Linear models I: linear regression and linear mixed models

Christoph Lippert¹ Oliver Stegle²

 1 Microsoft Research, Los Angeles, USA 2 Max-Planck-Institutes Tübingen, Germany



Basel 09. September 2012 Research

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 1

Outline

Outline

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 2

▲□▶ ▲圖▶ ▲≣▶ ▲≣▶ 三重 - のへの

Outline

Linear Regression

Hypothesis Testing

Multiple Hypothesis Testing

Population Structure

Population structure correction

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

▶ < ≣ ▶ ≣ ∽ Q ⊂ September 2012 3

Regression Noise model and likelihood

► Given a dataset D = {xⁿ, yⁿ}^N_{n=1}, where xⁿ = {xⁿ₁,...,xⁿ_S} is S dimensional, fit parameters θ of a regressor f with added Gaussian noise:

$$y^n = f(\mathbf{x}^n; \boldsymbol{\theta}) + \epsilon^n \quad \text{where} \quad p(\epsilon \,|\, \sigma^2) = \mathcal{N}\left(\epsilon \,\big|\, 0, \sigma^2
ight).$$

Equivalent likelihood formulation:

$$p(\mathbf{y} \,|\, \mathbf{X}) = \prod_{n=1}^{N} \mathcal{N}\left(y^n \,\big|\, f(\mathbf{x}^n), \sigma^2\right)$$

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 4

イロト 不得下 イヨト イヨト 二日

Regression Choosing a regressor

Choose f to be linear:

$$p(\mathbf{y} \mid \mathbf{X}) = \prod_{n=1}^{N} \mathcal{N} \left(y^n \mid \mathbf{x}^n \cdot \boldsymbol{\theta} + c, \sigma^2 \right)$$

Consider bias free case, c = 0, otherwise include an additional column of ones in each xⁿ.

C. Lippert & O. Stegle

Regression Choosing a regressor

Choose f to be linear:

$$p(\mathbf{y} | \mathbf{X}) = \prod_{n=1}^{N} \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta} + c, \sigma^2 \right)$$

Consider bias free case, c = 0, otherwise include an additional column of ones in each xⁿ.



Equivalent graphical model

(日) (同) (三) (三)

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 5

Linear Regression Maximum likelihood

Taking the logarithm, we obtain

$$\ln p(\mathbf{y} \mid \boldsymbol{\theta} \sigma^2) = \sum_{n=1}^{N} \ln \mathcal{N} \left(y^n \mid \mathbf{x}^n \cdot \boldsymbol{\theta}, \sigma^2 \right)$$
$$= -\frac{N}{2} \ln 2\pi \sigma^2 - \frac{1}{2\sigma^2} \underbrace{\sum_{n=1}^{N} (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta})^2}_{\text{Sum of squares}}$$

The likelihood is maximized when the squared error is minimized.

Least squares and maximum likelihood are equivalent.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

소리가 소문가 소문가 소문가 ...

Linear Regression Maximum likelihood

Taking the logarithm, we obtain

$$\ln p(\mathbf{y} \mid \boldsymbol{\theta} \sigma^2) = \sum_{n=1}^{N} \ln \mathcal{N} \left(y^n \mid \mathbf{x}^n \cdot \boldsymbol{\theta}, \sigma^2 \right)$$
$$= -\frac{N}{2} \ln 2\pi \sigma^2 - \frac{1}{2\sigma^2} \underbrace{\sum_{n=1}^{N} (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta})^2}_{\text{Sum of squares}}$$

The likelihood is maximized when the squared error is minimized.

Least squares and maximum likelihood are equivalent.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

イロト 不得 トイヨト イヨト

Linear Regression Maximum likelihood

Taking the logarithm, we obtain

$$\ln p(\mathbf{y} \mid \boldsymbol{\theta} \sigma^2) = \sum_{n=1}^{N} \ln \mathcal{N} \left(y^n \mid \mathbf{x}^n \cdot \boldsymbol{\theta}, \sigma^2 \right)$$
$$= -\frac{N}{2} \ln 2\pi \sigma^2 - \frac{1}{2\sigma^2} \underbrace{\sum_{n=1}^{N} (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta})^2}_{\text{Sum of squares}}$$

► The likelihood is maximized when the squared error is minimized.

• Least squares and maximum likelihood are equivalent.

C. Lippert & O. Stegle

< ロ > < 同 > < 回 > < 回 > < 回 > <

Linear Regression and Least Squares



(C.M. Bishop, Pattern Recognition and Machine Learning)

$$E(\boldsymbol{\theta}) = \frac{1}{2} \sum_{n=1}^{N} (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta})^2$$

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 7

イロン 不聞と 不同と 不同と

Linear Regression and Least Squares

• Derivative w.r.t. a single weight entry θ_i

$$\frac{d}{\mathrm{d}\theta_i} \ln p(\mathbf{y} \mid \boldsymbol{\theta}, \sigma^2) = \frac{d}{\mathrm{d}\theta_i} \left[-\frac{1}{2\sigma^2} \sum_{n=1}^N (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta})^2 \right]$$
$$= \frac{1}{\sigma^2} \sum_{n=1}^N (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta}) x_i$$

Set gradient w.r.t. θ to zero

$$\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{y} \mid \boldsymbol{\theta}, \sigma^2) = \frac{1}{\sigma^2} \sum_{n=1}^{N} (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta}) \mathbf{x}^{n\mathrm{T}} = 0$$
$$\implies \boldsymbol{\theta}_{\mathsf{ML}} = \underbrace{(\mathbf{X}^{\mathrm{T}} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}}}_{\mathbf{y}} \mathbf{y}$$

Pseudo inverse

• Here, the matrix \mathbf{X} is defined as $\mathbf{X} = \begin{bmatrix} x_1^1 & \dots & x_S^1 \\ \dots & \dots & \dots \\ x_1^N & \dots & x_S^N \end{bmatrix}$

C. Lippert & O. Stegle

Linear Regression and Least Squares

• Derivative w.r.t. a single weight entry θ_i

$$\frac{d}{\mathrm{d}\theta_i} \ln p(\mathbf{y} \mid \boldsymbol{\theta}, \sigma^2) = \frac{d}{\mathrm{d}\theta_i} \left[-\frac{1}{2\sigma^2} \sum_{n=1}^N (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta})^2 \right]$$
$$= \frac{1}{\sigma^2} \sum_{n=1}^N (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta}) x_i$$

• Set gradient w.r.t. θ to zero

$$\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{y} \mid \boldsymbol{\theta}, \sigma^2) = \frac{1}{\sigma^2} \sum_{n=1}^{N} (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta}) \mathbf{x}^{n\mathrm{T}} = 0$$
$$\implies \boldsymbol{\theta}_{\mathrm{ML}} = \underbrace{(\mathbf{X}^{\mathrm{T}} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}}}_{\mathrm{Pseudo inverse}} \mathbf{y}$$

Here, the matrix X is defined as X =

C. Lippert & O. Stegle

Linear Regression and Least Squares

• Derivative w.r.t. a single weight entry θ_i

$$\frac{d}{\mathrm{d}\theta_i} \ln p(\mathbf{y} \mid \boldsymbol{\theta}, \sigma^2) = \frac{d}{\mathrm{d}\theta_i} \left[-\frac{1}{2\sigma^2} \sum_{n=1}^N (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta})^2 \right]$$
$$= \frac{1}{\sigma^2} \sum_{n=1}^N (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta}) x_i$$

• Set gradient w.r.t. θ to zero

$$\nabla_{\boldsymbol{\theta}} \ln p(\mathbf{y} \mid \boldsymbol{\theta}, \sigma^2) = \frac{1}{\sigma^2} \sum_{n=1}^{N} (y^n - \mathbf{x}^n \cdot \boldsymbol{\theta}) \mathbf{x}^{n\mathrm{T}} = 0$$

$$\implies \boldsymbol{\theta}_{\mathsf{ML}} = \underbrace{(\mathbf{X}^{\mathrm{T}} \mathbf{X})^{-1} \mathbf{X}^{\mathrm{T}}}_{\mathsf{Pseudo inverse}} \mathbf{y}$$

$$\blacktriangleright \text{ Here, the matrix } \mathbf{X} \text{ is defined as } \mathbf{X} = \begin{bmatrix} x_1^1 & \dots & x_S^1 \\ \dots & \dots & \dots \\ x_1^N & \dots & x_S^N \end{bmatrix}$$

C. Lippert & O. Stegle

Testing in Linear Regression Likelihood Ratio Test

$$p(\mathbf{y} | \mathbf{X}) = \prod_{n=1}^{N} \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta} + x_s^n \beta, \sigma^2 \right)$$

- x_s^n : SNP to be tested
- xⁿ: regression covariates (including bias term)
 - Race
 - Known background SNPs
 - Environment



Equivalent graphical model

 x^n : regression covariates

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 9

Testing in Linear Regression Likelihood Ratio Test

$$p(\mathbf{y} | \mathbf{X}) = \prod_{n=1}^{N} \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta} + x_s^n \beta, \sigma^2 \right)$$

• x_s^n : SNP to be tested

 xⁿ: regression covariates (including bias term)

- Race
- Known background SNPs
- Environment



Equivalent graphical model

 x^n : regression covariates

イロト イポト イヨト イヨト

Testing in Linear Regression Likelihood Ratio Test

$$p(\mathbf{y} | \mathbf{X}) = \prod_{n=1}^{N} \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta} + x_s^n \beta, \sigma^2 \right)$$

- x_s^n : SNP to be tested
- xⁿ: regression covariates (including bias term)
 - Race
 - Known background SNPs
 - Environment



Equivalent graphical model

 x^n : regression covariates

(日) (周) (三) (三)

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 9

Testing in Linear Regression Likelihood Ratio Test

$$p(\mathbf{y} | \mathbf{X}) = \prod_{n=1}^{N} \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta} + x_s^n \beta, \sigma^2 \right)$$

- Test $\mathcal{H}_0: \beta = 0$
- ► The ratio of the ML estimator and the ML₀ estimator restricted to H₀ is a common test statistic.



Equivalent graphical model

 x^n : regression covariates

Testing in Linear Regression Likelihood Ratio Test

$$p(\mathbf{y} | \mathbf{X}) = \prod_{n=1}^{N} \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta} + x_s^n \beta, \sigma^2 \right)$$

- Test $\mathcal{H}_0: \beta = 0$
- ► The ratio of the ML estimator and the ML₀ estimator restricted to H₀ is a common test statistic.



