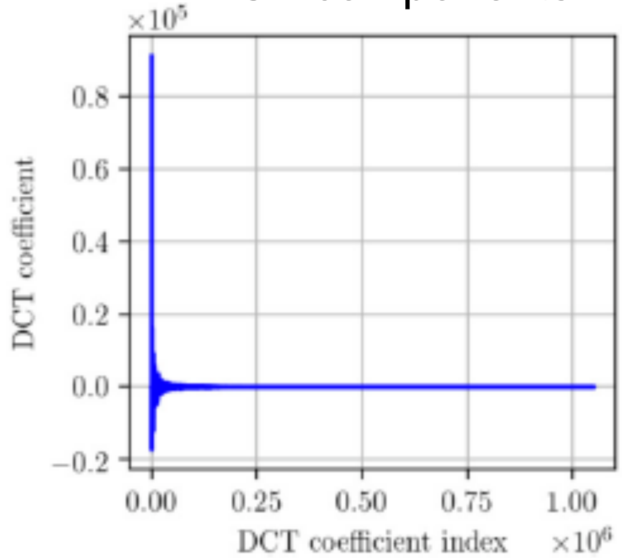
Demonstration on a real world data set

- Task: Estimation of Fourier (DCT) components from sub-sampled real world signals
  - Many demands in spectral analysis
  - Here, we handle a megapixel image for ease of visual understanding

Original



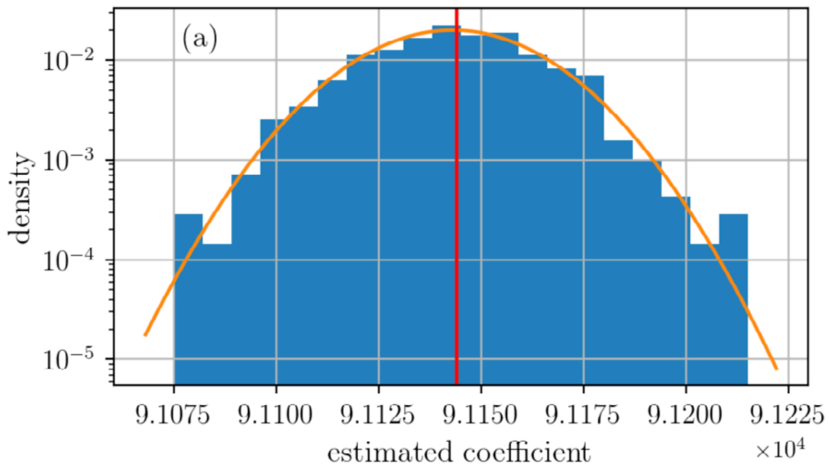
DCT components



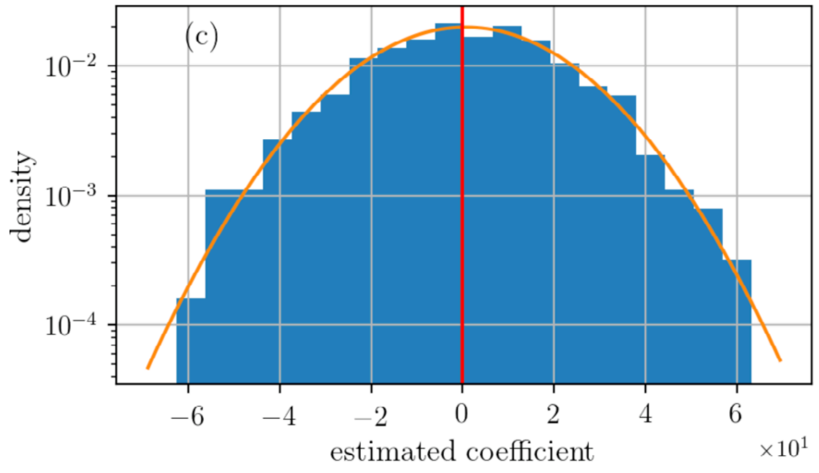
50% subsampled

Demonstration on a real world data set

Largest component



Smallest component



Histograms: evaluated from 1000 samples of random DCTs and noises
Curves: evaluated from a single sample of random DCT and noise

Summary

- We developed a method for de-biasing and uncertainty estimation in LASSO in the case of rotationally invariant design/observation matrices.
  - Earlier studies: fixed matrix, vanishing sparsity ratio
    - Javanmard and Montanari (2014), van de Geer et al (2014), Zhang and Zhang (2014)
  - Ours: random matrix, finite sparsity ratio

- Although we here focused on LASSO, the extension to other regularized estimation is straightforward.

Summary

- Advantages of the method
  - Computationally feasible
  - Ability for constructing confidence interval
  - Better performance for statistical testing (signal detection) than naïve LASSO
- Possible unsatisfaction
  - Limited applicability to a special class of random matrices although random Fourier ensembles may potentially have wide application domains

- Semi-analytic resampling (bootstrap) may be useful for resolving this issue.
  - Obuchi and YK, arXiv:1802.10254 => ``pair bootstrap''
  - Takahashi and YK, in preparation => ``residual bootstrap''

References

- T. Takahashi and YK,
  "A statistical mechanics approach to de-biasing and uncertainty estimation in LASSO for random measurements",
  J. Stat. Mech. (2018) 073405 (open access)
  - Demo:
    https://github.com/takashi-takahashi/debiasing_lasso_demo