Integrative sparse reduced-rank regression via orthogonal rotation for analysis of high-dimensional multi-source data

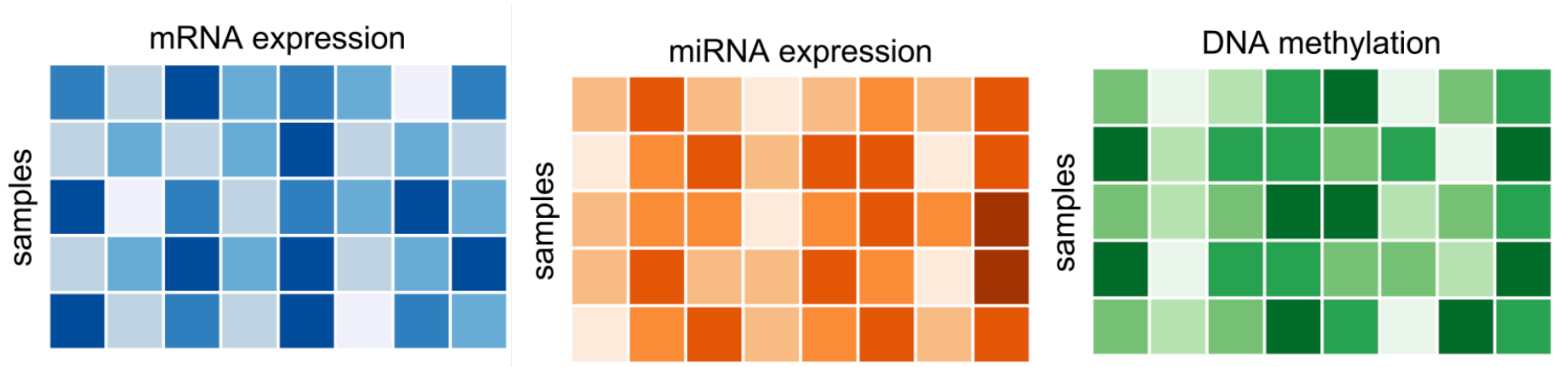
Kipoong Kim and Sungkyu Jung

Department of Statistics, Seoul National University

December 3, 2022

Multi-source data

- Recent technological advances have enabled us to collect massive amounts of multi-source data measured from the same individuals.
- For example, The Cancer Genome Atlas (TCGA) collects multi-omics data from various genomic technologies on the same samples
  - Multi-omics datasets



  - Multiple outcomes: drug responses
- Main goal is to investigate the relationship between multi-source data and multiple responses.

(1) Multivariate linear regression

- Suppose that we observed d multi-source datasets and q response variables from n individuals
  - $\mathbf{X}_i \in \mathbb{R}^{n \times p_i}$: the design matrix from the i-th source
  - $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_d] \in \mathbb{R}^{n \times p}$: the concatenated design matrix
  - $\mathbf{Y} \in \mathbb{R}^{n \times q}$: the response matrix

- Let us consider a general multivariate regression framework as

$$\mathbf{Y} = \mathbf{X}\mathbf{C} + \mathbf{E}, \tag{1}$$

  - $\mathbf{C} \in \mathbb{R}^{p \times q}$ is the coefficient matrix and $\mathbf{E} \in \mathbb{R}^{n \times q}$ is the error matrix containing independent random errors with mean zero.
  - Without loss of generality, we assumed that the predictor and response variables are centered to have mean zero.

- The multivariate linear regression is equivalent to performing q separate univariate linear regressions, so it does not take into account correlation among response variables.

(2) Reduced-rank regression (RRR)

- One way to handle this problem is to restrict the rank of the coefficient matrix $\mathbf{C}$ as $\text{rank}(\mathbf{C}) = r \leq \min\{n, p, q\}$, resulting in the following reduced-rank regression model[1] as

$$\mathbf{Y} = \mathbf{X}\mathbf{B}\mathbf{A}^T + \mathbf{E}, \tag{2}$$

  where $\mathbf{C} = \mathbf{B}\mathbf{A}^T$ for $\mathbf{A} \in \mathbb{R}^{q \times r}$ and $\mathbf{B} \in \mathbb{R}^{p \times r}$.

