A Central Limit Theorem for Regularized M-Estimators

Abstract

We prove a quantitative central limit theorem for linear functionals of regularized empirical-risk minimizers in the proportional-dimensional regime $p = O(n)$. The data columns are independent, not necessarily identically distributed, and satisfy a uniform columnwise Poincaré inequality. Under uniform curvature and smoothness assumptions, and for a quadratic regularizer, we show that every nondegenerate statistic $\sqrt{n}\, u^\top \hat{\theta}$, centered by its expectation and normalized by its standard deviation, converges to a standard normal random variable in Wasserstein distance, with rate $O((\log n)^7 n^{-1/4})$. The proof is based on moment and stability bounds for the minimizer, a second-order leave-one-out expansion, and a perturbative normal-approximation argument for functions of independent variables. We also prove the variance upper bound $\mathrm{Var}(u^\top \hat{\theta}) \le C\, |u|_2^2 / n$, identifying the $\sqrt{n}$ fluctuation scale.

1 Introduction

We consider a regularized empirical-risk minimizer

$$\hat{\theta} = \mathrm{argmin}_{\theta \in \mathbb{R}^p} \left\{ \frac{1}{n} \sum_{i=1}^{n} L(x_i^\top \theta) + \rho(\theta) \right\}.$$

Here $X = (x_1, \ldots, x_n) \in \mathbb{R}^{p \times n}$ has independent columns, the loss $L : \mathbb{R} \to \mathbb{R}$ is smooth and uniformly convex, and the regularizer $\rho : \mathbb{R}^p \to \mathbb{R}$ is strongly convex. Problems of this form are central in classical M-estimation, robust statistics, and convex learning theory. In fixed dimension, consistency and asymptotic normality for smooth M-estimators are classical; see Huber [11] and van der Vaart [20].

When the dimension grows with the sample size, the fixed-dimensional theory no longer applies directly. Classical increasing-dimension normal approximation results, such as those of Portnoy [16], Mammen [14], and He–Shao [10], require the dimension to grow sufficiently slowly compared with the sample size. In the proportional regime, where p/n is of constant order, central limit theorems and exact-asymptotic descriptions have been obtained in more structured settings. Donoho–Montanari [6], Sur–Candès [19], Zhao–Sur–Candès [21], and Bellec–Shen–Zhang [3] work under Gaussian-design assumptions, possibly with general covariance. Lei–Bickel–El Karoui [13] prove coordinate-wise normality under fixed-design regularity conditions. El Karoui and collaborators [8, 7] allow non-Gaussian designs, but the assumptions still impose iid, isotropic, or independent-entry structure together with strong concentration properties.

The present paper proves a quantitative central limit theorem under weaker distributional assumptions on the design. We work in the regime $p = O(n)$, and the columns x_i are independent but need not be identically distributed, Gaussian, centered, isotropic, or have iid entries. The main concentration assumption is instead a uniform columnwise Poincaré inequality. Under the smoothness and curvature assumptions stated below, and for a quadratic regularizer, we prove a Wasserstein CLT for arbitrary deterministic linear projections of the minimizer. More precisely, for deterministic directions $u = u_n \in \mathbb{R}^p$ with $|u|_2 \le 1$, set

$$\sigma_n^2 := \mathrm{Var}\big(\sqrt{n}\, u^\top \hat{\theta}\big).$$

If $\inf_n \sigma_n^2 > 0$, then

$$\frac{\sqrt{n}\, u^\top \hat{\theta} - \mathbb{E}[\sqrt{n}\, u^\top \hat{\theta}]}{\sigma_n} \Rightarrow \mathcal{N}(0,1)$$

quantitatively in Wasserstein distance, with rate $O((\log n)^7 n^{-1/4})$. We also prove the variance upper bound

$$\mathrm{Var}(u^\top \hat{\theta}) \lesssim \frac{|u|_2^2}{n},$$

so that the only variance assumption in the CLT is the matching lower bound.

The proof combines leave-one-out expansions with the perturbative normal-approximation method of Chatterjee [5]; see also the quantitative Berry–Esseen theory for nonlinear statistics and M-estimators of Shao–Zhang [17]. The key step is a second-order leave-one-out approximation for the increments of $\sqrt{n}\, u^\top \hat{\theta}$. This expansion uses an averaged-leverage score to remove the leading dependence of the active score on the active column. Once this approximation is established, moment estimates and conditional-covariance bounds verify a perturbative Wasserstein criterion for asymptotic normality.

Contents

2 Setting, assumptions, and main result

2.1 Notation

We use the following notation. For a measurable function $f : \mathbb{R}^p \to \mathbb{R}$, a finite-dimensional vector $a = (a_i)$, and $A \in \mathcal{M}_{p,n}$, set

- $|f|_\infty := \sup_{x \in \mathbb{R}^p} |f(x)|$;

- $|a|_2 := \big(\sum_i a_i^2\big)^{1/2}$ and $|a|_\infty := \sup_i |a_i|$;
- $|A|_{\mathrm{op}} := \sup_{|x|_2 \le 1} |Ax|_2$ and $|A|_F := \sqrt{\mathrm{Tr}(AA^\top)}$.

Given $f \in \mathcal{C}^2(\mathbb{R}^p)$ and $z \in \mathbb{R}^p$,

$$\nabla f(z) = \Big(\tfrac{\partial f(z)}{\partial z_1}, \dots, \tfrac{\partial f(z)}{\partial z_p}\Big)^\top, \qquad \nabla^2 f(z) = \left(\frac{\partial^2 f(z)}{\partial z_a \partial z_b}\right)_{a,b=1}^p.$$

For ℓ indexing a column of $Z = (z_1, \dots, z_n) \in \mathcal{M}_{p,n}$, we write $\mathbb{D}_\ell$ for the Fréchet derivative with respect to z_ℓ. If $T = T(Z)$ is real-, vector-, or matrix-valued, then $\mathbb{D}_\ell[T(Z)][h]$ denotes the derivative in the direction $h \in \mathbb{R}^p$. For instance, when T is vector-valued,

$$\Big|T(Z + he_\ell^\top) - T(Z) - \mathbb{D}_\ell[T(Z)][h]\Big|_2 = o(|h|_2).$$

When T is matrix-valued, we write

$$|\mathbb{D}_\ell[T]|_{\mathrm{op}}^* := \sup_{|h|_2 \le 1} |\mathbb{D}_\ell[T][h]|_{\mathrm{op}},$$

with analogous notation for the other norms and for real- or vector-valued maps. For example, given $u \in \mathbb{R}^p$,

$$\Big|u^\top \mathbb{D}_\ell[T]\Big|_2^* := \sup_{|h|_2 \le 1} \Big|u^\top \mathbb{D}_\ell[T][h]\Big|_2.$$

We will keep the $\|\cdot\|$ notation for the integrated norms defined for any $k > 0$ and any random variable $Z \in \mathbb{R}$ as:

$$\|Z\|_{L^k} := \mathbb{E}[|Z|^k]^{\frac{1}{k}}.$$

Given a statistic $g((z_1, \dots, z_n))$ depending on n independent variables $z_1, \dots, z_n$, we denote:

$$\|g((z_1, \dots, z_n))\|_{L_i^k} := \mathbb{E}[|g((z_1, \dots, z_n))|^k \mid (z_j)_{j \in [n] \setminus \{i\}}]^{\frac{1}{k}}$$

Similarly, $\mathbb{E}_i$, Var_i, and Cov_i denote conditional expectation, variance, and covariance with respect to z_i only, conditionally on all other variables.

For two nonnegative sequences (a_n) and (b_n), we write $a_n = O(b_n)$ if there exists $C > 0$ such that $a_n \le C b_n$ for all large enough n, and $a_n = o(b_n)$ if $a_n / b_n \to 0$. All implicit constants are deterministic and independent of n and p, unless explicitly stated otherwise.

2.2 Setting

Let $X = (x_1, \dots, x_n) \in \mathbb{R}^{p \times n}$ have independent columns $x_1, \dots, x_n \in \mathbb{R}^p$. For $\theta \in \mathbb{R}^p$, set

$$\Psi_X(\theta) := \frac{1}{n} \sum_{i=1}^n L(x_i^\top \theta) + \rho(\theta), \qquad \hat{\theta} := \mathrm{argmin}_{\theta \in \mathbb{R}^p} \Psi_X(\theta).$$

The assumptions below imply that Ψ_X is strongly convex, so this minimizer is unique. We view $(p, X, L, \rho) = (p_n, X_n, L_n, \rho_n)$ as a sequence indexed by n, and we suppress the dependence on n in the notation.

Assumption 1. $p = O(n)$.

Assumption 2. $|L'(0)| = O(1)$ and $|\nabla \rho(0)|_2 = O(1)$.

Assumption 3. $\sup_{i \in [n]} |\mathbb{E}[x_i]|_2 = O(1)$.

Assumption 4 (Poincaré inequality for the columns)**.** There exists $C_P = O(1)$ such that, for every $i \in [n]$ and every C^1 function $f : \mathbb{R}^p \to \mathbb{R}$ for which the right-hand side is finite,

$$\mathrm{Var}(f(x_i)) \le C_P \, \mathbb{E} \left| \nabla f(x_i) \right|^2 .$$

Assumption 5. The loss $L : \mathbb{R} \to \mathbb{R}$ is C^2, $|L''|_\infty = O(1)$, and there exists a constant $\kappa_L > 0$, with $\kappa_L^{-1} = O(1)$, such that, for all $t \in \mathbb{R}$,

$$L''(t) \ge \kappa_L.$$

The regularizer $\rho : \mathbb{R}^p \to \mathbb{R}$ is C^2 and κ-strongly convex for some constant $\kappa > 0$, with $\kappa^{-1} = O(1)$, namely

$$\nabla^2 \rho(\theta) \succeq \kappa I_p, \qquad \theta \in \mathbb{R}^p.$$

These assumptions ensure well-posedness and provide the basic moment and concentration bounds for $\hat{\theta}$. To derive a first-order expansion of the leave-one-out perturbation $\hat{\theta} - \hat{\theta}_{-i}$, we also require the following smoothness condition.

Assumption 6. The second derivative L'' is globally Lipschitz with constant $C_L = O(1)$:

$$|L''(t) - L''(t')| \le C_L |t - t'|, \qquad t, t' \in \mathbb{R}.$$

Moreover, $\nabla^2 \rho$ is globally Lipschitz in operator norm with constant $C_\rho = O(1)$:

$$\left| \nabla^2 \rho(\theta) - \nabla^2 \rho(\theta') \right|_{\mathrm{op}} \le C_\rho \left| \theta - \theta' \right|_2 , \qquad \theta, \theta' \in \mathbb{R}^p.$$

Finally, the CLT proof uses a second-order expansion of the leave-one-out perturbation. For this step we impose the following additional curvature condition.

Assumption 7 (Third-order curvature)**.** The loss L is C^4 with[1]:

$$|L'''|_\infty = O(1), \qquad \text{and} \qquad |L''''|_\infty = O(1).$$

The regularizer ρ is C^3. We denote its third-order differential tensor by $\nabla^3 \rho(\theta)$, so that, for any $h \in \mathbb{R}^p$, $\nabla^3 \rho(\theta)[h]$ is a symmetric matrix. There exists a constant $\gamma_\rho = O(1)$ such that,

$$\sup_{\theta \in \mathbb{R}^p} \sup_{|h|_2 \le 1} \left| \nabla^3 \rho(\theta)[h] \right|_{\mathrm{op}} \le O(1),$$

and for all $\theta, \theta', h \in \mathbb{R}^p$:

$$\left| \nabla^3 \rho(\theta)[h] - \nabla^3 \rho(\theta')[h] \right|_{\mathrm{op}} \le \gamma_\rho \left| \theta - \theta' \right| \, |h|_2 .$$

Fix a deterministic sequence of directions $u = u_n \in \mathbb{R}^p$ such that

$$|u|_2 \le 1.$$

Define

$$f_n(X) := \sqrt{n} \, u^\top \hat{\theta}(X), \qquad \sigma_n^2 := \mathrm{Var}\big(f_n(X)\big). \tag{2.1}$$

Whenever $\sigma_n > 0$, set

$$W_n := \frac{f_n(X) - \mathbb{E} f_n(X)}{\sigma_n}.$$

[1]The bound $|L''''|_\infty = O(1)$ will actually only be used in Section 5 but to stay simple we did not isolate it in an independent assumption.

2.3 Main result

Theorem 2.1 (CLT for linear functionals of the minimizer)**.** *Under Assumptions 1–7, assume in addition that ρ is quadratic, in the sense that*

$$\nabla^3 \rho \equiv 0, \qquad \text{and} \qquad \inf_n \sigma_n^2 > 0.$$

Then

$$d_W(W_n, \mathcal{N}) = O\left(\frac{(\log n)^8}{n^{1/4}}\right),$$

where $\mathcal{N}$ denotes a standard normal random variable and

$$d_W(Y, Z) := \sup_{|h|_{\mathrm{Lip}} \leq 1} |\mathbb{E}h(Y) - \mathbb{E}h(Z)|$$

is the Wasserstein distance.

The proof is based on Chatterjee's normal-approximation method [5] and on the simplified Chatterjee-type Wasserstein bound of Shao and Zhang [18], recalled in Appendix A as Theorem A.1. We use the following leave-one-out reformulation.

Let $Z = (z_1, \dots, z_n)$ have independent coordinates, and let $Z' = (z_1', \dots, z_n')$ be an independent copy of Z. For $A \subset [n]$, let Z^A be obtained from Z by replacing z_i by z_i' for every $i \in A$, and set

$$A_i := \{1, \dots, i-1\}.$$

Let $g_n = g_n(Z)$ be a real-valued statistic. For each $i \in [n]$, let g_n^{-i} be a leave-one-out statistic depending on all coordinates except z_i. We define

$$\delta_i g_n(Z^A) := g_n(Z^A) - g_n^{-i}(Z^A_{-i}),$$

where Z^A_{-i} denotes Z^A with its ith coordinate removed.

Theorem 2.2 (Leave-one-out Wasserstein bound)**.** *Assume that $z_1, \dots, z_n$ are independent. No identical-distribution or exchangeability assumption is imposed. Define*

- $\sigma_{g,n}^2 := \mathrm{Var}(g_n(Z))$,
- $m_n^{(4)} := \sup_{i \in [n]} \|\delta_i g_n(Z)\|_{L^4}$,
- $c_n^z := \sup_{\substack{i,j \in [n] \\ i \neq j}} \left\|\delta_i g_n(Z) - \mathbb{E}_j\left[\delta_i g_n(Z)\right]\right\|_{L^4}$,
- $c_n^{\bar{z}} := \sup_{i \in [n]} \sqrt{\mathrm{Var}(\mathrm{Cov}_i(\delta_i g_n(Z), \delta_i g_n(Z^{A_i})))}$.

Assume that these quantities are finite and that $\sigma_{g,n}^2 > 0$. Let

$$W_{g,n} := \frac{g_n(Z) - \mathbb{E}g_n(Z)}{\sigma_{g,n}}.$$

Then, for a standard normal random variable $\mathcal{N}$,

$$d_W(W_{g,n}, \mathcal{N}) \leq \frac{C}{\sigma_{g,n}^2}\left[n(m_n^{(4)})^4 + n(n-1)\left((m_n^{(4)})^3 c_n^z + (m_n^{(4)})^2 c_n^{\bar{z}}\right)\right]^{1/2} + \frac{8n(m_n^{(4)})^3}{\sigma_{g,n}^3},$$

where $C > 0$ is a universal constant.

With the notation of (2.1), we apply Theorem 2.2 with $Z = X$ and $g_n = f_n$. The corresponding leave-one-out statistic is

$$f_n^{-i}(X_{-i}) := \sqrt{n}\, u^\top \hat{\theta}_{-i},$$

where

$$\hat{\theta}_{-i} := \operatorname{argmin}_{\theta \in \mathbb{R}^p} \Psi_{X_{-i}}(\theta), \qquad \Psi_{X_{-i}}(\theta) := \frac{1}{n} \sum_{j \neq i} L(x_j^\top \theta) + \rho(\theta). \tag{2.2}$$

The normalization is kept equal to $1/n$, so that $\Psi_{X_{-i}}$ is the original objective with the ith loss term removed.

To prove Theorem 2.1, it remains to establish the following estimates:

1. $\sigma_n = O(1)$ in Section 3, Remark 3.16;
2. $m_n^{(4)} = O\left(\frac{\log n}{\sqrt{n}}\right)$ in Subection 4.1, Remark 4.11;
3. $c_n^x = O\left(\frac{(\log n)^7}{n}\right)$ in Subsection 5.2, Proposition 5.13
4. $c_n^{\bar{x}} = O\left(\frac{(\log n)^{13}}{n^{3/2}}\right)$ in Subsection 5.3, Corollary 5.22

Together with the nondegeneracy assumption $\inf_n \sigma_n^2 > 0$, these bounds imply the Wasserstein rate stated in Theorem 2.1.

The argument proceeds in three main stages. First, we prove moment, stability, and variance bounds for the minimizer and its leave-one-out analogues. Second, we derive first- and second-order leave-one-out expansions for $u^\top(\hat{\theta} - \hat{\theta}_{-i})$. Third, we insert these expansions into the perturbative normal-approximation bound and verify the required sensitivity and covariance estimates.

3 Preliminary concentration properties

3.1 Basic concentration results on the data

We first collect the concentration estimates for the data and the minimizer that will be used throughout the proof. The general Poincaré facts are recalled in Appendix B; here we only record their consequences for the independent columns $x_1, \dots, x_n$.

Lemma 3.1 (Tensorized Poincaré inequality for the data)**.** *Under Assumption 4, for every sufficiently smooth $f : \mathcal{M}_{p,n} \to \mathbb{R}$,*

$$\operatorname{Var}(f(X)) \le C_P \sum_{i=1}^n \| \, |\mathbb{D}_i[f(X)]|_2^* \, \|_{L^2}^2.$$

Proof. This is Proposition B.1 applied to the independent columns $x_1, \dots, x_n$. □

The next lemma gives the data moment bounds used below. It is a direct consequence of Lemma B.3, Lemma B.6, and the boundedness of the column means.

Lemma 3.2 (Column moment bounds)**.** *Under Assumptions 3 and 4, for any sequence of positive integers $k_n \ge 1$, and any deterministic sequence $u = u_n \in \mathbb{R}^p$ satisfying $|u| \le 1$,*

$$\sup_{i \in [n]} \left\| u^\top x_i \right\|_{L^{k_n}} = O(k_n).$$

If, in addition, Assumption 1 holds, then

$$\sup_{i \in [n]} \| |x_i|_2 \|_{L^{k_n}} = O(k_n \sqrt{n}).$$

Proof. Lemma B.3 gives

$$\sup_{i\in[n]} \left\| u^\top (x_i - \mathbb{E}x_i) \right\|_{L^{k_n}} = O(k_n), \qquad |u| \le 1.$$

Since

$$|u^\top \mathbb{E}x_i| \le |\mathbb{E}x_i|_2 = O(1)$$

uniformly in i by Assumption 3, the first bound follows by the triangle inequality. The second bound follows from Lemma B.6 and $p = O(n)$. □

Combining Lemma 3.2 with Lemma B.7 also gives the logarithmic control of columnwise maxima that will be used later. For instance, if $(u_i)_{i\in[n]}$ is deterministic and $|u_i| \le 1$, then, for every $k \ge 1$,

$$\left\| \max_{i\in[n]} |u_i^\top x_i| \right\|_{L^k} = O\big(\max(k, \log(en))\big).$$

Lemma 3.3 (Conditional quadratic-form bound)**.** *Let*

$$\Sigma_i := \mathbb{E}[x_i x_i^\top]$$

denote the second-moment matrix of x_i. Under Assumptions 3 and 4, for every fixed $k \ge 1$, every deterministic symmetric matrix A, and every $i \in [n]$,

$$\left\| x_i^\top A x_i - \mathrm{Tr}(\Sigma_i A) \right\|_{L_i^k} \le O_k(|A|_F).$$

The same bound holds conditionally whenever A is independent of x_i.

Proof. It is enough to prove the result for deterministic A, since the conditional version follows by freezing the variables on which A depends. Write $x_i = m_i + y_i$, where $m_i = \mathbb{E}x_i$ and $\mathbb{E}y_i = 0$. Then

$$x_i^\top A x_i - \mathrm{Tr}(\Sigma_i A) = \big(y_i^\top A y_i - \mathbb{E}[y_i^\top A y_i]\big) + 2m_i^\top A y_i.$$

The linear term satisfies

$$\left\| m_i^\top A y_i \right\|_{L^k} \le O_k(|Am_i|_2) \le O_k(|A|_{\mathrm{op}}) \le O_k(|A|_F),$$

by Lemma B.3 and Assumption 3. For the centered quadratic term, we use the standard L^k-form of the Poincaré inequality, obtained by applying the Poincaré inequality to powers of the centered function:

$$\|F - \mathbb{E}F\|_{L^k} \le O_k\left(\| |\nabla F|_2 \|_{L^k}\right).$$

Applying this to $F(y) = y^\top A y$, whose gradient is $2Ay$, gives

$$\left\| y_i^\top A y_i - \mathbb{E}[y_i^\top A y_i] \right\|_{L^k} \le O_k\left(\| |Ay_i|_2 \|_{L^k}\right).$$

If $A = \sum_a \sigma_a v_a w_a^\top$ is a singular-value decomposition, then Minkowski's inequality and Lemma B.3 yield

$$\| |Ay_i|_2 \|_{L^k} \le \left(\sum_a \sigma_a^2 \left\| w_a^\top y_i \right\|_{L^k}^2 \right)^{1/2} \le O_k(|A|_F).$$

Combining the two estimates proves the lemma. □

Let us finally mention, as a direct consequence of Lemma B.9 that will only be used once to set Lemma 4.12:

Lemma 3.4. *Under Assumptions 1,3 and 4:*

$$\| |X|_{\mathrm{op}} \| = O(\sqrt{n}).$$

3.2 Control of the minimizer and its leave-one-out displacement

We start with a deterministic stability estimate for strongly convex functions.

Lemma 3.5 (Stability of minimizers)**.** *Let $\phi : \mathbb{R}^p \to \mathbb{R}$ be differentiable and κ-strongly convex, and let*

$$\mu_\phi := \mathrm{argmin}_{x\in\mathbb{R}^p}\, \phi(x).$$

Then, for every $x \in \mathbb{R}^p$,

$$|\mu_\phi - x|_2 \le \frac{1}{\kappa}\, |\nabla\phi(x)|_2\,.$$

Proof. Strong convexity implies the κ-strong monotonicity of $\nabla\phi$:

$$\langle \nabla\phi(x) - \nabla\phi(y), x-y\rangle \ge \kappa\, |x-y|_2^2\,.$$

Taking $y = \mu_\phi$ and using $\nabla\phi(\mu_\phi) = 0$, we obtain

$$\kappa\, |x-\mu_\phi|^2 \le \langle \nabla\phi(x), x-\mu_\phi\rangle \le |\nabla\phi(x)|_2\, |x-\mu_\phi|_2\,.$$

The claim follows by canceling one factor $|x-\mu_\phi|_2$, the case $x = \mu_\phi$ being immediate. □

Lemma 3.6 (Empirical gradient at the origin)**.** *Under Assumptions 1–4, for any sequence $k = (k_n)_{n\in\mathbb{N}}$,*

$$\| |\nabla\Psi_X(0)|_2 \|_{L^{k_n}} = O(k_n).$$

Proof. We have

$$\nabla\Psi_X(0) = \frac{L'(0)}{n}\sum_{i=1}^n x_i + \nabla\rho(0).$$

Thus

$$\| |\nabla\Psi_X(0)|_2 \|_{L^{k_n}} \le |\nabla\rho(0)|_2 + |L'(0)| \left| \frac{1}{n}\sum_{i=1}^n \mathbb{E}x_i \right|_2 + |L'(0)| \left\| \left| \frac{1}{n}\sum_{i=1}^n (x_i - \mathbb{E}x_i) \right|_2 \right\|_{L^{k_n}}.$$

The first two terms are $O(1)$ by Assumptions 2 and 3.

Set

$$S_n := \frac{1}{n}\sum_{i=1}^n (x_i - \mathbb{E}x_i).$$

For each deterministic $u \in \mathbb{R}^p$ with $|u| \le 1$, the map

$$X \mapsto u^\top S_n$$

is a centered linear functional of X with Euclidean coefficient norm at most $n^{-1/2}$. By Lemma 3.1 and Lemma B.3,

$$\sup_{|u|_2\le 1} \left\| u^\top S_n \right\|_{L^{k_n}} = O(k_n n^{-1/2}).$$

Lemma B.6 and $p = O(n)$ then give

$$\| |S_n|_2 \|_{L^{k_n}} \le \sqrt{p}\, \sup_{|u|_2\le 1} \left\| u^\top S_n \right\|_{L^{k_n}} = O(k_n).$$

Since $|L'(0)| = O(1)$, the claim follows. □

Theorem 3.7 (Moment bound for the minimizer)**.** *Under Assumptions 1–5, for every sequence of positive integers $k_n \ge 1$*

$$\left\| |\hat\theta|_2 \right\|_{L^{k_n}} + \sup_{i\in[n]} \left\| \left| \hat\theta_{-i} \right| \right\|_{L^{k_n}} = O(k_n).$$

Proof. By Assumption 5, the objective Ψ_X is κ-strongly convex. Lemma 3.5, applied to $\phi = \Psi_X$ and $x = 0$, yields

$$|\hat{\theta}|_2 \leq \kappa^{-1} |\nabla \Psi_X(0)|_2 .$$

Taking L^{k_n}-norms and applying Lemma 3.6 proves the claim. The bound for $\hat{\theta}_{-i}$ is obtained in the same way, applied to

$$\nabla \Psi_{X_{-i}}(0) = \frac{L'(0)}{n} \sum_{j \neq i} x_j + \nabla \rho(0),$$

which satisfies the same bound as $\nabla \Psi_X(0)$, uniformly in i. □

We next control the leave-one-out displacement

$$\delta_i \hat{\theta} := \hat{\theta} - \hat{\theta}_{-i}, \qquad i \in [n].$$

We require two preliminary results. For $i \in [n]$, define the test score

$$s_i^- := L'(x_i^\top \hat{\theta}_{-i}).$$

Lemma 3.8 (Test score control)**.** *Under Assumptions 1–5, for every sequence of positive integers* k_n*,*

$$\sup_{i \in [n]} \left\| s_i^- \right\|_{L^{k_n}} = O(k_n^2).$$

Proof. Conditionally on X_{-i}, the vector $\hat{\theta}_{-i}$ is deterministic and independent of x_i. Lemma 3.2 gives

$$\left\| x_i^\top \hat{\theta}_{-i} \right\|_{L_i^{k_n}} \leq O(k_n) |\hat{\theta}_{-i}|_2.$$

Taking the L^{k_n}-norm over X_{-i} and using Theorem 3.7, we obtain

$$\left\| x_i^\top \hat{\theta}_{-i} \right\|_{L^{k_n}} = O(k_n^2).$$

The claim then follows from Assumptions 2 and 5 that provide $|L'(t)| \leq |L'(0)| + |L''|_\infty |t| \leq O(1)(1 + |t|)$. □

Lemma 3.9 (Strong monotonicity with prediction curvature)**.** *Under Assumption 5, for every* $t \in \mathbb{R}^p$*,*

$$\frac{\kappa_L}{n} \left| X^\top (\hat{\theta} - t) \right|_2^2 + \kappa \left| \hat{\theta} - t \right|_2^2 \leq -\nabla \Psi_X(t)^\top (\hat{\theta} - t).$$

Proof. For all $\theta, \theta' \in \mathbb{R}^p$,

$$\begin{aligned} &\left\langle \nabla \Psi_X(\theta) - \nabla \Psi_X(\theta'), \theta - \theta' \right\rangle \\ &\quad = \frac{1}{n} \sum_{j=1}^n \left(L'(x_j^\top \theta) - L'(x_j^\top \theta') \right) x_j^\top (\theta - \theta') + \left\langle \nabla \rho(\theta) - \nabla \rho(\theta'), \theta - \theta' \right\rangle. \end{aligned}$$

Since $L'' \geq \kappa_L$,

$$\left(L'(x_j^\top \theta) - L'(x_j^\top \theta') \right) x_j^\top (\theta - \theta') \geq \kappa_L \left(x_j^\top (\theta - \theta') \right)^2.$$

Since ρ is κ-strongly convex,

$$\left\langle \nabla \rho(\theta) - \nabla \rho(\theta'), \theta - \theta' \right\rangle \geq \kappa \left| \theta - \theta' \right|_2^2.$$

As a consequence, for all $\theta, \theta' \in \mathbb{R}^p$,

$$\left\langle \nabla \Psi_X(\theta) - \nabla \Psi_X(\theta'), \theta - \theta' \right\rangle \geq \frac{\kappa_L}{n} \left| X^\top (\theta - \theta') \right|_2^2 + \kappa \left| \theta - \theta' \right|_2^2.$$

Taking $\theta = \hat{\theta}$, $\theta' = t$, and using $\nabla \Psi_X(\hat{\theta}) = 0$, gives the result. □

Lemma 3.10 (Prediction stability of the leave-one-out displacement)**.** *Under Assumptions 1–5, for every sequence of positive integers k_n,*

$$\sup_{i\in[n]} \left\| \left|\delta_i\hat{\theta}\right|_2 \right\|_{L^{k_n}} = O\left(\frac{k_n^2}{\sqrt{n}}\right),$$

and

$$\sup_{i\in[n]} \left\| \left|X^\top\delta_i\hat{\theta}\right|_2 \right\|_{L^{k_n}} + \sup_{i\in[n]} \left\| \left|X_{-i}^\top\delta_i\hat{\theta}\right|_2 \right\|_{L^{k_n}} = O(k_n^2).$$

Proof. Since

$$\nabla\Psi_X(\hat{\theta}_{-i}) = \frac{1}{n} s_i^- x_i,$$

Lemma 3.9, applied with $t = \hat{\theta}_{-i}$, gives

$$\frac{\kappa_L}{n}\left|X^\top\delta_i\hat{\theta}\right|_2^2 + \kappa\left|\delta_i\hat{\theta}\right|_2^2 \le -\frac{1}{n}s_i^-\, x_i^\top\delta_i\hat{\theta}.$$

In particular,

$$\frac{\kappa_L}{n}\left|X^\top\delta_i\hat{\theta}\right|_2^2 \le \frac{1}{n}|s_i^-|\,|x_i^\top\delta_i\hat{\theta}| \le \frac{1}{n}|s_i^-|\,\left|X^\top\delta_i\hat{\theta}\right|_2,$$

because $x_i^\top\delta_i\hat{\theta}$ is one coordinate of $X^\top\delta_i\hat{\theta}$. If $\left|X^\top\delta_i\hat{\theta}\right|_2 = 0$, the prediction bound is immediate. Otherwise, dividing by $\left|X^\top\delta_i\hat{\theta}\right|_2$ yields

$$\left|X^\top\delta_i\hat{\theta}\right|_2 \le \kappa_L^{-1}|s_i^-|.$$

Returning to the same inequality and dropping the nonnegative prediction term,

$$\kappa\left|\delta_i\hat{\theta}\right|_2^2 \le \frac{1}{n}|s_i^-|\,|x_i^\top\delta_i\hat{\theta}| \le \frac{1}{n}|s_i^-|\,\left|X^\top\delta_i\hat{\theta}\right|_2 \le \frac{|s_i^-|^2}{\kappa_L n}.$$

Taking L^{k_n}-norms and using Lemma 3.8 gives the first two estimates. Finally,

$$\left|X_{-i}^\top\delta_i\hat{\theta}\right|_2 \le \left|X^\top\delta_i\hat{\theta}\right|_2,$$

which gives the last one. □

3.3 Concentration of the minimizer

We next derive the sensitivity estimates for $\hat{\theta}$ with respect to the columns of X. These estimates are the input needed to apply the tensorized Poincaré inequality to functions of $\hat{\theta}$.