Equivalent graphical model

 x^n : regression covariates

Testing in Linear Regression Likelihood Ratio Test

$$p(\mathbf{y} | \mathbf{X}) = \prod_{n=1}^{N} \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta} + x_s^n \beta, \sigma^2 \right)$$

- Test $\mathcal{H}_0: \beta = 0$
- The ratio of the ML estimator and the ML₀ estimator restricted to H₀ is a common test statistic.

$$\frac{\prod_{n=1}^{N} \mathcal{N}\left(y^{n} \left| \mathbf{x}_{n} \cdot \boldsymbol{\theta}_{\mathsf{ML}} + x_{s}^{n} \beta_{\mathsf{ML}}, \sigma_{\mathsf{ML}}^{2}\right)}{\prod_{n=1}^{N} \mathcal{N}\left(y^{n} \left| \mathbf{x}^{n} \cdot \boldsymbol{\theta}_{\mathsf{ML}_{0}} + x_{s}^{n} 0, \sigma_{\mathsf{ML}_{0}}^{2}\right.\right)}$$



Equivalent graphical model

 x^n : regression covariates

(日) (同) (三) (三)

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 10

Outline

Linear Regression

Hypothesis Testing

Multiple Hypothesis Testing

Population Structure

Population structure correction

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Hypothesis Testing

Example:

- Given a sample $\mathcal{D} = \{x^1, \dots, x^N\}.$
- ► Test whether \mathcal{H}_0 : $\beta_s = 0$ (null hypothesis) or \mathcal{H}_1 : $\beta_s \neq 0$ (alternative hypothesis) is true.
- ► To show that β_s ≠ 0 we can perform a statistical test that tries to reject H₀.
- ► type 1 error: H₀ is rejected but does hold.
- ▶ **type 2 error:** \mathcal{H}_0 is accepted but does not hold.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

≣ ► < ≣ ► ≡ ∽ Q ⊂ September 2012 12

Hypothesis Testing

Example:

- Given a sample $\mathcal{D} = \{x^1, \dots, x^N\}.$
- ► Test whether $\mathcal{H}_0 : \beta_s = 0$ (null hypothesis) or $\mathcal{H}_1 : \beta_s \neq 0$ (alternative hypothesis) is true.
- ► To show that β_s ≠ 0 we can perform a statistical test that tries to reject H₀.
- ► type 1 error: H₀ is rejected but does hold.
- ▶ **type 2 error:** \mathcal{H}_0 is accepted but does not hold.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

≣ ► < ≣ ► ≡ </td>< ⊙ < ⊙</p>September 201212

Hypothesis Testing

Example:

- Given a sample $\mathcal{D} = \{x^1, \dots, x^N\}.$
- ► Test whether $\mathcal{H}_0 : \beta_s = 0$ (null hypothesis) or $\mathcal{H}_1 : \beta_s \neq 0$ (alternative hypothesis) is true.
- ► To show that β_s ≠ 0 we can perform a statistical test that tries to reject H₀.
- ► type 1 error: H₀ is rejected but does hold.
- ▶ **type 2 error:** \mathcal{H}_0 is accepted but does not hold.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Hypothesis Testing

Example:

- Given a sample $\mathcal{D} = \{x^1, \dots, x^N\}.$
- ► Test whether $\mathcal{H}_0: \beta_s = 0$ (null hypothesis) or $\mathcal{H}_1: \beta_s \neq 0$ (alternative hypothesis) is true.
- ► To show that β_s ≠ 0 we can perform a statistical test that tries to reject H₀.

	π_0 noius	π_0 upesit t holu
\mathcal{H}_0 accepted	true negatives	false negatives
		type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

- ► **type 1 error:** \mathcal{H}_0 is rejected but does hold.
- ▶ **type 2 error:** \mathcal{H}_0 is accepted but does not hold.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 12

1 11 deseminated

Hypothesis Testing

Example:

- Given a sample $\mathcal{D} = \{x^1, \dots, x^N\}.$
- ► Test whether H₀ : β_s = 0 (null hypothesis) or H₁ : β_s ≠ 0 (alternative hypothesis) is true.
- ► To show that β_s ≠ 0 we can perform a statistical test that tries to reject H₀.

		5	•		
•	type 1 e	error: 7	\mathcal{H}_0 is	rejected	but
	does hole	d.			

▶ **type 2 error:** \mathcal{H}_0 is accepted but does not hold.

	\mathcal{H}_0 holds	\mathcal{H}_0 doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

A (B) < A (B) < A (B) </p>

September 2012 12

Hypothesis Testing

- Given a sample $\mathcal{D} = \{x^1, \dots, x^N\}.$
- ► Test whether H₀: β_s = 0 (null hypothesis) or H₁: β_s ≠ 0 (alternative hypothesis) is true.
- The significance level α defines the threshold and the sensitivity of the test. This equals the probability of a type-1 error.
- Usually decision is based on a test statistic.
- The critical region defines the values of the test statistic that lead to a rejection of the test.



Hypothesis Testing

- Given a sample $\mathcal{D} = \{x^1, \dots, x^N\}.$
- ► Test whether H₀ : β_s = 0 (null hypothesis) or H₁ : β_s ≠ 0 (alternative hypothesis) is true.
- The significance level α defines the threshold and the sensitivity of the test. This equals the probability of a type-1 error.
- Usually decision is based on a test statistic.
- The critical region defines the values of the test statistic that lead to a rejection of the test.



Hypothesis Testing

- Given a sample $\mathcal{D} = \{x^1, \dots, x^N\}.$
- ► Test whether H₀ : β_s = 0 (null hypothesis) or H₁ : β_s ≠ 0 (alternative hypothesis) is true.
- The significance level α defines the threshold and the sensitivity of the test. This equals the probability of a type-1 error.
- Usually decision is based on a test statistic.
- The critical region defines the values of the test statistic that lead to a rejection of the test.



Hypothesis Testing

- Given a sample $\mathcal{D} = \{x^1, \dots, x^N\}.$
- ► Test whether H₀ : β_s = 0 (null hypothesis) or H₁ : β_s ≠ 0 (alternative hypothesis) is true.
- The significance level α defines the threshold and the sensitivity of the test. This equals the probability of a type-1 error.
- Usually decision is based on a test statistic.
- The critical region defines the values of the test statistic that lead to a rejection of the test.



P-value definition

- Probability of observing a test statistic at least as extreme (e.g. likelihood ratio statistic), given that H₀ is true.
- Significance level α becomes threshold on P-value.
- Need to know the null distribution of test statistics. (usually unknown)
- Possible to generate artificial null-distribution by permutations

C. Lippert & O. Stegle

P-value definition

- Probability of observing a test statistic at least as extreme (e.g. likelihood ratio statistic), given that H₀ is true.
- Significance level α becomes threshold on *P*-value.
- Need to know the null distribution of test statistics. (usually unknown)
- Possible to generate artificial null-distribution by permutations

C. Lippert & O. Stegle

P-value definition

- ▶ Probability of observing a test statistic at least as extreme (e.g. likelihood ratio statistic), given that H₀ is true.
- Significance level α becomes threshold on *P*-value.
- Need to know the null distribution of test statistics. (usually unknown)
- Possible to generate artificial null-distribution by permutations

P-value definition

- Probability of observing a test statistic at least as extreme (e.g. likelihood ratio statistic), given that H₀ is true.
- Significance level α becomes threshold on *P*-value.
- Need to know the null distribution of test statistics. (usually unknown)
- Possible to generate artificial null-distribution by permutations

P-value Permutation procedure

Repeat M times:

- Permute phenotype y and covariates x jointly over individuals.
- Compute permuted test statistic
- Add test statistic to emprirical null distribution

P-value Permutation procedure

Repeat M times:

- Permute phenotype y and covariates x jointly over individuals.
- Compute permuted test statistic
- Add test statistic to emprirical null distribution



- N

P-value Permutation procedure

Repeat M times:

- Permute phenotype y and covariates x jointly over individuals.
- Compute permuted test statistic
- Add test statistic to emprirical null distribution


P-value Permutation procedure

Repeat M times:

- Permute phenotype y and covariates x jointly over individuals.
- Compute permuted test statistic
- Add test statistic to emprirical null distribution



P-value Permutation procedure

- Repeat M times:
 - Permute phenotype y and covariates x jointly over individuals.
 - Compute permuted test statistic
 - Add test statistic to emprirical null distribution
- The P-value is the quantile of real test statistic in artificial null distribution.



Testing in Linear Regression Likelihood Ratio Test revisited

 Can equivalently compute log-likelihood ratio:



- Wilks' theorem: 2LR follows a Chi-square distribution with 1 degree of freedom.
- *P*-value = 1-CDF(2LR).

Equivalent graphical model x^n : regression covariates

- 4 ≣ ▶

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

4 Ten

Testing in Linear Regression Likelihood Ratio Test revisited

 Can equivalently compute log-likelihood ratio:

$$\mathsf{LR} = \sum_{n=1}^{N} \log \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta}_{\mathsf{ML}} + x_s^n \beta_{\mathsf{ML}}, \sigma_{\mathsf{ML}}^2 \right) \\ - \sum_{n=1}^{N} \log \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta}_{\mathsf{ML}_0}, \sigma_{\mathsf{ML}_0}^2 \right)$$



 Wilks' theorem: 2LR follows a Chi-square distribution with 1 degree of freedom.

• P-value = 1-CDF(2LR).

Equivalent graphical model x^n : regression covariates

- 4 ≣ ▶

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 16

4 Ten

Testing in Linear Regression Likelihood Ratio Test revisited

 Can equivalently compute log-likelihood ratio:

$$\mathsf{LR} = \sum_{n=1}^{N} \log \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta}_{\mathsf{ML}} + x_s^n \beta_{\mathsf{ML}}, \sigma_{\mathsf{ML}}^2 \right) \\ - \sum_{n=1}^{N} \log \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta}_{\mathsf{ML}_0}, \sigma_{\mathsf{ML}_0}^2 \right)$$



 Wilks' theorem: 2LR follows a Chi-square distribution with 1 degree of freedom.

• P-value = 1-CDF(2LR).

(source: Wikipedia)

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 16

- E - N

Testing in Linear Regression Likelihood Ratio Test revisited

 Can equivalently compute log-likelihood ratio:

$$\mathsf{LR} = \sum_{n=1}^{N} \log \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta}_{\mathsf{ML}} + x_s^n \beta_{\mathsf{ML}}, \sigma_{\mathsf{ML}}^2 \right) \\ - \sum_{n=1}^{N} \log \mathcal{N} \left(y^n \, \big| \, \mathbf{x}^n \cdot \boldsymbol{\theta}_{\mathsf{ML}_0}, \sigma_{\mathsf{ML}_0}^2 \right)$$

- Wilks' theorem: 2LR follows a Chi-square distribution with 1 degree of freedom.
- P-value = 1-CDF(2LR).





Outline

Linear Regression

Hypothesis Testing

Multiple Hypothesis Testing

Population Structure

Population structure correction

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 17

3

イロト イポト イヨト イヨト

Multiple Hypothesis Testing Motivation

- Significance level α equals probability of type-1 error.
- In GWAS we perform $S = 10^6$ tests
- At α = 0.01 we would expect 10000 type-1 errors!
- Probability of at least 1 type-1 error is 1 − (1 − α)^S → 1.

	\mathcal{H}_0 holds	$ \mathcal{H}_0$ doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

イロト イポト イヨト イヨト

Multiple Hypothesis Testing Motivation

- Significance level α equals probability of type-1 error.
- In GWAS we perform $S = 10^6$ tests
- At α = 0.01 we would expect 10000 type-1 errors!
- Probability of at least 1 type-1 error is 1 − (1 − α)^S → 1.
- Individual P-values < 0.01 are not significant anymore.

	\mathcal{H}_0 holds	\mathcal{H}_0 doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

(日) (周) (三) (三)

Multiple Hypothesis Testing Motivation

- Significance level α equals probability of type-1 error.
- ▶ In GWAS we perform $S = 10^6$ tests
- At $\alpha = 0.01$ we would expect 10000 type-1 errors!
- Probability of at least 1 type-1 error is 1 − (1 − α)^S → 1.
- Individual *P*-values < 0.01 are not significant anymore.

	\mathcal{H}_0 holds	$ \mathcal{H}_0$ doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

(日) (周) (三) (三)

Multiple Hypothesis Testing Motivation

- Significance level α equals probability of type-1 error.
- ▶ In GWAS we perform $S = 10^6$ tests
- At $\alpha = 0.01$ we would expect 10000 type-1 errors!
- ► Probability of at least 1 type-1 error is $1 (1 \alpha)^S \rightarrow 1$.

	\mathcal{H}_0 holds	$ \mathcal{H}_0$ doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

- 4 同 6 4 日 6 4 日 6

Multiple Hypothesis Testing Motivation

- Significance level α equals probability of type-1 error.
- In GWAS we perform $S = 10^6$ tests
- At $\alpha = 0.01$ we would expect 10000 type-1 errors!
- ► Probability of at least 1 type-1 error is $1 (1 \alpha)^S \rightarrow 1$.
- Individual *P*-values < 0.01 are not significant anymore.