- Advantages
  - RRR can take into account the correlation between response variables through the latent variable, say $\mathbf{Z} = \mathbf{X}\mathbf{B}$.
  - RRR is more interpretable due to the latent variables
  - RRR can dramatically reduce the number of parameters to be estimated, and thus the estimates are more precise

[1] A. J. Izenman, *Journal of Multivariate Analysis* **5**, 248–264 (1975).

Structural learning in reduced-rank regression

- RRR model:

$$\mathbf{Y} = \mathbf{X}\mathbf{B}\mathbf{A}^T + \mathbf{E}, \tag{3}$$

- In this work, we assume
  - structured sparsity on $\mathbf{B} = (\boldsymbol{b}_{ik})$, where $\boldsymbol{b}_{ik} \in \mathbb{R}^{p_i \times 1}$.
  - row-wise sparsity on $\mathbf{A} = (A_{jk})$, where $A_{jk} \in \mathbb{R}$.

- e.g. if $d = 3$ data sources are given,

$$\mathbf{Y} = \left[\mathbf{X}_{(1)},\ \mathbf{X}_{(2)},\ \mathbf{X}_{(3)}\right] \begin{bmatrix} \boldsymbol{b}_{11} & \boldsymbol{b}_{12} & \mathbf{0} \\ \boldsymbol{b}_{21} & \boldsymbol{b}_{22} & \mathbf{0} \\ \boldsymbol{b}_{31} & \mathbf{0} & \boldsymbol{b}_{33} \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \\ 0 & 0 & 0 \end{bmatrix}^\top + \mathbf{E}, \tag{4}$$

  where $\boldsymbol{b}_{ik} \in \mathbb{R}^{p_i \times 1}$ and each column produces three types of components: (1) Joint, (2) Partially-joint, (3) Individual.

Identifiability problem in reduced-rank regression

- The reduced-rank regression model has a limitation that the decomposition $\mathbf{C} = \mathbf{B}\mathbf{A}^T$ is not unique up to an orthogonal matrix.

- For example,
  - suppose that we have a parameter pair $(\mathbf{A}, \mathbf{B})$ which leads to $\mathbf{C} = \mathbf{B}\mathbf{A}^T$.
  - However, for an $r \times r$ orthogonal matrix $\mathbf{Q}$,
    another parameter pair $(\mathbf{A}\mathbf{Q}, \mathbf{B}\mathbf{Q})$ also produces $\mathbf{C} = \mathbf{B}\mathbf{Q}\mathbf{Q}^T\mathbf{A}^T$.

Quartimax-simple structure

- In identifiability problem, $\mathbf{AQ}$ can be thought of as the orthogonal rotation in factor analysis.
- Quartimax criterion: $\mathcal{F}(\mathbf{A}) = \sum_{j=1}^{q} \sum_{k=1}^{r} A_{jk}^4$.
- $\mathcal{O}(r) = \left\{ \mathbf{Q} \in \mathbb{R}^{r \times r} : \mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}_r \right\}$
  $\mathcal{O}(q, r) = \left\{ \mathbf{A} \in \mathbb{R}^{q \times r} : \mathbf{A}^T\mathbf{A} = \mathbf{I}_r \right\}$ (also called Stiefel manifold)

Definition 1 (Quartimax-simple structure)

Given $\mathbf{A} \in \mathbb{R}^{q \times r}$, the rotated matrix $\mathbf{AQ}$ is said to have a *quartimax-simple structure* if $\mathbf{Q}$ maximizes the quartimax criterion $\mathcal{F}(\mathbf{AQ})$ over all $\mathbf{Q} \in \mathcal{O}(r)$. Also, a set of semi-orthogonal matrices with simple structure is defined as

$$\mathcal{O}_S(q, r) = \left\{ \mathbf{A}\hat{\mathbf{Q}} : \hat{\mathbf{Q}} = \underset{\mathbf{Q} \in \mathcal{O}(r)}{\arg\max}\ \mathcal{F}(\mathbf{AQ}),\ \mathbf{A} \in \mathcal{O}(q, r) \right\}.$$

Constrained reduced-rank regression model

- We consider the constrained reduced-rank regression model under the quartimax-simple loading matrix $\mathbf{A}$:

$$\mathbf{Y} = \mathbf{X}\mathbf{B}\mathbf{A}^T + \mathbf{E}, \quad \mathbf{A} \in \mathcal{O}_S(q, r), \tag{5}$$

  where $\mathbf{E} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)^T$ with $\boldsymbol{e}_l \sim \mathcal{N}_q(\mathbf{0}, \sigma^2\mathbf{I})$, $l = 1, \ldots, n$.