Fix $i \in [n]$, freeze all columns except x_i, and perturb x_i in the direction $h \in \mathbb{R}^p$:

$$x_i(t) = x_i + th, \qquad x_j(t) = x_j \quad (j \ne i), \qquad \theta(t) = \hat{\theta}(X(t)).$$

The first-order condition is

$$0 = \nabla\Psi_{X(t)}(\theta(t)) = \frac{1}{n}\sum_{j=1}^n L'(x_j(t)^\top\theta(t))x_j(t) + \nabla\rho(\theta(t)).$$

Since Ψ_X is strongly convex and C^2, the implicit function theorem allows us to differentiate this identity at $t = 0$. Define

$$G := \nabla^2\Psi_X(\hat{\theta})^{-1}.$$

Writing $\theta'(0) = \mathbb{D}_i[\hat{\theta}][h]$, we obtain

$$G^{-1}\mathbb{D}_i[\hat{\theta}][h] + \frac{1}{n}\left[L''(x_i^\top\hat{\theta})(h^\top\hat{\theta})x_i + L'(x_i^\top\hat{\theta})h\right] = 0.$$

Equivalently,

$$\mathbb{D}_i[\hat{\theta}] = -\frac{1}{n}\left[L''(x_i^\top\hat{\theta})Gx_i\hat{\theta}^\top + s_iG\right]. \tag{3.1}$$

where we introduced the following notation for the train score:

$$s_i := L'(x_i^\top\hat{\theta})$$

Lemma 3.11 (Uniform prediction moments)**.** *Under Assumptions 1–5, for every sequence of positive integers k_n,*

$$\sup_{i\in[n]}\left\|x_i^\top\hat{\theta}\right\|_{L^{k_n}} = O(k_n^2), \qquad \sup_{i\in[n]}\|s_i\|_{L^{k_n}} = O(k_n^2).$$

Proof. Fix $i \in [n]$. Since $\hat{\theta}_{-i}$ is independent of x_i, conditioning on $\hat{\theta}_{-i}$, applying Lemma 3.2, and then using Theorem 3.7 give

$$\left\|x_i^\top\hat{\theta}_{-i}\right\|_{L^{k_n}} = O(k_n^2).$$

One can then deduce from Lemma 3.10 that:

$$\begin{aligned}\left\|x_i^\top\hat{\theta}\right\|_{L^{k_n}} &\leq \left\|x_i^\top\hat{\theta}_{-i}\right\|_{L^{k_n}} + \left\|x_i^\top(\hat{\theta} - \hat{\theta}_{-i})\right\|_{L^{k_n}} \\ &\leq O(k_n^2) + \left\|\left|X^\top(\hat{\theta} - \hat{\theta}_{-i})\right|_2\right\|_{L^{k_n}} = O(k_n^2),\end{aligned}$$

Finally, since

$$|L'(t)| \leq O(1)(1 + |t|),$$

the same bound holds for $s_i = L'(x_i^\top\hat{\theta})$. □

Lemma 3.12 (Differential bound for the minimizer)**.** *Under Assumptions 1–5, for every $i \in [n]$, every random vector $z = z(X)$, and any sequence $k = (k_n)_{n\in\mathbb{N}}$,*

$$\left\|\left|z^\top\mathbb{D}_i[\hat{\theta}]\right|^*\right\|_{L^{k_n}} \leq O\left(\frac{k_n^2}{n}\right)\left(\left\|z^\top Gx_i\right\|_{L^{2k_n}} + \||z|_2\|_{L^{2k_n}}\right).$$

Proof. Multiplying the differential identity (3.1) by $z^\top$, we get, for every $h \in \mathbb{R}^p$,

$$z^\top\mathbb{D}_i[\hat{\theta}]\,[h] = -\frac{1}{n}h^\top\left[L''(x_i^\top\hat{\theta})(x_i^\top Gz)\hat{\theta} + s_iGz\right].$$

Since $|G|_{\text{op}} \leq \kappa^{-1}$ and $|L''|_\infty = O(1)$,

$$\left|z^\top\mathbb{D}_i[\hat{\theta}]\right|^* \leq \frac{O(1)}{n}\left(|x_i^\top Gz|\,\left|\hat{\theta}\right|_2 + |s_i|\,|z|_2\right).$$

Taking L^{k_n}-norms and applying Hölder's inequality, together with Theorem 3.7 and Lemma 3.11, proves the claim. □

Lemma 3.13 (Empirical prediction energy)**.** *Under Assumptions 1–5, for every sequence of positive integers k_n,*

$$\left\|\frac{1}{n}\sum_{i=1}^n(x_i^\top\hat{\theta})^2\right\|_{L^{k_n}} = O(k_n^2), \qquad \left\|\frac{1}{n}\sum_{i=1}^n s_i^2\right\|_{L^{k_n}} = O(k_n^2).$$

Proof. Applying Lemma 3.9 with $t = 0$, we get

$$\frac{\kappa_L}{n}\sum_{i=1}^{n}(x_i^\top\hat{\theta})^2 + \kappa\left|\hat{\theta}\right|_2^2 \leq -\nabla\Psi_X(0)^\top\hat{\theta}.$$

By Young's inequality,

$$-\nabla\Psi_X(0)^\top\hat{\theta} \leq \frac{1}{2\kappa}\left|\nabla\Psi_X(0)\right|_2^2 + \frac{\kappa}{2}\left|\hat{\theta}\right|_2^2.$$

After cancellation,

$$\frac{1}{n}\sum_{i=1}^{n}(x_i^\top\hat{\theta})^2 \leq O(1)\left|\nabla\Psi_X(0)\right|_2^2.$$

Taking L^{k_n}-norms and using Lemma 3.6 at exponent $2k_n$, we obtain the first bound.

Since $|L''|_\infty = O(1)$ and $|L'(0)| = O(1)$, $|L'(t)|^2 \leq O(1)(1+t^2)$, and therefore:

$$\frac{1}{n}\sum_{i=1}^{n}s_i^2 \leq O(1)\left(1 + \frac{1}{n}\sum_{i=1}^{n}(x_i^\top\hat{\theta})^2\right),$$

which gives the second bound. □

Lemma 3.14 (Square-summed background sensitivity)**.** *Under Assumptions 1–5, for every random vector $z = z(X)$ and any sequence of integers $k_n \geq 1$,*

$$\left\|\sum_{i=1}^{n}\left(\left|z^\top\mathbb{D}_i[\hat{\theta}]\right|^*\right)^2\right\|_{L^{k_n}} \leq O\left(\frac{k_n^2}{n}\right)\||z|_2\|_{L^{4k_n}}^2.$$

Proof. The pointwise bound in the proof of Lemma 3.12 gives

$$\left(\left|z^\top\mathbb{D}_i[\hat{\theta}]\right|^*\right)^2 \leq \frac{O(1)}{n^2}\left[(x_i^\top Gz)^2|\hat{\theta}|_2^2 + s_i^2\,|z|_2^2\right].$$

By Assumption 5,

$$G^{-1} = \frac{1}{n}\sum_{i=1}^{n}L''(x_i^\top\hat{\theta})x_ix_i^\top + \nabla^2\rho(\hat{\theta}) \succeq \frac{\kappa_L}{n}\sum_{i=1}^{n}x_ix_i^\top.$$

Hence

$$G\left(\frac{1}{n}\sum_{i=1}^{n}x_ix_i^\top\right)G \preceq \kappa_L^{-1}G \preceq O(1)I_p,$$

and therefore

$$\sum_{i=1}^{n}(x_i^\top Gz)^2 \leq O(n)\,|z|_2^2.$$

Summing the pointwise bound over i yields

$$\sum_{i=1}^{n}\left(\left|z^\top\mathbb{D}_i[\hat{\theta}]\right|^*\right)^2 \leq \frac{O(1)}{n}\left[|\hat{\theta}|_2^2 + \frac{1}{n}\sum_{i=1}^{n}s_i^2\right]|z|^2.$$

Taking L^{k_n}-norms and using Hölder's inequality gives

$$\left\|\sum_{i=1}^{n}\left(\left|z^\top\mathbb{D}_i[\hat{\theta}]\right|^*\right)^2\right\|_{L^{k_n}} \leq \frac{O(1)}{n}\left\|\left|\hat{\theta}\right|_2^2 + \frac{1}{n}\sum_{i=1}^{n}s_i^2\right\|_{L^{2k_n}}\left\||z|_2^2\right\|_{L^{2k_n}}.$$

By Theorem 3.7 and Lemma 3.13, applied at exponent $2k_n$, the first factor is $O(k_n^2)$. Since

$$\left\||z|_2^2\right\|_{L^{2k_n}} = \||z|_2\|_{L^{4k_n}}^2,$$

the claim follows. □

Corollary 3.15 (Variance bounds for smooth functions of the minimizer)**.** *Under Assumptions 1–5, for every sufficiently smooth* $f:\mathbb{R}^p\to\mathbb{R}$,

$$\mathrm{Var}(f(\hat\theta)) \le O\left(\frac{1}{n}\left\|\left|\nabla f(\hat\theta)\right|_2\right\|_{L^4}^2\right).$$

Proof. By the chain rule,

$$\left|\mathbb{D}_i[f(\hat\theta)]\right|_2^* = \left|\nabla f(\hat\theta)^\top \mathbb{D}_i[\hat\theta]\right|^*.$$

Applying Lemma 3.1 and then Lemma 3.14 with $z=\nabla f(\hat\theta)$ and $k_n=1$ gives

$$\mathrm{Var}(f(\hat\theta)) \le C_P\,\mathbb{E}\sum_{i=1}^n \left(\left|\nabla f(\hat\theta)^\top \mathbb{D}_i[\hat\theta]\right|^*\right)^2 \le O\left(\frac{1}{n}\left\|\left|\nabla f(\hat\theta)\right|_2\right\|_{L^4}^2\right).$$

□

Remark 3.16. Corollary 3.15 gives, in the setting of Theorem 2.1,

$$\sigma_n^2 = n\ \mathrm{Var}(u^\top\hat\theta) = O(1).$$

Thus the additional variance assumption in Theorem 2.1 is only the lower bound $\inf_n \sigma_n^2 > 0$.

Until now, we were able to show in Lemma 3.10 that $\||\delta_i\hat\theta|_2\|_{L^k} \le O(1/\sqrt{n})$ however that do not allow us to reach already the bound $\|\sqrt{n}u^\top\delta_i\hat\theta|\|_{L^k}\le O(1/\sqrt{n})$ that is needed to be able to use the Wasserstein bound of theorem 2.2. This estimate will be though provided through a precise estimation of the leave-one-out displacement in next section

4 Leave-one-out perturbation expansion

The goal of this section is to obtain quantitative expansions of the leave-one-out displacement

$$\delta_i\hat\theta := \hat\theta - \hat\theta_{-i}, \qquad i\in[n].$$

The prediction stability estimates proved below give direct control of $X^\top\delta_i\hat\theta$, while the Wasserstein bound requires a more precise description of deterministic projections $u^\top\delta_i\hat\theta$. We first derive a first-order expansion, then introduce an averaged-leverage score which removes the leading dependence of the score on the random leverage $n^{-1}x_i^\top G_{-i}x_i$. Finally, under the third-order curvature assumptions and the condition $\nabla^3\rho\equiv 0$, we derive a projected second-order expansion based on this averaged score.

For $i\in[n]$, write

$$\Psi_X^{-i}(\theta) := \frac{1}{n}\sum_{\substack{j\in[n]\\ j\neq i}} L(x_j^\top\theta) + \rho(\theta), \qquad \hat\theta_{-i} := \mathrm{argmin}_{\theta\in\mathbb{R}^p}\,\Psi_X^{-i}(\theta),$$

and

$$G_{-i} := \left(\nabla^2\Psi_X^{-i}(\hat\theta_{-i})\right)^{-1}, \qquad s_i := L'(x_i^\top\hat\theta).$$

By Assumption 5, G_{-i} is well-defined and

$$|G_{-i}|_{\mathrm{op}} \le \kappa^{-1}.$$

4.1 First-order expansion

Lemma 4.1 (Exact leave-one-out identity). *Under Assumption 5, for every $i \in [n]$,*

$$\delta_i\hat{\theta} = -\frac{s_i}{n}\bar{G}_i x_i, \qquad \bar{G}_i := \left[\int_0^1 \nabla^2\Psi_X^{-i}\big(\hat{\theta}_{-i} + t\delta_i\hat{\theta}\big)\,dt\right]^{-1}.$$

Proof. Since $\nabla\Psi_X(\hat{\theta}) = 0$ and

$$\Psi_X(\theta) = \Psi_X^{-i}(\theta) + \frac{1}{n}L(x_i^\top\theta),$$

we have

$$\nabla\Psi_X^{-i}(\hat{\theta}) = -\frac{1}{n}L'(x_i^\top\hat{\theta})x_i = -\frac{s_i}{n}x_i.$$

Moreover, $\nabla\Psi_X^{-i}(\hat{\theta}_{-i}) = 0$. Hence

$$-\frac{s_i}{n}x_i = \nabla\Psi_X^{-i}(\hat{\theta}) - \nabla\Psi_X^{-i}(\hat{\theta}_{-i}) = \left[\int_0^1 \nabla^2\Psi_X^{-i}\big(\hat{\theta}_{-i} + t\delta_i\hat{\theta}\big)\,dt\right]\delta_i\hat{\theta}.$$

The integrated Hessian is invertible by Assumption 5, and the claim follows. □

Write

$$\bar{G}_i^{-1} = G_{-i}^{-1} + R_i, \qquad R_i = R_i^\rho + R_i^L,$$

where

$$R_i^\rho := \int_0^1 \left[\nabla^2\rho(\hat{\theta}_{-i} + t\delta_i\hat{\theta}) - \nabla^2\rho(\hat{\theta}_{-i})\right] dt, \tag{4.1}$$

and

$$R_i^L := \frac{1}{n}X_{-i}\Gamma_i X_{-i}^\top. \tag{4.2}$$

Here Γ_i is the diagonal matrix, indexed by $j \neq i$, with entries

$$(\Gamma_i)_j := \int_0^1 \left[L''\Big(x_j^\top(\hat{\theta}_{-i} + t\delta_i\hat{\theta})\Big) - L''(x_j^\top\hat{\theta}_{-i})\right] dt.$$

The resolvent identity gives

$$G_{-i} - \bar{G}_i = \bar{G}_i R_i G_{-i}. \tag{4.3}$$

Lemma 4.2 (Curvature bounds for the resolvents). *Under Assumption 5, for every $i \in [n]$ and every realization,*

$$\left|\bar{G}_i\right|_{\mathrm{op}} + \left|G_{-i}\right|_{\mathrm{op}} = O(1),$$
$$\left|\bar{G}_i X_{-i}\right|_{\mathrm{op}} + \left|G_{-i}X_{-i}\right|_{\mathrm{op}} = O(\sqrt{n}),$$

and

$$\left|X_{-i}^\top\bar{G}_i X_{-i}\right|_{\mathrm{op}} + \left|X_{-i}^\top G_{-i}X_{-i}\right|_{\mathrm{op}} = O(n).$$

Proof. For every θ,

$$\nabla^2\Psi_X^{-i}(\theta) \succeq \kappa I_p + \frac{\kappa_L}{n}X_{-i}X_{-i}^\top.$$

The same lower bound holds after averaging in the definition of $\bar{G}_i^{-1}$. Thus G_{-i}^{-1} and $\bar{G}_i^{-1}$ are both bounded below by the right-hand side. This first gives

$$\left|G_{-i}\right|_{\mathrm{op}} + \left|\bar{G}_i\right|_{\mathrm{op}} = O(1).$$

If $M \in \{G_{-i}, \bar{G}_i\}$, then

$$M^{-1} \succeq \frac{\kappa_L}{n} X_{-i} X_{-i}^\top.$$

Therefore

$$M^{1/2} X_{-i} X_{-i}^\top M^{1/2} \preceq \frac{n}{\kappa_L} I_p.$$

It follows that

$$|M X_{-i}|_{\mathrm{op}}^2 = \left| M X_{-i} X_{-i}^\top M \right|_{\mathrm{op}} \leq \frac{n}{\kappa_L} |M|_{\mathrm{op}} = O(n),$$

and

$$\left| X_{-i}^\top M X_{-i} \right|_{\mathrm{op}} = \left| M^{1/2} X_{-i} X_{-i}^\top M^{1/2} \right|_{\mathrm{op}} \leq \frac{n}{\kappa_L}.$$

This proves the claim. □

Lemma 4.3 (Score comparison and score moments)**.** *Under Assumptions 1–5, define*

$$s_i^- := L'(x_i^\top \hat{\theta}_{-i}).$$

Then, for every $i \in [n]$,

$$|s_i| \leq |s_i^-|.$$

Moreover, for every fixed $k \geq 1$,

$$\sup_{i \in [n]} \|s_i\|_{L^k} + \sup_{i \in [n]} \left\| s_i^- \right\|_{L^k} = O_k(1).$$

Proof. Let

$$\alpha_i := \int_0^1 L''\left(x_i^\top (\hat{\theta}_{-i} + t \delta_i \hat{\theta}) \right) dt.$$

Then

$$s_i = s_i^- + \alpha_i x_i^\top \delta_i \hat{\theta}.$$

By Lemma 4.1,

$$x_i^\top \delta_i \hat{\theta} = -\frac{s_i}{n} x_i^\top \bar{G}_i x_i.$$

Thus

$$s_i \left(1 + \frac{\alpha_i}{n} x_i^\top \bar{G}_i x_i \right) = s_i^-.$$

Since $\alpha_i \geq 0$ and $\bar{G}_i \succeq 0$, we obtain

$$|s_i| \leq |s_i^-|.$$

It remains to bound s_i^-. By Assumptions 2 and 5,

$$|L'(t)| \leq |L'(0)| + \left| L'' \right|_\infty |t| \leq O(1)(1 + |t|).$$

Conditionally on X_{-i}, the vector $\hat{\theta}_{-i}$ is deterministic and independent of x_i. Hence Lemma 3.2 gives

$$\left\| x_i^\top \hat{\theta}_{-i} \right\|_{L_i^k} \leq O_k(1) \left| \hat{\theta}_{-i} \right|_2.$$

Taking the L^k-norm over X_{-i} and using Theorem 3.7, we get

$$\sup_{i \in [n]} \left\| x_i^\top \hat{\theta}_{-i} \right\|_{L^k} = O_k(1).$$

Therefore

$$\sup_{i \in [n]} \left\| s_i^- \right\|_{L^k} = O_k(1),$$

and the comparison $|s_i| \leq |s_i^-|$ gives the same bound for s_i. □

Lemma 4.4 (Cross-leverage bounds). *Under Assumptions 1–5, for every fixed $k \geq 1$,*

$$\sup_{i\in[n]} \left\| \left| X_{-i}^\top G_{-i} x_i \right|_\infty \right\|_{L^k} = O_k(\sqrt{n}\,\log n),$$

and

$$\sup_{i\in[n]} \left\| \left| X_{-i}^\top G_{-i} x_i \right|_2 \right\|_{L^k} = O_k(n).$$

Proof. Condition on X_{-i}. Then G_{-i} and $(x_j)_{j\neq i}$ are fixed and independent of x_i. For every $j \neq i$, Lemma 4.2 gives

$$|G_{-i}x_j|_2 \leq |G_{-i}X_{-i}|_{\text{op}} = O(\sqrt{n}).$$

Thus, conditionally on X_{-i}, the linear moment bound under the Poincaré assumption gives, for every integer $q \geq 1$,

$$\max_{j\neq i} \left\| x_i^\top G_{-i} x_j \right\|_{L_i^q} \leq O(q\sqrt{n}).$$

Applying Lemma B.7 conditionally on X_{-i} gives

$$\left\| \max_{j\neq i} |x_i^\top G_{-i} x_j| \right\|_{L_i^k} \leq O_k(\sqrt{n}\,\log n).$$

Taking the L^k-norm over X_{-i} proves the first estimate.

For the second estimate, Lemma 4.2 gives

$$\left| X_{-i}^\top G_{-i} x_i \right|_2 \leq |G_{-i}X_{-i}|_{\text{op}} \, |x_i|_2 = O(\sqrt{n}) \, |x_i|_2 \, .$$

The claim follows from Lemma 3.2. □

Lemma 4.5 (Frobenius bound for the loss-curvature drift). *Under Assumptions 1–6, for every fixed $k \geq 1$,*

$$\sup_{i\in[n]} \| |\Gamma_i|_F \|_{L^k} = O_k(1).$$

Proof. By Assumption 6, for $j \neq i$,

$$|(\Gamma_i)_j| \leq O(1) |x_j^\top \delta_i \hat{\theta}|.$$

Therefore

$$|\Gamma_i|_F \leq O(1) \left| X_{-i}^\top \delta_i \hat{\theta} \right|_2 .$$

The result follows from Lemma 3.10. □

Remark 4.6. At this stage we only need a Frobenius bound on Γ_i. The sharper entrywise estimate

$$\sup_{i\in[n]} \left\| |\Gamma_i|_{\text{op}} \right\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}} \right)$$

will follow later from the projected first-order estimate (see Lemma 4.18). The Frobenius bound is sufficient for the first-order expansion because it is used through

$$|\Gamma_i v|_2 \leq |\Gamma_i|_F \, |v|_\infty \, , \qquad v \in \mathbb{R}^{n-1}.$$

Lemma 4.7 (Regularizer drift term). *Under Assumptions 1–6, for every fixed $k \geq 1$,*

$$\sup_{i\in[n]} \left\| \left| \frac{s_i}{n} \bar{G}_i R_i^\rho G_{-i} x_i \right|_2 \right\|_{L^k} = O_k\left(\frac{1}{n} \right).$$

Proof. By Assumption 6 and Lemma 3.10,

$$|R_i^\rho|_{\mathrm{op}} \le O(1)\left|\delta_i\hat{\theta}\right|_2 \le O(1)\frac{|s_i^-|}{\sqrt{n}}. \tag{4.4}$$

Consequently,

$$\left|\frac{s_i}{n}\bar{G}_i R_i^\rho G_{-i}x_i\right|_2 \le \frac{O(1)}{n\sqrt{n}}|s_i|\,|s_i^-|\,|x_i|_2\,.$$

The claim follows from Lemmas 4.3 and 3.2, together with Hölder's inequality. ☐

Lemma 4.8 (Loss drift term)**.** *Under Assumptions 1–6, for every fixed $k \ge 1$,*

$$\sup_{i\in[n]}\left\|\left|\frac{s_i}{n}\bar{G}_i R_i^L G_{-i}x_i\right|_2\right\|_{L^k} = O_k\left(\frac{\log n}{n}\right).$$

Proof. Since $R_i^L = n^{-1}X_{-i}\Gamma_i X_{-i}^\top$,

$$\left|\frac{s_i}{n}\bar{G}_i R_i^L G_{-i}x_i\right|_2 \le \frac{|s_i|}{n^2}\left|\bar{G}_i X_{-i}\right|_{\mathrm{op}}\left|\Gamma_i\right|_F\left|X_{-i}^\top G_{-i}x_i\right|_\infty.$$

By Lemmas 4.3, 4.2, 4.5, and 4.4, Hölder's inequality gives

$$\sup_{i\in[n]}\left\|\left|\frac{s_i}{n}\bar{G}_i R_i^L G_{-i}x_i\right|_2\right\|_{L^k} = O_k\left(\frac{\log n}{n}\right).$$

☐

Define the first-order leave-one-out displacement based on the train score by

$$\delta_{i,\mathrm{tr}}^{(0)}\hat{\theta} := -\frac{s_i}{n}G_{-i}x_i.$$

Theorem 4.9 (First-order vector leave-one-out expansion)**.** *Under Assumptions 1–6, for every fixed $k \ge 1$,*

$$\sup_{i\in[n]}\left\|\left|\delta_i\hat{\theta} - \delta_{i,\mathrm{tr}}^{(0)}\hat{\theta}\right|_2\right\|_{L^k} = O_k\left(\frac{\log n}{n}\right),$$

and

$$\sup_{i\in[n]}\left\|\left|X_{-i}^\top(\delta_i\hat{\theta} - \delta_{i,\mathrm{tr}}^{(0)}\hat{\theta})\right|_2\right\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}}\right).$$

Proof. Using Lemma 4.1 and (4.3),

$$\delta_i\hat{\theta} - \delta_{i,\mathrm{tr}}^{(0)}\hat{\theta} = \frac{s_i}{n}\bar{G}_i R_i G_{-i}x_i = \frac{s_i}{n}\bar{G}_i R_i^\rho G_{-i}x_i + \frac{s_i}{n}\bar{G}_i R_i^L G_{-i}x_i.$$

The first estimate follows from Lemmas 4.7 and 4.8.

For the projected estimate, multiply the preceding identity by $X_{-i}^\top$. For the regularizer contribution, using (4.4), Lemma 4.2, and $|G_{-i}x_i|_2 \le O(1)\,|x_i|_2$, we get

$$\left\|\left|X_{-i}^\top\frac{s_i}{n}\bar{G}_i R_i^\rho G_{-i}x_i\right|_2\right\|_{L^k} = O_k\left(\frac{1}{\sqrt{n}}\right).$$

For the loss contribution,

$$\left|X_{-i}^\top\frac{s_i}{n}\bar{G}_i R_i^L G_{-i}x_i\right|_2 \le \frac{|s_i|}{n^2}\left|X_{-i}^\top\bar{G}_i X_{-i}\right|_{\mathrm{op}}\left|\Gamma_i\right|_F\left|X_{-i}^\top G_{-i}x_i\right|_\infty.$$

Using Lemmas 4.3, 4.2, 4.5, and 4.4, we obtain

$$\sup_{i\in[n]}\left\|\left|X_{-i}^\top\frac{s_i}{n}\bar{G}_i R_i^L G_{-i}x_i\right|_2\right\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}}\right).$$

Combining the two contributions proves the projected estimate. ☐

Corollary 4.10 (Projected leave-one-out bounds). *Under Assumptions 1–6, for every fixed $k \geq 1$ and every deterministic $u \in \mathbb{R}^p$ with $|u|_2 \leq 1$,*

$$\sup_{i\in[n]} \left\| u^\top \delta_{i,\mathrm{tr}}^{(0)} \hat{\theta} \right\|_{L^k} = O_k\left(\frac{1}{n}\right), \qquad \text{and} \qquad \sup_{i\in[n]} \left\| u^\top \delta_i \hat{\theta} \right\|_{L^k} = O_k\left(\frac{\log n}{n}\right).$$

Moreover,

$$\sup_{i\in[n]} \left\| \left| X_{-i}^\top \delta_{i,\mathrm{tr}}^{(0)} \hat{\theta} \right|_2 \right\|_{L^k} = O_k(1), \qquad \text{and} \qquad \sup_{i\in[n]} \left\| \left| X_{-i}^\top \delta_i \hat{\theta} \right|_2 \right\|_{L^k} = O_k(1).$$

Proof. For the first scalar bound,

$$|u^\top \delta_{i,\mathrm{tr}}^{(0)} \hat{\theta}| \leq \frac{|s_i|}{n} |u^\top G_{-i} x_i|.$$

Conditionally on X_{-i}, the vector $G_{-i}u$ is deterministic and has bounded Euclidean norm. Hence Lemma 3.2 gives

$$\left\| u^\top G_{-i} x_i \right\|_{L^k} = O_k(1).$$

Together with Lemma 4.3 and Hölder's inequality, this proves

$$\sup_{i\in[n]} \left\| u^\top \delta_{i,\mathrm{tr}}^{(0)} \hat{\theta} \right\|_{L^k} = O_k\left(\frac{1}{n}\right).$$

The bound for $u^\top \delta_i \hat{\theta}$ follows from Theorem 4.9.

Next,

$$\left| X_{-i}^\top \delta_{i,\mathrm{tr}}^{(0)} \hat{\theta} \right|_2 \leq \frac{|s_i|}{n} \left| X_{-i}^\top G_{-i} x_i \right|_2.$$

Using Lemmas 4.3 and 4.4 gives the $O_k(1)$ bound. The final estimate follows from Lemma 3.10. ☐

Remark 4.11 (Consequence for the leave-one-out Wasserstein bound). For $g_n = f_n$, the leave-one-out increment in Theorem 2.2 is

$$\delta_i f_n(X) = f_n(X) - f_n^{-i}(X_{-i}) = \sqrt{n}\, u^\top \delta_i \hat{\theta}(X).$$

Taking $k = 4$ in Corollary 4.10 gives

$$m_n^{(4)} = \sup_{i\in[n]} \|\delta_i f_n(X)\|_{L^4} = O\left(\frac{\log n}{\sqrt{n}}\right).$$

For the two covariance quantities in Theorem 2.2, the corresponding expressions are

$$c_n^x = \sqrt{n} \sup_{\substack{i,j\in[n]\\ i\neq j}} \left\| u^\top \delta_i \hat{\theta}(X) - \mathbb{E}_j\left[u^\top \delta_i \hat{\theta}(X)\right] \right\|_{L^4},$$

and

$$c_n^{\bar{x}} = n \sup_{j\in[n]} \sqrt{\mathrm{Var}\Big(\mathrm{Cov}_j\Big(u^\top \delta_j \hat{\theta}(X), u^\top \delta_j \hat{\theta}(X^{A_j})\Big)\Big)}.$$

The first-order expansion is not accurate enough for these two quantities. Indeed, Theorem 4.9 implies only

$$\left| c_n^x - \sqrt{n} \sup_{\substack{i,j\in[n]\\ i\neq j}} \left\| u^\top \delta_{i,\mathrm{tr}}^{(0)} \hat{\theta}(X) - \mathbb{E}_j\left[u^\top \delta_{i,\mathrm{tr}}^{(0)} \hat{\theta}(X)\right] \right\|_{L^4} \right| = O\left(\frac{\log n}{\sqrt{n}}\right),$$

which is too large for the target bound $c_n^x = O((\log n)^2/n)$. This motivate the second-order expansions below. But before that, to better disentangle the dependence of s_i on x_i, we will introduce an approximation of s_i that expresses as a Lipschitz functional of $x_i^\top \hat{\theta}_{-i}$, namely s_i^ζ. That will be crucial to bound efficiently $c_i^{\bar{z}}$. To be more precise, we need to work with s_i^ζ to be able to set Lemma 5.16 with sufficiently small bounds.

