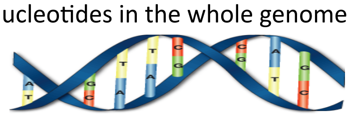
Structured Sparsity in Genetics

Seyoung Kim
sssykim@cs.cmu.edu
Lane Center for Computational Biology
Carnegie Mellon University

1

Datasets in Genetics

- Analysis of high-dimensional genomic datasets
  - Human genome
    - 3.2 billion nucleotides in the whole genome



    - >3 million genetic polymorphisms

    - >25,000 genes, whose expression-levels can be measured with microarray technology

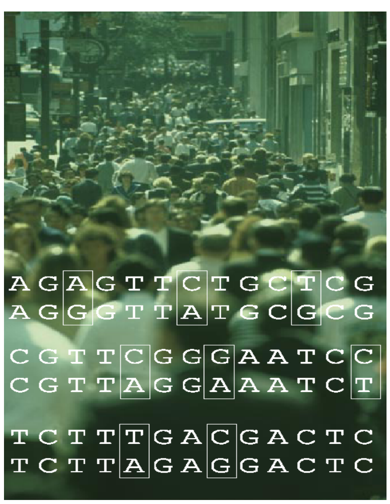
2

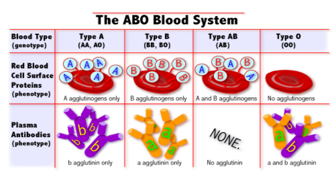
Sparsity and Genetics

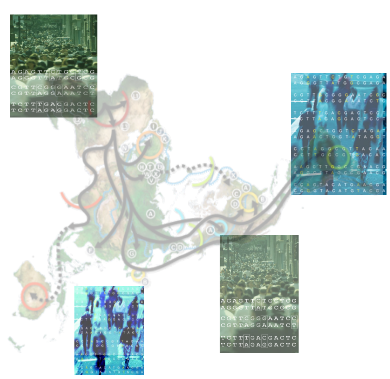
- Why sparsity in analysis of genomic datasets?
  - Biological systems are highly modular
  - Genes are organized into functional modules or pathways
  - Although genome data are very high-dimensional, each meaningful piece of information often involves much fewer entities
  - Sample size is significantly smaller than the number of dimensions
- The key research question
  - How is the information encoded in the genome expressed into observed phenotypes?
  - Which genes are responsible for each phenotypes?
    - E.g., height, obesity, disease susceptibility (cancer, diabetes, etc.), gene expression levels

3

Genome Polymorphisms

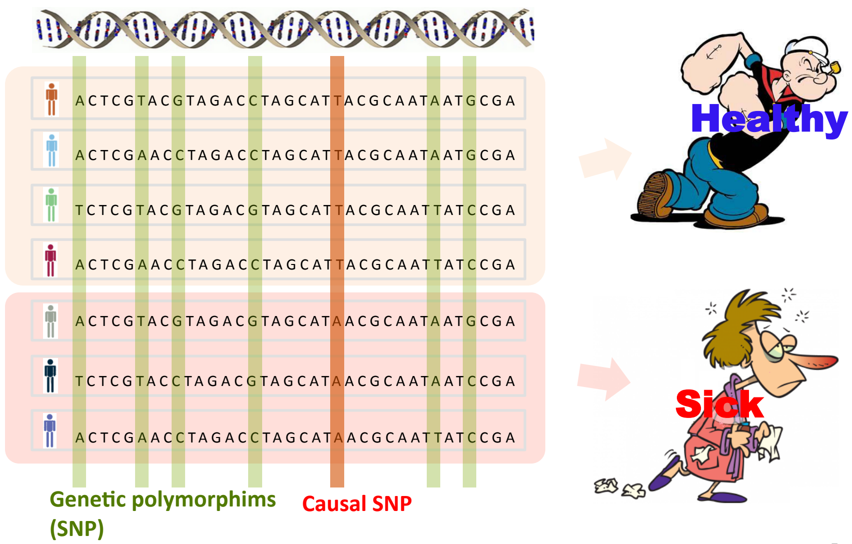






4

Genetic Basis of Diseases



5

Overview

- **Sparsity**: lasso for analysis of genome-phenome data
  - Linear regression with L_1 penalty
- From sparsity to **structured sparsity**: extending the lasso approach
  - Fused lasso and group lasso for simple structures
  - Fused lasso and group lasso as building blocks for more complex structures
- Although we assume a linear regression model, the approach can be applied to other types of models

6

Multivariate Regression Model

Trait		Genotype		Association Strength
2.1	=	T G A A C C A T G A A G T A	x	?
y	=	X	x	β

7

Regression Estimation

Trait		Genotype		Association Strength
2.1	=	T G A A C C A T G A A G T A	x	

$$\underset{\beta}{\operatorname{argmin}}\ (y - X\beta)' \cdot (y - X\beta)$$

Many non-zero associations:
Which SNPs are truly significant?

8

Lasso for Selecting Relevant Inputs

(Tibshirani, 1996)

Trait | Genotype | Association Strength

2.1 = T G A A C C A T G A A G T A x

Lasso Penalty
for sparsity

$$\underset{\beta}{\arg\min}\ (y - X\beta)'\cdot(y - X\beta) + \lambda \sum_{j=1}^{J} |\beta_j|$$

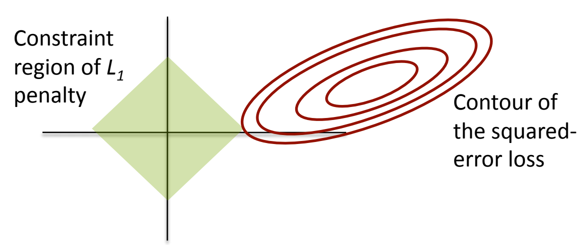
What if there are multiple related traits?

9

L1 Regularization (Lasso)

- Enforcing sparsity

$$\hat{\beta} = \arg\min_{\beta} \|\mathbf{Y} - \mathbf{X}\beta\|^2 + \lambda\|\beta\|_1$$



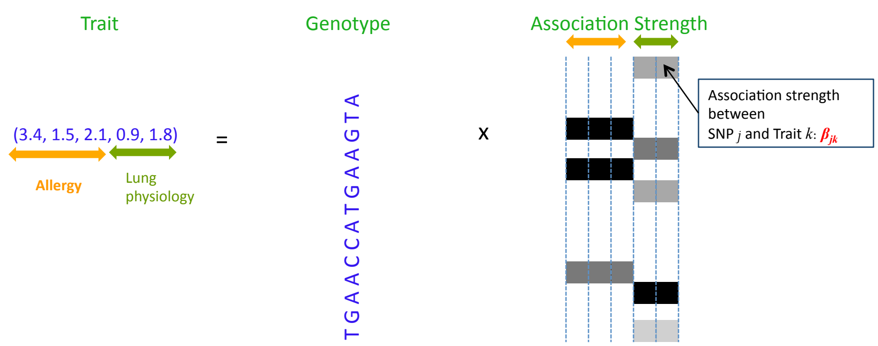
10

Sparsity

- **Makes statistical sense:** Learning is now feasible in high dimensions with small sample size
- **Makes biological sense:** each phenotype is likely to be influenced by a small number of SNPs, rather than all the SNPs.

