A Dimension Reduction Approach to Multivariate Mediation Analysis

Xinpei Shen

Department of Statistics

University of Michigan

Supervisor

Professor Kerby Shedden

An honors thesis submitted in partial fulfillment of the requirements for the degree of Bachelor of Science (Honors Data Science)

April 2024

Acknowledgement

Too many thanks in my heart to express in plain words.

To people who love me, support me, and help me.

Abstract

Mediation analysis is an important tool for understanding possible mechanistic relationships among three variables representing an exposure (X), a mediator (M), and an outcome (Y). Here we propose a method for conducting multivariate mediation analysis via dimension reduction. Specifically, we aim to project X, M, and Y to subspaces of minimal dimension that capture all correlations that could reflect mediation. In particular, the reduced M must simultaneously correlate with the reduced X and the reduced Y. We identify these mediation subspaces by minimizing a loss function based on determinants of covariance matrices. To aid in the interpretation, we rotate the reduced spaces to approximate "parallel mediation", in which the multivariate mediation approximately splits into separate univariate mediations on projected one-dimensional subspaces. We also develop a type of biplot to aid in the visualization of the findings. We present simulation studies showing that our method effectively identifies mediation structures, and we illustrate the approach by analyzing data from a mouse model of kidney disease, considering gene expression as a mediator between exposures and disease outcomes.

1 Introduction and motivation

Mediation analysis is an approach to understanding possible mechanistic relationships among three variables termed the *exposure*, the *mediator*, and the *outcome*, here denoted X, M, and Y. The premise of mediation analysis is that M is intermediate between X and Y, ideally in a causal sense such that X causally impacts M which in turn causally impacts Y, constituting the so-called *indirect effect* of X on Y. In addition, X may also have a direct causal effect on Y, referred to simply as the *direct effect*. In an ideal mediation analysis, no other causal effects exist, e.g. there is no causal effect of M on X.

As a concrete example, considered below in our data analysis, we have data on kidney pathology in a mouse model of human kidney disease. The focus is on the role of aging and diet, along with other possible sources of variation including sex differences. For each mouse, we have data on exposures X, including age, sex, diet, and a genetic manipulation conferring disease susceptibility. We also have data on mediators M, reflecting gene expression in the kidney, and on outcomes Y, consisting of various direct measures of kidney pathology. It is natural to consider that gene expression in kidney tissues mediates the relationship between exposures (aging, diet) and outcomes (pathology in the kidney tissue). We will explore this in detail in Section 6 employing the method we develop in Section 3.

Many methods for mediation analysis with univariate exposure, mediator, and outcome variables have been developed, including methods based on fitting separate regression models for Y regressed on M and X, and for M regressed on X [9], as well as methods based on path analysis and graphical modeling [12]. Here we consider the *multivariate mediation analysis* setting where exposures, mediators, and outcomes are vector-valued. Only a few works considering this setting have been published, e.g. [3]. In contrast to earlier work, we take a dimension reduction approach aiming to identify low-dimensional projections that capture the mediation structures. At the population level we have random vectors $X \in \mathbb{R}^p$, $M \in \mathbb{R}^q$, and $Y \in \mathbb{R}^r$,

and we seek matrices $A \in \mathbb{R}^{p\times d}$, $B \in \mathbb{R}^{q\times d}$, $C \in \mathbb{R}^{r\times d}$ such that the reduced d-dimensional variates $A'X$, $B'M$, and $C'Y \in \mathbb{R}^d$ capture the correlations among X, M, and Y. A special role is played by the reduced variate $B'M$, which must simultaneously capture associations between X and M and between M and Y.

An outline of the thesis is as follows. In Section 2 we consider existing formulations of mediation analysis for univariate data and review correlation analysis for multivariate data. In Section 3 we motivate an approach for multivariate mediation analysis using dimension reduction and consider how to express the mediation structure in an identified and interpretable way. In Section 4 we consider computational issues. Section 5 contains simulation studies assessing the ability of our approach to recover the mediation structure for various sample sizes and dimensions. In Section 6 we analyze the mouse data discussed above and also discuss a biplot approach to aid in interpreting the results. We conclude with a brief discussion in Section 7.

2 Overview of univariate mediation and correlation analysis methods

In this section, we briefly summarize some existing methods for mediation analysis. In addition, since our focus is the multivariate setting, we also review a few classic methods for correlation analysis of multivariate data. Our proposed approach to multivariate mediation analysis is inspired by the specific methods summarized below.

2.1 Univariate mediation analysis

Univariate mediation analysis examines the role of one mediator variable M in the relationship between one independent variable X and one outcome variable Y. The variable M acts as a mediator if the impact of X on Y is partially or fully transmitted through M. Examples in applied research include pain catastrophizing mediating the relationship between clinical exposure and disability status [14] and intention mediating the connection between attitudes and behavior [5].

The product coefficient estimator is a widely used method to assess the presence of mediation [2]. In the context of regression-based mediation analysis, two separate regression models are fit, as follows.

1. *Influence of the exposure (X) on the mediator (M)*: the regression model $E[M|X] = \alpha_0+\alpha_1 X$ captures the relationship between the exposure X and the mediator M.

2. *Influence of the exposure (X) and the mediator (M) on the outcome (Y)*: the regression model $E[Y|X, M] = \gamma_0 + \gamma_1 X + \gamma_2 M$ captures the relationship between the mediator M and the outcome Y while holding the exposure X fixed.

The indirect effect is captured by the product $\alpha_1 \cdot \gamma_2$, which is estimated by the plug-in estimator $\hat{\alpha}_1 \cdot \hat{\gamma}_2$. In one popular paradigm for assessing mediation effects, we first do three significance tests to check whether (1)

X is a statistically significant predictor of M, (2) X is a statistically significant predictor of Y, and (3) M is a statistically significant predictor of Y when controlling for X. Another paradigm for mediation analysis involves directly testing the null hypothesis $\alpha_1 \cdot \gamma_2 = 0$, known as a *Sobel test* [11]. This is an irregular test due to the fact that the null hypothesis is a composite of the simple null in which only one of α_1 and γ_2 are zero, and the compound null in which both of these parameters are zero. Bootstrapping approaches are generally used to address this challenge.

While this framework has proven to be effective in mediation analysis, it also has its limitations. First, it is based on the assumption of linear regression, which might not be suitable for real problems. This further leads to another problem regarding the sample size. To capture the relationship between two sets of variables, traditional regression models usually require the sample size to be greater than the number of parameters for good performance, and this requirement makes the framework inaccurate in high-dimensional settings like gene studies or blood measurements where the number of parameters might exceed the sample size by a lot. Additionally, the framework is designed with a single mediator, but in real problems, multiple mediators might engage in the mediation structure together.

2.2 Dimension reduction for correlation analysis

Two widely used methods in correlation analysis of vector-valued data are principal component analysis (PCA) [6] and canonical correlation analysis (CCA) [8]. We discuss both below and note that CCA has a strong relationship with the approach for multivariate mediation analysis that we propose below, in that our multivariate mediation analysis can be seen as comprising two coupled CCA's.

Principal component analysis (PCA) converts the raw data to a new coordinate system in which the coordinate axes (principal components) correspond to the directions in the original data with the greatest variance. By using PCA as a pre-processing step, we guarantee that any findings downstream of it reflect high-variance characteristics of the source population.

Canonical correlation analysis (CCA) studies the correlation between two collections of random vectors. Specifically, it aims to find the linear combinations of each collection that are maximally correlated. Suppose we have two sets of variables $A = (A_1, A_2, \ldots, A_m)$ and $B = (B_1, B_2, \ldots, B_r)$. The sample version of CCA aims to find linear combinations $U = a_1A_1 + \cdots + a_mA_m$ and $V = b_1B_1 + + \cdots + b_rB_r$ such that the (sample) correlation coefficient between U and V is maximized, where the maximization takes place over the coefficient vectors $(a_1, \ldots, a_m)$ and $(b_1, \ldots, b_r)$. We call U and V a pair of canonical variates.

