**Causal Inference for Gene-Gene Interaction**

**Model for Gene-Gene Interaction:**

 (1)







And

.

Then, equation (1) can be written in a matrix form:

. (2)



For the i-th equation, we have

. (3)

Dividing both sides of the above equation by and replacing , by we obtain

 (4)

 (5)

where .

Multiplying by the matrix on both sides of the equation (3), we obtain

. (6)

Let . Then, equation (6) can be written as

. (7)

It is known that

. (8)

The generalized least square estimate is given by

. (9)

**Sparse SEMs and Alternating Direction Method of Multipliers**

In general, the genotype-phenotype networks are sparse. Therefore, and are sparse

matrices. In order to obtain sparse estimates of and , the natural approach is the -norm penalized regression of equation (7). Using weighted least square and -norm penalization, we can form the following optimization problem:

 (10)

The optimization problem (10) can be further reduced to

 (A1)

To solve the optimization problem (A1), we form the augmented Lagrangian

. (A2)

The alternating direction method of multipliers (ADMM) consists of the iterations:

 (A3)

 (A4)

 (A5)

where . Let . Equations (A1-A3) can be reduced to

 (A6)

 (A7)

 (A8)

Solving minimization problem (A6), we obtain

,

which can be reduced to

(A9)

The optimization problem (A7) is non-differentiable. Although the first term in (A7) is not differentiable, we still can obtain a simple closed-form solution to the problem (A7) using subdiffenrential calculus.29 Let be a generalized derivative of the -th component  of the vector  and  where



Then, we have



which implies that

 (A10)

where



The size of the genotype-phenotype network may be large. An efficient alternating direction method of multipliers (ADMM)29 is used to solve the optimization problem (7). The procedure f Algorithm:

For 

Step 1. Initialization



Carry out steps 2,3 and 4 until convergence

Step 2.



Step 3.



where



Step 4.



Under some assumptions convergence of ADMM can be proved.29 In practice, although it can be slow to converge to high accuracy, ADMM converges to modest accuracy within a few tens of iterations. When we consider large-scale problems and parameter estimation problems, modest accuracy is sufficient. Therefore, ADMM may work very well for structure and parameter estimation in the genotype-phenotype networks.

Most of the elements of matrices and are equal to zero. The regularized Lasso for two stage least squares approach and ADMM algorithms are expected to shrink most of the coefficient matricesand  toward zero, yielding sparse network structures. The sparsity-controlling parameter  will be estimated via cross validation.

**Sparse Structural Functional Linear Models for Gene-Gene Interaction**

Let *t* be a genomic position. Define a genotype profile  of the i-th individual as

 (2-1)

where  and  are two alleles of the marker at the genomic position *t*, and are the frequencies of the alleles  and , respectively. Suppose that we are interested in G genomic regions or genes [denoted as . We consider the following SFEMs:

 (2-2)

where are genetic effect functions and are interaction function.

For each genomic region or gene, we use functional principal component analysis to calculate principal component function.14 We expand in each genomic region in terms of orthogonal principal component functions:



where are the -th principal component function in the -th genomic region or gene and are the functional principal component scores of the -th individual. Equations (2.2) are reduced to

 (2-3)

Let

,

  
,

,

,



. The structural functional equations can be reduced in terms of functional principal component scores:

,

which can be rewritten as

 (2-4)

where .

Let . Then, equation (2-4) can be written as

. (2-5)

It is known that

. (2-6)

The generalized least square estimate is given by

. (2-7)

In general, the genotype-phenotype networks are sparse. Therefore, and are sparse

matrices. In order to obtain sparse estimates of and , the natural approach is the -norm penalized regression of equation (2-5). Using weighted least square and -norm penalization, we can form the following optimization problem:

 (2-8)