11

What is Structured Sparsity?



12

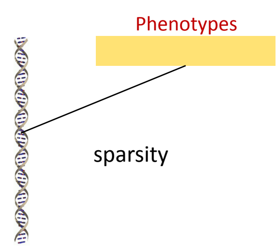
What is Structured Sparsity?

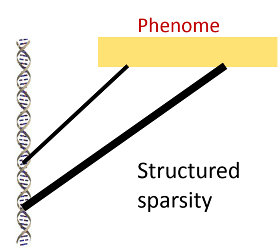
- Leveraging prior knowledge
  - Structure on the inputs
    - Inputs are ordered in time, and adjacent inputs are jointly relevant
    - Inputs are grouped, and inputs in the same group are jointly relevant
  - Structure on the outputs
    - Outputs are ordered in time, and adjacent outputs have the same relevant inputs
    - Outputs are grouped, and outputs in the same group are jointly relevant

13

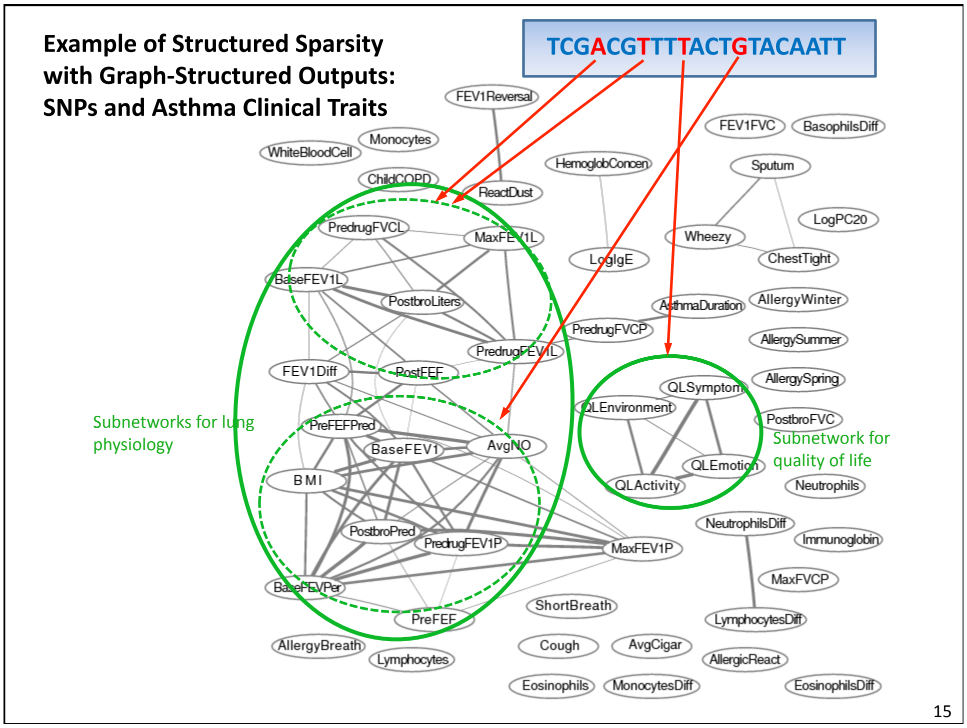
Why Structured Sparsity?

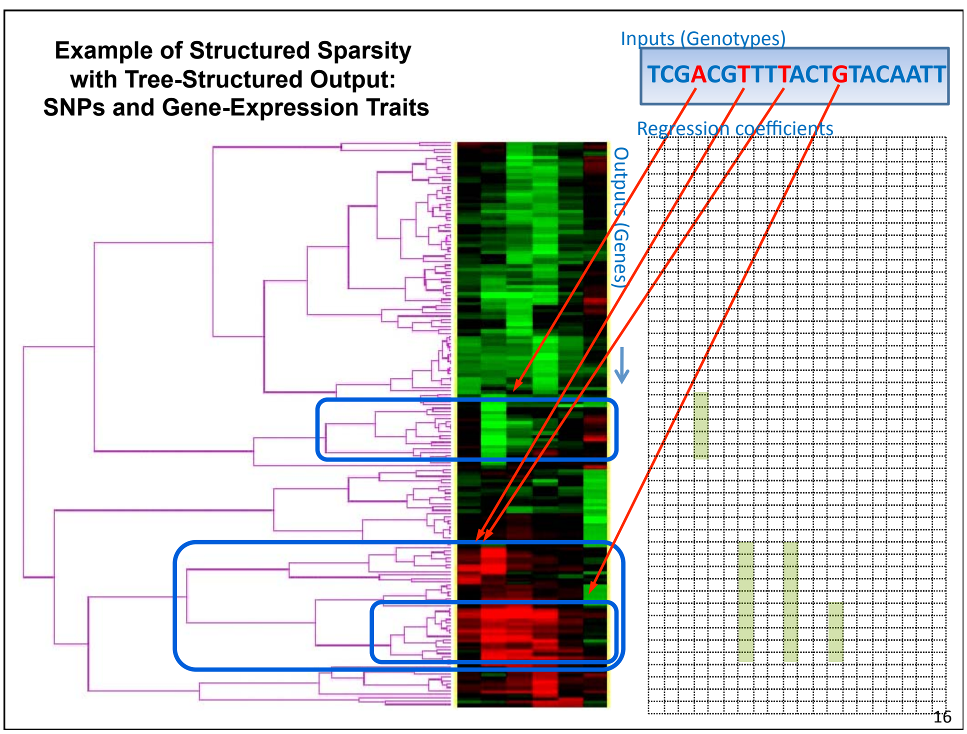
- Advantages of learning structured sparsity
  - Improve the power for detecting true relevant inputs by
    - Leveraging prior information
    - Combining statistical strength across multiple inputs and outputs
  - Reduce false positives in the estimated set of relevant inputs
  - Sparsity pattern in the parameters with structure can lead to more meaningful and interpretable results



