A Convex Formulation for Path Analysis in Structural Equation Modeling

Anupon Pruttiakaravanich[1†] and Jitkomut Songsiri[2]

[1]Department of Electrical Engineering, Chulalongkorn University, Bangkok, Thailand
(Tel: +66-2-424-9116; E-mail: anupon106@gmail.com)

[2]Department of Electrical Engineering, Chulalongkorn University, Bangkok, Thailand
(Tel: +66-2-218-6487; E-mail: jitkomut.s@chula.ac.th)

Abstract: Path analysis is a special problem in Structural Equation Modeling (SEM) where its model describes causal relations among measured variables in a form of multiple linear regression. This work presents an alternative estimation formulation for a special case problem in path analysis as a convex framework by relaxing an equality constraint of the original problem. Under a condition on problem parameters, we show that, our optimal solution is low rank and provides an estimate of the path matrix of the original problem. To solve our estimation problem in a convex framework, we apply the alternating direction method of multipliers (ADMM) algorithm which is suitable for large-scale implementation. The performance of this algorithm is demonstrated in numerical experiments.

Keywords: path analysis, confirmatory SEM, convex program

1. INTRODUCTION

Structural equation modeling (SEM) is a statistical technique used for seeking a statistical causal multivariate model in a form of multivariate linear regression of latent and measured variables. SEM has a long history since 1980s and is widely used in many behavioral researches such as in psychology [1], sociology [2], business [3], and many more; a history background can be found in [4, §1].

Path analysis is a special problem in SEM where it provides a model for explaining relationships among measured (or observed) variables only with additive error terms. The problem in path analysis starts from constructing a *hypothetical model* where directional paths from one variable to another are assumed from a prior knowledge about relationship structure of variables of interest. The formulation is then to estimate the value of nonzero entries in the path matrix and the covariance matrix of model residual errors so that the model-reproduced covariance matrix fits well with the sample covariance matrix in an optimal sense, evaluated by various types of criterion functions such as maximum likelihood, ordinary or weighted least-squares [4, §4]. When the zero structure of the path matrix is hypothetically given, the resulting estimation problem is called *confirmatory SEM* which can be found from many applications in behavioral research. In contrast, one may seek for a zero structure of the path matrix that best fits the data since its pattern reveals a causal structure of variables such as a problem of learning brain connectivity in [5-7]. The latter type of estimation problem is referred to as *exploratory SEM*.

Confirmatory and exploratory SEM problems are nonlinear optimization problems in matrix variables with quadratic equality and positive definite cone constraints. Many existing SEM commercial softwares, such as LISREL, EQS, Mplus [4, 8, 9], have been developed and they implement iterative methods, for instance Newton-Raphson or gradient descent, to estimate the model parameters [10, §7], [4, §4], so a starting value for the update is required. Though these numerical methods work well under normal conditions, it is also known that some initial values may not lead to the convergence in the optimal solution or may stuck into a local minima, hence several strategies for selecting initial values have been proposed [4, §4].

In this work, we present an alternative estimation formulation for a path analysis in confirmatory SEM. We relax the original nonlinear equality constraint of the model parameters to an inequality, allowing us to transform the problem into a convex formulation that lead to many efficient numerical methods. Our problem can be solved by the alternating direction method of multipliers (ADMM) algorithm which is suitable for large-scale implementation. We also show that, under a condition on problem parameter, our optimal covariance error is diagonal, meaning that errors are uncorrelated, and the optimal solution has low rank, providing an estimate of the path matrix for the original problem. Our estimate agrees with the original solution when the assumption on uncorrelated noise holds. When this does not hold, our solution is not optimal for the original problem but it can be served as a starting value for the iterative algorithm used in the original one in case that the convergence is not obtained.

Our paper is organized as follows. Section 2 summarizes the mathematical formulation of path analysis problem. Section 3 describes our convex formulation for confirmatory SEM and shows that the solution can be further used under the condition of having a low rank solution at optimum. Numerical methods for solving our formulation in large-scale framework are explained in section 4. All numerical experiments are demonstrated in section 5 and the conclusions are given in section 6.

Notation. $\mathbf{S}^n$ and $\mathbf{S}^n_+$ denote the set of $n \times n$ symmetric and positive definite matrices, respectively. For a square matrix X, $\mathbf{tr}(X)$ is the trace of X and $\mathbf{diag}(X)$ is a diagonal matrix containing diagonal entries of X.

† Anupon Pruttiakaravanich is the presenter of this paper.

