PARAMETRIC RATES FOR NONPARAMETRIC CONDITIONAL-MOMENT BILEVEL OPTIMIZATION

Anonymous authors
Paper under double-blind review

ABSTRACT

We study functional bilevel optimization with conditional moments, where plug-in hypergradients can inherit first-order errors from nonparametrically learned conditional expectations. We introduce OBiGrad, an orthogonalized hypergradient estimator whose bias is a product of nuisance errors. Under our regression and complexity conditions, projected gradient descent with the direct plug-in hypergradient yields expected stationarity error $O(T^{-1/2} + n^{-1/2} + n^{-\nu/b})$ for the best iterate, whereas OBiGrad achieves $O(T^{-1/2} + n^{-1/2} + n^{-2\nu/b})$, where $b = \nu(2+p) + \gamma$ combines the approximation and complexity exponents. Consequently, when $\nu(2-p) \geq \gamma$, OBiGrad attains the parametric statistical rate $O(T^{-1/2} + n^{-1/2})$, improving the sufficient sample-size bound relative to the direct plug-in bound for a fixed iteration budget. We illustrate these gains on instrumental-variable and fitted-Q-regression problems, examining stationary-point estimation, returned population stationarity, and regularization sensitivity.

1 INTRODUCTION

Bilevel optimization captures a common hierarchical structure in machine learning, in which one learning task is constrained by the solution of another: an outer parameter is chosen to optimize a criterion that depends on the solution of a separate, inner learning problem. Familiar examples include hyperparameter optimization and meta-learning (Pedregosa, 2016; Franceschi et al., 2018). In *functional bilevel optimization*, the inner variable is a prediction function, and differentiation is performed in function space (Petrulionyte et al., 2024). This viewpoint separates the learning problem from the particular kernel or neural-network model used to approximate its solution.

We focus on problems where the inner function is a conditional expectation: the average of a signal given observed covariates. Such functions arise in instrumental-variable regression (Newey & Powell, 2003; Hartford et al., 2017; Singh et al., 2019) and the regression step underlying fitted value or Q-function methods (Ernst et al., 2005; Farahmand et al., 2009). Specifically, we study a quadratic problem: the outer parameter determines a transformation of an observed input, and the inner least-squares regression predicts this signal from conditioning covariates. The population outer objective measures the prediction's mean squared error against the outcome on new observations. Its gradient with respect to the outer parameter, the *hypergradient*, depends on both the inner function and how that function changes with the parameter.

The standard way to compute this hypergradient first estimates the inner solution from data, then differentiates that estimate to obtain the gradient update (El Khoury et al., 2025). The inner function and its derivative are *nuisance functions*: auxiliary functions learned to compute the outer gradient. Limited data, restricted model capacity, and regularization can leave estimation error even when the empirical inner problem is solved accurately. At a fixed outer parameter, averaging the plug-in formula over independent evaluation data is unbiased for its own mean, which can differ from the true population gradient. We call this difference the estimator's *conditional bias*, holding the fitted nuisance functions fixed. Its expansion can contain *first-order bias* terms (Chernozhukov et al., 2018), linear in either function's estimation error, unless additional alignment or cancellation removes them. The optimizer can therefore make small estimated updates while the population objective still admits improvement.

We construct *OBiGrad* by adding a residual correction to the plug-in hypergradient. Besides the inner function and its derivative, this correction uses a regression of the outcome on the same con-

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

ditioning covariates. We call the resulting hypergradient *orthogonal* because its expectation is insensitive, to first order, to errors in these functions at their population values (Chernozhukov et al., 2022; Foster & Syrgkanis, 2023). An exact mixed-bias identity bounds the corrected gradient's population bias by the derivative-estimation error multiplied by the sum of the two regression errors, all measured in root-mean-square norm. For errors of comparable size, generic plug-in sensitivity is linear and orthogonal sensitivity is quadratic in that size, separately from sampling fluctuation. This reduces the accuracy required of the individual regressions to obtain a useful outer update.

