`cmenet` - A new method for bi-level variable selection of conditional main effects (CMEs)

C. F. Jeff Wu

Georgia Institute of Technology

Georgia Tech ISyE
H. Milton Stewart School of Industrial and Systems Engineering

---

Mak, S. and Wu, C. F. J. (2018). `cmenet`: a new method for bi-level variable selection of conditional main effects. *Journal of the American Statistical Association.* 114(526): 844–856.

Section 1

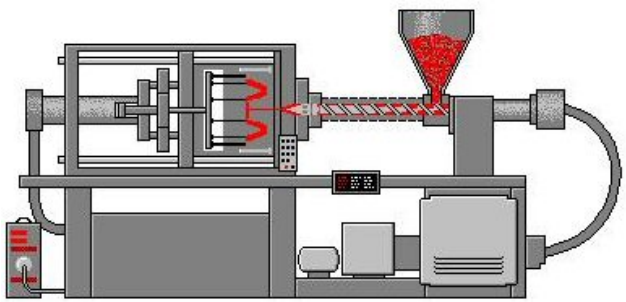
Introduction: CME analysis in designed experiments

"What happened to unconditional love?!"

https://www.andertoons.com/cartoons/dog

Conditional main effects

- A **conditional main effect (CME)** is the conditional effect of a factor at a fixed level of another factor
- CMEs have a direct interpretation in many applications:
  - **Genomics**:
    - E.g., which genes are conditionally active, which genes activate other genes
  - **Engineering**:
    - E.g., effect of mold temperature only at a high level of holding pressure
  - **Social sciences**:
    - E.g., effect of income on GPA, conditional on different ethnic backgrounds



Background on CMEs

- First introduced by Wu (2015) (following 2011 Fisher Lecture) as a way to disentangle aliased effects in a **designed experiment**
  - Believed to be impossible since the pioneering work (Finney, 1945) on fractional factorial designs
- Su and Wu (2017) developed a **variable selection** framework for CMEs in designed experiments:
  - Exploits group structure of CMEs under an orthogonal model
  - Selected models are more parsimonious, with aliased interactions untangled

Wu, C. F. J. (2015). Post-Fisherian experimentation: from physical to virtual. *Journal of the American Statistical Association*, 110(510):612–620.

Constructive definition of CMEs

Consider two factors A and B, each with two levels $+$ and $-$:

- **Main effect (ME)** of A:

$$\begin{aligned} \mathrm{ME}(A) &= \bar{y}(A+) - \bar{y}(A-) \\ &= \frac{1}{2}\Big\{\bar{y}(A+|B+) + \bar{y}(A+|B-)\Big\} - \frac{1}{2}\Big\{\bar{y}(A-|B+) + \bar{y}(A-|B-)\Big\} \end{aligned}$$

- **Two-factor interaction (2FI)** of A and B:

$$\mathrm{INT}(A,B) = \frac{1}{2}\Big\{\bar{y}(A+|B+)+\bar{y}(A-|B-)\Big\} - \frac{1}{2}\Big\{\bar{y}(A+|B-)+\bar{y}(A-|B+)\Big\}$$

- **Conditional main effect** of A given B at level $+$:

$$\mathrm{CME}(A|B+) = \bar{y}(A+|B+) - \bar{y}(A-|B+)$$

- **Conditional main effect** of A given B at level $-$:

$$\mathrm{CME}(A|B-) = \bar{y}(A+|B-) - \bar{y}(A-|B-)$$

Constructive definition of CMEs

- From this, one can derive the following identities:

$$\mathrm{CME}(A|\mathrm{B}+) = \frac{1}{2}\Big\{\mathrm{ME}(A) + \mathrm{INT}(A,\mathrm{B})\Big\}$$

$$\mathrm{CME}(A|\mathrm{B}-) = \frac{1}{2}\Big\{\mathrm{ME}(A) - \mathrm{INT}(A,\mathrm{B})\Big\}$$

$\mathrm{ME}(\mathrm{A})$	$\mathrm{ME}(\mathrm{B})$	$\mathrm{INT}(\mathrm{A},\mathrm{B})$	$\mathrm{CME}(\mathrm{A}\|\mathrm{B}+)$	$\mathrm{CME}(\mathrm{A}\|\mathrm{B}-)$
$+1$	$+1$	$+1$	$+1$	0
$+1$	-1	-1	0	$+1$
-1	$+1$	-1	-1	0
-1	-1	$+1$	0	-1

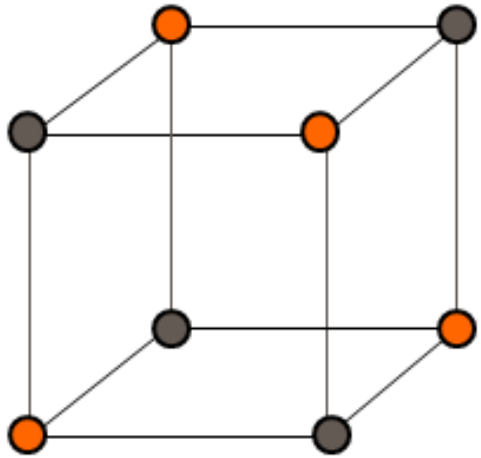
Table 1: Construction of the CMEs $A|\mathrm{B}+$ and $A|\mathrm{B}-$.