4.2 Averaged-leverage score

We now replace the train score $s_i = L'(x_i^\top \hat{\theta})$ by a leave-one-out score depending on x_i only through $x_i^\top \hat{\theta}_{-i}$. Let us introduce the averaged leverage:

$$\gamma_i := \mathbb{E}\left[\frac{1}{n} x_i^\top G_{-i} x_i\right].$$

Since x_i is independent of G_{-i}, this can also be written as

$$\gamma_i = \mathbb{E}\left[\frac{1}{n}\operatorname{Tr}(\Sigma_i G_{-i})\right], \qquad \Sigma_i := \mathbb{E}[x_i x_i^\top].$$

By Assumptions 4 and 3,

$$|\Sigma_i|_{\text{op}} = O(1),$$

and since $|G_{-i}|_{\text{op}} = O(1)$ and $p = O(n)$,

$$0 \le \gamma_i = O(1).$$

The following lemma allows to replace the conditional averaged leverage by its full expectation.

Lemma 4.12 (Concentration of the averaged leverage)**.** *Under Assumptions 1–7, for every fixed $k \ge 1$,*

$$\sup_{i\in[n]} \left\| \frac{1}{n} x_i^\top G_{-i} x_i - \gamma_i \right\|_{L^k} = O_k\left(\frac{1}{\sqrt{n}}\right).$$

Proof. We decompose

$$\frac{1}{n} x_i^\top G_{-i} x_i - \gamma_i = \frac{1}{n}\left(x_i^\top G_{-i} x_i - \operatorname{Tr}(\Sigma_i G_{-i})\right) + \left(\frac{1}{n}\operatorname{Tr}(\Sigma_i G_{-i}) - \mathbb{E}\left[\frac{1}{n}\operatorname{Tr}(\Sigma_i G_{-i})\right]\right).$$

We first treat the conditional quadratic fluctuation. Conditionally on X_{-i}, the matrix G_{-i} is deterministic and independent of x_i. Thus Lemma 3.3 gives

$$\sup_{i\in[n]} \left\| \frac{1}{n}\left(x_i^\top G_{-i} x_i - \operatorname{Tr}(\Sigma_i G_{-i})\right) \right\|_{L^k} \le \sup_{i\in[n]} O_k\left(\frac{|G_{-i}|_F}{n}\right) = O_k\left(\frac{1}{\sqrt{n}}\right),$$

since $\sup_{i\in[n]} |G_{-i}|_{\text{op}} = O(1)$ and $p = O(n)$.

It remains to control

$$\frac{1}{n}\operatorname{Tr}(\Sigma_i G_{-i}) - \mathbb{E}\left[\frac{1}{n}\operatorname{Tr}(\Sigma_i G_{-i})\right].$$

By Lemma B.2, it is enough to prove

$$\sup_{i\in[n]} \left\| \left(\sum_{j\neq i} \left(\left| \mathbb{D}_j \left[\frac{1}{n}\operatorname{Tr}(\Sigma_i G_{-i})\right] \right|_2^* \right)^2 \right)^{1/2} \right\|_{L^k} = O_k\left(\frac{1}{\sqrt{n}}\right).$$

Fix $j \neq i$. By the resolvent differential identity,

$$\mathbb{D}_j\left[\frac{1}{n}\operatorname{Tr}(\Sigma_i G_{-i})\right][h] = \frac{1}{n}\operatorname{Tr}(\Sigma_i \mathbb{D}_j[G_{-i}][h]) = -\frac{1}{n}\operatorname{Tr}\big(G_{-i}\Sigma_i G_{-i}\, \mathbb{D}_j[G_{-i}^{-1}][h]\big).$$

We now write explicitly the derivative of G_{-i}^{-1}. Since

$$G_{-i}^{-1} = \frac{1}{n}\sum_{\ell\neq i} L''(x_\ell^\top \hat{\theta}_{-i}) x_\ell x_\ell^\top + \nabla^2 \rho(\hat{\theta}_{-i}),$$

we decompose

$$\mathbb{D}_j[G_{-i}^{-1}][h] = \mathcal{L}_{ij}[h] + \mathcal{B}_{ij}[h],$$

where the local part, coming from the direct differentiation of the column x_j, is

$$\mathcal{L}_{ij}[h] := \frac{1}{n} L''(x_j^\top \hat{\theta}_{-i}) \left(h x_j^\top + x_j h^\top \right) + \frac{1}{n} L'''(x_j^\top \hat{\theta}_{-i})(h^\top \hat{\theta}_{-i}) x_j x_j^\top,$$

and the background part, coming from the variation of $\hat{\theta}_{-i}$, is

$$\mathcal{B}_{ij}[h] := \frac{1}{n} \sum_{\ell \neq i} L'''(x_\ell^\top \hat{\theta}_{-i}) \left(x_\ell^\top \mathbb{D}_j[\hat{\theta}_{-i}][h] \right) x_\ell x_\ell^\top + \nabla^3 \rho(\hat{\theta}_{-i}) \big[\mathbb{D}_j[\hat{\theta}_{-i}][h] \big].$$

Thus the derivative of the trace splits into a local and a background contribution.

We first bound the local contribution. Using $|L''|_\infty = O(1)$, $|L'''|_\infty = O(1)$, $|G_{-i}\Sigma_i G_{-i}|_{\mathrm{op}} = O(1)$, and $|h|_2 \le 1$, we have

$$\left| \frac{1}{n} \operatorname{Tr}(G_{-i}\Sigma_i G_{-i} \mathcal{L}_{ij}[h]) \right| \le \frac{O(1)}{n^2} \left(|G_{-i}\Sigma_i G_{-i} x_j|_2 + \left| \hat{\theta}_{-i} \right|_2 \left| x_j^\top G_{-i}\Sigma_i G_{-i} x_j \right| \right).$$

Consequently,

$$\begin{aligned} &\left\| \left(\sum_{j \neq i} \left[\sup_{|h|_2 \le 1} \left| \frac{1}{n} \operatorname{Tr}(G_{-i}\Sigma_i G_{-i} \mathcal{L}_{ij}[h]) \right| \right]^2 \right)^{1/2} \right\|_{L^k} \\ &\quad \le \frac{O(1)}{n^2} \left\| \left(\sum_{j \neq i} |G_{-i}\Sigma_i G_{-i} x_j|_2^2 \right)^{1/2} \right\|_{L^k} + \frac{O(1)}{n^2} \left\| \left| \hat{\theta}_{-i} \right|_2 \left(\sum_{j \neq i} \left| x_j^\top G_{-i}\Sigma_i G_{-i} x_j \right|^2 \right)^{1/2} \right\|_{L^k}. \end{aligned}$$

The first term is $O_k(n^{-1})$, since

$$\left(\sum_{j \neq i} |G_{-i}\Sigma_i G_{-i} x_j|_2^2 \right)^{1/2} \le O(1) \, |X_{-i}|_F$$

and $\| |X_{-i}|_F \|_{L^k} = O_k(n)$. For the second term,

$$\left| x_j^\top G_{-i}\Sigma_i G_{-i} x_j \right| \le |G_{-i}\Sigma_i G_{-i}|_{\mathrm{op}} |x_j|_2^2 \le |G_{-i}|_{\mathrm{op}}^2 |\Sigma_i|_{\mathrm{op}} |x_j|_2^2 \le O(1) \, |x_j|_2^2,$$

so

$$\begin{aligned} \left\| \left(\sum_{j \neq i} \left| x_j^\top G_{-i}\Sigma_i G_{-i} x_j \right|^2 \right)^{1/2} \right\|_{L^{2k}} &\le O(1) \left(\sum_{j \neq i} \left\| |x_j|_2^2 \right\|_{L^{2k}}^2 \right)^{1/2} \\ &= O(1) \left(\sum_{j \neq i} \| |x_j|_2 \|_{L^{4k}}^4 \right)^{1/2} = O_k(n^{3/2}). \end{aligned}$$

Together with

$$\left\| \left| \hat{\theta}_{-i} \right|_2 \right\|_{L^{2k}} = O_k(1),$$

this gives

$$\left\| \left(\sum_{j \neq i} \left[\sup_{|h|_2 \le 1} \left| \frac{1}{n} \operatorname{Tr}(G_{-i}\Sigma_i G_{-i} \mathcal{L}_{ij}[h]) \right| \right]^2 \right)^{1/2} \right\|_{L^k} = O_k\!\left(\frac{1}{\sqrt{n}} \right).$$

We now treat the background contribution. Define the random vector $z_i^\gamma \in \mathbb{R}^p$ by

$$z_i^\gamma := \frac{1}{n}\sum_{\ell \neq i} L'''(x_\ell^\top \hat{\theta}_{-i}) \left(x_\ell^\top G_{-i}\Sigma_i G_{-i} x_\ell\right) x_\ell + z_i^{\gamma,\rho},$$

where $z_i^{\gamma,\rho}$ is the representing vector of the linear functional

$$v \mapsto \operatorname{Tr}\Big(G_{-i}\Sigma_i G_{-i}\nabla^3\rho(\hat{\theta}_{-i})[v]\Big),$$

that is,

$$(z_i^{\gamma,\rho})^\top v = \operatorname{Tr}\Big(G_{-i}\Sigma_i G_{-i}\nabla^3\rho(\hat{\theta}_{-i})[v]\Big).$$

Then, for every $h \in \mathbb{R}^p$,

$$\operatorname{Tr}(G_{-i}\Sigma_i G_{-i}\mathcal{B}_{ij}[h]) = (z_i^\gamma)^\top \mathbb{D}_j[\hat{\theta}_{-i}][h].$$

Thus the background contribution to the square-summed derivative is

$$\left(\sum_{j\neq i}\left[\sup_{|h|_2\leq 1}\left|\frac{1}{n}\operatorname{Tr}(G_{-i}\Sigma_i G_{-i}\mathcal{B}_{ij}[h])\right|\right]^2\right)^{1/2} = \frac{1}{n}\left(\sum_{j\neq i}\left(\left|(z_i^\gamma)^\top \mathbb{D}_j[\hat{\theta}_{-i}]\right|^*\right)^2\right)^{1/2}.$$

We first record the bound

$$\big\||z_i^\gamma|_2\big\|_{L^{4k}} = O_k(n).$$

Indeed, set

$$A_i := G_{-i}\Sigma_i G_{-i}.$$

For the regularizer contribution, Assumption 7 gives

$$\begin{aligned}|z_i^{\gamma,\rho}|_2 &= \sup_{|v|_2\leq 1}\left|\operatorname{Tr}\Big(A_i\nabla^3\rho(\hat{\theta}_{-i})[v]\Big)\right| \\ &\leq |A_i|_* \sup_{|v|_2\leq 1}\left|\nabla^3\rho(\hat{\theta}_{-i})[v]\right|_{\mathrm{op}} \leq O(1)\,|A_i|_*.\end{aligned}$$

Since $|A_i|_{\mathrm{op}} = O(1)$ and $\operatorname{rank}(A_i) \leq p = O(n)$,

$$|z_i^{\gamma,\rho}|_2 = O(n).$$

This is sufficient for the argument below, since the loss contribution is also $O_k(n)$ in L^{4k}. For the loss part, using $|L'''|_\infty = O(1)$, $|G_{-i}\Sigma_i G_{-i}|_{\mathrm{op}} = O(1)$, and Lemma 3.4,

$$\left\|\frac{1}{n}\sum_{\ell\neq i} L'''(x_\ell^\top\hat{\theta}_{-i})\left(x_\ell^\top G_{-i}\Sigma_i G_{-i}x_\ell\right)x_\ell\right\|_2 \leq \frac{O(1)}{n}\,|X_{-i}|_{\mathrm{op}}\left(\sum_{\ell\neq i}|x_\ell|_2^4\right)^{1/2}.$$

Taking the L^{4k}-norm and using

$$\big\||X_{-i}|_{\mathrm{op}}\big\|_{L^{8k}} = O_k(\sqrt{n}), \qquad \left\|\left(\sum_{\ell\neq i}|x_\ell|_2^4\right)^{1/2}\right\|_{L^{8k}} = O_k(n^{3/2}),$$

we obtain the claimed $O_k(n)$ bound for $\big\||z_i^\gamma|_2\big\|_{L^{4k}}$.

We may now invoke Lemma 3.14, applied to the leave-one-out minimizer $\hat{\theta}_{-i}$. The same proof applies verbatim after replacing $X, \hat{\theta}, G$ by $X_{-i}, \hat{\theta}_{-i}, G_{-i}$, with the normalization still equal to $1/n$. Hence

$$\left\| \left(\sum_{j \neq i} \left(\left| (z_i^\gamma)^\top \mathbb{D}_j[\hat{\theta}_{-i}] \right|^* \right)^2 \right)^{1/2} \right\|_{L^k} \leq O_k\left(\frac{1}{\sqrt{n}}\right) \left\| |z_i^\gamma|_2 \right\|_{L^{4k}} = O_k(\sqrt{n}).$$

Multiplying by the outer factor $1/n$, we get

$$\left\| \left(\sum_{j \neq i} \left[\sup_{|h|_2 \leq 1} \left| \frac{1}{n} \operatorname{Tr}(G_{-i} \Sigma_i G_{-i} \mathcal{B}_{ij}[h]) \right| \right]^2 \right)^{1/2} \right\|_{L^k} = O_k\left(\frac{1}{\sqrt{n}}\right).$$

Combining the local and background contributions gives

$$\sup_{i \in [n]} \left\| \left(\sum_{j \neq i} \left(\left| \mathbb{D}_j \left[\frac{1}{n} \operatorname{Tr}(\Sigma_i G_{-i}) \right] \right|_2^* \right)^2 \right)^{1/2} \right\|_{L^k} = O_k\left(\frac{1}{\sqrt{n}}\right).$$

The tensorized L^k-Poincaré inequality therefore yields

$$\sup_{i \in [n]} \left\| \frac{1}{n} \operatorname{Tr}(\Sigma_i G_{-i}) - \mathbb{E}\left[\frac{1}{n} \operatorname{Tr}(\Sigma_i G_{-i}) \right] \right\|_{L^k} = O_k\left(\frac{1}{\sqrt{n}}\right).$$

Together with the conditional quadratic-form estimate at the beginning of the proof, this proves

$$\sup_{i \in [n]} \left\| \frac{1}{n} x_i^\top G_{-i} x_i - \gamma_i \right\|_{L^k} = O_k\left(\frac{1}{\sqrt{n}}\right),$$

□

For $t \in \mathbb{R}$, define $\zeta_i(t)$ as the unique solution of (see Lemma 4.13 for a justification of this definition)

$$z + \gamma_i L'(z) = t,$$

where

$$\gamma_i := \mathbb{E}\left[\frac{1}{n} x_i^\top G_{-i} x_i \right].$$

We also set

$$s_i^\zeta := L'\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big).$$

Lemma 4.13 (Definition and elementary properties of ζ_i)**.** *Under Assumption 5, for every $i \in [n]$ and every $t \in \mathbb{R}$, the equation*

$$z + \gamma_i L'(z) = t$$

admits a unique solution $z = \zeta_i(t)$. Moreover, $t \mapsto \zeta_i(t)$ is 1-Lipschitz and

$$|\zeta_i(t)| \leq |t| + \gamma_i |L'(0)|.$$

Consequently, under Assumptions 1–5, for every fixed $k \geq 1$,

$$\sup_{i \in [n]} \left\| s_i^\zeta \right\|_{L^k} = O_k(1).$$

Proof. Let

$$H_i(z) := z + \gamma_i L'(z).$$

Then

$$H_i'(z) = 1 + \gamma_i L''(z) \geq 1 + \gamma_i \kappa_L \geq 1.$$

Hence H_i is strictly increasing. Moreover, since $H_i'(z) \geq 1$,

$$H_i(z) \to +\infty \quad \text{as } z \to +\infty, \qquad H_i(z) \to -\infty \quad \text{as } z \to -\infty.$$

Thus H_i is a bijection from $\mathbb{R}$ to $\mathbb{R}$, and $\zeta_i = H_i^{-1}$ is well-defined. Since $H_i' \geq 1$, its inverse is 1-Lipschitz. In particular,

$$|\zeta_i(t)| = |\zeta_i(t) - \zeta_i(H_i(0))| \leq |t - H_i(0)| \leq |t| + \gamma_i |L'(0)|.$$

Taking $t = x_i^\top \hat{\theta}_{-i}$, conditionally on X_{-i}, the vector $\hat{\theta}_{-i}$ is deterministic and independent of x_i. Hence Lemma 3.2 and Theorem 3.7 give

$$\sup_{i \in [n]} \left\| x_i^\top \hat{\theta}_{-i} \right\|_{L^k} = O_k(1).$$

Since $\gamma_i = O(1)$, $|L'(0)| = O(1)$, and

$$|L'(z)| \leq |L'(0)| + \left| L'' \right|_\infty |z|,$$

we obtain

$$\sup_{i \in [n]} \left\| L'\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big) \right\|_{L^k} = O_k(1),$$

which is the desired bound on s_i^ζ. □

Lemma 4.14 (Derivative of the averaged inverse score)**.** *Under Assumption 5, for each $i \in [n]$, the deterministic map*

$$t \mapsto \zeta_i(t), \qquad \zeta_i(t) + \gamma_i L'(\zeta_i(t)) = t,$$

is C^1, and

$$\zeta_i'(t) = \frac{1}{1 + \gamma_i L''(\zeta_i(t))}.$$

In particular, for $j \neq i$,

$$\mathbb{D}_j \Big[\zeta_i(x_i^\top \hat{\theta}_{-i}) \Big] [h] = \frac{x_i^\top \mathbb{D}_j[\hat{\theta}_{-i}][h]}{1 + \gamma_i L''\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big)},$$

and

$$\mathbb{D}_j[s_i^\zeta][h] = \frac{L''\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big)}{1 + \gamma_i L''\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big)} x_i^\top \mathbb{D}_j[\hat{\theta}_{-i}][h].$$

Proof. The derivative of ζ_i follows from the implicit function theorem applied to

$$F(z, t) := z + \gamma_i L'(z) - t.$$

Indeed,

$$\partial_z F(z, t) = 1 + \gamma_i L''(z) \geq 1,$$

so the implicit derivative is well-defined and gives

$$\zeta_i'(t) = \frac{1}{1 + \gamma_i L''(\zeta_i(t))}.$$

Since γ_i is deterministic, differentiating

$$\zeta_i(x_i^\top \hat{\theta}_{-i}) + \gamma_i L'\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big) = x_i^\top \hat{\theta}_{-i}$$

with respect to the column x_j, $j \neq i$, gives

$$\Big[1 + \gamma_i L''\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big)\Big] \mathbb{D}_j \Big[\zeta_i(x_i^\top \hat{\theta}_{-i})\Big] [h] = x_i^\top \mathbb{D}_j[\hat{\theta}_{-i}][h].$$

This proves the formula for $\mathbb{D}_j[\zeta_i(x_i^\top \hat{\theta}_{-i})]$. The formula for $\mathbb{D}_j[s_i^\zeta]$ follows from the chain rule. $\square$

Lemma 4.15 (Approximation of the train score)**.** *Under Assumptions 1–7, for every fixed $k \geq 1$,*

$$\sup_{i \in [n]} \left\| s_i - s_i^\zeta \right\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}}\right).$$

Proof. Set

$$\lambda_i := \frac{1}{n} x_i^\top \bar{G}_i x_i.$$

By Lemma 4.1,

$$x_i^\top \hat{\theta} = x_i^\top \hat{\theta}_{-i} - \lambda_i L'(x_i^\top \hat{\theta}).$$

On the other hand, by definition of $\zeta_i(x_i^\top \hat{\theta}_{-i})$,

$$\zeta_i(x_i^\top \hat{\theta}_{-i}) = x_i^\top \hat{\theta}_{-i} - \gamma_i L'\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big).$$

Let

$$H_i^\lambda(z) := z + \lambda_i L'(z).$$

Since

$$\lambda_i = \frac{1}{n} x_i^\top \bar{G}_i x_i \geq 0$$

and $L'' \geq \kappa_L > 0$, we have

$$(H_i^\lambda)'(z) = 1 + \lambda_i L''(z) \geq 1.$$

Hence H_i^λ is increasing with inverse Lipschitz constant at most one: for all $z, z' \in \mathbb{R}$,

$$|z - z'| \leq |H_i^\lambda(z) - H_i^\lambda(z')|.$$

By Lemma 4.1,

$$H_i^\lambda(x_i^\top \hat{\theta}) = x_i^\top \hat{\theta}_{-i}.$$

On the other hand, by definition of ζ_i,

$$\begin{aligned} H_i^\lambda\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big) &= \zeta_i(x_i^\top \hat{\theta}_{-i}) + \lambda_i L'\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big) \\ &= x_i^\top \hat{\theta}_{-i} + (\lambda_i - \gamma_i) L'\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big). \end{aligned}$$

Consequently,

$$\Big| x_i^\top \hat{\theta} - \zeta_i(x_i^\top \hat{\theta}_{-i}) \Big| \leq \Big| H_i^\lambda(x_i^\top \hat{\theta}) - H_i^\lambda\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big) \Big| = |\lambda_i - \gamma_i| \Big| L'\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big) \Big|.$$

Since the map $z \mapsto z + \lambda_i L'(z)$ has derivative at least one,

$$\Big| x_i^\top \hat{\theta} - \zeta_i(x_i^\top \hat{\theta}_{-i}) \Big| \leq |\lambda_i - \gamma_i| \; \Big| L'\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big) \Big|.$$

Therefore,

$$|s_i - s_i^\zeta| = \left|L'(x_i^\top \hat\theta) - L'\Big(\zeta_i(x_i^\top \hat\theta_{-i})\Big)\right| \le \left|L''\right|_\infty \left|x_i^\top \hat\theta - \zeta_i(x_i^\top \hat\theta_{-i})\right| \le O(1)|\lambda_i - \gamma_i|\,|s_i^\zeta|.$$

It remains to bound $\lambda_i - \gamma_i$. We decompose

$$\lambda_i - \gamma_i = \left(\frac{1}{n} x_i^\top G_{-i} x_i - \gamma_i\right) + \frac{1}{n} x_i^\top (\bar G_i - G_{-i}) x_i.$$

The first term is bounded by Lemma 4.12:

$$\left\|\frac{1}{n} x_i^\top G_{-i} x_i - \gamma_i\right\|_{L^k} = O_k\left(\frac{1}{\sqrt{n}}\right).$$

For the second term, use

$$\bar G_i - G_{-i} = -\bar G_i R_i G_{-i}.$$

The regularizer contribution is bounded by

$$\left|\frac{1}{n} x_i^\top \bar G_i R_i^\rho G_{-i} x_i\right| \le \frac{1}{n}\, |x_i|_2^2 \left|\bar G_i\right|_{\mathrm{op}} |R_i^\rho|_{\mathrm{op}}\, |G_{-i}|_{\mathrm{op}} \le \frac{O(1)}{n\sqrt{n}} |s_i^-|\, |x_i|_2^2\,,$$

where we used (4.4). Hence

$$\left\|\frac{1}{n} x_i^\top \bar G_i R_i^\rho G_{-i} x_i\right\|_{L^k} = O_k(n^{-1/2}).$$

For the loss contribution,

$$\left|\frac{1}{n} x_i^\top \bar G_i R_i^L G_{-i} x_i\right| \le \frac{1}{n^2} \left|X_{-i}^\top \bar G_i x_i\right|_2 |\Gamma_i|_F \left|X_{-i}^\top G_{-i} x_i\right|_\infty .$$

By Lemma 4.2,

$$\left|X_{-i}^\top \bar G_i x_i\right|_2 \le \left|\bar G_i X_{-i}\right|_{\mathrm{op}} |x_i|_2 = O(\sqrt{n})\, |x_i|_2\,.$$

Using Lemmas 3.2, 4.5, and 4.4, we get

$$\left\|\frac{1}{n} x_i^\top \bar G_i R_i^L G_{-i} x_i\right\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}}\right).$$

Thus

$$\|\lambda_i - \gamma_i\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}}\right).$$

Since $\left\|s_i^\zeta\right\|_{L^{2k}} = O_k(1)$ by Lemma 4.13, Hölder's inequality gives

$$\left\|s_i - s_i^\zeta\right\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}}\right).$$

□

From now on, in the scalar leave-one-out approximation, we use the averaged first-order displacement

$$\delta_i^{(0)} \hat\theta := -\frac{s_i^\zeta}{n} G_{-i} x_i.$$

Lemma 4.16 (First-order bounds with deterministic averaged score)**.** *Under Assumptions 1–7, for every fixed* $k \geq 1$*,*

$$\sup_{i\in[n]} \left\| \left| \delta_i\hat{\theta} - \delta_i^{(0)}\hat{\theta} \right|_2 \right\|_{L^k} = O_k\left(\frac{\log n}{n}\right),$$

and

$$\sup_{i\in[n]} \left\| \left| X_{-i}^\top(\delta_i\hat{\theta} - \delta_i^{(0)}\hat{\theta}) \right|_2 \right\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}}\right).$$

Proof. Write

$$\delta_i\hat{\theta} - \delta_i^{(0)}\hat{\theta} = \delta_i\hat{\theta} - \delta_{i,\mathrm{tr}}^{(0)}\hat{\theta} - \frac{s_i - s_i^\zeta}{n} G_{-i}x_i.$$

By Lemma 4.15 and Lemma 3.2,

$$\left\| \frac{s_i - s_i^\zeta}{n} G_{-i}x_i \right\|_{L^k} = O_k\left(\frac{\log n}{n}\right).$$

Combining this with Theorem 4.9 gives the first bound.

For the projected bound,

$$\left| X_{-i}^\top \frac{s_i - s_i^\zeta}{n} G_{-i}x_i \right|_2 \leq \frac{|s_i - s_i^\zeta|}{n} \left| X_{-i}^\top G_{-i}x_i \right|_2.$$

Lemma 4.15 and Lemma 4.4 give

$$\left\| \left| X_{-i}^\top \frac{s_i - s_i^\zeta}{n} G_{-i}x_i \right|_2 \right\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}}\right).$$

The projected estimate follows again from Theorem 4.9. □

Corollary 4.17 (First-order bounds with deterministic averaged score)**.** *Under Assumptions 1–7, for every fixed* $k \geq 1$*,*

$$\sup_{i\in[n]} \left\| \left| X_{-i}^\top \delta_i^{(0)}\hat{\theta} \right|_2 \right\|_{L^k} = O_k(1), \qquad \textit{and} \qquad \sup_{i\in[n]} \left\| \left| X_{-i}^\top \delta_i^{(0)}\hat{\theta} \right|_\infty \right\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}}\right).$$

Consequently,

$$\sup_{i\in[n]} \left\| \left| X_{-i}^\top \delta_i\hat{\theta} \right|_\infty \right\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}}\right).$$

Proof. Let us bound:

$$\left| X_{-i}^\top \delta_i^{(0)}\hat{\theta} \right|_2 \leq \frac{|s_i^\zeta|}{n} \left| X_{-i}^\top G_{-i}x_i \right|_2,$$

and

$$\left| X_{-i}^\top \delta_i^{(0)}\hat{\theta} \right|_\infty \leq \frac{|s_i^\zeta|}{n} \left| X_{-i}^\top G_{-i}x_i \right|_\infty.$$

The L^k-bounds follow from Lemma 4.13 and Lemma 4.4. The final bound for $X_{-i}^\top\delta_i\hat{\theta}$ follows by adding the projected error bound proved above.

□

4.3 Second-order expansion

In this subsection we work under Assumptions 1–7 and under the supplementary condition $\nabla^3 \rho \equiv 0$. Thus $\nabla^2 \rho$ is constant and $R_i^\rho = 0$.