- The following proposition illustrates the identifiability of (5).

Proposition 1 (Identifiability of the constrained RRR model)

In model (5), if $\mathbf{B}^T\mathbf{X}^T\mathbf{X}\mathbf{B}$ has r distinct positive eigenvalues for the fixed design matrix $\mathbf{X}$, then the parameter set $(\mathbf{A}, \mathbf{X}\mathbf{B}, \sigma^2)$ is identifiable up to simultaneous signed permutations of the columns of $\mathbf{A}$ and $\mathbf{X}\mathbf{B}$.

Restricted eigenvalue condition

- However, in the high-dimensional setting with $n < p$, $\mathbf{XB} = \mathbf{XB}'$ does not imply that $\mathbf{B} = \mathbf{B}'$ for $\mathbf{B}, \mathbf{B}' \in \mathbb{R}^{p \times r}$.
- We consider the following restricted eigenvalue condition for sparse $\mathbf{B}$:

Condition 1 (Restricted eigenvalue condition)

Let $J \subseteq \{(u, v) : 1 \leq u \leq p, 1 \leq v \leq r\}$ be any index set. Consider the following set of matrices:

$$\mathbb{C}(s, \xi) = \left\{ \mathbf{\Delta} \in \mathbb{R}^{p \times r} \backslash \{\mathbf{0}\} : \|\mathbf{\Delta}_{J^c}\|_{1,1} \leq \xi \|\mathbf{\Delta}_J\|_{1,1}, \ |J| \leq s \right\},$$

for some $1 \leq s \leq pr$ and $\xi > 0$, where $\|\cdot\|_{1,1}$ is the entry-wise ℓ_1 norm. The matrix $\mathbf{X} \in \mathbb{R}^{n \times p}$ is said to satisfy the restricted eigenvalue (RE) condition over $\mathbb{C}(s, \xi)$ if there exists $\kappa(s, \xi) > 0$ such that

$$\kappa(s, \xi) = \min_{\mathbf{\Delta} \in \mathbb{C}(s, \xi)} \frac{\|\mathbf{X\Delta}\|_F}{\sqrt{n} \|\mathbf{\Delta}\|_F}.$$

Identifiability under RE condition

- Under the restricted eigenvalue condition, we have the following corollary.

Corollary 2

In Proposition 1, assume that $\mathbf{B}$ *has at most* s *nonzero elements. If the design matrix* $\mathbf{X} \in \mathbb{R}^{n \times p}$ *satisfies the RE condition over* $\mathbb{C}(2s, \xi)$ *for some* $\xi > 0$*, the set of parameters* $(\mathbf{A}, \mathbf{B}, \sigma^2)$ *is identifiable up to simultaneous signed permutations of the columns.*

Integrative sparse reduced-rank regression (iSRRR)

- We propose to estimate $\mathbf{A}$ and $\mathbf{B}$ for integrative sparse reduced-rank regression (iSRRR) by solving the constrained optimization problem

$$\min_{\mathbf{A},\mathbf{B}} \frac{1}{2n}\|\mathbf{Y} - \mathbf{X}\mathbf{B}\mathbf{A}^T\|_F^2 + \lambda \sum_{i=1}^{d}\sum_{k=1}^{r} \sqrt{p_i}\|\boldsymbol{b}_{ik}\|_2 \tag{6}$$
$$\text{subject to } \mathbf{A} \in \mathcal{O}_S(q,r) \text{ and } \mathbf{A} \in \mathcal{T}(\nu),$$

  where $r \geq 1$ is a given rank and

$$\mathcal{T}(\nu) = \left\{ \mathbf{A} \in \mathcal{O}(q,r) : \min_{j:\boldsymbol{a}_{j\cdot} \neq \mathbf{0}} \|\boldsymbol{a}_{j\cdot}\|_2 \geq \nu \right\}.$$

- Tuning parameters:
  - Tuning parameter $\lambda \geq 0$ controls the structured sparsity of $\mathbf{B}$
  - Row-wise sparsity level ν controls the entry-wise sparsity of $\mathbf{A}$.