- CMEs can be viewed as a component of an interaction effect

De-aliasing via CME reparametrization

For illustration, take the 2^{6-2}_{IV} fractional factorial design with aliasing relation:

$$\mathrm{I} = \mathrm{ABCE} = \mathrm{BCDF} = \mathrm{ADEF}$$



- Interactions AB and CE are fully aliased (Wu and Hamada, 2009) – there's no way to separate their effects from designed data
- But AB and CE can be reparametrized via their CMEs (e.g., $\mathrm{A}|\mathrm{B}+$ and $\mathrm{C}|\mathrm{E}+$), which are only partially aliased and can be estimated
- Goal is to analyze designed data via the reparametrized CMEs, which bypasses the fully-aliased structure in interaction effects

De-aliasing via CME reparametrization

Key **selection rule** (Rule 1) in Su and Wu (2017):

Suppose main effect A and interaction AB are selected via traditional analysis (e.g. a half-normal plot):

- If A and AB have same signs and similar magnitudes, then replace both A and AB with the CME $A|B+$
  - **Intuition**: $\mathrm{CME}(A|B+) = \frac{1}{2}\Big\{\mathrm{ME}(A) + \mathrm{INT}(A,B)\Big\}$ has greater effect than both A and AB
- If A and AB have opposite signs and similar magnitudes, then replace both A and AB with the CME $A|B-$
  - **Intuition**: $\mathrm{CME}(A|B-) = \frac{1}{2}\Big\{\mathrm{ME}(A) - \mathrm{INT}(A,B)\Big\}$ has greater effect than both A and AB

Su, H. and Wu, C. F. J. (2017). Cme analysis: a new method for unraveling aliased effects in two-level fractional factorial experiments. *Journal of Quality Technology*, 49(1):1–10.

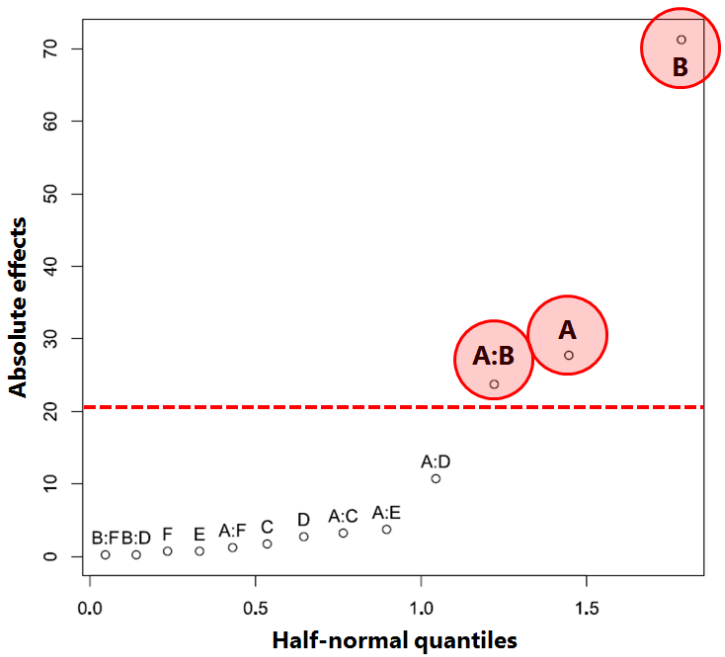
A simple example

Consider an **injection molding** experiment (Montgomery, 1991):

- 2^{6-2}_{IV} fractional factorial design ($n = 16$ runs) with

$$I = ABCE = BCDF = ADEF$$

- Traditional analysis (half-normal plot) selects A, B and AB as active effects



- Fitted model:

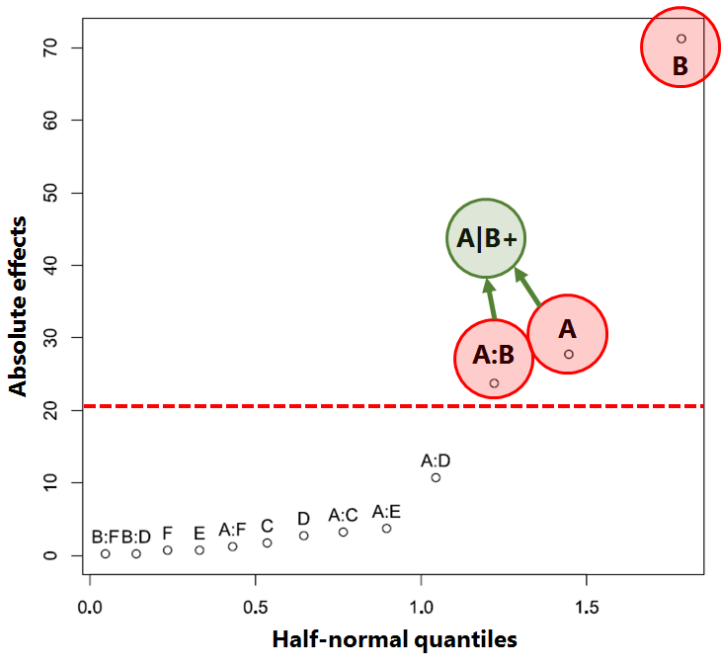
$$y \sim (2.4\times10^{-9})B + (5.4\times10^{-5})A + (2.2\times10^{-4})AB \quad (R^2 = 96.2\%)$$

A simple example

With **CME analysis**:

- Since A and AB have same signs, replace both with the CME $A|B+$
- CME model:

$$y \sim (6.1 \times 10^{-10})B + (1.7 \times 10^{-6})A|B+ \quad (R^2 = 96.1\%)$$



- **New** model more parsimonious, smaller effect p-values and similar R^2 to traditional model
- Good engineering interpretation: pressure (A) has a significant effect on shrinkage (y) at high screw speed $(B+)$, but not low speed $(B-)$

Section 2

CME selection for observational data

Onto observational data

- CMEs equally as valuable for analyzing **observational data** – these basis functions are more **interpretable** than traditional interactions
  - E.g., in genetics, which genes are conditionally active, and which genes activate other genes

"Examining the consequence of how one mutation behaves when in the presence of a second mutation forms the basis of our understanding of genetic interactions, and is part of the fundamental toolbox of genetic analysis."