2. BACKGROUND ON PATH ANALYSIS

In this section, we describe the mathematical formulation of path analysis in SEM. The model is given by a multiple linear regression:

$$Y = c + AY + \epsilon, \tag{1}$$

where $Y \in \mathbf{R}^n$ is the measured (or observed) variables, $c \in \mathbf{R}^n$ is a baseline vector, and $\epsilon \in \mathbf{R}^n$ is the model error, assumed to be Gaussian distributed. The matrix $A \in \mathbf{R}^{n \times n}$ denotes the *path matrix* where a_{ij} represents a causal relationship among variables in the model, *i.e.*, if $a_{ij} = 0$ then there is no path from Y_j to Y_i. If this structure is assumed from a *priori* knowledge, then the problem of estimating A is called *confirmatory SEM*.

Let S be a sample covariance matrix of Y, computed from sample measurements and Σ be the covariance matrix of Y, derived from (1).

$$\Sigma = (I - A)^{-1}\Psi(I - A)^{-T}, \tag{2}$$

where $\Psi = \mathbf{cov}(\epsilon)$. An estimation in SEM is to seek for A and Ψ such that the estimated Σ is closest to S in the sense that the Kullback-Leiber divergence function,

$$d(S, \Sigma) = \log\det\Sigma + \mathbf{tr}(S\Sigma^{-1}) - \log\det S - n,$$

is minimized, while maintaining that Σ, A and Ψ are related by (2). Moreover, the structure of the path matrix is presumably encoded by a model hypothesis: i) $A_{ij} = 0$ if there is no path from Y_j to Y_i and ii) we always have $\mathbf{diag}(A) = 0$, meaning that there is no path from Y_i to itself. To specify the zero structure of A, we then define the associated index set $I_A \subseteq \{1, 2, \ldots, n\} \times \{1, 2, \ldots, n\}$ with properties that i) $(i, j) \in I_A$ if $A_{ij} = 0$ and ii) $\{(1, 1), (2, 2), \ldots, (n, n)\} \subseteq I_A$ since $\mathbf{diag}(A) = 0$.

Given the index set I_A, we define a *projection operator* $P : \mathbf{R}^{n \times n} \to \mathbf{R}^{n \times n}$,

$$P(X) = \begin{cases} X_{ij}, & (i, j) \in I_A, \\ 0, & \text{otherwise}, \end{cases}$$

and denote $P^c = I - P$. The operators P^c and P are both self-adjoint, *i.e.*, $\mathbf{tr}(Y^T P(X)) = \mathbf{tr}(P(Y)^T X)$ and that $P^c(P(X)) = 0$. This projection operator will be used repeatedly in our analysis.

Therefore, with the definition of P and a change of variable $X = \Sigma^{-1}$, the estimation problem corresponding to confirmatory SEM is

$$\begin{array}{ll} \text{minimize} & -\log\det X + \mathbf{tr}(SX), \\ \text{subject to} & X = (I - A)^T\Psi^{-1}(I - A), \\ & P(A) = 0, \end{array} \tag{3}$$

with variables $A \in \mathbf{R}^{n \times n}, \Psi \in \mathbf{S}^n_+$ and $\Sigma \in \mathbf{S}^n_+$. The condition $P(A) = 0$ basically explains the zero constraint on the entries of A, and when there is no extra information on the path matrix, this condition reduces to $\mathbf{diag}(A) = 0$.

3. CONVEX FORMULATION FOR CONFIRMATORY SEM

The problem (3) is obviously nonconvex due to the quadratic equality constraint. In this section, we propose an alternative convex formulation and derive its dual problem. Our main result is that the optimal solution is useful only when it is low rank, which happens at optimum under a condition on problem parameter and results in a diagonal covariance error, meaning that the residual errors are uncorrelated.

Consider a convex relaxation of constraint (2) to $X \succeq (I - A)^T\Psi^{-1}(I - A)$. Using a property of Schur complement on this relaxed constraint, we propose an alternative convex formulation:

$$\begin{array}{ll} \text{minimize} & -\log\det X + \mathbf{tr}(SX), \\ \text{subject to} & \begin{bmatrix} X & (I - A)^T \\ I - A & \Psi \end{bmatrix} \succeq 0, \\ & 0 \preceq \Psi \preceq \alpha I, \quad P(A) = 0, \end{array} \tag{4}$$

with variables $X \in \mathbf{S}^n, A \in \mathbf{R}^{n \times n}$ and $\Psi \in \mathbf{S}^n$, where $\alpha > 0$ is a given parameter. We note that the inequality constraint $\Psi \preceq \alpha I$ is added to prevent (4) from having a trivial solution, *e.g.*, Ψ can be arbitrarily large. We justify that α can serve as an upper bound on the covariance error in SEM. Throughout this paper, we refer to (4) as the *primal convex SEM* which falls into a type of semidefinite programming. We can see that for a given α, a numerical solution can be solved by many existing convex program solvers. One example is `CVX` which is a MATLAB package for specifying and solving convex programs [11].