	\mathcal{H}_0 holds	$\mid \mathcal{H}_0$ doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

- 4 同 6 4 日 6 4 日 6

Multiple Hypothesis Testing Motivation

- Significance level α equals probability of type-1 error.
- ▶ In GWAS we perform $S = 10^6$ tests
- At $\alpha = 0.01$ we would expect 10000 type-1 errors!
- ► Probability of at least 1 type-1 error is $1 (1 \alpha)^S \rightarrow 1$.
- Individual *P*-values < 0.01 are not significant anymore.

Need to correct for multiple hypothesis testing!

	\mathcal{H}_0 holds	$ \mathcal{H}_0$ doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

- 4 回 ト 4 三 ト 4 三 ト

Multiple Hypothesis Testing Family-Wise Error Rate (FWER)

Probability of at least one type-1 error.

- Correct by bounding the FWER.
- Bonferroni correction: $P_B = P \cdot S$
- Equivalently $P < \frac{\alpha}{s}$ significant.

Bounds the FWER 1 − (1 − α/S)^S by α

	\mathcal{H}_0 holds	$ \mathcal{H}_0$ doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

(日) (周) (三) (三)

Multiple Hypothesis Testing Family-Wise Error Rate (FWER)

- Probability of at least one type-1 error.
- Correct by bounding the FWER.
- Bonferroni correction: $P_B = P \cdot S$
- Equivalently $P < \frac{\alpha}{g}$ significant.

Bounds the FWER 1 − (1 − α/S)^S by α

	\mathcal{H}_0 holds	$ \mathcal{H}_0$ doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

イロト イポト イヨト イヨト

Multiple Hypothesis Testing Family-Wise Error Rate (FWER)

- Probability of at least one type-1 error.
- Correct by bounding the FWER.
- Bonferroni correction: $P_B = P \cdot S$
- Equivalently $P < \frac{\alpha}{\alpha}$ significant.
- ► Bounds the FWER 1 − (1 − α/S)^S by α

	\mathcal{H}_0 holds	\mathcal{H}_0 doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 19

Multiple Hypothesis Testing Family-Wise Error Rate (FWER)

- Probability of at least one type-1 error.
- Correct by bounding the FWER.
- Bonferroni correction: $P_B = P \cdot S$
- Equivalently $P < \frac{\alpha}{S}$ significant.
- ▶ Bounds the FWER $1 (1 \alpha/S)^S$ by α

	\mathcal{H}_0 holds	\mathcal{H}_0 doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

- 4 同 6 4 日 6 4 日 6

Multiple Hypothesis Testing Family-Wise Error Rate (FWER)

- Probability of at least one type-1 error.
- Correct by bounding the FWER.
- Bonferroni correction: $P_B = P \cdot S$
- Equivalently $P < \frac{\alpha}{S}$ significant.
- Bounds the FWER $1 (1 \alpha/S)^S$ by α

	\mathcal{H}_0 holds	\mathcal{H}_0 doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

(人間) トイヨト イヨト

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 19

False Discovery Rate (FDR)

- FWER based correction (Bonferroni) leads to very conservative significance thresholds.
- Because of the abundance of tests we might be willing to accept a few false positives.
- Intuitive definition of the FDR:
 - $\blacktriangleright \mathbb{E}\left[\frac{FP}{FP+TP}\right]$
- ▶ But: this can not be bounded when \mathcal{H}_0 always true (FN + TP = 0). In this case $\mathbb{E}\left[\frac{FP}{FP + TP}\right] = 1$.

	\mathcal{H}_0 holds	$ \mathcal{H}_0$ doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 20

False Discovery Rate (FDR)

- FWER based correction (Bonferroni) leads to very conservative significance thresholds.
- Because of the abundance of tests we might be willing to accept a few false positives.
- ▶ Intuitive definition of the FDR: ▶ $\mathbb{E}\left[\frac{FP}{FP+TP}\right]$
- ▶ But: this can not be bounded when \mathcal{H}_0 always true (FN + TP = 0). In this case $\mathbb{E}\left[\frac{FP}{FP + TP}\right] = 1$.

	\mathcal{H}_0 holds	\mathcal{H}_0 doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

False Discovery Rate (FDR)

- FWER based correction (Bonferroni) leads to very conservative significance thresholds.
- Because of the abundance of tests we might be willing to accept a few false positives.
- Intuitive definition of the FDR:

$$\blacktriangleright \mathbb{E}\left[\frac{FP}{FP+TP}\right]$$

▶ But: this can not be bounded when \mathcal{H}_0 always true (FN + TP = 0). In this case $\mathbb{E}\left[\frac{FP}{FP + TP}\right] = 1$.

	\mathcal{H}_0 holds	\mathcal{H}_0 doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

False Discovery Rate (FDR)

- FWER based correction (Bonferroni) leads to very conservative significance thresholds.
- Because of the abundance of tests we might be willing to accept a few false positives.
- Intuitive definition of the FDR:

$$\blacktriangleright \mathbb{E}\left[\frac{FP}{FP+TP}\right]$$

▶ But: this can not be bounded when \mathcal{H}_0 always true (FN + TP = 0). In this case $\mathbb{E}\left[\frac{FP}{FP + TP}\right] = 1$.

	\mathcal{H}_0 holds	\mathcal{H}_0 doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

False Discovery Rate (FDR) q-value estimation

$$\blacktriangleright \mathbb{E}\left[\frac{FP}{FP+TP}\right]$$

estimate number of false discoveries
 (FP) at given α cutoff

- Under $\mathcal{H}_0: \alpha \cdot S$
- ▶ FP+TP: number positives

	\mathcal{H}_0 holds	\mathcal{H}_0 doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

イロト 不得 トイヨト イヨト 二日

[Benjamini and Hochberg, 1995]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 21

False Discovery Rate (FDR) q-value estimation

$$\blacktriangleright \mathbb{E}\left[\frac{FP}{FP+TP}\right]$$

estimate number of false discoveries
 (FP) at given α cutoff

- Under \mathcal{H}_0 : $\alpha \cdot S$
- FP+TP: number positives

	\mathcal{H}_0 holds	\mathcal{H}_0 doesn't hold
\mathcal{H}_0 accepted	true negatives	false negatives type-2 error
\mathcal{H}_0 rejected	false positives type-1 error	true positives

イロト 不得 トイヨト イヨト

[Benjamini and Hochberg, 1995]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

False Discovery Rate (FDR) q-value estimation

$$\blacktriangleright \mathbb{E}\left[\frac{FP}{FP+TP}\right]$$

- estimate number of false discoveries
 (FP) at given α cutoff
- Under \mathcal{H}_0 : $\alpha \cdot S$
- ► FP+TP: number positives



False Discovery Rate (FDR) q-value estimation

•
$$\mathbb{E}\left[\frac{FP}{FP+TP}\right]$$

- estimate number of false discoveries
 (FP) at given α cutoff
- Under \mathcal{H}_0 : $\alpha \cdot S$
- ► FP+TP: number positives

Definition: q-value of a SNP

- $q(P) = \min_{FDR}$ with $P < \alpha_{FDR}$
- report all SNPs with q-value i allowed FDR.



C. Lippert & O. Stegle

Model Checking

- Do my estimated P-values match the true null distribution?
 - By definition uniformly distributed under null distribution.
- Do the empirical results match my assumptions on the null model?
- ► In GWAS we perform a large number of tests. (usually in the order of 10⁶)
- Use the strong prior knowledge that in GWAS almost all of the test SNPs have no effect on the phenotype.
- Empirical test statistics should follow the null distribution





A (10) A (10)

C. Lippert & O. Stegle

Model Checking

- Do my estimated P-values match the true null distribution?
 - By definition uniformly distributed under null distribution.
- Do the empirical results match my assumptions on the null model?
- ► In GWAS we perform a large number of tests. (usually in the order of 10⁶)
- Use the strong prior knowledge that in GWAS almost all of the test SNPs have no effect on the phenotype.
- Empirical test statistics should follow the null distribution





• • • • • • • • • • • •

C. Lippert & O. Stegle

Model Checking

- Do my estimated P-values match the true null distribution?
 - By definition uniformly distributed under null distribution.
- Do the empirical results match my assumptions on the null model?
- ► In GWAS we perform a large number of tests. (usually in the order of 10⁶)
- Use the strong prior knowledge that in GWAS almost all of the test SNPs have no effect on the phenotype.
- Empirical test statistics should follow the null distribution





C. Lippert & O. Stegle

Model Checking

- Do my estimated P-values match the true null distribution?
 - By definition uniformly distributed under null distribution.
- Do the empirical results match my assumptions on the null model?
- ► In GWAS we perform a large number of tests. (usually in the order of 10⁶)
- Use the strong prior knowledge that in GWAS almost all of the test SNPs have no effect on the phenotype.
- Empirical test statistics should follow the null distribution





C. Lippert & O. Stegle

Model Checking

- Do my estimated P-values match the true null distribution?
 - By definition uniformly distributed under null distribution.
- Do the empirical results match my assumptions on the null model?
- ► In GWAS we perform a large number of tests. (usually in the order of 10⁶)
- Use the strong prior knowledge that in GWAS almost all of the test SNPs have no effect on the phenotype.
- Empirical test statistics should follow the null distribution





Model Checking QQ-plot

Compare quantiles of the empirical test statistic distribution to assumed null distribution.

- Sort test statistics
- Plot test statisitcs against (y-axis) quantiles of the theoretical null-distribution (x-axis)
 - for example: 2LR vs. χ_1^2
- If the plot is close to the diagonal, the distributions match up
- Deviation from the diagonal indicates inflation or deflation of test statistics.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

- 4 同 6 4 日 6 4 日 6

Model Checking QQ-plot

Compare quantiles of the empirical test statistic distribution to assumed null distribution.

Sort test statistics

 Plot test statisitcs against (y-axis) quantiles of the theoretical null-distribution (x-axis)

• for example: 2LR vs. χ_1^2

- If the plot is close to the diagonal, the distributions match up
- Deviation from the diagonal indicates inflation or deflation of test statistics.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

- 4 同 6 4 日 6 4 日 6

Model Checking QQ-plot

Compare quantiles of the empirical test statistic distribution to assumed null distribution.

- Sort test statistics
- Plot test statisitcs against (y-axis) quantiles of the theoretical null-distribution (x-axis)
 - for example: 2LR vs. χ_1^2
- If the plot is close to the diagonal, the distributions match up
- Deviation from the diagonal indicates inflation or deflation of test statistics.



Model Checking QQ-plot

Compare quantiles of the empirical test statistic distribution to assumed null distribution.

- Sort test statistics
- Plot test statisitcs against (y-axis) quantiles of the theoretical null-distribution (x-axis)
 - for example: 2LR vs. χ_1^2
- If the plot is close to the diagonal, the distributions match up
- Deviation from the diagonal indicates inflation or deflation of test statistics.



Model Checking QQ-plot

Compare quantiles of the empirical test statistic distribution to assumed null distribution.

- Sort test statistics
- Plot test statisitcs against (y-axis) quantiles of the theoretical null-distribution (x-axis)
 - for example: 2LR vs. χ_1^2
- If the plot is close to the diagonal, the distributions match up
- Deviation from the diagonal indicates inflation or deflation of test statistics.