– Chari and Dworkin (2013, PLoS Genetics)

Chari, S. and Dworkin, I. (2013). The conditional nature of genetic interactions: the consequences of wild-type backgrounds on mutational interactions in a genome-wide modifier screen. *PLoS Genetics*, 9(8):e1003661.

Conditional definition of CMEs

Definition (Conditional main effect)

Let $\tilde{\mathbf{x}}_j \in \{-1, +1\}^n$ be the covariate vector for **main effect (ME)** J, $j = 1, \cdots, p$. The **CME** J|K+ quantifies the effect of $\tilde{\mathbf{x}}_j$ conditional on $\tilde{\mathbf{x}}_k = +1$.

- J and K are the **parent** and **conditioned** effects of CME J|K+

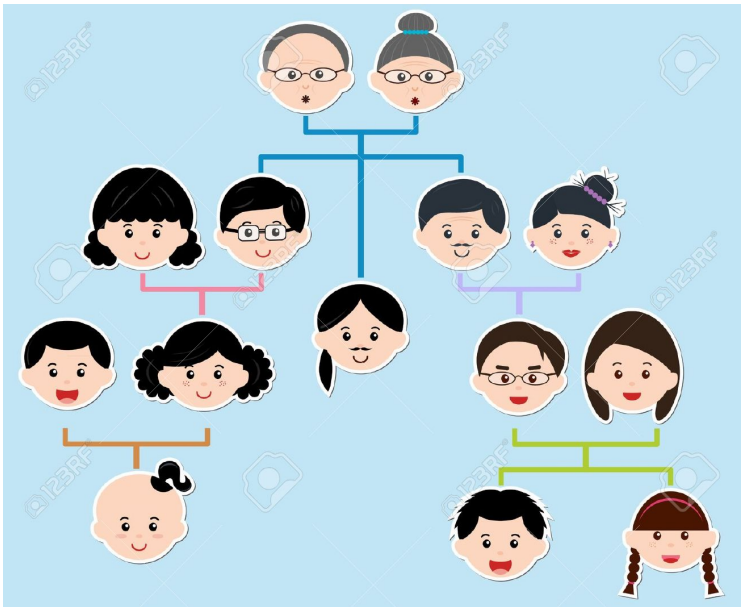
A	B	$A\|B+$	$A\|B-$	$B\|A+$	$B\|A-$
+1	+1	+1	0	+1	0
+1	-1	0	+1	-1	0
-1	+1	-1	0	0	+1
-1	-1	0	-1	0	-1

Table 2: MEs A and B, and its four CMEs A|B+, A|B−, B|A+, B|A−.

CME groupings

Consider the following **effect groups**:

- **Siblings**: CMEs with same parent effect, e.g., $A|B+$ and $A|C+$
- **Cousins**: CMEs with same conditioned effect, e.g., $B|A+$ and $C|A+$
- **Parent-child**: A CME and its parent, e.g., $A|B+$ and A



The need for new methodology

Why not use off-the-shelf methods for selecting CMEs?

- Standard procedure:
  - Normalize each CME to zero mean and unit variance
  - Apply **LASSO** (Tibshirani, 1996), or your favorite **non-convex** penalty, e.g., SCAD (Fan and Li, 2001) or MC+ (Zhang, 2010)
- But this **ignores** the implicit group structure of CMEs!

Why not **Group LASSO** (Yuan and Lin, 2006)?

- This select all effects in a group, whereas only a handful of effects may be active in a CME group
- We need a **bi-level** selection framework (Breheny, 2015), which selects both active CME groups and CMEs within groups

---

Breheny, P. (2015). The group exponential lasso for bi-level variable selection. *Biometrics*, 71(3):731–740.

Sibling and cousin groups

We will group CMEs into **sibling** and **cousin groups**:

- **Sibling group** of J:

$$\mathcal{S}(j) = \left\{J, J|A+, J|A-, J|B+, J|B-, \cdots\right\}$$

  - Consists of J and all CMEs with parent J
- **Cousin group** of J:

$$\mathcal{C}(j) = \left\{J, A|J+, A|J-, B|J+, B|J-, \cdots\right\}$$

  - Consists of J and all CMEs with condition J

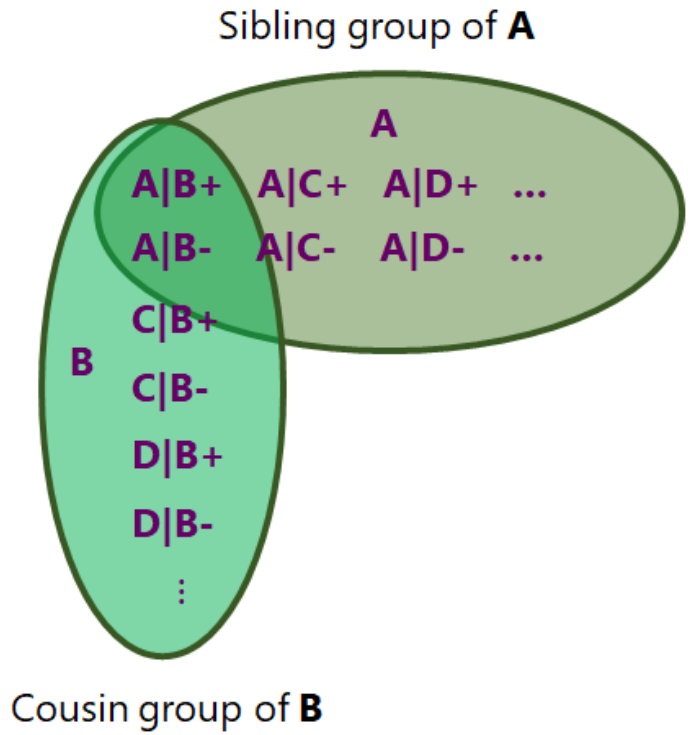


Figure 1: Sibling group of A, cousin group of B.

