Center for Innovative Design & Analysis
colorado school of public health

"But what about interactions, are any of those significant?"
Resolving the collaborative nightmare of covariate interactions through regularization

RYAN A. PETERSON, PHD; JOSEPH E. CAVANAUGH, PHD

WHAT IS RANKED SPARSITY?

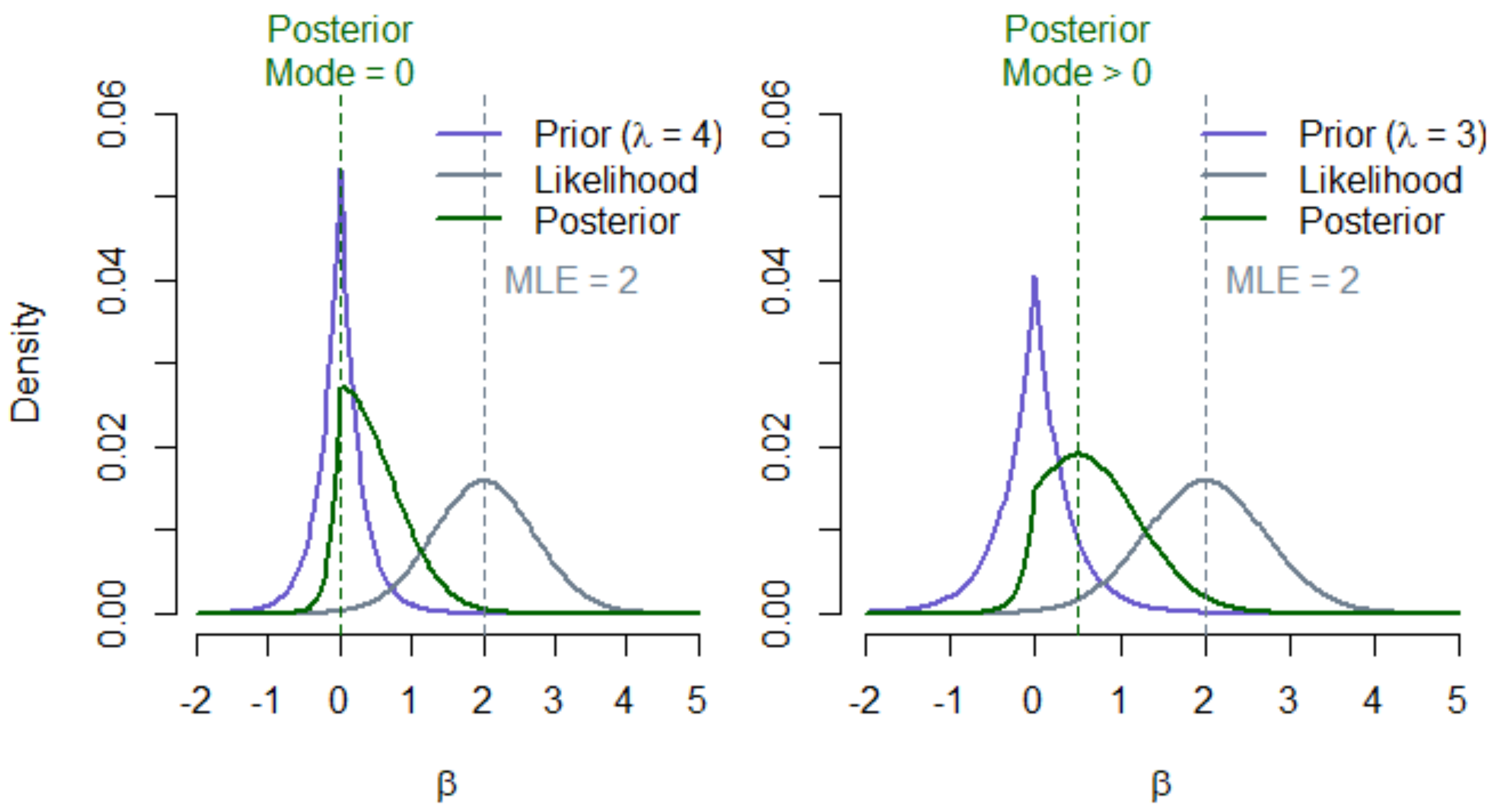
- Selecting interactions is an arduous process. The sheer number of them means one is at high risk of false discoveries, which if found, will necessitate a difficult interpretation of an unnecessarily opaque model.
- Ranked sparsity is the idea that some covariates deserve more skepticism *a priori* than others.
- In a model with polynomial terms or interactions, does presuming "covariate equipoise" make sense?
- Ranked sparsity allows us to require a higher degree of evidence for certain features to get selected, which is especially important for *derived* features.
- We have developed a framework that can adequately and flexibly account for disparities in the prior information contained in multiple covariate subspaces: *the sparsity-ranked lasso (SRL)*.
- Prioritizing main effects over derived variables, as visualized below, improves final model transparency.