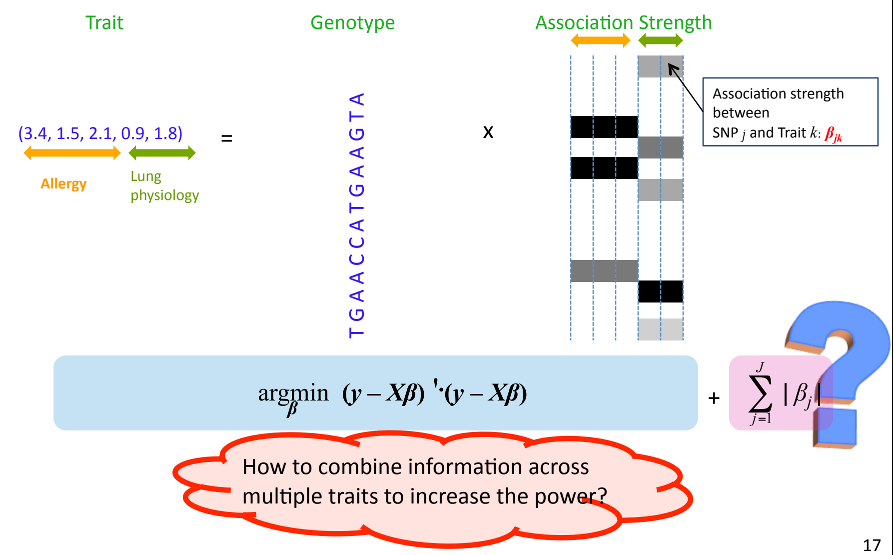


14


Example of Structured Sparsity
with Graph-Structured Outputs:
SNPs and Asthma Clinical Traits
TCGACGTTTTACTGTACAATT
FEV1Reversal
FEV1FVC
BasophilsDiff
Monocytes
WhiteBloodCell
HemoglobConcen
Sputum
ChildCOPD
ReactDust
LogPC20
PredrugFVCL
MaxFEV1L
Wheezy
LogIgE
ChestTight
BaseFEV1L
PostbroLiters
AsthmaDuration
AllergyWinter
PredrugFVCP
AllergySummer
PredrugFEV1L
FEV1Diff
PostFEF
QLSymptom
AllergySpring
QLEnvironment
PostbroFVC
Subnetworks for lung physiology
PreFEFPred
BaseFEV1
AvgNO
Subnetwork for quality of life
QLEmotion
BMI
QLActivity
Neutrophils
NeutrophilsDiff
PostbroPred
Immunoglobin
PredrugFEV1P
MaxFEV1P
MaxFVCP
BaseFEVPer
ShortBreath
PreFEF
LymphocytesDiff
AllergyBreath
Cough
AvgCigar
Lymphocytes
AllergicReact
Eosinophils
MonocytesDiff
EosinophilsDiff
15


Example of Structured Sparsity
with Tree-Structured Output:
SNPs and Gene-Expression Traits
Inputs (Genotypes)
TCGACGTTTTACTGTACAATT
Regression coefficients
Outputs (Genes)
16

Learning Structured Sparsity



$$\arg\min_{\beta} (y - X\beta)'\cdot(y - X\beta) + \sum_{j=1}^{J} |\beta_j|$$

17

Structured Sparsity

- Structured sparsity for simple structures
  - Fused lasso
  - Group lasso
- We will use fused lasso and group lasso as a building block to construct other penalty functions for incorporating more complex structures

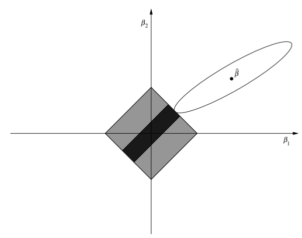
18

Fused Lasso

- Fused lasso (Tibshirani et al., 2004)
  - Assumes a chain structure over inputs (ordered in time)
  - Goal: select adjacent inputs as relevant jointly – structured sparsity!

minimize $L(\lambda_1, \lambda_2, \beta) = |\mathbf{y} - \mathbf{X}\beta|^2 + \lambda_2|\beta|^2 + \lambda_1 \sum_{j=2}^{p} |\beta_j - \beta_{j-1}|$

  - Penalizes the difference between β_j and β_{j-1}
  - Encourages β_j and β_{j-1} to take similar values



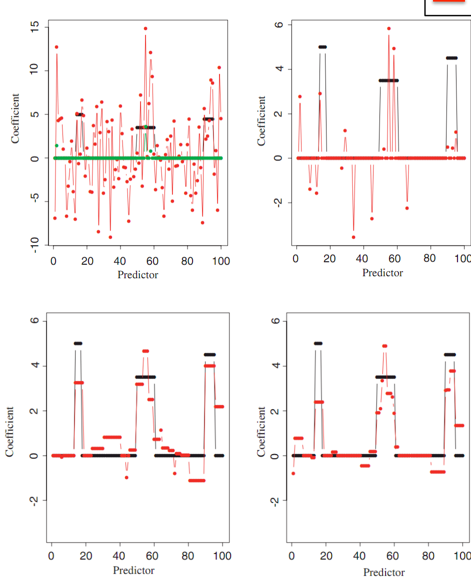
19

Fused Lasso

True regression coefficients
Estimated

Standard regression

lasso



Fusion penalty only

Fused lasso

20

Group Lasso

- Group lasso (Yuan and Lin, 2006)
  - Assumes the groups over inputs are known
  - Goal: select groups of inputs (rather than individual inputs) as relevant to the output – structured sparsity!

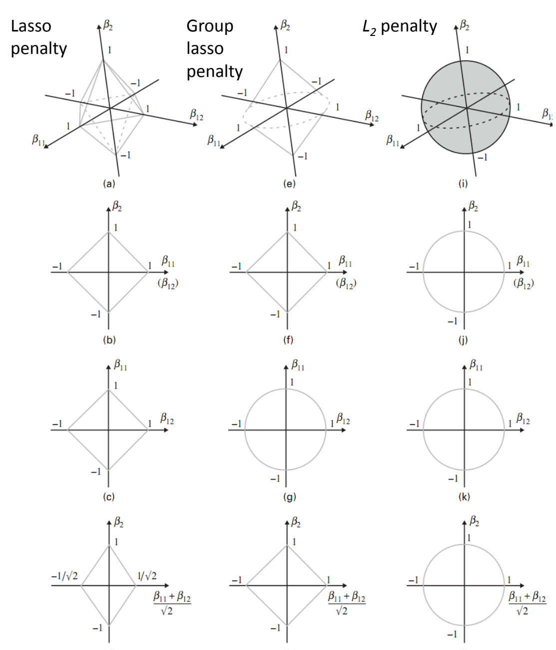
minimize $$L(\lambda_1, \lambda_2, \beta) = |\mathbf{y} - \mathbf{X}\beta|^2 + \lambda \sum_{j=1}^{J} \|\beta_j\|_{L1/L2}$$

$$\|\beta_j\|_{L1/L2} = \sqrt{\sum_k \beta_{jk}^{\ 2}}$$

  - L1/L2 penalty
    - L1 component (lasso penalty) performs sparse selection
    - L2 component (ridge penalty) enforces the β_{jk}'s in the same group to be selected jointly as non-zero