Section 3

Bi-level variable selection criterion

Bi-level selection criterion

We propose the following **selection criterion**:

$$\min_{\beta} Q(\beta) \equiv \min_{\beta} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + P_{\mathcal{S}}(\beta) + P_{\mathcal{C}}(\beta) \right\}$$

- $y \in \mathbb{R}^n$ is the observed response vector
- $X \in \mathbb{R}^{n \times p'}$ is the normalized model matrix, where $p' = p + 4\binom{p}{2}$ is the total # of MEs and CMEs
- $\beta \in \mathbb{R}^{p'}$ is the coefficient vector for MEs and CMEs
- $P_{\mathcal{S}}(\beta)$ and $P_{\mathcal{C}}(\beta)$ are the sibling and cousin penalty functions:

$$P_{\mathcal{S}}(\beta) = \sum_{j=1}^{p} f_{o,\mathcal{S}} \left\{ \sum_{k \in \mathcal{S}(j)} f_{i,\mathcal{S}}(\beta_k) \right\}, P_{\mathcal{C}}(\beta) = \sum_{j=1}^{p} f_{o,\mathcal{C}} \left\{ \sum_{k \in \mathcal{C}(j)} f_{i,\mathcal{C}}(\beta_k) \right\}$$

Outer and inner penalties

$$P_{\mathcal{S}}(\boldsymbol{\beta}) = \sum_{j=1}^{p} f_{o,\mathcal{S}} \left\{ \sum_{k \in \mathcal{S}(j)} f_{i,\mathcal{S}}\left(\beta_k\right) \right\}, P_{\mathcal{C}}(\boldsymbol{\beta}) = \sum_{j=1}^{p} f_{o,\mathcal{C}} \left\{ \sum_{k \in \mathcal{C}(j)} f_{i,\mathcal{C}}\left(\beta_k\right) \right\}$$

The **outer** and **inner** penalties f_o and f_i parametrize the **bi-level selection** of CMEs:

- f_o controls **between-group** selection (selecting CME groups):

$$f_o(\theta) = \frac{\lambda^2}{\tau}\left(1 - \exp\left\{-\frac{\tau\theta}{\lambda}\right\}\right) \quad \text{(Exponential penalty; Breheny, 2015)}$$

- f_i controls **within-group** selection (selecting CMEs within a group):

$$f_i(\beta) = \int_0^{|\beta|} \left(1 - \frac{x}{\lambda\gamma}\right)_+ dx \quad \text{(MC+ non-convex penalty; Zhang, 2010)}$$

- Different penalties λ_s and λ_c for sibling and cousin groups

CME coupling

The **bi-level** formulation gives two appealing selection principles:

- **CME coupling**: If $A|B+$ is active, then its siblings $A|C+$, $A|D+$, $\cdots$ are more likely to be active as well

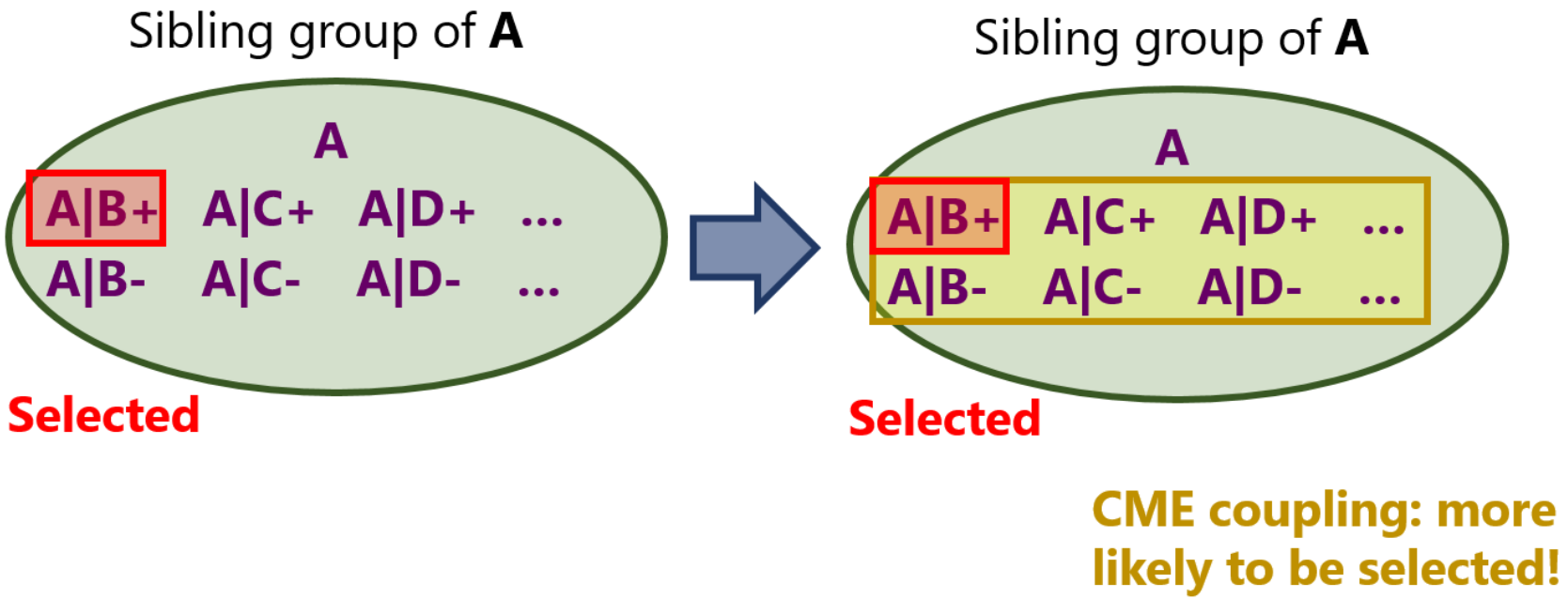


Figure 2: An illustration of CME coupling.

