

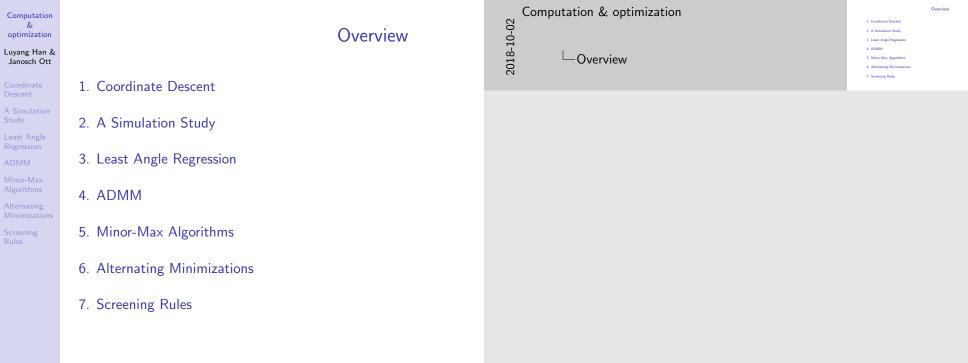
# 2018-10

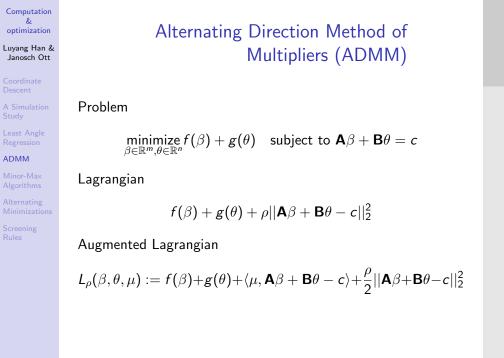
Computation & optimization for Lasso - part 2

Luyang Han & Janosch Ott

22 October 2018

Computation & optimization







-ADMM

Augmented: scalar product with  $\mu$  gets added

 $L_{\mu}(\beta, \theta, \mu) := f(\beta)+g(\theta)+(\mu, \mathbf{A}\beta + \mathbf{B}\theta - c)+\frac{\rho}{2}||\mathbf{A}\beta + \mathbf{B}\theta - c||_{2}^{2}$ 

Alternating Direction Method of

minimize  $f(\beta) + g(\theta)$  subject to  $\mathbf{A}\beta + \mathbf{B}\theta = c$ 



Luyang Han &

# Dual variable update

Janosch Ott

**ADMM** 



 $heta^{t+1} = \arg\min_{ heta \in \mathbb{R}^m} L_{
ho}(eta^{t+1}, heta, \mu^t)$ 

$$\mu^{t+1} = \mu^t + \rho(\mathbf{A}\beta^{t+1} + \mathbf{B}\theta^{t+1} - c)$$

Computation & optimization 2018-10-02 -ADMM └─Dual variable update

 $\beta^{t+1} = \arg \min_{\alpha, \dots, \mu} L_{\rho}(\beta, \theta^t, \mu^t)$  $\theta^{t+1} = \arg \min_{\rho = 0} L_{\rho}(\beta^{t+1}, \theta, \mu^t)$  $\mu^{t+1} = \mu^t + \rho(\mathbf{A}\beta^{t+1} + \mathbf{B}\theta^{t+1} - c)$ 

Dual variable update

Luyang Han &

Janosch Ott

ADMM - Why?

- Coordina Descent
- A Simulation Study
- Least Angl

. \_ . . . .

#### ADMM

Minor-Max Algorithms

Alternating

Screening

- convex problems with nondifferentiable constraints
- blockwise computation
  - sample blocks
  - feature blocks

Computation & optimization CONTROL COMPUTATION ADMM

COMPUTATION ADMM - Why?

convex problems with nondifferentiable constraints
 blockwise computation
 sample blocks
 feature blocks

ADMM - Why?



Computation

#### ADMM for the Lasso

**ADMM** 





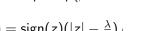




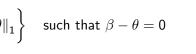
$$\mu^{t+1} = \mu^t + \rho(\beta^{t+1} - \theta^{t+1})$$

Problem in Lagrangian form

where 
$$S_{\lambda/\rho}(z) = \text{sign}(z)(|z| - \frac{\lambda}{\rho})_+$$
.