If we define a variable

$$X = \begin{bmatrix} X_1 & X_2^T \\ X_2 & X_4 \end{bmatrix}, \quad X_4 = \Psi, \quad X_2 = I - A,$$

we see that $P(X_2) = P(I) - P(A) = P(I) - 0 = I$ (note that the P projects the entries assigned by I_A which includes the diagonal terms). Another equivalent formulation of the primal is

$$\begin{array}{ll} \text{minimize} & -\log\det X_1 + \mathbf{tr}(SX_1), \\ \text{subject to} & X = \begin{bmatrix} X_1 & X_2^T \\ X_2 & X_4 \end{bmatrix} \succeq 0, \\ & 0 \preceq X_4 \preceq \alpha I, \quad P(X_2) = I, \end{array} \tag{5}$$

with variable $X \in \mathbf{S}^{2n}$ where each X_k has size $n \times n$. The dual problem of (4) is

$$\begin{array}{ll} \text{minimize} & -\log\det(S - Z_1) - 2\,\mathbf{tr}(Z_2) - \alpha\,\mathbf{tr}(Z_4), \\ \text{subject to} & Z = \begin{bmatrix} Z_1 & Z_2^T \\ Z_2 & Z_4 \end{bmatrix} \succeq 0, \quad P^c(Z_2) = 0, \end{array} \tag{6}$$

with variable $Z \in \mathbf{S}^{2n}$ where each Z_k has size $n \times n$. The constraint $P^c(Z_2) = 0$ explains that the corresponding entries of Z_2 to the zero entries in A are free, otherwise they are all zeros. If $P(A) = 0$ reduces to $\mathbf{diag}(A) = 0$ in the primal convex SEM, then $P^c(Z_2) = 0$ in the dual is simplified to that Z_2 is diagonal.

An important assumption of the problem (4) is that S must be positive definite. Otherwise, the problem could be unbounded below.

3.1. KKT conditions

The KKT conditions, derived as the optimality condition for the optimal solution to (4), are:

Zero gradient of the Lagragian

$$X = (S - Z_1)^{-1}. \tag{7}$$

Complementary Slackness

$$0 = \mathbf{tr}\left(Z \begin{bmatrix} X & (I-A)^T \\ I-A & \Psi \end{bmatrix}\right), \quad 0 = \mathbf{tr}(Z_4(\Psi - \alpha I)). \tag{8}$$

Primal Feasibility

$$(I-A)^T\Psi^{-1}(I-A) \preceq X, \quad 0 \prec \Psi \preceq \alpha I, \quad P(A) = 0.$$

Dual feasibility

$$Z \succeq 0, \quad P^c(Z_2) = 0.$$

We will use these conditions to analyze the solution properties later in the next section.

3.2. Trivial solution

The following proposition states an important result of our paper:

Proposition 1: Let $\alpha_c = n/\,\mathbf{tr}(S^{-1})$ (the harmonic mean of the eigenvalues of $S \succ 0$). If $\alpha \leq \alpha_c$ then

$$S^{-1} \succeq (I-A)^T\Psi^{-1}(I-A) \tag{9}$$

is *infeasible* in Ψ and A, or that $Z = 0$ cannot be an optimal solution for the dual problem (6).

This explains that there is a critical value α_c such that if the optimal dual $Z^\star = 0$ then $\alpha \geq \alpha_c$, *i.e.*, if the trivial solution in the dual occurs then we have used too large value of α. From the KKT conditions, if $Z^\star = 0$ then they are reduced to (9), which means $\Psi^\star$ can be chosen to be sufficiently large (if α is arbitrarily large) and the RHS of the above inequality can be sufficiently small. Then $X^\star = S^{-1}$ is then not equal to $(I - A^\star)^T\Psi^{\star -1}(I - A^\star)$ as opposed to the desired equality.

The proof of the above proposition is obtained by applying a generalization of Farka's lemma to semidefinite programming [12]. We opt to omit the detail due to space limit.