When observations are costly to collect, obtaining fresh data at every outer step may be impractical. We split the fixed dataset into two independent samples of size n, using the training sample to learn the regression functions and the other sample to evaluate hypergradients. Both samples are reused at every outer step. Each iterate depends on previous gradient evaluations. We bound the gradient error across all feasible parameters, as in Kernel Bilevel Optimization (KBO) (El Khoury et al., 2025), to cover every adaptive iterate. This turns second-order robustness into a population optimization guarantee without an error term growing linearly with the number of updates.

Population optimization guarantees. Suppose the population outer objective is smooth and bounded below on a compact convex feasible set. Run projected gradient descent with a fixed step size η no larger than the reciprocal of the gradient Lipschitz constant, using either OBiGrad or the direct plug-in hypergradient. The population gradient mapping G_η is the difference between an iterate and its projected step using the exact population gradient, divided by η; its Euclidean norm measures stationarity. Under Assumptions A to D, use the tuning of Theorem 3.2 and set $b = \nu(2+p)+\gamma$. For the respective trajectories (ω_t^{PI}) and $(\omega_t^{\mathrm{OBi}})$, the statistical bounds combined with Proposition 2.1 give

$$\begin{aligned} \text{Direct plug-in:} \quad & \mathbb{E}\left[\min_{0\le t<T} \|G_\eta(\omega_t^{\mathrm{PI}})\|\right] \lesssim T^{-1/2} + n^{-1/2} + n^{-\nu/b}, \\ \text{OBiGrad:} \quad & \mathbb{E}\left[\min_{0\le t<T} \|G_\eta(\omega_t^{\mathrm{OBi}})\|\right] \lesssim T^{-1/2} + n^{-1/2} + n^{-2\nu/b}. \end{aligned}$$

Expectations are over both samples and any fitting randomness; $\lesssim$ suppresses constants independent of T and n for fixed step size and initial objective gap. Each bound concerns the smallest population stationarity value among the first T iterates. Here ν is the approximation exponent, while p and γ control class complexity. The OBiGrad bound uses Theorem 3.2; the direct plug-in bound follows by applying the same nuisance-moment and localization controls to its first-order bias decomposition (Appendices A.7 and B.3). Thus OBiGrad attains $O(T^{-1/2}+n^{-1/2})$ stationarity when $\nu(2-p) \ge \gamma$.

Existing bilevel analyses control the effect of approximate hypergradients on optimization (Ghadimi & Wang, 2018; Ji et al., 2021). Kernel bilevel optimization (KBO) instantiates this program in a reproducing kernel Hilbert space (RKHS), using the representer theorem to obtain finite-dimensional empirical solutions (El Khoury et al., 2025). KBO provides uniform gradient and population-stationarity guarantees for a regularized problem. Here the target is the unregularized quadratic conditional-moment objective, and the question is how to cancel first-order sensitivity to regression error before it enters the optimizer. The correction also has a statistical debiasing interpretation (Chernozhukov et al., 2018; Foster & Syrgkanis, 2023), developed in the appendix.

Contributions. First, we give an explicit orthogonal hypergradient with an exact product-of-errors bias identity. Second, we establish uniform control for fixed-sample population optimization, yielding projected-stationarity guarantees along adaptive trajectories and a qualified comparison with plug-in bounds. Third, matched instrumental-variable and fitted-Q-regression experiments examine stationary-point estimation, population stationarity along optimization trajectories, and robustness to nonzero conditional residuals.

Organization of the paper. Section 2 defines the quadratic problem, projected optimization, and plug-in bias; Section 3 develops the orthogonal hypergradient and its uniform error bound; and Section 4 reports the main experiments. Appendix A contains technical proofs and the hypergradient lower bound. Appendix B gives nuisance-learning conditions, the uniform direct plug-in comparison, and sample-complexity bounds; Appendices C–E give additional statistical derivations and fixed-parameter inference; and Appendix F gives experimental details and robustness analyses.

