A PROOFS

A.1 PROPOSITION 4.1

Proof. We begin by observing that we can write $W_m$ as

$$W_m = \left( m^{-1}X_i^\top X_i + \frac{n}{m}m^{-1}X_0^\top X_0 \right)^{-1} \left( m^{-1}X_i^\top X_i \right). \quad (A)$$

We apply the strong law of large numbers to obtain that

$$m^{-1}X_i^\top X_i \xrightarrow{a.s.} \text{Cov}(X_i) \quad \text{and} \quad n^{-1}X_0^\top X_0 \xrightarrow{a.s.} \text{Cov}(X_0).$$

Due to the fact that $\lim_{m \to \infty} \frac{n(m)}{m} = c$ for some $c > 0$, we conclude

$$W_m \xrightarrow{a.s.} W_\infty := \left( \text{Cov}(X_i) + c \cdot \text{Cov}(X_0) \right)^{-1} \text{Cov}(X_i).$$

We observe that

$$(I - W_\infty) = \left( \text{Cov}(X_i) + c \cdot \text{Cov}(X_0) \right)^{-1} c \cdot \text{Cov}(X_0).$$

Since both covariance matrices are positive definite, so is $\text{Cov}(X_i) + c \cdot \text{Cov}(X_0)$. We conclude that the smallest singular value of $I - W_\infty$ is strictly greater than 0. This means

$$\left\| \mathbb{E}[\hat{\alpha}_{W_\infty}^m] - \alpha \right\|_2^2 = \left\| (I_p - W_\infty) \Delta \right\|_2^2 \geq c' \left\| \Delta \right\|_2^2,$$

for some fixed constant $c' > 0$. We obtain therefore

$$0 < \lim_{m \to \infty} \left\| \mathbb{E}[\hat{\alpha}_{W_\infty}^m] - \alpha \right\|_2^2 \leq \lim_{m \to \infty} \text{MSE} \left( \hat{\alpha}_{W_\infty}^m \right),$$

where we invoked Jensen’s inequality. We see that $W_\infty$ is constant and bounded. We note that almost sure convergence implies convergence in probability. We can thus apply Lemma B.1, which yields the desired result

$$0 < \lim_{m \to \infty} \text{MSE} \left( \hat{\alpha}_{W_\infty}^m \right) \leq \lim_{m \to \infty} \text{MSE} \left( \hat{\alpha}_{W_p}^m \right).$$

\[\square\]
A.2 PROPOSITION 4.2

**Proposition 4.2.** Let \( \lim_{m \to \infty} \frac{n(m)}{m} = 0 \). Then, it holds that

\[
\lim_{m \to \infty} \text{MSE} \left( \hat{\alpha}_m^m \right) = 0.
\]

**Proof.** Similar to the proof of Proposition 4.1, we employ the formulation of (A) and consider the term

\[
\frac{n}{m} \cdot n^{-1} X_o^\top X_o.
\]

We see that \( \lim_{m \to \infty} \frac{n(m)}{m} = 0 \) and by the strong law of large numbers, \( n^{-1} X_o^\top X_o \overset{a.s.}{\to} \text{Cov}(X_o) \). Hence, we obtain that

\[
\frac{n}{m} \cdot n^{-1} X_o^\top X_o \overset{a.s.}{\to} 0.
\]

By the continuous mapping theorem, we conclude that

\[
\hat{\alpha}_m^m \overset{a.s.}{\to} \text{I}_p,
\]

and by Lemma B.2, this implies that

\[
\lim_{m \to \infty} \text{MSE} \left( \hat{\alpha}_m^m \right) \leq \lim_{m \to \infty} \text{MSE} \left( \hat{\alpha}_1^m \right) = 0.
\]

\( \square \)

A.3 PROPOSITION 4.3

**Proof.** We rewrite \( \hat{\alpha}_m^m \) as follows:

\[
\hat{\alpha}_m^m = \left( n^{-1} \frac{n^{-1} X_o^\top X_o}{n-1} \sigma_{Y|X}^2 + \hat{\Delta} \hat{\Delta}^\top + \epsilon I_p \right)^{-1}. 
\]

where we insert any almost surely converging estimators for \( \Delta, \sigma_{Y|X}^2 \) instead of their ground-truth values. By almost sure convergence of linear estimators individually, we see that this holds specifically for \( \hat{\Delta} = \hat{\alpha}_0^m - \hat{\alpha}_1^m \). Also, we can use the strong law of large numbers to conclude almost sure convergence of \( \hat{\sigma}_{Y|do(X)}^2 \) and \( \hat{\sigma}_{Y|X}^2 \).

