

EAAP 2018 – 69th Annual Meeting, Dubrovnik, Croatia, 29 August 2018

A sparse-group lasso variant for whole-genome regression models in half sibs

Jan Klosa¹, Noah R. Simon², Dörte Wittenburg¹

¹ Leibniz Institute for Farm Animal Biology, Dummerstorf, Germany

² University of Washington, Seattle, WA, USA



LEIBNIZ INSTITUTE
FOR FARM ANIMAL BIOLOGY

DFG Deutsche
Forschungsgemeinschaft
German Research Foundation

Member of
Leibniz
Leibniz
Association

Background

“phenotype = genotype + environment”

Predict future
performance
of animals

- Estimate genotypic effects
- “ $p > n$ ”
- Selection and shrinkage approach

Objectives

Develop a statistical approach that jointly allows to:

- 1 **Exploit relationships** among predictors
- 2 Achieve high-precision **estimation of breeding values**
- 3 **Identify causative regions** on the genome



Whole-genome regression model

- Model equation:

$$y = X\beta + e$$

y phenotype vector

X marker-genotype matrix

β vector of additive-genetic effects

e vector of i.i.d. residuals

- Optimization problem (n -invariant formulation):

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \text{penalty} \right\}$$

What is the regression model

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \text{penalty} \right\}$$

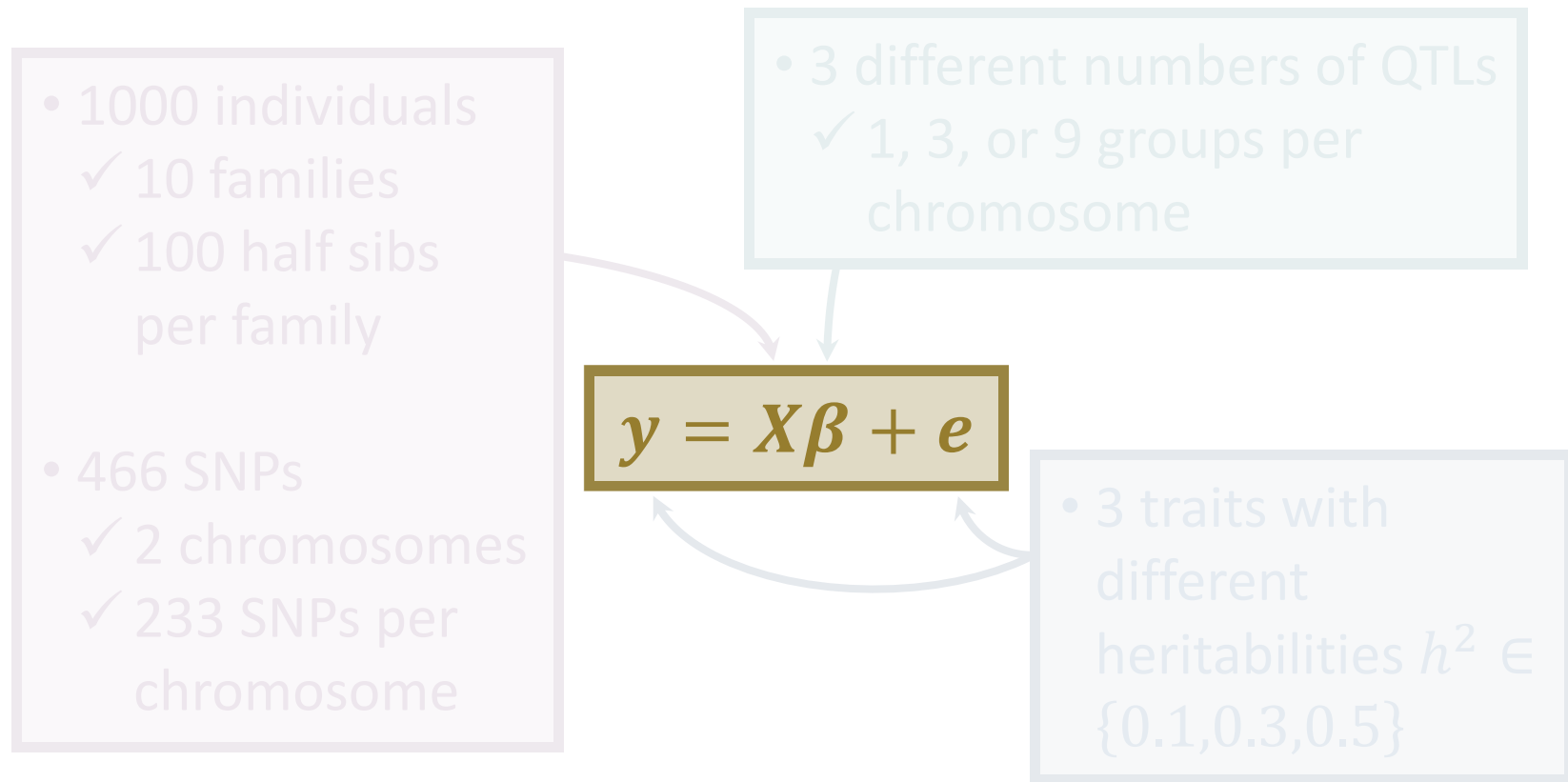
Name	Penalty term
1 lasso (Tibshirani, 1996)	$\lambda \ \beta\ _1$
2 group lasso (Yuan and Lin, 2007)	$\lambda \sum_l \sqrt{p_l} \ \beta^{(l)}\ _2$
3 sparse-group lasso (Simon et al., 2013)	$\alpha \lambda \ \beta\ _1 + (1 - \alpha) \lambda \sum_l \sqrt{p_l} \ \beta^{(l)}\ _2$
4 modified sparse-group lasso $\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{2n} \ y - X\beta\ _2^2 + \text{penalty} \right\}$	$\alpha \lambda \ \beta\ _1 + (1 - \alpha) \lambda \sum_l \sqrt{p_l} g_{\varepsilon} \left(\ X^{(l)} \beta^{(l)}\ _2 \right)$ <small>g_{ε}... Moreau envelope</small>