Since we frame CCA as the goal of maximizing two pairwise correlations simultaneously, we devise an approach that essentially has the form of two coupled CCA tasks. In addition, to accommodate high dimensionality where present, we will use PCA to pre-process the data, so that we don't end up finding highly correlated variates that reflect trivial levels of variation in the data.

3 Multivariate mediation analysis via dimension reduction

Multivariate mediation at the population level is defined by a joint distribution of random vectors (X, M, Y). At the population level, we will be interested in linearly reducing these vectors to $A'X$, $B'M$, and $C'Y$ in a way that preserves the correlations between X and M and simultaneously preserves the correlations between M and Y.

3.1 Identification of mediation subspaces

The correlation between two random vectors can be measured in terms of determinants of cross-covariance and covariance matrices. Given two random vectors U and V, both of dimension p, the correlation between U and V can be quantified as

$$\frac{|\mathrm{Cov}(U, V)|}{|\mathrm{Cov}(U)|^{1/2} \cdot |\mathrm{Cov}(V)|^{1/2}} \tag{1}$$

where $|M|$ represents the determinant of the square matrix M. Note that the value of expression 1 is invariant to the non-singular linear transformation of U and V.

Given two random vectors X and M, and corresponding reduced variates $A'X$ and $B'M$ of the same dimension d, the correlation between the reduced variates is captured by the expression

$$\frac{|A'\Sigma_{XM}B|}{|A'\Sigma_X A|^{1/2} \cdot |B'\Sigma_M B|^{1/2}}, \tag{2}$$

where $\Sigma_{XM} = \mathrm{cov}(X, M) \in \mathbb{R}^{p\times q}$, $\Sigma_X = \mathrm{cov}(X) \in \mathbb{R}^{p\times p}$, and $\Sigma_M = \mathrm{cov}(M) \in \mathbb{R}^{q\times q}$. We will use expression 2 to quantify the aggregate correlation between the reduced variates $A'X$ and $B'M$.

In practice, we observe data matrices $\mathbf{X} \in \mathbb{R}^{n\times p}$, $\mathbf{M} \in \mathbb{R}^{n\times q}$, and $\mathbf{Y} \in \mathbb{R}^{n\times r}$ such that the i^{th} rows of $\mathbf{X}$, $\mathbf{M}$, and $\mathbf{Y}$ constitute a sample from the joint distribution (X, M, Y) described above. Moreover, these observations are independent of the row indices $i = 1, \ldots, n$. Covariance matrices Σ_{XM}, Σ_{MY}, Σ_X, Σ_M, and Σ_Y can be estimated by the usual moment-based estimates, and substituted into expressions such as (2).

Our approach to multivariate mediation analysis is as follows. Given the population level random vectors $X \in \mathbb{R}^p$, $M \in \mathbb{R}^q$, and $Y \in \mathbb{R}^r$, we aim to find matrices $A \in \mathbb{R}^{p\times d}$, $B \in \mathbb{R}^{q\times d}$, and $C \in \mathbb{R}^{r\times d}$ such that $A'X$, $B'M$, and $C'Y \in \mathbb{R}^d$ captures the correlations among X, M, and Y. Referring to (2), the mediation structure we aim to capture through dimension reduction can be identified by optimizing

$$\frac{|A'\Sigma_{XM}B|}{|A'\Sigma_X A|^{1/2} \cdot |B'\Sigma_M B|^{1/2}} \times \frac{|B'\Sigma_{MY}C|}{|B'\Sigma_M B|^{1/2} \cdot |C'\Sigma_Y C|^{1/2}} \tag{3}$$

where $\Sigma_{XM} = \mathrm{cov}(X, M) \in \mathbb{R}^{p\times q}$, $\Sigma_{MY} = \mathrm{cov}(M, Y) \in \mathbb{R}^{q\times r}$, $\Sigma_X = \mathrm{cov}(X) \in \mathbb{R}^{p\times p}$, $\Sigma_M = \mathrm{cov}(M) \in \mathbb{R}^{q\times q}$, $\Sigma_Y = \mathrm{cov}(Y) \in \mathbb{R}^{r\times r}$, and $|M|$ represents the determinant of square matrix M. The product in criteria 3 ensures that we identify subspaces capturing correlation between X and M and correlations between M and

Y simultaneously. Notably, both factors of this criteria take exactly the form of the objective function in canonical correlation analysis.

We note that (3) only depends on A, B, and C through the column spaces span(A), span(B), and span(C). Specifically, if we replace A, B, and C by AR_A, BR_B, and CR_C, respectively, for invertible matrices R_A, R_B, and $R_C \in \mathbb{R}^{d\times d}$, the value of criteria 3 is unchanged. The following is a simple proof of the invariance of its first factor under non-degenerate linear transformation.

$$\begin{aligned}
\frac{|(AR_A)'\Sigma_{XM}BR_B|}{|(AR_A)'\Sigma_X AR_A|^{1/2}\cdot|(BR_B)'\Sigma_M BR_B|^{1/2}} &= \frac{|R_A|\cdot|A'\Sigma_{XM}B|\cdot|R_B|}{|R_A|^{1/2}\cdot|A'\Sigma_X A|^{1/2}\cdot|R_A|^{1/2}\cdot|R_B|^{1/2}\cdot|B'\Sigma_M B|\cdot|R_B|^{1/2}} \\
&= \frac{|R_A|\cdot|R_B|}{|R_A|^{1/2}\cdot|R_A|^{1/2}\cdot|R_B|^{1/2}\cdot|R_B|^{1/2}}\cdot\frac{|A'\Sigma_{XM}B|}{|A'\Sigma_X A|^{1/2}\cdot|B'\Sigma_M B|^{1/2}} \\
&= 1\cdot\frac{|A'\Sigma_{XM}B|}{|A'\Sigma_X A|^{1/2}\cdot|B'\Sigma_M B|^{1/2}}
\end{aligned}$$

Thus, we can view this factor as a function on the product Stiefel manifold $S_X \otimes S_M$, where S_X and S_M respectively are the Stiefel manifolds of d dimensional subspaces of $\mathbb{R}^p$ and $\mathbb{R}^q$ respectively. The other term in criteria 3 involving B and C works in the same way, and we will discuss later in Section 4 how we take advantage of this invariance to simplify computation and interpretation.

3.2 Rotation for interpretability

Since the mediation structure is only identified as subspaces, we are free to choose any bases for these subspaces that aid in interpretation. This is related to the classical topic of factor rotation. For example, Thurstone's varimax rotation in factor analysis has been recently reconsidered [13, 10]. However, while the varimax rotations are orthogonal, we allow for any non-singular change of basis. Specifically, given any matrices A, B, and C that identify the multivariate mediation structure, we can replace these matrices with AR_A, BR_B, and CR_C, for non-singular matrices R_A, R_B, and R_C. In this section, we consider how this can be done to simplify interpretation.

We posit that the following conditions will lead to a simple interpretation of the mediation structure:

- *Uncorrelated factors*: we seek rotations such that $A'\Sigma_X A$, $B'\Sigma_M B$, and $C'\Sigma_Y C$ are all approximately identity matrices. This implies that within each group of variables (exposures, mediators, outcomes), the factors are mutually uncorrelated.

- *Parallel mediation*: we seek rotations such that $A'\Sigma_{XM}B$ and $B'\Sigma_{MY}C$ are approximately diagonal. This gives rise to a parallel mediation structure in which mediation primarily exists along the rotated axes, with minimal "cross-loading" between axes.

Taking $d = 2$ as an example, Figure 1 depicts the parallel mediation structure. Solid lines correspond to strong covariances and dashed lines correspond to weak (ideally zero) covariances.

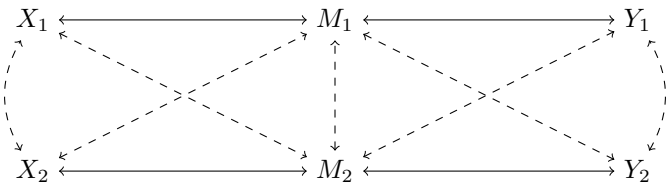


Figure 1: Parallel mediation structure for two-factor mediation. Solid lines correspond to strong covariances among exposure and mediator factors.