21

Contour Surfaces of Penalty Functions

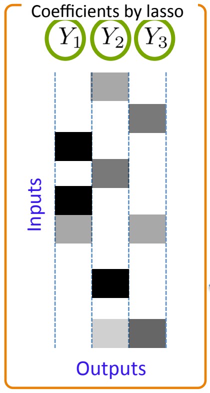


22

Extending Group Lasso

- L_1/L_2 penalty for group lasso can be used in various different ways
  - L_1/L_2-regularized multli-task regression

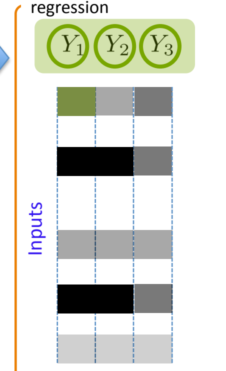
Regression Coefficients by lasso



$$\text{argmin} \sum_k (\mathbf{y}_k - X\beta_k)' \cdot (\mathbf{y}_k - X\beta_k) + \lambda \sum_j \sqrt{\sum_k \beta_{jk}^2}$$

$$\text{argmin} \sum_k (\mathbf{y}_k - X\beta_k)' \cdot (\mathbf{y}_k - X\beta_k) + \lambda \sum_k \sum_j |\beta_{jk}|$$

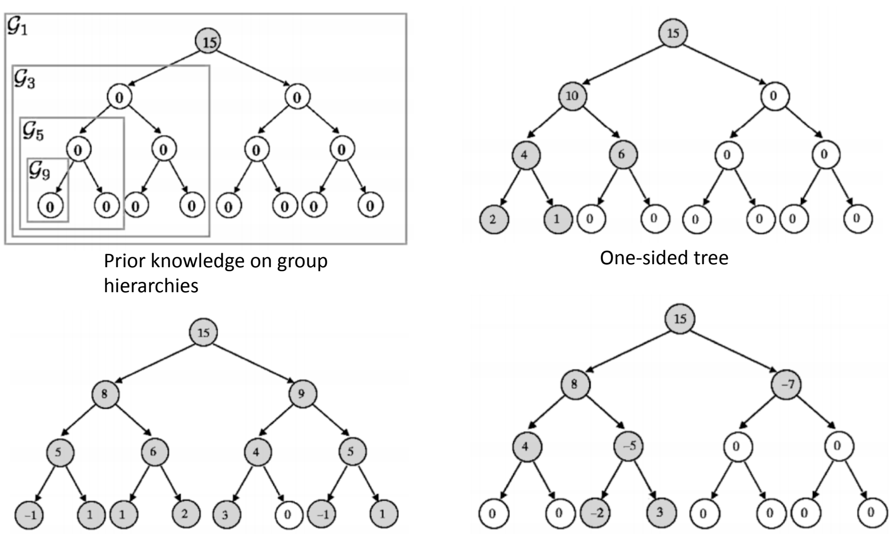
Regression Coefficients By L1/L2 multi-task regression



23

Hierarchical Selection with Nested Groups

(Zhao, Rocha, and Yu, 2009)



Prior knowledge on group hierarchies

One-sided tree

Almost-complete tree

Regular tree

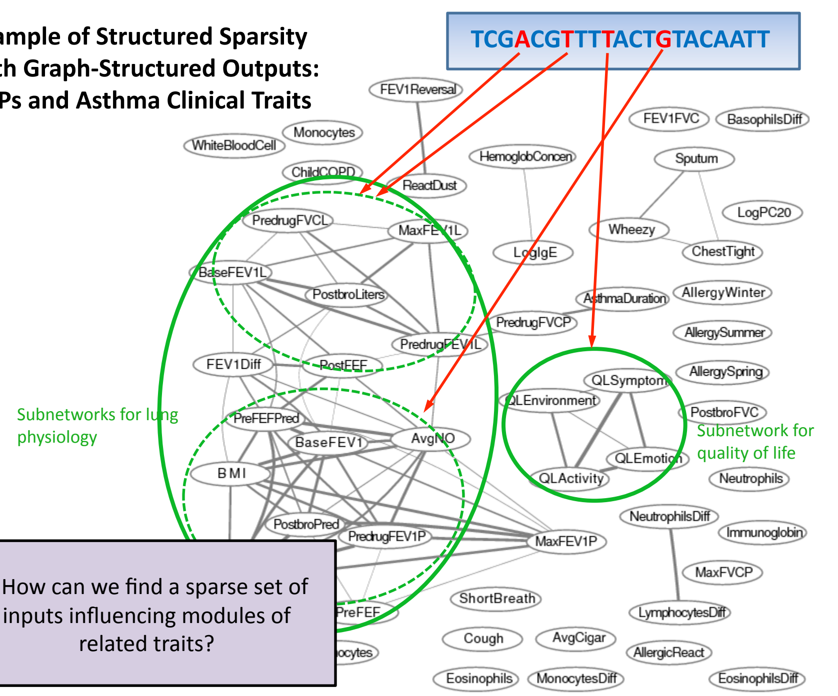
24

Structured Sparsity

- Structured sparsity for simple structures
  - Fused lasso
  - Group lasso
- We will use fused lasso and group lasso as a building block to construct other penalty functions for incorporating more complex structure
  - Graph-guided fused lasso
  - Temporally smoothed lasso
  - Tree-guided group lasso

25

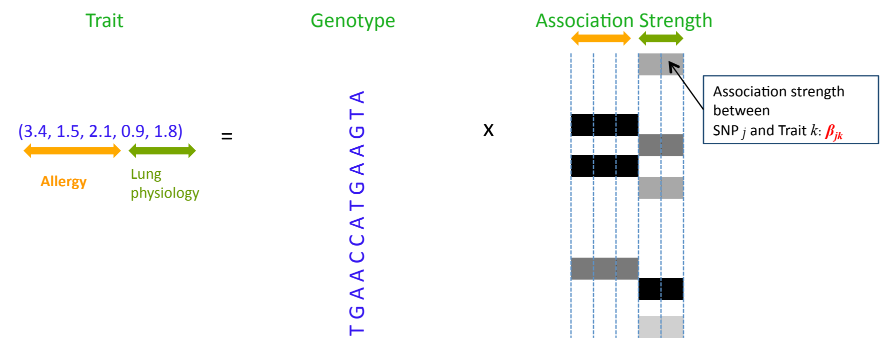
Example of Structured Sparsity with Graph-Structured Outputs: SNPs and Asthma Clinical Traits



How can we find a sparse set of inputs influencing modules of related traits?

