Supplements

A. Proof of Theorem 1

Following the definition, we can derive the PL for the PLA-GGM as:

$$\ell_{PL} \left( \{z_i, g_i\}_{i \in [n]} ; R(\cdot), \Omega_0 \right) \propto \sum_{i=1}^{n} \sum_{j=1}^{p} \left\{ z_{ij} \left( \Omega_{ij} + \sum_{j' \neq j} \Omega_{ijj'} z_{ij} z_{ij'} \right) - \frac{1}{2} z_{ij}^2 \right\}.$$

Then Lemma 1 can be proved by the definition of $z_{i,-j}$.

B. Proof of Lemma 2

According to the analysis in Section 3.1, we treat the PL as $p$ partially-linear additive linear regressions. Then, for each regression, we can derive $\hat{M}_{ij}$ as the estimation to the smooth part following the rationale in (Fan et al., 2005). Combining the results for every regression, we can derive Lemma 2.

C. Proof of Theorem 1

In this Section, we prove the $\sqrt{n}$-sparsistency of the $L_1$-regularized MPPLE by following the widely-used primal-dual witness proof technique (Wainwright, 2009; Ravikumar et al., 2010; Yang & Ravikumar, 2011; Yang et al., 2015a). PDW is characterized by the following Lemma 3:

**Lemma 3.** Let $\hat{\Omega}_0$ be an optimal solution to (5), and $\hat{Z}$ be the corresponding dual solution. If $\hat{Z}$ satisfies $\|\hat{Z}_N\|_\infty < 1$, then any given optimal solution to (5) $\tilde{\Omega}_0$ satisfies $\tilde{\Omega}_0 I = 0$. Moreover, if $H_{SS}$ is positive definite, then the solution to (5) is unique.

**Proof.** Specifically, following the same rationale as Lemma 1 in Wainwright 2009, Lemma 1 in Ravikumar et al. 2010, and Lemma 2 in Yang & Ravikumar 2011, we can derive Lemma 3 characterizing the optimal solution of (5).

**Bound $\|\nabla F(\Omega_0^*)\|_\infty$**

Before we use the PDW, we first provide a Lemma bounding $\|\nabla F(\Omega_0^*)\|_\infty$, which has been shown to be vital for PDW (Wainwright, 2009; Ravikumar et al., 2010; Yang & Ravikumar, 2011; Yang et al., 2015a).

**Lemma 4.** Let $r := 4C_5 \lambda$. For any $\epsilon_d > 0$, with probability of at least $1 - \epsilon_d$, there exists $C_4 > 0$ and $N_d > 0$ satisfying the following two inequalities:

$$\|\nabla F(\Omega_0^*)\|_\infty \leq C_4 \sqrt{\log p / n}, \quad (8)$$

$$\|\tilde{\Theta}_S - \Theta_0^*\|_\infty \leq r, \quad (9)$$

for $n > N_d$.

**Proof.** We prove (8) and (9) in turn.

**PROOF OF (8)**

To begin with, we prove (8). We define

$$\lambda_{ij}^* = (1_i - S_{ij})^\top x_j \Omega_0^*.$$


We first study \( \frac{\partial F(\Omega^*_0)}{\partial \Omega^*_0} \) is:

\[
\frac{\partial F(\Omega^*_0)}{\partial \Omega^*_0} = \frac{1}{n} \sum_{i=1}^{n} \left\{ -\begin{pmatrix} 1 - S_{ij} \end{pmatrix} \begin{pmatrix} y_j \end{pmatrix} \right\}_{j'} + \lambda^*_{ij} \left[ \begin{pmatrix} 1 - S_{ij} \end{pmatrix} \begin{pmatrix} x_j \end{pmatrix} \right]_{j'} \right\} \\
+ \frac{1}{n} \sum_{i=1}^{n} \left\{ -\begin{pmatrix} 1 - S_{ij} \end{pmatrix} \begin{pmatrix} y_{j'} \end{pmatrix} \right\}_{j'} \left[ \begin{pmatrix} 1 - S_{ij} \end{pmatrix} \begin{pmatrix} x_{j'} \end{pmatrix} \right]_{j'} + \lambda^*_{ij'} \left[ \begin{pmatrix} 1 - S_{ij} \end{pmatrix} \begin{pmatrix} x_{j'} \end{pmatrix} \right]_{j'} \right\},
\]