We note that if we were conducting a canonical correlation analysis between X and M, it would be possible to simultaneously diagonalize the cross-covariance Σ_{XM} and both of the covariances Σ_X and Σ_M. However, we are working with two coupled CCA tasks, using a shared dimension reduction $B'M$ that must capture correlations between X and M and between M and Y simultaneously. For this reason, it is not possible to simultaneously diagonalize the two cross-covariances $A'\Sigma_{XM}B$, $B'\Sigma_{MY}C$, and all three covariances Σ_X, Σ_M, and Σ_Y. In Section 4.2 we discuss an approach for achieving approximate diagonalization.

We also note that in canonical correlation analysis, diagonalized covariances occur together with orthogonal factor loadings. For example, we diagonalize $A'\Sigma_X A$, with a matrix A that satisfies $A'A = I$, and similarly for B and C. Diagonalization of the covariance $A'\Sigma_X A$ guarantees that the scores are uncorrelated, while achieving $A'A = I$ is arguably a less important property for statistical interpretation. In the multivariate mediation setting, where we are compromising between two CCA's, diagonalization with orthogonal factor loadings is not possible in general. We, therefore, prioritize diagonalizing the covariance matrices, e.g. $A'\Sigma_X A$, thereby producing scores that are nearly uncorrelated, while allowing the loadings to be non-orthogonal.

As a final note on how we formulate the multivariate mediation analysis, a conventional univariate mediation analysis considers the association between X and M, and the association between M and Y given X. Here we are not partialing out the effect of X in the latter regression, considering instead the marginal covariance $\Sigma_{MX} = \mathrm{Cov}(M, X)$. To partial out the effect of X, we could first residualize Y and M with respect to X before estimating their cross-covariance, yielding an estimate $\Sigma_{MY|X}$ of the covariance between M and Y given X. For simplicity, we do not use this residualization here.

4 Computation of multivariate mediation estimators

Without loss of generality, we optimize criteria 3 on the log scale, yielding the smooth objective function

$$\begin{aligned} g(A, B, C) = \log|A'\Sigma_{XM}B| - \frac{1}{2}(\log|A'\Sigma_X A| + \log|B'\Sigma_M B|) \\ + \log|B'\Sigma_{MY}C| - \frac{1}{2}(\log|B'\Sigma_M B| + \log|C'\Sigma_Y C|) \end{aligned} \tag{4}$$

which has a relatively simple gradient. Since g is effectively a function of the column spaces of A, B, and C,

without loss of generality, the optimization can be restricted to A, B, and C which are orthogonal matrices:

$$A'A = I,\ B'B = I,\ C'C = I. \tag{5}$$

Imposing this constraint is advantageous during optimization in that it reduces the dimension of the parameter space, but it places the optimization onto a non-Euclidean Stiefel manifold, so conventional gradient-based optimization methods cannot be directly applied. We also note that as discussed above, we ultimately do not seek a mediation structure with orthogonal factor loadings, and instead aim to prioritize parallel meditation and uncorrelated factors. However, at the stage of estimating the mediation subspaces themselves, any basis can be used, and it is convenient for computation to work with orthogonal bases, as discussed further below.

One possible approach for the optimization of g is the use of conjugate gradient methods adapted to the setting of smooth manifolds [4]. However, we take a different approach here employing a projected gradient algorithm that is similar to the family of alternating direction method of multiplier (ADMM) algorithms [1]. This is developed in Section 4.1 below.

The rotation preferences expressed above in Section 3.2 can be expressed as an optimization problem, but not one that is amenable to gradient descent. We therefore propose a rotation approach based on gradient-free optimization in Section 4.2 below.

4.1 Computation of mediation subspace estimators

We wish to take advantage of the fact that g in Equation (4) has a simple gradient when its domain is embedded into the Euclidean vector space comprising all possible values of A, B, and C. It is natural to use this gradient as a search direction, however standard gradient-descent algorithms will not preserve the constraint in Equation (5) which is not a linear submanifold of the parent vector space. Here we present an approach for resolving this difficulty by changing the constraints to penalty terms.

Suppose that at the kth step of the optimization we have values A_k, B_k, and C_k, which are not necessarily orthogonal. We can calculate orthogonal matrices $\tilde{A}_k$, $\tilde{B}_k$, and $\tilde{C}_k$ whose columns spaces are the same as the column spaces of A_k, B_k, and C_k respectively, e.g. using the Q factor of the QR decomposition. In the next iteration, we wish to continue optimizing g in Equation 4, while approximately maintaining the orthogonality of the parameter matrices. We do this by augmenting the objective function with penalties

$$h(A_{k+1}, B_{k+1}, C_{k+1}; \tilde{A}_k, \tilde{B}_k, \tilde{C}_k) \equiv \|A_{k+1} - \tilde{A}_k\|_F^2 + \|B_{k+1} - \tilde{B}_k\|_F^2 + \|C_{k+1} - \tilde{C}_k\|_F^2. \tag{6}$$

Since h is smooth and easy to differentiate, one can use gradient-based optimization methods such as BFGS to minimize the augmented objective function.

$$\text{obj}(A, B, C, \rho) = g(A_{k+1}, B_{k+1}, C_{k+1}) - \rho_{k+1} h(A_{k+1}, B_{k+1}, C_{k+1}; \tilde{A}_k, \tilde{B}_k, \tilde{C}_k). \tag{7}$$

As a type of annealing, we optimize the objective in equation 7 using an ascending sequence of values for ρ_k to gradually introduce the orthogonality penalties. For sufficiently large values of ρ_{k+1}, the orthogonality constraints in (5) will almost exactly hold.

To start the optimization, we set the initial values of A, B, and C to be the principal components obtained from the sample estimates of Σ_X, Σ_M, and Σ_Y. The coefficient of penalty term ρ reaches $\rho = 500$ to make the resulting A, B, and C nearly orthogonal.

4.2 Estimation of optimal rotations

After estimating A, B, and C, we rotate (reparameterize) them to achieve uncorrelated factors and parallel mediation, as discussed in Section 3.2. Perfectly uncorrelated factors result if $A'\Sigma_X A$, $B'\Sigma_M B$, and $C'\Sigma_Y C$ are identity matrices. Perfectly parallel mediation results if $A'\Sigma_{XM}B$ and $B'\Sigma_{MY}C$ are diagonal matrices. However in general it is impossible to achieve both perfectly uncorrelated factors and perfectly parallel mediation at the same time. Notice that we can reparameterize the model by replacing A with AR_A, B with BR_B, and C with CR_C, where R_A, R_B, and R_C are non-singular. Taking advantage of this invariance, we aim to find R_A, R_B, and R_C by taking singular value decomposition on $A'\Sigma_X A$, $B'\Sigma_M B$, and $C'\Sigma_Y C$, and we set them as the initial value when we solve the following optimization problem regarding R_A, R_B, and R_C.

As discussed in Section 3.2, we optimize a composite measure of uncorrelatedness and parallelism. Uncorrelatedness is measured as

$$\delta_u(R_A, R_B, R_C) \equiv \|(AR_A)'\Sigma_X AR_A - I\|_F^2 + \|(BR_B)'\Sigma_M BR_B - I\|_F^2 + \|(CR_C)'\Sigma_Y CR_C - I\|_F^2. \quad (8)$$

Parallelism is more complicated. The special position of B makes it difficult to find analytical solutions of R_A, R_B, and R_C to diagonalize both $(AR_A)'\Sigma_{XM}BR_B$ and $(BR_B)'\Sigma_{XM}CR_C$ simultaneously, since R_B required for different covariance matrices are different. Rather, we convert this to an optimization problem to find R_A, R_B, and R_C where both covariance matrices are "almost" diagonal.