Estimation: Alternating algorithm

- For an initial value $(\mathbf{A}^{(0)}, \mathbf{B}^{(0)})$,
  - *A*-step: Given $\mathbf{B}^{(m)}$, solve the problem in three substeps

$$\min_{\mathbf{A}} \frac{1}{2n}\|\mathbf{Y} - \mathbf{X}\mathbf{B}\mathbf{A}^T\|_F^2 \text{ s.t. } \mathbf{A} \in \mathcal{T}(\nu) \text{ and } \mathbf{A} \in \mathcal{O}_S(q, r)$$

    (1) Orthogonal Procrustes Problem
    (2) Hard Thresholding
    (3) Orthogonal Projection

  - *Q*-step: Given $(\mathbf{A}^{(m)}, \mathbf{B}^{(m)})$, obtain the rotation matrix

$$\hat{\mathbf{Q}} = \arg\max_{\mathbf{Q}} \mathcal{F}(\mathbf{A}^{(m)}\mathbf{Q}),$$

    and rotate $\mathbf{A}^{(m)} \leftarrow \mathbf{A}^{(m)}\hat{\mathbf{Q}}$ and $\mathbf{B}^{(m)} \leftarrow \mathbf{B}^{(m)}\hat{\mathbf{Q}}$.

  - *B*-step: Given $\mathbf{A}^{(m)}$, perform the group-lasso regression estimation

$$\min_{\mathbf{B}} \sum_{k=1}^{r} \left[\frac{1}{2n}\|\mathbf{Y}\mathbf{A}_k - \mathbf{X}\mathbf{B}_k\|_2^2 + \lambda \sum_{i=1}^{d} \sqrt{p_i}\|\boldsymbol{b}_{ik}\|_2 \right],$$

    where $\mathbf{A}_k$ and $\mathbf{B}_k$ denote the kth column of $\mathbf{A}$ and $\mathbf{B}$, respectively.

Rank selection and parameter tuning

- The rank selection can be treated as estimating the rank of $\mathbf{C}$.
  - Self-Tuning Rank Selection (STRS)[2].
  - BIC for rank selection (BICk)[3].
    - Both methods enjoy nice asymptotic properties.

- Our model includes two tuning parameters: $\lambda \geq 0$ and $\nu \in [0, 1]$.
  - For choosing the optimal (λ, ν), we propose to use the Bayesian information criterion:

$$BIC(\lambda) = \frac{1}{nq}\|\mathbf{Y} - \mathbf{X}\hat{\mathbf{B}}_\lambda \hat{\mathbf{A}}_\lambda^T\|_F^2 + \frac{\log(nq)}{nq}\hat{\mathrm{df}}(\lambda),$$

where $\hat{\mathrm{df}}(\lambda) = \|\hat{\mathbf{A}}_\lambda\|_0 + \|\hat{\mathbf{B}}_\lambda\|_0$.

[2] X. Bing, M. H. Wegkamp, *The Annals of Statistics* **47**, 3157–3184 (2019).

[3] C. Zou *et al.*, *Journal of the American Statistical Association* **117**, 693–703 (2022).

Simulation design

- The simulated data were generated by the model

$$\mathbf{Y} = \mathbf{X}\mathbf{B}\mathbf{A}^T + \mathbf{E},$$

  - $\mathbf{X} = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)^T$ with $\boldsymbol{x}_l \sim \mathcal{N}(0, \boldsymbol{\Psi}), l = 1, \ldots, n,$
  - $\boldsymbol{\Psi} = \{\psi_{uv}\}$ with $\psi_{uv} = 0.5^{|u-v|}$ for $u, v = 1, \ldots, p.$
  - Each row of the random error matrix $\mathbf{E} = (\boldsymbol{e}_1, \ldots, \boldsymbol{e}_n)^T$ was generated from the multivariate normal distribution as $\boldsymbol{e}_l \sim \mathcal{N}(\mathbf{0}, \sigma^2\mathbf{I})$,
  - σ^2 was set so that $SNR = \|\mathbf{X}\mathbf{C}\|_F / \|\mathbf{E}\|_F = 0.5$, where $\mathbf{C} = \mathbf{B}\mathbf{A}^T$