We now show \( \hat{\alpha}_m^m \overset{a.s.}{\to} \text{I}_p \). First, we see that

\[
(c m)^{-1} \left( n^{-1} X_o^\top X_o \right)^{-1} \sigma_{Y|X}^2 \overset{a.s.}{\to} 0 \quad \text{and} \quad m^{-1} \left( m^{-1} X_i^\top X_i \right)^{-1} \sigma_{Y|do(X)}^2 \overset{a.s.}{\to} 0,
\]

since \( m^{-1} X_i^\top X_i \sigma_{Y|do(X)}^2 \) and \( n^{-1} X_o^\top X_o \sigma_{Y|X}^2 \) converge almost surely to constants and \( m^{-1} \) vanishes. Hence,

\[
\hat{\alpha}_m^m \overset{a.s.}{\to} \left( \Delta \Delta^\top + \epsilon I_p \right) \left( \Delta \Delta^\top + \epsilon I_p \right)^{-1} = \text{I}_p.
\]

\( \square \)

A.4 THEOREM 4.4

**Proof.** We have that \( \text{I}_p \) is bounded in norm, almost surely. So we can apply Lemma B.2 to see that

\[
\lim_{m \to \infty} \text{MSE} \left( \hat{\alpha}_m^m \right) \leq \lim_{m \to \infty} \text{MSE} \left( \hat{\alpha}_1^m \right) = 0.
\]

\( \square \)
A.5 PROPOSITION 4.5

Proof. By Theorem 4.4, it suffices to show that \( \hat{\mathbf{W}}_m \overset{a.s.}{\to} \mathbf{I}_p \). Since the other quantities \( \text{Cov}(\hat{\alpha}_i^m) \), \( \text{Cov}(\hat{\alpha}_i^m) \) for estimating \( \mathbf{W}_m \) remain unchanged compared to \( \hat{\mathbf{W}}_m \), it suffices to show that the modified computation of \( \Delta_m \) we call \( \Delta_m^\ell \) converges almost surely to the true \( \Delta = \alpha_i - \alpha_o \), where \( \alpha_i \) and \( \alpha_o \) are short-hand for \( E_{\text{est}}[Y \mid \mathbf{X} = \mathbf{x}] \) and \( E_{\text{obs}}[Y \mid \mathbf{X} = \mathbf{x}] \), respectively. We observe that \( \Delta_m^\ell \) has a closed-form solution

\[
\Delta_m^\ell = -(\mathbf{X}_i^\top \mathbf{X}_i + \lambda_i \mathbf{I}_p)^{-1} \mathbf{X}_i^\top (\mathbf{y}_i - \mathbf{X}_i \hat{\alpha}_o^m) = (\mathbf{X}_i^\top \mathbf{X}_i + \lambda_i \mathbf{I}_p)^{-1} \mathbf{X}_i^\top \mathbf{X}_i \hat{\alpha}_o^m - (\mathbf{X}_i^\top \mathbf{X}_i + \lambda_i \mathbf{I}_p)^{-1} \mathbf{X}_i^\top \mathbf{y}_i,
\]

(B) since \( \hat{\alpha}_o^m \) is again a closed-form solution to an ordinary least squares problem. Considering the first term in \( (C) \), we conclude almost sure convergence with respect to \( \alpha_i \) (it is simply the ridge regression solution on the interventional data, which is well-known to converge almost surely for fixed \( \lambda_i \)). The second term satisfies

\[
(\mathbf{X}_i^\top \mathbf{X}_i + \lambda_i \mathbf{I}_p)^{-1} \mathbf{X}_i^\top \mathbf{X}_i \overset{a.s.}{\to} \mathbf{I}_p \quad \text{and} \quad \hat{\alpha}_o^m \overset{a.s.}{\to} \alpha_o.
\]

