Adaptive Sparse Estimation with Side Information

Trambak Banerjee (joint with Gourab Mukherjee & Wenguang Sun)

Data Sciences and Operations, University of Southern California

Information Pooling in High-Dimensional Analysis

- Vast amounts of data with various types of side information being collected nowadays.
- Multiple testing - side information on the hypotheses is often used to improve power of multiple testing procedures.
- In high dimensional estimation problems, such additional information regarding the signal sparsity may yield more accurate results.

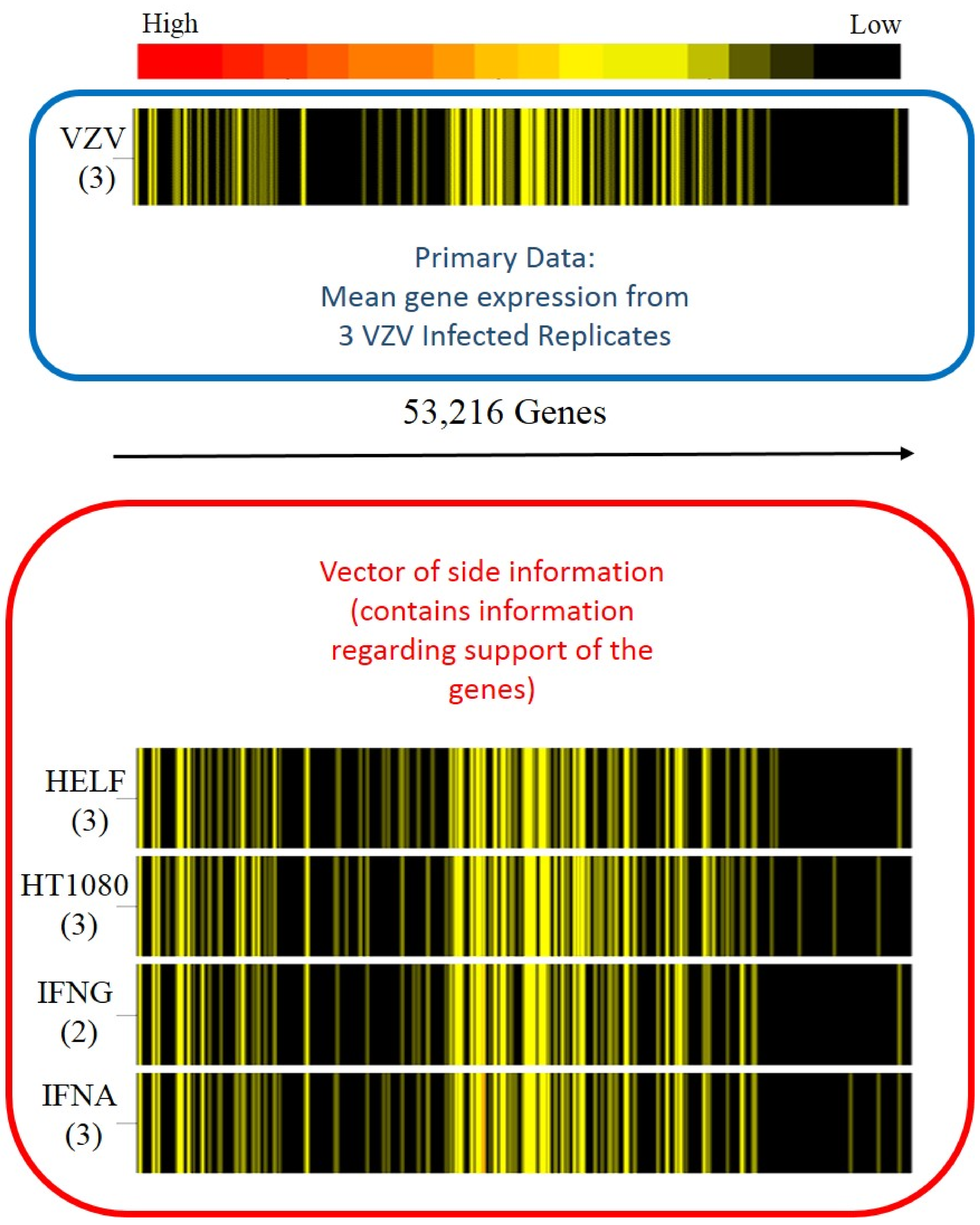
Aim

Construct an estimator that includes side information such that:

- new estimator is adaptive to the strength of side information and robust to it's non-usefulness.
- if the side information is imperfect then the estimator is not too far away from state-of-the-art sparse estimators built on using no side information
- this extra information is used in a model agnostic fashion.