2 CONDITIONAL MOMENTS AS FUNCTIONAL BILEVEL OPTIMIZATION

2.1 PROBLEM STATEMENT

We consider the problem of finding a transformation of a covariate Z that best matches an outcome Y conditional on an auxiliary variable X. More precisely, let (X, Y, Z) be drawn from a probability distribution P supported on a product space $\mathcal{X} \times \mathcal{Y} \times \mathcal{Z}$ of Euclidean subsets $\mathcal{X} \subset \mathbb{R}^{d_X}$, $\mathcal{Y} \subset \mathbb{R}^q$ and $\mathcal{Z} \subset \mathbb{R}^r$. We model the transformation using a parametric family $g_\omega : \mathcal{Z} \to \mathbb{R}^q$, indexed by $\omega \in \Omega \subset \mathbb{R}^d$, where Ω is compact and convex. Our goal is to choose ω so that the conditional means of $g_\omega(Z)$ and Y given X are as close as possible. Such a conditional moment problem can be naturally formulated as Functional Bilevel Optimization:

$$\min_{\omega\in\Omega} \mathcal{F}(\omega) \coloneqq \frac{1}{2}\mathbb{E}_P\big[\|Y - h^\star_{\omega,P}(X)\|^2\big] \quad \text{s.t.} \quad h^\star_{\omega,P} = \operatorname*{argmin}_{h\in L^2(P_X)} \frac{1}{2}\mathbb{E}_P\big[\|h(X) - g_\omega(Z)\|^2\big], \tag{FBO}$$

where $L^2(P_X)$ is the space of $\mathbb{R}^q$-valued functions on $\mathcal{X}$ that are square-integrable under the law of X, whereas $\|\cdot\|$ denotes the Euclidean norm with associated Euclidean inner product $\langle\cdot,\cdot\rangle$ on $\mathbb{R}^q$. The inner problem (rightmost problem) regresses $g_\omega(Z)$ from X for a given ω over all square integrable functions of X and results in the conditional mean $h^\star_{\omega,P}(X) = \mathbb{E}_P[g_\omega(Z) \mid X]$ of $g_\omega(Z)$ given X. Writing $m^\star_P(X) := \mathbb{E}_P[Y \mid X]$, orthogonality gives $\mathcal{F}(\omega) = \frac{1}{2}\mathbb{E}_P\|Y - m^\star_P(X)\|^2 + \frac{1}{2}\|m^\star_P - h^\star_{\omega,P}\|^2_{L^2(P_X)}$. The first term is independent of ω, so minimizing $\mathcal{F}$ matches the two conditional means. This problem formulation covers many cases of interest. We discuss two of them below.

Instrumental variable regression. A canonical instance is nonparametric instrumental variable (IV) regression, where X is an instrument, Z is the treatment, and Y is the outcome. Under the quadratic losses, the inner minimizer projects the treatment-driven signal $g_\omega(Z)$ onto the instrument X. The bilevel problem thus learns the structural signal $g_\omega(Z)$ and evaluates it through a downstream prediction criterion on the outcome Y, connecting to nonparametric IV and modern deep IV procedures (Newey & Powell, 2003; Hartford et al., 2017; Bennett et al., 2019; Shen et al., 2026).

Fitted Bellman regression in reinforcement learning. Quadratic inner losses also arise in reinforcement learning for fitted value iteration (FVI) and fitted Q-function iteration (FQI) methods. For a state–action pair $X = (S, A)$ and transition tuple $Z = (S, A, R, S')$, the Bellman target is $g_\omega(Z) = R + \gamma V_\omega(S')$ for FVI or $g_\omega(Z) = R + \gamma \max_{a'} Q_\omega(S', a')$ for FQI, with discount factor $\gamma \in (0, 1)$. The inner minimizer averages over transition randomness in S' given (S, A), yielding the projected Bellman backup $h^\star_{\omega,P}(S, A) = \mathbb{E}_P[g_\omega(Z) \mid S, A]$, which the outer objective fits to returns or policy-performance targets. This covers the regression step in FQI and least-squares Bellman methods (Ernst et al., 2005; Antos et al., 2007; Farahmand et al., 2009). For FQI, the hard maximum can fail to be differentiable at action ties; the theory below applies to instances satisfying Assumption A.