3.3. Low rank solution of the primal convex SEM

The solution of the primal convex SEM is useful if $X = (I-A)^T\Psi^{-1}(I-A)$ at optimum or that X is a *low rank* solution, so that we can use $\Sigma = X^{-1}$. In this section, we aim to find a relation between the parameter α and the low rank optimal solution of (4). To show this, consider complementary slackness condition (8). Using a property of trace: $\mathbf{tr}(AB) = 0 \iff AB = 0$ for $A, B \succeq 0$, provides

$$\underbrace{\begin{bmatrix} Z_1 & Z_2^T \\ Z_2 & Z_4 \end{bmatrix}}_{Z} \underbrace{\begin{bmatrix} X & (I-A)^T \\ I-A & \Psi \end{bmatrix}}_{W} = 0. \tag{10}$$

The result in (10) further shows that the columns of W are in the nullspace of Z. Therefore, we have $\mathbf{rank}(W) = \mathbf{nullity}(Z)$ and that $\mathbf{rank}(Z) = 2n - \mathbf{rank}(W)$. Since $X \succ 0$ is an implicit constraint, this implies that X must be full rank. The rank of W must satisfy $n \leq \mathbf{rank}(W) \leq 2n$ and therefore $0 \leq \mathbf{rank}(Z) \leq n$.

We obtain a low rank solution when the optimal primal of (4) and the optimal dual of (6) satisfy

$$X = (I-A)^T\Psi^{-1}(I-A) \text{ or equivalently } \mathbf{rank}(Z) = n,$$

(because $\mathbf{rank}(W) = n$). Furthermore, when this holds, $\mathbf{rank}(Z_4) = n$ and from (8), it gives $\Psi = \alpha I$, *i.e.*, the estimated covariance error becomes a diagonal matrix. From section 3.2, we have shown that if α is smaller than $\alpha_c = n/\,\mathbf{tr}(S^{-1})$, then the optimal dual solution is not zero. We can show that the minimum eigenvalue of S satisfies $\lambda_{\min}(S) < \alpha_c$ and this suggests that we can consider three ranges of α where the rank of Z varies as shown in Figure 1. The value of α_c lies somewhere in the interval of condition that $0 < \mathbf{rank}(Z) < n$. If

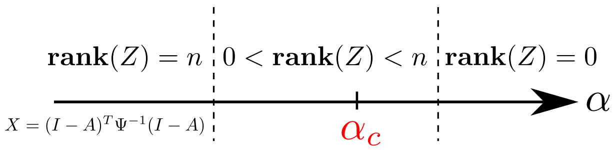

Fig. 1: $\mathbf{rank}(Z)$ as α varied.

$\alpha = \lambda_{\min}(S)$, then it is often the case that $\mathbf{rank}(Z) = n$ which will be shown in the numerical experiment section. This suggests us that we have put a constraint $\Psi \preceq \lambda_{\min}(S)I \preceq S$ into the estimation problem. Our justification is that we control the covariance error to be less than the covariance of the variables. Therefore, throughout our numerical experiments, we select $\alpha = \lambda_{\min}(S)$ so that the problem is likely to return a low rank solution.

4. ALGORITHM

To solve large-scale convex optimization problems, we opt to use the *alternating direction method of multipliers* (ADMM) algorithm which has the superior convergence properties [13]. From Section 3.3, the low rank solution holds when $\alpha = \lambda_{\min}(S)$ and it provides $\Psi = \alpha I$ at optimum, so that, in this case, we can solve (5) by using a constraint $X_4 = \alpha I$ instead of $0 \preceq X_4 \preceq \alpha I$. To arrange the problem (5) in ADMM format, let us define the constraint sets: $\mathcal{C}_1 = \mathbf{S}^n_+$ (positive definite cone),

$$\mathcal{C}_2 = \left\{ V \in \mathbf{S}^{2n} \;\middle|\; V = \begin{bmatrix} V_1 & V_2^T \\ V_2 & \alpha I \end{bmatrix} \;\middle|\; P(V_2) = I \right\},$$

and define $f : \mathbf{S}^{2n} \to \mathbf{R}$ given by $f(X) = -\log\det(X_1) + \mathbf{tr}(SX_1)$, g_1 and g_2 are indicator functions of sets $\mathcal{C}_1, \mathcal{C}_2$, respectively. The problem (5), with a replacement of $0 \preceq X_4 \preceq \alpha I$ by $X_4 = \alpha I$, can then be rearranged into ADMM format as

$$\begin{array}{ll} \underset{X,U,V}{\text{minimize}} & f(X) + g_1(U) + g_2(V), \\ \text{subject to} & X - U = 0, \quad X - V = 0, \end{array}$$