CME reduction

The **bi-level** formulation gives two appealing selection principles:

- **CME reduction**: If many siblings $A|B+$, $A|C+$, $\cdots$ are selected, its parent effect A may be active instead

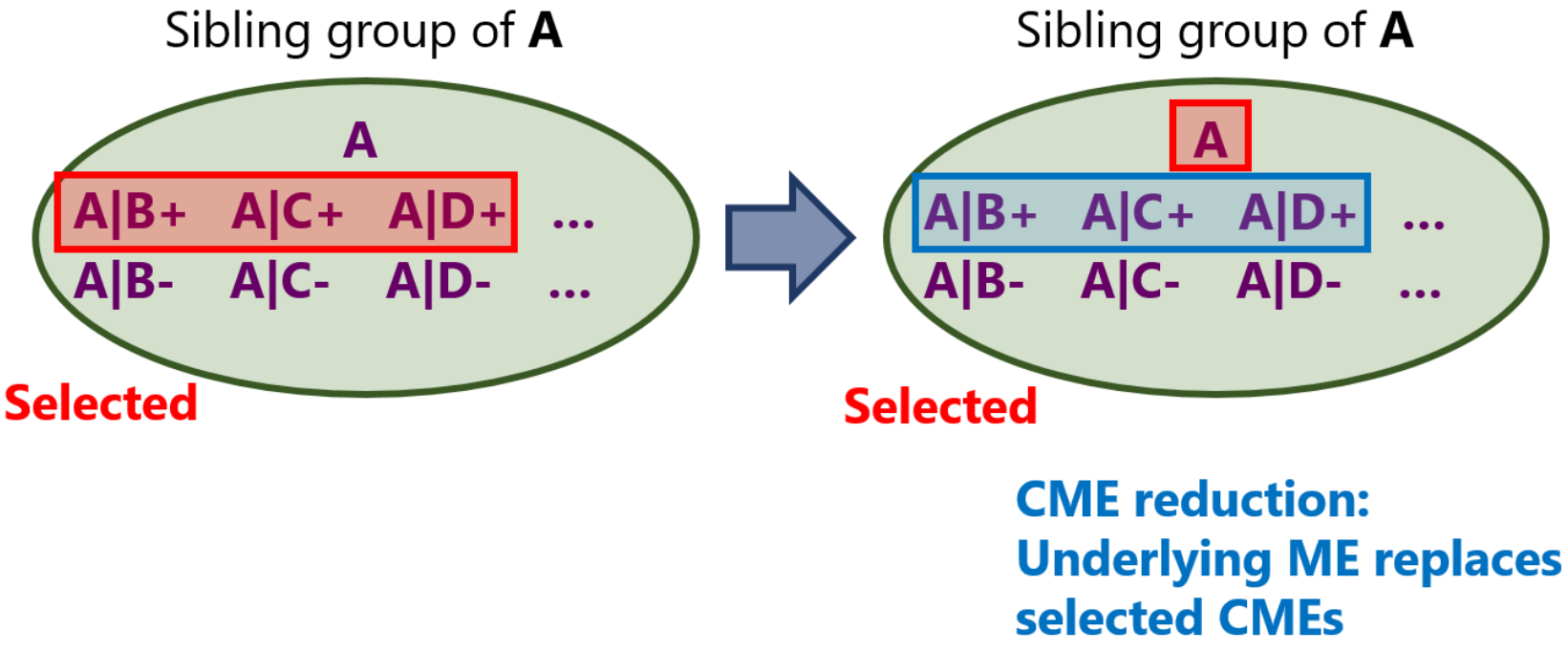


Figure 3: An illustration of CME reduction.

Connection to design principles

CME coupling and CME reduction can be viewed as extensions of **effect heredity** and **effect hierarchy** – two fundamental principles in factorial design (Wu and Hamada, 2009):

- **Effect heredity (weak)**: An interaction is present only when one of its components are active
  - CME coupling: heredity-like principle which encourages the selection of a CME when its siblings / cousins are in the model
- **Effect hierarchy**: Lower-order interactions are more likely active than higher-order ones
  - CME reduction: encourages reduction of selected siblings / cousins CMEs (higher-order) to its underlying ME (lower-order)
- `cmenet` extends these fundamental principles to the novel CMEs setting at hand

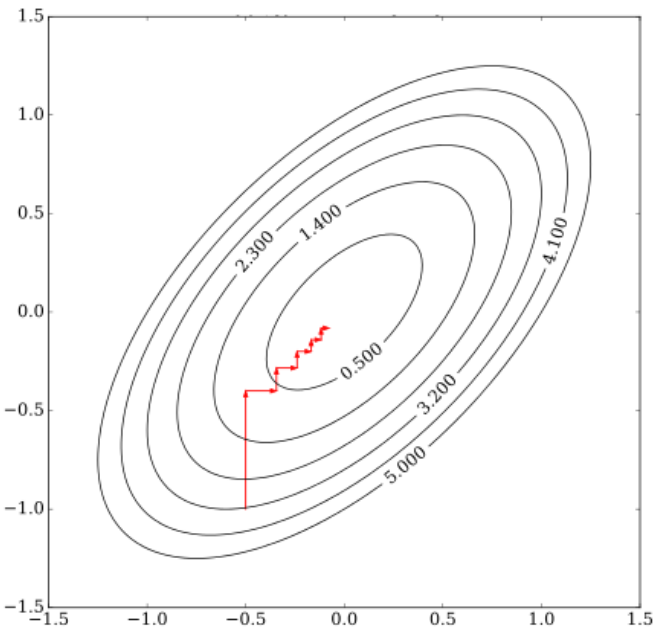
Section 4

Optimization & Simulations

Coordinate descent

We minimize $Q(\boldsymbol{\beta})$ using a technique called **coordinate descent** (Bertsekas, 1999):