For any matrix M, we define the following measure of off-diagonal mass

$$q(M) = \sum_{i \neq j} M_{ij}^2 - (p-1) \cdot \sum_{i=j} M_{ij}^2 \quad (9)$$

where p is the dimension of matrix M. We do not use the Frobenius norm here because it counts diagonal and off-diagonal elements in the same direction while we want to count them oppositely. We also give diagonal elements $p-1$ weights because there are p diagonal elements and $p \cdot (p-1)$ off-diagonal elements in a square matrix. A small $q(M)$ indicates a high level of diagonalization. Recall that we want to find R_A, R_B, and R_C such that $(AR_A)'\Sigma_{XM}BR_B$, $(BR_B)'\Sigma_{MY}CR_C$ are diagonal, so with help of equation (9), the parallelism is

measured as

$$\delta_p(R_A, R_B, R_C) = q((AR_A)'\Sigma_{XM}BR_B) \times q((BR_B)'\Sigma_{MY}CR_C). \tag{10}$$

We seek off-diagonal scores for both matrices simultaneously, so we use multiplication to express the trade-off between them. Combining two measurements for uncorrelatedness and parallelism, we derive the objective function for the rotation optimization problem as

$$f(R_A, R_B, R_C) = \delta_u(R_A, R_B, R_C) + \delta_p(R_A, R_B, R_C). \tag{11}$$

Given that this function does not have a simple gradient, we optimize it with Powell's method to find a local minimizer.

5 Simulation study

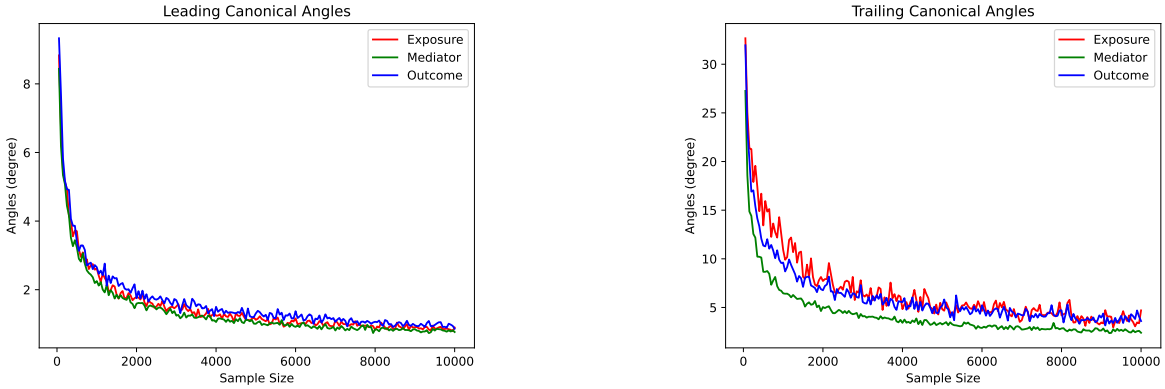
In this section, we conduct simulation studies to assess the effectiveness of our proposed multivariate mediation analysis methods. We construct population mediation structures, simulate datasets, and assess how well our method captures the known mediation structure. We use Jordan's canonical (principal) angles between subspaces to measure whether the estimated and true subspaces are similar. We then assess effectiveness of the rotation to approximate parallel mediation, by quantifying the off-diagonal versus the on-diagonal mass of matrices that we are aiming to approximately diagonalize.

5.1 Population mediation structure

We consider a simple two-dimensional mediation structure, with latent random vectors $u_j \sim N(0, I_{n\times n})$ and observed arrays $X = [u_1\ u_2\ 0_n\ 0_n\ 0_n] + E_X$, $M = [0_n\ u_1\ u_2\ 0_n\ 0_n] + E_M$, and $Y = [u_1\ 0_n\ u_2\ 0_n\ 0_n] + E_Y$, where E_X, E_M, and E_Y represent $n \times 5$ arrays of standard normal noise. In this way, we make two mediation paths:

- The first factor in X correlates with the first factor in Y through the second factor in M.
- The second factor in X correlates with the third factor in Y through the third factor in M.

Based on this construction, we can derive the population mediation subspace bases A, B, and C as:



(a) Primary Angles

(b) Secondary Angles

Figure 2: Canonical angles for various sample sizes.

$$A_{\text{true}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \end{bmatrix}^T$$

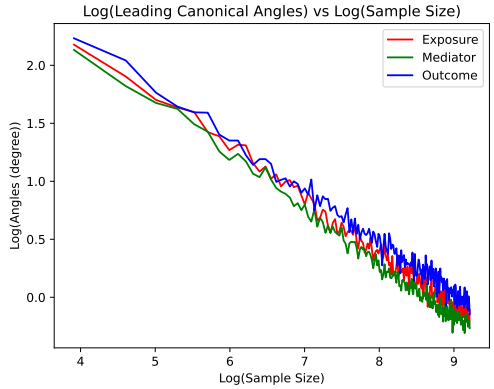
$$B_{\text{true}} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}^T$$

$$C_{\text{true}} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix}^T$$

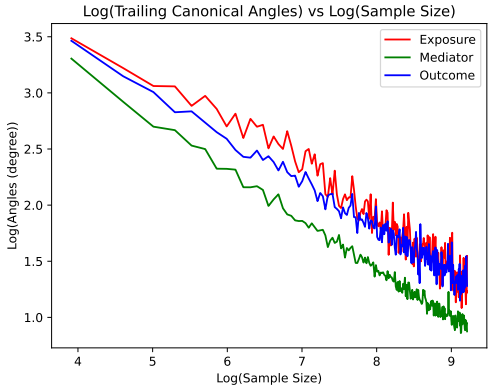
5.2 Subspace identification and measurement of fitness

The estimation accuracy for mediation subspaces is assessed based on canonical angles. For each simulation replication, we compare $A'X$ to $A'_{\text{true}}X$ using canonical angles, which are invariant to the choice of basis in A_{true}. Similarly, we compare $B'M$ to $B'_{\text{true}}M$, and $C'Y$ to $C'_{\text{true}}Y$. Since the dimension in our simulations is $d = 2$, each simulation yields two canonical angles, which we refer to as the leading and trailing angles. The leading (smaller) angle reflects the easier direction within a given target (e.g. $\text{col}(A_{\text{true}})$) to identify, and the trailing (larger) angle reflects the more difficult direction to identify. In our simulation, with sample size reaching $n = 10000$, the mean leading canonical angle was less than 1 degree, and the mean trailing canonical angle was less than 5 degrees, suggesting consistency for large sample sizes. Furthermore, the mean leading canonical angle was less than 5 degrees for much smaller sample sizes ($n < 250$). Figure 2 shows the relationship between sample size and canonical angles.

The simulation study also reveals the scaling of estimation accuracy with sample size. Figure 3 shows the average estimation error (canonical angle) with respect to sample size in the log/log scale. The linear relationship with slopes ranging from -0.46 to -0.4, suggests a power-law relationship with approximate $1/\sqrt{n}$ scaling.



(a) Primary Angles



(b) Secondary Angles

Figure 3: Canonical angles for various sample sizes in log coordinates.

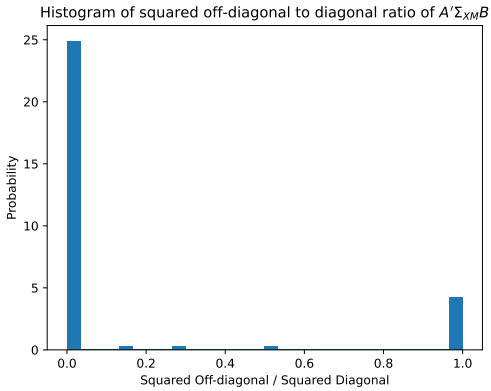
5.3 Rotation effectiveness

Having identified the mediation subspaces accurately, the next step is to assess whether our proposed rotation methods achieve the desired parallel/uncorrelated mediation structure. Here we focus on assessing parallelism.