 $\theta^{t+1} = \mathcal{S}_{\lambda/\rho}(\beta^{t+1} + \mu^t/\rho)$ 



# Update

$$\in \mathbb{R}^p \left( 2^{n^2} \right)^{n^2}$$

 $\beta^{t+1} = (\mathbf{X}^T \mathbf{X} + \rho \mathbf{I})^{-1} (\mathbf{X}^T \mathbf{y} + \rho \theta^t - \mu^t)$ 





## Computation & optimization 2018-10-02 -ADMM





ADMM for the Lasso

Computational cost: Initially  $\mathcal{O}(p^3)$ , which is a lot, for the SVD(singular value decomposition of X), after that comparable to coordinate descent or composite gradient from earlier

Luyang Han & Janosch Ott

Minor-Max Algorithms

### Minorization-Maximization Algorithms (MMA)

- Problem: minimize  $f(\beta)$  over  $\beta \in \mathbb{R}^p$ for *f* possibly non-convex
- Introduce additional variable  $\theta$
- Use  $\theta$  to majorize (bound from above) the objective function to be minimized

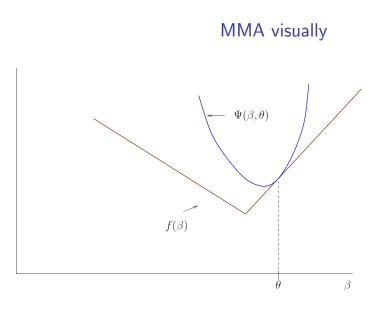
Majorization-Minimization Algorithms work analoguosly.

Computation & optimization 10-02 Minor-Max Algorithms 2018-1

☐ Minorization-Maximization Algorithms (MMA)

Minorization-Maximization Algorithms (MMA)

- Problem: minimize f(β) over β ∈ R<sup>β</sup>
- Introduce additional variable θ
- Use θ to majorize (bound from above) the objective
- Majorization-Minimization Algorithms work analoguosi



Computation

optimization

Luyang Han &

Janosch Ott

A Simulation

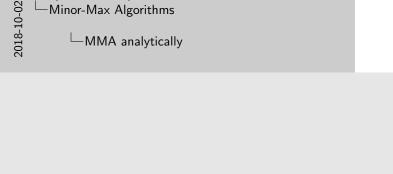
Minor-Max Algorithms

Screening

Figure: Figure 5.10 from [Hastie et al., 2015]







MMA analytically

Computation & optimization

Dual Polytope Projection (DPP)

Luyang Han & Janosch Ott

Screening

Rules

Suppose we want to calculate a lasso solution at  $\lambda < \lambda_{max}$ . The DPP rule discards the *i*<sup>th</sup> variable if

$$\left\|\mathbf{x}_{j}^{\mathcal{T}}\mathbf{y}\right\|<\lambda_{\mathsf{max}}-\left\|\mathbf{x}_{j}
ight\|_{2}\left\|\mathbf{y}
ight\|_{2}rac{\lambda_{\mathsf{max}}-\lambda}{\lambda}$$

Sequential DPP rule

Suppose we have the lasso solution  $\hat{\beta}(\lambda')$  at  $\lambda'$  and want to screen variables for solutions at  $\lambda < \lambda'$ . We discard the  $i^{th}$ variable if

$$\left|\mathbf{x}_{j}^{T}(\mathbf{y} - \mathbf{X}\hat{eta}(\lambda'))\right| < \lambda' - \left\|\mathbf{x}_{j}\right\|_{2} \left\|\mathbf{y}\right\|_{2} rac{\lambda_{\mathsf{max}} - \lambda}{\lambda}$$

Computation & optimization -Screening Rules

□ Dual Polytope Projection (DPP)

 $\left|\mathbf{x}_{j}^{T}\mathbf{y}\right|<\lambda_{\max}-\left\|\mathbf{x}_{j}\right\|_{2}\left\|\mathbf{y}\right\|_{2}\frac{\lambda_{\max}-\lambda}{\epsilon}$ 

Sequential DPP rule

Suppose we have the lasso solution  $\hat{\beta}(\lambda')$  at  $\lambda'$  and want to screen variables for solutions at  $\lambda < \lambda'$ . We discard the  $i^{ti}$ 

 $|\mathbf{x}_{i}^{T}(\mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}(\lambda'))| < \lambda' - ||\mathbf{x}_{i}||_{2} ||\mathbf{y}||_{2} \frac{\lambda_{\max} - \lambda}{2}$ 