Outline

Linear Regression

Hypothesis Testing

Multiple Hypothesis Testing

Population Structure

Population structure correction

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

 Image: Image:

イロト イポト イヨト イヨト

Genome wide association studies (GWAS)

- Identify associations between variable genetic loci and phenotypes.
 - Linear and logistic regression
 - Statistical dependence tests

```
(F-test, likelihood ratio)
```



C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Genome wide association studies (GWAS)

- Identify associations between variable genetic loci and phenotypes.
 - Linear and logistic regression
 - Statistical dependence tests
 - (F-test, likelihood ratio)

$$\frac{\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{e}}^{2}\mathbf{I}\right)}{\mathcal{N}\left(\mathbf{y}|\mathbf{0};\sigma_{\mathrm{e}}^{2}\mathbf{I}\right)} \qquad (1)$$



(日) (同) (日) (日)

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Population stratification

- Confounding structure leads to false positives.
 - Population structure
 - Family structure
 - Cryptic relatedness



Population stratification

- Confounding structure leads to false positives.
 - Population structure
 - Family structure
 - Cryptic relatedness



Population stratification

- Confounding structure leads to false positives.
 - Population structure
 - Family structure
 - Cryptic relatedness



Population stratification

- Confounding structure leads to false positives.
 - Population structure
 - Family structure
 - Cryptic relatedness



Population stratification

- Confounding structure leads to false positives.
 - Population structure
 - Family structure
 - Cryptic relatedness





Population stratification

GWA on inflammatory bowel disease (WTCCC)

▶ 3.4k cases, 11.9k controls

- Methods
 - Linear regression
 - Likelihood ratio test

[Burton et al., 2007]

イロト イポト イヨト イヨト

Population stratification

GWA on inflammatory bowel disease (WTCCC)

- ▶ 3.4k cases, 11.9k controls
- Methods
 - Linear regression
 - Likelihood ratio test

[Burton et al., 2007]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Population stratification

GWA on inflammatory bowel disease (WTCCC)

- ▶ 3.4k cases, 11.9k controls
- Methods
 - Linear regression
 - Likelihood ratio test



C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Outline

Linear Regression

Hypothesis Testing

Multiple Hypothesis Testing

Population Structure

Population structure correction

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

イロト イポト イヨト イヨト

Genomic control [Devlin and Roeder, Biometrics 1999]

• Genomic control λ

 $\lambda = \frac{\mathrm{median}(2LR)}{\mathrm{median}(\chi^2)}.$

- $\lambda = 1$: Calibrated *p*-values
- $\lambda > 1$: Inflation
- $\lambda < 1$: Deflation
- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of *p*-values.
- Very conservative.



C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Genomic control [Devlin and Roeder, Biometrics 1999]

• Genomic control λ

$$\lambda = \frac{\text{median}(2LR)}{\text{median}(\chi^2)}.$$

- $\lambda = 1$: Calibrated *p*-values
- $\lambda > 1$: Inflation
- $\lambda < 1$: Deflation
- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of *p*-values.
- Very conservative.



A D > 4 B >

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Genomic control [Devlin and Roeder, Biometrics 1999]

• Genomic control λ

$$\lambda = \frac{\text{median}(2LR)}{\text{median}(\chi^2)}.$$

- $\lambda = 1$: Calibrated *p*-values
- $\lambda > 1$: Inflation
- $\lambda < 1$: Deflation
- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of *p*-values.
- Very conservative.



C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Genomic control [Devlin and Roeder, Biometrics 1999]

• Genomic control λ

$$\lambda = \frac{\text{median}(2LR)}{\text{median}(\chi^2)}.$$

- $\lambda = 1$: Calibrated *p*-values
- ▶ λ > 1: Inflation
- $\lambda < 1$: Deflation
- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of *p*-values.
- Very conservative.



C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Genomic control [Devlin and Roeder, Biometrics 1999]

• Genomic control λ

$$\lambda = \frac{\text{median}(2LR)}{\text{median}(\chi^2)}.$$

- $\lambda = 1$: Calibrated *p*-values
- ▶ λ > 1: Inflation
- $\lambda < 1$: Deflation
- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of *p*-values.

Very conservative.



C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Genomic control [Devlin and Roeder, Biometrics 1999]

• Genomic control λ

$$\lambda = \frac{\text{median}(2LR)}{\text{median}(\chi^2)}.$$

- $\lambda = 1$: Calibrated *p*-values
- ▶ λ > 1: Inflation
- ▶ λ < 1: Deflation</p>
- Correct by dividing test statistic by λ.
- Applicable in combination with every method.
- Does not change (non-)uniformity of *p*-values.
- Very conservative.



Linear models I: linear regression and linear mixed models

- Population structure causes genome-wide correlations between SNPs
- A large part of the total variation in the SNPs can be explained by population differences.
- Novembre et al. [2008] show that the eigenvectors of the SNP covariance matrix reflect population structure.
- Eigenstrat uses this property to correct for population structure in GWAS.



[Price et al., 2006, Patterson et al., 2006, Novembre et al., 2008] C. Lippert & O. Stegle Linear models I: linear regression and linear mixed models

- Population structure causes genome-wide correlations between SNPs
- A large part of the total variation in the SNPs can be explained by population differences.
- Novembre et al. [2008] show that the eigenvectors of the SNP covariance matrix reflect population structure.
- Eigenstrat uses this property to correct for population structure in GWAS.



[Price et al., 2006, Patterson et al., 2006, Novembre et al., 2008] C. Lippert & O. Stegle Linear models I: linear regression and linear mixed models

- Population structure causes genome-wide correlations between SNPs
- A large part of the total variation in the SNPs can be explained by population differences.
- Novembre et al. [2008] show that the eigenvectors of the SNP covariance matrix reflect population structure.
- Eigenstrat uses this property to correct for population structure in GWAS.



[Price et al., 2006, Patterson et al., 2006, Novembre et al., 2008] C. Lippert & O. Stegle Linear models I: linear regression and linear mixed models

- Population structure causes genome-wide correlations between SNPs
- A large part of the total variation in the SNPs can be explained by population differences.
- Novembre et al. [2008] show that the eigenvectors of the SNP covariance matrix reflect population structure.
- Eigenstrat uses this property to correct for population structure in GWAS.



Eigenstrat

Eigenstrat procedure:

- Compute covariance matrix based on SNPs
- Compute eigenvectors of covaraince matrix
- Add largest eigenvector as covariate to regression.
- Repeat until *P*-values are uniform.



C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models



enome-wide SNP covariance

Eigenstrat procedure:

- Compute covariance matrix based on SNPs
- Compute eigenvectors of covaraince matrix
- Add largest eigenvector as covariate to regression.
- Repeat until *P*-values are uniform.



C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

(人間) トイヨト イヨト

Eigenstrat procedure:

- Compute covariance matrix based on SNPs
- Compute eigenvectors of covaraince matrix
- Add largest eigenvector as covariate to regression.
- Repeat until *P*-values are uniform.



C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Eigenstrat procedure:

- Compute covariance matrix based on SNPs
- Compute eigenvectors of covaraince matrix
- Add largest eigenvector as covariate to regression.
- Repeat until *P*-values are uniform.



C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Eigenstrat procedure:

- Compute covariance matrix based on SNPs
- Compute eigenvectors of covaraince matrix
- Add largest eigenvector as covariate to regression.
- Repeat until *P*-values are uniform.



C. Lippert & OnStegle

Linear models I: linear regression and linear mixed models

September 2012 31

Linear mixed models (LMM)

► Kernel matrix K

- Estimated from SNP data
- Kinship coefficients
 - Identity by state
 Identity by descent
- Realized relationship matrix (linear)
- Sample random effect **u**.
- Sample phenotype y.



Linear mixed models (LMM)

► Kernel matrix K

- Estimated from SNP data
- Kinship coefficients
 - Identity by stateIdentity by descent
- Realized relationship matrix (linear)
- Sample random effect **u**.
- Sample phenotype y.



C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 32

Linear mixed models (LMM)

- ► Kernel matrix K
 - Estimated from SNP data
 - Kinship coefficients
 - Identity by state
 - Identity by descent
 - Realized relationship matrix (linear)
- Sample random effect u.Sample phenotype y.



(日) (同) (三) (三)

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 32

Linear mixed models (LMM)

- ► Kernel matrix K
 - Estimated from SNP data
 - Kinship coefficients
 - Identity by state
 - Identity by descent
 - Realized relationship matrix (linear)
- Sample random effect **u**.
- Sample phenotype y.



Linear mixed models (LMM)

- ► Kernel matrix K
 - Estimated from SNP data
 - Kinship coefficients
 - Identity by state
 - Identity by descent
 - Realized relationship matrix (linear)
- Sample random effect **u**.
- Sample phenotype y.



Linear mixed models (LMM)

- ► Kernel matrix K
 - Estimated from SNP data
 - Kinship coefficients
 - Identity by state
 - Identity by descent
 - Realized relationship matrix (linear)
- Sample random effect **u**.
- Sample phenotype y.

$$\int_{\mathbf{u}} \mathcal{N}\left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta} + \mathbf{u}; \sigma_{e}^{2} \mathbf{I}\right) \mathcal{N}\left(\mathbf{u} | \mathbf{0}; \sigma_{g}^{2} \mathbf{K}\right)$$



Linear mixed models (LMM)

- ► Kernel matrix K
 - Estimated from SNP data
 - Kinship coefficients
 - Identity by state
 - Identity by descent
 - Realized relationship matrix (linear)
- Sample random effect **u**.
- Sample phenotype y.

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\mathbf{K}+\sigma_{e}^{2}\mathbf{I}\right)$$



Linear mixed models (LMM)

- Corrects for all levels of population structure.
- ML estimation is computationally demanding

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\mathbf{K}+\sigma_{e}^{2}\mathbf{I}\right)$$



Linear mixed models (LMM)

- Corrects for all levels of population structure.
- ML estimation is computationally demanding

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\mathbf{K}+\sigma_{e}^{2}\mathbf{I}\right)$$




Linear mixed models (LMM)

- Corrects for all levels of population structure.
- ML estimation is computationally demanding
 Non-convex in a² and a²

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\mathbf{K}+\sigma_{e}^{2}\mathbf{I}\right)$$





(日) (同) (三) (三)

Linear mixed models (LMM)

- Corrects for all levels of population structure.
- ML estimation is computationally demanding
 - ▶ Non-convex in σ_{g}^{2} and σ_{e}^{2}

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\mathbf{K}+\sigma_{e}^{2}\mathbf{I}\right)$$





(日) (同) (日) (日)

Linear mixed models (LMM)

- Corrects for all levels of population structure.
- ML estimation is computationally demanding
 - ▶ Non-convex in $\sigma_{\rm g}^2$ and $\sigma_{\rm e}^2$.

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\mathbf{K}+\sigma_{e}^{2}\mathbf{I}\right)$$





(日) (同) (日) (日)

GWAS for Flowering Time in Arabidopsis thaliana

Linear Model:



QQ-plot:



GWAS for Flowering Time in Arabidopsis thaliana

Linear Model:



Linear Mixed Model:



GWAS for Flowering Time in Arabidopsis thaliana

Linear Mixed Model:



QQ-plot:



C. Lippert & O. Stegle

Linear mixed models (LMM) EMMA

LMM log likelihood

$$LL(\boldsymbol{\beta}, \sigma_{g}^{2}, \sigma_{e}^{2}) = \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \mathbf{K} + \sigma_{e}^{2} \mathbf{I} \right).$$

• Change of variables, introducing $\delta = \sigma_{
m e}^2/\sigma_{
m g}^2$:

 $LL(\boldsymbol{\beta}, \sigma_{g}^{2}, \delta) = \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right).$

- ML-parameters $\hat{\beta}$ and $\hat{\sigma}_{g}^{2}$ follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.
 O(N³) per SNP.

[Kang et al., 2008]

C. Lippert & O. Stegle

Linear mixed models (LMM) EMMA

LMM log likelihood

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^{2}, \sigma_{\mathrm{e}}^{2}) = \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{\mathrm{g}}^{2} \mathbf{K} + \sigma_{\mathrm{e}}^{2} \mathbf{I} \right).$$

• Change of variables, introducing $\delta = \sigma_{\rm e}^2/\sigma_{\rm g}^2$:

$$LL(\boldsymbol{\beta}, \sigma_{g}^{2}, \delta) = \log \mathcal{N}\left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I}\right)\right).$$

- ML-parameters $\hat{\boldsymbol{\beta}}$ and $\hat{\sigma_{g}^{2}}$ follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.
 O(N³) per SNP.