- **Idea**: Cyclic minimization of $Q(\boldsymbol{\beta})$ for each variable $\beta_1, \beta_2, \cdots, \beta_p$ until $\boldsymbol{\beta}$ converges
- Very fast if this coordinate-wise optimization has **closed form** solution
- A first-order Taylor expansion of outer penalty f_o reduces the coordinate-wise problem to a LASSO-like problem:
  - Closed-form solution as a **threshold function**



`cv.cmenet`: Parameter tuning and speed-ups

- Parameters λ_s, λ_c, γ and τ tuned via **cross-validation**
  - Can be computationally expensive
- Three computational speed-ups for large problems:
  - **Warm starts:** Using previous coefficient sol'n to initialize current optimization
  - **Active sets:** Optimize only on a subset of potentially active variables
  - **Strong rules:** Use previous sol'ns to screen out inactive effects for current optimization

Simulation set-up

Simulation set-up:

- $(n, p) = (50, 50), (100, 100), (150, 150)$
  - $p' = 4950, 19900, 44850$ MEs & CMEs
- $\mathbf{X}$ simulated from latent Gaussian model, with correlation $\rho = 0$ or $1/\sqrt{2}$
- Active groups = siblings, cousins or MEs
- `cmenet` compared with:
  - **LASSO** (Tibshirani, 1996)
  - **SparseNet** (Mazumder et al., 2011)
  - `hierNet` (Bien et al., 2013) (state-of-the-art interaction method)

  All methods select the same MEs and CMEs
- Compared on:
  - # of misspecified effects: false-positives + true-negatives
  - Mean-squared prediction error (MSPE)

No correlation ($\rho = 0$)

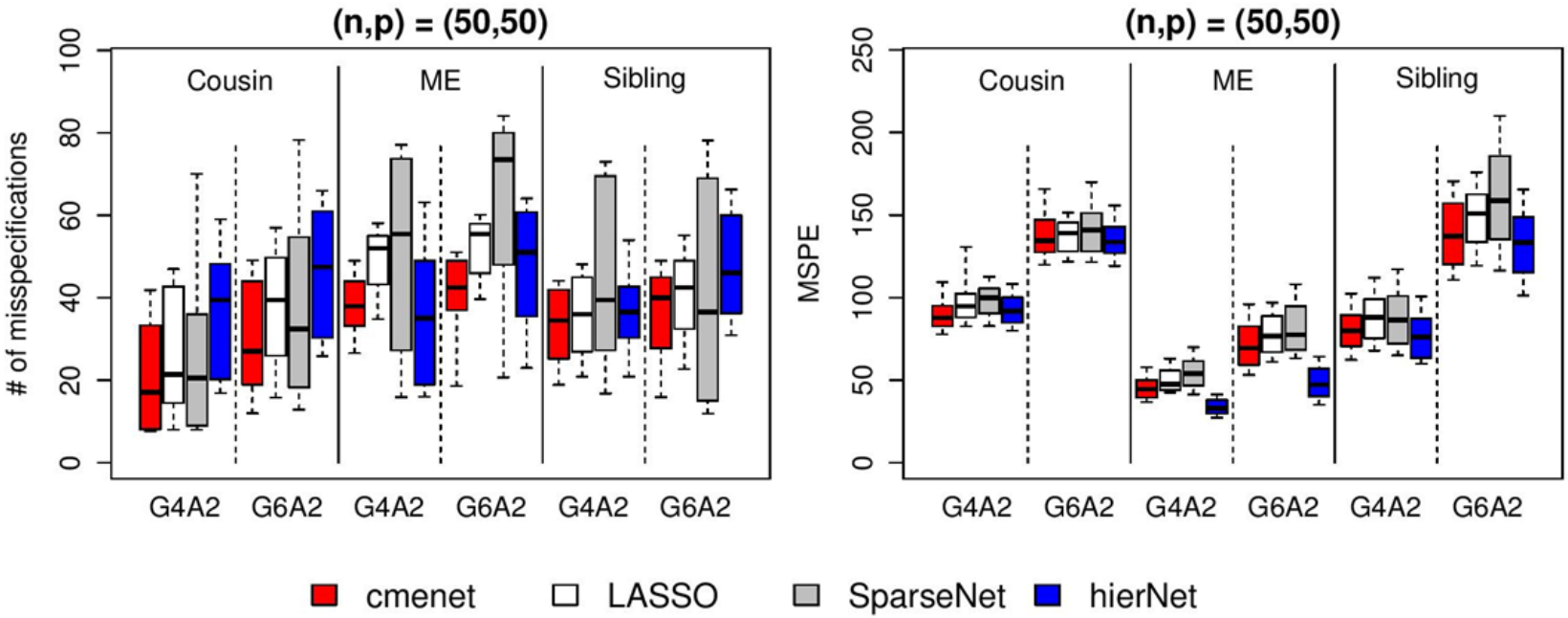


Figure 4: # of misspecifications and MSPE for $\rho = 0$. G4A2 means 4 active groups with 2 active effects in each (same for G6A2).