Define the diagonal matrix $\tilde{\Gamma}_i$, indexed by $j \neq i$, by

$$(\tilde{\Gamma}_i)_j := \frac{1}{2} L'''(x_j^\top \hat{\theta}_{-i})\, x_j^\top \delta_i^{(0)} \hat{\theta} = -\frac{s_i^\zeta}{2n} L'''(x_j^\top \hat{\theta}_{-i})\, x_j^\top G_{-i} x_i,$$

and set

$$\tilde{R}_i := \frac{1}{n} X_{-i} \tilde{\Gamma}_i X_{-i}^\top.$$

The second-order correction is

$$\delta_i^{(1)} \hat{\theta} := \frac{s_i^\zeta}{n} G_{-i} \tilde{R}_i G_{-i} x_i.$$

Lemma 4.18 (Second-order curvature remainder with deterministic averaged score)**.** *Let $E_i := \Gamma_i - \tilde{\Gamma}_i$, equivalently*

$$R_i^L - \tilde{R}_i = \frac{1}{n} X_{-i} E_i X_{-i}^\top.$$

Then, for every fixed $k \geq 1$,

$$\sup_{i \in [n]} \| |E_i|_F \|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}} \right).$$

Moreover,

$$\sup_{i \in [n]} \left\| \left| \tilde{\Gamma}_i \right|_F \right\|_{L^k} = O_k(1),$$

and

$$\sup_{i \in [n]} \left\| |\Gamma_i|_{\mathrm{op}} \right\|_{L^k} + \sup_{i \in [n]} \left\| \left| \tilde{\Gamma}_i \right|_{\mathrm{op}} \right\|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}} \right).$$

Proof. By Taylor's formula and the Lipschitz bound on L''',

$$|(E_i)_j| \leq O(1) |x_j^\top (\delta_i \hat{\theta} - \delta_i^{(0)} \hat{\theta})| + O(1) |x_j^\top \delta_i \hat{\theta}|^2.$$

Therefore

$$|E_i|_F \leq O(1) \left| X_{-i}^\top (\delta_i \hat{\theta} - \delta_i^{(0)} \hat{\theta}) \right|_2 + O(1) \left| X_{-i}^\top \delta_i \hat{\theta} \right|_\infty \left| X_{-i}^\top \delta_i \hat{\theta} \right|_2.$$

The first term is $O_k((\log n)/\sqrt{n})$ by Lemma 4.16. The second has the same order by Lemma 4.16, Corollary 4.17, and Hölder's inequality. Hence

$$\sup_{i \in [n]} \| |E_i|_F \|_{L^k} = O_k\left(\frac{\log n}{\sqrt{n}} \right).$$

The Frobenius bound follows from

$$\left| \tilde{\Gamma}_i \right|_F \leq \frac{O(1) |s_i^\zeta|}{n} \left| X_{-i}^\top G_{-i} x_i \right|_2,$$

and the operator bounds follows from

$$|\Gamma_i|_{\mathrm{op}} \leq O(1) \left| X_{-i}^\top \delta_i \hat{\theta} \right|_\infty \qquad \text{and} \qquad \left| \tilde{\Gamma}_i \right|_{\mathrm{op}} \leq \frac{O(1) |s_i^\zeta|}{n} \left| X_{-i}^\top G_{-i} x_i \right|_\infty.$$

We conclude using Lemmas 4.16, 4.13, 4.4 and Corollary 4.17. □

Theorem 4.19 (Projected second-order expansion with deterministic averaged score)**.** *Under Assumptions 1–7 and $\nabla^3\rho \equiv 0$, for every fixed $k \geq 1$ and every deterministic $u \in \mathbb{R}^p$ with $|u|_2 \leq 1$,*

$$\sup_{i\in[n]} \left\| u^\top \left(\delta_i\hat{\theta} - \delta_i^{(0)}\hat{\theta} - \delta_i^{(1)}\hat{\theta} \right) \right\|_{L^k} = O_k\left(\frac{(\log n)^2}{n^{3/2}} \right).$$

Furthermore,

$$\sup_{i\in[n]} \left\| \left| \delta_i\hat{\theta} - \delta_i^{(0)}\hat{\theta} - \delta_i^{(1)}\hat{\theta} \right|_2 \right\|_{L^k} = O_k\left(\frac{\log n}{n} \right).$$

Proof. Since $\nabla^3\rho \equiv 0$, $R_i = R_i^L$. Starting from

$$\delta_i\hat{\theta} = -\frac{s_i}{n}\bar{G}_i x_i, \qquad \delta_i^{(0)}\hat{\theta} = -\frac{s_i^\zeta}{n} G_{-i} x_i,$$

and using

$$\bar{G}_i - G_{-i} = -\bar{G}_i R_i^L G_{-i},$$

we get

$$\begin{aligned} \delta_i\hat{\theta} - \delta_i^{(0)}\hat{\theta} - \delta_i^{(1)}\hat{\theta} = &-\frac{s_i - s_i^\zeta}{n} G_{-i} x_i \\ &+ \frac{s_i}{n}\bar{G}_i (R_i^L - \tilde{R}_i) G_{-i} x_i \\ &+ \frac{s_i}{n}(\bar{G}_i - G_{-i})\tilde{R}_i G_{-i} x_i \\ &+ \frac{s_i - s_i^\zeta}{n} G_{-i}\tilde{R}_i G_{-i} x_i. \end{aligned}$$

We bound the deterministic projection of these four terms.

First,

$$\left| \frac{s_i - s_i^\zeta}{n} u^\top G_{-i} x_i \right| \leq \frac{|s_i - s_i^\zeta|}{n} |u^\top G_{-i} x_i|.$$

By Lemma 4.15 and Lemma 3.2,

$$\left\| \frac{s_i - s_i^\zeta}{n} u^\top G_{-i} x_i \right\|_{L^k} = O_k\left(\frac{\log n}{n^{3/2}} \right).$$

Second, since $R_i^L - \tilde{R}_i = n^{-1} X_{-i} E_i X_{-i}^\top$,

$$\left| \frac{s_i}{n} u^\top \bar{G}_i (R_i^L - \tilde{R}_i) G_{-i} x_i \right| \leq \frac{|s_i|}{n^2} \left| X_{-i}^\top \bar{G}_i u \right|_2 |E_i|_F \left| X_{-i}^\top G_{-i} x_i \right|_\infty = O_k\left(\frac{(\log n)^2}{n^{3/2}} \right),$$

using $\left| X_{-i}^\top \bar{G}_i u \right|_2 = O(\sqrt{n})$ and Lemmas 4.3, 4.18, 4.4.

For the third term, use

$$\bar{G}_i - G_{-i} = -\frac{1}{n}\bar{G}_i X_{-i}\Gamma_i X_{-i}^\top G_{-i}, \qquad \tilde{R}_i = \frac{1}{n} X_{-i}\tilde{\Gamma}_i X_{-i}^\top.$$

Then

$$\begin{aligned} &\left| \frac{s_i}{n} u^\top (\bar{G}_i - G_{-i})\tilde{R}_i G_{-i} x_i \right| \\ &\quad \leq \frac{|s_i|}{n^3} \left| X_{-i}^\top \bar{G}_i u \right|_2 |\Gamma_i|_{\mathrm{op}} \left| X_{-i}^\top G_{-i} X_{-i} \right|_{\mathrm{op}} \left| \tilde{\Gamma}_i \right|_F \left| X_{-i}^\top G_{-i} x_i \right|_\infty. \end{aligned}$$

By Lemmas 4.3, 4.2, 4.18, and 4.4, this term is

$$O_k\left(\frac{(\log n)^2}{n^{3/2}} \right)$$

in L^k.

For the fourth term,

$$\left| \frac{s_i - s_i^\zeta}{n} u^\top G_{-i} \tilde{R}_i G_{-i} x_i \right| \le \frac{|s_i - s_i^\zeta|}{n^2} \left| X_{-i}^\top G_{-i} u \right|_2 \left| \tilde{\Gamma}_i \right|_{\mathrm{op}} \left| X_{-i}^\top G_{-i} x_i \right|_2 .$$

Using

$$\left| X_{-i}^\top G_{-i} u \right|_2 = O(\sqrt{n}), \qquad \left\| \left| X_{-i}^\top G_{-i} x_i \right|_2 \right\|_{L^k} = O_k(n),$$

together with Lemmas 4.15 and 4.18, we get

$$\left\| \frac{s_i - s_i^\zeta}{n} u^\top G_{-i} \tilde{R}_i G_{-i} x_i \right\|_{L^k} = O_k\left(\frac{(\log n)^2}{n^{3/2}} \right).$$

This proves the projected estimate.

The vector estimate follows from the same decomposition. The first term is bounded by

$$\left\| \frac{s_i - s_i^\zeta}{n} |G_{-i} x_i|_2 \right\|_{L^k} = O_k\left(\frac{\log n}{n} \right),$$

and the other three terms are $O_k((\log n)^2/n^{3/2})$ in vector norm by the same estimates as above, replacing $\left| X_{-i}^\top \bar{G}_i u \right|_2$ by $\left| \bar{G}_i X_{-i} \right|_{\mathrm{op}}$ that has the same order. Hence the vector remainder is $O_k((\log n)/n)$. □

Remark 4.20 (Consequence for the Wasserstein bound)**.** Define

$$\begin{aligned} \mathfrak{d}_i(X) &:= \sqrt{n}\, u^\top \left(\delta_i^{(0)} \hat{\theta}(X) + \delta_i^{(1)} \hat{\theta}(X) \right) \\ &= -\frac{s_i^\zeta}{\sqrt{n}} u^\top G_{-i} x_i + \frac{s_i^\zeta}{n^{3/2}} u^\top G_{-i} X_{-i} \tilde{\Gamma}_i X_{-i}^\top G_{-i} x_i. \end{aligned}$$

By Theorem 4.19,

$$\sup_{i \in [n]} \|\delta_i f_n(X) - \mathfrak{d}_i(X)\|_{L^4} = O\left(\frac{(\log n)^2}{n} \right).$$

The same estimate holds with X replaced by any X^A, since X^A has the same distribution as X. Consequently,

$$\left| c_n^x - \sup_{\substack{i,j \in [n] \\ i \ne j}} \left\| \mathfrak{d}_i(X) - \mathbb{E}_j\left[\mathfrak{d}_i(X) \right] \right\|_{L^4} \right| = O\left(\frac{(\log n)^2}{n} \right).$$

Moreover,

$$\sup_{j \in [n]} \|\delta_j f_n(X)\|_{L^4} + \sup_{j \in [n]} \|\mathfrak{d}_j(X)\|_{L^4} = O\left(\frac{\log n}{\sqrt{n}} \right).$$

Using conditional Cauchy–Schwarz and Jensen's inequality, the two preceding bounds imply, uniformly in j,

$$\left\| \mathrm{Cov}_j\left(\delta_j f_n(X), \delta_j f_n(X^{A_j}) \right) - \mathrm{Cov}_j\left(\mathfrak{d}_j(X), \mathfrak{d}_j(X^{A_j}) \right) \right\|_{L^2} = O\left(\frac{(\log n)^3}{n^{3/2}} \right).$$

Therefore

$$\left| c_n^{\bar{x}} - \sup_{j \in [n]} \sqrt{\mathrm{Var}\left(\mathrm{Cov}_j\left(\mathfrak{d}_j(X), \mathfrak{d}_j(X^{A_j}) \right) \right)} \right| \le O\left(\frac{(\log n)^3}{n^{3/2}} \right).$$

5 Sensitivity analysis for the cross-covariance terms

5.1 Notation and preliminary bounds

We now prove the estimate needed to control c_n^x. The estimates are first written in a generic form, without the leave-one-out index. They apply to the leave-one-out quantities by replacing

$$X, \hat{\theta}, G, \tilde{\Gamma}_x, x, s_x^\zeta \quad \text{with} \quad X_{-i}, \hat{\theta}_{-i}, G_{-i}, \tilde{\Gamma}_i, x_i, s_i^\zeta.$$

This replacement does not change the estimates: the columns remain independent, the normalization is still $1/n$, the curvature lower bounds are unchanged, and all moment and Poincaré constants are uniform. Throughout this section $r \geq 1$ is fixed, and we write $\log n$ instead of $\log(en)$, which is equivalent for our asymptotic bounds.

Let $x \in \mathbb{R}^p$ be an auxiliary column, independent of X, satisfying the same columnwise assumptions as the columns of X. We fix a deterministic scalar γ_x satisfying

$$0 \leq \gamma_x = O(1),$$

corresponding in the leave-one-out application to

$$\gamma_i = \mathbb{E}\left[\frac{1}{n} x_i^\top G_{-i} x_i\right].$$

With this convention, the map

$$z \mapsto z + \gamma_x L'(z)$$

has derivative $1 + \gamma_x L''(z) \geq 1$, and hence is a bijection of $\mathbb{R}$ onto $\mathbb{R}$. Thus $\zeta_x : \mathbb{R} \to \mathbb{R}$ satisfying

$$\zeta_x(t) + \gamma_x L'(\zeta_x(t)) = t,$$

is well-defined and 1-Lipschitz. Further set

$$s_x^\zeta := L'\Big(\zeta_x(x^\top \hat{\theta})\Big).$$

Under the replacement

$$(X, \hat{\theta}, G, x, \gamma_x) \mapsto (X_{-i}, \hat{\theta}_{-i}, G_{-i}, x_i, \gamma_i),$$

this is precisely

$$s_i^\zeta = L'\Big(\zeta_i(x_i^\top \hat{\theta}_{-i})\Big).$$

Throughout the generic estimates, set

$$G := \Big(\nabla^2 \Psi_X(\hat{\theta})\Big)^{-1}, \qquad s_j := L'(x_j^\top \hat{\theta}), \qquad j \in [n].$$

Here s_j denotes the usual train score of the background sample X; it is not replaced by an averaged score.

We introduce the diagonal matrix $\tilde{\Gamma}_x$, indexed by the columns of X, by

$$(\tilde{\Gamma}_x)_\ell := -\frac{s_x^\zeta}{2n} L'''(x_\ell^\top \hat{\theta})\, x_\ell^\top G x.$$

With $X = X_{-i}$ and $x = x_i$, this is exactly the matrix $\tilde{\Gamma}_i$ used in the second-order leave-one-out expansion.

We shall repeatedly use the following deterministic consequences of the curvature argument in Lemma 4.2:

$$|G|_{\text{op}} = O(1), \qquad |GX|_{\text{op}} = \Big|X^\top G\Big|_{\text{op}} = O(\sqrt{n}), \qquad \Big|X^\top G X\Big|_{\text{op}} = O(n).$$

Similarly, the generic version of Lemma 4.4 gives, for every fixed $r \geq 1$,

$$\||Gx|_2\|_{L^r} = O_r(\sqrt{n}), \qquad \left\|\left|X^\top Gx\right|_2\right\|_{L^r} = O_r(n), \qquad \left\|\left|X^\top Gx\right|_\infty\right\|_{L^r} = O_r(\sqrt{n}\log n).$$

Moreover, by Lemma 4.13,

$$\left\|s_x^\zeta\right\|_{L^r} = O_r(1).$$

Lemma 5.1 (Generic leverage bounds)**.** *Under Assumptions 1–7 and $\nabla^3\rho \equiv 0$, for every fixed integer $r \geq 1$, and every deterministic $u \in \mathbb{R}^p$ with $|u|_2 \leq 1$,*

$$\left\|u^\top Gx\right\|_{L^r} = O_r(1), \qquad \left\|\left|X^\top Gu\right|_2\right\|_{L^r} = O_r(\sqrt{n}).$$

Proof. Conditionally on X, Gu is deterministic and $|Gu|_2 = O(1)$. Lemma 3.2 gives

$$\left\|u^\top Gx\right\|_{L^r_x} = O_r(1),$$

and the first estimate follows by taking the L^r-norm over X. The second estimate is deterministic:

$$\left|X^\top Gu\right|_2 \leq \left|X^\top G\right|_{\mathrm{op}} |u|_2 = O(\sqrt{n}).$$

□

Lemma 5.2 (Bounds involving $\tilde{\Gamma}_x$)**.** *Under Assumptions 1–7 and $\nabla^3\rho \equiv 0$, for every fixed integer $r \geq 1$,*

$$\left\|\left|\tilde{\Gamma}_x\right|_{\mathrm{op}}\right\|_{L^r} = O_r\left(\frac{\log n}{\sqrt{n}}\right), \qquad \left\|\left|\tilde{\Gamma}_x X^\top Gx\right|_2\right\|_{L^r} = O_r(\sqrt{n}\log n),$$

and

$$\left\|\left|X^\top GX\tilde{\Gamma}_x X^\top Gx\right|_2\right\|_{L^r} = O_r(n^{3/2}\log n).$$

Proof. By the definition of $\tilde{\Gamma}_x$ and the boundedness of L''',

$$\left|\tilde{\Gamma}_x\right|_{\mathrm{op}} \leq \frac{O(1)|s_x^\zeta|}{n}\left|X^\top Gx\right|_\infty.$$

Using $\left\|s_x^\zeta\right\|_{L^r} = O_r(1)$ and the generic cross-leverage bound, we obtain

$$\left\|\left|\tilde{\Gamma}_x\right|_{\mathrm{op}}\right\|_{L^r} = O_r\left(\frac{\log n}{\sqrt{n}}\right).$$

Next,

$$\left|\tilde{\Gamma}_x X^\top Gx\right|_2 \leq \left|\tilde{\Gamma}_x\right|_{\mathrm{op}}\left|X^\top Gx\right|_2,$$

and the second estimate follows from Hölder's inequality and $\||X^\top Gx|_2\|_{L^r} = O_r(n)$. Finally,

$$\left|X^\top GX\tilde{\Gamma}_x X^\top Gx\right|_2 \leq \left|X^\top GX\right|_{\mathrm{op}}\left|\tilde{\Gamma}_x X^\top Gx\right|_2,$$

and the deterministic bound $\left|X^\top GX\right|_{\mathrm{op}} = O(n)$ gives the last estimate. □

Lemma 5.3 (Uniform background drift)**.** *Under Assumptions 1–7 and $\nabla^3 \rho \equiv 0$, define, for $j \in [n]$,*

$$\Delta_j := \mathrm{diag}_{\ell \neq j} \left(L''(x_\ell^\top \hat{\theta}) - L''(x_\ell^\top \hat{\theta}_{-j}) \right).$$

Then, for every fixed integer $r \geq 1$,

$$\left\| \sup_{j \in [n]} |\Delta_j|_{\mathrm{op}} \right\|_{L^r} = O_r\left(\frac{(\log n)^6}{\sqrt{n}} \right).$$

Proof. By the Lipschitz property of L'',

$$|\Delta_j|_{\mathrm{op}} \leq O(1) \left| X_{-j}^\top (\hat{\theta} - \hat{\theta}_{-j}) \right|_\infty.$$

Using the train-score first-order term

$$\delta_{j,\mathrm{tr}}^{(0)} \hat{\theta} := -\frac{s_j}{n} G_{-j} x_j,$$

we get

$$\sup_{j \in [n]} |\Delta_j|_{\mathrm{op}} \leq O(1) \sup_{j \in [n]} \left| X_{-j}^\top \delta_{j,\mathrm{tr}}^{(0)} \hat{\theta} \right|_\infty + O(1) \sup_{j \in [n]} \left| X_{-j}^\top \left(\delta_j \hat{\theta} - \delta_{j,\mathrm{tr}}^{(0)} \hat{\theta} \right) \right|_2.$$

For the leading term,

$$\left| X_{-j}^\top \delta_{j,\mathrm{tr}}^{(0)} \hat{\theta} \right|_\infty \leq \frac{|s_j|}{n} \left| X_{-j}^\top G_{-j} x_j \right|_\infty.$$

The sequence-level score bound from Lemma 3.11 gives

$$\forall q \geq 1: \qquad \sup_{j \in [n]} \|s_j\|_{L^q} = O(q^2).$$

Thus Lemma B.7 gives

$$\left\| \sup_{j \in [n]} |s_j| \right\|_{L^r} = O_r((\log n)^2).$$

Moreover, conditionally on X_{-j}, the variables $(x_\ell^\top G_{-j} x_j)_{\ell \neq j}$ are linear forms in x_j, with coefficients of Euclidean norm at most $O(\sqrt{n})$. Therefore,

$$\forall q \geq 1: \qquad \sup_{\substack{j,\ell \in [n] \\ \ell \neq j}} \left\| x_\ell^\top G_{-j} x_j \right\|_{L^q} \leq O(q\sqrt{n}).$$

Applying Lemma B.7 to the $O(n^2)$ family $(x_\ell^\top G_{-j} x_j)_{\ell \neq j}$, we obtain

$$\left\| \sup_{j \in [n]} \left| X_{-j}^\top G_{-j} x_j \right|_\infty \right\|_{L^r} = O_r(\sqrt{n} \log n).$$

Hence

$$\left\| \sup_{j \in [n]} \left| X_{-j}^\top \delta_{j,\mathrm{tr}}^{(0)} \hat{\theta} \right|_\infty \right\|_{L^r} = O_r\left(\frac{(\log n)^3}{\sqrt{n}} \right).$$

It remains to control the first-order remainder uniformly in j. Under $\nabla^3 \rho \equiv 0$, the regularizer drift vanishes, and the proof of Theorem 4.9 gives

$$\left| X_{-j}^\top \left(\delta_j \hat{\theta} - \delta_{j,\mathrm{tr}}^{(0)} \hat{\theta} \right) \right|_2 \leq \frac{|s_j|}{n^2} \left| X_{-j}^\top \bar{G}_j X_{-j} \right|_{\mathrm{op}} |\Gamma_j|_F \left| X_{-j}^\top G_{-j} x_j \right|_\infty.$$

By Lemma 4.2,

$$\left|X_{-j}^\top \bar{G}_j X_{-j}\right|_{\mathrm{op}} = O(n).$$

Moreover, as in the proof of the Frobenius bound for Γ_j,

$$|\Gamma_j|_F \le O(1)\left|X_{-j}^\top \delta_j\hat{\theta}\right|_2 \le \frac{O(1)|s_j|}{n}\left|X_{-j}^\top \bar{G}_j\right|_{\mathrm{op}} |x_j|_2 \le O(1)\frac{|s_j|}{\sqrt{n}}|x_j|_2\,.$$

Using

$$\left\|\sup_{j\in[n]}|s_j|\right\|_{L^r} = O_r((\log n)^2),$$

and

$$\left\|\sup_{j\in[n]}|x_j|_2\right\|_{L^r} + \left\|\sup_{j\in[n]}\left|X_{-j}^\top G_{-j}x_j\right|_\infty\right\|_{L^r} = O_r(\sqrt{n}\log n),$$

Hölder's inequality yields

$$\left\|\sup_{j\in[n]}\left|X_{-j}^\top\left(\delta_j\hat{\theta} - \delta_{j,\mathrm{tr}}^{(0)}\hat{\theta}\right)\right|_2\right\|_{L^r} = O_r\left(\frac{(\log n)^6}{\sqrt{n}}\right).$$

Combining the leading term and the first-order remainder gives the claim. □

Lemma 5.4 (Fixed-direction and auxiliary-column full-resolvent leverage)**.** *Under Assumptions 1–7 and $\nabla^3\rho\equiv 0$, let $u\in\mathbb{R}^p$ be deterministic with $|u|_2\le 1$. Then, for every fixed integer $r\ge 1$,*

$$\left\|\left|X^\top G u\right|_\infty\right\|_{L^r} = O_r\big((\log n)^6\big) \qquad \textit{and} \qquad \left\|\left|X^\top G x\right|_\infty\right\|_{L^r} = O_r(\sqrt{n}\log n).$$

Proof. We first prove the fixed-direction bound. Fix $j\in[n]$. We use the decomposition

$$G^{-1} = G_{-j}^{-1} + B_j^{\mathrm{bg}} + B_j^{\mathrm{loc}},$$

where

$$B_j^{\mathrm{loc}} = \frac{1}{n}L''(x_j^\top\hat{\theta})x_jx_j^\top = a_jx_jx_j^\top, \qquad a_j := \frac{1}{n}L''(x_j^\top\hat{\theta}) \ge \frac{\kappa_L}{n},$$

and

$$B_j^{\mathrm{bg}} = \frac{1}{n}X_{-j}\Delta_jX_{-j}^\top.$$

Set

$$\widetilde{G}_j := \left(G_{-j}^{-1} + B_j^{\mathrm{bg}}\right)^{-1}.$$

Since $\nabla^3\rho\equiv 0$,

$$\widetilde{G}_j^{-1} = \nabla^2\Psi_X^{-j}(\hat{\theta}) \succeq \kappa I_p + \frac{\kappa_L}{n}X_{-j}X_{-j}^\top.$$

By the Sherman–Morrison formula,

$$G = \widetilde{G}_j - \frac{a_j\widetilde{G}_jx_jx_j^\top\widetilde{G}_j}{1 + a_jx_j^\top\widetilde{G}_jx_j}.$$

Therefore

$$x_j^\top G u = \frac{x_j^\top\widetilde{G}_ju}{1 + a_jx_j^\top\widetilde{G}_jx_j}.$$

Using the resolvent identity

$$\widetilde{G}_j - G_{-j} = -\widetilde{G}_jB_j^{\mathrm{bg}}G_{-j},$$

we obtain

$$x_j^\top G u = \frac{x_j^\top G_{-j} u}{1 + a_j x_j^\top \widetilde{G}_j x_j} - \frac{x_j^\top \widetilde{G}_j B_j^{\mathrm{bg}} G_{-j} u}{1 + a_j x_j^\top \widetilde{G}_j x_j}.$$

Since the denominator is at least one,

$$|x_j^\top G u| \le |x_j^\top G_{-j} u| + \frac{|x_j^\top \widetilde{G}_j B_j^{\mathrm{bg}} G_{-j} u|}{1 + a_j x_j^\top \widetilde{G}_j x_j}.$$

For the first term, conditionally on X_{-j}, $G_{-j}u$ is deterministic and has Euclidean norm $O(1)$. Hence, for every integer $q \ge 1$,

$$\sup_{j \in [n]} \left\| x_j^\top G_{-j} u \right\|_{L^q} = O(q).$$

Lemma B.7 gives

$$\left\| \sup_{j \in [n]} |x_j^\top G_{-j} u| \right\|_{L^r} = O_r(\log n).$$

For the background correction, the curvature lower bound on $\widetilde{G}_j^{-1}$ gives

$$\widetilde{G}_j X_{-j} X_{-j}^\top \widetilde{G}_j \preceq \frac{n}{\kappa_L} \widetilde{G}_j.$$

Thus

$$\left| X_{-j}^\top \widetilde{G}_j x_j \right|_2^2 = x_j^\top \widetilde{G}_j X_{-j} X_{-j}^\top \widetilde{G}_j x_j \le \frac{n}{\kappa_L} x_j^\top \widetilde{G}_j x_j.$$

Since $a_j \ge \kappa_L / n$, for $t := x_j^\top \widetilde{G}_j x_j \ge 0$,

$$\frac{\left| X_{-j}^\top \widetilde{G}_j x_j \right|_2}{1 + a_j x_j^\top \widetilde{G}_j x_j} \le \sqrt{\frac{n}{\kappa_L}} \frac{\sqrt{t}}{1 + (\kappa_L/n) t} \le O(n).$$

Therefore

$$\frac{|x_j^\top \widetilde{G}_j B_j^{\mathrm{bg}} G_{-j} u|}{1 + a_j x_j^\top \widetilde{G}_j x_j} = \frac{1}{n} \frac{|x_j^\top \widetilde{G}_j X_{-j} \Delta_j X_{-j}^\top G_{-j} u|}{1 + a_j x_j^\top \widetilde{G}_j x_j} \le O(1) \, |\Delta_j|_{\mathrm{op}} \left| X_{-j}^\top G_{-j} u \right|_2.$$

Moreover,

$$\left| X_{-j}^\top G_{-j} u \right|_2 \le \left| X_{-j}^\top G_{-j} \right|_{\mathrm{op}} |u|_2 = O(\sqrt{n}).$$

Hence

$$\sup_{j \in [n]} \frac{|x_j^\top \widetilde{G}_j B_j^{\mathrm{bg}} G_{-j} u|}{1 + a_j x_j^\top \widetilde{G}_j x_j} \le O(\sqrt{n}) \sup_{j \in [n]} |\Delta_j|_{\mathrm{op}}.$$

Lemma 5.3 gives

$$\left\| \sup_{j \in [n]} \frac{|x_j^\top \widetilde{G}_j B_j^{\mathrm{bg}} G_{-j} u|}{1 + a_j x_j^\top \widetilde{G}_j x_j} \right\|_{L^r} = O_r((\log n)^6).$$

Combining the leave-one-out term and the background correction proves

$$\left\| \left| X^\top G u \right|_\infty \right\|_{L^r} = O_r((\log n)^6).$$

We now prove the auxiliary-column estimate. Since x is independent of X, we condition on X. Then G and the vectors $(Gx_j)_{j \in [n]}$ are fixed, and

$$(X^\top G x)_j = x_j^\top G x = x^\top G x_j.$$

By the curvature bound,

$$\max_{j\in[n]} |Gx_j|_2 \le |GX|_{\mathrm{op}} = O(\sqrt{n}).$$

Therefore, conditionally on X, Lemma 3.2 gives, for every integer $q \ge 1$,

$$\max_{j\in[n]} \left\| x^\top G x_j \right\|_{L^q_x} \le O(q\sqrt{n}).$$

Applying Lemma B.7 conditionally on X, we obtain

$$\left\| \left| X^\top G x \right|_\infty \right\|_{L^r_x} = \left\| \max_{j\in[n]} |x^\top G x_j| \right\|_{L^r_x} \le O_r(\sqrt{n}\log n).$$

The right-hand side is deterministic, so taking the L^r-norm over the background variables proves

$$\left\| \left| X^\top G x \right|_\infty \right\|_{L^r} = O_r(\sqrt{n}\log n).$$

□

5.2 Sensitivity to column-wise differentiation

Recall from (3.1) that

$$\mathbb{D}_j[\hat\theta][h] = -\frac{1}{n} G \left[L''(x_j^\top \hat\theta)(h^\top \hat\theta) x_j + s_j h \right].$$

Lemma 5.5 (Derivative of the minimizer)**.** *Under Assumptions 1–5, for every fixed integer $r \ge 1$,*

$$\sup_{j\in[n]} \left\| \left| \mathbb{D}_j[\hat\theta] \right|_2^* \right\|_{L^r} = O_r\left(\frac{1}{\sqrt{n}} \right),$$

and

$$\sup_{j\in[n]} \left\| \left| X^\top \mathbb{D}_j[\hat\theta] \right|_2^* \right\|_{L^r} = O_r(1).$$

Consequently,

$$\left\| \left(\sum_{j=1}^n \left(\left| \mathbb{D}_j[\hat\theta] \right|_2^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r(1),$$

and

$$\left\| \left(\sum_{j=1}^n \left(\left| X^\top \mathbb{D}_j[\hat\theta] \right|_2^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r(\sqrt{n}).$$

Proof. For $|h|_2 \le 1$, the derivative identity gives

$$\mathbb{D}_j[\hat\theta][h] = -\frac{1}{n} G \left[L''(x_j^\top \hat\theta)(h^\top \hat\theta) x_j + s_j h \right].$$

Therefore

$$\left| \mathbb{D}_j[\hat\theta][h] \right|_2 \le \frac{O(1)}{n} \left(\left| \hat\theta \right|_2 |x_j|_2 + |s_j| \right).$$

The first estimate follows from Theorem 3.7, Lemma 3.2, and Lemma 3.11.

Multiplying the same identity by $X^\top$, we get

$$X^\top \mathbb{D}_j[\hat\theta][h] = -\frac{1}{n} \left[L''(x_j^\top \hat\theta)(h^\top \hat\theta) X^\top G x_j + s_j X^\top G h \right].$$

Using

$$\left|X^\top G h\right|_2 \le O(\sqrt{n})\,|h|_2\,, \qquad \left|X^\top G x_j\right|_2 \le \left|X^\top G\right|_{\mathrm{op}} |x_j|_2 = O(\sqrt{n})\,|x_j|_2\,,$$

we obtain

$$\left|X^\top \mathbb{D}_j[\hat\theta]\right|_2^* \le \frac{O(1)}{\sqrt{n}} \left(\left|\hat\theta\right|_2 |x_j|_2 + |s_j|\right).$$

The second estimate follows again from Theorem 3.7, Lemma 3.2, and Lemma 3.11.