- Settings: $n \in \{10, 50, 100\}$, $p_i = 500$, $q = 60$, $d = 4$, $r = 6$
- Simulation was repeated 100 times

Simulation design

- Three different types of $\mathbf{B} = \{\boldsymbol{b}_{ik}\}$ with $\boldsymbol{b}_{ik} \in \mathbb{R}^{p_i \times 1}$ were considered as structured sparsity patterns as

  (a) Individual: $\mathbf{B} = \begin{bmatrix} \boldsymbol{b}_{11} & \boldsymbol{b}_{12} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{b}_{23} & \boldsymbol{b}_{24} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{b}_{35} & \boldsymbol{b}_{36} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$

  (b) Mixed: $\mathbf{B} = \begin{bmatrix} \boldsymbol{b}_{11} & \boldsymbol{b}_{12} & \mathbf{0} & \boldsymbol{b}_{14} & \boldsymbol{b}_{15} & \mathbf{0} \\ \boldsymbol{b}_{21} & \mathbf{0} & \boldsymbol{b}_{23} & \boldsymbol{b}_{24} & \mathbf{0} & \boldsymbol{b}_{26} \\ \boldsymbol{b}_{31} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \boldsymbol{b}_{35} & \boldsymbol{b}_{36} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{bmatrix}$,

  where each entry was generated from $\mathcal{U}(0.1, 0.3)$.

- A semi-orthogonal and quartimax-simple $\mathbf{A}$ with sparse entries was generated by repeating *(1) orthogonalization, (2) hard-thresholding, (3) rotation steps* on random matrix from $N(0, 1)$, with true sparsity-level $\nu^* = 0,\ 0.2,\ 0.4$.
  - The sparsity levels produce approximately 0%, 15%, 80% zero rows, respectively.

Evaluation criteria

- Recovery of the underlying structure
  - Element-wise accuracy of $\hat{\mathbf{C}}$.
- Estimation performances:
  - The first principal angles:

$$\angle(\mathbf{A}, \hat{\mathbf{A}}) \quad \text{and} \quad \angle(\mathbf{B}, \hat{\mathbf{B}})$$

- Predictive performance
  - RMSPE on an independent test dataset:

$$\text{RMSPE} = \sqrt{\|\mathbf{Y}_{\text{test}} - \mathbf{X}_{\text{test}}\hat{\mathbf{C}}\|_F^2 / n_{\text{test}} q}$$

- All evaluation criteria were averaged over 100 simulation replicates.

Result: Element-wise accuracy of $\hat{\mathbf{C}}$

		Individual				Mixed			
ν^*	n	iSRRR	RSSVD	SOFAR	SECURE	iSRRR	RSSVD	SOFAR	SECURE
0	10	0.974 (0.07)	0.254 (0.00)	0.750 (0.00)	0.250 (0.00)	0.943 (0.11)	0.252 (0.00)	0.750 (0.00)	0.250 (0.00)
	50	0.998 (0.01)	0.291 (0.01)	0.750 (0.00)	0.255 (0.00)	0.998 (0.01)	0.254 (0.01)	0.750 (0.00)	0.262 (0.01)
	100	0.999 (0.00)	0.343 (0.02)	0.750 (0.00)	0.261 (0.00)	0.998 (0.00)	0.264 (0.02)	0.750 (0.00)	0.279 (0.01)
0.2	10	0.920 (0.08)	0.340 (0.03)	0.663 (0.03)	0.337 (0.03)	0.873 (0.12)	0.339 (0.03)	0.663 (0.03)	0.337 (0.03)
	50	0.991 (0.01)	0.375 (0.03)	0.663 (0.03)	0.343 (0.03)	0.979 (0.02)	0.343 (0.03)	0.663 (0.03)	0.351 (0.03)
	100	0.999 (0.00)	0.421 (0.03)	0.663 (0.03)	0.348 (0.03)	0.995 (0.01)	0.352 (0.03)	0.663 (0.03)	0.367 (0.03)
0.4	10	0.931 (0.12)	0.899 (0.02)	0.198 (0.27)	0.899 (0.02)	0.865 (0.16)	0.896 (0.02)	0.146 (0.19)	0.898 (0.02)
	50	0.987 (0.04)	0.904 (0.02)	0.174 (0.24)	0.901 (0.02)	0.952 (0.09)	0.899 (0.02)	0.140 (0.18)	0.902 (0.02)
	100	0.995 (0.02)	0.911 (0.02)	0.119 (0.13)	0.903 (0.02)	0.993 (0.02)	0.902 (0.02)	0.149 (0.20)	0.906 (0.02)

Table: The corresponding standard errors are presented in parentheses.