Computational implementation

$$\hat{\beta} = \arg \min_{\beta} \left\{ \frac{1}{2n} \|y - X\beta\|_2^2 + \text{penalty} \right\}$$

- | | Name | Penalty term |
|---|--|---|
| • | Iterative algorithm: | Proximal gradient descent |
| ① | lasso
(Tibshirani, 1996) | $\lambda \ \beta\ _1$ |
| ② | group lasso
(Yuan and Lin, 2007) | $\lambda \sum_l \sqrt{p_l} \ \beta^{(l)}\ _2$ |
| • | Moreau envelope to circumvent | potential troublesome |
| ③ | sparse-group lasso
(Simon et al., 2013) | $\alpha \lambda \ \beta\ _1 + (1 - \alpha) \lambda \sum_l \sqrt{p_l} \ \beta^{(l)}\ _2$ |
| ④ | modified sparse-group lasso | $g_{\varepsilon}(z) := \begin{cases} \alpha \lambda \ z\ _2 + (1 - \alpha) \lambda \sum_l \sqrt{p_l} g_{\varepsilon}(\ X^{(l)} \beta^{(l)}\ _2) \\ \frac{\ z\ _2^2}{2\varepsilon} + \frac{\varepsilon}{2} g_{\varepsilon}(\ z\ _2) \end{cases}$ |