where \([\cdot]_{j}\) denotes the \(j\)th component of the vector. For the ease of presentation, we define

\[
y_j' = \begin{pmatrix} (1 - S_{1ij}) \begin{pmatrix} y_j \end{pmatrix} \\ \vdots \\ (1 - S_{nij}) \begin{pmatrix} y_j \end{pmatrix} \end{pmatrix} \quad \text{and} \quad x_j' = \begin{pmatrix} (1 - S_{1ij}) \begin{pmatrix} x_j \end{pmatrix} \\ \vdots \\ (1 - S_{nij}) \begin{pmatrix} x_j \end{pmatrix} \end{pmatrix}.
\]

Then, we consider

\[
\frac{x_j'^T x_j'^T \Omega^*_0 - x_j'^T y_j'}{n},
\]
whose \(j\)th component is just the target value

\[
\frac{\sum_{i=1}^{n} \left\{ -\begin{pmatrix} 1 - S_{ij} \end{pmatrix} \begin{pmatrix} y_j \end{pmatrix} \right\}_{j'} + \lambda^*_{ij} \left[ \begin{pmatrix} 1 - S_{ij} \end{pmatrix} \begin{pmatrix} x_j \end{pmatrix} \right]_{j'} \right\}}{n}.
\]

Therefore, we focus on bounding (11). Then,

\[
\frac{x_j'^T x_j'^T \Omega^*_0 - x_j'^T y_j'}{n} = \frac{x_j'^T x_j' \left[ \Omega^*_0 - \left( x_j'^T x_j' \right)^{-1} x_j'^T y_j' \right]}{n} = \frac{x_j'^T (I - S_{ij}) (M_j + \epsilon_j)}{n},
\]

where the second equality is due to Lemma 5.

We first study \( \frac{x_i'^T (I - S_{ij}) M_j}{n} \) : according to Lemma 8, for any \( \epsilon_a > 0 \), there exists \( \delta_a > 0 \) and \( N_a > 0 \) satisfying

\[
P \left\{ \left\| \frac{x_i'^T (I - S_{ij}) M_j}{n} \right\|_{\infty} > \delta_a \frac{\log \left( \frac{1}{\epsilon_a} \right)}{n} + h^4 + 2h^2 \sqrt{\frac{\log \left( \frac{1}{\epsilon_a} \right)}{n}} \right\} < \epsilon_a,
\]

with \( n > N_a \).

According to Assumption 1

\[
P \left\{ \left\| \frac{x_i'^T (I - S_{ij}) M_j}{n} \right\|_{\infty} > \delta_a C_1 \sqrt{\frac{\log p}{n}} \right\} < \epsilon_a.
\]

Now, we study \( \frac{x_i'^T (I - S_{ij})}{n} \epsilon_j \). According to Lemma 9, we have

\[
\frac{x_i'^T (I - S_{ij})}{n} \epsilon_j = \sum_{i=1}^{n} \left\{ x_{ij} - E^T \left[ \mathbb{1}_{g_i > g_j} Z_{i,-j} Z_{i,-j}^T \right] \frac{g_i}{g_i} \right\} E^{-1} \left[ \mathbb{1}_{g_i > g_j} Z_{i,-j} Z_{i,-j}^T \right] x_{ij},
\]

\[
\epsilon_{ij} (1 + o_p(1))/n,
\]

Partially Linear Additive Gaussian Graphical Models
uniformly for \( j \). Note that \( \frac{X_j^T (I - S_j) \epsilon_j}{n} \) is a \( p \times 1 \) vector. Therefore, for the \( j^{th} \) component, we have

\[
\begin{align*}
\left| \frac{X_j^T (I - S_j) \epsilon_j}{n} \right|_{j'} & \leq \sum_{i: g_i' \leq g^*} \left( 1 - \mathbb{I}_{g_i' > g^*} \right) \left| z_{ij'} \epsilon_{ij'} \right| (1 + |o_p(1)|) / n \\
& = \frac{1}{2n} \sum_{i=1} \left( 1 - \mathbb{I}_{g_i > g^*} \right) \left[ \left( \frac{z_{ij'} + \epsilon_{ij}}{\sqrt{2}} \right)^2 - 1 \right] \\
& \quad \quad - \sum_{i=1} \left( 1 - \mathbb{I}_{g_i > g^*} \right) \left[ \left( \frac{z_{ij'} - \epsilon_{ij}}{\sqrt{2}} \right)^2 - 1 \right] \geq \frac{1}{2} \left( \frac{\epsilon_{ij}}{n} + \frac{\epsilon_c}{n} \right) (1 + |o_p(1)|) \\
& \quad \quad - 4 \sqrt{n} |\epsilon_{ij}| - 2 \epsilon_c \\
& \leq \exp(-2 \epsilon_c) \\
\end{align*}
\]  