This leads to the desired conclusion. \(\Box\)

B ADDITIONAL LEMMAS

Lemma B.1. Let \( \hat{\mathbf{W}}_m - \mathbf{W}_m \overset{p}{\to} 0 \) and let there exist \( c > 0, m' \in \mathbb{N} \), such that \( \| \mathbf{W}_m \|_2 \leq c \), for all \( m \geq m' \), almost surely. Then, it holds that

\[
\lim_{m \to \infty} \text{MSE} \left( \hat{\alpha}_m^{\mathbf{W}_m} \right) \leq \lim_{m \to \infty} \text{MSE} \left( \hat{\alpha}_m^{\hat{\mathbf{W}}_m} \right),
\]

where \( \overset{p}{\to} \) denotes convergence in probability.

Proof. We derive a lower bound on \( \text{MSE} \left( \hat{\alpha}_m^{\hat{\mathbf{W}}_m} \right) \) by using the formulation

\[
\text{MSE} \left( \hat{\alpha}_m^{\hat{\mathbf{W}}_m} \right) = \mathbb{E} \left[ \mathbb{I} \left\{ \| \hat{\mathbf{W}}_m - \mathbf{W}_m \|_2 \leq \epsilon \right\} \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha \|_2^2 \right] + \mathbb{E} \left[ \mathbb{I} \left\{ \| \hat{\mathbf{W}}_m - \mathbf{W}_m \|_2 > \epsilon \right\} \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha \|_2^2 \right], \quad \forall \epsilon > 0.
\]

We bound the second summand of \( (D) \) from below by zero. For the first summand, we use reverse triangle inequality, which yields

\[
\mathbb{E} \left[ \mathbb{I} \left\{ \| \hat{\mathbf{W}}_m - \mathbf{W}_m \|_2 \leq \epsilon \right\} \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha \|_2^2 \right] = \mathbb{E} \left[ \mathbb{I} \left\{ \| \hat{\mathbf{W}}_m - \mathbf{W}_m \|_2 \leq \epsilon \right\} \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha^{\hat{\mathbf{W}}_m} - (\alpha - \alpha^{\hat{\mathbf{W}}_m}) \|_2^2 \right] + \mathbb{E} \left[ \mathbb{I} \left\{ \| \hat{\mathbf{W}}_m - \mathbf{W}_m \|_2 > \epsilon \right\} \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha^{\hat{\mathbf{W}}_m} \|_2^2 \right] - \frac{\epsilon^2}{2} \mathbb{E} \left[ \mathbb{I} \left\{ \| \hat{\mathbf{W}}_m - \mathbf{W}_m \|_2 \leq \epsilon \right\} \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha \|_2^2 \right] + \frac{\epsilon^2}{2} \mathbb{E} \left[ \mathbb{I} \left\{ \| \hat{\mathbf{W}}_m - \mathbf{W}_m \|_2 > \epsilon \right\} \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha \|_2^2 \right] - \frac{\epsilon^2}{2} \mathbb{E} \left[ \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha \|_2^2 \right] \geq \mathbb{E} \left[ \mathbb{I} \left\{ \| \hat{\mathbf{W}}_m - \mathbf{W}_m \|_2 > \epsilon \right\} \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha \|_2^2 \right] - \frac{\epsilon^2}{2} \mathbb{E} \left[ \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha \|_2^2 \right] + \frac{\epsilon^2}{2} \mathbb{E} \left[ \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha \|_2^2 \right]. \quad \tag{E}
\]