$$
\begin{aligned}
E(y|X) = {} & \beta_0+\beta_1 x_1+\beta_2 x_2+\beta_3 x_3+\beta_4 x_4+\beta_5 x_5+\beta_6 x_6 + \beta_7 x_7+ \beta_8 x_8 \\
& + \beta_{11} x_1{}^2 + \beta_{22} x_2{}^2 + \beta_{33} x_3{}^2 + \beta_{44} x_4{}^2 + \beta_{55} x_5{}^2 + \beta_{66} x_6{}^2 + \beta_{77} x_7{}^2 + \beta_{88} x_8{}^2 \\
& + \beta_{111} x_1{}^3 + \beta_{222} x_2{}^3 + \beta_{333} x_3{}^3 + \beta_{444} x_4{}^3 + \beta_{555} x_5{}^3 + \beta_{666} x_6{}^3 + \beta_{777} x_7{}^3 + \beta_{888} x_8{}^3 \\
& + \beta_{12} x_1x_2 +\beta_{13} x_1x_3 +\beta_{14} x_1x_4 +\beta_{15} x_1x_5 +\beta_{16} x_1x_6 +\beta_{17} x_1x_7 +\beta_{18} x_1x_8 \\
& + \beta_{23} x_2x_3 +\beta_{24} x_2x_4 +\beta_{25} x_2x_5 +\beta_{26} x_2x_6 +\beta_{27} x_1x_7 +\beta_{28} x_2x_8+\beta_{34} x_3x_4 \\
& + \beta_{35} x_3x_5 +\beta_{36} x_3x_6 +\beta_{37} x_3x_7 +\beta_{38} x_3x_8 +\beta_{45} x_4x_5 +\beta_{46} x_4x_6 +\beta_{47} x_4x_7 \\
& + \beta_{48} x_4x_8 +\beta_{56} x_5x_6 +\beta_{57} x_5x_7 +\beta_{58} x_5x_8 + \beta_{67} x_6x_7 +\beta_{68} x_6x_8 +\beta_{78} x_7x_8 \\
& + \beta_{123} x_1x_2x_3 + \beta_{124} x_1x_2x_4 + \beta_{125} x_1x_2x_5 + \beta_{126} x_1x_2x_6 + \beta_{127} x_1x_2x_7 + \beta_{128} x_1x_2x_8 \\
& + \beta_{134} x_1x_3x_4 + \beta_{135} x_1x_3x_5 + \beta_{136} x_1x_3x_6 + \beta_{137} x_1x_3x_7 + \beta_{138} x_1x_3x_8 + \beta_{145} x_1x_4x_5 \\
& + \beta_{146} x_1x_4x_6 + \beta_{147} x_1x_4x_7 + \beta_{148} x_1x_4x_8 + \beta_{156} x_1x_5x_6 + \beta_{157} x_1x_5x_7 + \beta_{158} x_1x_5x_8 \\
& + \beta_{167} x_1x_6x_7 + \beta_{168} x_1x_6x_8 + \beta_{178} x_1x_7x_8 + \beta_{234} x_2x_3x_4 + \beta_{235} x_2x_3x_5 + \beta_{236} x_2x_3x_6 \\
& + \beta_{237} x_2x_3x_7 + \beta_{238} x_2x_3x_8 + \beta_{245} x_2x_4x_5 + \beta_{246} x_2x_4x_6 + \beta_{247} x_2x_4x_7 + \beta_{248} x_2x_4x_8 \\
& + \beta_{256} x_2x_5x_6 + \beta_{257} x_2x_5x_7 + \beta_{258} x_2x_5x_8 + \beta_{267} x_2x_6x_7 + \beta_{268} x_2x_6x_8 + \beta_{278} x_2x_7x_8 \\
& + \beta_{345} x_3x_4x_5 + \beta_{346} x_3x_4x_6 + \beta_{347} x_3x_4x_7 + \beta_{348} x_3x_4x_8 + \beta_{356} x_3x_5x_6 + \beta_{357} x_3x_5x_7 \\
& + \beta_{358} x_3x_5x_8 + \beta_{367} x_3x_6x_7 + \beta_{368} x_3x_6x_8 + \beta_{378} x_3x_7x_8 + \beta_{456} x_4x_5x_6 + \beta_{457} x_4x_5x_7 \\
& + \beta_{458} x_4x_5x_8 + \beta_{467} x_4x_6x_7 + \beta_{468} x_4x_6x_8 + \beta_{478} x_4x_7x_8 + \beta_{567} x_5x_6x_7 + \beta_{568} x_5x_6x_8 \\
& + \beta_{578} x_5x_7x_8 + \beta_{678} x_6x_7x_8
\end{aligned}
$$