It can be shown that \( \left( \frac{z_{ij'} + \epsilon_{ij}}{\sqrt{2}} \right)^2 \) and \( \left( \frac{z_{ij'} - \epsilon_{ij}}{\sqrt{2}} \right)^2 \) are independent and follow chi-squared distribution with degree equal to 1.

By Lemma 1 in (Laurent & Massart, 2000), the linear combination of chi-squared random variables satisfies:

\[
\begin{align*}
P \left\{ \sum_{i=1} \left( 1 - \mathbb{I}_{g_i > g^*} \right) \left[ \left( \frac{z_{ij'} + \epsilon_{ij}}{\sqrt{2}} \right)^2 - 1 \right] \geq 2 \sqrt{n} \epsilon_c + 2 \epsilon_c \right\} & \leq \exp(-\epsilon_c), \\
P \left\{ \sum_{i=1} \left( 1 - \mathbb{I}_{g_i > g^*} \right) \left[ \left( \frac{z_{ij'} - \epsilon_{ij}}{\sqrt{2}} \right)^2 - 1 \right] \leq -2 \sqrt{n} \epsilon_c - 2 \epsilon_c \right\} & \leq \exp(-\epsilon_c), \\
\end{align*}
\]

and

\[
\begin{align*}
P \left\{ \sum_{i=1} \left( 1 - \mathbb{I}_{g_i > g^*} \right) \left[ \left( \frac{z_{ij'} - \epsilon_{ij}}{\sqrt{2}} \right)^2 - 1 \right] \geq 2 \sqrt{n} \epsilon_c + 2 \epsilon_c \right\} & \leq \exp(-2 \epsilon_c). \\
\end{align*}
\]

for any \( \epsilon_c > 0 \). Combining the previous four probabilistic bounds, we can derive

\[
\begin{align*}
P \left\{ \sum_{i=1} \left( 1 - \mathbb{I}_{g_i > g^*} \right) \left[ \left( \frac{z_{ij'} + \epsilon_{ij}}{\sqrt{2}} \right)^2 - 1 \right] - \sum_{i=1} \left( 1 - \mathbb{I}_{g_i > g^*} \right) \left[ \left( \frac{z_{ij'} - \epsilon_{ij}}{\sqrt{2}} \right)^2 - 1 \right] \right. \\
& \left. \quad \quad \geq \frac{1}{2} \left( \frac{\epsilon_{ij}}{n} + \frac{\epsilon_c}{n} \right) (1 + |o_p(1)|) \right\} & \leq \exp(-2 \epsilon_c).
\end{align*}
\]

Taking (14) and (15) into (13), we can derive

\[
P \left\{ \left| \frac{X_j^T (I - S_j) \epsilon_j}{n} \right|_{j'} \geq \left( \frac{\epsilon_{ij}}{n} + \frac{\epsilon_c}{n} \right) (1 + |o_p(1)|) \right\} \leq 2 \exp(-2 \epsilon_c).
\]

Then, by the definition of \( o_p(1) \), for any \( \epsilon_b > 0 \), there exists \( N_b \) so that for \( n > N_b \):

\[
P \left\{ |o_p(1)| \geq 1 \right\} \leq \epsilon_b.
\]
Combining (16) and (17), we derive

$$\mathbb{P}\left\{ \left\| \frac{X_j^T (I - S_j) M_j}{n} \right\|_{\infty} \geq \left( 4 \sqrt{\frac{\epsilon_a}{n}} + 2 \frac{\epsilon_c}{n} \right) \right\} \leq 2p \exp(-2\epsilon_c) + \epsilon_b,$$

(18)

by a union bound. Eventually, according to (12) and (18), and by setting $\epsilon_c = 2 \log p$ we prove:

$$\left\| \frac{x_j^T X_j' \Omega_{0,j} - x_j^T y_j}{n} \right\|_{\infty} \leq (6 + \delta_c C_1) \frac{2 \log p}{n},$$

with probability larger than $1 - \epsilon_b - \epsilon_a - 2p^{-1}$. Thus, for any $\epsilon_d > 0$, there exists $C_4 > 0$ and $N_d > 0$

$$\| \nabla F(\Theta^*) \|_{\infty} \leq C_4 \sqrt{\frac{\log p}{n}},$$

with probability larger than $1 - \epsilon_d$, for $n > N_d$.