Result: Principal angles and accuracies of $(\hat{\mathbf{A}}, \hat{\mathbf{B}})$

ν^*	n	Individual				Mixed			
		$\angle(\hat{\mathbf{A}})$	$\angle(\hat{\mathbf{B}})$	$\overline{acc}(\hat{\mathbf{A}})$	$acc(\hat{\mathbf{B}})$	$\angle(\hat{\mathbf{A}})$	$\angle(\hat{\mathbf{B}})$	$\overline{acc}(\hat{\mathbf{A}})$	$acc(\hat{\mathbf{B}})$
0	10	0.285 (0.63)	20.968 (2.38)	0.992 (0.02)	0.839 (0.07)	0.263 (0.08)	21.696 (2.59)	0.997 (0.01)	0.843 (0.06)
	50	0.096 (0.17)	13.757 (0.56)	0.998 (0.01)	0.768 (0.10)	0.120 (0.03)	12.981 (1.02)	0.997 (0.01)	0.789 (0.06)
	100	0.056 (0.01)	12.189 (0.41)	0.998 (0.01)	0.747 (0.11)	0.086 (0.02)	10.678 (0.61)	0.998 (0.01)	0.780 (0.05)
0.2	10	0.262 (0.48)	20.957 (2.37)	0.928 (0.03)	0.850 (0.06)	0.306 (0.31)	21.804 (2.39)	0.910 (0.03)	0.844 (0.06)
	50	0.072 (0.01)	13.706 (0.53)	0.988 (0.02)	0.788 (0.10)	0.105 (0.02)	12.993 (1.09)	0.971 (0.03)	0.788 (0.06)
	100	0.051 (0.01)	12.183 (0.42)	0.998 (0.01)	0.750 (0.11)	0.076 (0.01)	10.698 (0.65)	0.993 (0.01)	0.780 (0.05)
0.4	10	0.024 (0.04)	20.186 (2.16)	0.913 (0.16)	0.941 (0.06)	0.072 (0.09)	21.707 (2.96)	0.834 (0.19)	0.895 (0.07)
	50	0.002 (0.01)	13.299 (0.48)	0.985 (0.05)	0.968 (0.06)	0.014 (0.03)	12.036 (0.64)	0.937 (0.12)	0.945 (0.06)
	100	0.000 (0.00)	11.855 (0.37)	0.997 (0.03)	0.955 (0.08)	0.002 (0.01)	10.225 (0.44)	0.991 (0.02)	0.932 (0.08)

Table: Here, $\angle(\cdot)$, $\overline{acc}(\cdot)$ and $acc(\cdot)$ refer to the largest principal angle, row-wise accuracy and element-wise accuracy, respectively. The principal angles are given in degrees.

NCI-60 dataset

- We applied the proposed method to an NCI-60 multi-omics dataset with drug responses[4] that is publicly available in the R package "rcellminer"
- We applied the sure independence screening (SIS)[5] for dense signals.
- After some pre-processing steps, we finally have

$$\mathbf{X}_{exp} \in \mathbb{R}^{57\times 502},\ \mathbf{X}_{mut} \in \mathbb{R}^{57\times 409},\ \mathbf{X}_{met} \in \mathbb{R}^{57\times 476},\ \mathbf{X}_{cop} \in \mathbb{R}^{57\times 558},$$

$$\mathbf{Y}_{drug} \in \mathbb{R}^{57\times 53}$$

  where *exp*, *mut*, *met* and *cop* represent expression, mutation, methylation, and copy number variation, respectively.
- The rank was estimated at 2 by STRS , and (λ, ν) was chosen by BIC.

---

[4] S. Dietrich *et al.*, *The Journal of Clinical Investigation* **128**, 427–445 (2018).