Dual Polytope Projection (DPP)



#### Global Strong Rule

oordina escent

Simulation

Least Angl

Regressi

ADMM

Algorithms

Alternating Minimization

Screening Rules Suppose we want to calculate a lasso solution at  $\lambda < \lambda_{\text{max}}$ . The global strong rule discards the  $j^{th}$  variable if

$$\left|\mathbf{x}_{j}^{T}\mathbf{y}\right|<\lambda-\left(\lambda_{\mathsf{max}}-\lambda\right)=2\lambda-\lambda_{\mathsf{max}}$$

### Sequential Strong Rule

Suppose we have the lasso solution  $\hat{\beta}(\lambda')$  at  $\lambda'$  and want to screen variables for solutions at  $\lambda < \lambda'$ . We discard the  $j^{th}$  variable if

$$\left|\mathbf{x}_{j}^{T}(\mathbf{y}-\mathbf{X}\hat{eta}(\lambda'))\right|<2\lambda-\lambda'$$



#### Global Strong Rule

uppose we want to calculate a lasks solution at  $\lambda$ he global strong rule discards the  $j^{th}$  variable if  $\left|\mathbf{x}_{j}^{T}\mathbf{y}\right| < \lambda - \left(\lambda_{\max} - \lambda\right) = 2\lambda - \lambda_{\max}$ 

Sequential Strong Rule

Suppose we have the lasso solution  $\hat{\beta}(\lambda')$  at  $\lambda'$  and want to screen variables for solutions at  $\lambda < \lambda'$ . We discard the  $j^{\text{th}}$  variable if  $\left|\mathbf{x}_j^T(\mathbf{y} - \mathbf{X}\hat{\beta}(\lambda'))\right| < 2\lambda - \lambda'$ 

Luyang Han &

Paragraphs of Text

Janosch Ott

Simulatio

Least Ang Regressio

ADMM Minor-Max

ternating

Minimization Screening

Rules

Sed diam enim, sagittis nec condimentum sit amet, ullamcorper sit amet libero. Aliquam vel dui orci, a porta odio. Nullam id suscipit ipsum. Aenean lobortis commodo sem, ut commodo leo gravida vitae. Pellentesque vehicula ante iaculis arcu pretium rutrum eget sit amet purus. Integer ornare nulla quis neque ultrices lobortis. Vestibulum ultrices tincidunt libero, quis commodo erat ullamcorper id.

Sed iaculis dapibus gravida. Morbi sed tortor erat, nec interdum arcu. Sed id lorem lectus. Quisque viverra augue id sem ornare non aliquam nibh tristique. Aenean in ligula nisl. Nulla sed tellus ipsum. Donec vestibulum ligula non lorem

vulputate fermentum accumsan neque mollis.

Computation & optimization

Paragraphs of Text

#### Paragraphs of Text

Sed isculis dapibus gravida. Morbi sed tortor erat, nec interdum arcu. Sed id forem hectus. Quisque viverra augue id sem ornare non aliquam nibh tristique. Annean in ligula nisl. Nulla sed tellus ipsum. Donec vestibulum ligula non lorem vulputate fermentum accumsan neque mollis.

Sed d'am enim, sajettis nec condimentum sit amer, ulluncops ist amet Bhero. Alquam vel dis cris, a post acifo. Null'acim vel suscipit l'ipsum. Amesan laborisi commodo sem, ut commodo los gavado vitas. Pallentesque velicitula ante iscalis sem pertium ratrum eget sit amet parse. Integer omare nulla quis neque attrices bloristi. Vestiblum ritriosi sincidunt libero, quis commodo erat ulluncorper id.

Luyang Han &

Janosch Ott

**Bullet Points** 

Coordina

Study

rtegress

Minor M

Alternating

Minimizatio

Screening Rules

- Lorem ipsum dolor sit amet, consectetur adipiscing elit
- Aliquam blandit faucibus nisi, sit amet dapibus enim tempus eu
- Nulla commodo, erat quis gravida posuere, elit lacus lobortis est, quis porttitor odio mauris at libero
- Nam cursus est eget velit posuere pellentesque
- Vestibulum faucibus velit a augue condimentum quis convallis nulla gravida

 $\begin{array}{c} \text{Computation \& optimization} \\ \text{--Screening Rules} \\ \text{--Bullet Points} \end{array}$ 