26

Graph-Guided Fused Lasso



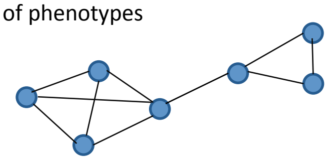
$$\underset{\beta}{\operatorname{argmin}}\ (y - X\beta)' \cdot (y - X\beta) + \sum_{j=1}^{J} |\beta_j|$$

+ We introduce graph-guided fusion penalty

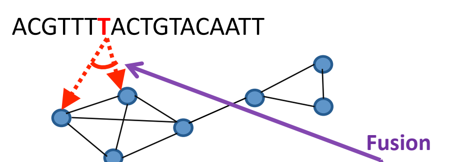
27

Graph-Constrained Fused Lasso

Step 1: Thresholded correlation graph of phenotypes



Step 2: Graph-constrained fused lasso



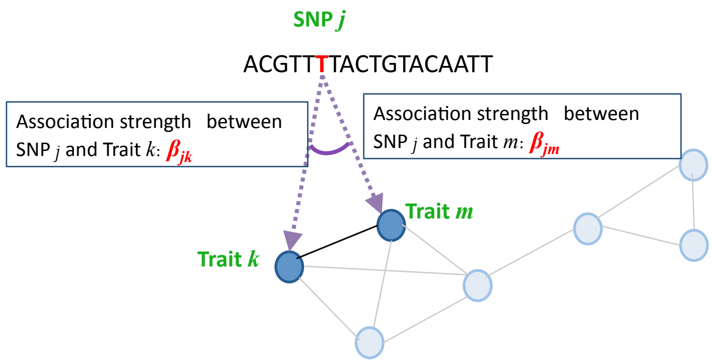
$$\hat{\mathbf{B}}^{GC} = \operatorname{argmin} \sum_k (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k)^T \cdot (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k) + \lambda \sum_k \sum_j |\beta_{jk}| + \gamma \sum_{(m,l)\in E} \sum_j |\beta_{jm} - \operatorname{sign}(r_{ml})\beta_{jl}|$$

Lasso Penalty

Graph-constrained fusion penalty

28

Fusion Penalty

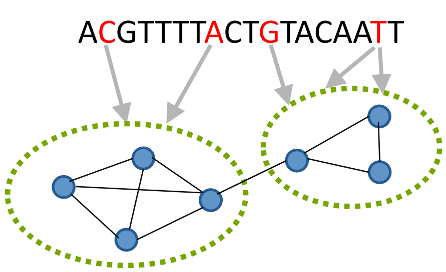


- Fusion Penalty: $|\beta_{jk} - \beta_{jm}|$
- For two correlated traits (connected in the network), the association strengths may have similar values.

29

Graph-Constrained Fused Lasso

Overall effect

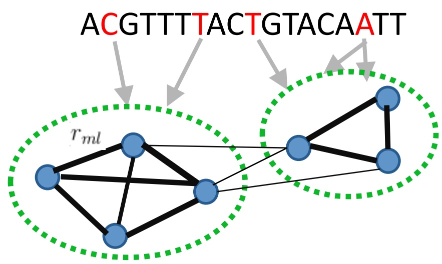


- Fusion effect propagates to the entire network
- Association between SNPs and subnetworks of traits

30

Graph-Weighted Fused Lasso

Overall effect



- Subnetwork structure is embedded as a densely connected nodes with large edge weights
- Edges with small weights are effectively ignored

31

Estimating Parameters

- Quadratic programming formulation
  - Graph-constrained fused lasso

$$\hat{\mathbf{B}}^{\mathrm{GC}} = \operatorname{argmin} \sum_k (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k)^T \cdot (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k)$$

$$\text{s. t.} \quad \sum_k \sum_j |\beta_{jk}| \le s_1 \text{ and } \sum_{(m,l)\in E} \sum_j |\beta_{jm} - \mathrm{sign}(r_{ml})\beta_{jl}| \le s_2$$

  - Graph-weighted fused lasso

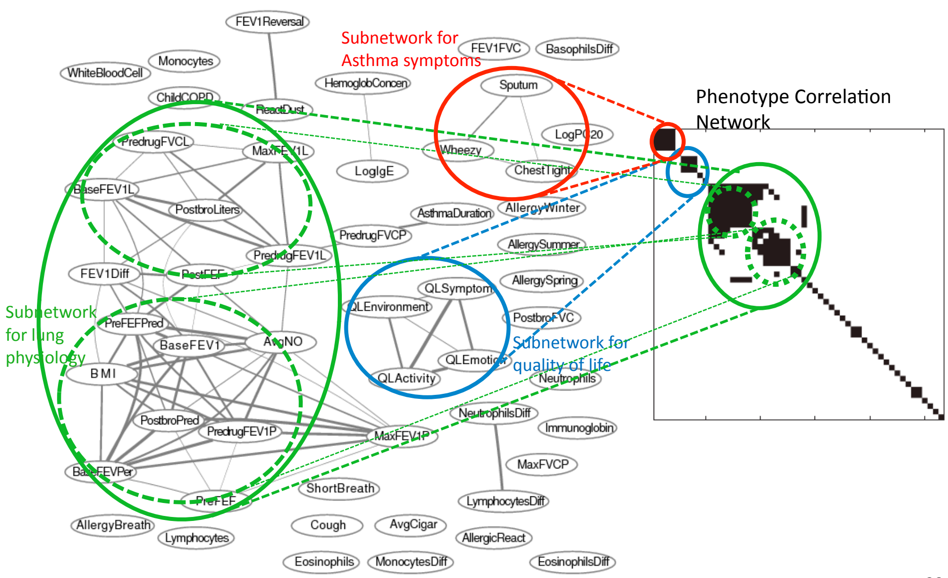
$$\hat{\mathbf{B}}^{\mathrm{GW}} = \operatorname{argmin} \sum_k (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k)^T \cdot (\mathbf{y}_k - \mathbf{X}\boldsymbol{\beta}_k)$$

$$\text{s. t.} \quad \sum_k \sum_j |\beta_{jk}| \le s_1 \text{ and } \sum_{(m,l)\in E} f(r_{ml}) \sum_j |\beta_{jm} - \mathrm{sign}(r_{ml})\beta_{jl}| \le s_2$$

- Many publicly available software packages for solving convex optimization problems can be used
- Slow! - we will discuss more efficient proximal gradient method later.