Finally, by Minkowski's inequality in $L^r(\ell_2)$,

$$\left\|\left(\sum_{j=1}^n \left(\left|\mathbb{D}_j[\hat\theta]\right|_2^*\right)^2\right)^{1/2}\right\|_{L^r} \le \left(\sum_{j=1}^n \left\|\left|\mathbb{D}_j[\hat\theta]\right|_2^*\right\|_{L^r}^2\right)^{1/2} = O_r(1),$$

and similarly

$$\left\|\left(\sum_{j=1}^n \left(\left|X^\top\mathbb{D}_j[\hat\theta]\right|_2^*\right)^2\right)^{1/2}\right\|_{L^r} \le \left(\sum_{j=1}^n \left\|\left|X^\top\mathbb{D}_j[\hat\theta]\right|_2^*\right\|_{L^r}^2\right)^{1/2} = O_r(\sqrt{n}).$$

□

Lemma 5.6 (Derivative of the averaged auxiliary score)**.** *Under Assumptions 1–7, for every fixed integer $r \ge 1$,*

$$\sup_{j\in[n]} \left\|\left|\mathbb{D}_j[s_x^\zeta]\right|^*\right\|_{L^r} = O_r\left(\frac{1}{\sqrt{n}}\right).$$

Proof. By Lemma 4.14, applied in the generic setting,

$$\mathbb{D}_j[s_x^\zeta][h] = \frac{L''(\zeta_x(x^\top\hat\theta))}{1+\gamma_x L''(\zeta_x(x^\top\hat\theta))} x^\top \mathbb{D}_j[\hat\theta][h].$$

Therefore

$$|\mathbb{D}_j[s_x^\zeta][h]| \le O(1)|x^\top \mathbb{D}_j[\hat\theta][h]|.$$

Using the explicit derivative of $\hat\theta$,

$$|x^\top \mathbb{D}_j[\hat\theta][h]| \le \frac{O(1)}{n}\left(|x^\top G x_j|\,\left|\hat\theta\right|_2 + |s_j|\;|Gx|_2\right),$$

uniformly over $|h|_2 \le 1$. The generic cross-leverage bound gives

$$\left\||x^\top G x_j|\right\|_{L^r} = O_r(\sqrt{n}), \qquad \||Gx|_2\|_{L^r} = O_r(\sqrt{n}),$$

and Theorem 3.7 together with Lemma 3.11 gives

$$\left\|\left|\hat\theta\right|_2\right\|_{L^r} + \sup_{j\in[n]} \|s_j\|_{L^r} = O_r(1).$$

The claim follows by Hölder's inequality. □

Differentiating G^{-1}, we use the decomposition

$$\mathbb{D}_j[G^{-1}][h] = \mathcal{L}_j[h] + \mathcal{B}_j[h],$$

where the local part, coming from the direct differentiation of the column x_j, is

$$\mathcal{L}_j[h] := \frac{1}{n} L''(x_j^\top\hat\theta)\left(hx_j^\top + x_j h^\top\right) + \frac{1}{n} L'''(x_j^\top\hat\theta)(h^\top\hat\theta)x_j x_j^\top,$$

and the background part, coming from the variation of $\hat{\theta}$, is

$$\mathcal{B}_j[h] := \frac{1}{n}\sum_{\ell=1}^{n} L'''(x_\ell^\top \hat{\theta}) \left(x_\ell^\top \mathbb{D}_j[\hat{\theta}][h]\right) x_\ell x_\ell^\top + \nabla^3\rho(\hat{\theta})\left[\mathbb{D}_j[\hat{\theta}][h]\right].$$

Under the standing condition $\nabla^3\rho \equiv 0$, the last term in $\mathcal{B}_j[h]$ vanishes. The resolvent derivative is therefore

$$\mathbb{D}_j[G][h] = -G\left(\mathcal{L}_j[h] + \mathcal{B}_j[h]\right)G.$$

In the leave-one-out application, after replacing $X, \hat{\theta}, G$ by $X_{-i}, \hat{\theta}_{-i}, G_{-i}$, the corresponding objects are denoted by $\mathcal{L}_{ij}$ and $\mathcal{B}_{ij}$.

Proposition 5.7 (First-order resolvent derivative). *Under Assumptions 1–7 and $\nabla^3\rho \equiv 0$, for every fixed integer $r \geq 1$,*

$$\sup_{j\in[n]} \left\| \left| \frac{1}{\sqrt{n}} u^\top \mathbb{D}_j[G] x \right|^* \right\|_{L^r} = O_r\left(\frac{(\log n)^7}{n}\right).$$

Proof. For the local part, using the explicit form of $\mathcal{L}_j[h]$,

$$\left|\frac{1}{\sqrt{n}} u^\top G\mathcal{L}_j[h]Gx\right| \leq \frac{O(1)}{n\sqrt{n}}\left(|u^\top Gh|\,|x_j^\top Gx| + |u^\top Gx_j|\,|h^\top Gx|\right) + \frac{O(1)}{n\sqrt{n}}|u^\top Gx_j|\,|x_j^\top Gx|\,\left|\hat{\theta}\right|_2.$$

The bounds

$$\left\| |x_j^\top Gx| \right\|_{L^r} = O_r(\sqrt{n}\log n), \qquad \left\| |Gx|_2 \right\|_{L^r} = O_r(\sqrt{n}),$$

together with Lemma 5.4 and $|Gu|_2 = O(1)$, yield

$$\sup_{j\in[n]} \left\| \left| \frac{1}{\sqrt{n}} u^\top G\mathcal{L}_j[h]Gx \right| \right\|_{L^r} = O_r\left(\frac{(\log n)^7}{n}\right),$$

uniformly over $|h|_2 \leq 1$.

For the background part, set

$$b := X^\top Gu, \qquad y := X^\top Gx, \qquad d_j := X^\top Gx_j,$$

and let

$$D_L := \operatorname{diag}_{\ell\in[n]}\left(L'''(x_\ell^\top \hat{\theta})\right).$$

Then

$$u^\top G\mathcal{B}_j[h]Gx = \frac{1}{n} b^\top D_L \operatorname{diag}(y)\, X^\top \mathbb{D}_j[\hat{\theta}][h].$$

Using the explicit derivative of the minimizer,

$$X^\top \mathbb{D}_j[\hat{\theta}][h] = -\frac{1}{n}\left[L''(x_j^\top \hat{\theta})(h^\top \hat{\theta})d_j + s_j X^\top Gh\right].$$

Consequently,

$$\sup_{|h|_2\leq 1} \left|\frac{1}{\sqrt{n}} u^\top G\mathcal{B}_j[h]Gx\right| \leq \frac{O(1)}{n^2\sqrt{n}}\left(\left|\hat{\theta}\right|_2 \left|b^\top D_L \operatorname{diag}(y) d_j\right| + |s_j|\,\left|GXD_L \operatorname{diag}(b)y\right|_2\right).$$

Conditionally on X, the first scalar term is a linear form in the auxiliary column x, since

$$b^\top D_L \operatorname{diag}(y) d_j = x^\top G X D_L \operatorname{diag}(b) d_j.$$

Thus the linear moment bound for x, together with

$$|GX|_{\mathrm{op}} = O(\sqrt{n}), \qquad |b|_\infty = \left|X^\top G u\right|_\infty, \qquad |d_j|_2 = \left|X^\top G x_j\right|_2 \le \left|X^\top G X\right|_{\mathrm{op}} = O(n),$$

gives

$$\left\|b^\top D_L \operatorname{diag}(y) d_j\right\|_{L^r} \le O_r\left(n^{3/2} (\log n)^6\right).$$

For the second term,

$$|GXD_L \operatorname{diag}(b) y|_2 \le |GX|_{\mathrm{op}} |b|_\infty |y|_2 .$$

Using

$$\left\|\left|X^\top G u\right|_\infty\right\|_{L^r} = O_r((\log n)^6), \qquad \left\|\left|X^\top G x\right|_2\right\|_{L^r} = O_r(n),$$

we obtain

$$\||GXD_L \operatorname{diag}(b) y|_2\|_{L^r} = O_r\left(n^{3/2} (\log n)^6\right).$$

Since

$$\left\|\left|\hat{\theta}\right|_2\right\|_{L^r} + \sup_{j \in [n]} \|s_j\|_{L^r} = O_r(1),$$

Hölder's inequality yields

$$\sup_{j \in [n]} \left\|\left|\frac{1}{\sqrt{n}} u^\top G \mathcal{B}_j G x\right|^*\right\|_{L^r} = O_r\left(\frac{(\log n)^6}{n}\right).$$

□

Lemma 5.8 (Resolvent part of the derivative of $X^\top G x$)**.** *Under Assumptions 1–7 and $\nabla^3 \rho \equiv 0$, for every fixed integer $r \ge 1$,*

$$\sup_{j \in [n]} \left\|\left|X^\top \mathbb{D}_j[G] x\right|_2^*\right\|_{L^r} = O_r(\sqrt{n} \log n).$$

Proof. Using

$$\mathbb{D}_j[G][h] = -G\left(\mathcal{L}_j[h] + \mathcal{B}_j[h]\right) G,$$

we bound the local and background terms separately. For the local contribution, the three terms in $\mathcal{L}_j[h]$, together with

$$\left|X^\top G h\right|_2 = O(\sqrt{n}), \qquad \||Gx|_2\|_{L^r} = O_r(\sqrt{n}), \qquad \left\|\left|X^\top G x_j\right|_2\right\|_{L^r} = O_r(n),$$

give

$$\sup_{j \in [n]} \left\|\left|X^\top G \mathcal{L}_j[h] G x\right|_2\right\|_{L^r} = O_r(\sqrt{n} \log n),$$

uniformly over $|h|_2 \le 1$. For the background contribution,

$$\left|X^\top G \mathcal{B}_j[h] G x\right|_2 \le \frac{1}{n} \left|X^\top G X\right|_{\mathrm{op}} \left|\Lambda_j(h) X^\top G x\right|_2 .$$

Moreover,

$$\left|\Lambda_j(h) X^\top G x\right|_2 \le O(1) \left|X^\top \mathbb{D}_j[\hat{\theta}][h]\right|_2 \left|X^\top G x\right|_\infty .$$

Using Lemma 5.5 and the generic cross-leverage bound, we obtain

$$\sup_{j \in [n]} \left\|\left|X^\top G \mathcal{B}_j[h] G x\right|_2\right\|_{L^r} = O_r(\sqrt{n} \log n).$$

Taking the supremum over $|h|_2 \le 1$ proves the claim. □

Proposition 5.9 (Resolvent derivatives in the second-order term)**.** *Under Assumptions 1–7 and $\nabla^3\rho \equiv 0$, for every fixed integer $r \ge 1$,*

$$\sup_{j\in[n]} \left\| \left| \frac{1}{n^{3/2}} u^\top \mathbb{D}_j[G] X\tilde{\Gamma}_x X^\top G x \right|^* \right\|_{L^r} = O_r\left(\frac{(\log n)^7}{n}\right).$$

The same bound holds when the derivative falls on the final resolvent:

$$\sup_{j\in[n]} \left\| \left| \frac{1}{n^{3/2}} u^\top G X\tilde{\Gamma}_x X^\top \mathbb{D}_j[G] x \right|^* \right\|_{L^r} = O_r\left(\frac{(\log n)^7}{n}\right).$$

Proof. We first consider the derivative of the first resolvent:

$$u^\top \mathbb{D}_j[G][h] X\tilde{\Gamma}_x X^\top G x = -u^\top G\mathcal{L}_j[h] G X\tilde{\Gamma}_x X^\top G x - u^\top G\mathcal{B}_j[h] G X\tilde{\Gamma}_x X^\top G x.$$

For the local part, the three terms in $\mathcal{L}_j[h]$ give, uniformly over $|h|_2 \le 1$,

$$\begin{aligned}
&\left| \frac{1}{n^{3/2}} u^\top G\mathcal{L}_j[h] G X\tilde{\Gamma}_x X^\top G x \right| \\
&\quad \le \frac{O(1)}{n^{5/2}} \Big(|x_j^\top G X\tilde{\Gamma}_x X^\top G x| + |x_j^\top G u| \left| G X\tilde{\Gamma}_x X^\top G x \right|_2 \\
&\qquad\qquad + |x_j^\top G u|\, |x_j^\top G X\tilde{\Gamma}_x X^\top G x| \left|\hat{\theta}\right|_2 \Big).
\end{aligned}$$

Now

$$\left\| \left| G X\tilde{\Gamma}_x X^\top G x \right|_2 \right\|_{L^r} = O_r(n \log n),$$

and

$$\left\| \left| X^\top G X\tilde{\Gamma}_x X^\top G x \right|_2 \right\|_{L^r} = O_r(n^{3/2} \log n)$$

by Lemmas 4.2 and 5.2. Together with

$$\left\| |x_j^\top G u| \right\|_{L^r} = O_r((\log n)^6), \qquad \left\| \left|\hat{\theta}\right|_2 \right\|_{L^r} = O_r(1),$$

this gives

$$\sup_{j\in[n]} \left\| \left| \frac{1}{n^{3/2}} u^\top G\mathcal{L}_j[h] G X\tilde{\Gamma}_x X^\top G x \right| \right\|_{L^r} = O_r\left(\frac{(\log n)^7}{n}\right).$$

For the background part,

$$u^\top G\mathcal{B}_j[h] G X\tilde{\Gamma}_x X^\top G x = \frac{1}{n} (X^\top G u)^\top \Lambda_j(h) X^\top G X\tilde{\Gamma}_x X^\top G x,$$

where $\Lambda_j(h)$ is as above. Thus

$$\begin{aligned}
&\left| \frac{1}{n^{3/2}} u^\top G\mathcal{B}_j[h] G X\tilde{\Gamma}_x X^\top G x \right| \\
&\quad \le \frac{O(1)}{n^{5/2}} \left| X^\top G u \right|_\infty \left| X^\top \mathbb{D}_j[\hat{\theta}][h] \right|_2 \left| X^\top G X\tilde{\Gamma}_x X^\top G x \right|_2.
\end{aligned}$$

Lemmas 5.4, 5.5, and 5.2 yield

$$\sup_{j\in[n]} \left\| \left| \frac{1}{n^{3/2}} u^\top G\mathcal{B}_j[h] G X\tilde{\Gamma}_x X^\top G x \right| \right\|_{L^r} = O_r\left(\frac{(\log n)^7}{n}\right).$$

This proves the first bound.

For the term with the final resolvent differentiated, write

$$\frac{1}{n^{3/2}} u^\top G X \tilde{\Gamma}_x X^\top \mathbb{D}_j[G][h] x = \frac{1}{n^{3/2}} (X^\top G u)^\top \tilde{\Gamma}_x X^\top \mathbb{D}_j[G][h] x.$$

Therefore

$$\left| \frac{1}{n^{3/2}} u^\top G X \tilde{\Gamma}_x X^\top \mathbb{D}_j[G][h] x \right| \le \frac{1}{n^{3/2}} \left| X^\top G u \right|_2 \left| \tilde{\Gamma}_x \right|_{\mathrm{op}} \left| X^\top \mathbb{D}_j[G][h] x \right|_2 .$$

Using Lemmas 5.1, 5.2, and 5.8, we obtain

$$\sup_{j \in [n]} \left\| \left| \frac{1}{n^{3/2}} u^\top G X \tilde{\Gamma}_x X^\top \mathbb{D}_j[G] x \right|^* \right\|_{L^r} = O_r\left(\frac{(\log n)^7}{n} \right).$$

□

Proposition 5.10 (Derivative of the curvature diagonal)**.** *Under Assumptions 1–7 and* $\nabla^3 \rho \equiv 0$, *for every fixed integer* $r \ge 1$,

$$\sup_{j \in [n]} \left\| \left| \frac{1}{n^{3/2}} u^\top G X \, \mathbb{D}_j[\tilde{\Gamma}_x] \, X^\top G x \right|^* \right\|_{L^r} = O_r\left(\frac{(\log n)^7}{n} \right).$$

Proof. For $|h|_2 \le 1$,

$$\frac{1}{n^{3/2}} u^\top G X \, \mathbb{D}_j[\tilde{\Gamma}_x][h] \, X^\top G x = \frac{1}{n^{3/2}} \sum_{\ell=1}^{n} (x_\ell^\top G u) \, \mathbb{D}_j[(\tilde{\Gamma}_x)_\ell][h] \, (x_\ell^\top G x).$$

Since

$$(\tilde{\Gamma}_x)_\ell = -\frac{s_x^\zeta}{2n} L'''(x_\ell^\top \hat{\theta}) \, x_\ell^\top G x,$$

we decompose

$$\begin{aligned}
\mathbb{D}_j[(\tilde{\Gamma}_x)_\ell][h] = &-\frac{\mathbb{D}_j[s_x^\zeta][h]}{2n} L'''(x_\ell^\top \hat{\theta}) x_\ell^\top G x \\
&- \frac{s_x^\zeta}{2n} \mathbb{D}_j\left[L'''(x_\ell^\top \hat{\theta}) \right][h] \, x_\ell^\top G x \\
&- \frac{s_x^\zeta}{2n} L'''(x_\ell^\top \hat{\theta}) \left(\mathbf{1}_{\ell=j} h^\top G x + x_\ell^\top \mathbb{D}_j[G][h] x \right).
\end{aligned}$$

The derivative of s_x^ζ gives a contribution bounded by

$$\begin{aligned}
&\left\| \frac{1}{n^{3/2}} \sum_{\ell=1}^{n} (x_\ell^\top G u) \, \frac{\mathbb{D}_j[s_x^\zeta][h]}{2n} L'''(x_\ell^\top \hat{\theta}) \, (x_\ell^\top G x)^2 \right\|_{L^r} \\
&\qquad \le \frac{O(1)}{n^{5/2}} \left\| \left| \mathbb{D}_j[s_x^\zeta] \right|^* \left| X^\top G u \right|_\infty \left| X^\top G x \right|_2^2 \right\|_{L^r} = O_r\left(\frac{(\log n)^6}{n} \right).
\end{aligned}$$

thanks to Lemmas 5.6, 5.4, and the generic leverage bounds.

Next consider the derivative of $L'''(x_\ell^\top \hat{\theta})$. Define

$$\eta_\ell(h) := \mathbf{1}_{\ell=j} h^\top \hat{\theta} + x_\ell^\top \mathbb{D}_j[\hat{\theta}][h].$$

The Lipschitz bound on L''' gives

$$\left| \mathbb{D}_j \left[L'''(x_\ell^\top \hat{\theta}) \right][h] \right| \le O(1) |\eta_\ell(h)|.$$

Thus the corresponding contribution T_2 satisfies

$$T_2 \le \frac{O(1)|s_x^\zeta|}{n^{5/2}} \sum_{\ell=1}^n |x_\ell^\top Gu|\, |x_\ell^\top Gx|^2\, |\eta_\ell(h)|$$
$$\le \frac{O(1)|s_x^\zeta|}{n^{5/2}} \left|X^\top Gu\right|_\infty \left|X^\top Gx\right|_\infty \left|X^\top Gx\right|_2 |\eta(h)|_2 .$$

By Lemma 5.5,

$$\left\| \sup_{|h|_2\le 1} |\eta(h)|_2 \right\|_{L^r} \le \left\| \left|\hat\theta\right|_2 \right\|_{L^r} + \left\| \sup_{|h|_2\le 1} \left| X^\top \mathbb{D}_j[\hat\theta][h] \right|_2 \right\|_{L^r} = O_r(1)$$

Using $\left\| s_x^\zeta \right\|_{L^r} = O_r(1)$, Lemma 5.4, and the generic cross-leverage bounds, we get

$$\|T_2\|_{L^r} = O_r\left(\frac{(\log n)^7}{n} \right).$$

Finally, the derivative of $x_\ell^\top Gx$ gives two contributions. The explicit derivative of the column x_j contributes

$$T_{3,\mathrm{exp}} \le \frac{O(1)|s_x^\zeta|}{n^{5/2}} |x_j^\top Gu|\ |Gx|_2\ |x_j^\top Gx|.$$

Using Lemma 5.4,

$$\left\| \left| X^\top Gu \right|_\infty \right\|_{L^r} = O_r((\log n)^6),$$

and the generic bounds

$$\||Gx|_2\|_{L^r} = O_r(\sqrt{n}), \qquad \left\| \left| X^\top Gx \right|_\infty \right\|_{L^r} = O_r(\sqrt{n}\log n),$$

we get

$$\|T_{3,\mathrm{exp}}\|_{L^r} = O_r\left(\frac{(\log n)^7}{n^{3/2}} \right) = O_r\left(\frac{(\log n)^7}{n} \right).$$

The derivative through the resolvent contributes

$$T_{3,G} \le \frac{O(1)|s_x^\zeta|}{n^{5/2}} \left|X^\top Gu\right|_\infty \left|X^\top Gx\right|_2 \left|X^\top \mathbb{D}_j[G][h]x\right|_2 .$$

By Lemma 5.8, together with the previous bounds,

$$\|T_{3,G}\|_{L^r} = O_r\left(\frac{(\log n)^7}{n} \right).$$

Combining the bounds for T_1, T_2, $T_{3,\mathrm{exp}}$, and $T_{3,G}$ proves the proposition. □

Lemma 5.11 (Derivatives of the explicit X-factors)**.** *Under Assumptions 1–7 and $\nabla^3\rho \equiv 0$, for every fixed integer $r \ge 1$,*

$$\sup_{j\in[n]} \left\| \left| \frac{1}{n^{3/2}} u^\top G\, \mathbb{D}_j[X]\, \tilde\Gamma_x X^\top Gx \right|^* \right\|_{L^r} = O_r\left(\frac{(\log n)^2}{n^{3/2}} \right),$$

and

$$\sup_{j\in[n]} \left\| \left| \frac{1}{n^{3/2}} u^\top GX\tilde\Gamma_x\, \mathbb{D}_j[X^\top]Gx \right|^* \right\|_{L^r} = O_r\left(\frac{(\log n)^7}{n^{3/2}} \right).$$

Proof. Set

$$y := X^\top Gx.$$

For $|h|_2 \le 1$, the matrix $\mathbb{D}_j[X][h]$ has column h in position j and zero elsewhere. Hence

$$u^\top G\,\mathbb{D}_j[X][h]\,\tilde{\Gamma}_x X^\top Gx = u^\top Gh\,(\tilde{\Gamma}_x y)_j.$$

Therefore

$$\sup_{|h|_2\le 1}\left|\frac{1}{n^{3/2}}u^\top G\,\mathbb{D}_j[X][h]\,\tilde{\Gamma}_x X^\top Gx\right| \le \frac{1}{n^{3/2}}\,|Gu|_2\left|\tilde{\Gamma}_x\right|_{\mathrm{op}}\left|X^\top Gx\right|_\infty.$$

Using $|Gu|_2 = O(1)$, Lemma 5.2, and Lemma 5.4, specifically the auxiliary-column bound

$$\left\|\left|X^\top Gx\right|_\infty\right\|_{L^r} = O_r(\sqrt{n}\log n),$$

we obtain

$$\sup_{j\in[n]}\left\|\left|\frac{1}{n^{3/2}}u^\top G\,\mathbb{D}_j[X]\,\tilde{\Gamma}_x X^\top Gx\right|^*\right\|_{L^r} = O_r\left(\frac{(\log n)^2}{n^{3/2}}\right).$$

For the second explicit factor, $\mathbb{D}_j[X^\top][h]$ has row $h^\top$ in position j and zero elsewhere. Thus

$$u^\top GX\tilde{\Gamma}_x\,\mathbb{D}_j[X^\top][h]Gx = (x_j^\top Gu)\,(\tilde{\Gamma}_x)_j\,h^\top Gx.$$

Consequently,

$$\sup_{|h|_2\le 1}\left|\frac{1}{n^{3/2}}u^\top GX\tilde{\Gamma}_x\,\mathbb{D}_j[X^\top][h]Gx\right| \le \frac{1}{n^{3/2}}\left|X^\top Gu\right|_\infty\left|\tilde{\Gamma}_x\right|_{\mathrm{op}}|Gx|_2.$$

By Lemmas 5.4 and 5.2, and by the generic bound

$$\||Gx|_2\|_{L^r} = O_r(\sqrt{n}),$$

the right-hand side has L^r-norm

$$O_r\left(\frac{1}{n^{3/2}}(\log n)^6\frac{\log n}{\sqrt{n}}\sqrt{n}\right) = O_r\left(\frac{(\log n)^7}{n^{3/2}}\right).$$

This proves the second estimate. ☐

Corollary 5.12 (Derivative of the explicit approximation)**.** *Under Assumptions 1–7 and $\nabla^3\rho \equiv 0$, for every fixed integer $r \ge 1$,*

$$\sup_{j\in[n]}\||\mathbb{D}_j[\mathfrak{d}_x(X)]|^*\|_{L^r} = O_r\left(\frac{(\log n)^7}{n}\right),$$

where we recall the notation:

$$\mathfrak{d}_x(X) := -\frac{s_x^\zeta}{\sqrt{n}}u^\top Gx + \frac{s_x^\zeta}{n^{3/2}}u^\top GX\tilde{\Gamma}_x X^\top Gx.$$

Proof. Differentiate the two terms in $\mathfrak{d}_x$.

For the first-order part,

$$\mathbb{D}_j\left[-\frac{s_x^\zeta}{\sqrt{n}}u^\top Gx\right][h] = -\frac{\mathbb{D}_j[s_x^\zeta][h]}{\sqrt{n}}u^\top Gx - \frac{s_x^\zeta}{\sqrt{n}}u^\top\mathbb{D}_j[G][h]x.$$

In L^r-norm, the first term is $O_r(1/n)$ by Lemma 5.6 and Lemma 5.1. The second term is $O_r((\log n)^7/n)$ by Proposition 5.7 and $\left\|s_x^\zeta\right\|_{L^r} = O_r(1)$.

For the second-order part, the derivative can hit s_x^ζ, the first resolvent G, the two explicit X-factors, the diagonal $\tilde{\Gamma}_x$, or the final resolvent G. The derivative of s_x^ζ gives

$$\begin{aligned}\left\| \sup_{|h|_2 \le 1} \frac{\mathbb{D}_j[s_x^\zeta][h]}{n^{3/2}} u^\top G X \tilde{\Gamma}_x X^\top G x \right\|_{L^r} &\le \frac{1}{n^{3/2}} \left\| \left|\mathbb{D}_j[s_x^\zeta]\right|^* \left|X^\top G u\right|_2 \left|\tilde{\Gamma}_x X^\top G x\right|_2 \right\|_{L^r} \\ &= \frac{1}{n^{3/2}} O_r\left(n^{-1/2}\sqrt{n}\sqrt{n}\log n\right) = O_r((\log n)/n).\end{aligned}$$

thanks to Lemmas 5.1, 5.2 and 5.6.

The derivatives of the two resolvents are controlled by Proposition 5.9; the derivative of $\tilde{\Gamma}_x$ is controlled by Proposition 5.10; and the derivatives of the explicit X-factors are controlled by Lemma 5.11. Combining these estimates gives

$$\sup_{j\in[n]} \| |\mathbb{D}_j[\mathfrak{d}_x(X)]|^* \|_{L^r} = O_r\left(\frac{(\log n)^7}{n}\right).$$

□

Consequently, replacing

$$X, \hat{\theta}, G, x, \tilde{\Gamma}_x, s_x^\zeta \quad \text{by} \quad X_{-i}, \hat{\theta}_{-i}, G_{-i}, x_i, \tilde{\Gamma}_i, s_i^\zeta,$$

we obtain uniformly over $i \neq j$,

$$\| |\mathbb{D}_j[\mathfrak{d}_i(X)]|^* \|_{L^r} = O_r\left(\frac{(\log n)^7}{n}\right).$$

Combining this bound with Lemma B.8, applied with $k = 4$, and with Remark 4.20 from the section on the second-order leave-one-out displacement approximation yields the following bound.

Proposition 5.13 (Bound on c_n^x)**.** *Under Assumptions 1–7 and* $\nabla^3 \rho \equiv 0$,

$$c_n^x = O\left(\frac{(\log n)^7}{n}\right).$$

Proof. Fix $i \neq j$. By Lemma B.8, applied conditionally with respect to the column x_j,

$$\left\|\mathfrak{d}_i(X) - \mathbb{E}_j\left[\mathfrak{d}_i(X)\right]\right\|_{L^4} \le O(1)\, \| |\mathbb{D}_j[\mathfrak{d}_i(X)]|^* \|_{L^4} = O\left(\frac{(\log n)^7}{n}\right),$$

uniformly over $i \neq j$.

By Remark 4.20,

$$\sup_{i\in[n]} \|\delta_i f_n(X) - \mathfrak{d}_i(X)\|_{L^4} = O\left(\frac{(\log n)^2}{n}\right).$$

Since conditional expectation is a contraction in L^4,

$$\begin{aligned}&\|(\delta_i f_n(X) - \mathbb{E}_j[\delta_i f_n(X)]) - (\mathfrak{d}_i(X) - \mathbb{E}_j[\mathfrak{d}_i(X)])\|_{L^4} \\ &\le 2\, \|\delta_i f_n(X) - \mathfrak{d}_i(X)\|_{L^4} = O\left(\frac{(\log n)^2}{n}\right).\end{aligned}$$

Taking the supremum over $i \neq j$ gives

$$c_n^x = \sup_{i\neq j} \|\delta_i f_n(X) - \mathbb{E}_j[\delta_i f_n(X)]\|_{L^4} = O\left(\frac{(\log n)^7}{n}\right).$$

□

5.3 Contracted active-background term

It remains to bound the contribution to $c_n^{\bar{x}}$. We use the generic notation introduced in Subsection 5.2. Thus $x \in \mathbb{R}^p$ denotes the active column, independent of the background matrix $X = (x_1, \ldots, x_n)$, and

$$G = \left(\nabla^2 \Psi_X(\hat{\theta})\right)^{-1}.$$

The deterministic averaged leverage associated with the active column is denoted by γ_x, and we define

$$s_x^\zeta := L'\Big(\zeta_x(x^\top \hat{\theta})\Big), \qquad \zeta_x(t) + \gamma_x L'(\zeta_x(t)) = t.$$

We use a check accent for quantities evaluated at the checked background X^{A_j}:

$$\check{G} := G(X^{A_j}), \qquad \check{\hat{\theta}} := \hat{\theta}(X^{A_j}), \qquad \check{s}_x^\zeta := s_x^\zeta(X^{A_j}), \qquad \check{\tilde{\Gamma}}_x := \tilde{\Gamma}_x(X^{A_j}).$$

The active column x is the same in the checked and unchecked expressions; only the background is checked.

Define

$$y := Gu, \qquad \check{y} := \check{G}u.$$

Let D be the diagonal matrix, indexed by the columns of X, with entries

$$D_\ell := L'''(x_\ell^\top \hat{\theta})\, x_\ell^\top y,$$

and set

$$Y := GXDX^\top G.$$

The checked matrices $\check{D}$ and $\check{Y}$ are defined analogously. Recall that

$$\tilde{\Gamma}_x = \mathrm{diag}_{\ell \in [n]} - \frac{s_x^\zeta}{2n} L'''(x_\ell^\top \hat{\theta})\, x_\ell^\top G x,$$

Therefore the explicit approximation can be written as

$$\mathfrak{d}_x = \mathfrak{d}_x^{(0)} + \mathfrak{d}_x^{(1)},$$

where

$$\mathfrak{d}_x^{(0)} = -\frac{s_x^\zeta}{\sqrt{n}}\, y^\top x, \qquad \mathfrak{d}_x^{(1)} = -\frac{(s_x^\zeta)^2}{2n^{5/2}}\, x^\top Y x.$$

The checked statistics $\check{\mathfrak{d}}_x, \check{\mathfrak{d}}_x^{(0)}, \check{\mathfrak{d}}_x^{(1)}$ are obtained by replacing s_x^ζ, y, Y with $\check{s}_x^\zeta, \check{y}, \check{Y}$.