[Kang et al., 2008]

C. Lippert & O. Stegle

Linear mixed models (LMM) EMMA

LMM log likelihood

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^{2}, \sigma_{\mathrm{e}}^{2}) = \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{\mathrm{g}}^{2} \mathbf{K} + \sigma_{\mathrm{e}}^{2} \mathbf{I} \right).$$

• Change of variables, introducing $\delta = \sigma_{\rm e}^2/\sigma_{\rm g}^2$:

$$LL(\boldsymbol{\beta}, \sigma_{g}^{2}, \delta) = \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right).$$

- ML-parameters $\hat{\beta}$ and $\hat{\sigma}_{g}^{2}$ follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.
 O(N³) per SNP.

[Kang et al., 2008]

C. Lippert & O. Stegle

Linear mixed models (LMM) EMMA

LMM log likelihood

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^{2}, \sigma_{\mathrm{e}}^{2}) = \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{\mathrm{g}}^{2} \mathbf{K} + \sigma_{\mathrm{e}}^{2} \mathbf{I} \right).$$

• Change of variables, introducing $\delta = \sigma_{\rm e}^2/\sigma_{\rm g}^2$:

$$LL(\boldsymbol{\beta}, \sigma_{g}^{2}, \delta) = \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right).$$

- ML-parameters $\hat{\beta}$ and $\hat{\sigma}_{g}^{2}$ follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.
 O(N³) per SNP.

[Kang et al., 2008]

C. Lippert & O. Stegle

イロン イボン イヨン イヨン 三日

Linear mixed models (LMM) EMMA

LMM log likelihood

$$LL(\boldsymbol{\beta}, \sigma_{\mathrm{g}}^{2}, \sigma_{\mathrm{e}}^{2}) = \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{\mathrm{g}}^{2} \mathbf{K} + \sigma_{\mathrm{e}}^{2} \mathbf{I} \right).$$

• Change of variables, introducing $\delta = \sigma_{\rm e}^2/\sigma_{\rm g}^2$:

$$LL(\boldsymbol{\beta}, \sigma_{g}^{2}, \delta) = \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right).$$

- ML-parameters $\hat{\beta}$ and $\hat{\sigma}_{g}^{2}$ follow in closed form.
- Use optimizer to solve 1-dimensional optimization problem over δ.
 O(N³) per SNP.

[Kang et al., 2008]

C. Lippert & O. Stegle

Linear mixed models (LMM)

ML parameters

Gradient of the LMM log likelihood w.r.t. m eta

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

E ▶ 《 E ▶ E ∽ Q ⊂ September 2012 38

< ロ > < 同 > < 回 > < 回 > < 回 > <

Linear mixed models (LMM)

ML parameters

Gradient of the LMM log likelihood w.r.t. m eta

$$\nabla_{\boldsymbol{\beta}} \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right) = \nabla_{\boldsymbol{\beta}} - \frac{1}{2\sigma_{g}^{2}} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)$$
$$= \frac{1}{\sigma_{g}^{2}} \left[-\mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{y} + \mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{X} \right]$$

C. Lippert & O. Stegle

< ロ > < 同 > < 回 > < 回 > < 回 > <

Linear mixed models (LMM) ML parameters

Gradient of the LMM log likelihood w.r.t. β

$$\nabla_{\boldsymbol{\beta}} \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right) = \nabla_{\boldsymbol{\beta}} - \frac{1}{2\sigma_{g}^{2}} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)$$
$$= \frac{1}{\sigma_{g}^{2}} \left[-\mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{y} + \mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{X} \right]$$

set gradient to zero:

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

E ► < E ► E </p>
September 2012 38

Linear mixed models (LMM) ML parameters

Gradient of the LMM log likelihood w.r.t. β

$$\nabla_{\boldsymbol{\beta}} \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right) = \nabla_{\boldsymbol{\beta}} - \frac{1}{2\sigma_{g}^{2}} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)$$
$$= \frac{1}{\sigma_{g}^{2}} \left[-\mathbf{X}^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{y} + \mathbf{X}^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{X} \right]$$

set gradient to zero:

$$\mathbf{0} = \frac{1}{\sigma_{g}^{2}} \left[\mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{y} - \mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{X} \boldsymbol{\beta} \right]$$
$$\mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{y}$$
$$\boldsymbol{\beta}_{\mathsf{ML}} = \left(\mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{y}$$

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Linear mixed models (LMM) ML parameters

Gradient of the LMM log likelihood w.r.t. β

$$\nabla_{\boldsymbol{\beta}} \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right) = \nabla_{\boldsymbol{\beta}} - \frac{1}{2\sigma_{g}^{2}} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)$$
$$= \frac{1}{\sigma_{g}^{2}} \left[-\mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{y} + \mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{X} \right]$$

set gradient to zero:

$$\mathbf{0} = \frac{1}{\sigma_{g}^{2}} \left[\mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{y} - \mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{X} \boldsymbol{\beta} \right]$$
$$\mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{y}$$
$$\boldsymbol{\beta}_{\mathsf{ML}} = \left(\mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{X} \right)^{-1} \mathbf{X}^{T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{y}$$

Note that this solution is analogous to the ML solution of the linear regression $(\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{y}$.

C. Lippert & O. Stegle

Linear mixed models (LMM) ML parameters

Derivative of the LMM log likelihood w.r.t. $\sigma_{\rm g}^2$

Note that For every SNP we need to calculate (K + δI)⁻¹, which is an O(N³) operation.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 39

Linear mixed models (LMM) ML parameters

Derivative of the LMM log likelihood w.r.t. $\sigma_{\rm g}^2$

$$d\sigma_{g}^{2} \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right)$$
$$= -\frac{1}{2} \left[\frac{N}{\sigma_{g}^{2}} - \frac{N}{\sigma_{g}^{4}} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right) \right]$$

Note that For every SNP we need to calculate (K + δI)⁻¹, which is an O(N³) operation.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 39

Linear mixed models (LMM) ML parameters

Derivative of the LMM log likelihood w.r.t. $\sigma_{\rm g}^2$

$$d\sigma_{g}^{2} \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right)$$
$$= -\frac{1}{2} \left[\frac{N}{\sigma_{g}^{2}} - \frac{N}{\sigma_{g}^{4}} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right) \right]$$

set derivative to zero:

Note that For every SNP we need to calculate (K + δI)⁻¹, which is an O(N³) operation.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

▶ ৰ ≣ ▶ ≣ ৩ ৭ ৫ September 2012 39

イロト 不得 トイヨト イヨト

Linear mixed models (LMM) ML parameters

Derivative of the LMM log likelihood w.r.t. $\sigma_{\rm g}^2$

$$d\sigma_{g}^{2} \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right)$$
$$= -\frac{1}{2} \left[\frac{N}{\sigma_{g}^{2}} - \frac{N}{\sigma_{g}^{4}} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right) \right]$$

set derivative to zero:

$$0 = -\frac{1}{2} \left[\frac{n}{\sigma_g^2} - \frac{N}{\sigma_g^4} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right) \right]$$
$$\sigma_{g\mathsf{ML}}^2 = \frac{1}{N} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)$$

Note that For every SNP we need to calculate (K + δI)⁻¹, which is an O(N³) operation.

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

Linear mixed models (LMM) ML parameters

Derivative of the LMM log likelihood w.r.t. $\sigma_{\rm g}^2$

$$d\sigma_{g}^{2} \log \mathcal{N} \left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta}; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I} \right) \right)$$
$$= -\frac{1}{2} \left[\frac{N}{\sigma_{g}^{2}} - \frac{N}{\sigma_{g}^{4}} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right)^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right) \right]$$

set derivative to zero:

$$0 = -\frac{1}{2} \left[\frac{n}{\sigma_{g}^{2}} - \frac{N}{\sigma_{g}^{4}} \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right)^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right) \right]$$
$$\sigma_{g\mathsf{ML}}^{2} = \frac{1}{N} \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right)^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta} \right)$$

Note that For every SNP we need to calculate (K + δI)⁻¹, which is an O(N³) operation.

C. Lippert & O. Stegle

FaST LMM

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\left(\mathbf{K}+\delta\mathbf{I}\right)\right).$$
(2)

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 40

イロト イポト イヨト イヨト 二日

FaST LMM

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\left(\mathbf{K}+\delta\mathbf{I}\right)\right).$$
(2)

$$= \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{g}^{2}\left(\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} + \delta\mathbf{I}\right)\right).$$
(3)

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

イロト イポト イヨト イヨト

FaST LMM

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\left(\mathbf{K}+\delta\mathbf{I}\right)\right).$$
(2)

$$= \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{g}^{2}\left(\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} + \delta\mathbf{I}\right)\right).$$
(3)

$$= \mathcal{N} \left(\mathbf{U}^{\mathrm{T}} \mathbf{y} | \mathbf{U}^{\mathrm{T}} \mathbf{X} \boldsymbol{\beta}; \sigma_{\mathrm{g}}^{2} \left(\mathbf{S} + \delta \mathbf{I} \right) \right).$$
(4)

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

E ► < E ► E </p>
September 2012 40

・ロト ・ 四ト ・ ヨト ・ ヨト

FaST LMM

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\left(\mathbf{K}+\delta\mathbf{I}\right)\right).$$
(2)

$$= \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\left(\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} + \delta\mathbf{I}\right)\right).$$
(3)

$$= \mathcal{N} \left(\mathbf{U}^{\mathrm{T}} \mathbf{y} | \mathbf{U}^{\mathrm{T}} \mathbf{X} \boldsymbol{\beta}; \sigma_{\mathrm{g}}^{2} \left(\mathbf{S} + \delta \mathbf{I} \right) \right).$$
(4)



[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 40

3

<ロ> (日) (日) (日) (日) (日)

FaST LMM

$$\mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{K}+\delta\mathbf{I}\right)\right).$$
 (2)

$$= \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta}; \sigma_{g}^{2}\left(\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} + \delta\mathbf{I}\right)\right).$$
(3)

$$= \mathcal{N} \left(\mathbf{U}^{\mathrm{T}} \mathbf{y} | \mathbf{U}^{\mathrm{T}} \mathbf{X} \boldsymbol{\beta}; \sigma_{\mathrm{g}}^{2} \left(\mathbf{S} + \delta \mathbf{I} \right) \right).$$
(4)

- 4 ≣ ▶



[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

3 September 2012 40

- ∢ ≣ →

FaST LMM

$$\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{eta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{S}+\delta\mathbf{I}
ight)
ight).$$

Factored Spectrally Transformed LMM

► O(N³) once for spectral decomposition.

Exact LMM solution N times faster than EMMA.

• Bottlenecks: $O(N^3)$ runtime, $O(N^2)$ memory for **K**.



[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 41

(5)

FaST LMM

$$\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}oldsymbol{eta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{S}+\delta\mathbf{I}
ight)
ight).$$

Factored Spectrally Transformed LMM

- $O(N^3)$ once for spectral decomposition.
- Exact LMM solution N times faster than EMMA.
- Bottlenecks: $O(N^3)$ runtime, $O(N^2)$ memory for **K**.



[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

(5)

FaST LMM

$$\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}oldsymbol{eta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{S}+\delta\mathbf{I}
ight)
ight).$$

Factored Spectrally Transformed LMM

- $O(N^3)$ once for spectral decomposition.
- Exact LMM solution N times faster than EMMA.
- Bottlenecks: $O(N^3)$ runtime, $O(N^2)$ memory for ${f K}.$



[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 41

(5)

FaST LMM

$$\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{S}+\delta\mathbf{I}\right)\right).$$
(5)

Factored Spectrally Transformed LMM

- $O(N^3)$ once for spectral decomposition.
- Exact LMM solution N times faster than EMMA.
- ▶ Bottlenecks: $O(N^3)$ runtime, $O(N^2)$ memory for K.



[Lippert et al., 2011]

C. Lippert & O. Stegle

FaST linear-time mixed models Using linear genetic similarity matrices

► In general: FaST-LMM has O(N³) runtime and O(N²) memory requirement.