- For models with **active CMEs** (cousins or siblings), `cmenet` gives the best selection performance
- For models with only **active MEs**, `cmenet` comparable with `hierNet`
  - Not surprising: `cmenet` tackles CME selection

Moderate correlation ($\rho = 1/\sqrt{2}$)

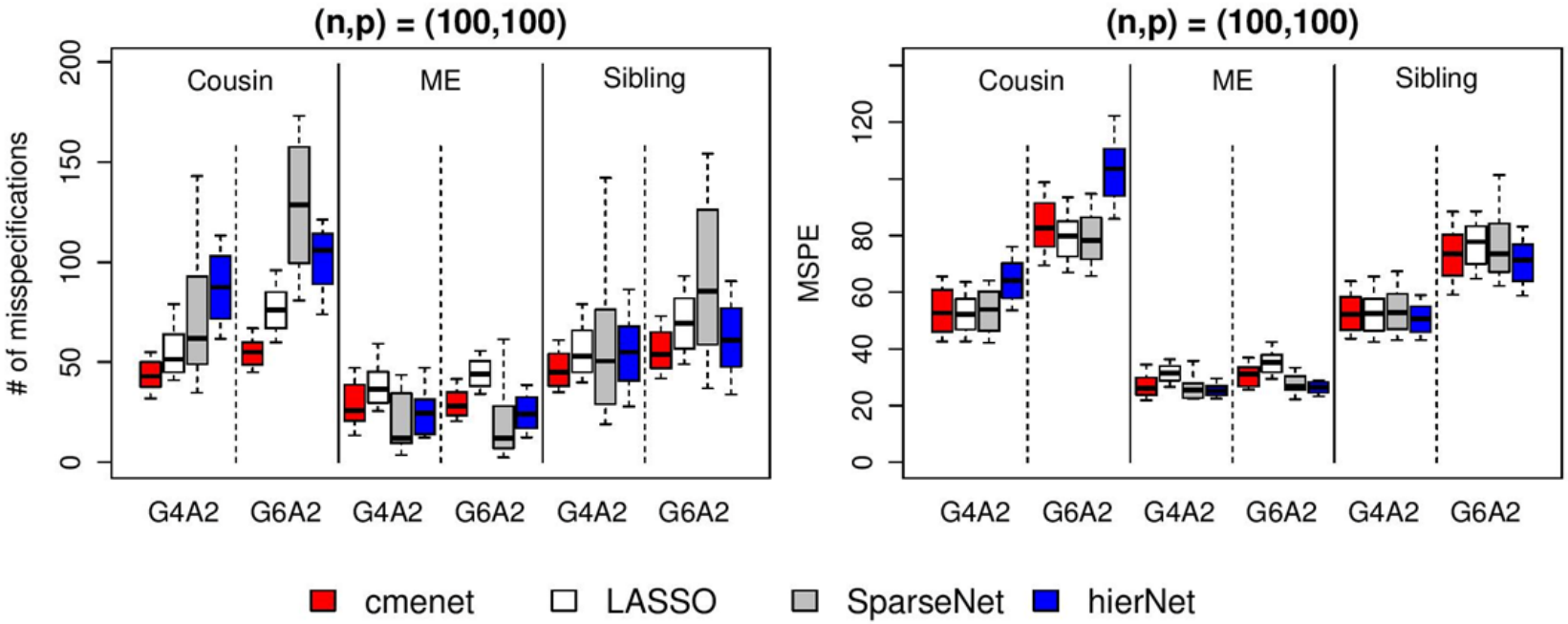


Figure 5: # of misspecifications and MSPE for $\rho = 1/\sqrt{2}$. G4A2 means 4 active groups with 2 active effects in each (same for G6A2).

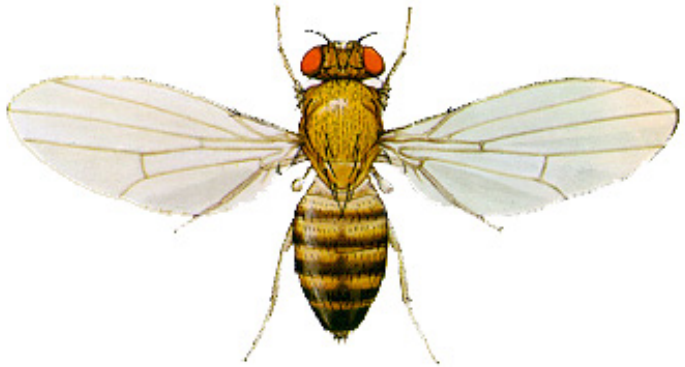
- For models with **active CMEs** (cousins or siblings), `cmenet` gives the best selection performance
  - Improvement gap much larger than for $\rho = 0$
  - CME group structure more prominent for large ρ
- For (only) **active ME** models, `cmenet` comparable with others