**PROOF OF (9)**

To prove (9), we use the fixed point method by defining a map $G(\Delta_S) := -H^{-1}_{SS} \left[ \nabla_S F(\Omega_{0,S}^* + \Delta_S) + \lambda \hat{Z}_S \right] + \Delta_S$. If $\| \Delta \|_{\infty} \leq r$, by Taylor expansion of $\nabla_S F(\Omega_0 + \Delta)$ centered at $\nabla_S F(\Omega_0)$,

$$\| G(\Delta_S) \|_{\infty} = \left\| -H^{-1}_{SS} \left[ \nabla_S F(\Omega_{0,S}^* + \Delta_S + R_S(\Delta)) + \lambda \hat{Z}_S \right] + \Delta_S \right\|_{\infty}$$

$$\leq \| H^{-1}_{SS} \left[ \nabla_S F(\Omega_{0,S}^* + \Delta) + \lambda \hat{Z}_S \right] \|_{\infty} + \| R_S(\Delta) \|_{\infty} + \lambda \| \hat{Z}_S \|_{\infty}$$

$$\leq C_2 (\lambda + C_3 r^2 + \lambda) = C_2 C_3 r^2 + 2C_2 \lambda,$$

where the inequality is due to Assumption 4 and Assumption 5, and $\| \nabla_S F(\Theta^*) \|_{\infty} \leq \lambda$ with a high probability, according to (8). Then, based on the definition of $r$, we can derive the upper bound of $\| G(\Delta_S) \|_{\infty}$ as $\| G(\Delta_S) \|_{\infty} \leq r/2 + r/2 = r$.

Therefore, according to the fixed point theorem (Ortega & Rheinboldt, 2000; Yang & Ravikumar, 2011), there exists $\Delta_S$ satisfying $G(\Delta_S) = \Delta_S$, which indicates $\nabla_S F(\Omega_{0,S}^* + \Delta) + \lambda \hat{Z}_S = 0$. The optimal solution to (20) is unique, and thus $\hat{\Delta}_S = \Delta_S$. Therefore, $\| \hat{\Delta}_S \|_{\infty} \leq r$, with probability larger than $1 - \epsilon$.

**PDW**

By Lemma 3, we can prove the sparsistency by building an optimal solution to (5) satisfying the strict dual feasibility (SDF) defined as $\| \hat{Z}_N \|_{\infty} < 1$, which is summarized. Therefore, we now build a solution by solving a restricted problem.

**SOLVE A RESTRICTED PROBLEM**

First of all, we derive the KKT condition of (5):

$$\nabla F(\Omega_0) + \lambda \hat{Z} = 0.$$

(19)

To construct an optimal primal-dual pair solution, we define $\tilde{\Omega}_0$ as an optimal solution to the restricted problem:

$$\tilde{\Omega}_0 := \min_{\Omega_0} F(\Omega_0) + \lambda \| \Omega_0 \|_1,$$

(20)

with $\Omega_{0,N} = 0$. $\tilde{\Omega}_0$ is unique due to Lemma 3. Then, we define the subgradient corresponding to $\tilde{\Omega}_0$ as $\hat{Z}$. Therefore, $(\tilde{\Omega}_0, \hat{Z})$ is a pair of optimal solutions to the restricted problem (20). $\hat{Z}_S$ is determined according to the values of $\Omega_{0,S}$ via the KKT conditions of (20). Thus we have

$$\nabla_S F(\Theta) + \lambda \hat{Z}_S = 0,$$

(21)

where $\nabla_S$ represents the gradient components with respect to $S$. Letting $\tilde{\Omega}_0 = \hat{\Omega}_0$, we determine $\hat{Z}_N$ according to (19). It now remains to show that $\hat{Z}_N$ satisfies SDF.
Now, we demonstrate that $\tilde{\Theta}$ and $\tilde{Z}$ satisfy SDF. We define $\tilde{\Delta} := \tilde{\Theta} - \Theta^*$. By (21), and by the Taylor expansion of $\nabla_S F(\Omega_0)$, we have that

$$H_{SS} \Delta_S + \nabla_S F(\Omega_0^*) + R_S(\Delta) + \lambda \tilde{Z}_S = 0,$$

which means

$$\Delta_S = H_{SS}^{-1} \left[ - \nabla_S F(\Omega_0^*) - R_S(\Delta) - \lambda \tilde{Z}_S \right],$$

where $H_{SS}$ is positive definite and hence invertible.