with variables $X, U, V \in \mathbf{S}^{2n}$. The ADMM algorithm starts with forming the augmented Lagrangian defined by

$$\begin{aligned} L_\rho(X, U, V, Y_1, Y_2) &= \log\det(X_1) + \mathbf{tr}(SX_1) \\ &+ \mathbf{tr}(Y_1^T(X-U)) + \mathbf{tr}(Y_2^T(X-V)) \\ &+ \frac{\rho}{2}\|X-U\|_F^2 + \frac{\rho}{2}\|X-V\|_F^2, \end{aligned}$$

where $\rho > 0$ is called the *penalty parameter* in which its value relates to the speed of convergence and enforcing the equality constraints. Let us denote X and X^+ the variables in current and next iteration. The update rule of ADMM is to minimize L_ρ with respect to X, U, V independently and can be described as follows.

X-update. Minimizing L_ρ w.r.t. X is the problem

$$X^+ = \underset{X}{\text{argmin}}\ \ f(X) + \rho\|X - M\|_F^2,$$

where $M = \frac{1}{2}(U+V) - \frac{1}{2\rho}(Y_1 + Y_2)$. The zero gradient condition is

$$\begin{bmatrix} -X_1^{-1} + S & 0 \\ 0 & 0 \end{bmatrix} + 2\rho(X - M) = 0,$$

with an implicit constraint from the domain of f that $X_1 \succ 0$. We can apply the method based on eigenvalue decomposition from [13, §6.5] to show that the zero gradient condition on the $(1,1)$ block: $2\rho X_1 - X_1^{-1} = M_1 - S$, can be achieved with a positive definite X_1. Other blocks of X are simply $X_k = M_k$, for $k = 2, 4$.

U-update. Minimizing L_ρ w.r.t. U is the problem

$$U^+ = \underset{U \succ 0}{\text{argmin}}\ \ \|U - M\|_F^2, \quad M = X + \frac{1}{\rho}Y_1.$$

This is a projection problem onto $\mathbf{S}_+^{2n}$, the positive definite cone, and has a closed-form solution [13, §4].

V-update. Minimizing L_ρ w.r.t. V is the problem

$$V^+ = \underset{V \in \mathcal{C}_2}{\text{argmin}}\ \ \|V - M\|_F^2, \quad M = X + \frac{1}{\rho}Y_2.$$

The optimal V is the projection of M onto $\mathcal{C}_2$, given by

$$V = \begin{bmatrix} M_1 & (P^c(M_2) + I)^T \\ P^c(M_2) + I & \alpha I \end{bmatrix}.$$

Y-update. The update of dual variables are

$$Y_1^+ = Y_1 + \rho(X^+ - U^+), \quad Y_2^+ = Y_2 + \rho(X^+ - V^+).$$

Here U, V are auxiliary variables. At the end X, U and V must be equal so that we iterate this algorithm until the residual errors, $X - U$ and $X - V$, are small enough. The main cost of ADMM in solving (5) is from the eigenvalue decomposition in the X-update and U-update only.

5. NUMERICAL EXPERIMENTS

All numerical experiments and corresponding results are demonstrated in this section, such as simulated examples under the condition that low rank solutions are obtained, the effect of our problem parameter on such condition, and the performance of the algorithm.

5.1. Low rank solution of the primal convex SEM

To find the condition that leads to a low rank solution, our simulation starts with generating a sample covariance matrix (S) as a positive definite matrix in which its eigenvalues are in the range of $[1, 20]$. Since A has special structures, *i.e.*, some entries of A are zero including diagonal entries, the structure of A is randomly generated by setting the sparsity of A about 10%. To vary α, we set $\alpha \in [0.5\lambda_{\min}(S), 5\lambda_{\min}(S)]$ with step size 0.02. Using the problem parameters: S, α and sparsity pattern of A, we then solve the primal convex SEM (4) for each α. Solving this problem has been done by CVX package in MATLAB [11].

Figure 2 shows the difference between X (supposed to be the estimated Σ^{-1}) and $(I-A)^T\Psi^{-1}(I-A)$ using 50 runs of S with the same n, *i.e.*, solving the primal convex SEM with one sample of S produces a line in the figure. The norm of error, $\|X - (I-A)^T\Psi^{-1}(I-A)\|$, is zero when we obtain the low rank solution. We notice that the range of α resulting in the low rank solutions does not depend on n. This often occurs when $\alpha \leq \lambda_{\min}(S)$. Therefore, if we solve the primal convex SEM (4) instead of the original problem (3), we can heuristically choose $\alpha = \lambda_{\min}(S)$ to obtain a low rank solution.