Bullet Points

Lorem ipsum dolor sit amet, consectetur adipiscing elit
 Aliquam blandit faucibus nisi, sit amet dapibus enim

 Nulla commodo, erat quis gravida posuere, elit lacus lobortis est, quis porttitor odio mauris at libero

Nam cursus est eget velit posuere pellentesque

 Vestibulum faucibus velit a augue condimentum quis convalis nulla gravida

Computation Blocks of Highlighted Text optimization Luyang Han & Janosch Ott Block 1 Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer lectus nisl, ultricies in feugiat rutrum, porttitor sit amet augue. Aliquam ut tortor mauris. Sed volutpat ante purus, quis accumsan dolor. Block 2 Pellentesque sed tellus purus. Class aptent taciti sociosqu ad litora torquent per conubia nostra, per inceptos himenaeos. Screening Vestibulum quis magna at risus dictum tempor eu vitae velit. Rules Block 3 Suspendisse tincidunt sagittis gravida. Curabitur condimentum, enim sed venenatis rutrum, ipsum neque consectetur orci, sed blandit justo nisi ac lacus.

#### Block 1 Lorem iosum dolor sit arnet, consectetur adioiscine elit, Interesr

Lorem ipsum dolor sit amest, consectetur adipiscing elit. Integende lectus nisil, ultriciss in feugiat nutrum, portitior sit amest augue Aliquam ut tortor mauris. Sed volutpat ante punus, quis accumsan dolor.

Block 2
Pellentesque sed tellus purus. Class aptent taciti sociosqu ad litora torquent per consibia nostra, per inceptos himenaeos. Vestibulum quis magna at risus dictum tempor eu vitae velit

Blocks of Highlighted Text

Block 3

Computation & optimization

-Blocks of Highlighted Text

-Screening Rules

Block 3 Suspendisse tincidunt sagittis gravida. Curabitur condimentun enim sed venenatis rutrum, ipsum neque consectetur orci, sed blandit justo nisi ac lacus.

Luyang Han & Janosch Ott

Descent

Study

Heading

Statement

2 Explanation

3 Example

Regressi

ADMM

Minor-Max

Alternating

Screening Rules

### Multiple Columns

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Integer lectus nisl, ultricies in feugiat rutrum, porttitor sit amet augue. Aliquam ut tortor mauris. Sed volutpat ante purus, quis accumsan dolor.

Computation & optimization
Screening Rules

Multiple Columns

Heading

Statement

Explanation

Example

Lorem ipsum dolor sit amet, consectatur adipiccing elft. Integer lectus nist, ultricies ir feugiat rutrum, portifor sit amet augue. Aliquam ut tort mauris. Sed volutpat ante purus, quis accumsan dolor.