By the definition of $\tilde{\Omega}_0$ and $\tilde{Z}$,

$$\nabla F(\tilde{\Omega}_0) + \lambda \tilde{Z} = 0 \Rightarrow \nabla F(\tilde{\Omega}_0^*) + H \Delta + R(\tilde{\Omega}_0) + \lambda \tilde{Z} = 0 \Rightarrow \nabla_S F(\tilde{\Theta}) + H_{NS} \Delta_S + R_N(\Delta) + \lambda \tilde{Z}_N = 0. \quad (23)$$

Due to (22),

$$\lambda \left\| \tilde{Z}_N \right\|_\infty = \left\| -H_{NS} \Delta_S - \nabla_S F(\Omega_0^*) - R_N(\Delta) \right\|_\infty \leq \left\| H_{NS} H_{SS}^{-1} \left[ - \nabla_S F(\Omega_0^*) - R_S(\Delta) - \lambda \tilde{Z}_S \right] \right\|_\infty + \left\| \nabla_S F(\Omega_0^*) + R_N(\Delta) \right\|_\infty.$$

Further, we use the Assumption 4,

$$\lambda \left\| \tilde{Z}_N \right\|_\infty \leq (1 - \alpha) \left( \left\| \nabla_S F(\Omega_0^*) \right\|_\infty + \left\| R_S(\Delta) \right\|_\infty \right) + (1 - \alpha) \lambda + \left( \left\| \nabla_S F(\Omega_0^*) \right\|_\infty + \left\| R_N(\Delta) \right\|_\infty \right),$$

where we have used in the first inequality, and the third inequality is due to Assumption 4.

Now, we study $\left\| \nabla F(\Omega_0^*) \right\|_\infty$. By Lemma 4 and the assumption on $\lambda$ in Theorem 1, $\left\| \nabla F(\Theta^*) \right\|_\infty \leq \frac{\alpha C_2}{4} \sqrt{\frac{\log p}{n}} \leq \frac{\alpha \lambda}{4}$, with probability larger than $1 - \epsilon_d$.

It remains to control $\left\| R(\Delta) \right\|_\infty$. According to Assumption 5 and Lemma 4,

$$\left\| R(\Delta) \right\|_\infty \leq C_3 \left\| \Delta \right\|_\infty^2 \leq C_3 r^2 \leq C_3 (4C_2 \lambda)^2 \leq \frac{\lambda 64 C_2^2 C_3 \alpha \lambda}{4} \leq \left( C_5 \sqrt{\frac{\log p}{n}} \right) \frac{64 C_2^2 C_3 \alpha \lambda}{4}. \quad (25)$$

where in the last inequality we have used the assumption $\lambda \leq C_5 \sqrt{\frac{\log p}{n}}$ in Theorem 1. Therefore, when we choose $n \geq \left( 64 C_2^2 C_3 / \alpha \right)^2 \log p$ in Theorem 1, from (25), we can conclude that $\left\| R(\Delta) \right\|_\infty \leq \frac{\alpha \lambda}{4}$. As a result, $\lambda \left\| \tilde{Z}_N \right\|_\infty$ can be bounded by $\lambda \left\| \tilde{Z}_N \right\|_\infty < \alpha \lambda / 2 + \alpha \lambda / 2 + (1 - \alpha) \lambda = \lambda$. Combined with Lemma 3, we demonstrate that any optimal solution of (5) satisfies $\Theta_N = 0$. Furthermore, (9) controls the difference between the optimal solution of (5) and the real parameter by $\left\| \Delta_S \right\|_\infty \leq r$, by the fact that $r \leq \left\| \Theta_N^* \right\|_\infty$ in Theorem 1, $\Theta_S$ shares the same sign with $\Theta_S^*$.

**Auxiliary Lemmas**

In this section, we provide and prove the used auxiliary lemmas.