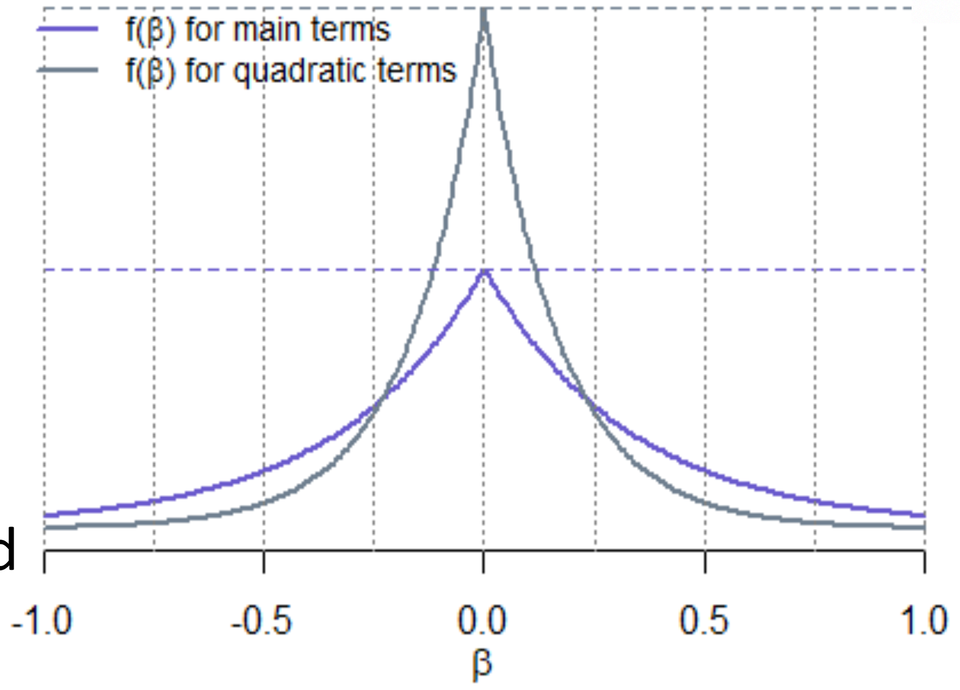
- With the SRL, coefficients are penalized based on the dimension of their "parent" covariate group.