Motivating Data - estimating gene expression

- Primary data - expression level Y_i of $n = 53,216$ genes infected with VZV (varicella) virus.
- Goal - estimate the true expression level θ_i of these n genes.
- Side information - expression levels corresponding to 4 disparate experimental conditions; HELF, HT1080, IFNG and IFNA, for the same n genes.



A Framework for Information Pooling

- ξ - latent noiseless side information encoding the sparsity of $\boldsymbol{\theta}$.
- $\mathbf{S}$ - noisy or observed side information.
- Relate $\boldsymbol{\theta}$ and $\mathbf{S}$ to ξ via unknown real-valued functions h_θ and h_s.

$$\theta_i = h_\theta(\xi_i, \eta_{1i})$$
$$S_i = h_s(\xi_i, \eta_{2i})$$
$$Y_i = \theta_i + \epsilon_i,\ \ \epsilon_i \sim N(0, \sigma_i^2) \text{ with } \sigma_i^2 \text{ known}$$

η_{1i}, η_{2i} - random perturbations independent of ξ_i.

- Flexible framework
- No assumption on any *particular* functional relationship between $\boldsymbol{\theta}$ and ξ
- S_i conditionally independent of Y_i given laten ξ_i

ASUS - Adaptive SURE Thresholding using Side Information

Key idea - construct optimal groups and soft threshold separately in each group.

Let $\mathcal{I} = \{1, \dots, n\}$ and $\mathcal{T} = \{\tau, t_1, t_2\}$. Define

$$\mathcal{I}_1^\tau = \{i : 0 < |S_i| \le \tau\},$$
$$\mathcal{I}_2^\tau = \mathcal{I} \setminus \mathcal{I}_1^\tau$$

Class of soft thresholding estimators:

$$\hat{\theta}_i^{SI}(\mathcal{T}) := Y_i + \sigma_i \eta_{t_k}(Y_i) \text{ if } i \in \mathcal{I}_k^\tau$$

Then the ASUS estimator is given by $\hat{\theta}_i^{SI}(\hat{\mathcal{T}})$ where

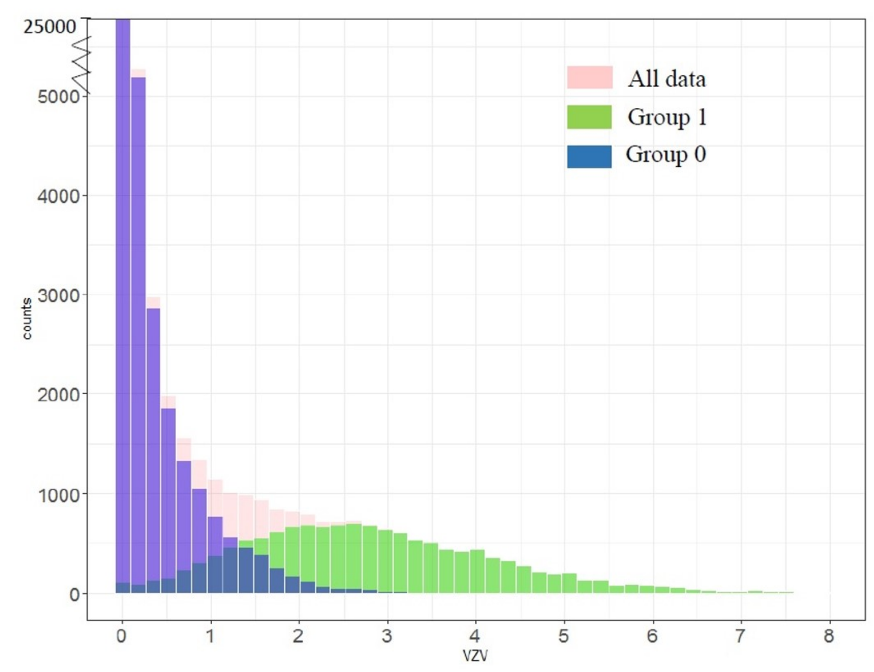
$$\hat{\mathcal{T}} = \arg\min_{\mathcal{T}} S(\mathcal{T}, \mathbf{Y}, \mathbf{S})$$