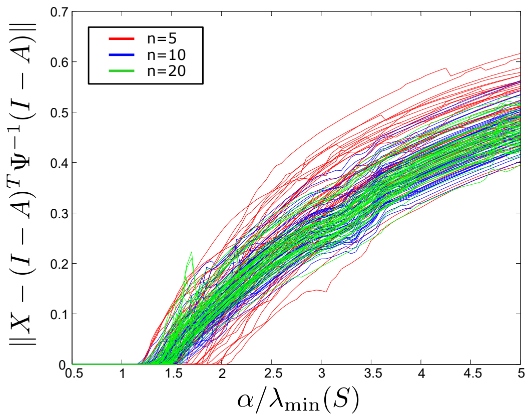

Fig. 2: Lines with the same color correspond to the result from using the same n. Each line in the same color is distinguished by each run of S. The error between X and $(I-A)^T\Psi^{-1}(I-A)$ increases as α increases and is zero when α is sufficiently small relatively to the minimum eigenvalue of S.

5.2. Large value of α

In this section we show the result of Proposition 1. The experiment is setup with $n = 5$ and varying $\alpha \in [0.5\lambda_{\min}(S), 5\lambda_{\min}(S)]$, but in this experiment we generate each S as a positive definite matrix having the same α_c (the harmonic mean of eigenvalues of S) to be 0.5. We then solve the dual of primal convex SEM and plot a relationship between $\mathbf{rank}(Z)$ and α.

From Figure 3, the experiment has been done with 50 samples of S and the result illustrates that for $\alpha \leq \alpha_c$, Z^* cannot be zero. This plot can provide other information, for instance, $Z^* = 0$, when α is large enough, the portion that $\mathbf{rank}(Z) = n$ is approximately 74% and the portion that $\mathbf{rank}(Z) < n$ is approximately 36%, computed from 50 samples of S. Although, we cannot guarantee the relationship between the low rank solution and $\mathbf{rank}(Z)$ with α_c but this result can guide us that if we choose $\alpha < \alpha_c$, we have more chances to get the condition $\mathbf{rank}(Z) = n$ (or more chances to get a low rank solution).

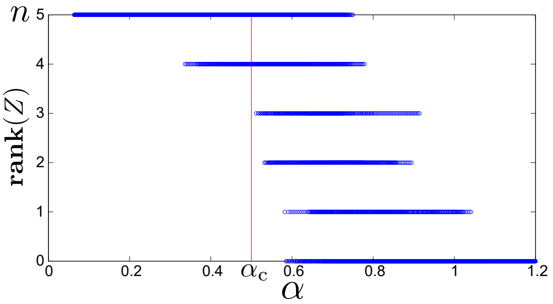

Fig. 3: $\mathbf{rank}(Z)$ as α varies. For each S, the condition $\mathbf{rank}(Z) = 0$ lies on RHS of α_c, meaning that if $Z^* = 0$, $\alpha > \alpha_c$.

5.3. Estimation result

In this section, we verify that if we suppose to know about true path matrix, defined by A_{true}, and variance of noise, defined by σ^2, our estimation formulation can provide that our estimate, $\hat{A}$, is equal to A_{true}. Suppose we fix $S = \sigma^2(I - A_{\text{true}})^{-1}(I - A_{\text{true}})^{-T}$. For our approach, the result of simulation is illustrated in Figure 4. This simulation has been done by setting $n = 5$ as α varies in range $[0.0001, 0.02]$, using step size 0.0001.

In this plot, Figure 4 (top) shows the value of $\|A_{\text{true}} - \hat{A}\|$ as α varies. We observe that $\hat{A}$ reaches to A_{true} when α reaches to σ^2, meaning that our approach can provide $\hat{A}$ which is equal to A_{true} if we choose $\alpha = \sigma^2$. Figure 4 (middle and bottom) shows the result of perfect fitting, $X = S^{-1}$ and the value of objective of (4) is zero ($p^* = 0$). This result illustrates that we can get the perfect fitting when α reaches to σ^2. But in the real application, we do not have information about σ^2 from noise, therefore we can still choose $\alpha = \lambda_{\min}(S)$ that guarantees to obtain a low rank solution. From this choice of α, the value of estimated A is not significantly different from A_{true} and X reaches to S^{-1} but it is not exactly equal.