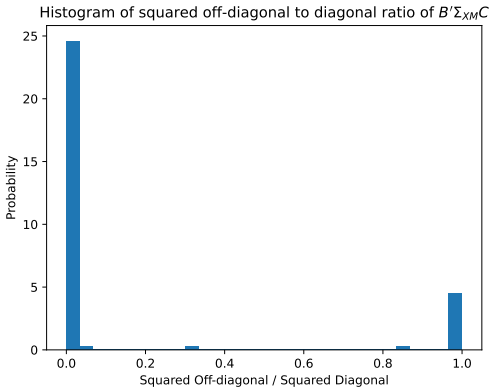
Recalling the goal of parallel mediation and Equation 11, we want to make $A'\Sigma_{XM}B$ and $B'\Sigma_{MY}C$ as diagonal as possible. In other words, we want the mass of elements to concentrate on the diagonal, so we introduce the ratio criteria $z(M)$ for a squared matrix M.

$$z(M) = \frac{\sum_{i \neq j} M_{ij}^2}{\sum_{i=j} M_{ij}^2} \tag{12}$$

Notice that $q(M)$ in Equation 9 could be negative, which means δ_p in Equation 10 could be negative, and the optimization will divert from the desired direction. This feature makes the result of optimization depends on the initial states of R_A, R_B, and R_C such that we would like the initial value Equation 10 to be positive. Nevertheless, randomness is involved in our methods and simulation studies, so it is hard to control the initial states of Equation 10, which might make the rotation results unstable. However, based on our simulation studies, this unstableness is not a big problem in real scenarios. Referring to Figure 4, the results will be very undesired ($z(M) \approx 1$) when the optimization starts with a negative value, making the problem easily realizable. Since our rotation is independent of the sample size, it is efficient to run it again. Still referring to Figure 4, the rotation algorithm performs well in most cases in concentrating the mass to diagonal elements ($z(M) \approx 0$).



(a) $z(A'\Sigma_{XM}B)$



(b) $z(B'\Sigma_{MY}C)$

Figure 4: Distribution of the ratio of squared off-diagonal elements to squared diagonal elements

6 Multivariate mediation analysis of aging and kidney pathology in a mouse experiment

To illustrate the mediation methods developed above, we analyze data from an experimental study of aging and kidney disease in mice. We have data on 90 mice from two experiments, with different ages, sexes, diets, TGFB (transforming growth factor β) gene manipulation, and degrees of kidney pathology. For all mice, we also have gene expression data from the kidney glomeruli obtained at the time of sacrifice. Glomeruli are clusters of blood vessels involved with filtering waste products from the blood. We view gene expression as the mediator connecting exposures (to be discussed below) to outcomes relating to kidney pathology in maturity.

The mice were part of two experiments, with distinct experimental manipulations. In one experiment, diet was manipulated in 60 mice, and in the other experiment, the TGFB gene was manipulated in 30 mice. In the diet experiment, mice were fed either a calorie-restricted diet (CR) or were allowed to eat at will (ad-lib). In the TGFB experiment, a genetic manipulation was used to render mice prone to early kidney tissue scarring, similar to what occurs in humans with chronic kidney disease (CKD). The TGFB manipulation varied in intensity and was coded as "mild" or "severe" accordingly. The mice in the diet experiment did not undergo TGFB manipulation so are coded "none" for the TGFB variable. The mice in the TGFB experiment were all on an ad-lib diet.

The exposure variables in the mediation analysis include the experimental manipulations discussed above, along with variables that reflect the life histories of the animals from birth to the point of sacrifice. In all, there are five exposure variables, which when dummy-coded yield 10 variables in all. The exposure variables are age at sacrifice (26-685 days), sex (female, male), diet (calorie-restricted, ad-lib), TGFB mutation status (none, mild, severe), and experimental cohort (diet, TGFB-mutation). The cohort variable is included to account for batch effects between the two experiments whose data are pooled.

Gene expression is taken as the mediator, taking the perspective that exposures discussed above causally impact the glomerular gene expression, which in turn causally impacts the pathologic state of the kidneys later in the life course. Gene expression is measured with microarrays and thus is high-dimensional. Therefore we took the 10,000 most variable genes and reduced them to several linear factors using principal components analysis (PCA). The selected number of PC's is discussed below.

The outcome variables (Y) are derived from post-mortem assessments of kidney tissue pathology, ascertained by an expert. The outcome variables are all quantitative. They are mesangial volume, glomerular volume, podocyte density, left kidney weight, and mesangial index. Due to heavy right skew, the mesangial and glomerular volumes are log-transformed. Poorer kidney health is reflected in greater values of mesangial volume, glomerular volume, and mesangial index, and lower values of podocyte density. Left kidney weight is strongly influenced by sex, with male mice having larger kidneys.

6.1 Gene expression dimension reduction

Gene expression was evaluated on the log scale, and after selecting the 10,000 most variable genes, we standardized all genes to have a mean equal to 0 and a standard deviation equal to 1. Figure 5 shows a scree plot of the eigenvalues λ_n of the gene expression correlation matrix. Figure 6 plots the eigenvalues in log space, aiming to assess whether the "tail eigenvalues" $\lambda_n, n > n^*$ exhibit a power-law structure.

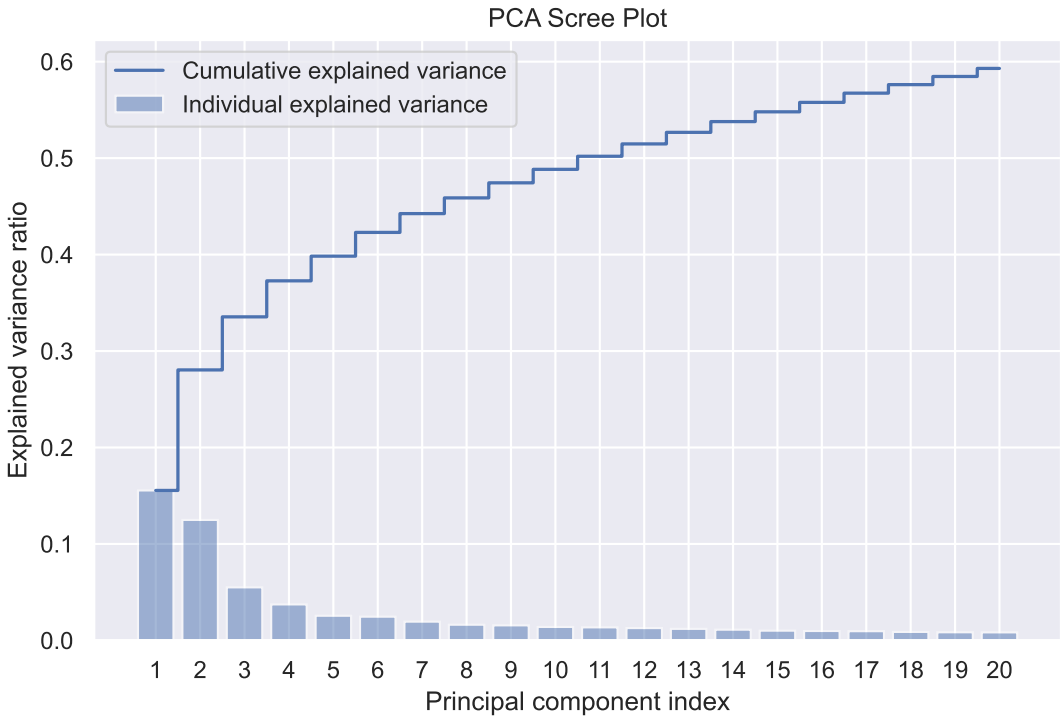


Figure 5: Scree plot of PCA eigenvalues.