In this subsection, $\mathbb{E}_x$, Var_x, and Cov_x denote conditional expectation, variance, and covariance with respect to the active column x only, conditionally on all background variables. For a background-measurable matrix B, and for fixed $q \geq 1$, write

$$\mathcal{Q}_q(B) := \left\| x^\top B x \right\|_{L_x^q}.$$

We also write

$$\Sigma_x := \mathbb{E}_x[xx^\top]$$

for the second-moment matrix of the active column. By Assumptions 4 and 3,

$$|\Sigma_x|_{\mathrm{op}} = O(1).$$

For background-measurable vectors $a, b \in \mathbb{R}^p$ and symmetric matrices $B, C \in \mathbb{R}^{p\times p}$, introduce the active contractions

$$\mathcal{C}^{00}(a,b) := \mathrm{Cov}_x\Big(s_x^\zeta\, a^\top x, \check{s}_x^\zeta\, b^\top x\Big)\,,$$
$$\mathcal{C}^{01}(a,C) := \mathrm{Cov}_x\Big(s_x^\zeta\, a^\top x, (\check{s}_x^\zeta)^2\, x^\top Cx\Big)\,,$$
$$\mathcal{C}^{10}(B,b) := \mathrm{Cov}_x\Big((s_x^\zeta)^2\, x^\top Bx, \check{s}_x^\zeta\, b^\top x\Big)\,,$$
$$\mathcal{C}^{11}(B,C) := \mathrm{Cov}_x\Big((s_x^\zeta)^2\, x^\top Bx, (\check{s}_x^\zeta)^2\, x^\top Cx\Big)\,.$$

Then

$$\mathrm{Cov}_x\big(\mathfrak{d}_x, \check{\mathfrak{d}}_x\big) = \frac{1}{n}\mathcal{C}^{00}(y,\check{y}) + \frac{1}{2n^3}\mathcal{C}^{01}(y,\check{Y}) + \frac{1}{2n^3}\mathcal{C}^{10}(Y,\check{y}) + \frac{1}{4n^5}\mathcal{C}^{11}(Y,\check{Y}). \tag{5.1}$$

Lemma 5.14 (Square-summed derivatives of the active score)**.** *Under Assumptions 1–7, for every fixed integer $r \geq 1$,*

$$\left\| \left(\sum_{\ell=1}^n \left(\left| \mathbb{D}_\ell[s_x^\zeta] \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r(1).$$

The same estimate holds for the checked score:

$$\left\| \left(\sum_{\ell=1}^n \left(\left| \mathbb{D}'_\ell[\check{s}_x^\zeta] \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r(1).$$

Proof. By Lemma 4.14, in the generic background notation,

$$\mathbb{D}_\ell[s_x^\zeta][h] = \frac{L''(\zeta_x(x^\top\hat{\theta}))}{1+\gamma_x L''(\zeta_x(x^\top\hat{\theta}))}x^\top\mathbb{D}_\ell[\hat{\theta}][h].$$

The prefactor is uniformly bounded. Hence

$$\left|\mathbb{D}_\ell[s_x^\zeta]\right|^* \leq O(1)\left|x^\top\mathbb{D}_\ell[\hat{\theta}]\right|^*.$$

Applying Lemma 3.14 conditionally on the active column x, with $z = x$, gives

$$\left\| \left(\sum_{\ell=1}^n \left(\left| x^\top\mathbb{D}_\ell[\hat{\theta}] \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} \leq \frac{O_r(1)}{\sqrt{n}}\, \||x|_2\|_{L^{4r}} = O_r(1),$$

using Lemma 3.2. The checked estimate is identical. □

Lemma 5.15 (Background matrix bounds)**.** *Under Assumptions 1–7 and $\nabla^3\rho \equiv 0$, for every fixed integers $q, r \geq 1$,*

$$\||y|_2\|_{L^r} = O_r(1), \qquad \|\mathcal{Q}_q(Y)\|_{L^r} = O_{q,r}\big(n^2(\log n)^6\big)\,.$$

The same estimates hold for $\check{y}$ and $\check{Y}$.

Proof. Recall that

$$y = Gu, \qquad D = \mathrm{diag}_{m\in[n]}\Big(L'''(x_m^\top\hat{\theta})\, x_m^\top y\Big), \qquad Y = GXDX^\top G.$$

Since $|G|_{\text{op}} = O(1)$ and $|u|_2 \le 1$,

$$|y|_2 = O(1).$$

Set

$$z := X^\top G x.$$

Then

$$x^\top Y x = z^\top D z.$$

By the boundedness of L''',

$$|D|_{\text{op}} \le O(1) \left| X^\top y \right|_\infty .$$

Moreover, Lemma 5.4 gives

$$\left\| \left| X^\top y \right|_\infty \right\|_{L^r} = O_r((\log n)^6),$$

while the generic cross-leverage estimate gives, conditionally on the background,

$$\left\| \left| X^\top G x \right|_2 \right\|_{L_x^{2q}} = O_q(n).$$

Therefore

$$\begin{aligned} \mathcal{Q}_q(Y) = \left\| z^\top D z \right\|_{L_x^q} &\le |D|_{\text{op}} \left\| |z|_2^2 \right\|_{L_x^q} \\ &\le O_q(n^2) \left| X^\top y \right|_\infty . \end{aligned}$$

Taking the L^r-norm over the background gives

$$\|\mathcal{Q}_q(Y)\|_{L^r} = O_{q,r}\left(n^2 (\log n)^6\right).$$

The checked estimate is identical, since the checked background has the same law and satisfies the same bounds. □

Lemma 5.16 (Differentiated active-score contractions)**.** *Under Assumptions 1–7 and $\nabla^3 \rho \equiv 0$, for every fixed integer $r \ge 1$,*

- $$\left\| \left(\sum_{\ell=1}^n \left(\left| \mathrm{Cov}_x \left(\mathbb{D}_\ell[s_x^\zeta]\, y^\top x, \check{s}_x^\zeta\, \check{y}^\top x \right) \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r\left(\frac{1}{\sqrt{n}} \right).$$
- $$\left\| \left(\sum_{\ell=1}^n \left(\left| \mathrm{Cov}_x \left(\mathbb{D}_\ell[s_x^\zeta]\, y^\top x, (\check{s}_x^\zeta)^2\, x^\top \check{Y} x \right) \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r\left(n^{3/2} (\log n)^6 \right).$$
- $$\left\| \left(\sum_{\ell=1}^n \left(\left| \mathrm{Cov}_x \left(2 s_x^\zeta\, \mathbb{D}_\ell[s_x^\zeta]\, x^\top Y x, (\check{s}_x^\zeta)^2\, x^\top \check{Y} x \right) \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r\left(n^{7/2} (\log n)^{12} \right).$$

The same bound holds with the checked and unchecked roles interchanged, in particular for the score derivatives appearing in $\mathcal{C}^{10}(Y, \check{y})$.

Proof. We prove the unchecked estimates; the checked estimates are identical. By Lemma 4.14,

$$\mathbb{D}_\ell[s_x^\zeta][h] = \frac{L''(\zeta_x(x^\top \hat{\theta}))}{1 + \gamma_x L''(\zeta_x(x^\top \hat{\theta}))} x^\top \mathbb{D}_\ell[\hat{\theta}][h].$$

The prefactor is bounded.

For the 00-term, for every $|h|_2 \le 1$,

$$\operatorname{Cov}_x\Big(\mathbb{D}_\ell[s_x^\zeta][h]\, y^\top x, \check{s}_x^\zeta\, \check{y}^\top x\Big) = \left(\mathbb{E}_x\left[\frac{L''(\zeta_x(x^\top\hat\theta))}{1+\gamma_x L''(\zeta_x(x^\top\hat\theta))} x\,(y^\top x)\left(\check{s}_x^\zeta\, \check{y}^\top x - \mathbb{E}_x[\check{s}_x^\zeta\, \check{y}^\top x]\right)\right]\right)^\top \mathbb{D}_\ell[\hat\theta][h].$$

Consequently, after taking the supremum over $|h|_2 \le 1$ and summing in ℓ, Lemma 3.14 gives

$$\left\|\left(\sum_{\ell=1}^n \left(\left|\operatorname{Cov}_x\Big(\mathbb{D}_\ell[s_x^\zeta]\, y^\top x, \check{s}_x^\zeta\, \check{y}^\top x\Big)\right|^*\right)^2\right)^{1/2}\right\|_{L^r} \le \frac{O_r(1)}{\sqrt{n}}\left\|\left|\mathbb{E}_x\left[\frac{L''(\zeta_x(x^\top\hat\theta))}{1+\gamma_x L''(\zeta_x(x^\top\hat\theta))} x\,(y^\top x)\left(\check{s}_x^\zeta\, \check{y}^\top x - \mathbb{E}_x[\check{s}_x^\zeta\, \check{y}^\top x]\right)\right]\right|_2\right\|_{L^{4r}}$$

We bound the vector inside the last display by duality. For every background-measurable $v \in \mathbb{R}^p$ with $|v|_2 \le 1$, conditional Cauchy–Schwarz and the active linear moment bounds give

$$\left|\mathbb{E}_x\left[(v^\top x)(y^\top x)\left(\check{s}_x^\zeta\, \check{y}^\top x - \mathbb{E}_x[\check{s}_x^\zeta\, \check{y}^\top x]\right)\right]\right| \le O(1)\,|y|_2\,|\check{y}|_2\left\|\check{s}_x^\zeta\right\|_{L_x^4}.$$

Taking the supremum over $|v|_2 \le 1$, then the L^{4r}-norm over the background, yields

$$\left\|\left|\mathbb{E}_x\left[\frac{L''(\zeta_x(x^\top\hat\theta))}{1+\gamma_x L''(\zeta_x(x^\top\hat\theta))} x\,(y^\top x)\left(\check{s}_x^\zeta\, \check{y}^\top x - \mathbb{E}_x[\check{s}_x^\zeta\, \check{y}^\top x]\right)\right]\right|_2\right\|_{L^{4r}} = O_r(1).$$

Consequently the 00-term is $O_r(n^{-1/2})$. By conditional Cauchy–Schwarz, Lemmas 3.2, 4.13 and 5.15. The checked derivative gives the same estimate.

For the one-quadratic active slot, the same duality argument gives

$$\left|\mathbb{E}_x\left[\frac{L''(\zeta_x(x^\top\hat\theta))}{1+\gamma_x L''(\zeta_x(x^\top\hat\theta))} x\,(y^\top x)\left((\check{s}_x^\zeta)^2 x^\top \check{Y} x - \mathbb{E}_x[(\check{s}_x^\zeta)^2 x^\top \check{Y} x]\right)\right]\right|_2 \le O(1)\,|y|_2\,\mathfrak{Q}_4(\check{Y})\left\|\check{s}_x^\zeta\right\|_{L_x^8}^2.$$

Applying Lemmas 3.14 and 5.15 as above gives

$$\left\|\left(\sum_{\ell=1}^n \left(\left|\operatorname{Cov}_x\Big(\mathbb{D}_\ell[s_x^\zeta]\, y^\top x, (\check{s}_x^\zeta)^2\, x^\top \check{Y} x\Big)\right|^*\right)^2\right)^{1/2}\right\|_{L^r} \le O_r((\log n)^6 n^{3/2}).$$

The same argument gives the corresponding bound for $\mathcal{C}^{10}(Y, \check{y})$.

For the two-quadratic active slot, duality gives

$$\left|\mathbb{E}_x\left[x\,(x^\top Y x)\left((\check{s}_x^\zeta)^2 x^\top \check{Y} x - \mathbb{E}_x[(\check{s}_x^\zeta)^2 x^\top \check{Y} x]\right)\right]\right|_2 \le O(1)\mathfrak{Q}_8(Y)\mathfrak{Q}_8(\check{Y})\left\|\check{s}_x^\zeta\right\|_{L_x^{16}}^2.$$

By Lemma 5.15,

$$\|\mathfrak{Q}_8(Y)\|_{L^{8r}} + \left\|\mathfrak{Q}_8(\check{Y})\right\|_{L^{8r}} = O_r\left(n^2 (\log n)^6\right).$$

Hence

$$\left\|\left(\sum_{\ell=1}^n \left(\left|\operatorname{Cov}_x\Big(2 s_x^\zeta\, \mathbb{D}_\ell[s_x^\zeta]\, x^\top Y x, (\check{s}_x^\zeta)^2 x^\top \check{Y} x\Big)\right|^*\right)^2\right)^{1/2}\right\|_{L^r} = O_r\left(n^{7/2} (\log n)^{12}\right).$$

This proves all the stated estimates. $\square$

To bound the terms appearing when differentiating the terms $y, \check{y}, Y, \check{Y}$, we make use of the following preliminary Lemma.

Lemma 5.17 (Active-column contractions)**.** *Under Assumptions 1–7, for every fixed integer $r \geq 1$, there exists a nonnegative background-measurable random variable R, with*

$$\|R\|_{L^r} = O_r(1),$$

such that, for all background-measurable vectors $a, b \in \mathbb{R}^p$ and symmetric matrices $B, C \in \mathbb{R}^{p\times p}$,

$$|\mathcal{C}^{00}(a,b)| \leq R\,|a|_2\,|b|_2\,, \qquad |\mathcal{C}^{01}(a,C)| \leq R\,|a|_2\,\mathcal{Q}_4(C),$$
$$|\mathcal{C}^{10}(B,b)| \leq R\,\mathcal{Q}_4(B)\,|b|_2\,, \qquad |\mathcal{C}^{11}(B,C)| \leq R\,\mathcal{Q}_4(B)\mathcal{Q}_4(C).$$

Proof. Set

$$R := C\left(1 + \left\|s_x^\zeta\right\|_{L_x^8} + \left\|\check{s}_x^\zeta\right\|_{L_x^8}\right)^4.$$

By Lemma 4.13, and the same argument applied to the checked background,

$$\|R\|_{L^r} = O_r(1).$$

For the linear-linear contraction, conditional Cauchy–Schwarz gives

$$\begin{aligned}|\mathcal{C}^{00}(a,b)| &\leq \left\|s_x^\zeta\, a^\top x\right\|_{L_x^2}\left\|\check{s}_x^\zeta\, b^\top x\right\|_{L_x^2}\\ &\leq \left\|s_x^\zeta\right\|_{L_x^4}\left\|\check{s}_x^\zeta\right\|_{L_x^4}\left\|a^\top x\right\|_{L_x^4}\left\|b^\top x\right\|_{L_x^4} \leq R\,|a|_2\,|b|_2\,.\end{aligned}$$

For the mixed contractions,

$$|\mathcal{C}^{01}(a,C)| \leq \left\|s_x^\zeta\, a^\top x\right\|_{L_x^2}\left\|(\check{s}_x^\zeta)^2 x^\top C x\right\|_{L_x^2} \leq R\,|a|_2\,\mathcal{Q}_4(C),$$

and the proof of the bound for $\mathcal{C}^{10}$ is identical. Finally,

$$|\mathcal{C}^{11}(B,C)| \leq \left\|(s_x^\zeta)^2 x^\top B x\right\|_{L_x^2}\left\|(\check{s}_x^\zeta)^2 x^\top C x\right\|_{L_x^2} \leq R\,\mathcal{Q}_4(B)\mathcal{Q}_4(C).$$

This proves the lemma. □

Lemma 5.18 (Square-summed derivatives of y)**.** *Under Assumptions 1–7 and $\nabla^3\rho \equiv 0$, for every fixed integer $r \geq 1$, and for every background-measurable vector $v = v(X)$,*

$$\left\|\left(\sum_{\ell=1}^n\left(\left|v^\top \mathbb{D}_\ell[y]\right|^*\right)^2\right)^{1/2}\right\|_{L^r} \leq \frac{O_r((\log n)^6)}{\sqrt{n}}\,\||v|_2\|_{L^{4r}}\,.$$

The same estimates hold for the checked variables, with $\mathbb{D}'_\ell$ in place of $\mathbb{D}_\ell$.

Proof. Since $y = Gu$, for $|h|_2 \leq 1$,

$$\mathbb{D}_\ell[y][h] = -G\left(\mathcal{L}_\ell[h] + \mathcal{B}_\ell[h]\right)y,$$

where

$$\mathcal{L}_\ell[h] := \frac{1}{n}L''(x_\ell^\top\hat{\theta})\left(hx_\ell^\top + x_\ell h^\top\right) + \frac{1}{n}L'''(x_\ell^\top\hat{\theta})(h^\top\hat{\theta})x_\ell x_\ell^\top,$$

and

$$\mathcal{B}_\ell[h] := \frac{1}{n}\sum_{m=1}^n L'''(x_m^\top\hat{\theta})\left(x_m^\top\mathbb{D}_\ell[\hat{\theta}][h]\right)x_m x_m^\top + \nabla^3\rho(\hat{\theta})\left[\mathbb{D}_\ell[\hat{\theta}][h]\right].$$

Under $\nabla^3\rho \equiv 0$, the last term vanishes.

For the local part,

$$\left|v^\top G\mathcal{L}_\ell[h]y\right| \le \frac{O(1)}{n}\big[\,|v|_2\;|x_\ell^\top y| + |x_\ell^\top Gv|\;|y|_2 + \left|\hat\theta\right|_2 |x_\ell^\top Gv|\,|x_\ell^\top y|\big].$$

Therefore

$$\left(\sum_{\ell=1}^n \left(\left|v^\top G\mathcal{L}_\ell y\right|^*\right)^2\right)^{1/2} \le \frac{O(1)}{n}\left[|v|_2\left|X^\top y\right|_2 + |y|_2\left|X^\top Gv\right|_2 + \left|\hat\theta\right|_2\left|X^\top y\right|_\infty\left|X^\top Gv\right|_2\right].$$

Using

$$|y|_2 = O(1), \qquad \left|X^\top y\right|_2 = O(\sqrt{n}), \qquad \left|X^\top Gv\right|_2 \le O(\sqrt{n})\,|v|_2\,,$$

and

$$\left\|\left|X^\top y\right|_\infty\right\|_{L^r} = O_r((\log n)^6)$$

by Lemma 5.4, we get

$$\left\|\left(\sum_{\ell=1}^n \left(\left|v^\top G\mathcal{L}_\ell y\right|^*\right)^2\right)^{1/2}\right\|_{L^r} \le \frac{O_r((\log n)^6)}{\sqrt{n}}\,\||v|_2\|_{L^{4r}}\,.$$

For the background part, using the displayed expression of $\mathcal{B}_\ell[h]$,

$$v^\top G\mathcal{B}_\ell[h]y = \frac{1}{n}\sum_{m=1}^n (x_m^\top Gv)L'''(x_m^\top\hat\theta)(x_m^\top\mathbb{D}_\ell[\hat\theta][h])(x_m^\top y).$$

Using

$$\mathbb{D}_\ell[\hat\theta][h] = -\frac{1}{n}G\left[L''(x_\ell^\top\hat\theta)(h^\top\hat\theta)x_\ell + s_\ell h\right], \qquad s_\ell := L'(x_\ell^\top\hat\theta),$$

and writing $F = \mathrm{diag}_m(L'''(x_m^\top\hat\theta))$, we obtain

$$\begin{aligned} v^\top G\mathcal{B}_\ell[h]y = &-\frac{L''(x_\ell^\top\hat\theta)(h^\top\hat\theta)}{n^2}(X^\top Gv\odot X^\top y)^\top FX^\top Gx_\ell \\ &-\frac{s_\ell}{n^2}(X^\top Gv\odot X^\top y)^\top FX^\top Gh. \end{aligned}$$

Hence

$$\left(\sum_{\ell=1}^n \left(\left|v^\top G\mathcal{B}_\ell y\right|^*\right)^2\right)^{1/2} \le \frac{O(1)}{n^2}\left[\left|\hat\theta\right|_2\left|X^\top GXF(X^\top Gv\odot X^\top y)\right|_2 + \left(\sum_{\ell=1}^n s_\ell^2\right)^{1/2}\left|GXF(X^\top Gv\odot X^\top y)\right|_2\right].$$

Since L''' is bounded,

$$\left|F(X^\top Gv\odot X^\top y)\right|_2 \le O(1)\left|X^\top y\right|_\infty\left|X^\top Gv\right|_2 \le O(\sqrt{n})\left|X^\top y\right|_\infty |v|_2\,.$$

Therefore

$$\left|X^\top GXF(X^\top Gv\odot X^\top y)\right|_2 \le O(n^{3/2})\left|X^\top y\right|_\infty |v|_2\,,$$

and

$$\left|GXF(X^\top Gv \odot X^\top y)\right|_2 \le O(n)\left|X^\top y\right|_\infty |v|_2 .$$

Using

$$\left\|\left|X^\top y\right|_\infty\right\|_{L^r} = O_r((\log n)^6), \qquad \left\|\left|\hat{\theta}\right|_2\right\|_{L^r} = O_r(1), \qquad \left\|\left(\sum_{\ell=1}^n s_\ell^2\right)^{1/2}\right\|_{L^r} = O_r(\sqrt{n}),$$

we obtain

$$\left\|\left(\sum_{\ell=1}^n \left(\left|v^\top G\mathcal{B}_\ell y\right|^*\right)^2\right)^{1/2}\right\|_{L^r} \le \frac{O_r((\log n)^6)}{\sqrt{n}} \||v|_2\|_{L^{4r}} .$$

This proves the first square-summed derivative estimate.

□

Lemma 5.19 (Background-resolvent active quadratic estimate)**.** *Under Assumptions 1–7 and* $\nabla^3\rho \equiv 0$*, for every fixed integers* $q, r \ge 1$*,*

$$\left\|\left(\sum_{\ell=1}^n \left[\sup_{|h|_2\le 1}\left\|x^\top GXDX^\top G\mathcal{B}_\ell[h]Gx\right\|_{L_x^q}\right]^2\right)^{1/2}\right\|_{L^r} = O_{q,r}\left(n^{3/2}(\log n)^6\right).$$

where for $\ell \in [n]$ *and* $h \in \mathbb{R}^p$*, we denoted,*

$$\mathcal{B}_\ell[h] := \frac{1}{n}X\Lambda_\ell(h)X^\top, \qquad \textit{with } \Lambda_\ell(h) := \mathrm{diag}_{m\in[n]}\, L'''(x_m^\top\hat{\theta})\, x_m^\top\mathbb{D}_\ell[\hat{\theta}][h].$$

Proof. Let us recall the notations:

$$y = Gu, \qquad D = \mathrm{diag}_{m\in[n]}\left(L'''(x_m^\top\hat{\theta})\, x_m^\top y\right),$$

we have

$$x^\top GXDX^\top G\mathcal{B}_\ell[h]Gx = x^\top A_\ell[h]x,$$

where

$$A_\ell[h] := \frac{1}{n}GXDX^\top GX\Lambda_\ell(h)X^\top G.$$

The conditional quadratic-form bound, applied conditionally on the background and to the symmetric part of $A_\ell[h]$, gives

$$\left\|x^\top A_\ell[h]x\right\|_{L_x^q} \le O_q\left(|A_\ell[h]|_F + |\mathrm{Tr}(A_\ell[h]\Sigma_x)|\right).$$

It is therefore enough to bound the Frobenius and trace contributions.

- For the Frobenius contribution, using

$$|GX|_{\mathrm{op}} = O(\sqrt{n}), \qquad \left|X^\top GX\right|_{\mathrm{op}} = O(n), \qquad \left|X^\top G\right|_{\mathrm{op}} = O(\sqrt{n}),$$

  we get, uniformly over $|h|_2 \le 1$,

$$|A_\ell[h]|_F \le \frac{1}{n}|GX|_{\mathrm{op}}|D|_{\mathrm{op}}\left|X^\top GX\right|_{\mathrm{op}}|\Lambda_\ell(h)|_F\left|X^\top G\right|_{\mathrm{op}} \le O(n)\,|D|_{\mathrm{op}}\,|\Lambda_\ell(h)|_F .$$

  Since L''' is bounded,

$$|\Lambda_\ell(h)|_F \le O(1)\left|X^\top\mathbb{D}_\ell[\hat{\theta}][h]\right|_2 .$$

Thus

$$\left\| \left(\sum_{\ell=1}^{n} \left[\sup_{|h|_2 \le 1} |A_\ell[h]|_F \right]^2 \right)^{1/2} \right\|_{L^r} \le O(n) \left\| |D|_{\mathrm{op}} \left(\sum_{\ell=1}^{n} \left(\left| X^\top \mathbb{D}_\ell[\hat{\theta}] \right|_2^* \right)^2 \right)^{1/2} \right\|_{L^r}.$$

By Lemmas 5.4 and 5.5,

$$\left\| |D|_{\mathrm{op}} \right\|_{L^{2r}} = O_r((\log n)^6),$$

and

$$\left\| \left(\sum_{\ell=1}^{n} \left(\left| X^\top \mathbb{D}_\ell[\hat{\theta}] \right|_2^* \right)^2 \right)^{1/2} \right\|_{L^{2r}} = O_r(\sqrt{n}).$$

Therefore

$$\left\| \left(\sum_{\ell=1}^{n} \left[\sup_{|h|_2 \le 1} |A_\ell[h]|_F \right]^2 \right)^{1/2} \right\|_{L^r} = O_r\left(n^{3/2} (\log n)^6 \right).$$

- We now treat the trace contribution. By cyclicity of the trace,

$$\mathrm{Tr}(A_\ell[h]\Sigma_x) = \frac{1}{n} \mathrm{Tr}\Big(\Lambda_\ell(h) X^\top G \Sigma_x G X D X^\top G X \Big).$$

Let

$$F := \mathrm{diag}_{m \in [n]} \left(L'''(x_m^\top \hat{\theta}) \right), \qquad \text{and} \qquad c := \mathrm{diag}\Big(X^\top G \Sigma_x G X D X^\top G X \Big) \in \mathbb{R}^n.$$

Since $\mathrm{diag}(\Lambda_\ell(h)) = F X^\top \mathbb{D}_\ell[\hat{\theta}][h]$, we have

$$\mathrm{Tr}(A_\ell[h]\Sigma_x) = \frac{1}{n} (XFc)^\top \mathbb{D}_\ell[\hat{\theta}][h].$$

Using the explicit derivative formula

$$\mathbb{D}_\ell[\hat{\theta}][h] = -\frac{1}{n} G \left[L''(x_\ell^\top \hat{\theta})(h^\top \hat{\theta}) x_\ell + s_\ell h \right],$$

we obtain, uniformly over $|h|_2 \le 1$,

$$\left| (XFc)^\top \mathbb{D}_\ell[\hat{\theta}][h] \right| \le \frac{O(1)}{n} \left(\left| \hat{\theta} \right|_2 |e_\ell^\top X^\top G X F c| + |s_\ell| \, |GXFc|_2 \right).$$

Therefore

$$\left(\sum_{\ell=1}^{n} \left[\sup_{|h|_2 \le 1} |\mathrm{Tr}(A_\ell[h]\Sigma_x)| \right]^2 \right)^{1/2} \le \frac{O(1)}{n^2} \left(\left| \hat{\theta} \right|_2 \left| X^\top G X F c \right|_2 + |s|_2 \, |GXFc|_2 \right).$$

Since $|F|_{\mathrm{op}} \le |L'''|_\infty \le O(1)$,

$$|GXFc|_2 \le |GX|_{\mathrm{op}} \, |c|_2 = O(\sqrt{n}) \, |c|_2, \qquad \text{and} \qquad \left| X^\top G X F c \right|_2 = O(n) \, |c|_2.$$

Moreover,

$$\left\| \left| \hat{\theta} \right|_2 \right\|_{L^{2r}} = O_r(1), \qquad \| |s|_2 \|_{L^{2r}} = O_r(\sqrt{n}).$$

Thus

$$\left\| \left(\sum_{\ell=1}^{n} \left[\sup_{|h|_2 \le 1} |\mathrm{Tr}(A_\ell[h]\Sigma_x)| \right]^2 \right)^{1/2} \right\|_{L^r} \le \frac{O_r(1)}{n} \, \||c|_2\|_{L^{2r}} .$$

It remains to bound c. Since

$$|c|_2 \le \left| X^\top G \Sigma_x G X D X^\top G X \right|_F \le \left| X^\top G \Sigma_x G X \right|_{\mathrm{op}} |D|_F \left| X^\top G X \right|_{\mathrm{op}} = O(n^{5/2}),$$

since

$$|\Sigma_x|_{\mathrm{op}} = O(1), \qquad \left| X^\top G \Sigma_x G X \right|_{\mathrm{op}} = O(n), \qquad \left| X^\top G X \right|_{\mathrm{op}} = O(n),$$

and

$$|D|_F \le O(1) \left| X^\top y \right|_2 = O(\sqrt{n}).$$

Consequently,

$$\left\| \left(\sum_{\ell=1}^{n} \left[\sup_{|h|_2 \le 1} |\mathrm{Tr}(A_\ell[h]\Sigma_x)| \right]^2 \right)^{1/2} \right\|_{L^r} = O_r(n^{3/2}).$$

Combining the Frobenius and trace bounds in the conditional quadratic-form estimate yields

$$\left\| \left(\sum_{\ell=1}^{n} \left[\sup_{|h|_2 \le 1} \left\| x^\top G X D X^\top G \mathcal{B}_\ell[h] G x \right\|_{L^q_x} \right]^2 \right)^{1/2} \right\|_{L^r} = O_{q,r}\left(n^{3/2} (\log n)^6 \right).$$

□

Lemma 5.20 (Active quadratic norm of the background matrix)**.** *Under Assumptions 1–7 and* $\nabla^3 \rho \equiv 0$*, for every fixed integers* $q, r \ge 1$*,*

$$\left\| \left(\sum_{\ell=1}^{n} (\mathcal{Q}_q(\mathbb{D}_\ell[Y])^*)^2 \right)^{1/2} \right\|_{L^r} = O_{q,r}\left(n^{3/2} (\log n)^7 \right),$$

where, for a matrix-valued statistic M*, we denote*

$$\mathcal{Q}_q(\mathbb{D}_\ell[M])^* := \sup_{|h|_2 \le 1} \left\| x^\top \mathbb{D}_\ell[M][h] x \right\|_{L^q_x} .$$

The same estimates hold for $\check{Y}$*, with* $\mathbb{D}'_\ell$ *in place of* $\mathbb{D}_\ell$*.*

Proof. Set

$$z := X^\top G x, \qquad a := X^\top y, \qquad D = \mathrm{diag}_{m \in [n]} \left(L'''(x_m^\top \hat{\theta}) a_m \right).$$

Then

$$x^\top Y x = x^\top G X D X^\top G x = z^\top D z.$$

For $|h|_2 \le 1$,

$$\mathbb{D}_\ell[z][h] = e_\ell h^\top G x + X^\top \mathbb{D}_\ell[G][h] x,$$

and therefore

$$\mathbb{D}_\ell[x^\top Y x][h] = 2 z^\top D e_\ell h^\top G x + z^\top \mathbb{D}_\ell[D][h] z + 2 z^\top D X^\top \mathbb{D}_\ell[G][h] x.$$

We bound these three terms.