- Special case: Linear kernel
- Spectral decomposition directly from SVD of X

Computation of **K** can be avoided.

- ▶ For N > S_c, runtime and storage become linear.
- Total runtime $O(N \cdot S_c^2)$.
- Total storage $O(N \cdot S_c)$.



C. Lippert & O. Stegle

FaST linear-time mixed models Using linear genetic similarity matrices

- ► In general: FaST-LMM has O(N³) runtime and O(N²) memory requirement.
- Special case: Linear kernel
- \blacktriangleright Spectral decomposition directly from SVD of $\widetilde{\mathbf{X}}$

Computation of **K** can be avoided.

- ▶ For N > S_c, runtime and storage become linear.
- Total runtime $O(N \cdot S_c^2)$.
- Total storage $O(N \cdot S_c)$.



C. Lippert & O. Stegle

FaST linear-time mixed models Using linear genetic similarity matrices

- In general: FaST-LMM has O(N³) runtime and O(N²) memory requirement.
- Special case: Linear kernel
- Spectral decomposition directly from SVD of X

Computation of ${\bf K}$ can be avoided.

- ▶ For N > S_c, runtime and storage become linear.
- Total runtime $O(N \cdot S_c^2)$.
- Total storage $O(N \cdot S_c)$.



(人間) とうきょうきょう

C. Lippert & O. Stegle

FaST linear-time mixed models Using linear genetic similarity matrices

- In general: FaST-LMM has O(N³) runtime and O(N²) memory requirement.
- Special case: Linear kernel
- Spectral decomposition directly from SVD of X

Computation of ${\bf K}$ can be avoided.

- ▶ For N > S_c, runtime and storage become linear.
- ► Total runtime $O(N \cdot S_c^2)$. ► Total storage $O(N \cdot S_c)$



FaST linear-time mixed models Using linear genetic similarity matrices

- In general: FaST-LMM has O(N³) runtime and O(N²) memory requirement.
- Special case: Linear kernel
- Spectral decomposition directly from SVD of X

Computation of ${\bf K}$ can be avoided.

- ▶ For N > S_c, runtime and storage become linear.
- Total runtime $O(N \cdot S_c^2)$.
- Total storage $O(N \cdot S_c)$.



C. Lippert & O. Stegle

FaST linear-time mixed models Using linear genetic similarity matrices

- In general: FaST-LMM has O(N³) runtime and O(N²) memory requirement.
- Special case: Linear kernel
- Spectral decomposition directly from SVD of X

Computation of ${\bf K}$ can be avoided.

- ▶ For N > S_c, runtime and storage become linear.
- Total runtime $O(N \cdot S_c^2)$.
- Total storage $O(N \cdot S_c)$.



C. Lippert & O. Stegle
FaST low rank Mixed Models Using low rank genetic similarity matrices

Let K be rank k.

 $\mathbf{K} = \mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}}.$

- ▶ S has k non-zero diagonal elements (S₁), N k zero diagonal elements (S₂ = 0).
- ▶ $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$, with $\mathbf{U}_1 \in \mathcal{R}^{N imes k}$ and $\mathbf{U}_2 \in \mathcal{R}^{N imes N k}$
- k-SVD of **K** can be computed in $O(N^2k)$.
- $\mathbf{U}_1^{\mathrm{T}}\mathbf{y}$ and $\mathbf{U}_1^{\mathrm{T}}\mathbf{X}$ can be computed in $O(N \cdot k)$ per SNP.

[Lippert et al., 2011]

C. Lippert & O. Stegle

FaST low rank Mixed Models Using low rank genetic similarity matrices

Let K be rank k.

$$\mathbf{K} = \mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} = \mathbf{U}_{1}\mathbf{S}_{1}\mathbf{U}_{1}^{\mathrm{T}} + \mathbf{U}_{2}\mathbf{S}_{2}\mathbf{U}_{2}^{\mathrm{T}}$$

- S has k non-zero diagonal elements (S₁), N − k zero diagonal elements (S₂ = 0).
- ▶ $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$, with $\mathbf{U}_1 \in \mathcal{R}^{N imes k}$ and $\mathbf{U}_2 \in \mathcal{R}^{N imes N k}$
- k-SVD of **K** can be computed in $O(N^2k)$.
- $\mathbf{U}_1^{\mathrm{T}}\mathbf{y}$ and $\mathbf{U}_1^{\mathrm{T}}\mathbf{X}$ can be computed in $O(N \cdot k)$ per SNP.

[Lippert et al., 2011]

C. Lippert & O. Stegle

FaST low rank Mixed Models Using low rank genetic similarity matrices

Let K be rank k.

$$\mathbf{K} = \mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} = \mathbf{U}_{1}\mathbf{S}_{1}\mathbf{U}_{1}^{\mathrm{T}} + \mathbf{U}_{2}\mathbf{S}_{2}\mathbf{U}_{2}^{\mathrm{T}}$$

- S has k non-zero diagonal elements (S₁), N − k zero diagonal elements (S₂ = 0).
- ▶ $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$, with $\mathbf{U}_1 \in \mathcal{R}^{N \times k}$ and $\mathbf{U}_2 \in \mathcal{R}^{N \times N k}$
- k-SVD of **K** can be computed in $O(N^2k)$.
- ▶ $\mathbf{U}_1^{\mathrm{T}}\mathbf{y}$ and $\mathbf{U}_1^{\mathrm{T}}\mathbf{X}$ can be computed in $O(N \cdot k)$ per SNP.

[Lippert et al., 2011]

C. Lippert & O. Stegle

FaST low rank Mixed Models Using low rank genetic similarity matrices

Let K be rank k.

$$\mathbf{K} = \mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} = \mathbf{U}_{1}\mathbf{S}_{1}\mathbf{U}_{1}^{\mathrm{T}} + \mathbf{U}_{2}\mathbf{S}_{2}\mathbf{U}_{2}^{\mathrm{T}} = \mathbf{U}_{1}\mathbf{S}_{1}\mathbf{U}_{1}^{\mathrm{T}}.$$

S has k non-zero diagonal elements (S₁), N − k zero diagonal elements (S₂ = 0).

- ▶ $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$, with $\mathbf{U}_1 \in \mathcal{R}^{N \times k}$ and $\mathbf{U}_2 \in \mathcal{R}^{N \times N k}$
- *k*-SVD of **K** can be computed in $O(N^2k)$.
- $\mathbf{U}_1^{\mathrm{T}}\mathbf{y}$ and $\mathbf{U}_1^{\mathrm{T}}\mathbf{X}$ can be computed in $O(N \cdot k)$ per SNP.

[Lippert et al., 2011]

C. Lippert & O. Stegle

FaST low rank Mixed Models Using low rank genetic similarity matrices

Let K be rank k.

$$\mathbf{K} = \mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} = \mathbf{U}_{1}\mathbf{S}_{1}\mathbf{U}_{1}^{\mathrm{T}} + \mathbf{U}_{2}\mathbf{S}_{2}\mathbf{U}_{2}^{\mathrm{T}} = \mathbf{U}_{1}\mathbf{S}_{1}\mathbf{U}_{1}^{\mathrm{T}}.$$

S has k non-zero diagonal elements (S₁), N − k zero diagonal elements (S₂ = 0).

- ▶ $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$, with $\mathbf{U}_1 \in \mathcal{R}^{N \times k}$ and $\mathbf{U}_2 \in \mathcal{R}^{N \times N k}$
- k-SVD of **K** can be computed in $O(N^2k)$.
- $\mathbf{U}_1^{\mathrm{T}}\mathbf{y}$ and $\mathbf{U}_1^{\mathrm{T}}\mathbf{X}$ can be computed in $O(N \cdot k)$ per SNP.

[Lippert et al., 2011]

C. Lippert & O. Stegle

FaST low rank Mixed Models Using low rank genetic similarity matrices

Let K be rank k.

$$\mathbf{K} = \mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} = \mathbf{U}_{1}\mathbf{S}_{1}\mathbf{U}_{1}^{\mathrm{T}} + \mathbf{U}_{2}\mathbf{S}_{2}\mathbf{U}_{2}^{\mathrm{T}} = \mathbf{U}_{1}\mathbf{S}_{1}\mathbf{U}_{1}^{\mathrm{T}}.$$

S has k non-zero diagonal elements (S₁), N − k zero diagonal elements (S₂ = 0).

- ▶ $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$, with $\mathbf{U}_1 \in \mathcal{R}^{N \times k}$ and $\mathbf{U}_2 \in \mathcal{R}^{N \times N k}$
- k-SVD of **K** can be computed in $O(N^2k)$.
- $\mathbf{U}_1^{\mathrm{T}}\mathbf{y}$ and $\mathbf{U}_1^{\mathrm{T}}\mathbf{X}$ can be computed in $O(N \cdot k)$ per SNP.

$$\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{S}+\delta\mathbf{I}_{n}\right)\right)$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

FaST low rank Mixed Models Using low rank genetic similarity matrices

Let K be rank k.

$$\mathbf{K} = \mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} = \mathbf{U}_{1}\mathbf{S}_{1}\mathbf{U}_{1}^{\mathrm{T}} + \mathbf{U}_{2}\mathbf{S}_{2}\mathbf{U}_{2}^{\mathrm{T}} = \mathbf{U}_{1}\mathbf{S}_{1}\mathbf{U}_{1}^{\mathrm{T}}.$$

S has k non-zero diagonal elements (S₁), N − k zero diagonal elements (S₂ = 0).

- ▶ $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2]$, with $\mathbf{U}_1 \in \mathcal{R}^{N \times k}$ and $\mathbf{U}_2 \in \mathcal{R}^{N \times N k}$
- k-SVD of **K** can be computed in $O(N^2k)$.
- $\mathbf{U}_1^{\mathrm{T}}\mathbf{y}$ and $\mathbf{U}_1^{\mathrm{T}}\mathbf{X}$ can be computed in $O(N \cdot k)$ per SNP.

$$\mathcal{N}\left(\mathbf{U}^{\mathrm{T}}\mathbf{y}|\mathbf{U}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{S}+\delta\mathbf{I}_{n}
ight)
ight)$$

$$\neq \mathcal{N}\left(\mathbf{U}_{1}^{\mathrm{T}}\mathbf{y} | \mathbf{U}_{1}^{\mathrm{T}}\mathbf{X}\boldsymbol{\beta}; \sigma_{g}^{2}\left(\mathbf{S}_{1} + \delta\mathbf{I}_{d}\right)\right).$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

FaST low rank Mixed Models Using low rank genetic similarity matrices

 $\log \mathcal{N}\left(\mathbf{y} | \mathbf{X} \boldsymbol{\beta} ; \sigma_{g}^{2} \left(\mathbf{K} + \delta \mathbf{I}\right)\right).$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 44

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\log \mathcal{N}\left(\mathbf{y}|\mathbf{X}oldsymbol{eta};\sigma_{\mathrm{g}}^{2}\left(\mathbf{K}+\delta\mathbf{I}
ight)
ight).$$

$$= -\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right).$$

[Lippert et al., 2011]

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\log \mathcal{N}\left(\mathbf{y}|\mathbf{X}\boldsymbol{\beta};\sigma_{g}^{2}\left(\mathbf{K}+\delta\mathbf{I}\right)\right).$$

$$= -\frac{1}{2} \left(n \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right).$$

[Lippert et al., 2011]

3

・ロト ・四ト ・ヨト ・ヨト

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$= -\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X}\beta)^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X}\beta) \right).$$

 $\log \mathcal{N} \left(\mathbf{v} | \mathbf{X} \boldsymbol{\beta} \cdot \boldsymbol{\sigma}^2 \left(\mathbf{K} + \delta \mathbf{I} \right) \right)$

[Lippert et al., 2011]