5.4. Algorithm performance

To see the performance of ADMM algorithm in solving primal convex SEM, we generate data with $n = 50, 100, \ldots, 1000$, using 50 samples of S for each n, then solve (5) with a modified constraint: $X_4 = \alpha I$, and plot the averaged CPU time against with n. The computer's specification used in this experiment is: CPU : Intel Core I5-6400 (2.7 GHz), RAM : 16GB DDR4 BUS2133, HDD : SATA III 7200 RPM (1TBs), OS : WINDOWS10-64bit Education. Solving a primal convex SEM with dimension n involves total number of variables of $n(n+1)/2$ (number of variables in X) plus number of paths in A. A trial problem with $n = 1000$ and a given pattern in A,

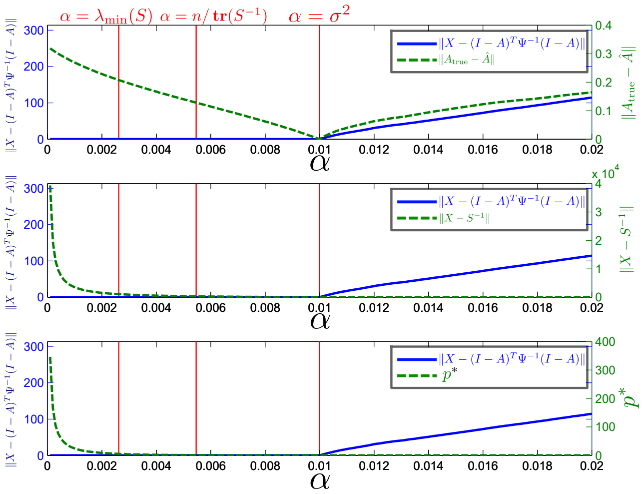

Fig. 4: Simulation result with $n = 5$ as α varies. When $\alpha = \sigma^2$, we can get a low rank solution and perfect fitting.

resulting in totally $1,000,000$ variables, requires approximately about 11 minutes. A large-scale setting like this may not be feasible when implemented with an iterative method based on the use of Hessian matrix.

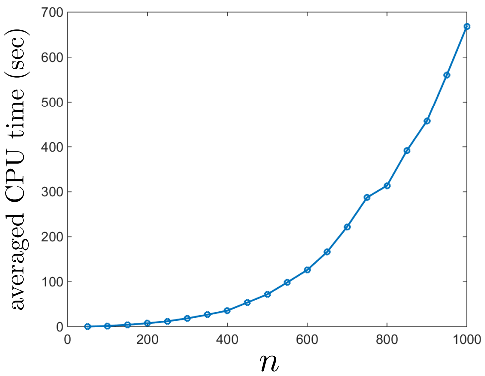

Fig. 5: Averaged cpu time used to solve primal convex SEM. With $n = 1000$, it takes around 11 minutes.

5.5. Implementation on real fMRI data

To show a promising application of our framework, in this experiment, we provide a preliminary result of estimating path coefficients when two causal structures of brain signals are assumed. We use the data from StarPlus fMRI database [14] and average the data over spatial domain, resulting in measurements from 23 regions of interest (ROIs), or $n = 23$. We estimate two path matrices according the two assumed causal structures illustrated in Figure 6. We comment that we provide an applicability of our work to learn a contemporaneous causal relation among brain regions where the true network is not known and we are not trying to validate the results quantitatively. The results can be further used in exploratory SEM where a trade off between the estimated maximum likelihood and the zero structure in the path matrix of the best structure is chosen by some model selection criterion based on chi-squared test or goodness of fit index (GFI).