Simulated data

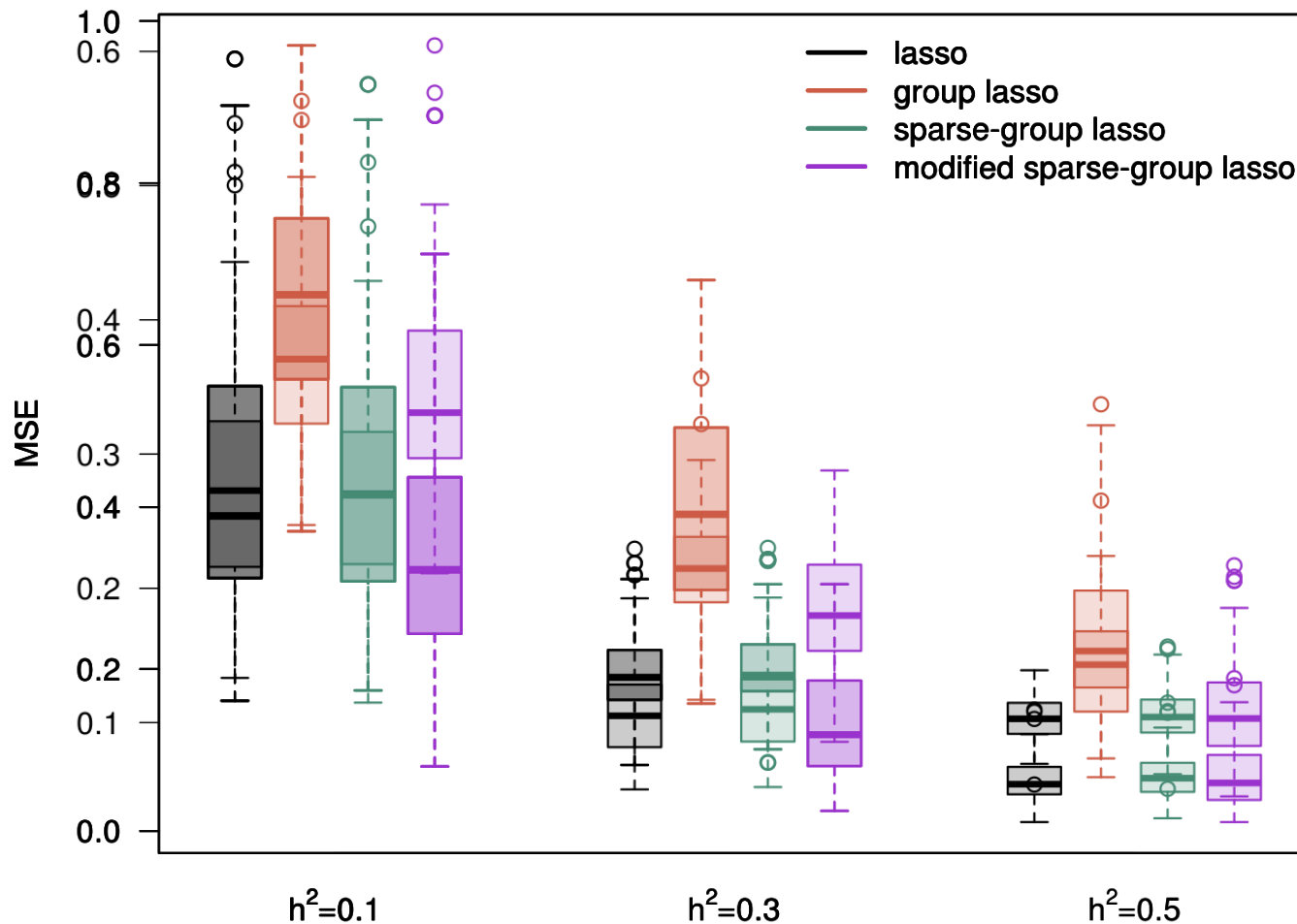


For statistical evaluation: 100 experiments in total

Evaluation criteria

- 1 Accuracy of estimation of variance components
(additive-genetic and residual variance)
- 2 Accuracy of prediction of breeding values (BVs)
(correlation between true and predicted BVs)
- 3 Sensitivity and specificity of QTL detection
- 4 **Mean squared errors: $MSE(\hat{\beta})$**

Mean squared errors



- **3 groups of** QTLs per chromosome
- $\alpha = 0.9$

Discussion and outlook

- ① Similarity measures within the grouping:
 - Population-wide linkage disequilibrium
 - Theoretical correlation (Wittenburg et al., 2016)

- ② Modified sparse-group lasso:
 - Choose ε for $g_\varepsilon \left(\|X^{(l)} \beta^{(l)}\|_2 \right)$
 - $\|X^{(l)} \beta^{(l)}\|_2 \rightarrow \sqrt{\langle \beta^{(l)}, P^{(l)} \beta^{(l)} \rangle_2}$
(P ... precision matrix)

Summary

With respect to the MSE and compared to the lasso:

- 1 Modified sparse-group lasso: severe improvements if 1 group of QTLs per chromosome
- 2 Sparse-group lasso: no or small improvements if 3 or more groups of QTLs per chromosome
- 3 Group lasso: no improvements

Thank you very much for your attention.