**Lemma 5.** For the graphical model defined in Section 2 parameterized by $\Omega_0^*$, the conditional distribution of $Z_{ij}$ follows

$$Z_{ij} \mid G_i = g_i \sim Z_{i,j^-}\Omega_{0,j} + M_{ij} + \epsilon_{ij},$$

where

$$[Z_{i,j^-}]_j' = \begin{cases} Z_{ij'} & j' \neq j' \ni j \ni \epsilon_{ij'} \end{cases}$$

$\epsilon_{ij'}$s follow the standard normal distribution, and $\epsilon_{ij'}$ is independent with $\epsilon_{i'j'}$ for $j \neq j'$ in $[p]$. 

Proof. According to Lemma 1, the node-wise conditional distribution of a PLA-GGM follows a Gaussian distribution. Then, Lemma 5 can be proved.

**Lemma 6.** For a kernel regression on \( \{x_i, y_i\}_{i=1}^n \) as the IID samples of \((X, Y)\). Assume that \( \mathbb{E}|Y|^s \leq \infty \) and \( \sup_X \mathbb{E}|Y|^8 f(X, Y) dY \leq \infty \). Given that \( n^{2\epsilon - 1} h \to \infty \) for \( \epsilon < 1 - s^{-1} \), we have

\[
\sum_i \left| \frac{1}{n} \sum_{i=1}^n \left[ K_h(x_i - x) - \mathbb{E} \{ K_h(x_i - x) y_i \} \right] \right| = O_p \left( \left( \frac{\log(1/h)}{nh} \right)^{1/2} \right).
\]

Proof. Lemma 6 follows (Mack & Silverman, 1982).

**Lemma 7.** Suppose \( Y = \{Y_1, Y_2, \cdots, Y_n\} \) follows a multivariate Gaussian distribution, then \( \max |Y_i| \) follows a sub-Gaussian distribution with variance \( \max \text{var}(Y_i) \). Further, for any \( t > 0 \), the tail probability can be controlled via

\[
P \left\{ \max |\epsilon_{ij}| \geq t \right\} \leq \exp \left( -\frac{t^2}{2} \right).
\]

**Lemma 8.** For any \( \epsilon > 0 \), there exists \( \delta > 0 \) and \( N > 0 \), so that when \( n > N \), we have

\[
P \left\{ \left\| \frac{X_i'(I-S_j)M_j}{n} \right\|_{\infty} \geq \delta \epsilon_n^2 \right\} \leq \epsilon,
\]

uniformly for \( j \in [p] \).

Proof. To start with, we review the definition of \( S_{ij} \)

\[
S_{ij} = \left[ \mathbb{I}_{g_i > g^*} Z_{i,-j}^T \right] (D_{ij}^T W_i D_{ij})^{-1} D_{ij}^T W_i.
\]

We first study \( D_{ij}^T W_i D_{ij} \):

\[
D_{ij}^T W_i D_{ij} = \begin{bmatrix}
\sum_{i' = 1}^n \mathbb{I}_{g_{i'} > g^*} Z_{i',j}^T \psi \left( |g_{i'} - g_i| / h \right) \\
\sum_{i' = 1}^n \mathbb{I}_{g_{i'} > g^*} Z_{i',j}^T \psi \left( |g_{i'} - g_i| / h \right)
\end{bmatrix}
\]

To bound \( D_{ij}^T W_i D_{ij} \) uniformly over \( j \), we consider a random vector \( B_i = \left[ \mathbb{I}_{g_i > g^*} Z_{i}^T, 1 \right]^T \), with observations

\[
\begin{bmatrix}
b_1 = \mathbb{I}_{g_i > g^*} Z_{1}^T, 1 \\
b_n = \mathbb{I}_{g_i > g^*} Z_{n}^T, 1
\end{bmatrix}
\]

Then, we study an auxiliary matrix

\[
O_i = \begin{bmatrix}
\sum_{i' = 1}^n \mathbb{I}_{g_{i'} > g^*} b_{i'} b_{i'}^T \psi \left( |g_{i'} - g_i| / h \right) \\
\sum_{i' = 1}^n \mathbb{I}_{g_{i'} > g^*} b_{i'} b_{i'}^T \psi \left( |g_{i'} - g_i| / h \right)
\end{bmatrix}
\]