For any constant \( \mathbf{W}, \mathbf{W}' \in \mathbb{R}^{p \times p} \), we rewrite

\[
\mathbb{E} \left[ \| \hat{\alpha}_m^{\hat{\mathbf{W}}_m} - \alpha \|_2^2 \right] = \mathbb{E} \left[ \| \hat{\mathbf{W}}_m - \mathbf{W} \|_2^2 \right] + \mathbb{E} \left[ \| \mathbf{W} - \mathbf{W}' \|_2^2 \right] \leq 2 \| \mathbf{W} - \mathbf{W}' \|_2^2 \operatorname{Tr} \left( \mathbb{E} \left[ \hat{\alpha}_m^{\hat{\mathbf{W}}_m} \hat{\alpha}_m^{\hat{\mathbf{W}}_m} \right] \right) + 2 \| \mathbf{W} - \mathbf{W}' \|_2^2 \operatorname{Tr} \left( \mathbb{E} \left[ \hat{\alpha}_o^m \hat{\alpha}_o^m \right] \right) + 2 \| \mathbf{W} - \mathbf{W}' \|_2^2 \left( \| \mathbb{E} \left[ \hat{\alpha}_i^m \right] \|_2^2 + \operatorname{Tr} \left( \text{Cov} \left( \hat{\alpha}_i^m \right) \right) + \| \mathbb{E} \left[ \hat{\alpha}_o^m \right] \|_2^2 + \operatorname{Tr} \left( \text{Cov} \left( \hat{\alpha}_o^m \right) \right) \right),
\]

\(\text{We note that } \mathbf{W}_m \text{ may be random.}\)
where we have used Young’s inequality in the first step. We see that both \( |E[\tilde{\alpha}^m_{W}]|_2 \) and \( |E[\tilde{\alpha}^m_{O}]|_2 \) remain bounded \( \forall m \), while Tr \( (Cov(\tilde{\alpha}^m_{O})) \) and Tr \( (Cov(\tilde{\alpha}^m_{W})) \) decrease monotonically in \( m \). Hence, we conclude that for any \( \epsilon' > 0 \), there exists an \( \epsilon > 0 \) such that
\[
E \left[ |\tilde{\alpha}^m_{W} - \hat{\alpha}^m_{W} |_2^2 \right] \leq \epsilon', \ \forall m \in \mathbb{N} \text{ and } \forall W, W' \in \mathbb{R}^{p \times p} \text{ s.t. } |W - W'|_2 \leq \epsilon.
\] (F)

Since \( |W^m|_2 \leq c \) for all \( m \geq m' \), we have that \( |\tilde{\alpha}^m_{W} - \alpha|_2 \) is also bounded by some constant \( \epsilon' > 0 \), for all \( m \geq m' \), almost surely. We now fix an \( \epsilon' > 0 \) and choose a corresponding \( \epsilon \) such that (F) holds. We then conclude from (F) that
\[
\text{MSE} \left( \tilde{\alpha}^m_{Wm} \right) \geq E \left[ \mathbb{I} \left\{ |\tilde{\alpha}^m_{Wm} - \alpha|_2 \right\}, \forall m \geq m' \text{ and } \forall \alpha \right] - 2\epsilon \epsilon',
\] (G)

for all \( m \geq m' \). Thus, we conclude
\[
\lim_{m \to \infty} \text{MSE} \left( \tilde{\alpha}^m_{Wm} \right) \geq \lim_{m \to \infty} \text{MSE} \left( \tilde{\alpha}^m_{Wm} \right) - 2\epsilon \epsilon'.
\]

We can repeat this procedure for any \( \epsilon' > 0 \) and therefore conclude
\[
\lim_{m \to \infty} \text{MSE} \left( \tilde{\alpha}^m_{Wm} \right) \geq \lim_{m \to \infty} \text{MSE} \left( \tilde{\alpha}^m_{Wm} \right),
\]
which is the desired result. \( \Box \)

**Lemma B.2.** Let \( \tilde{W}^m - W^m \xrightarrow{a.s.} 0 \) and let there exist some \( c > 0, m' \in \mathbb{N} \), such that \( |W^m|_2 \leq c, \forall m \geq m' \), almost surely. Then, it holds that
\[
\lim_{m \to \infty} \text{MSE} \left( \tilde{\alpha}^m_{Wm} \right) \leq \lim_{m \to \infty} \text{MSE}(\tilde{\alpha}^m_{Wm}).
\]