First,

$$\left\| \left(\sum_{\ell=1}^n \left[\sup_{|h|_2 \le 1} \left\| D_\ell z_\ell h^\top G x \right\|_{L_x^q} \right]^2 \right)^{1/2} \right\|_{L^r} \le O_q \left\| \||Gx|_2\|_{L_x^{2q}} \left(\sum_{\ell=1}^n D_\ell^2 \, |Gx_\ell|_2^2 \right)^{1/2} \right\|_{L^r}$$

$$\le O_q \left\| \||Gx|_2\|_{L_x^{2q}} \, |D|_{\mathrm{op}} \, |GX|_F \right\|_{L^r} = O_{q,r}\left(n^{3/2} (\log n)^6 \right),$$

using

$$\||Gx|_2\|_{L_x^{2q}} = O_q(\sqrt{n}), \qquad \left\| |D|_{\mathrm{op}} \right\|_{L^{3r}} = O_r((\log n)^6), \qquad \||GX|_F\|_{L^{3r}} = O_r(n).$$

We now handle the term $z^\top \mathbb{D}_\ell[D] z$. For $m \in [n]$,

$$\mathbb{D}_\ell[D_m][h] = \mathbb{D}_\ell \left[L'''(x_m^\top \hat\theta) \right] [h] a_m + L'''(x_m^\top \hat\theta) \left(\mathbf{1}_{m=\ell} h^\top y + x_m^\top \mathbb{D}_\ell[y][h] \right),$$

with

$$\left| \mathbb{D}_\ell \left[L'''(x_m^\top \hat\theta) \right] [h] \right| \le O(1) \left| \mathbf{1}_{m=\ell} h^\top \hat\theta + x_m^\top \mathbb{D}_\ell[\hat\theta][h] \right|.$$

Thus $z^\top \mathbb{D}_\ell[D][h] z$ is the sum of the following three contributions:

$$\begin{aligned} T_\ell^{D,\mathrm{loc}}[h] &:= O(1) \left(h^\top \hat\theta \, a_\ell + h^\top y \right) z_\ell^2, \\ T_\ell^{D,\theta}[h] &:= O(1) \sum_{m=1}^n L'''(x_m^\top \hat\theta) a_m z_m^2 \, x_m^\top \mathbb{D}_\ell[\hat\theta][h], \\ T_\ell^{D,y}[h] &:= \sum_{m=1}^n L'''(x_m^\top \hat\theta) z_m^2 \, x_m^\top \mathbb{D}_\ell[y][h]. \end{aligned}$$

For the local contribution,

$$\left\| \left(\sum_{\ell=1}^n \left[\sup_{|h|_2 \le 1} \left\| T_\ell^{D,\mathrm{loc}}[h] \right\|_{L_x^q} \right]^2 \right)^{1/2} \right\|_{L^r}$$

$$\le O_q \left\| \left(\left|\hat\theta\right|_2 |a|_\infty + |y|_2 \right) \left(\sum_{\ell=1}^n \||z_\ell|^2\|_{L_x^q}^2 \right)^{1/2} \right\|_{L^r} \le O_q \left\| \left(\left|\hat\theta\right|_2 |a|_\infty + |y|_2 \right) \left(\sum_{\ell=1}^n |Gx_\ell|_2^4 \right)^{1/2} \right\|_{L^r}$$

$$\le O_q \left\| \left(\left|\hat\theta\right|_2 |a|_\infty + |y|_2 \right) |GX|_{\mathrm{op}} \, |GX|_F \right\|_{L^r} = O_{q,r}\left(n^{3/2} (\log n)^6 \right).$$

For the $\hat\theta$-background contribution, set

$$w_m := L'''(x_m^\top \hat\theta) a_m z_m^2, \qquad w = (w_m)_{m \in [n]}.$$

Then

$$T_\ell^{D,\theta}[h] = O(1) \sum_{m=1}^n w_m \, x_m^\top \mathbb{D}_\ell[\hat\theta][h].$$

Using

$$\mathbb{D}_\ell[\hat\theta][h] = -\frac{1}{n} G \left[L''(x_\ell^\top \hat\theta)(h^\top \hat\theta) x_\ell + s_\ell h \right],$$

we have

$$\left\| \left(\sum_{\ell=1}^{n} \left[\sup_{|h|_2 \le 1} \left\| T_\ell^{D,\theta}[h] \right\|_{L_x^q} \right]^2 \right)^{1/2} \right\|_{L^r}$$
$$\le \frac{O(1)}{n} \left\| \left|\hat{\theta}\right|_2 \left\| \left| X^\top G X w \right|_2 \right\|_{L_x^q} + |s|_2 \, \| |GXw|_2 \|_{L_x^q} \right\|_{L^r}.$$

Moreover,

$$\left\| \| |w|_2 \|_{L_x^q} \right\|_{L^{3r}} \le O_q \left\| |a|_\infty \left(\sum_{m=1}^{n} \| |z_m|^2 \|_{L_x^q}^2 \right)^{1/2} \right\|_{L^{3r}}$$
$$\le O_q \left\| |a|_\infty \, |GX|_{\mathrm{op}} \, |GX|_F \right\|_{L^{3r}} = O_{q,r}\left(n^{3/2} (\log n)^6 \right).$$

Hence, using

$$\left| X^\top G X \right|_{\mathrm{op}} = O(n), \qquad |GX|_{\mathrm{op}} = O(\sqrt{n}), \qquad \left\| \left|\hat{\theta}\right|_2 \right\|_{L^{3r}} = O_r(1), \qquad \| |s|_2 \|_{L^{3r}} = O_r(\sqrt{n}),$$

we obtain

$$\left\| \left(\sum_{\ell=1}^{n} \left[\sup_{|h|_2 \le 1} \left\| T_\ell^{D,\theta}[h] \right\|_{L_x^q} \right]^2 \right)^{1/2} \right\|_{L^r} = O_{q,r}\left(n^{3/2} (\log n)^6 \right).$$

For the y-background contribution, set

$$v_m := L'''(x_m^\top \hat{\theta}) z_m^2, \qquad v = (v_m)_{m \in [n]}.$$

Then

$$T_\ell^{D,y}[h] = \sum_{m=1}^{n} v_m \, x_m^\top \mathbb{D}_\ell[y][h].$$

Since

$$\mathbb{D}_\ell[y][h] = -G \left(\mathcal{L}_\ell[h] + \mathcal{B}_\ell[h] \right) y,$$

the local part satisfies

$$\left\| \left(\sum_{\ell=1}^{n} \left[\sup_{|h|_2 \le 1} \left\| \sum_{m=1}^{n} v_m x_m^\top G \mathcal{L}_\ell[h] y \right\|_{L_x^q} \right]^2 \right)^{1/2} \right\|_{L^r}$$
$$\le \frac{O(1)}{n} \left\| \| |GXv|_2 \|_{L_x^q} \left| X^\top y \right|_2 + |y|_2 \left\| \left| X^\top G X v \right|_2 \right\|_{L_x^q} + \left|\hat{\theta}\right|_2 |a|_\infty \left\| \left| X^\top G X v \right|_2 \right\|_{L_x^q} \right\|_{L^r}.$$

Also,

$$\left\| \| |v|_2 \|_{L_x^q} \right\|_{L^{3r}} \le O_q \left\| \left(\sum_{m=1}^{n} \| |z_m|^2 \|_{L_x^q}^2 \right)^{1/2} \right\|_{L^{3r}}$$
$$\le O_q \left\| |GX|_{\mathrm{op}} \, |GX|_F \right\|_{L^{3r}} = O_{q,r}(n^{3/2}).$$

Therefore the local y-background contribution is

$$\left\| \left(\sum_{\ell=1}^{n} \left[\sup_{|h|_2 \le 1} \left\| \sum_{m=1}^{n} v_m x_m^\top G \mathcal{L}_\ell[h] y \right\|_{L_x^q} \right]^2 \right)^{1/2} \right\|_{L^r} = O_{q,r}\left(n^{3/2} (\log n)^6 \right).$$

For the background part of $\mathbb{D}_\ell[y]$, we argue conditionally on the active column x. Then the coefficient vector

$$v_m = L'''(x_m^\top \hat{\theta}) z_m^2$$

is frozen, and the background-part computation in the proof of Lemma 5.18 applies with v in place of the deterministic coefficient vector. It gives, conditionally on x,

$$\begin{aligned}
&\left(\sum_{\ell=1}^n \left[\sup_{|h|_2 \le 1} \left| \sum_{m=1}^n v_m x_m^\top G \mathcal{B}_\ell[h] y \right| \right]^2 \right)^{1/2} \\
&\quad \le \frac{O(1)}{n^2} \left[\left|\hat{\theta}\right|_2 \left| X^\top G X \right|_{\mathrm{op}} \left| X^\top y \right|_\infty \left| X^\top G X v \right|_2 + |s|_2 \, |GX|_{\mathrm{op}} \left| X^\top y \right|_\infty \left| X^\top G X v \right|_2 \right].
\end{aligned}$$

Using

$$\left| X^\top G X v \right|_2 \le \left| X^\top G X \right|_{\mathrm{op}} |v|_2 = O(n) \, |v|_2 \, ,$$

together with

$$\left\| \| |v|_2 \|_{L_x^q} \right\|_{L^{3r}} = O_{q,r}(n^{3/2}),$$

and the bounds

$$\left\| \left| X^\top y \right|_\infty \right\|_{L^{3r}} = O_r((\log n)^6), \qquad \left\| \left| \hat{\theta} \right|_2 \right\|_{L^{3r}} = O_r(1), \qquad \| |s|_2 \|_{L^{3r}} = O_r(\sqrt{n}),$$

we obtain

$$\left\| \left(\sum_{\ell=1}^n \left[\sup_{|h|_2 \le 1} \left\| \sum_{m=1}^n v_m x_m^\top G \mathcal{B}_\ell[h] y \right\|_{L_x^q} \right]^2 \right)^{1/2} \right\|_{L^r} = O_{q,r}\left(n^{3/2} (\log n)^6 \right).$$

Thus

$$\left\| \left(\sum_{\ell=1}^n \left[\sup_{|h|_2 \le 1} \left\| z^\top \mathbb{D}_\ell[D][h] z \right\|_{L_x^q} \right]^2 \right)^{1/2} \right\|_{L^r} = O_{q,r}\left(n^{3/2} (\log n)^6 \right).$$

It remains to bound the resolvent contribution

$$2 z^\top D X^\top \mathbb{D}_\ell[G][h] x = -2 z^\top D X^\top G \mathcal{L}_\ell[h] G x - 2 z^\top D X^\top G \mathcal{B}_\ell[h] G x.$$

For the local part,

$$\begin{aligned}
&\left\| \left(\sum_{\ell=1}^n \left[\sup_{|h|_2 \le 1} \left\| z^\top D X^\top G \mathcal{L}_\ell[h] G x \right\|_{L_x^q} \right]^2 \right)^{1/2} \right\|_{L^r} \\
&\quad \le \frac{O(1)}{n} \left\| \left\| |GXDz|_2 \, |z|_2 + \left| X^\top G X D z \right|_2 |Gx|_2 + \left| \hat{\theta} \right|_2 \left| X^\top G X D z \right|_2 |z|_\infty \right\|_{L_x^q} \right\|_{L^r} \\
&\quad \le O_{q,r}\left(n^{3/2} (\log n)^7 \right).
\end{aligned}$$

Indeed, we used

$$|GXDz|_2 \le |GX|_{\mathrm{op}} \, |D|_{\mathrm{op}} \, |z|_2 \, , \qquad \left| X^\top G X D z \right|_2 \le \left| X^\top G X \right|_{\mathrm{op}} |D|_{\mathrm{op}} \, |z|_2 \, ,$$

together with

$$\| |z|_2 \|_{L_x^{3q}} = O_q(n), \qquad \| |z|_\infty \|_{L_x^{3q}} = O_q(\sqrt{n} \log n), \qquad \| |Gx|_2 \|_{L_x^{3q}} = O_q(\sqrt{n}).$$

For the background part, Lemma 5.19 gives directly

$$\left\| \left(\sum_{\ell=1}^{n} \left[\sup_{|h|_2 \le 1} \left\| z^\top D X^\top G \mathcal{B}_\ell[h] G x \right\|_{L^q_x} \right]^2 \right)^{1/2} \right\|_{L^r} = O_{q,r}\left(n^{3/2} (\log n)^6 \right).$$

Combining the three displayed bounds yields

$$\left\| \left(\sum_{\ell=1}^{n} (\mathcal{Q}_q(\mathbb{D}_\ell[Y])^*)^2 \right)^{1/2} \right\|_{L^r} = O_{q,r}\left(n^{3/2} (\log n)^7 \right).$$

The checked estimates are identical, since the checked background satisfies the same assumptions and has the same distribution. □

Lemma 5.21 (Background gradients of the active contractions)**.** *Under Assumptions 1–7 and* $\nabla^3 \rho \equiv 0$*, for every fixed integer* $r \ge 1$*,*

$$\left\| \left(\sum_{\ell=1}^{n} \left(|\mathbb{D}_\ell[\mathcal{C}^{00}(y, \check{y})]|^* \right)^2 + \sum_{\ell=1}^{n} \left(|\mathbb{D}'_\ell[\mathcal{C}^{00}(y, \check{y})]|^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r\left(\frac{(\log n)^6}{\sqrt{n}} \right),$$

$$\left\| \left(\sum_{\ell=1}^{n} \left(|\mathbb{D}_\ell[\mathcal{C}^{01}(y, \check{Y})]|^* \right)^2 + \sum_{\ell=1}^{n} \left(|\mathbb{D}'_\ell[\mathcal{C}^{01}(y, \check{Y})]|^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r\left(n^{3/2} (\log n)^{12} \right),$$

and the same bound holds with $\mathcal{C}^{01}(y, \check{Y})$ *replaced by* $\mathcal{C}^{10}(Y, \check{y})$*. Finally,*

$$\left\| \left(\sum_{\ell=1}^{n} \left(|\mathbb{D}_\ell[\mathcal{C}^{11}(Y, \check{Y})]|^* \right)^2 + \sum_{\ell=1}^{n} \left(|\mathbb{D}'_\ell[\mathcal{C}^{11}(Y, \check{Y})]|^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r\left(n^{7/2} (\log n)^{13} \right).$$

Proof. We will skip some of the terms appearing from the derivations $\mathbb{D}'_\ell$ when they are symmetric to those appearing through derivations $\mathbb{D}_\ell$ (for the terms $\mathcal{C}^{00}$ and $\mathcal{C}^{11}$). We use when useful the active-contraction bounds given by Lemma 5.17:

$$|\mathcal{C}^{00}(a, b)| \le R\, |a|_2\, |b|_2,$$

$$|\mathcal{C}^{01}(a, C)| \le R\, |a|_2\, \mathcal{Q}_4(C), \qquad |\mathcal{C}^{10}(B, b)| \le R \mathcal{Q}_4(B)\, |b|_2,$$

and

$$|\mathcal{C}^{11}(B, C)| \le R \mathcal{Q}_4(B) \mathcal{Q}_4(C), \qquad \|R\|_{L^r} = O_r(1).$$

For $\mathcal{C}^{00}(y, \check{y})$, the product rule gives

$$\mathbb{D}_\ell[\mathcal{C}^{00}(y, \check{y})][h] = \mathrm{Cov}_x\left(\mathbb{D}_\ell[s^\zeta_x][h]\, y^\top x, \check{s}^\zeta_x\, \check{y}^\top x \right) + \mathcal{C}^{00}(\mathbb{D}_\ell[y][h], \check{y}).$$

By Lemma 5.16,

$$\left\| \left(\sum_{\ell=1}^{n} \left(\left| \mathrm{Cov}_x\left(\mathbb{D}_\ell[s^\zeta_x]\, y^\top x, \check{s}^\zeta_x\, \check{y}^\top x \right) \right|^* \right)^2 + \sum_{\ell=1}^{n} \left(\left| \mathrm{Cov}_x\left(s^\zeta_x\, y^\top x, \mathbb{D}'_\ell[\check{s}^\zeta_x]\, \check{y}^\top x \right) \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} \le O_r(n^{-1/2}).$$

Moreover, by duality applied to $\mathcal{C}^{00}$, there exist background-measurable vectors v_{00} and $\check{v}_{00}$ such that

$$\mathcal{C}^{00}(a, \check{y}) = \check{v}_{00}^\top a, \qquad \mathcal{C}^{00}(y, b) = v_{00}^\top b,$$

with

$$|\check{v}_{00}|_2 \le R\,|\check{y}|_2\,, \qquad |v_{00}|_2 \le R\,|y|_2\,.$$

Therefore, by Lemma 5.18 and Lemma 5.15,

$$\begin{aligned}\left\| \left(\sum_{\ell=1}^{n} \left(\left| \mathcal{C}^{00}(\mathbb{D}_\ell[y], \check{y}) \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} &\le \frac{O_r((\log n)^6)}{\sqrt{n}} \, \| |\check{v}_{00}|_2 \|_{L^{4r}} \\ &\le \frac{O_r((\log n)^6)}{\sqrt{n}} \, \| R\,|\check{y}|_2 \|_{L^{4r}} = O_r\left(\frac{(\log n)^6}{\sqrt{n}} \right),\end{aligned}$$

and similarly

$$\left\| \left(\sum_{\ell=1}^{n} \left(\left| \mathcal{C}^{00}(y, \mathbb{D}'_\ell[\check{y}]) \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r\left(\frac{(\log n)^6}{\sqrt{n}} \right).$$

This proves the 00-bound.

We now treat $\mathcal{C}^{01}(y, \check{Y})$. The product rule gives

$$\begin{aligned}\mathbb{D}_\ell[\mathcal{C}^{01}(y, \check{Y})][h] = \mathrm{Cov}_x\Big(&\mathbb{D}_\ell[s_x^\zeta][h]\, y^\top x, (\check{s}_x^\zeta)^2 x^\top \check{Y} x \Big) \\ &+ \mathcal{C}^{01}(\mathbb{D}_\ell[y][h], \check{Y}).\end{aligned}$$

By Lemma 5.16,

$$\left\| \left(\sum_{\ell=1}^{n} \left(\left| \mathrm{Cov}_x\Big(\mathbb{D}_\ell[s_x^\zeta]\, y^\top x, (\check{s}_x^\zeta)^2 x^\top \check{Y} x \Big) \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} \le O_r(n^{3/2} (\log n)^6).$$

By duality applied to $\mathcal{C}^{01}$, there exists $\check{v}_{01}$ such that

$$\mathcal{C}^{01}(a, \check{Y}) = \check{v}_{01}^\top a, \qquad |\check{v}_{01}|_2 \le R\mathcal{Q}_4(\check{Y}).$$

Thus, by Lemmas 5.18 and 5.15,

$$\left\| \left(\sum_{\ell=1}^{n} \left(\left| \mathcal{C}^{01}(\mathbb{D}_\ell[y], \check{Y}) \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} \le \frac{O_r((\log n)^6)}{\sqrt{n}} \, \| R\mathcal{Q}_4(\check{Y}) \|_{L^{4r}} \le O_r\Big(n^{3/2} (\log n)^{12} \Big).$$

For the checked matrix derivative, the active contraction bound gives

$$\left(\sum_{\ell=1}^{n} \left(\left| \mathrm{Cov}_x\Big(s_x^\zeta\, y^\top x, (\check{s}_x^\zeta)^2 x^\top \mathbb{D}'_\ell[\check{Y}] x \Big) \right|^* \right)^2 \right)^{1/2} \le R\,|y|_2 \left(\sum_{\ell=1}^{n} \left(\mathcal{Q}_4(\mathbb{D}'_\ell[\check{Y}])^* \right)^2 \right)^{1/2}.$$

Consequently, by Lemmas 5.20 and 5.15,

$$\begin{aligned}\left\| \left(\sum_{\ell=1}^{n} \left(\left| \mathrm{Cov}_x\Big(s_x^\zeta\, y^\top x, (\check{s}_x^\zeta)^2 x^\top \mathbb{D}'_\ell[\check{Y}] x \Big) \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} &\le \left\| R\,|y|_2 \left(\sum_{\ell=1}^{n} \left(\mathcal{Q}_4(\mathbb{D}'_\ell[\check{Y}])^* \right)^2 \right)^{1/2} \right\|_{L^r} \\ &\le O_r\Big(n^{3/2} (\log n)^7 \Big).\end{aligned}$$

Combining the three estimates yields

$$\left\| \left(\sum_{\ell=1}^{n} \left(\left| \mathbb{D}_\ell[\mathcal{C}^{01}(y, \check{Y})] \right|^* \right)^2 + \sum_{\ell=1}^{n} \left(\left| \mathbb{D}'_\ell[\mathcal{C}^{01}(y, \check{Y})] \right|^* \right)^2 \right)^{1/2} \right\|_{L^r} = O_r\Big(n^{3/2} (\log n)^{12} \Big).$$

The proof for $\mathcal{C}^{10}(Y,\check y)$ is identical, with checked and unchecked quantities interchanged.

It remains to treat $\mathcal{C}^{11}(Y,\check Y)$. The differentiated-score terms are controlled by Lemma 5.16:

$$\left\|\left(\sum_{\ell=1}^n\left(\left|\mathrm{Cov}_x\left(2s_x^\zeta\mathbb{D}_\ell[s_x^\zeta]\,x^\top Yx,(\check s_x^\zeta)^2x^\top\check Yx\right)\right|^*\right)^2\right)^{1/2}\right\|_{L^r}\le O_r(n^{7/2}(\log n)^{12}),$$

and the checked differentiated-score term is controlled in the same way. For the unchecked matrix derivative,

$$\left(\sum_{\ell=1}^n\left(\left|\mathrm{Cov}_x\left((s_x^\zeta)^2x^\top\mathbb{D}_\ell[Y]x,(\check s_x^\zeta)^2x^\top\check Yx\right)\right|^*\right)^2\right)^{1/2}\le R\mathcal{Q}_4(\check Y)\left(\sum_{\ell=1}^n(\mathcal{Q}_4(\mathbb{D}_\ell[Y])^*)^2\right)^{1/2}.$$

Therefore, by Lemma 5.15 and5.20,

$$\begin{aligned}\left\|\left(\sum_{\ell=1}^n\left(\left|\mathrm{Cov}_x\left((s_x^\zeta)^2x^\top\mathbb{D}_\ell[Y]x,(\check s_x^\zeta)^2x^\top\check Yx\right)\right|^*\right)^2\right)^{1/2}\right\|_{L^r}&\le\left\|R\mathcal{Q}_4(\check Y)\left(\sum_{\ell=1}^n(\mathcal{Q}_4(\mathbb{D}_\ell[Y])^*)^2\right)^{1/2}\right\|_{L^r}\\&\le O_r\left(n^{7/2}(\log n)^{13}\right).\end{aligned}$$

The checked matrix derivative satisfies the same bound:

$$\left\|\left(\sum_{\ell=1}^n\left(\left|\mathrm{Cov}_x\left((s_x^\zeta)^2x^\top Yx,(\check s_x^\zeta)^2x^\top\mathbb{D}'_\ell[\check Y]x\right)\right|^*\right)^2\right)^{1/2}\right\|_{L^r}\le O_r\left(n^{7/2}(\log n)^{13}\right).$$

Combining the differentiated-score and matrix-derivative estimates gives

$$\left\|\left(\sum_{\ell=1}^n\left(\left|\mathbb{D}_\ell[\mathcal{C}^{11}(Y,\check Y)]\right|^*\right)^2+\sum_{\ell=1}^n\left(\left|\mathbb{D}'_\ell[\mathcal{C}^{11}(Y,\check Y)]\right|^*\right)^2\right)^{1/2}\right\|_{L^r}=O_r\left(n^{7/2}(\log n)^{13}\right).$$

This completes the proof. □

Combining Lemma 5.21 with (5.1) and Remark 4.20, we get the following bound.

Corollary 5.22 (Bound on $c_n^{\bar x}$). *Under Assumptions 1–7 and* $\nabla^3\rho\equiv0$*,*

$$c_n^{\bar x}=O\left(\frac{(\log n)^{13}}{n^{3/2}}\right).$$

Proof. Fix the active index j. In the generic notation of this subsection,

$$\Phi_j:=\mathrm{Cov}_j\left(\mathfrak{d}_j(X),\mathfrak{d}_j(X^{A_j})\right)$$

has the same form as $\mathrm{Cov}_x(\mathfrak{d}_x,\check{\mathfrak{d}}_x)$. By (5.1),

$$\Phi_j=\frac1n\mathcal{C}^{00}(y,\check y)+\frac{1}{2n^3}\mathcal{C}^{01}(y,\check Y)+\frac{1}{2n^3}\mathcal{C}^{10}(Y,\check y)+\frac{1}{4n^5}\mathcal{C}^{11}(Y,\check Y).$$

We apply the tensorized Poincaré inequality to Φ_j as a function of the background variables and their independent copies. For coordinates that are shared by X and X^{A_j}, the derivative is the sum of the unchecked and checked derivatives; hence it is bounded by the square-sum of the two contributions controlled in Lemma 5.21. Therefore

$$\sqrt{\mathrm{Var}(\Phi_j)}\le O(1)\left\|\left(\sum_\ell(|\mathbb{D}_\ell[\Phi_j]|^*)^2+\sum_\ell\left(|\mathbb{D}'_\ell[\Phi_j]|^*\right)^2\right)^{1/2}\right\|_{L^2}.$$

Using Lemma 5.21 and the decomposition above, we obtain

$$\sqrt{\mathrm{Var}(\Phi_j)} \le \frac{1}{n} O\Big(\frac{(\log n)^6}{\sqrt{n}}\Big) + \frac{1}{n^3} O\Big(n^{3/2}(\log n)^{12}\Big) + \frac{1}{n^5} O\Big(n^{7/2}(\log n)^{13}\Big)$$
$$= O\Big(\frac{(\log n)^{13}}{n^{3/2}}\Big),$$

uniformly in j. Hence

$$\sup_{j\in[n]} \sqrt{\mathrm{Var}\big(\mathrm{Cov}_j\big(\mathfrak{d}_j(X), \mathfrak{d}_j(X^{A_j})\big)\big)} = O\Big(\frac{(\log n)^{13}}{n^{3/2}}\Big).$$

This bound transfers to $c_n^{\bar{x}}$ since, by Remark 4.20, replacing $\delta_j f_n$ by $\mathfrak{d}_j$ changes the corresponding $c_n^{\bar{x}}$ quantity by only and order $O\Big(\frac{(\log n)^3}{n^{3/2}}\Big)$. □

A Perturbative Wasserstein bound

With the notation introduced around Theorem 2.2, for $i \in [n]$ and $i \notin A$, define the replacement increment

$$\Delta_i g_n(Z^A) := g_n(Z^A) - g_n(Z^{A\cup\{i\}}).$$

Theorem A.1 (Wasserstein part of the perturbative bound of Shao–Zhang [18])**.** *Assume that* $\sigma_{g,n}^2 := \mathrm{Var}(g_n(Z)) > 0$*, and set*

$$W_{g,n} := \frac{g_n(Z) - \mathbb{E}g_n(Z)}{\sigma_{g,n}}.$$

Let $\mathcal{N}$ *be a standard normal random variable and define*

$$V_{g,n} := \frac{1}{2\sigma_{g,n}^2} \sum_{i=1}^n \Delta_i g_n(Z) \Delta_i g_n(Z^{A_i}).$$

Assume that the quantities on the right-hand side below are finite. Then

$$d_W(W_{g,n}, \mathcal{N}) \le \sqrt{\mathrm{Var}(V_{g,n})} + \frac{1}{\sigma_{g,n}^3} \sum_{i=1}^n \mathbb{E}|\Delta_i g_n(Z)|^3.$$

For $A \subset [n] \setminus \{i\}$, the replacement increment can be written in terms of the leave-one-out increments as

$$\Delta_i g_n(Z^A) = \delta_i g_n(Z^A) - \delta_i g_n(Z^{A\cup\{i\}}).$$

In particular, define

$$Y_i := \Delta_i g_n(Z) \Delta_i g_n(Z^{A_i}).$$

Then

$$Y_i = \big(\delta_i g_n(Z) - \delta_i g_n(Z^{\{i\}})\big)\big(\delta_i g_n(Z^{A_i}) - \delta_i g_n(Z^{A_i\cup\{i\}})\big).$$

For $j \in [n]$, let

$$\mathcal{F}_{-\mathbf{j}} := \sigma\big((z_\ell, z'_\ell) : \ell \ne j\big),$$

and let $\mathbb{E}_{\mathbf{j}}$ denote conditional expectation with respect to the pair (z_j, z'_j), with all other pairs fixed. We shall use the following consequence of the definition of c_n^z. For every $i \ne j$ and every $B \subset [n]$,

$$\big\|\delta_i g_n(Z^B) - \mathbb{E}_{\mathbf{j}}\big[\delta_i g_n(Z^B)\big]\big\|_{L^4} \le c_n^z.$$

Indeed, if $j \notin B$, then $\delta_i g_n(Z^B)$ depends on z_j but not on z'_j, so $\mathbb{E}_{\mathbf{j}}$ reduces to conditional expectation with respect to z_j. The bound is then exactly the definition of c_n^z, applied to the vector Z^B, whose coordinates have the same joint law as those of Z. If $j \in B$, then $\delta_i g_n(Z^B)$ depends on z'_j but not on z_j, and the same argument applies with z'_j in place of z_j.