Throughout the rest of the paper we impose the following smoothness and boundedness conditions, which keep all of the population objects above well defined and square-integrable.

Assumption A (Smoothness and boundedness)**.** The map $\omega \mapsto g_\omega(z)$ is twice continuously differentiable for every z. Moreover, there exist finite constants $A, B, D, L > 0$ such that $\|Y\| \le A$ almost surely and

$$\sup_{\omega\in\Omega} \|g_\omega(z)\| \le B, \quad \sup_{\omega\in\Omega} \|D_\omega g_\omega(z)\|_F \le D, \quad \sup_{\omega\in\Omega} \|D^2_\omega g_\omega(z)\|_{\mathrm{op}} \le L, \qquad \forall z \in \mathcal{Z}. \tag{1}$$

Here $D_\omega g_\omega(z)$ and $D^2_\omega g_\omega(z)$ are the Jacobian and second derivative, $\|\cdot\|_F$ is the Frobenius norm, and $\|D^2_\omega g_\omega(z)\|_{\mathrm{op}} := \sup_{\|u\|=1}(\sum_{k=1}^d \|\sum_{j=1}^d u_j \partial_{\omega_j}\partial_{\omega_k} g_\omega(z)\|^2)^{1/2}$.

2.2 FIRST-ORDER OPTIMIZATION FOR SOLVING FBO

When solving (FBO) in practice, one only has access to i.i.d. samples from the distribution P. Throughout, expectations are taken with respect to these samples and any auxiliary randomness used by the estimator. When $\mathcal{F}(\omega)$ is differentiable, a natural approach to solving (FBO) is to optimize ω using a first-order method based on an estimator $\widehat{\nabla\mathcal{F}}(\omega)$ of the hypergradient $\nabla\mathcal{F}(\omega)$ of the value function $\mathcal{F}$. We view $\widehat{\nabla\mathcal{F}} : \Omega \to \mathbb{R}^d$ as a random gradient field constructed from these data and

reused throughout the optimization. To keep the iterates in Ω, we can use the projected update:

$$\omega_{t+1} = \mathrm{Proj}_{\Omega}(\omega_t - \eta\widehat{\nabla\mathcal{F}}(\omega_t)), \tag{2}$$

where $\eta > 0$ is a step size and Proj_{Ω} denotes Euclidean projection onto the compact convex set Ω. We hope to approach a stationary point $\omega^\star$ of the value function $\mathcal{F}$ on Ω. Since the gradient need not vanish at a constrained stationary point, stationarity is typically measured by the norm of the *gradient mapping*, defined for $\eta > 0$ by

$$G_\eta(\omega; g) \coloneqq \frac{\omega - \mathrm{Proj}_{\Omega}(\omega - \eta g)}{\eta}, \qquad G_\eta(\omega) \coloneqq G_\eta(\omega; \nabla\mathcal{F}(\omega)).$$

This mapping vanishes exactly at stationary points of $\mathcal{F}$ on Ω and reduces to the ordinary gradient when $\Omega = \mathbb{R}^d$. The following convergence result bounds the stationarity measure by the sum of an iteration-dependent term and a gradient estimation error term.

Proposition 2.1 (Projected gradient descent convergence)**.** *Suppose Assumption A holds, $\widehat{\nabla\mathcal{F}} : \Omega \to \mathbb{R}^d$ is measurable and satisfies $\mathbb{E}\left[\sup_{\omega\in\Omega} \|\widehat{\nabla\mathcal{F}}(\omega) - \nabla\mathcal{F}(\omega)\|\right] < \infty$, and the iterates are given by equation* (2) *with step-size η satisfying $0 < \eta \le \frac{1}{L_{\mathcal{F}}}$, where $L_{\mathcal{F}} \coloneqq D^2 + (A+B)L$ is the smoothness constant of $\mathcal{F}$ on Ω provided in Lemma A.1. Then, for every $T \ge 1$:*

$$\mathbb{E}\left[\min_{0\le t<T} \|G_\eta(\omega_t)\|\right] \lesssim \left(\frac{\mathcal{F}(\omega_0) - F_{\inf}}{\eta T}\right)^{1/2} + \overbrace{\mathbb{E}\left[\sup_{\omega\in\Omega} \|\widehat{\nabla\mathcal{F}}(\omega) - \nabla\mathcal{F}(\omega)\|\right]}^{\mathfrak{s}}.$$