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$-\frac{1}{2}\left(N\log\left(2\pi\sigma_{\rm g}^2\right) + \log|\mathbf{K} + \delta\mathbf{I}| + \frac{1}{\sigma_{\rm g}^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\rm T}(\mathbf{K} + \delta\mathbf{I})^{-1}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right).$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$-\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right).$$
$$\log |\mathbf{K} + \delta \mathbf{I}|$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

・ロト ・ 四ト ・ モト ・ モト

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{aligned} -\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right) .\\ \log |\mathbf{K} + \delta \mathbf{I}| \\\log \left| \mathbf{U} \mathbf{S} \mathbf{U}^{\mathrm{T}} + \delta \mathbf{I} \right| \end{aligned}$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 45

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$-\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right).$$
$$\log |\mathbf{K} + \delta \mathbf{I}|$$
$$\log |\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} + \delta \mathbf{I}|$$
$$\log |\mathbf{U}| |\mathbf{S} + \delta \mathbf{I}| \left| \mathbf{U}^{\mathrm{T}} \right|$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{aligned} -\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right) .\\ \log |\mathbf{K} + \delta \mathbf{I}| \\ \log |\mathbf{U}\mathbf{S}\mathbf{U}^{\mathrm{T}} + \delta \mathbf{I}| \\ \log |\mathbf{U}| |\mathbf{S} + \delta \mathbf{I}| \left| \mathbf{U}^{\mathrm{T}} \right| \end{aligned}$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 45

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{aligned} -\frac{1}{2} \left(N \log \left(2\pi \sigma_{\rm g}^2 \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{\rm g}^2} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\rm T} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right) \cdot \\ \log |\mathbf{K} + \delta \mathbf{I}| \\ \log |\mathbf{U}\mathbf{S}\mathbf{U}^{\rm T} + \delta \mathbf{I}| \\ \log |\mathbf{U}\mathbf{S}\mathbf{U}^{\rm T} + \delta \mathbf{I}| \\ \log |\mathbf{U}| |\mathbf{S} + \delta \mathbf{I}| \left| \mathbf{U}^{\rm T} \right| \end{aligned}$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{aligned} -\frac{1}{2} \left(N \log \left(2\pi \sigma_{\rm g}^2 \right) + \log \left| \mathbf{K} + \delta \mathbf{I} \right| + \frac{1}{\sigma_{\rm g}^2} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\rm T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \left(\mathbf{y} - \mathbf{X} \boldsymbol{\beta} \right) \right). \\ \log \left| \mathbf{K} + \delta \mathbf{I} \right| \\ \log \left| \mathbf{U} \mathbf{S} \mathbf{U}^{\rm T} + \delta \mathbf{I} \right| \\ \log \left| \mathbf{U} \mathbf{I} \right| + \delta \mathbf{I} \right| \\ \log \left| \mathbf{U} \right| \left| \mathbf{S} + \delta \mathbf{I} \right| \left| \mathbf{U}^{\rm T} \right| \\ \frac{1}{2} \sum_{i=1}^{N} \log \left(\frac{1}{s_{n,n} + \delta} \right) \\ \sum_{i=1}^{k} \log \left(\frac{1}{s_{n,n} + \delta} \right) + (N - k) \log \delta \end{aligned}$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$-\frac{1}{2}\left(N\log\left(2\pi\sigma_{\rm g}^2\right) + \log|\mathbf{K} + \delta\mathbf{I}| + \frac{1}{\sigma_{\rm g}^2}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\rm T}\left(\mathbf{K} + \delta\mathbf{I}\right)^{-1}\left(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\right)\right).$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

▶ ◀ ≣ ▶ ≣ ∽ (September 2012 46

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$-\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X}\beta)^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X}\beta) \right).$$
$$\mathbf{a} = (\mathbf{y} - \mathbf{X}\beta)$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 46

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{split} -\frac{1}{2} \left(N \log \left(2\pi \sigma_{\rm g}^2 \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{\rm g}^2} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\rm T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right) . \\ \mathbf{a} &= (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \\ \mathbf{a}^{\rm T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{a} \end{split}$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 46

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{split} &-\frac{1}{2} \left(N \log \left(2 \pi \sigma_{\rm g}^2 \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{\rm g}^2} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\rm T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right) \\ & \mathbf{a} = (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \\ & \mathbf{a}^{\rm T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{a} \\ & \mathbf{a}^{\rm T} \left(\mathbf{U} \mathbf{S} \mathbf{U}^{\rm T} + \delta \mathbf{I} \right)^{-1} \mathbf{a} \end{split}$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 46

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{split} -\frac{1}{2} \left(N \log \left(2\pi \sigma_{\rm g}^2 \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{\rm g}^2} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\rm T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right), \\ \mathbf{a} &= (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \\ \mathbf{a}^{\rm T} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{a} \\ \mathbf{a}^{\rm T} \left(\mathbf{U} \mathbf{S} \mathbf{U}^{\rm T} + \delta \mathbf{I} \right)^{-1} \mathbf{a} \\ \mathbf{a}^{\rm T} \mathbf{U} \left(\mathbf{S} + \delta \mathbf{I} \right)^{-1} \mathbf{U}^{\rm T} \mathbf{a} \end{split}$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 46

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{aligned} -\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right), \\ \mathbf{a} &= (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \qquad \mathbf{a}^{\mathrm{T}} [\mathbf{U}_{1}, \mathbf{U}_{2}] (\mathbf{S} + \delta \mathbf{I})^{-1} [\mathbf{U}_{1}, \mathbf{U}_{2}]^{\mathrm{T}} \mathbf{a} \\ \mathbf{a}^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} \mathbf{a} \\ \mathbf{a}^{\mathrm{T}} \left(\mathbf{U} \mathbf{S} \mathbf{U}^{\mathrm{T}} + \delta \mathbf{I} \right)^{-1} \mathbf{a} \\ \mathbf{a}^{\mathrm{T}} \mathbf{U} (\mathbf{S} + \delta \mathbf{I})^{-1} \mathbf{U}^{\mathrm{T}} \mathbf{a} \end{aligned}$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

: ▶ ◀ ≣ ▶ ≣ ∽ ९ (September 2012 46

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{aligned} -\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X}\beta)^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} (\mathbf{y} - \mathbf{X}\beta) \right). \\ \mathbf{a} &= (\mathbf{y} - \mathbf{X}\beta) \\ \mathbf{a}^{\mathrm{T}} \left(\mathbf{U}, \mathbf{U}_{2} \right) \left(\mathbf{S} + \delta \mathbf{I} \right)^{-1} \left[\mathbf{U}_{1}, \mathbf{U}_{2} \right]^{\mathrm{T}} \mathbf{a} \\ \mathbf{a}^{\mathrm{T}} \left(\mathbf{K} + \delta \mathbf{I} \right)^{-1} \mathbf{a} \\ \mathbf{a}^{\mathrm{T}} \left(\mathbf{U} \mathbf{S} \mathbf{U}^{\mathrm{T}} + \delta \mathbf{I} \right)^{-1} \mathbf{a} \\ \mathbf{a}^{\mathrm{T}} \mathbf{U} \left(\mathbf{S} + \delta \mathbf{I} \right)^{-1} \mathbf{U}^{\mathrm{T}} \mathbf{a} \end{aligned}$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{aligned} -\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X}\beta)^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X}\beta) \right) . \\ \mathbf{a} &= (\mathbf{y} - \mathbf{X}\beta) \\ \mathbf{a}^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} \mathbf{a} \\ \mathbf{a}^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} \mathbf{a} \\ \mathbf{a}^{\mathrm{T}} \left(\mathbf{U} \mathbf{S} \mathbf{U}^{\mathrm{T}} + \delta \mathbf{I} \right)^{-1} \mathbf{a} \\ \mathbf{a}^{\mathrm{T}} \left(\mathbf{U} \mathbf{S} \mathbf{U}^{\mathrm{T}} + \delta \mathbf{I} \right)^{-1} \mathbf{a} \\ \mathbf{a}^{\mathrm{T}} \mathbf{U} (\mathbf{S} + \delta \mathbf{I})^{-1} \mathbf{U}^{\mathrm{T}} \mathbf{a} \\ (\mathbf{U}_{1}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\mathbf{S}_{1} + \delta \mathbf{I}_{k})^{-1} (\mathbf{U}_{1}^{\mathrm{T}} \mathbf{a}) + \\ (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\mathbf{S}_{2} + \delta \mathbf{I}_{N-k})^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \end{aligned}$$

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

・ロト ・ 四ト ・ ヨト ・ ヨト

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$-\frac{1}{2}\left(N\log\left(2\pi\sigma_{g}^{2}\right)+\log\left|\mathbf{K}+\delta\mathbf{I}\right|+\frac{1}{\sigma_{g}^{2}}(\mathbf{y}-\mathbf{X}\boldsymbol{\beta})^{\mathrm{T}}\left(\mathbf{K}+\delta\mathbf{I}\right)^{-1}\left(\mathbf{y}-\mathbf{X}\boldsymbol{\beta}\right)\right).$$
$$(\mathbf{U}_{2}^{\mathrm{T}}\mathbf{a})^{\mathrm{T}}\left(\mathbf{S}_{2}+\delta\mathbf{I}_{N-k}\right)^{-1}\left(\mathbf{U}_{2}^{\mathrm{T}}\mathbf{a}\right)$$

 $\blacktriangleright O(N \cdot k)$ per SNP

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 47

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$-\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right).$$
$$(\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} \left(\delta \mathbf{I}_{N-k} \right)^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})$$

 $\blacktriangleright O(N \cdot k)$ per SNP

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 47

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$-\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X}\beta)^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X}\beta) \right).$$
$$(\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\delta \mathbf{I}_{N-k})^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})$$
$$\delta^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{U}_{2} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})$$

 $\blacktriangleright O(N \cdot k)$ per SNP

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

▶ ◀ ≣ ▶ ≣ ∽ (September 2012 47

< ロ > < 同 > < 回 > < 回 > < 回 > <

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{split} &-\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right). \\ & \left(\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a} \right)^{\mathrm{T}} \left(\delta \mathbf{I}_{N-k} \right)^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ & \delta^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{U}_{2} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ & \delta^{-1} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \end{split}$$

► $O(N \cdot k)$ per SNP

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 47

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

• $O(N \cdot k)$ per SNP

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{aligned} &-\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right). \\ & \left(\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a} \right)^{\mathrm{T}} (\delta \mathbf{I}_{N-k})^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ & \mathbf{U} \mathbf{U}^{\mathrm{T}} = \mathbf{I} \\ & \delta^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{U}_{2} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ & \left[\mathbf{U}_{1}, \mathbf{U}_{2} \right] [\mathbf{U}_{1}, \mathbf{U}_{2}]^{\mathrm{T}} = \mathbf{I} \\ & \delta^{-1} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \end{aligned}$$

 $\triangleright O(N \cdot k)$ per SNP

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 47

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{aligned} &-\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right). \\ & \left(\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a} \right)^{\mathrm{T}} \left(\delta \mathbf{I}_{N-k} \right)^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ & \mathbf{U} \mathbf{U}^{\mathrm{T}} = \mathbf{I} \\ & \delta^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{U}_{2} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ & \left[\mathbf{U}_{1}, \mathbf{U}_{2} \right] [\mathbf{U}_{1}, \mathbf{U}_{2}]^{\mathrm{T}} = \mathbf{I} \\ & \delta^{-1} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ & \mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}} + \mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} = \mathbf{I} \end{aligned}$$

 $\blacktriangleright O(N \cdot k)$ per SNP

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

イロト イロト イヨト イヨト 三日

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{aligned} -\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X}\beta)^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X}\beta) \right). \\ (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} \left(\delta \mathbf{I}_{N-k} \right)^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ \delta^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{U}_{2} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ \delta^{-1} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ \delta^{-1} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ \mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}} + \mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} = \mathbf{I} \\ \mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} = \mathbf{I} - \mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}} \end{aligned}$$