DERIVATION

- A Bayesian motivation for the lasso (Tibshirani, 1996):



- We show that the *prior* Fisher information (based on the 2nd derivative of the joint prior distribution for our features) increases with the dimension of the parameter space.
- We can set the prior information for each covariate subspace to be equal by scaling the prior distributions separately.
- We can also further tune our prior distributions via functionalized sparsity patterns, where the prior information contribution of each subspace is formulated to coincide with potentially prominent features suggested by the phenomenon at hand.

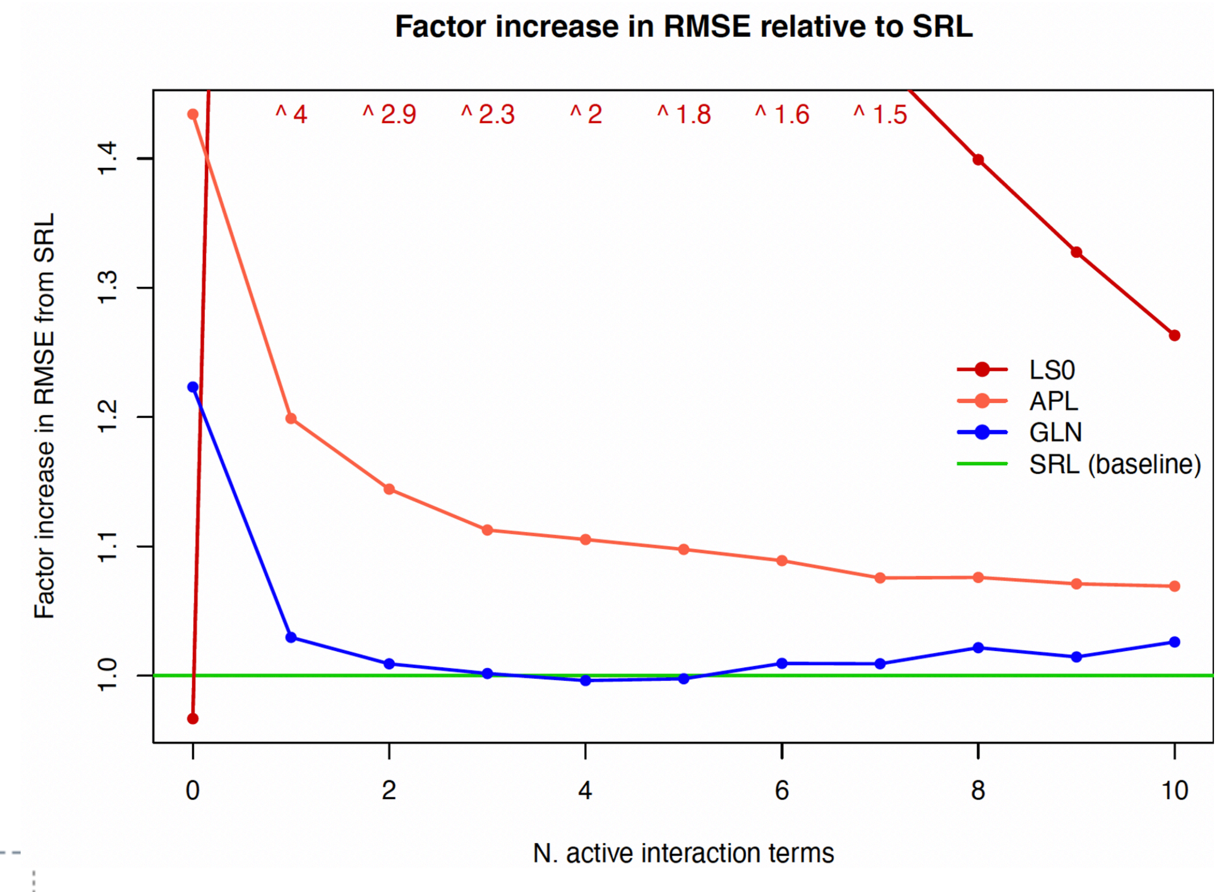


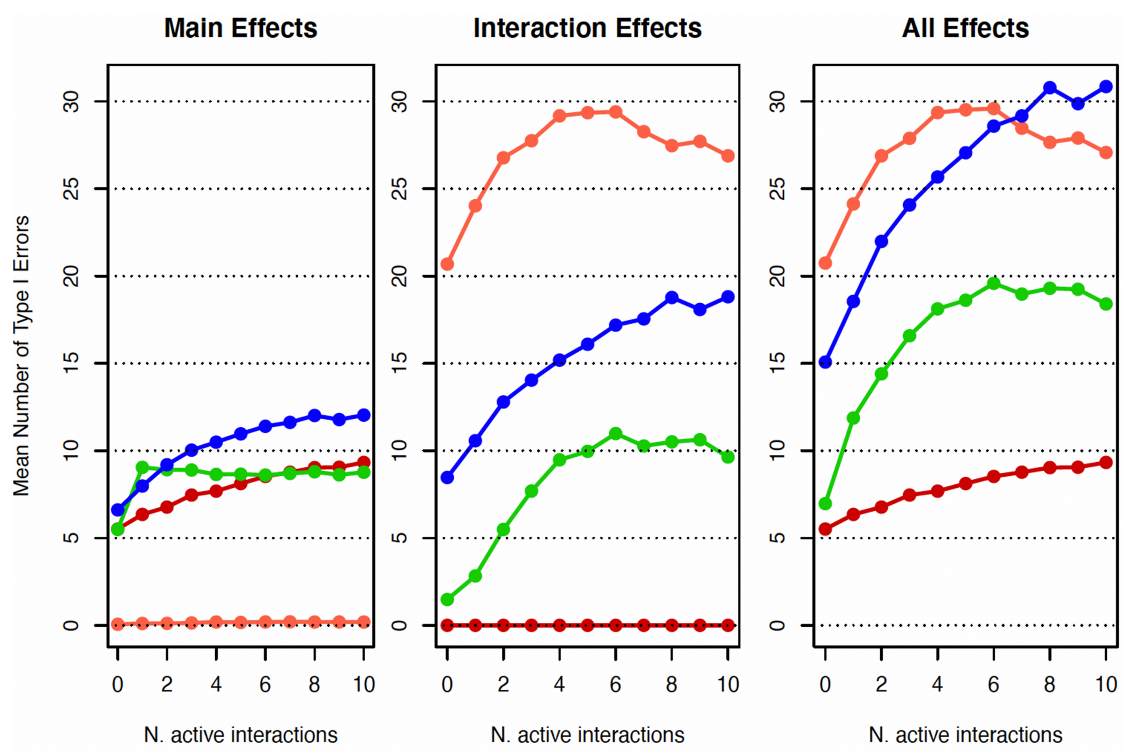
- If p_j refers to the dimension of a covariate's parent subspace, we the SRL is a scaling of the penalty by $\sqrt{p_j}$ or by $\sqrt{\sum_{k=1}^{j} p_k}$.

SIMULATION SET-UP

- Another method of using regularization to select important interactions is called "glinternet" (Lim and Hastie 2014).
- We generate data where between 0 and 10 interactions are truly "active" and compare the following fitting methods: LS0 (lasso on original terms only), APL (lasso with all pairwise terms), GLN (glinternet method), and SRL (sparsity-ranked lasso).

SIMULATION RESULTS





CONCLUSIONS

- The process of seeking out novel interactions and derived variables can be treacherous.
- Using ranked sparsity tools, you can be equipped to look for these without a preponderance of false discoveries.
- Compared to glinternet, the SRL predicts much better when there are no truly active interactions, and about the same when there are active interactions.
- SRL is optimal in terms of Type I error, especially for interactions.