Proposition 2.1 quantifies the effect of gradient estimation error on convergence and is proven in Appendix A.2. The same proof controls the population mapping norm of the implementable iterate minimizing the estimated mapping norm; Appendix F.1 records the corresponding inequality. With exact gradients ($\mathfrak{s} = 0$), the result recovers the standard $O(T^{-1/2})$ stationarity rate for gradient descent on smooth nonconvex objectives. We next show that *plug-in estimators*, a natural choice for gradient estimation, can yield unfavorable bounds on $\mathfrak{s}$, whereas a class of debiased estimators admits substantially sharper bounds.

2.3 Plug-In Hypergradient Estimation and its First-Order Bias

Plug-in estimators of the hypergradient $\nabla\mathcal{F}(\omega)$ can be constructed by leveraging its expression (Petrulionyte et al., 2024; El Khoury et al., 2025). Below, we present two of them.

Direct plug-in estimator. The chain rule allows expressing the hypergradient through $h^\star_{\omega,P}$ and its derivatives $j^\star_{k,\omega,P} \coloneqq \partial_{\omega_k} h^\star_{\omega,P}$. Specifically, defining $\varphi^{\mathrm{PI}}(h, j) \coloneqq \langle h(X) - Y, j(X)\rangle$ for $h, j \in L^2(P_X)$, we obtain, under suitable differentiability and integrability conditions, the *direct form*

$$\partial_{\omega_k}\mathcal{F}(\omega) = \mathbb{E}_P\left[\varphi^{\mathrm{PI}}(h^\star_{\omega,P}, j^\star_{k,\omega,P})\right]. \tag{3}$$

Furthermore, the partial derivative $j^\star_{k,\omega,P}(X)$ can be expressed as a conditional expectation of the model's partial derivatives $\partial_{\omega_k} g_\omega$, which suggests estimating it by regression:

$$j^\star_{k,\omega,P}(X) \coloneqq \mathbb{E}_P[\partial_{\omega_k} g_\omega(Z) \mid X], \qquad k = 1, \dots, d. \tag{4}$$

Substituting regression estimates h and j_k of $h^\star_{\omega,P}$ and $j^\star_{k,\omega,P}$ into equation (3) suggests the plug-in estimator $[\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega)]_k \coloneqq \mathbb{E}_{\hat{P}}[\varphi^{\mathrm{PI}}(h, j_k)]$, where $\hat{P}$ is the empirical distribution of an i.i.d. sample from P, independent of the training data used for computing the estimates h and j_k. The hypergradient estimator $\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega) = ([\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega)]_1, \dots, [\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega)]_d)$ thus requires $d+1$ regression targets.

Adjoint plug-in estimator. A second expression for $\partial_{\omega_k}\mathcal{F}(\omega)$ follows from equations (3) and (4) by application of the tower property for conditional expectations and allows constructing a second estimator requiring only two separate regression functions $h^\star_{\omega,P}$ and the conditional mean $m^\star_P$ defined in Section 2.1. Specifically, defining $\varphi^{\mathrm{API}}_{k,\omega}(h, m) \coloneqq \langle \partial_{\omega_k} g_\omega(Z),\, h(X) - m(X)\rangle$ for arbitrary $m, h \in L^2(P_X)$, the *adjoint form* is given by

$$\partial_{\omega_k}\mathcal{F}(\omega) = \mathbb{E}_P\left[\varphi^{\mathrm{API}}_{k,\omega}(h^\star_{\omega,P}, m^\star_P)\right]. \tag{5}$$

Substituting regression estimates h and m of $h^\star_{\omega,P}$ and $m^\star_P(X)$ into the above equation suggests the plug-in estimator $[\widehat{\nabla\mathcal{F}}^{\mathrm{API}}(\omega)]_k \coloneqq \mathbb{E}_{\hat{P}}[\varphi^{\mathrm{API}}_{k,\omega}(h,m)]$, where $\hat{P}$ is the empirical distribution of an i.i.d. sample from P, independent of the training data used to compute the estimates h and m.