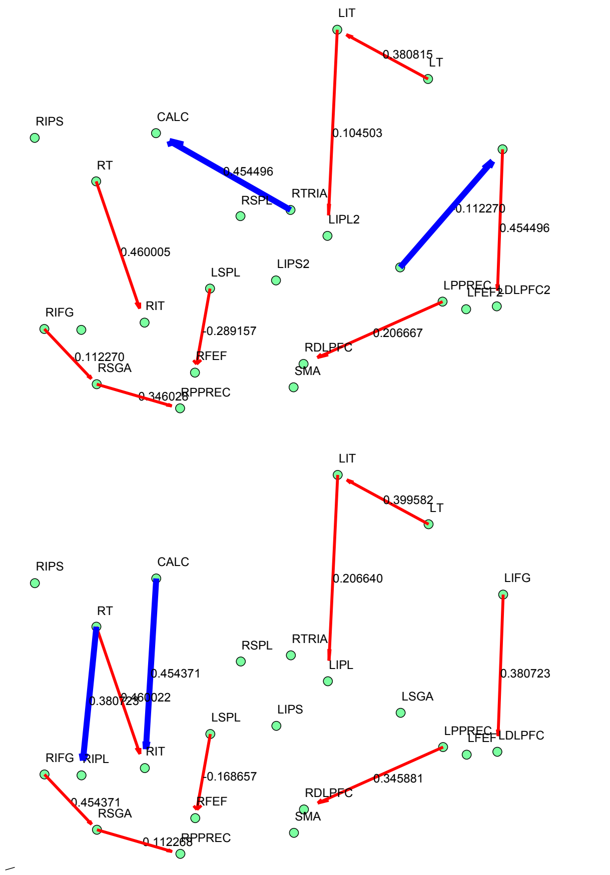

Fig. 6: The causal structures from our hypothesis. Capital letters represent the name of each ROI. The numbers represent path strength between ROIs.

6. CONCLUSION

In this work, we present an alternative estimation formulation for a special problem of path analysis as a convex framework in which it can be solved by efficient numerical methods. In our analysis, we suggest to choose $\alpha = \lambda_{\min}(S)$ to obtain a low rank solution which is useful since we can let X be the estimate of Σ^{-1}. Lastly, we solve primal convex SEM with a modified constraint: $X_4 = \alpha I$ by using an ADMM algorithm where the main computational cost of our estimation formulation directly depends on the cost of eigenvalue decomposition in which it can be performed efficiently. Therefore, our problem can be solved in a large-scale setting.

Despite a difference in our estimation formulation and the original one, we believe that our proposed formulations serve two folds. Firstly, unlike previous SEM applications that only a few variables are of interest, many applications tend to consider a much larger number of variables such as fMRI studies where the variables are neuronal activities and its number is up to thousand. Existing approaches of learning causal structures in the exploratory SEM may experience a computational difficulty in terms of memory storage or convergence. Secondly, our solution for confirmatory SEM is obtained under an assumption of homoscedasticity of residual errors, so if this assumption holds, ours and the original solution coincide. Even if it does not hold, so our solution is not optimal for the original problem but it can be served as a starting value for the iterative algorithm used in the original one in case that the convergence is not obtained.

REFERENCES

[1] R. MacCallum and J. Austin, "Applications of structural equation modeling in psychological research," *Annual review of psychology*, vol. 51, no. 1, pp. 201–226, 2000.

[2] N. Suki, "Passenger satisfaction with airline service quality in Malaysia: a structural equation modeling approach," *Research in transportation business & management*, vol. 10, pp. 26–32, 2014.

[3] U. Ramanathan and L. Muyldermans, "Identifying the underlying structure of demand during promotions: a structural equation modeling approach," *Expert systems with applications*, vol. 38, pp. 5544–5552, 2011.

[4] K. Bollen, *Structural equations with latent variables*. John Wiley & Sons, 1989.

[5] A. McIntosh and F. Gonzalez-Lima, "Structural equation modeling and its application to network analysis in functional brain imaging," *Human Brain Mapping*, vol. 2, no. 1-2, pp. 2–22, 1994.

[6] E. Bullmore, B. Horwitz, G. Honey, M. Brammer, S. Williams, and T. Sharma, "How good is good enough in path analysis of fmri data?" *NeuroImage*, vol. 11, no. 4, pp. 289–301, 2000.

[7] G. Chen, D. Glen, Z. Saad, J. Hamilton, M. Thomason, I. Gotlib, and R. Cox, "Vector autoregression, structural equation modeling, and their synthesis in neuroimaging data analysis," *Computers in biology and medicine*, vol. 41, pp. 1142–1155, 2011.

[8] T. Raykov and G. Marcoulides, *A First Course in Structural Equation Modeling*, 2nd ed. Lawrence Eribaum Associates, Inc., 2006.

[9] K. Jöreskog, D. Sörbom, S. D. Toit, and M. D. Toit, *LISREL 8: New statistical features*. Scientific Software International, 2000.

[10] S. Mulaik, *Linear causal modeling with structural equations*. CRC Press, 2009.

[11] M. Grant and S. Boyd, *CVX: Matlab software for disciplined convex programming (web page and software)*, `http://stanford.edu/~boyd/cvx`, 2007.

[12] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge University Press, 2004, Available: `www.stanford.edu/~boyd/cvxbook`.

[13] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," *Foundations and Trends in Machine Learning*, vol. 3, no. 1, pp. 1–122, 2010.

[14] M. Just, "StarPlus fMRI data," http://www.cs.cmu.edu/afs/cs.cmu.edu/project/theo-81/www/.