[5] J. Fan, J. Lv, *Journal of the Royal Statistical Society: Series B* **70**, 849–911 (2008).

The results of NCI-60 dataset

Heatmap of the estimated coefficient matrix $\hat{\mathbf{C}}$

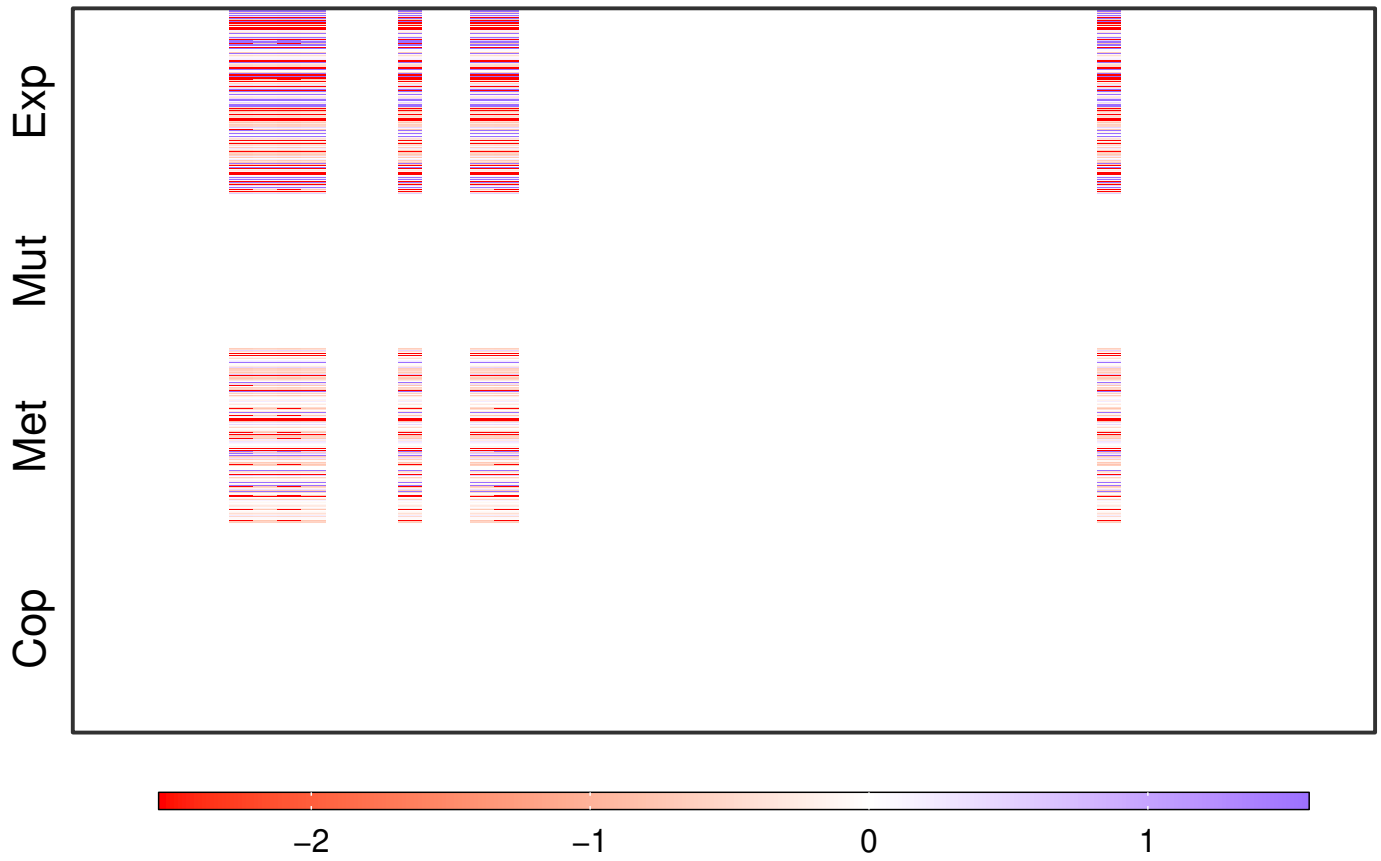


Figure: The rows and columns correspond to predictors and responses, respectively. Extreme values are truncated for better visibility