Section 5

Gene association study

www.redbubble.com/people/gardenicsalad/works/6325570-long-face

Gene association study

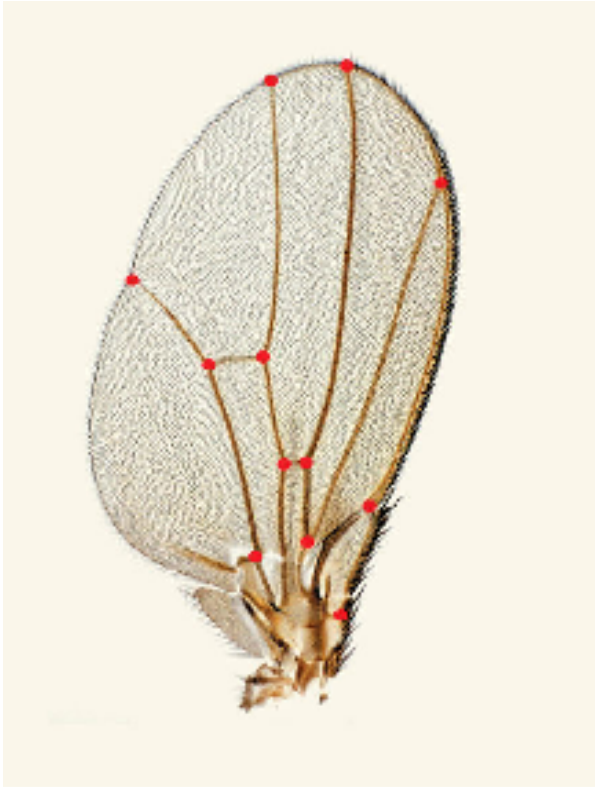


- **Single nucleotide polymorphisms** (SNPs) serve as biological markers for many organism characteristics
- `cmenet` can reveal **activation behavior** of gene-gene interactions:
  - Which genes are conditionally active?
  - Which genes activate other genes?
- We apply `cmenet` on a gene association study for the wing shape of *Drosophila Melanogaster*, the common fruit fly
  - $n = 701$ observations (fly wing shape indices)
  - $p = 48$ polygene markers
  - $p' = 4560$ MEs and CMEs

Gene association study: Set-up

- `cmenet` (selecting MEs and CMEs) is compared with:
  - **LASSO**
  - **SparseNet**
  - `hierNet`

  The latter three methods select MEs and 2FIs (no CMEs):
  - Standard approach for gene-gene interaction analysis
- Compared on:
  - MSPE (80% training, 20% testing)
  - Model interpretability



Gene association study: Predictive accuracy

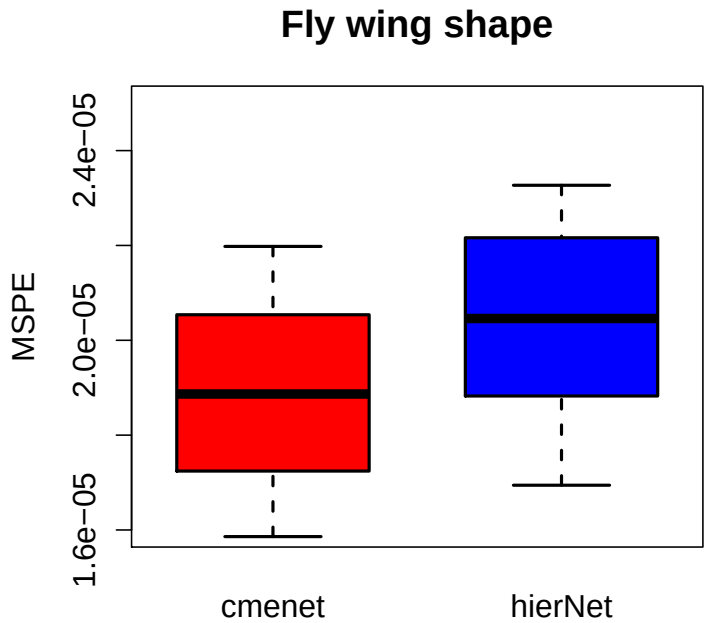


Figure 6: Out-of-sample MSPE boxplots for `cmenet` and `hierNet`

- `cmenet` gives lower MSPE to `hierNet` (state-of-the-art)
- LASSO and SparseNet have much higher MSPEs
- This suggests that the underlying gene structure is **conditional**:
  - Some genes are conditionally active
  - Some genes activate other genes

Gene association study: Gene selection

Method	*# of selected effects*	*Some selected effects (p-values)*	
`cmenet`	21	g14|g27- (6.1×10^{-4}), g14|g38+ (2.0×10^{-2}), g17|g14- (1.6×10^{-12}), g23|g14+ (2.5×10^{-30})	g45|g10+ (7.3×10^{-7})
`hierNet`	129	g14 (8.3×10^{-1})	g45 (1.5×10^{-1}), g45 $*$ g10 (8.1×10^{-1})

Table 3: Selected effects (p-values bracketed) from `cmenet` and `hierNet`.

- CME model from `cmenet` more parsimonious
- All three **standard** methods selected gene g14, whereas `cmenet` selected the CMEs g14|g27−, g14|g38+, g17|g14−, g23|g14+
  - `cmenet` gives a more **nuanced analysis** of g14:
    - conditionally active under genes g27- and g38+
    - activates gene g23 and inhibits gene g17
  - Greater insight into biological activation in gene-gene interactions

Gene association study: Gene selection

Method	*# of selected effects*	*Some selected effects (p-values)*	
`cmenet`	21	g14|g27- (6.1×10^{-4}), g14|g38+ (2.0×10^{-2}), g17|g14- (1.6×10^{-12}), g23|g14+ (2.5×10^{-30})	g45|g10+ (7.3×10^{-7})
`hierNet`	129	g14 (8.3×10^{-1})	g45 (1.5×10^{-1}), g45 $*$ g10 (8.1×10^{-1})

Table 3: Selected effects (p-values bracketed) from `cmenet` and `hierNet`.

- All three **standard** methods selected the ME g45 and the 2FI g45×g10, whereas `cmenet` selected only the CME g45|g10+:
  - Recall:
$$\mathrm{CME}(\mathrm{A}|\mathrm{B}+) = \frac{1}{2}\Big\{\mathrm{ME}(\mathrm{A}) + \mathrm{INT}(\mathrm{A},\mathrm{B})\Big\}$$
  - Replacing ME g45 and 2FI g45×g10 with CME g45|g10+ yields a more parsimonious and interpretable model
  - Rule 1 of Su and Wu (2017) for designed experiments

Summary

- CMEs are **interpretable** effects in many engineering and biological applications
- `cmenet` performs variable selection on CMEs in observational data, via the principles of CME coupling and CME reduction
- `cmenet` provides improved CME selection and better model interpretability over generic variable selection methods
- **R package** `cmenet` (out soon on CRAN)

Questions?