Multiple Columns



Table

Janosch Ott

Alternating

Screening Rules

Treatment 3

Treatment 1 Treatment 2

Table: Table caption

**Treatments Response 1** Response 2 0.0003262 0.562 0.0015681 0.910 0.296 0.0009271

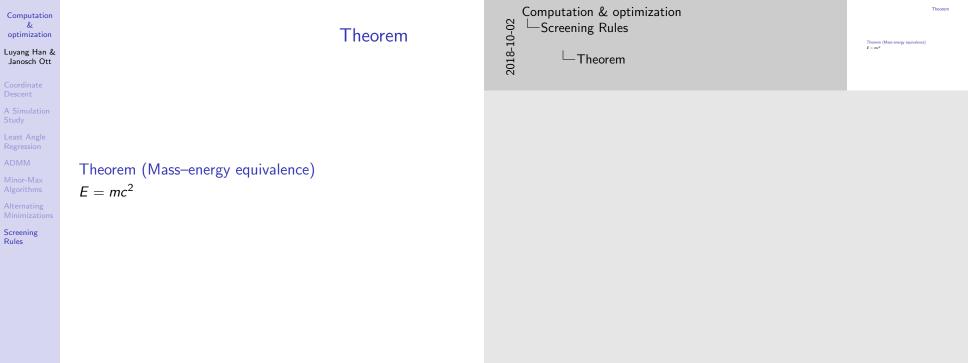
└─Table

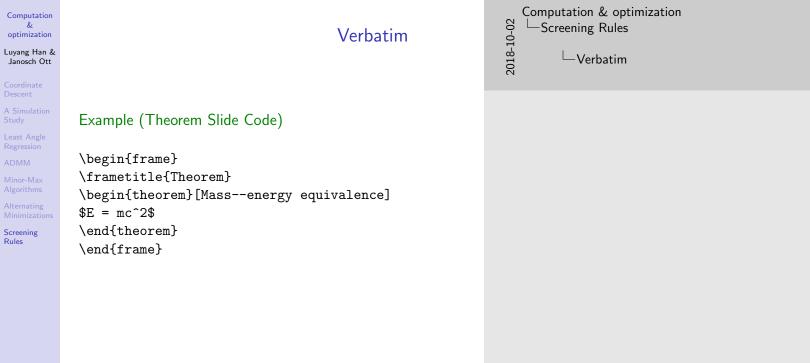
-Screening Rules

2018-10-02

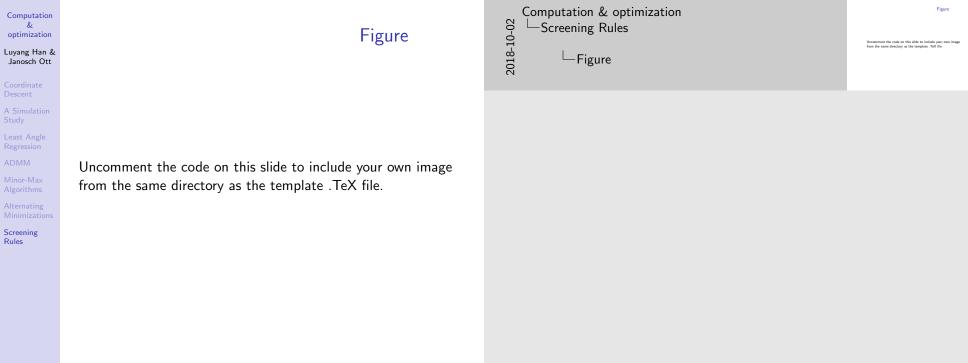
Computation & optimization

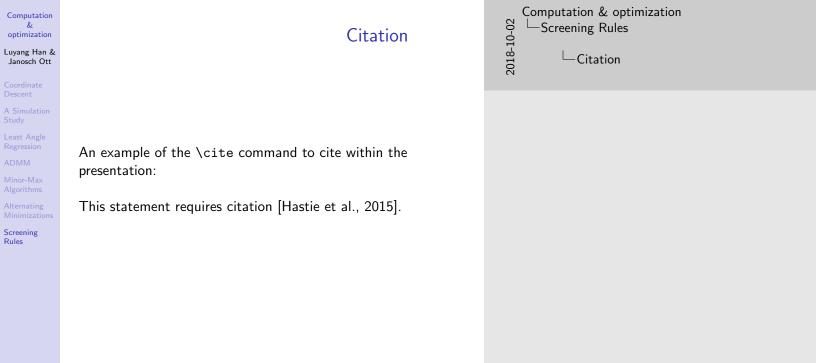
Treatments Response 1 Response 2 Treatment 1 0.0003262 0.562 Treatment 2 0.0015681 0.910 Treatment 3 0.0009271 0.296 Table: Table caption





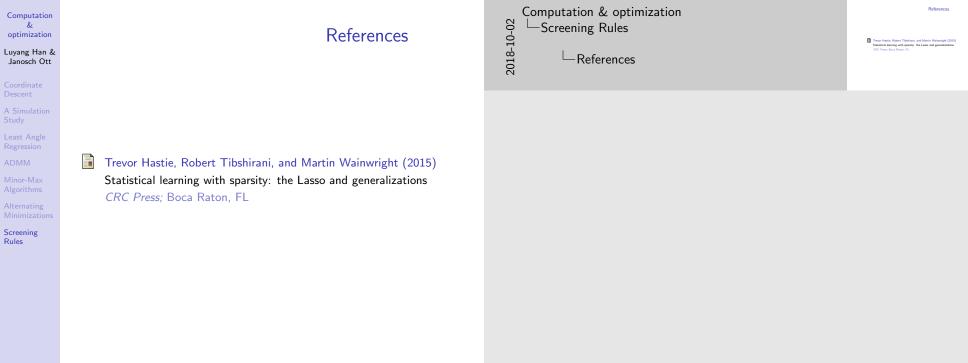
Verbatim





An example of the \cite command to cite within the

This statement requires citation [Hastie et al., 2015].





# Screening Rules

The End

Computation & optimization