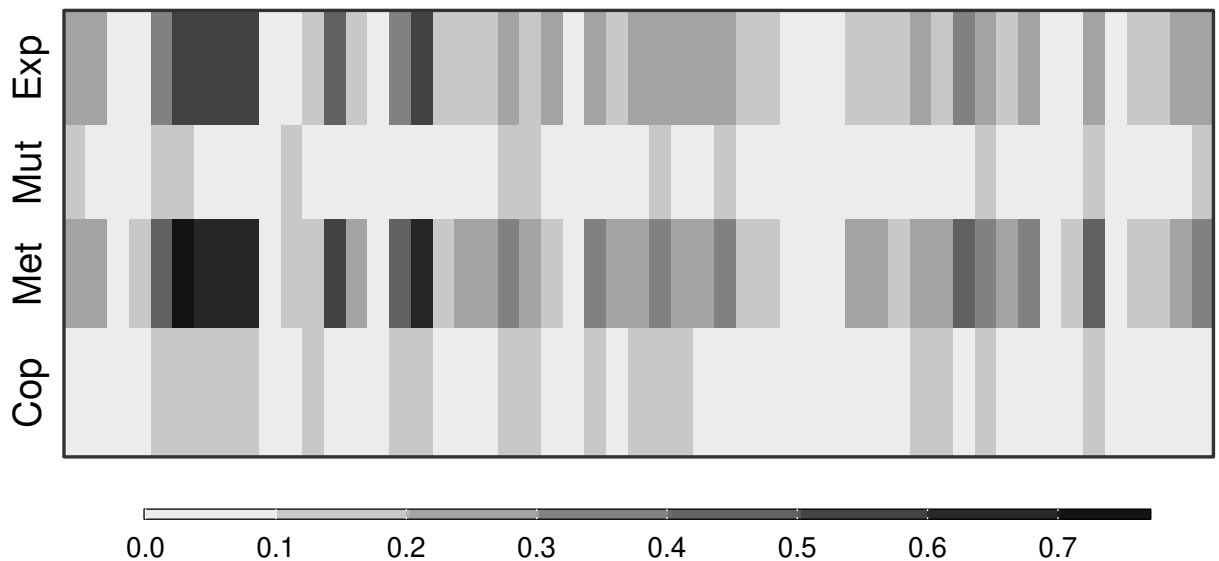
The results of NCI-60 dataset

The structure-wise selection frequencies from 100 bootstrap replicates

Structure	Selection Frequency
$(\ \cdot\ ,\ \ \cdot\ ,\ \ \cdot\ ,\ \ \cdot\ )$	0.37
$(exp,\ \ \cdot\ ,\ \ \cdot\ ,\ \ \cdot\ )$	0.21
$(\ \cdot\ , mut,\ \ \cdot\ ,\ \ \cdot\ )$	0.12
$(\ \cdot\ ,\ \ \cdot\ , met,\ \ \cdot\ )$	0.41
$(\ \cdot\ ,\ \ \cdot\ ,\ \ \cdot\ , cop)$	0.15
$(exp,\ \ \cdot\ , met,\ \ \cdot\ )$	0.51
$(\ \cdot\ , mut,\ \ \cdot\ , cop)$	0.01
$(\ \cdot\ ,\ \ \cdot\ , met,\ cop)$	0.02
$(exp,\ mut,\ \ \cdot\ , cop)$	0.01
$(\ \cdot\ , mut,\ met,\ cop)$	0.01

The results of NCI-60 dataset

The heatmap of selection frequencies for the estimated $\hat{\mathbf{C}}$ from 100 bootstrap replicates



Conclusion

- In this work, we have proposed a new method for both integrating multi-source data and recovering structured sparsity in the reduced-rank regression framework.
- The proposed method restricts the parameter space to a set of semi-orthogonal matrices that have the simple structure based on the quartimax criterion in order to solve the identifiability problem.
- Our method have advantages in the following two aspects:
  (1) it is more interpretable and can detect the structured sparsity patterns;
  (2) it can perform the simultaneous selection for both predictors and responses through achieving sparisty on both $\mathbf{A}$ and $\mathbf{B}$.
- Some interesting topics for future research:
  - the group-lasso penalty cannot detect variables within a group
  - non-gaussian data types of responses and/or predictors

Thank you for attending