Therefore, the components of \( D_{ij}^T W_i D_{ij} \) belong to \( O_i \), and each part of \( O_i \) is in the form of a kernel regression. By Lemma 6, we have

\[
O_i = n f(g_i) \mathbb{E} \left[ B_i B_i^T | g_i \right] \otimes \begin{bmatrix} 1 & 0 \\ 0 & \mu_2 \end{bmatrix} \{ 1 + O_p(c_n) \},
\]

which holds uniformly for \( i \). Therefore,

\[
D_{ij}^T W_i D_{ij} = n f(g_i) \mathbb{E} \left[ \mathbb{I}_{g_i > g^*} Z_{i,-j} Z_{i,-j}^T | g_i \right] \otimes \begin{bmatrix} 1 & 0 \\ 0 & \mu_2 \end{bmatrix} \{ 1 + O_p(c_n) \}
\]
holds uniformly for \( i \) with the same \( O_p(c_n) \) for every \( j \). Define
\[
\alpha_j(g_i) = \begin{bmatrix} \Omega_{1,j} & \cdots & \Omega_{n,j} \end{bmatrix}.
\]
By the same technique, uniformly for \( i \) and with the same \( O_p(c_n) \) for every \( j \), we can show
\[
D_{ij}^T W_i M_j = n f(g_i) E \left[ 1_{g_i > g_j} Z_{i,-j} Z_{i,-j}^T | g_i \right] \otimes \begin{bmatrix} 1 & 0 \end{bmatrix}^T \alpha_j(g_i) \{ 1 + O_p(c_n) \},
\]
(27)
and
\[
D_{ij}^T W_i x_j = n f(g_i) E \left[ 1_{g_i > g_j} Z_{i,-j} Z_{i,-j}^T | g_i \right] \otimes \begin{bmatrix} 1 & 0 \end{bmatrix}^T \{ 1 + O_p(c_n) \}.
\]
(28)
Combining (26) and (27) we have
\[
\begin{bmatrix} \bar{x}_{ij}^T & 0 \end{bmatrix} (D_{ij}^T W_i D_{ij})^{-1} D_{ij}^T W_i M_j = \bar{x}_{ij}^T \alpha_j(g_i) \{ 1 + O_p(c_n) \}.
\]
(29)
Similarly, combining (26) and (28), we have
\[
x_{ij}' = x_{ij} - \bar{x}_{ij} E^{-1} \left[ 1_{g_i > g_j} Z_{i,-j} Z_{i,-j}^T | g_i \right] E^{-1} \left[ 1_{g_i > g_j} Z_{i,-j} Z_{i,-j}^T | g_i \right] \bar{x}_{ij} \epsilon_{ij},
\]
(30)
Next, we follow the rationale of the Lemma A.4 in (Fan et al., 2005), and combine (29) and (30). Finally, we have
\[
\frac{x_{ij}' (I - S_j) M_j}{n} = O_p(c_n^2)
\]
uniformly for \( j \).

\textbf{Lemma 9.} For any \( \epsilon > 0 \), there exists \( N > 0 \), so that when \( n > N \), we have
\[
\| x_{ij}'^T (I - S_j) \epsilon_j \| \propto 2 \sum_{i=1}^{n} \left\{ x_{ij} - \bar{x}_{ij} \right\} \left( D_{ij}^T W_i D_{ij} \right)^{-1} D_{ij}^T W_i \epsilon_j \bigg\} \epsilon_{ij},
\]
uniformly for \( j \in [p] \) with probability less than \( \epsilon \).

\textbf{Proof.} By definition, we have
\[
x_{ij}'^T (I - S_j) \epsilon_j = \sum_{i=1}^{n} \left\{ x_{ij} - \bar{x}_{ij} \right\} \left( D_{ij}^T W_i D_{ij} \right)^{-1} D_{ij}^T W_i \epsilon_j \bigg\} \epsilon_{ij},
\]
Using the technique in (26), we have
\[
\begin{bmatrix} \bar{x}_{ij}^T & 0 \end{bmatrix} (D_{ij}^T W_i D_{ij})^{-1} D_{ij}^T W_i \epsilon_j = \bar{x}_{ij} E^{-1} \left[ 1_{g_i > g_j} Z_{i,-j} Z_{i,-j}^T | g_i \right] E \left[ \bar{x}_{ij}^T | g_i \right] O_p(c_n).
\]
Therefore,
\[
x_{ij}'^T (I - S_j) \epsilon_j = \sum_{i=1}^{n} \left\{ x_{ij} - \bar{x}_{ij} \right\} \left[ 1_{g_i > g_j} Z_{i,-j} Z_{i,-j}^T | g_i \right] E^{-1} \left[ 1_{g_i > g_j} Z_{i,-j} Z_{i,-j}^T | g_i \right] \bar{x}_{ij} \epsilon_{ij} [1 + o_p(1)],
\]
uniformly for \( j \). \qed
D. Proof of Theorem 2