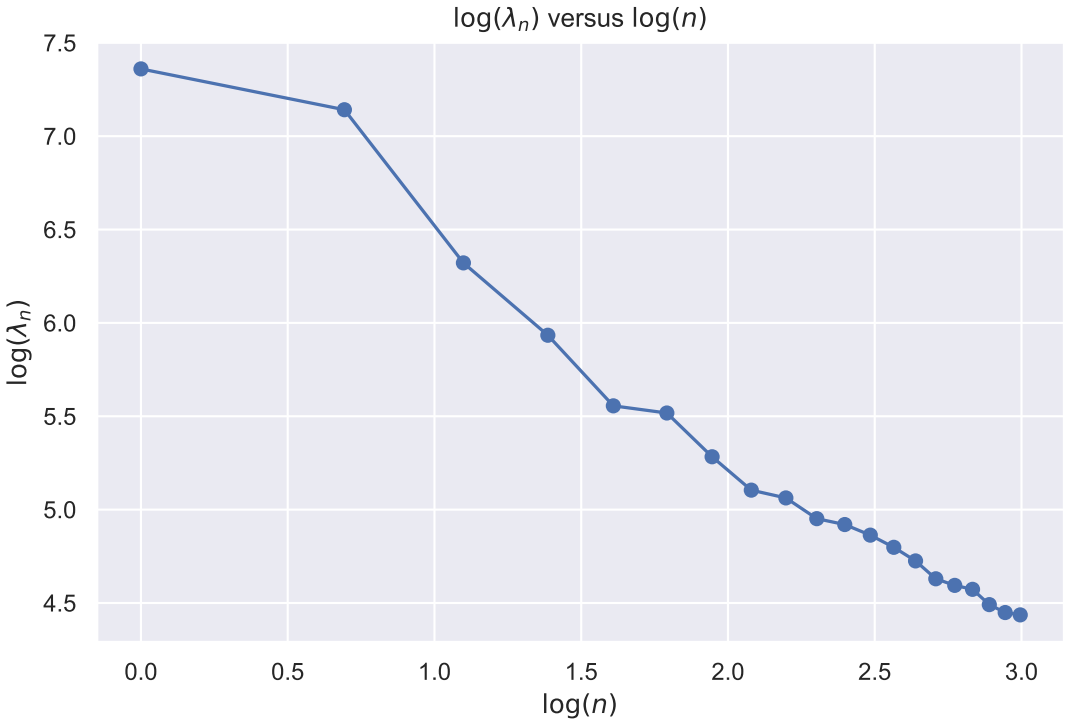


Figure 6: PCA eigenvalues in log space.

To select the number of components to include in the mediation analysis, we note that there is a trade-off between variance explained and parsimony. For example, 12 principal components explain over 50% of the variance in gene expression, but 6 of these components explain less than 2% of the variance individually. On the other hand, with 3 principal components, every component explains at least 5% of the variance, but the overall variance explained is less than 35%. We chose 6 as the number of principal components after balancing the variance explained and parsimony. We note as well that eigenvalues 7 and beyond arguably behave like a power-law based on figure 6. In a power-law relationship, the gaps between successive eigenvalues are modest and it is difficult to differentiate the roles of the corresponding eigenvectors.

The dominant 6 principal components of the gene expression data will be used as mediators (M). After fitting the mediation model we will consider the important issue of interpretation. Noting that while some of the PC's may reflect specific biological pathways, this may not always be the case. Thus we will aim to interpret the mediation structure in terms of individual genes as well as in terms of factors summarizing the behavior of many genes.

The PC loadings reflect the standardized contributions of each gene to each PC factor. In Table 1 we present the 20 leading loadings (10 positive and 10 negative) for each principal component in descending order of magnitudes. Among the genes in this table are genes with strong expression in the kidney (SLC4A1) and genes with strong sex-specific expression (Xist).

PC1	PC2	PC3	PC4	PC5	PC6
Gm15459	**Csad**	**Gm10344**	**Gm37074**	**Gm13301**	**Gm37108**
(-0.024164)	(-0.02503)	(-0.038966)	(-0.043471)	(-0.047811)	(-0.037652)
Ccl28	**Igha**	**Ank1**	**Gm42788**	**Ndufv1**	**Gm37317**
(-0.024146)	(-0.02462)	(-0.037837)	(-0.042809)	(-0.044893)	(-0.035751)
Serpina1d	**Gm10286**	**4930512B01Rik**	**Trpc2**	**Fbln7**	**Gm45512**
(-0.024086)	(-0.024541)	(-0.036938)	(-0.0419)	(-0.041528)	(-0.034702)
Pecr	**Slc4a1**	**Vasp**	**4930527F14Rik**	**Gm16754**	**Tdo2**
(-0.023998)	(-0.023681)	(-0.036906)	(-0.041137)	(-0.041336)	(-0.034597)
Akr1c14	**Dab2**	**Ppef1**	**Siglece**	**Arl4a**	**Gm11110**
(-0.023985)	(-0.023437)	(-0.036601)	(-0.041119)	(-0.040909)	(-0.033159)
Ugt2b38	**1700020N18Rik**	**Rbbp4**	**Nlrp3**	**Pbrm1**	**Flvcr2**
(-0.023947)	(-0.023138)	(-0.03622)	(-0.040954)	(-0.040862)	(-0.032909)
Cbr3	**Klk1b22**	**Tgm3**	**Rnf225**	**Moxd1**	**Sorl1**
(-0.023843)	(-0.023122)	(-0.035968)	(-0.040511)	(-0.040767)	(-0.032908)
Xist	**Slc25a31**	**Cd4**	**Gm29114**	**Ighv1-14**	**A430108G06Rik**
(-0.023793)	(-0.022921)	(-0.03566)	(-0.040143)	(-0.039873)	(-0.032719)
Frzb	**Atp13a4**	**Gm49713**	**1500005C15Rik**	**Fam83d**	**Gm12917**
(-0.023711)	(-0.022901)	(-0.035582)	(-0.040024)	(-0.039694)	(-0.032715)
Hmgb1-ps3	**A730091E23Rik**	**Gm9913**	**Gm26717**	**Ly6k**	**Gm43549**
(-0.023688)	(-0.022893)	(-0.035419)	(-0.039732)	(-0.039015)	(-0.031894)
Ephx2	**Ccl5**	**Psmb1**	**Klrg2**	**Rpl30-ps9**	**Gm12866**
(0.023931)	(0.026693)	(0.024377)	(0.032375)	(0.046591)	(0.045437)
Prcd	**Mat1a**	**Psmd1**	**Gm5767**	**Litaf**	**Rnf24**
(0.023656)	(0.026213)	(0.024013)	(0.031407)	(0.045585)	(0.0447)
Ighv8-8	**Gm12906**	**Mrc2**	**Gm38379**	**Cidec**	**Tbc1d1**
(0.02357)	(0.025778)	(0.023842)	(0.031128)	(0.044468)	(0.04301)
Acox3	**Hibadh**	**Tug1**	**Btnl10**	**Gm43693**	**Alpl**
(0.023388)	(0.025545)	(0.023821)	(0.030093)	(0.043391)	(0.042061)
Gm16284	**Rps2-ps10**	**Gm39526**	**Gm5948**	**Gm28231**	**Syne2**
(0.02338)	(0.025118)	(0.023425)	(0.029722)	(0.042478)	(0.04016)
Gm12989	**Rnase6**	**Etl4**	**4930590J08Rik**	**Bdh1**	**Rpl13-ps6**
(0.023329)	(0.025065)	(0.023387)	(0.029501)	(0.041886)	(0.03729)
Ccn1	**Plpp3**	**Wwc2**	**Gm50055**	**Gm13387**	**A730049H05Rik**
(0.023259)	(0.024965)	(0.022929)	(0.02916)	(0.040286)	(0.037151)
Tfcp2l1	**Gm6397**	**Insyn2b**	**Gm49314**	**Gm4828**	**Gm15513**
(0.023124)	(0.02496)	(0.022579)	(0.028534)	(0.040249)	(0.037075)
Abca3	**Zfp36l2**	**Gm37292**	**Gm7609**	**Gm5939**	**Irf6**
(0.022971)	(0.024898)	(0.022542)	(0.028345)	(0.037328)	(0.036923)
Bcat1	**AA413626**	**Adam10**	**Plpp4**	**Scn1a**	**Gm45924**
(0.022869)	(0.024825)	(0.02253)	(0.027811)	(0.03718)	(0.036155)

Table 1: Genes PC Loadings Table

6.2 Mediation estimates

We fit our multivariate mediation model to the mice data, reducing the data to $d = 2$ dimensions, and then optimally rotated the mediation structure using the methods discussed above. We begin by considering the extent to which we successfully identified uncorrelated factors that exhibit parallel mediation.