$\blacktriangleright O(N \cdot k)$ per SNP

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 47

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$\begin{aligned} -\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta})^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X} \boldsymbol{\beta}) \right) \\ & \left(\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a} \right)^{\mathrm{T}} \left(\delta \mathbf{I}_{N-k} \right)^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ & \mathbf{U} \mathbf{U}^{\mathrm{T}} = \mathbf{I} \\ & \delta^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{U}_{2} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ & \delta^{-1} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a}) \\ \delta^{-1} ((\mathbf{I} - \mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}}) \mathbf{a})^{\mathrm{T}} ((\mathbf{I} - \mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}}) \mathbf{a}) \\ \end{bmatrix} \end{aligned}$$

 $\triangleright O(N \cdot k)$ per SNP

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

3

FaST low rank Mixed Models Using low rank genetic similarity matrices

$$-\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X}\beta)^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X}\beta) \right).$$

$$(\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\delta \mathbf{I}_{N-k})^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})$$

$$\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$$

$$\delta^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{U}_{2} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})$$

$$[\mathbf{U}_{1}, \mathbf{U}_{2}] [\mathbf{U}_{1}, \mathbf{U}_{2}]^{\mathrm{T}} = \mathbf{I}$$

$$\delta^{-1} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})$$

$$\mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}} + \mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} = \mathbf{I}$$

$$\delta^{-1} ((\mathbf{I} - \mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}}) \mathbf{a})^{\mathrm{T}} ((\mathbf{I} - \mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}}) \mathbf{a})$$

$$\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} = \mathbf{I} - \mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}}$$

$$\delta^{-1} (\mathbf{a} - \mathbf{U}_{1} (\mathbf{U}_{1}^{\mathrm{T}} \mathbf{a}))^{\mathrm{T}} (\mathbf{a} - \mathbf{U}_{1} (\mathbf{U}_{1}^{\mathrm{T}} \mathbf{a}))$$

 $\triangleright O(N \cdot k)$ per SNP

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

3
FaST low rank Mixed Models Using low rank genetic similarity matrices

$$-\frac{1}{2} \left(N \log \left(2\pi \sigma_{g}^{2} \right) + \log |\mathbf{K} + \delta \mathbf{I}| + \frac{1}{\sigma_{g}^{2}} (\mathbf{y} - \mathbf{X}\beta)^{\mathrm{T}} (\mathbf{K} + \delta \mathbf{I})^{-1} (\mathbf{y} - \mathbf{X}\beta) \right).$$

$$(\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\delta \mathbf{I}_{N-k})^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})$$

$$\mathbf{U}\mathbf{U}^{\mathrm{T}} = \mathbf{I}$$

$$\delta^{-1} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{U}_{2} (\mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})$$

$$[\mathbf{U}_{1}, \mathbf{U}_{2}] [\mathbf{U}_{1}, \mathbf{U}_{2}]^{\mathrm{T}} = \mathbf{I}$$

$$\delta^{-1} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})^{\mathrm{T}} (\mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} \mathbf{a})$$

$$\mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}} + \mathbf{U}_{2} \mathbf{U}_{2}^{\mathrm{T}} = \mathbf{I}$$

$$\delta^{-1} ((\mathbf{I} - \mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}}) \mathbf{a})^{\mathrm{T}} ((\mathbf{I} - \mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}}) \mathbf{a})$$

$$U_{2} \mathbf{U}_{2}^{\mathrm{T}} = \mathbf{I} - \mathbf{U}_{1} \mathbf{U}_{1}^{\mathrm{T}}$$

$$\delta^{-1} (\mathbf{a} - \mathbf{U}_{1} (\mathbf{U}_{1}^{\mathrm{T}} \mathbf{a}))^{\mathrm{T}} (\mathbf{a} - \mathbf{U}_{1} (\mathbf{U}_{1}^{\mathrm{T}} \mathbf{a}))$$

▶ $O(N \cdot k)$ per SNP

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

3

< ロ > < 同 > < 回 > < 回 > < 回 > <

FaST Linear Mixed Models

Using covariance genetic similarity matrices

Bottleneck:

k-spectral decomposition of general K is $O(N^2 \cdot k)$.

k-spectral decomposition of Covariance be computed in O(N · S_c · k).
 x̄_c ∈ R^{N×S_c}.

• For $N>S_{
m c}$ and $k=S_{
m c},$ this is linear in N

▶ Total runtime becomes $O(N \cdot S_c^2)$ plus $O(N \cdot S_c)$ per SNP.

[Lippert et al., 2011]

イロト イポト イヨト イヨト

FaST Linear Mixed Models

Using covariance genetic similarity matrices

Bottleneck:

k-spectral decomposition of general K is $O(N^2 \cdot k)$.

- ► k-spectral decomposition of Covariance be computed in $O(N \cdot S_{c} \cdot k)$.
 - $\tilde{\mathbf{X}}_{c} \in \mathcal{R}^{N \times S_{c}}$.
 - For $N > S_c$ and $k = S_c$, this is linear in N.
- ▶ Total runtime becomes $O(N \cdot S_c^2)$ plus $O(N \cdot S_c)$ per SNP.

$$\mathbf{K}_{\mathsf{c}} = \frac{1}{S_{\mathsf{c}}} \tilde{\mathbf{X}}_{\mathsf{c}} \tilde{\mathbf{X}}_{\mathsf{c}}^{\mathrm{T}}$$
(6)

イロト イポト イヨト イヨト

[Lippert et al., 2011]

FaST Linear Mixed Models

Using covariance genetic similarity matrices

Bottleneck:

k-spectral decomposition of general **K** is $O(N^2 \cdot k)$.

- ► k-spectral decomposition of Covariance be computed in $O(N \cdot S_{c} \cdot k)$.
 - $\tilde{\mathbf{X}}_{c} \in \mathcal{R}^{N \times S_{c}}$.
 - For $N > S_{c}$ and $k = S_{c}$, this is linear in N.
- ▶ Total runtime becomes $O(N \cdot S_c^2)$ plus $O(N \cdot S_c)$ per SNP.

$$\mathbf{K}_{\mathsf{c}} = \frac{1}{S_{\mathsf{c}}} \tilde{\mathbf{X}}_{\mathsf{c}} \tilde{\mathbf{X}}_{\mathsf{c}}^{\mathrm{T}}$$
(6)

イロト イポト イヨト イヨト

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 48

FaST Linear Mixed Models

Using covariance genetic similarity matrices

Bottleneck:

k-spectral decomposition of general K is $O(N^2 \cdot k)$.

- ► k-spectral decomposition of Covariance be computed in $O(N \cdot S_{c} \cdot k)$.
 - $\tilde{\mathbf{X}}_{c} \in \mathcal{R}^{N \times S_{c}}$.
 - For $N > S_{c}$ and $k = S_{c}$, this is linear in N.
- ▶ Total runtime becomes $O(N \cdot S_c^2)$ plus $O(N \cdot S_c)$ per SNP.

$$\mathbf{K}_{\mathsf{c}} = \frac{1}{S_{\mathsf{c}}} \tilde{\mathbf{X}}_{\mathsf{c}} \tilde{\mathbf{X}}_{\mathsf{c}}^{\mathrm{T}}$$
(6)

[Lippert et al., 2011]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 48

Run-time comparison

- 32GB dual six-core AMD machine running Linux
- Run-times on single core.
- ► WTCCC
- GAW14.x smoking synthetic

 Generated:
 GUMM generative models
 GUMM (generative models)
 Joy, Sor, JON, 2054, 505 and 1005

[Burton et al., 2007]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 49

3

< ロ > < 同 > < 回 > < 回 > < 回 > <

Run-time comparison

- 32GB dual six-core AMD machine running Linux
- Run-times on single core.
- WTCCC
- ► GAW14.x smoking synthetic
 - Original:

 Generated:
 GLMM generative model
 GLMM generative model
 SL, Str. 20c, 30c and doi: 100

[Burton et al., 2007]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 49

3

< ロ > < 同 > < 回 > < 回 > < 回 > <

Run-time comparison

- 32GB dual six-core AMD machine running Linux
- Run-times on single core.
- ► WTCCC
- GAW14.x smoking synthetic
 - Original:

```
    Generated:
    GLMM generative model
    1x, 5x, 10x, 20x, 50x and
100x
```

[Burton et al., 2007]

train	4k SNPs	8k SNPs
EMMA-X	470 min	538 min
FaST-LMM	28 min	79 min

イロト イポト イヨト イヨト

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

▶ ▲ 重 ▶ 重 少 Q ○ September 2012 49

Run-time comparison

- 32GB dual six-core AMD machine running Linux
- Run-times on single core.
- WTCCC
- GAW14.x smoking synthetic
 - Original:
 - 1.2k individuals
 - 8k SNPs
 - Generated:
 - GLMM generative model
 - 1x, 5x, 10x, 20x, 50x and 100x

[Burton et al., 2007]

train	4k SNPs	8k SNPs
EMMA-X	470 min	538 min
FaST-LMM	28 min	79 min

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 49

- ∢ ≣ →

Run-time comparison

- 32GB dual six-core AMD machine running Linux
- Run-times on single core.
- WTCCC
- GAW14.x smoking synthetic
 - Original:
 - 1.2k individuals
 - 8k SNPs
 - Generated:
 - GLMM generative model
 - 1x, 5x, 10x, 20x, 50x and 100x

[Burton et al., 2007]

train	4k SNPs	8k SNPs
EMMA-X	470 min	538 min
FaST-LMM	28 min	79 min

C. Lippert & O. Stegle

Run-time comparison

- 32GB dual six-core AMD machine running Linux
- Run-times on single core.
- WTCCC
- GAW14.x smoking synthetic
 - Original:
 - 1.2k individuals
 - 8k SNPs
 - Generated:
 - GLMM generative model
 - 1x, 5x, 10x, 20x, 50x and 100x

train	4k SNPs	8k SNPs
EMMA-X	470 min	538 min
FaST-LMM	28 min	79 min



→ ∃ →

[Burton et al., 2007]

C. Lippert & O. Stegle

Linear models I: linear regression and linear mixed models

September 2012 49

- - E - N

References I

- Y. Benjamini and Y. Hochberg. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society. Series B* (*Methodological*), pages 289–300, 1995.
- P. Burton, D. Clayton, L. Cardon, N. Craddock, P. Deloukas, A. Duncanson, D. Kwiatkowski, M. McCarthy, W. Ouwehand, N. Samani, et al. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature*, 447(7145):661–678, 2007.
- B. Devlin and K. Roeder. Genomic control for association studies. *Biometrics*, 55(4):997–1004, 1999.
- H. M. Kang, N. A. Zaitlen, C. M. Wade, A. Kirby, D. Heckerman, M. J. Daly, and E. Eskin. Efficient control of population structure in model organism association mapping. *Genetics*, 107, 2008.
- C. Lippert, J. Listgarten, Y. Liu, C. Kadie, R. Davidson, and D. Heckerman. Fast linear mixed models for genome-wide association studies. *Nature Methods*, 8(10):838;835, 10 2011. doi: 10.1038/nmeth.1681.
- J. Novembre, T. Johnson, K. Bryc, Z. Kutalik, A. R. Boyko, A. Auton, A. Indap, K. S. King, S. Bergmann, M. R. Nelson, M. Stephens, and C. D. Bustamante. Genes mirror geography within europe. *Nature*, 456(7218):98–101, Nov. 2008.
- N. Patterson, A. L. Price, and D. Reich. Population structure and eigenanalysis. *PLoS Genetics*, 2(12):e190+, December 2006.

イロト 不得下 イヨト イヨト 二日

References II

A. L. Price, N. J. Patterson, R. M. Plenge, M. E. Weinblatt, N. A. Shadick, and D. Reich. Principal components analysis corrects for stratification in genome-wide association studies. *Nature genetics*, 38(8):904–909, August 2006.

イロト イポト イヨト イヨト