**Proof.** We again employ the formulation from [D], but this time to construct an upper bound. For the first term of (D), we see that
\[
E \left[ \mathbb{I} \left\{ |\tilde{W}^m - W^m|_2 \leq \epsilon \right\}, \forall m \geq m' \text{ and } \forall \alpha \right] \leq \text{MSE} \left( \tilde{\alpha}^m_{Wm} \right) + 2 \sqrt{E \left[ \mathbb{I} \left\{ |\tilde{W}^m - W^m|_2 \leq \epsilon \right\}, \forall m \geq m' \text{ and } \forall \alpha \right]} \left[ \text{E} \left[ \tilde{\alpha}^m_{Wm} - \tilde{\alpha}^m_{Wm} + \tilde{\alpha}^m_{Wm} - \alpha \right] \right]
\]
by triangle inequality and the Cauchy-Schwarz inequality. Since for \( m \geq m' \) it holds that \( |W^m|_2 \leq c \), almost surely, there exists a constant \( \epsilon' > 0 \) such that \( E \left[ |\tilde{\alpha}^m_{Wm} - \alpha|_2 \right] \leq \epsilon' \), for all \( m \geq m' \). This is true because the two estimators \( \tilde{\alpha}^m_{Wm} \) and \( \tilde{\alpha}^m_{O} \) have both bounded mean squared error for any sample size \( m \).

Analogously to the proof for Lemma B.1, we now fix an \( \epsilon' > 0 \) and choose a corresponding \( \epsilon \) such that (F) holds. For \( m \geq m' \), we then conclude from (G) that
\[
E \left[ \mathbb{I} \left\{ |\tilde{W}^m - W^m|_2 \leq \epsilon \right\}, \forall m \geq m' \text{ and } \forall \alpha \right] \leq \text{MSE} \left( \tilde{\alpha}^m_{Wm} \right) + 2 \sqrt{\epsilon' E \left[ |\tilde{\alpha}^m_{Wm} - \alpha|_2 \right] + \epsilon'}.
\] (H)

This bounds the first term of (D). For the second term of (D), we use almost sure convergence of \( \tilde{W}^m - W^m \). Since \( W^m \) is bounded in the limit, almost surely, so is \( \tilde{W}^m \). Formally, \( |\tilde{W}^m|_2 \leq c'', \forall m \geq m' \) for some \( m' \in \mathbb{N} \), almost surely.
We use this to bound \[ \| \hat{\alpha}_{W,m}^m - \alpha \|_2 < c'' \] for all \( m \geq m' \), almost surely, for some \( c'' > 0 \). Now, we apply iterated expectations to the second term of (I) to see that for all \( m \geq m' \)

\[
E \left[ \mathbb{1}\left\{ \| \hat{W}^m - W^m \|_2 > \epsilon \right\} \| \hat{\alpha}_{W,m}^m - \alpha \|_2 \right] \leq E_{\hat{W}^m} \left[ \mathbb{1}\left\{ \| \hat{W}^m - W^m \|_2 > \epsilon \right\} \| \hat{\alpha}_{W,m}^m - \alpha \|_2 \right] \leq P \left( \| \hat{W}^m - W^m \|_2 > \epsilon \right) c'' ,
\]

almost surely. Now, we can combine the inequalities (I) and (II) to obtain

\[
\text{MSE} \left( \hat{\alpha}_{W,m}^m \right) \leq \text{MSE} \left( \hat{\alpha}_{W,m}^m \right) + 2\sqrt{\epsilon c''} + \epsilon' + P \left( \| \hat{W}^m - W^m \|_2 > \epsilon \right) c'' ,
\]

for all \( m \geq m'' \). Almost sure convergence implies consistency of \( \hat{W}^m - W^m \) with respect to 0, so we see that \( P \left( \| \hat{W}^m - W^m \|_2 > \epsilon \right) \) vanishes in the limit \( m \to \infty \), for all \( \epsilon > 0 \). We can repeat this procedure for any \( \epsilon' > 0 \). This implies the desired result. \( \square \)

C  DETAILED DERIVATION OF OPTIMAL WEIGHTING SCHEMES

In general, we observe that

\[
\text{Bias}(\hat{\alpha}_{W}^m) = W\alpha + (I - W)(\alpha + \Delta) - \alpha = (I - W)\Delta,
\]

\[
\text{Cov}(\hat{\alpha}_{W}^m) = WCov(\hat{\alpha}_{I}^m)W^T + (I - W)\text{Cov}(\hat{\alpha}_{0}^m)(I - W)^T.
\]