We first study CON-GGMs. According to (6) and (Eaton, 1983), we have

$$[\text{cov}(Z \mid G = g)]^{-1} = \left[\Sigma_{ZZ} - \Sigma_{ZG}\Sigma_{GG}^{-1}\Sigma_{GZ}\right]^{-1},$$

whose right-hand side has nothing to do with $g$. Therefore, the conditional distribution of $Z \mid G = g$ follows a GGM with parameter $[\Sigma_{ZZ} - \Sigma_{ZG}\Sigma_{GG}^{-1}\Sigma_{GZ}]^{-1}$ irrelevant to $g$. In other words CON-GGM is equivalent to assuming that $G$ follows a normal distribution and $R(g) = 0$ on the basis of the proposed PLA-GGM.

Then, we study LR-GGMs. Again, given $G = g$ for any $g$, we have

$$[\text{cov}(Z \mid G = g)]^{-1} = \Omega_0,$$

which has nothing to do with $G$ either. Given $G = g$, the conditional distribution of $Z \mid G = g$ follows a GGM with the parameter $\Omega$. Therefore, LR-GGM is a special case of the proposed PLA-GGM by assuming $R(g) = 0$.

E. Experiments

Data Simulation

To simulate the samples from PLA-GGMs, we first define $f(g)$ as follows:

$$f(g) = \begin{cases} 
  g - 10 & g > 12 \\
  x + \frac{(x-12)^2}{4} - 11 & 10 < g \leq 12 \\
  0 & -10 < g \leq 10 \\
  x + \frac{(x+12)^2}{4} + 11 & -12 < g \leq -10 \\
  g + 10 & g \leq -12
\end{cases}$$

We provide the following procedure:

1. We consider $p = 10, 20, 50, 100$, and implement the following steps separately.
2. We randomly generate a sparse precision matrix as $\Omega_0$. Specifically, each element of $\Omega_0$ is drawn randomly to be non-zero with probability 0.3.
3. A dense precision matrix $W$ is generated to build the confounding.
4. We take $\{-400, \cdots, 0, \cdots, 399\}$ as the confounders. For each $g \in \{-400, \cdots, 0, \cdots, 399\}$, the precision matrix is selected to be $\Omega(g) = \Omega_0 + f(g)W$, and a sample is generated by a GGM with parameter $\Omega(g)$. Thus, we get 800 samples.

Note that the procedure is equivalent to selecting $g^* = 10$.

Glass Brains for Brain Function Connectivity Estimation

We report the glass brains from other angles for the brain function connectivity estimation experiment in Section 6.2.

Schizophrenia Diagnosis using Different $1_{\{|g| \geq g^*\}}$’s

We conduct the analysis in Section 6.2 using different $1_{\{|g| \geq g^*\}}$’s. Specifically, we consider the function $1 - \exp(-kx^2)/2$ using $k = 144, 150$. The achieved accuracy using the parameter selected by the 10-fold cross validation and AIC are reported in Figure 9. The performance of PLA-GGMs is not hugely affected when selecting $1_{\{|g| \geq g^*\}}$ in a reasonable range, which is consistent with our analysis in Theorem 1. Note that, if we select $k$ too large, the PPL method will be not applicable. The reason is that a large $k$ corresponds to a small $g^*$, and will induce few non-confounded samples observed. As a result, $(D^T_{ij}W_iD_{ij})$ will be singular. In practice, if we use a relative large $g^*$ corresponding to a small $k$, (2) will tend to be like $R(g) = 0$ used in CON-GGMs and LR-GGMs.
Partially Linear Additive Gaussian Graphical Models

Figure 5: Controls using PLA-GGMs

Figure 6: Patients using PLA-GGMs

Figure 7: Controls using LR-GGMs
Figure 8: Patients using LR-GGMs

Figure 9: Diagnosis using different $I_{\{|g| \geq g^*\}}$'s.