LEIBNIZ INSTITUTE
FOR FARM ANIMAL BIOLOGY



Dummerstorf

Leibniz Institute for Farm Animal Biology FBN

Leibniz Institute for Farm Animal Biology FBN

Wilhelm-Stahl-Allee 2
18196 Dummerstorf
Germany

Contact

Jan Klosa

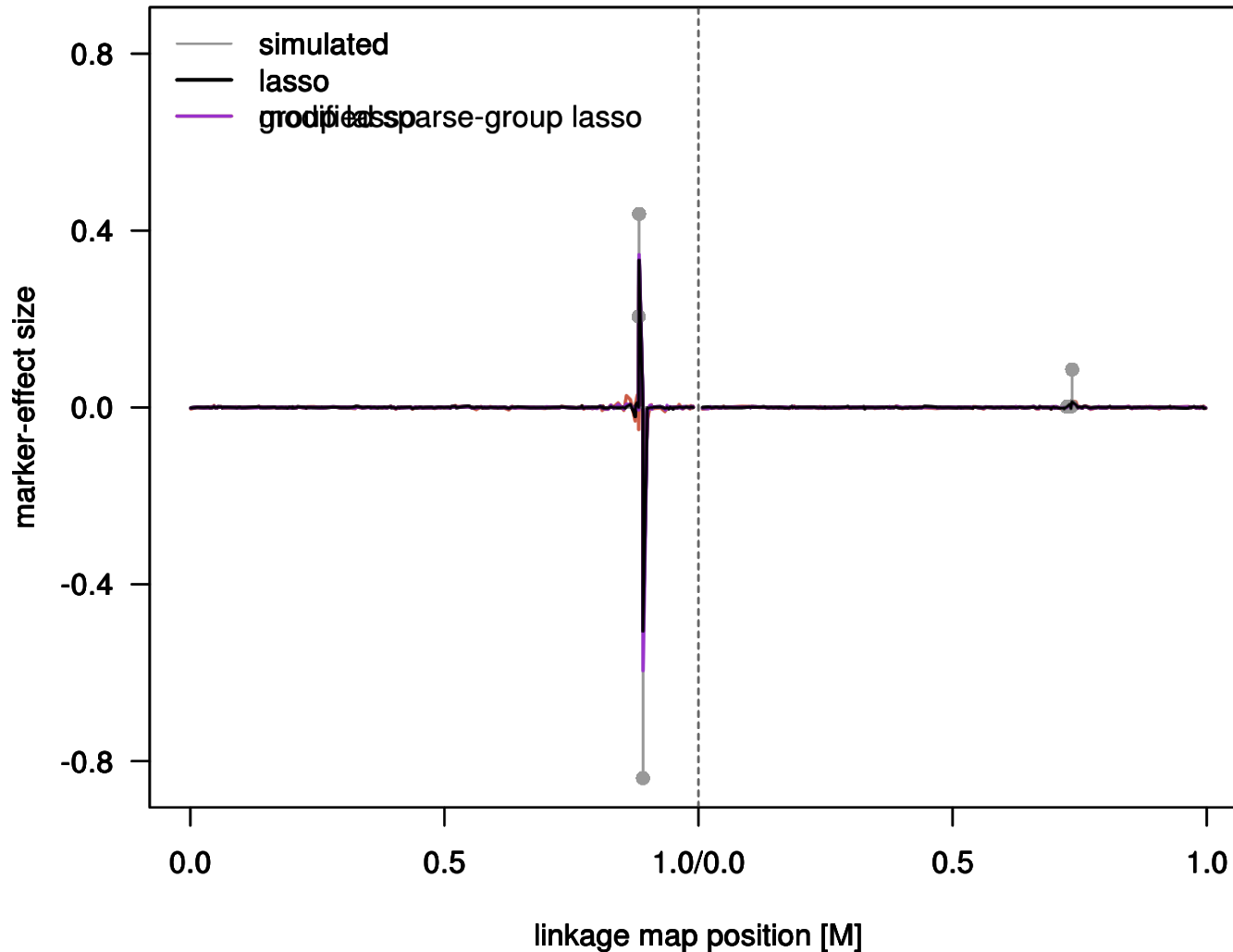
Phone: +49 38208 68 918

Fax: +49 38208 68 902

E-Mail: klosa@fbn-dummerstorf.de

Internet: www.fbn-dummerstorf.de

Average marker-effect estimation



- 1 group of QTLs per chromosome
- $h^2 = 0.3$
- $\alpha = 0.9$