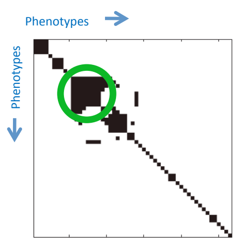
32

Asthma Trait Network



33

Comparison of Gflasso with Others



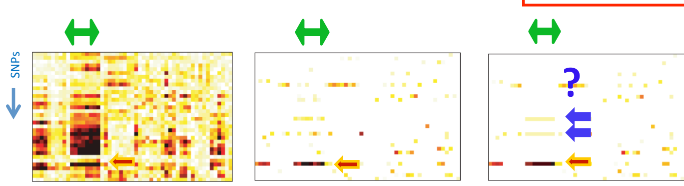
Trait Network

Lung physiology-related traits I

- Baseline FEV1 predicted value: MPVLung
- Pre FEF 25-75 predicted value
- Average nitric oxide value: online
- Body Mass Index
- Postbronchodilation FEV1, liters: Spirometry
- Baseline FEV1 % predicted: Spirometry
- Baseline predrug FEV1, % predicted
- Baseline predrug FEV1, % predicted

Q551R SNP

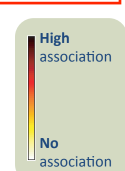
- Codes for amino-acid changes in the intracellular signaling portion of the receptor
- Exon 11



Single-Marker
Single-Trait Test

Lasso

Graph-guided
Fused Lasso



34

Structured Sparsity

- Structured sparsity for simple structures
  - Fused lasso
  - Group lasso
- We will use fused lasso and group lasso as a building block to construct other penalty functions for incorporating more complex structure
  - Graph-guided fused lasso
  - Temporally smoothed lasso
  - Tree-guided group lasso

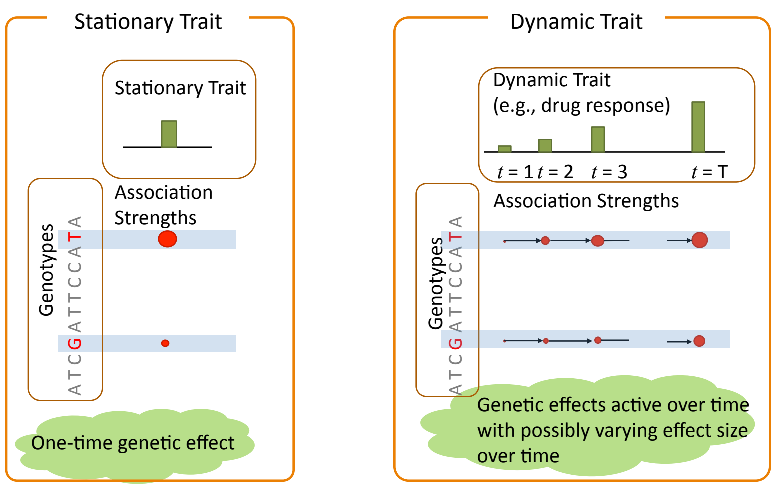
35

Time-Series Measurements of Traits

- Dynamic trait with a temporal trend
  - Growth of tumor over time
  - Height, weight over time
  - Gene expressions over time in cell cycle or embryonic development
- Are there underlying genetic variants (SNPs) that influence the overall trend over time?

36

Time-Series Measurements of Traits



37

Temporally-Smoothed Lasso

- Step 1: Autoregressive Model
  - Captures the shape of the temporal trend in the dynamic-trait data
  - Estimates the model parameters based on the dynamic-trait data only
- Step 2: Temporally-Smoothed Lasso
  - Penalized regression framework
  - Incorporates the estimated dynamic-trait shape parameters from Step 1
  - Detects time-varying genetic effects on the dynamic trait

38

Step 1: Autoregressive Model

Autoregressive Model :

$$\mathbf{y}_{k,t+1} = \alpha_{k,t}\mathbf{y}_{k,t} + \alpha^0_{k,t}\mathbf{1} + \epsilon$$

Estimating Model Parameters:

$$\hat{\alpha}_{k,t} = \text{argmin } (\mathbf{y}_{k,t+1} - \alpha_{k,t}\mathbf{y}_{k,t} - \alpha^0_{k,t})^T \cdot (\mathbf{y}_{k,t+1} - \alpha_{k,t}\mathbf{y}_{k,t} - \alpha^0_{k,t})$$

Estimates of the Model Parameters:

$$\hat{\alpha}_{k,t} = \frac{\mathbf{y}^T_{k,t} \cdot \mathbf{y}_{k,t+1}}{\mathbf{y}^T_{k,t} \cdot \mathbf{y}_{k,t}}$$

39

Step 2: Temporally-Smoothed Lasso

Autoregressive parameters from Step 1

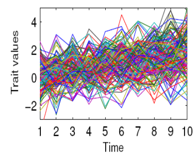
$$\hat{\mathrm{B}}^{\mathrm{dyn}} = \text{argmin} \sum_k \sum_t (\mathbf{y}_{k,t} - \mathrm{X}\beta_{k,t})^T \cdot (\mathbf{y}_{k,t} - \mathrm{X}\beta_{k,t}) + \lambda \cdot \sum_k \sum_t \sum_j \beta^j_{k,t} + \gamma \cdot \sum_j \sum_k \sum_{t=1}^{T-1} |\beta^j_{k,t+1} - \hat{\alpha}_{k,t}\beta^j_{k,t}|$$

Lasso Penalty

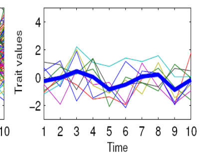
Temporally-smoothed Lasso Penalty

40

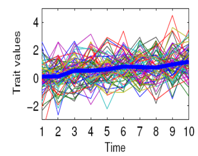
Simulation Study – Linear Dynamic



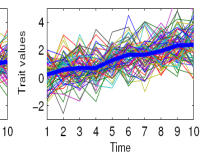
Trait data for all individuals



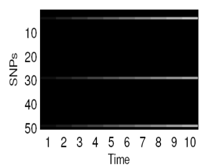
Trait data for individuals with no association SNPs



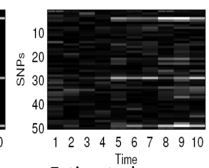
Trait data for individuals with 1-2 association SNPs



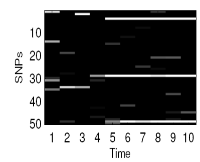
Trait data for individuals with >3 association SNPs



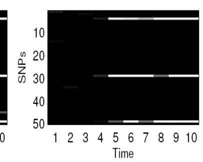
True association strength



Estimated association strength (single SNP analysis)

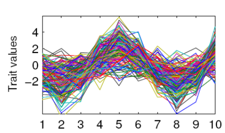


Estimated association strength (lasso)