The estimated correlation matrices for exposures $\Sigma_{A'X}$, for mediators $\Sigma_{B'M}$, and for outcomes $\Sigma_{C'Y}$ are all $\mathbb{R}^{2\times 2}$ matrices, and are shown in (13).

$$\begin{aligned}
\Sigma_{A'X} &= \begin{bmatrix} 1 & -3.21 \times 10^{-4} \\ -3.2 \times 10^{-4} & 1 \end{bmatrix} \\
\Sigma_{B'M} &= \begin{bmatrix} 1 & 9 \times 10^{-3} \\ 9 \times 10^{-3} & 1 \end{bmatrix} \\
\Sigma_{C'Y} &= \begin{bmatrix} 1 & 6.29 \times 10^{-5} \\ 6.29 \times 10^{-5} & 1 \end{bmatrix}
\end{aligned} \tag{13}$$

The near-diagonal structure of these matrices indicates that the rotation successfully identified two nearly uncorrelated factors for each of the three components of the mediation analysis (exposure, mediator, and outcome).

With $d = 2$, up to two parallel mediation structures can be identified, although there is no guarantee that a rotated structure can be found with negligible cross-loading between factors. These mediation structures are characterized by the cross-correlations between exposures and mediators, and between mediators and outcomes, all calculated in the reduced space. The estimated cross-correlation matrices are shown in (14).

$$\begin{aligned}
\mathrm{Cor}(A'X, B'M) &= \begin{bmatrix} 0.598 & -1.61 \times 10^{-5} \\ 0.112 & 0.651 \end{bmatrix} \\
\mathrm{Cor}(B'M, C'Y) &= \begin{bmatrix} 0.657 & 0.023 \\ 0.101 & 0.522 \end{bmatrix}
\end{aligned} \tag{14}$$

These findings indicate that we were able to find two strong mediation structures. Recall that the factors underlying $A'X$, $B'M$, and $C'Y$ are nearly uncorrelated. The cross-correlations in (14) are fairly weak (none is larger than 0.11), and the mediation correlations themselves are strong, all being greater than 0.5.

To study the mediation role played by the genes, we can characterize each gene mediation factor (a column of MB) in terms of the individual genes that strongly correlate with it. Thus, we selected the top 10 genes that are most negatively, and most positively correlated to the first and second dimension of the reduced mediators and show them in table 2. The genes that are negatively correlated with mediator 2 include 3 genes whose coding sequence is on the Y chromosome – Ddx3y, Eif2s3y, and Uty, suggesting that high scores on mediator 2 occur in female mice and low scores on mediator 2 occur in male mice.

1st Reduced Mediator		2nd Reduced Mediator	
Fgr	**Usp2**	**Cpe**	**Mcoln2**
(-0.700089)	(0.722084)	(-0.833298)	(0.895033)
Prex1	**Gm45819**	**Col19a1**	**Apmap**
(-0.6819)	(0.712365)	(-0.82262)	(0.894624)
Myo1g	**Ak4**	**Nwd2**	**Kyat3**
(-0.671555)	(0.71006)	(-0.82194)	(0.885233)
Foxj1	**Coq10b**	**Tent5a**	**Car5a**
(-0.661593)	(0.657081)	(-0.821193)	(0.870386)
Lpcat2	**Slc9a3r1**	**Ddx3y**	**Myh1**
(-0.660084)	(0.651162)	(-0.814544)	(0.861084)
Cdh11	**Hspe1**	**Atxn7l2**	**Slc17a2**
(-0.648689)	(0.646392)	(-0.808755)	(0.857896)
Dcdc2a	**Lonrf1**	**Rundc3a**	**9030619P08Rik**
(-0.647409)	(0.644853)	(-0.808484)	(0.856785)
Ltbp2	**Dnaja1**	**Eif2s3y**	**Mab21l4**
(-0.641689)	(0.643509)	(-0.807528)	(0.855217)
Myo1f	**Slc25a13**	**Kdm5d**	**Slc22a13**
(-0.639863)	(0.634844)	(-0.806102)	(0.853899)
Trp53i11	**Nus1**	**Uty**	**Acaa2**
(-0.639111)	(0.634354)	(-0.802896)	(0.853263)

Table 2: Correlation between Genes and Reduced Mediators

6.3 Visualization and interpretation

While univariate mediation analysis admits a relatively straightforward and direct interpretation, multivariate mediation analysis arguably does not. To aid in interpreting the fitted mediation structures, we employ biplots [7], which have long been used in interpreting factor analyses and other forms of multivariate data analysis based on linear dimension reduction.

Conventional biplots are not directly applicable since we are simultaneously analyzing three sets of variables. We construct biplots for the mediation setting as follows. First, to make the plot more concise and to reduce overplotting, we omit the redundant variables that are perfectly correlated with other variables. This leaves age at sacrifice, sex (female), diet (ad-lib), TGFB mutation status (mild, severe), and experimental cohort (diet) to be shown in the biplot.

Since we plot these biplots after we rotate the mediation space, employing methods discussed in Section 3.2, and moreover the correlations between factors in (14) are all positive, it follows that the axes in each plot will represent co-occurring characteristics of the mice. For example, if female mice tend to be positioned with positive vertical coordinates in the biplot for exposures (Figure 7), then female mice will also tend to have the characteristics associated with greater vertical positions in the mediator and outcome biplots (Figures 8 and 9).

Starting with the vertical axes, in Figure 7 we see that the sex indicators align with the vertical axis and in particular the female indicator has a positive loading while the (implicit) male indicator has a negative loading. The TGFB mild versus severe loadings also align with the vertical axis, likely due to the fact that