C.1  OPTIMAL SCALAR WEIGHT

Here, we have

\[
\frac{\partial}{\partial w} \text{MSE} \left( \hat{\alpha}_{wI}^m \right) = \frac{\partial}{\partial w} \left[ \text{Bias} \left( \hat{\alpha}_{wI}^m \right) \right]_2 + \frac{\partial}{\partial w} \text{Tr} \left( \text{Cov} \left( \hat{\alpha}_{wI}^m \right) \right) = -2(1 - w)\| \Delta \|_2^2 + 2w\text{Tr} \left( \text{Cov}(\hat{\alpha}_{I}^m) \right) - 2(1 - w)\text{Tr} \left( \text{Cov}(\hat{\alpha}_{0}^m) \right) = 0.
\]

By rearranging, we get

\[
w_m^* = \frac{\text{Tr} \left( \text{Cov}(\hat{\alpha}_{0}^m) \right) + \| \Delta \|_2^2}{\text{Tr} \left( \text{Cov}(\hat{\alpha}_{I}^m) \right) + \text{Tr} \left( \text{Cov}(\hat{\alpha}_{0}^m) \right) + \| \Delta \|_2^2}.
\]

C.2  OPTIMAL DIAGONAL WEIGHT MATRIX

Here, we see that the objective decouples into a sum over the individual dimensions

\[
\text{MSE} \left( \hat{\alpha}_{wI}^m \right) = \sum_{k=1}^{p} \left( 1 - w^{(k)} \right)^2 \Delta^{(k)} + w^{(k)} \text{Cov}^{(k,k)}(\hat{\alpha}_{I}^m) + \left( 1 - w^{(k)} \right)^2 \text{Cov}^{(k,k)}(\hat{\alpha}_{0}^m),
\]

Thus, we optimize for each dimension \( k \) separately and obtain

\[
w_m^{(k)} = \frac{\text{Cov}^{(k,k)}(\hat{\alpha}_{I}^m) + \Delta^{(k)} + \text{Cov}^{(k,k)}(\hat{\alpha}_{0}^m) + \Delta^{(k)}^2}{\text{Cov}^{(k,k)}(\hat{\alpha}_{I}^m) + \text{Cov}^{(k,k)}(\hat{\alpha}_{0}^m) + \Delta^{(k)}^2}.
\]
C.3 OPTIMAL WEIGHT MATRIX

Using \( \frac{\partial}{\partial \mathbf{W}} \text{Tr}(\mathbf{W}A\mathbf{W}^\top) = 2\mathbf{W} \) since \( \mathbf{A} \) is symmetric, we observe that
\[
\frac{\partial}{\partial \mathbf{W}} \text{MSE} \left( \mathbf{\alpha}_{\mathbf{W}}^{\mathbf{m}} \right) = 2 \mathbf{W} \left( \text{Cov} \left( \mathbf{\alpha}_{\mathbf{W}}^{\mathbf{m}} \right) + \text{Cov}(\mathbf{\alpha}_{\mathbf{W}}^{\mathbf{0}}) + \mathbf{\Delta}^\top \right) - 2 \left( \mathbf{\Delta}^\top + \text{Cov}(\mathbf{\alpha}_{\mathbf{W}}^{\mathbf{0}}) \right)
\]
\[
\seteq{0}
\]
We see that this minimum is attained for
\[
\left( \text{Cov} \left( \mathbf{\alpha}_{\mathbf{W}}^{\mathbf{0}} \right) + \mathbf{\Delta}^\top \right) \left( \text{Cov} \left( \mathbf{\alpha}_{\mathbf{W}}^{\mathbf{m}} \right) + \text{Cov}(\mathbf{\alpha}_{\mathbf{W}}^{\mathbf{0}}) + \mathbf{\Delta}^\top \right)^{-1}.
\]

D NON ZERO-MEAN EXOGENOUS VARIABLES

All results established here can readily be extended to settings, where any of the exogenous variables have non-zero mean, i.e., \( \mu_{\mathbf{N}_X}, \mu_{\tilde{\mathbf{N}}_X}, \mu_{\mathbf{N}_Y} \) (see (1)–(3)) may be non-zero. In order to extend the practical estimators introduced here, one needs to consider the following two pre-processing steps:

First, we center both treatment distributions separately, without scaling:
\[
x_i' \leftarrow x_i - n^{-1} \sum_{j=1}^{n} x_j, \quad \forall i \in 1, \ldots, n, \quad (J)
\]
\[
x_i' \leftarrow x_i - m^{-1} \sum_{j=n+1}^{n+m} x_j, \quad \forall i \in n+1, \ldots, n+m. \quad (K)
\]
In this manner, both treatment variables become zero-mean.

Furthermore, we add a dummy dimension with value one to all treatment vectors:
\[
x_i'' \leftarrow (x_i', 1), \quad \forall i \in 1, \ldots, n+m.
\]
This naturally adds one more dimension also to \( \mathbf{\alpha} \), which corresponds to the intercept term. We then use the constructed \( x_i'' \) to compute the weight matrices proposed in this work.

Finally, we see that the intercept term must be identical for both distributions, interventional and observational:
\[
\mathbb{E}[Y | \mathbf{X}' = \mathbf{x}'] = \gamma^\top \mathbb{E}[\mathbf{Z} | \mathbf{X}' = \mathbf{x}'] + \mathbf{\alpha}^\top \mathbf{x}' + \mu_{\mathbf{N}_Y}.
\]
We then have in the observational setting (data points 1, ..., \( n \)) that
\[
\gamma^\top \mathbb{E}[\mathbf{Z} | \mathbf{X}' = \mathbf{x}'] = \gamma^\top \mu_{\mathbf{N}_Z} + \gamma^\top \Sigma_{\mathbf{N}_Z} \mathbf{B}^\top (\Sigma_{\mathbf{N}_X} + \mathbf{B} \Sigma_{\mathbf{N}_Z} \mathbf{B}^\top)^{-1} (\mathbf{x}' - \mathbb{E}[\mathbf{X}'])
\]
\[
= \gamma^\top \mu_{\mathbf{N}_Z} + \mathbf{\Delta}^\top \mathbf{x}',
\]
where \( \mathbb{E}[\mathbf{X}'] = 0 \) due to (J).

For the interventional data, we have independence between \( \mathbf{X}' \) and \( \mathbf{Z} \) by definition and so we trivially get
\[
\gamma^\top \mathbb{E}[\mathbf{Z} | \mathbf{X}' = \mathbf{x}'] = \gamma^\top \mu_{\mathbf{N}_Z}
\]
here. Thus, the intercept is \( \gamma^\top \mu_{\mathbf{N}_Z} + \mu_{\mathbf{N}_Y} \) for both distributions and we fix \( \hat{\mathbf{\Delta}}^{p+1} = 0. \)
We see that the ground truth covariance matrices of $\hat{\alpha}_i^m$ and $\hat{\alpha}_o^m$ adapt to changes in the sample sizes, keeping the distributions of all variables fixed. For instance, we see that

$$\text{Cov}(\hat{\alpha}_i^m) = (X_i^\top X_i)^{-1}\sigma_Y^2|_{\text{do}(X)} = m^{-1}(m^{-1}X_i^\top X_i)^{-1}\sigma_Y^2|_{\text{do}(X)}.$$ 

The term $(m^{-1}X_i^\top X_i)^{-1}\sigma_Y^2$ is bounded in probability, for large enough $m$. Accordingly, this implies that $\text{Cov}(\hat{\alpha}_i^m) \xrightarrow{P} 0$. Thus, when keeping $n$ fixed, we obtain $W_i^m \xrightarrow{P} I_p$, for $m \to \infty$.

On the other hand, if we keep $m$ fixed and consider the limit $n \to \infty$ instead, we observe that

$$W_i^m \xrightarrow{P} \Delta^\top (\text{Cov}(\hat{\alpha}_i^m) + \Delta^\top)^{-1}.$$ 

We note that we do not have $W_i^m \xrightarrow{P} 0$ here in general, because the bias in $\hat{\alpha}_o^m$ remains, independent of the sample size $n$. 