Consequently, for every $i \neq j$ and every $B \subset [n]$,

$$\left\|\Delta_i g_n(Z^B) - \mathbb{E}_{\mathbf{j}}\left[\Delta_i g_n(Z^B)\right]\right\|_{L^4} \leq 2c_n^z.$$

Proof of Theorem 2.2. For any $k \geq 1$,

$$\|\Delta_i g_n(Z)\|_{L^k} = \left\|\delta_i g_n(Z) - \delta_i g_n(Z^{\{i\}})\right\|_{L^k} \leq 2\,\|\delta_i g_n(Z)\|_{L^k},$$

because $\delta_i g_n(Z^{\{i\}})$ has the same distribution as $\delta_i g_n(Z)$. Hence $\|\cdot\|_{L^3} \leq \|\cdot\|_{L^4}$ gives

$$\sum_{i=1}^n \mathbb{E}|\Delta_i g_n(Z)|^3 \leq 8n(m_n^{(4)})^3.$$

We next bound the variance term. Since

$$V_{g,n} = \frac{1}{2\sigma_{g,n}^2}\sum_{i=1}^n Y_i,$$

it is enough to control $\mathrm{Var}(\sum_i Y_i)$. We use

$$\mathrm{Var}\left(\sum_{i=1}^n Y_i\right) \leq n \sup_{i\in[n]} \mathrm{Var}(Y_i) + n(n-1) \sup_{\substack{i,j\in[n]\\ i\neq j}} |\operatorname{Cov}(Y_i, Y_j)|.$$

Since Z^{A_i} has the same distribution as Z, the variables $\Delta_i g_n(Z)$ and $\Delta_i g_n(Z^{A_i})$ have the same distribution. Therefore Hölder's inequality yields

$$\mathrm{Var}(Y_i) \leq \|Y_i\|_{L^2}^2 \leq \|\Delta_i g_n(Z)\|_{L^4}^2 \left\|\Delta_i g_n(Z^{A_i})\right\|_{L^4}^2 \leq 16(m_n^{(4)})^4.$$

It remains to control the off-diagonal covariances. Fix $i \neq j$. We decompose

$$\operatorname{Cov}(Y_i, Y_j) = \operatorname{Cov}(Y_i - \mathbb{E}_{\mathbf{j}} Y_i, Y_j) + \operatorname{Cov}(\mathbb{E}_{\mathbf{j}} Y_i, Y_j).$$

For the first term, recall that

$$\Delta_i g_n(Z) = \delta_i g_n(Z) - \delta_i g_n(Z^{\{i\}}), \qquad \Delta_i g_n(Z^{A_i}) = \delta_i g_n(Z^{A_i}) - \delta_i g_n(Z^{A_i\cup\{i\}}).$$

Applying the preceding observation with $B = \varnothing$ and $B = A_i$, we get uniformly over $i \neq j$,

$$\|\Delta_i g_n(Z) - \mathbb{E}_{\mathbf{j}}\Delta_i g_n(Z)\|_{L^4} + \left\|\Delta_i g_n(Z^{A_i}) - \mathbb{E}_{\mathbf{j}}\Delta_i g_n(Z^{A_i})\right\|_{L^4} \leq Cc_n^z.$$

Indeed, if z_j has been replaced by z'_j in the argument of $\delta_i g_n$, the same estimate follows because z'_j has the same law as z_j and is independent of the remaining variables. Moreover,

$$\|\Delta_i g_n(Z)\|_{L^4} + \left\|\Delta_i g_n(Z^{A_i})\right\|_{L^4} \leq Cm_n^{(4)}.$$

Set

$$H_i := \mathbb{E}_{\mathbf{j}}\Delta_i g_n(Z)\,\mathbb{E}_{\mathbf{j}}\Delta_i g_n(Z^{A_i}).$$

Then H_i is $\mathcal{F}_{-\mathbf{j}}$-measurable. Since conditional expectation is the L^2-projection onto $\mathcal{F}_{-\mathbf{j}}$,

$$\|Y_i - \mathbb{E}_{\mathbf{j}} Y_i\|_{L^2} \leq \|Y_i - H_i\|_{L^2}.$$

Since

$$Y_i - H_i = \big(\Delta_i g_n(Z) - \mathbb{E}_{\mathbf{j}}\Delta_i g_n(Z)\big)\Delta_i g_n(Z^{A_i}) + \mathbb{E}_{\mathbf{j}}\Delta_i g_n(Z)\big(\Delta_i g_n(Z^{A_i}) - \mathbb{E}_{\mathbf{j}}\Delta_i g_n(Z^{A_i})\big),$$

Hölder's inequality yields

$$\|Y_i - \mathbb{E}_{\mathbf{j}}Y_i\|_{L^2} \leq C m_n^{(4)} c_n^z.$$

Also $\|Y_j\|_{L^2} \leq C(m_n^{(4)})^2$. Hence

$$|\mathrm{Cov}(Y_i - \mathbb{E}_{\mathbf{j}}Y_i, Y_j)| \leq C(m_n^{(4)})^3 c_n^z.$$

For the second term, since $\mathbb{E}_{\mathbf{j}}Y_i$ is $\mathcal{F}_{-\mathbf{j}}$-measurable,

$$\mathrm{Cov}(\mathbb{E}_{\mathbf{j}}Y_i, Y_j) = \mathrm{Cov}(\mathbb{E}_{\mathbf{j}}Y_i, \mathbb{E}_{\mathbf{j}}Y_j).$$

Moreover,

$$\|\mathbb{E}_{\mathbf{j}}Y_i\|_{L^2} \leq \|Y_i\|_{L^2} \leq C(m_n^{(4)})^2.$$

It remains to identify $\mathbb{E}_{\mathbf{j}}Y_j$. Conditionally on $\mathcal{F}_{-\mathbf{j}}$, the pair

$$\big(\delta_j g_n(Z), \delta_j g_n(Z^{A_j})\big)$$

depends only on z_j, while

$$\big(\delta_j g_n(Z^{\{j\}}), \delta_j g_n(Z^{A_j\cup\{j\}})\big)$$

depends only on z'_j. These two pairs are conditionally independent copies. Therefore

$$\begin{aligned}\mathbb{E}_{\mathbf{j}}Y_j &= \mathbb{E}_{\mathbf{j}}\big[\big(\delta_j g_n(Z) - \delta_j g_n(Z^{\{j\}})\big)\big(\delta_j g_n(Z^{A_j}) - \delta_j g_n(Z^{A_j\cup\{j\}})\big)\big] \\ &= 2\,\mathrm{Cov}_j\big(\delta_j g_n(Z), \delta_j g_n(Z^{A_j})\big).\end{aligned}$$

Therefore, by the definition of $c_n^{\bar{z}}$,

$$\sqrt{\mathrm{Var}(\mathbb{E}_{\mathbf{j}}Y_j)} = 2\sqrt{\mathrm{Var}\big(\mathrm{Cov}_j\big(\delta_j g_n(Z), \delta_j g_n(Z^{A_j})\big)\big)} \leq 2c_n^{\bar{z}}.$$

On the other hand, the preceding L^2-bound gives

$$\sqrt{\mathrm{Var}(\mathbb{E}_{\mathbf{j}}Y_i)} \leq \|\mathbb{E}_{\mathbf{j}}Y_i\|_{L^2} \leq C(m_n^{(4)})^2.$$

Hence

$$|\mathrm{Cov}(\mathbb{E}_{\mathbf{j}}Y_i, \mathbb{E}_{\mathbf{j}}Y_j)| \leq \sqrt{\mathrm{Var}(\mathbb{E}_{\mathbf{j}}Y_i)}\sqrt{\mathrm{Var}(\mathbb{E}_{\mathbf{j}}Y_j)} \leq C(m_n^{(4)})^2 c_n^{\bar{z}}.$$

Combining the two off-diagonal estimates, we obtain

$$\sup_{i\neq j}|\,\mathrm{Cov}(Y_i, Y_j)| \leq C\left((m_n^{(4)})^3 c_n^z + (m_n^{(4)})^2 c_n^{\bar{z}}\right).$$

Therefore

$$\mathrm{Var}\left(\sum_{i=1}^n Y_i\right) \leq C\left[n(m_n^{(4)})^4 + n(n-1)\left((m_n^{(4)})^3 c_n^z + (m_n^{(4)})^2 c_n^{\bar{z}}\right)\right].$$

Using the definition of $V_{g,n}$ and Theorem A.1 gives the claimed bound. □

B Consequences of the Poincaré inequality

We say that a random vector $Z \in \mathbb{R}^m$ satisfies a Poincaré inequality with constant C_P if, for every sufficiently smooth mapping $F : \mathbb{R}^m \to \mathbb{R}$,

$$\operatorname{Var}(F(Z)) \le C_P \mathbb{E}\left[|\nabla_Z F(Z)|^2\right]. \tag{B.1}$$

We start with the standard tensorization property of the Poincaré inequality; see, for instance, [9, 12].

Proposition B.1 (Poincaré inequality tensorization)**.** *Let $Z_1, \dots, Z_n \in \mathbb{R}^p$ be independent random vectors, and denote $Z := (Z_1, \dots, Z_n) \in \mathcal{M}_{p,n}$. Assume that each Z_i satisfies a Poincaré inequality with constant C_P. Then, for every sufficiently smooth $F : \mathcal{M}_{p,n} \to \mathbb{R}$,*

$$\operatorname{Var}(F(Z)) \le C_P \mathbb{E}\left[|\nabla F(Z)|_F^2\right] = C_P \sum_{i=1}^{n} \mathbb{E}\left[|\nabla_i F(Z)|^2\right].$$

Proof. By the variance decomposition, or equivalently the Efron–Stein inequality for independent variables,

$$\operatorname{Var}(F(Z)) \le \sum_{i=1}^{n} \mathbb{E}\left[\operatorname{Var}_{Z_i}\left(F(Z_1, \dots, Z_n)\right)\right],$$

where Var_{Z_i} denotes variance with respect to Z_i only, all other coordinates being fixed. For each i, conditionally on $(Z_1, \dots, Z_{i-1}, Z_{i+1}, \dots, Z_n)$, the map $z_i \mapsto F(Z_1, \dots, Z_{i-1}, z_i, Z_{i+1}, \dots, Z_n)$ is a function on $\mathbb{R}^p$. Applying the Poincaré inequality for Z_i gives

$$\operatorname{Var}_{Z_i}\big(F(Z)\big) \le C_P \mathbb{E}_{Z_i}\left[|\nabla_i F(Z)|_2^2\right].$$

Taking expectation with respect to the remaining variables and summing over i yields

$$\operatorname{Var}(F(Z)) \le C_P \sum_{i=1}^{n} \mathbb{E}\left[|\nabla_i F(Z)|_2^2\right] = C_P \mathbb{E}\left[|\nabla F(Z)|_F^2\right].$$

□

The tensorization also works for higher moments than the variance.

Lemma B.2 (Tensorized L^k Poincaré inequality)**.** *Let $Z = (Z_1, \dots, Z_n)$ have independent blocks. Assume that each block Z_ℓ satisfies a Poincaré inequality with constant C_P. Then, for every integer $k \ge 2$ and every sufficiently smooth scalar function F,*

$$\|F(Z) - \mathbb{E}F(Z)\|_{L^k} \le C\sqrt{C_P}\, k \left\| \left(\sum_{\ell=1}^{n} (|\mathbb{D}_\ell[F(Z)]|_2^*)^2 \right)^{1/2} \right\|_{L^k},$$

where $C > 0$ is universal. In particular, for fixed k, the constant is $O_{k,P}(1)$.

Proof. By Proposition B.1, the product vector Z satisfies the Poincaré inequality

$$\operatorname{Var}(H(Z)) \le C_P \mathbb{E}\left[\sum_{\ell=1}^{n} (|\mathbb{D}_\ell[H(Z)]|_2^*)^2\right]$$

for every sufficiently smooth scalar H. Set

$$Y := F(Z) - \mathbb{E}F(Z), \qquad S := \left(\sum_{\ell=1}^{n} (|\mathbb{D}_\ell[F(Z)]|_2^*)^2 \right)^{1/2}.$$

For $q \geq 2$, write

$$a_q := \|Y\|_{L^q}, \qquad b_q := \|S\|_{L^q}.$$

Applying the tensorized Poincaré inequality to $H = |Y|^{q/2}$, with the usual smooth approximation if necessary, gives

$$\mathrm{Var}\Big(|Y|^{q/2}\Big) \leq C_P \frac{q^2}{4} \mathbb{E}\left[|Y|^{q-2} S^2\right].$$

By Hölder's inequality,

$$\mathbb{E}\left[|Y|^{q-2} S^2\right] \leq a_q^{q-2} b_q^2.$$

Since

$$\mathbb{E}|Y|^q = \mathrm{Var}\Big(|Y|^{q/2}\Big) + \Big(\mathbb{E}|Y|^{q/2}\Big)^2,$$

we obtain

$$a_q^q \leq C_P \frac{q^2}{4} a_q^{q-2} b_q^2 + a_{q/2}^q.$$

If $a_q = 0$, there is nothing to prove. Otherwise, using $a_{q/2} \leq a_q$,

$$a_q^2 \leq C_P \frac{q^2}{4} b_q^2 + a_{q/2}^2.$$

Iterating this inequality until the exponent lies in $(1, 2]$, and using the ordinary tensorized Poincaré inequality at exponent 2, gives

$$a_q \leq C\sqrt{C_P}\, q\, b_q$$

for a universal constant C. Taking $q = k$ proves the claim. □

Lemma B.3 (Linear moments under Poincaré)**.** *Let $m \in \mathbb{N}$ and let $Z \in \mathbb{R}^m$ be a random vector satisfying a Poincaré inequality with constant C_P. Then, for any $k \in \mathbb{N}$,*

$$\sup_{\|u\| \leq 1} \left\| u^\top (Z - \mathbb{E}[Z]) \right\|_{L^k} \leq 6\sqrt{C_P}\, k.$$

This lemma allows us to bound the operator norm of the kth-order centered moment tensor $M_k[Z] \in (\mathbb{R}^m)^{\otimes k}$ of Z, defined, for all $u_1, \ldots, u_k \in \mathbb{R}^m$, by

$$M_k[Z] \cdot (u_1, \ldots, u_k) = \mathbb{E}\left[u_1^\top \tilde{Z} \cdots u_k^\top \tilde{Z}\right],$$

where $\tilde{Z} := Z - \mathbb{E}[Z]$. Indeed, Hölder's inequality and Lemma B.3 give the following corollary.

Corollary B.4. *Let $m \in \mathbb{N}$ and let $Z \in \mathbb{R}^m$ be a random vector satisfying a Poincaré inequality with constant C_P. Then, for any $k \in \mathbb{N}$,*

$$|M_k[Z]|_{\mathrm{op}} := \sup_{\|u_1\|, \ldots, \|u_k\| \leq 1} |M_k[Z] \cdot (u_1, \ldots, u_k)| \leq \left(6\sqrt{C_P}\, k\right)^k.$$

Equivalently,

$$|M_k[Z]|_{\mathrm{op}}^{1/k} \leq 6\sqrt{C_P}\, k.$$

Lemma B.3 relies on the following classical result, taken from [4, Proposition 4.1].

Lemma B.5 (Bobkov–Ledoux exponential integrability bound)**.** *Let $m \in \mathbb{N}$ and let $Z \in \mathbb{R}^m$ be a random vector satisfying a Poincaré inequality with constant C_P. Let $\kappa > 0$. For any bounded κ-Lipschitz mapping $f : \mathbb{R}^m \to \mathbb{R}$,*

$$\forall 0 \leq \lambda < \frac{2}{\sqrt{C_P}\kappa}: \qquad \mathbb{E}\left[\exp\left(\lambda \left|f(Z) - \mathbb{E}[f(Z)]\right|\right)\right] \leq \frac{4/\sqrt{C_P} + 2\lambda\kappa}{2/\sqrt{C_P} - \lambda\kappa}.$$

Proof of Lemma B.3. Consider a deterministic vector $u \in \mathbb{R}^m$ satisfying $|u|_2 \le 1$. If $u = 0$, there is nothing to prove. Set

$$Y := u^\top Z - \mathbb{E}[u^\top Z].$$

To apply Bobkov–Ledoux's exponential integrability bound, we introduce a quantity $M > 0$ and the mapping $T_M(t) := (-M) \vee (t \wedge M)$, which allows us to truncate the linear form. Define

$$Y^{(M)} := T_M(u^\top Z) - \mathbb{E}[T_M(u^\top Z)].$$

Then $Y^{(M)}$ is bounded, centered, and $|u|_2$-Lipschitz as a function of Z. Lemma B.5 applied with $\lambda = 1/(\sqrt{C_P}\,|u|_2)$ yields

$$\mathbb{E}\left[\exp\left(\frac{|Y^{(M)}|}{\sqrt{C_P}\,|u|_2}\right)\right] \le 6. \tag{B.2}$$

By the Poincaré inequality applied to linear functions, $u^\top Z$ has finite variance. Hence $T_M(u^\top Z) \to u^\top Z$ in L^1, so $Y^{(M)}$ converges to Y in probability and along a subsequence almost surely. Fatou's lemma allows us to pass to the limit in (B.2), with Y in place of $Y^{(M)}$. Therefore, using $t^k \le k!e^t$ for $t \ge 0$,

$$\forall k \in \mathbb{N} : \qquad \mathbb{E}|Y|^k \le 6k!\,(\sqrt{C_P}\,|u|_2)^k.$$

Hence, since $k! \le k^k$,

$$\left\|u^\top(Z - \mathbb{E}[Z])\right\|_{L^k} \le 6^{1/k}\sqrt{C_P}\,k\,|u|_2\,.$$

Taking the supremum over $|u|_2 \le 1$ gives the claim. □

Lemma B.6 (Euclidean moments of a random vector)**.** *Let $m \in \mathbb{N}$ and let $Z \in \mathbb{R}^m$ be a random vector. Then, for any $k \ge 2$,*

$$\||Z|_2\|_{L^k} \le \sqrt{m} \sup_{\|u\| \le 1} \left\|u^\top Z\right\|_{L^k}.$$

Proof. Let $(e_a)_{a \le m}$ be an orthonormal basis of $\mathbb{R}^m$. For $k \ge 2$, Minkowski's inequality gives

$$\||Z|_2\|_{L^k} = \left\|\left(\sum_{a=1}^m |e_a^\top Z|^2\right)^{1/2}\right\|_{L^k} \le \left(\sum_{a=1}^m \left\|e_a^\top Z\right\|_{L^k}^2\right)^{1/2}.$$

The stated bound follows by taking the supremum over all unit vectors. □

The next lemma shows that maxima of a family of random variables satisfying similar linear moment-growth bounds grow only logarithmically with the size of the family. Note that this type of uniform control is not available from fixed-order moment assumptions alone.

Lemma B.7 (Maxima under polynomial moment growth)**.** *Let $\alpha > 0$. Assume that $Z_1, \dots, Z_m$ are real-valued random variables and that there exists $C > 0$ such that, for every integer $q \ge 1$,*

$$\max_{\ell \in [m]} \|Z_\ell\|_{L^q} \le Cq^\alpha.$$

Then, for every integer $r \ge 1$,

$$\left\|\max_{\ell \in [m]} |Z_\ell|\right\|_{L^r} \le e\,C\,(\lceil r \vee \log(em)\rceil)^\alpha\,.$$

Proof. Let

$$q := \lceil r \vee \log(em) \rceil .$$

Then $q \geq r$, and

$$\begin{aligned}\left\| \max_{\ell \in [m]} |Z_\ell| \right\|_{L^r} &\leq \left\| \max_{\ell \in [m]} |Z_\ell| \right\|_{L^q} \\ &\leq \left(\sum_{\ell=1}^m \|Z_\ell\|_{L^q}^q \right)^{1/q} \leq C q^\alpha m^{1/q}.\end{aligned}$$

Since $q \geq \log(em)$, we have $m^{1/q} \leq e$, which proves the claim. □

Lemma B.8 (Conditional L^k Poincaré bound)**.** *Let $Z = (Z_1, \dots, Z_n)$ be a vector of independent blocks. Assume that the block Z_j satisfies a Poincaré inequality with constant C_P. Then, for every integer $k \geq 2$ and every sufficiently smooth scalar function F,*

$$\|F(Z) - \mathbb{E}_j[F(Z)]\|_{L^k} \leq C_{k,P} \left\| |\mathbb{D}_j[F]|_2^* \right\|_{L^k} ,$$

where $C_{k,P}$ depends only on k and C_P. More precisely, one can take $C_{k,P} = O(k\sqrt{C_P})$.

Proof. It suffices to prove the corresponding conditional estimate. Fix all blocks except Z_j, and write

$$Y := F(Z) - \mathbb{E}_j[F(Z)], \qquad G := |\mathbb{D}_j[F]|_2^* .$$

For $q \geq 2$, set

$$a_q := \|Y\|_{L_j^q} , \qquad b_q := \|G\|_{L_j^q} .$$

The Poincaré inequality applied conditionally to $|Y|^{q/2}$ gives

$$\mathrm{Var}_j\left(|Y|^{q/2}\right) \leq C_P \frac{q^2}{4} \mathbb{E}_j\left[|Y|^{q-2} G^2\right] .$$

By Hölder's inequality,

$$\mathbb{E}_j\left[|Y|^{q-2} G^2\right] \leq a_q^{q-2} b_q^2.$$

Hence

$$\mathrm{Var}_j\left(|Y|^{q/2}\right) \leq C_P \frac{q^2}{4} a_q^{q-2} b_q^2.$$

Since

$$\mathbb{E}_j |Y|^q = \mathrm{Var}_j\left(|Y|^{q/2}\right) + \left(\mathbb{E}_j |Y|^{q/2}\right)^2 ,$$

we obtain

$$a_q^q \leq C_P \frac{q^2}{4} a_q^{q-2} b_q^2 + a_{q/2}^q.$$

If $a_q = 0$, there is nothing to prove. Otherwise, dividing by a_q^{q-2} and using $a_{q/2} \leq a_q$, we get

$$a_q^2 \leq C_P \frac{q^2}{4} b_q^2 + a_{q/2}^2.$$

Iterating this inequality until the exponent lies in $(1, 2]$, and using the ordinary Poincaré inequality at exponent 2, gives

$$a_q \leq C\sqrt{C_P}\, q\, b_q$$

for a universal constant C. Taking $q = k$, we have shown that, almost surely with respect to the remaining blocks,

$$\|F(Z) - \mathbb{E}_j[F(Z)]\|_{L_j^k} \leq C\sqrt{C_P}\, k \left\| |\mathbb{D}_j[F]|_2^* \right\|_{L_j^k} .$$

Raising this inequality to the power k and integrating over the remaining blocks yields

$$\|F(Z) - \mathbb{E}_j[F(Z)]\|_{L^k} \leq C\sqrt{C_P}\, k \left\| |\mathbb{D}_j[F]|_2^* \right\|_{L^k} .$$

This proves the claim. □

Lemma B.9 (Operator norm of a Poincaré data matrix)**.** *Let $Z = (z_1, \ldots, z_n) \in \mathcal{M}_{p,n}$ have independent columns. Assume that each z_i satisfies a Poincaré inequality with constant $C_P = O(1)$, and that*

$$\sup_{i\in[n]} |\mathbb{E}z_i|_2 = O(1).$$

Then, for every fixed integer $k \geq 1$,

$$\left\| |Z|_{\mathrm{op}} \right\|_{L^k} = O_k\left(\sqrt{p+n}\right).$$

This lemma relies on the following result taken from [1] (one could also use ε-nets to prove it)

Lemma B.10 (Empirical-covariance input)**.** *Let $Y = (y_1, \ldots, y_n) \in \mathbb{R}^{p\times n}$ have independent centered columns. Assume that, for some constant $K = O(1)$,*

$$\sup_{i\in[n]} \sup_{\|u\|_2\leq 1} \sup_{q\geq 1} q^{-1} \left\| u^\top y_i \right\|_{L^q} \leq K, \qquad \sup_{i\in[n]} \mathbb{E}\|y_i\|_2^2 \leq K^2 p.$$

Then there exists a constant $C = C(K)$ such that

$$\mathbb{P}\Big(\|YY^\top\|_{\mathrm{op}} > C(p+n) \Big) \leq \frac{1}{4}.$$

Proof of Lemma B.9. Write

$$m_i := \mathbb{E}z_i, \qquad M := (m_1, \ldots, m_n), \qquad Y := Z - M = (y_1, \ldots, y_n).$$

The deterministic mean part satisfies

$$|M|_{\mathrm{op}} \leq |M|_F = \left(\sum_{i=1}^n |m_i|_2^2 \right)^{1/2} = O(\sqrt{n}) \leq O\left(\sqrt{p+n}\right).$$

It remains to control Y.

For every $u \in \mathbb{R}^p$ with $|u|_2 \leq 1$, the Poincaré inequality applied to the linear map $x \mapsto u^\top x$ gives

$$\mathrm{Var}(u^\top z_i) \leq C_P.$$

Hence

$$\left| \sum_{i=1}^n \mathbb{E}[y_i y_i^\top] \right|_{\mathrm{op}} \leq C_P n.$$

Moreover, Lemma B.3 gives the uniform sub-exponential moment bound

$$\sup_{i\in[n]} \sup_{|u|_2\leq 1} \sup_{q\geq 1} \frac{1}{q} \left\| u^\top y_i \right\|_{L^q} = O(1).$$

Furthermore,

$$\mathbb{E}\,|y_i|_2^2 = \mathrm{Tr}\big(\mathrm{Cov}(z_i)\big) \leq C_P p,$$

and Lemma B.3 gives

$$\sup_{i\in[n]} \sup_{\|u\|_2\leq 1} \sup_{q\geq 1} q^{-1} \left\| u^\top y_i \right\|_{L^q} = O(1).$$

Therefore Lemma B.10 yields, for a constant C depending only on C_P,

$$\mathbb{P}\left(\left| YY^\top \right|_{\mathrm{op}} > C(p+n) \right) \leq \frac{1}{4}.$$

$$\mathbb{P}\left(\left|YY^\top\right|_{\mathrm{op}} > C(p+n)\right) \le \frac{1}{4}.$$

Equivalently, with

$$R := |Y|_{\mathrm{op}}\,,$$

we have

$$\mathbb{P}\big(R \le C\sqrt{p+n}\big) \ge \frac{3}{4}.$$

On the other hand, the map $Y \mapsto |Y|_{\mathrm{op}}$ is 1-Lipschitz with respect to the Frobenius norm. Since the columns of Y are independent and each satisfies a Poincaré inequality with constant C_P, tensorization and the Bobkov–Ledoux moment consequence give, for every fixed integer $k \ge 1$,

$$\|R - \mathbb{E}R\|_{L^k} = O_k(1).$$

Here one may apply Lemma B.5 first to the bounded truncations $R \wedge A$, and then let $A \to \infty$.

We now combine this concentration around the mean with the preceding positive-probability bound. Let

$$a := C\sqrt{p+n}.$$

If $\mathbb{E}R > a + t$, then

$$\frac{3}{4} \le \mathbb{P}(R \le a) \le \mathbb{P}(|R - \mathbb{E}R| \ge t) \le \frac{\|R - \mathbb{E}R\|_{L^k}^k}{t^k}.$$

Since $\|R - \mathbb{E}R\|_{L^k} = O_k(1)$, this implies $t = O_k(1)$. Therefore

$$\mathbb{E}R = O_k\big(\sqrt{p+n}\big)\,.$$

Consequently,

$$\|R\|_{L^k} \le \mathbb{E}R + \|R - \mathbb{E}R\|_{L^k} = O_k\big(\sqrt{p+n}\big)\,.$$

Finally,

$$\Big\||Z|_{\mathrm{op}}\Big\|_{L^k} \le |M|_{\mathrm{op}} + \Big\||Y|_{\mathrm{op}}\Big\|_{L^k} = O_k\big(\sqrt{p+n}\big)\,.$$

Under Assumption 1, $p = O(n)$, and the final bound becomes $O_k(\sqrt{n})$. □

References

[1] Radosław Adamczak, Alexander E. Litvak, Alain Pajor, and Nicole Tomczak-Jaegermann. Sharp bounds on the rate of convergence of the empirical covariance matrix. *Comptes Rendus Mathématique*, 349(3–4):195–200, 2011.

[2] Derek Bean, Peter J. Bickel, Noureddine El Karoui, and Bin Yu. Optimal M-estimation in high-dimensional regression. *Proceedings of the National Academy of Sciences*, 110(36):14563–14568, 2013.

[3] Pierre C. Bellec, Yiwei Shen, and Cun-Hui Zhang. Asymptotic normality of robust M-estimators with convex penalty. *Electronic Journal of Statistics*, 16(2):5591–5622, 2022.

[4] Sergey G. Bobkov and Michel Ledoux. Poincaré's inequalities and talagrand's concentration phenomenon for the exponential distribution. *Probability Theory and Related Fields*, 107(3):383–400, 1997.

[5] Sourav Chatterjee. A new method of normal approximation. *The Annals of Probability*, 36(4):1584–1610, 2008.

[6] David Donoho and Andrea Montanari. High dimensional robust M-estimation: Asymptotic variance via approximate message passing. *Probability Theory and Related Fields*, 166(3–4):935–969, 2016.

[7] Noureddine El Karoui. Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: Rigorous results, 2013.

[8] Noureddine El Karoui, Derek Bean, Peter J. Bickel, Chinghway Lim, and Bin Yu. On robust regression with high-dimensional predictors. *Proceedings of the National Academy of Sciences*, 110(36):14557–14562, 2013.

[9] Nathael Gozlan. Transport inequalities. a survey. *Markov Processes and Related Fields*, 16(4):635–736, 2010.

[10] Xuming He and Qi-Man Shao. On parameters of increasing dimensions. *Journal of Multivariate Analysis*, 73(1):120–135, 2000.

[11] Peter J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, volume 1, pages 221–233, Berkeley, CA, 1967. University of California Press.

[12] Michel Ledoux. The concentration of measure phenomenon. *Mathematical Surveys and Monographs*, 89, 2001.

[13] Lihua Lei, Peter J. Bickel, and Noureddine El Karoui. Asymptotics for high dimensional regression M-estimates: Fixed design results. *Probability Theory and Related Fields*, 172(3–4):983–1079, 2018.

[14] Enno Mammen. Asymptotics with increasing dimension for robust regression with applications to the bootstrap. *The Annals of Statistics*, 17(1):382–400, 1989.

[15] Stephen Portnoy. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. I. consistency. *The Annals of Statistics*, 12(4):1298–1309, 1984.

[16] Stephen Portnoy. Asymptotic behavior of M-estimators of p regression parameters when p^2/n is large. II. normal approximation. *The Annals of Statistics*, 13(4):1403–1417, 1985.

[17] Qi-Man Shao and Zhuo-Song Zhang. Berry–esseen bounds for multivariate nonlinear statistics with applications to M-estimators and stochastic gradient descent algorithms. *Bernoulli*, 28(3):1548–1576, 2022.

[18] Qi-Man Shao and Zhuo-Song Zhang. Berry–Esseen bounds for functionals of independent random variables. *Stochastic Processes and their Applications*, 183:104574, 2025.

[19] Pragya Sur and Emmanuel J. Candès. A modern maximum-likelihood theory for high-dimensional logistic regression. *Proceedings of the National Academy of Sciences*, 116(29):14516–14525, 2019.

[20] A. W. van der Vaart. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 1998.

[21] Qian Zhao, Pragya Sur, and Emmanuel J. Candès. The asymptotic distribution of the MLE in high-dimensional logistic models: Arbitrary covariance. *Bernoulli*, 28(3):1835–1861, 2022.