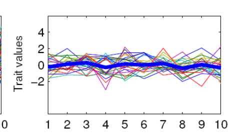


Estimated association strength (temporally-smoothed lasso)

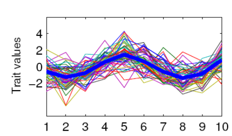
Simulation Study – Cyclic Dynamic



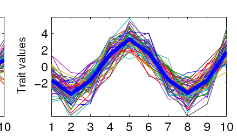
Trait data for all individuals



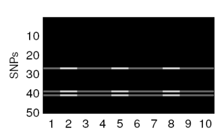
Trait data for individuals with no association SNPs



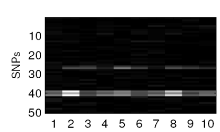
Trait data for individuals with 1-2 association SNPs



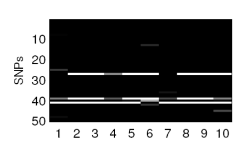
Trait data for individuals with >3 association SNPs



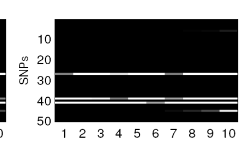
True association strength



Estimated association strength (single SNP analysis)



Estimated association strength (lasso)



Estimated association strength (temporally-smoothed lasso)

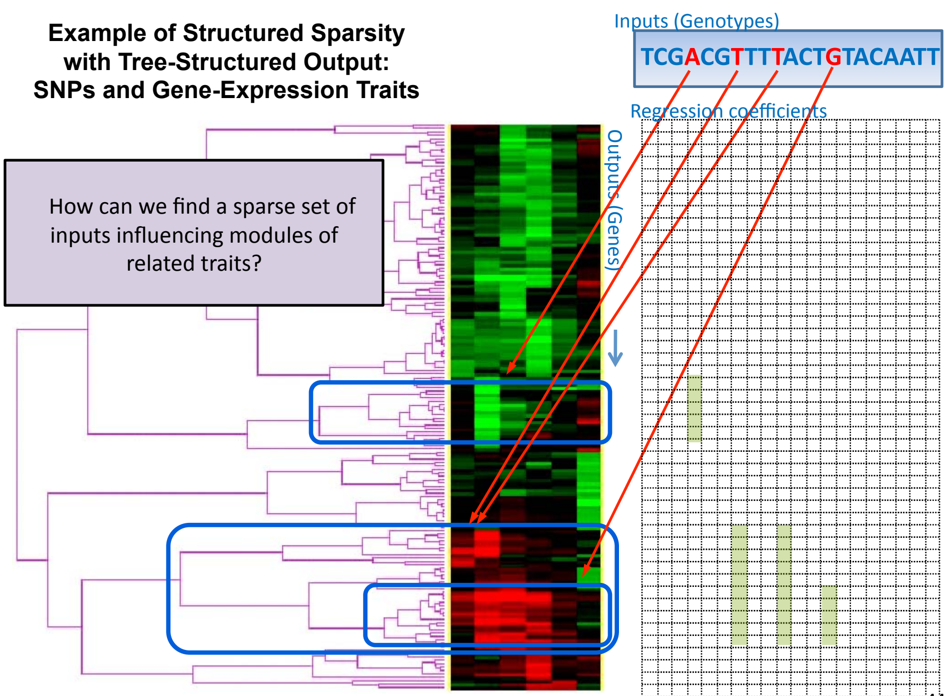
Structured Sparsity

- Structured sparsity for simple structures
  - Fused lasso
  - Group lasso
- We will use fused lasso and group lasso as a building block to construct other penalty functions for incorporating more complex structure
  - Graph-guided fused lasso
  - Temporally smoothed lasso
  - Tree-guided group lasso

43

Example of Structured Sparsity with Tree-Structured Output: SNPs and Gene-Expression Traits

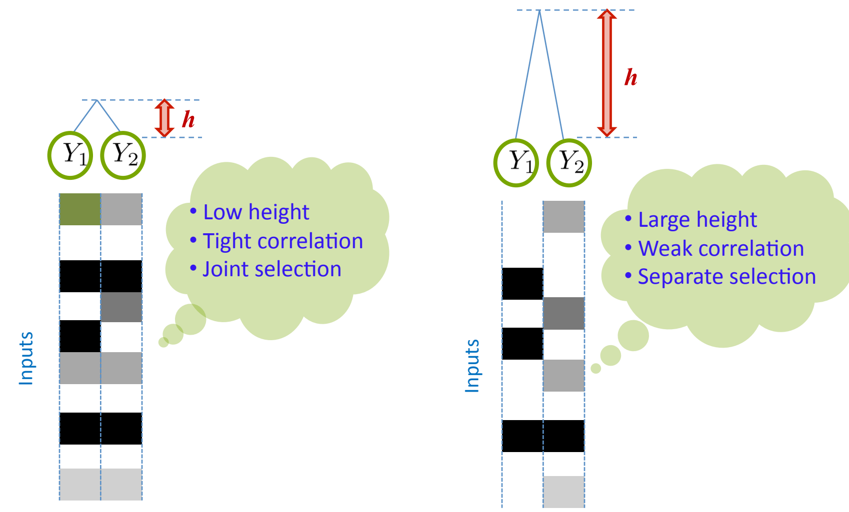
How can we find a sparse set of inputs influencing modules of related traits?