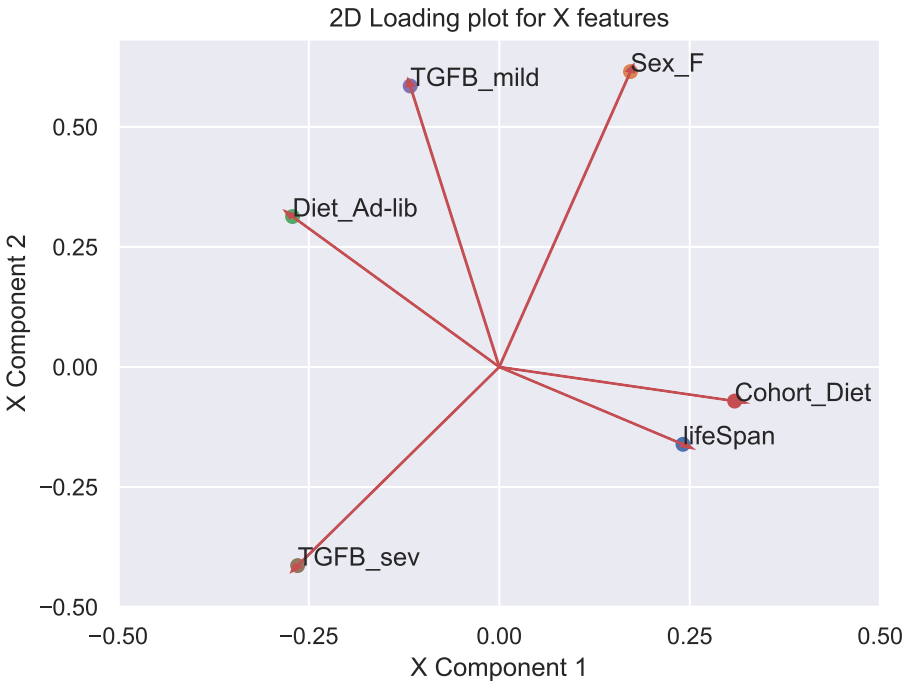


Figure 7: Biplot of reduced exposure variates

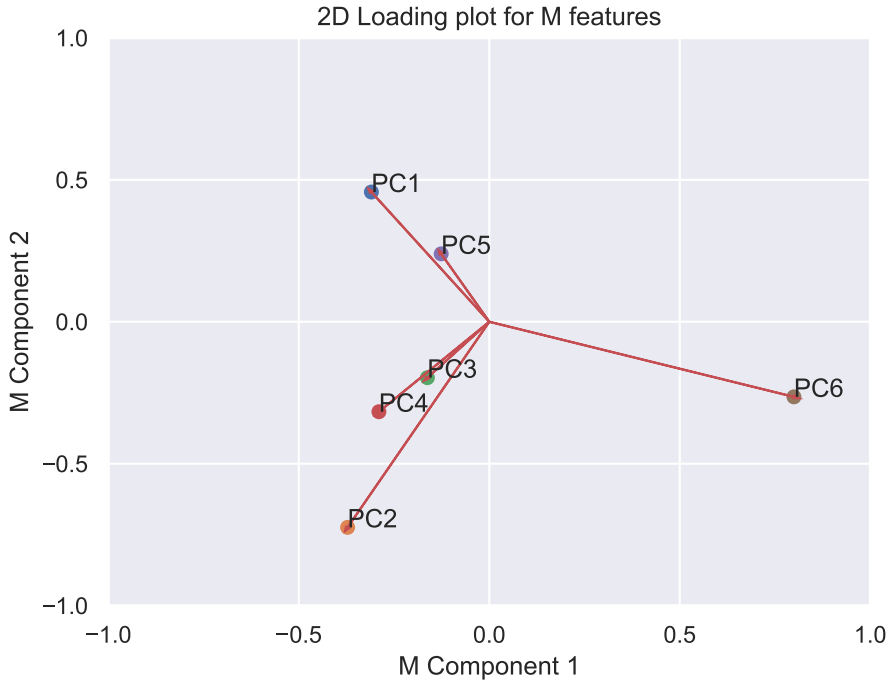


Figure 8: Biplot of reduced mediators (gene PCs)

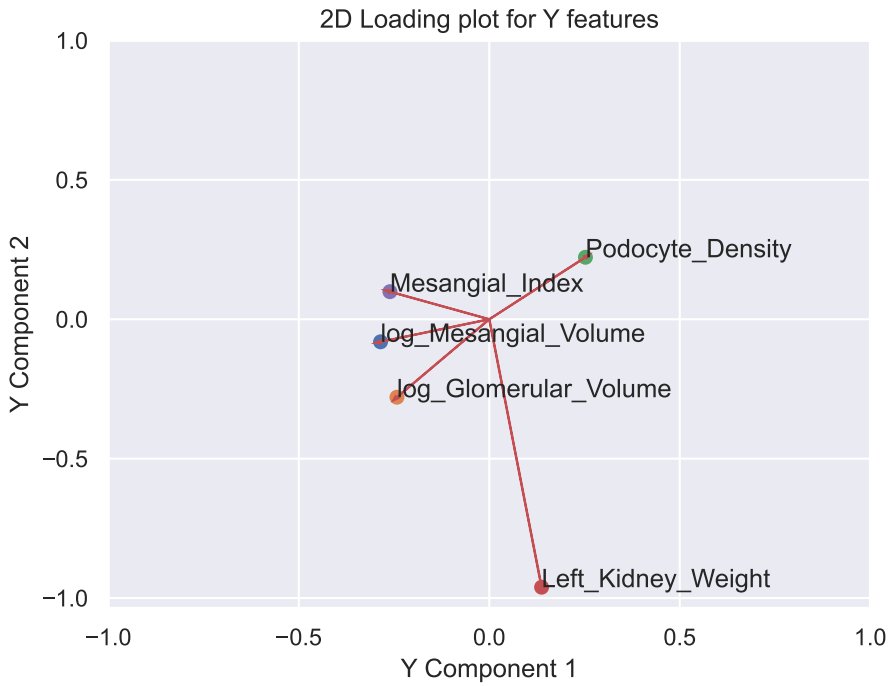


Figure 9: Biplot of outcome variates (kidney pathology)

there were fewer female mice in the severe TGFB group. We will interpret the vertical axis as primarily sex-related, with other factors being incident to imbalances in the design. This corresponds to the vertical axis of Figure 9, where the strongest loading is left kidney weight, which is a very sex-related attribute since male mice tend to have bigger kidneys than female mice.

The horizontal axes of the mediation structure appear to reflect kidney tissue pathology that could be related to aging, and to the diet and TGFB interventions. Calorie restriction is argued to be a mechanism for slowing down the effects of aging, and severe TGFB status results from an intervention that is intended to give mice a phenotype resembling aging in humans. For this reason, it is notable that the ad-lib diet and severe TGFB status loading vectors are similarly aligned with the horizontal axis in the biplot in Figure 7. Furthermore, the same horizontal deviation in the biplot in Figure 9 corresponds to greater mesangial index, mesangial volume, and glomerular volume, but lower podocyte density, all of which reflect poorer kidney health.

7 Discussion

This work provides methods for conducting mediation analysis in a multivariate, possibly high-dimensional setting. The mediation structure is identified by a dimension reduction that is rotated to exhibit approximate parallel mediation structures. We also demonstrate the use of biplots to aid in interpreting the fitted medi-

ation structure. Both the subspace identification and rotation steps are framed as optimization problems, with the former addressed using an ADMM-like algorithm, and the latter addressed using Powell's method for gradient-free optimization.

Simulation studies indicated that the methods developed here are capable of accurately identifying mediation subspaces with modest sample sizes. We then analyzed data from a genomic study of kidney pathology in mice and identified parallel mediation structures associated with the emergence of aging-related kidney pathology, and with sex differences.

This study provides a novel framework for multivariate mediation analysis. It may be especially useful in exploratory analysis of multivariate data through the lens of mediation analysis, to identify hypotheses that may reflect mechanistic mediation relationships.

References

[1] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine Learning*, 3(1):1–122, 2011.

[2] Oliver Y. Chén, Hengyi Cao, Huy Phan, Guy Nagels, Jenna M. Reinen, Jiangtao Gou, Tianchen Qian, Junrui Di, John Prince, Tyrone D. Cannon, and Maarten de Vos. Identifying neural signatures mediating behavioral symptoms and psychosis onset: High-dimensional whole brain functional mediation analysis. *NeuroImage*, 226:117508, 2021.

[3] Oliver Y Chén, Ciprian Crainiceanu, Elizabeth L Ogburn, Brian S Caffo, Tor D Wager, and Martin A Lindquist. High-dimensional multivariate mediation with application to neuroimaging data. *Biostatistics*, 19(2):121–136, 06 2017.

[4] Alan Edelman, T. A. Arias, and Steven T. Smith. The geometry of algorithms with orthogonality constraints, 1998.

[5] Martin Fishbein and Icek Ajzen. Belief, attitude, intention, and behavior: An introduction to theory and research. *Philosophy and Rhetoric*, 10(2):130–132, 1977.

[6] Karl Pearson F.R.S. Liii. on lines and planes of closest fit to systems of points in space. *Philosophical Magazine Series 1*, 2:559–572, 1901.

[7] J.C. Gower and D.J. Hand. *Biplots*. Chapman & Hall/CRC Monographs on Statistics & Applied Probability. Taylor & Francis, 1995.

[8] Harold Hotelling. Relations between two sets of variates. *Biometrika*, 28(3/4):321–377, 1936.

[9] Kosuke Imai, Luke Keele, and Dustin Tingley. A general approach to causal mediation analtsis. *American Psychological Association*, 2010.

[10] Karl Rohe and Muzhe Zeng. Vintage factor analysis with varimax performs statistical inference, 2020.

[11] Michael E. Sobel. Asymptotic confidence intervals for indirect effects in structural equation models. *Sociological Methodology*, 13:290–312, 1982.

[12] Johan Steen and Stijn Vansteelandt. *Mediation Analysis*, page 405–438. CRC Press, November 2018.

[13] Louis Leon Thurstone. Multiple factor analysis. *Psychological review*, 38(5):406, 1931.

[14] R. Whittle, G. Mansell, P. Jellema, and D. van der Windt. Applying causal mediation methods to clinical trial data: What can we learn about why our interventions (don't) work? *European Journal of Pain*, 21(4):614–622, 2017.