First-order bias of the direct plug-in estimator. The functions $h^\star_{\omega,P}$, $j^\star_{\omega,P}$ and $m^\star_P$ will be called *nuisances* as they are unknown functions that must themselves be estimated before estimates of the hypergradient can be computed. Such two-stage estimation induces a bias that can be quantified. The next proposition provides a decomposition of the bias of $[\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega)]_k$ into first order errors and a higher order term. A similar decomposition for $[\widehat{\nabla\mathcal{F}}^{\mathrm{API}}(\omega)]_k$ can be obtained but is omitted here for simplicity.

Proposition 2.2. *For any $h, j_k \in L^2(P_X)$, the estimator $[\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega)]_k = \mathbb{E}_{\hat{P}}[\varphi^{\mathrm{PI}}(h, j_k)]$ satisfies:*

$$\mathbb{E}_P\big[[\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega)]_k\big] - \partial_{\omega_k}\mathcal{F}(\omega) = \underbrace{\langle h - h^\star_{\omega,P}, j^\star_{k,\omega,P}\rangle_{L^2(P_X)}}_{\text{(I)}} + \underbrace{\langle h^\star_{\omega,P} - m^\star_P, j_k - j^\star_{k,\omega,P}\rangle_{L^2(P_X)}}_{\text{(II)}} + \underbrace{\langle h - h^\star_{\omega,P}, j_k - j^\star_{k,\omega,P}\rangle_{L^2(P_X)}}_{\text{(III)}}. \tag{6}$$

Terms (I) and (II) are each *linear* in one nuisance error, multiplied against the true, generically nonzero, value of the *other* nuisance, so each has an upper bound linear in that error; only the product term (III) is higher order and can vanish faster than each of the errors under nonparametric rates (Bickel et al., 1993; van der Vaart, 1998). Concretely, with $j^\star_{k,\omega,P}$ and $h^\star_{\omega,P} - m^\star_P$ fixed in $L^2(P_X)$, estimators based on n samples satisfying $\|h - h^\star_{\omega,P}\|_{L^2(P_X)} = O(n^{-\alpha})$ and $\|j_k - j^\star_{k,\omega,P}\|_{L^2(P_X)} = O(n^{-\beta})$, for $\alpha, \beta > 0$, yield (I)$= O(n^{-\alpha})$, (II)$= O(n^{-\beta})$, and (III)$= O(n^{-(\alpha+\beta)})$ by Cauchy–Schwarz. Thus, the linear terms determine the leading rate bound, and improving one regression alone does not generally remove the plug-in bias. Appendices B.2 and B.3 give pointwise and uniform plug-in bounds, respectively. Under Assumptions A to D and the tuning of Theorem 3.2, with $b = \nu(2+p)+\gamma$, the expected uniform plug-in gradient error satisfies $\mathfrak{s} = O(n^{-1/2} + n^{-\nu/b})$ (Appendix B.3). For fixed step size, these bounds give sufficient choices $T = O(\epsilon^{-2})$ and $n = O(\epsilon^{-b/\nu})$ for expected best-iterate stationarity error at most ϵ. The familiar nuisance exponent $1/(2+p)$ is the special case $\gamma = 0$; Corollary B.1 also records the corresponding conditional bound under the milder appendix assumptions. For standard classes, such as Sobolev balls, p is typically of the form $p = d_X/s$, where d_X is the dimension of X and s measures the smoothness of the nuisance functions. Thus, high dimensionality or low smoothness can make the gradient method particularly sample-intensive (Appendix B.4). The next section constructs estimators with second-order bias and derives the conditions under which this improves the optimization bound in Proposition 2.1.