44

Tree-Guided Group Lasso

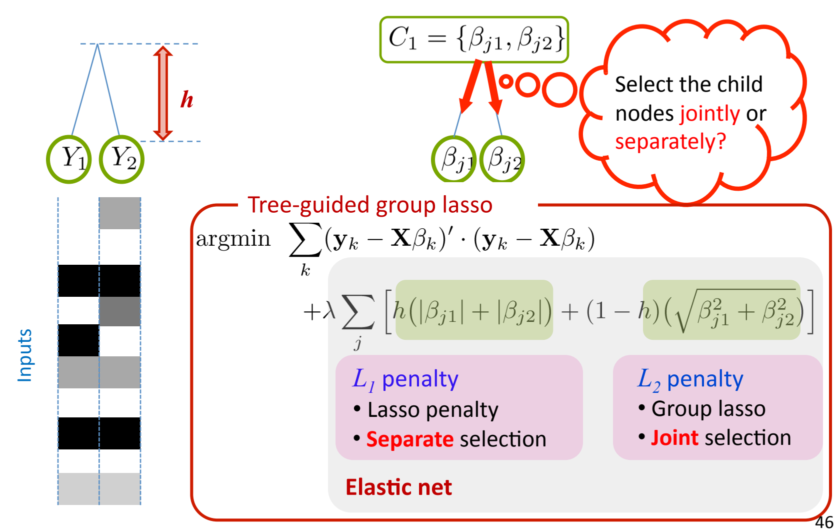
- In a simple case of two outputs



45

Tree-Guided Group Lasso

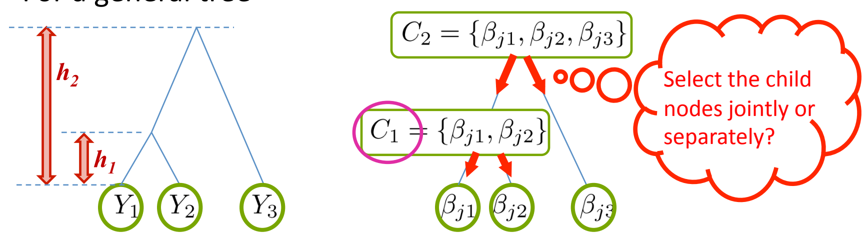
- In a simple case of two outputs



46

Tree-Guided Group Lasso

- For a general tree



Tree-guided group lasso

$$\text{argmin} \sum_k (\mathbf{y}_k - \mathbf{X}\beta_k)' \cdot (\mathbf{y}_k - \mathbf{X}\beta_k)$$

$$+\lambda \sum \left[(1-h_2)\left(\sqrt{\beta_{j1}^2 + \beta_{j2}^2 + \beta_{j3}^2}\right) + h_2\left(|C_1| + |\beta_{j3}|\right) \right]$$

$$(1-h_1)\left(\sqrt{\beta_{j1}^2 + \beta_{j2}^2}\right) + h_1\left(|\beta_{j1}| + |\beta_{j2}|\right)$$

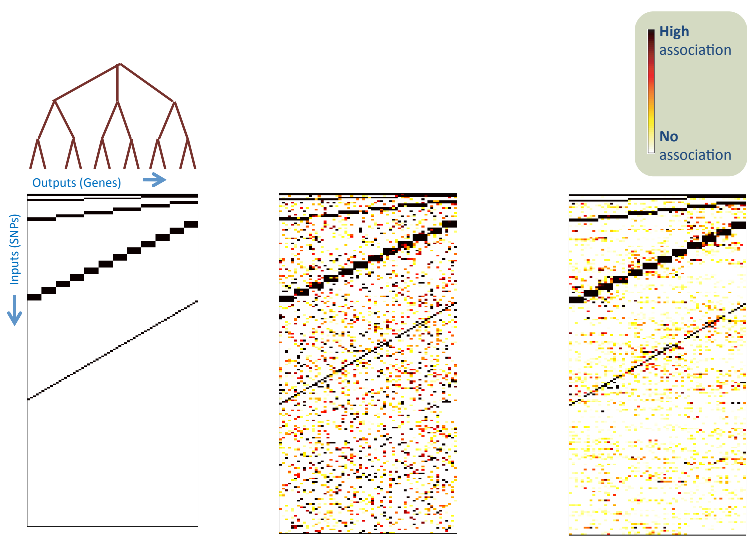
Joint selection

Separate selection

Note that the groups overlap!

47

Illustration with Simulated Data



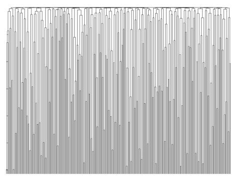
True regression coefficients

Lasso

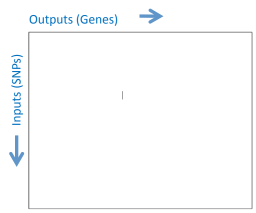
Tree-guided group lasso

48

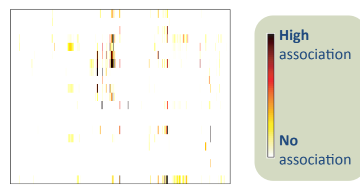
Analysis of Yeast Data



Hierarchical clustering tree for genes (outputs)



Lasso



Tree-guided group lasso

49

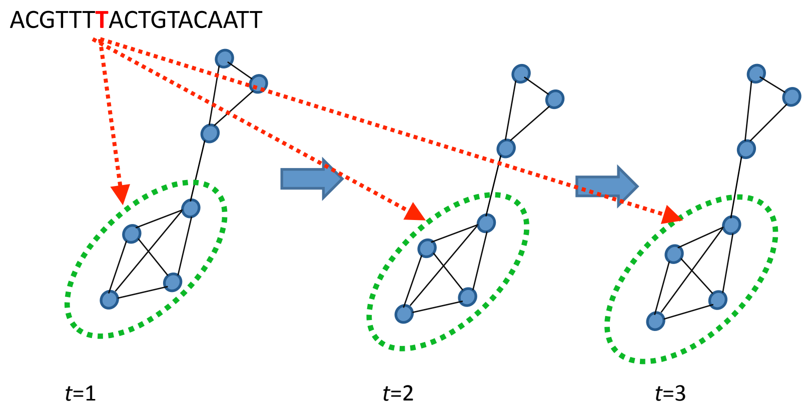
Towards More Complex Structured Sparsity Pattern

- Different types of penalties can be combined in a single objective function to achieve a more complex structured-sparsity pattern
- The regularization parameter for each penalty needs to be determined via cross-validation

50

Combining Penalties for Structured Sparsity

- How can we detect SNPs (inputs) influencing trait modules?
  - Graph-guided fused lasso + temporally-smoothed lasso



51

Improving Scalability

- Challenges in optimization
  - Non-smooth penalty
  - Non-separable penalty: each term in summation of the penalty function needs to contain non-overlapping set of parameters
- Convex program approach is too slow for genome-scale datasets
- Proximal-gradient method (Xi et al., submitted)
  - Handles non-separability by reformulation of the problem through dual norm
  - Handles non-smoothness by introducing smooth approximation of the non-smooth function
  - We can apply accelerated gradient method
  - General approach applicable that can be used for fused lasso, graph-guided fused lasso, tree-guided group lasso, temporally smoothed lasso, and many other penalized regression

52

Summary

- Key advantages of enforcing structured sparsity
  - Increase power for recovering true relevant features
  - Reduce false positives
  - Combine information across multiple inputs and outputs to produce more meaningful and interpretable results
- Methods for learning structured sparsity pattern in regression parameters
  - Assume the structure in inputs or outputs is known as prior knowledge
  - Otherwise, learn the structure from data in pre-processing step
  - Construct penalty functions that encode the structure information
    - Graph-guided fused lasso: graph structure over outputs
    - Temporally-smoothed lasso: chain structure for time-series outputs
    - Tree-guided group lasso: tree structure over outputs
  - Can be used in applications in genetics, computer vision, language modeling, etc.

53