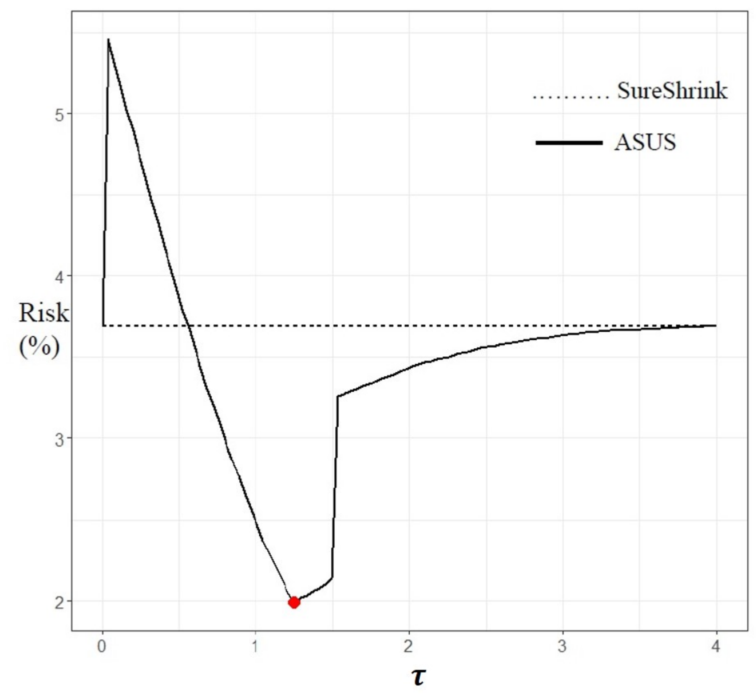
and

$$nS(\mathcal{T}, \mathbf{Y}, \mathbf{S}) = \sum_{i=1}^n \sigma_i^2 + \sum_{k=1}^2 \sum_{i \in \mathcal{I}_k^\tau} \Big\{ \sigma_i^2 \Big(\frac{|Y_i|}{\sigma_i} \wedge t_k\Big)^2 - 2\sigma_i^2 I\Big(\frac{|Y_i|}{\sigma_i} \le t_k\Big) \Big\}$$

is the SURE function.

Revisiting: estimating gene expression





- SURE estimate of risk of SureShrink estimator is 3.69% at $t = 0.61$
- At $\hat{\mathcal{T}} = \{1.25, 1.16, 0\}$, the SURE estimate of risk of ASUS is 1.99%.

Risk reduction by ASUS over SureShrink is about 30% in a predictive framework.

Risk Properties and Theoretical Analyses

- $S(\mathcal{T}, \mathbf{Y}, \mathbf{S})$ is uniformly close to the true risk (and loss)

With $c_n = n^{1/2}(\log n)^{-3/2}$,

$$c_n \sup_{\mathcal{T} \in \mathcal{H}_n} \Big| S(\mathcal{T}, \mathbf{Y}, \mathbf{S}) - r_n(\mathcal{T}; \boldsymbol{\theta}) \Big| \xrightarrow{L_1} 0,$$
$$c_n \sup_{\mathcal{T} \in \mathcal{H}_n} \Big| S(\mathcal{T}, \mathbf{Y}, \mathbf{S}) - l_n(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{SI}(\mathcal{T})) \Big| \xrightarrow{L_1} 0$$

where $\mathcal{H}_n = \mathbf{R}_+ \times [0, t_n]^2$ and $t_n = (2\log n)^{1/2}$

- $\mathcal{R}_n^{OS}$ - maximal risk of the oracle procedure
- $\mathcal{R}_n^{NS}$ - minimax risk of all soft thresh. estimators with no side information
- $\mathcal{R}_n^{AS}$ - maximal risk of ASUS
- $q_n^{jk}(\tau)$ - prob. of misclassifying coordinate i into class j (summed over i)

Asymptotic Optimality of ASUS

If $\exists$ a sequence $\{\tau_n\}_{n\ge1}$ such that $q_n^{12}(\tau_n)$ and $q_n^{21}(\tau_n)$ are appropriately controlled then

$$(\mathcal{R}_n^{NS} - \mathcal{R}_n^{AS})/(\mathcal{R}_n^{NS} - \mathcal{R}_n^{OS}) \to 1 \text{ as } n \to \infty$$

- Let $\mathcal{E}_n = (\mathcal{R}_n^{NS} - \mathcal{R}_n^{OS})/(\mathcal{R}_n^{AS} - \mathcal{R}_n^{OS})$

Robustness of ASUS

- We always have $\underline{\lim}\, \mathcal{E}_n \ge 1$.
- If for all sequence $\{\tau_n\}_{n\ge1}$, $q_n^{jk}(\tau_n)$ do not have the prescribed control then we must have

$$\mathcal{E}_n \to 1 \text{ as } n \to \infty$$

ASUS is atleast asymptotically as efficient as competitive methods when pooling non-informative auxiliary data.

Trambak.Banerjee.2020@marshall.usc.edu