3 Debiased Hypergradient Estimation

Section 2 showed that the plug-in hypergradient estimator suffers a first-order bias from the nonparametric estimation of its nuisance components: its bias contains terms that are *linear* in a single nuisance's estimation error, and such terms can dominate the higher-order terms under nonparametric rates. We now construct a corrected score that removes these first-order terms at the population level, and show that the resulting estimator's hypergradient error admits a finite-sample bound that holds uniformly over $\omega \in \Omega$. The key device is a *mixed bias identity*: we show that the bias of the corrected score, evaluated at arbitrary candidate nuisances, collapses to a single term that is a *product* of two nuisance errors, with no term linear in either error alone. This reduces the nuisance accuracy needed to control the hypergradient bias.

3.1 OBiGrad: One-fold doubly robust hypergradient estimator

The plug-in bias in Proposition 2.2 arises because substituting candidates h and j_k directly into $\varphi^{\mathrm{PI}}(h, j_k)$ leaves each candidate's error multiplied against the *true* value of the other nuisance. We remove this first order bias by constructing an alternative estimator built from all nuisances and chosen so that linear terms in the bias decomposition cancel exactly. To this end, we consider the following *debiased score* defined for candidate nuisances $h, j_k, m \in L^2(P_X)$

$$\varphi_{k,\omega}(h, j_k, m) \coloneqq \big\langle g_\omega(Z) - Y,\, j_k(X)\big\rangle + \big\langle \partial_{\omega_k} g_\omega(Z) - j_k(X),\, h(X) - m(X)\big\rangle. \tag{7}$$

When evaluating $\varphi_{k,\omega}$ at the triplet $(h^\star_{\omega,P}, j^\star_{k,\omega,P}, m^\star_P)$ and taking expectation under P, the first term in equation (7) recovers $\partial_{\omega_k}\mathcal{F}(\omega)$, while the second vanishes:

$$\mathbb{E}_P\left[\varphi_{k,\omega}(h^\star_{\omega,P}, j^\star_{k,\omega,P}, m^\star_P)\right] = \partial_{\omega_k}\mathcal{F}(\omega) + \underbrace{\mathbb{E}_P\left[\left\langle \partial_{\omega_k} g_\omega(Z) - j^\star_{k,\omega,P}(X),\, h^\star_{\omega,P}(X) - m^\star_P(X)\right\rangle\right]}_{=0}. \quad (8)$$

The second term is corrective as it has zero mean when using the exact nuisance, but will act as a first-order debiasing term when using nuisance estimators. We use this score to construct OBiGrad, a hypergradient estimator whose *double-robustness* property follows from the mixed-bias identity below. Specifically, given nuisance estimators h, $j_1, \dots, j_d$ and m of $h^\star_{\omega,P}$, $j^\star_{1,\omega,P}, \dots j^\star_{d,\omega,P}$ and $m^\star_P$ at some parameter value ω, we define OBiGrad as follows:

$$[\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)]_k = \mathbb{E}_{\hat{P}}\left[\varphi_{k,\omega}(h, j_k, m)\right], \quad (9)$$

where, similarly as for plug-in estimators, $\hat{P}$ is an empirical distribution obtained from i.i.d. samples from P and independent of the training data used to fit the nuisances. Equation (7) augments the plug-in moment by a regression-residual correction chosen so that, at the true nuisance functions, the derivative of its expectation with respect to errors in either j_k or the difference $h - m$ (which estimates $\mathbb{E}_P[g_\omega(Z) - Y \mid X]$) is zero. Thus small nuisance-estimation errors affect the expected hypergradient only at second order. This zero first-order sensitivity is the orthogonality property. The next theorem parallels Proposition 2.2 and shows that $[\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)]_k$ admits only second order bias, unlike its plug-in counterparts.

Theorem 3.1 (Mixed bias)**.** *Assume A, for $\omega \in \Omega$, $1 \le k \le d$ and $h, j_k, m \in L^2(P_X)$, the hypergradient estimator OBiGrad defined in equation* (9) *satisfies:*

$$\mathbb{E}_P\left[[\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)]_k\right] - \partial_{\omega_k}\mathcal{F}(\omega) = \left\langle j_k - j^\star_{k,\omega,P},\ (m - m^\star_P) - (h - h^\star_{\omega,P})\right\rangle_{L^2(P_X)}. \quad (10)$$

Equation (10) holds exactly, for every candidate $h, j_k, m \in L^2(P_X)$, and vanishes whenever *either* $j_k = j^\star_{k,\omega,P}$ or $h - m = h^\star_{\omega,P} - m^\star_P$, regardless of the error in the other nuisance block. This is the double-robustness property. If $\|j_{k,\omega} - j^\star_{k,\omega,P}\|_{L^2(P_X)} = O(n^{-\beta})$ and $\|h_\omega - h^\star_{\omega,P}\|_{L^2(P_X)} + \|m - m^\star_P\|_{L^2(P_X)} = O(n^{-\alpha})$, the decomposition in equation (10) implies a bound of order $O(n^{-\alpha-\beta})$, which is $o(n^{-1/2})$ under the product-rate condition $\alpha + \beta > 1/2$, a condition that can be satisfied even if both nuisances are estimated at nonparametric rates. This is in stark contrast with requirements when using a plug-in estimator, as discussed in section 2.3.

3.2 Fast uniform rates for OBiGrad

We establish a control on the uniform gradient estimation error $\mathfrak{s}$ when using OBiGrad as a gradient estimator under a general entropy–approximation scale condition on the hypothesis classes when the nuisances are fitted by least-squares up to some vanishing sub-optimality error. To allow a unified treatment of the nuisances, we introduce the index $\ell \in \{h, j, m\}$ which indicates which regression problem is being considered. Thus, we define $f^\star_{\ell,\omega}$ to be the nuisance function of index ℓ so that $(f^\star_{h,\omega}, f^\star_{j,\omega}, f^\star_{m,\omega}) = (h^\star_\omega, j^\star_\omega, m^\star)$, while T^ℓ_ω refers to its corresponding nuisance's target, i.e. $(T^h_\omega(O), T^j_\omega(O), T^m_0(O)) = (g_\omega(Z), \dot{g}_\omega(Z), Y)$ where $O = (X, Y, Z) \sim P$. Here $\dot{g}_\omega = (\partial_{\omega_k} g_\omega)_{k=1}^d$, $\omega_h = \omega_j = \omega$, and $\omega_m = 0$; quantities indexed by m do not depend on ω. Hence, each nuisance function $f^\star_{\ell,\omega}$ minimizes the regression objective

$$R^\ell_\omega(f) = \mathbb{E}_P\left[\ell^\ell_\omega(O; f)\right], \qquad \ell^\ell_\omega(O; f) = \frac{1}{2}\|f(X) - T^\ell_{\omega_\ell}(O)\|^2. \quad (11)$$

Each nuisance $f^\star_{\ell,\omega}$ is fitted in a class $\mathcal{F}_\ell(t)$ of functions defined on $\mathcal{X}$ and indexed by *approximation scale* $t \in (0, 1]$ that balances approximation error and statistical complexity and may depend on sample-size n[1]. Specifically, given a dataset S_1 of n, i.i.d. observations, denote by $P_{n,1}$ the empirical measure on S_1 and define the empirical risk of nuisance ℓ as $\widehat{R}^\ell_\omega(f) = \mathbb{E}_{P_{n,1}}\left[\ell^\ell_\omega(\cdot; f)\right]$. The nuisance estimate $\widehat{f}_{\ell,\omega}$ is then obtained by approximately minimizing the empirical risk $\widehat{R}^\ell_\omega$ in

[1] $\mathcal{F}_\ell(t)$ could be a ball of radius $\frac{R}{t}$ in some suitable function space. The approximation scale t will typically be indexed by training sample-size and decrease to 0 as the sample-size increases.

the class $\mathcal{F}_\ell(t_n)$ for a suitable choice of approximation scale $t_n \in (0,1]$ to be defined later. The estimates $\widehat{f}_{\ell,\omega} \in \mathcal{F}_\ell(t_n)$ are measurable functions of S_1 and, optionally, of an additional random source U independent of all data. Theorem 3.1 suggests controlling the estimation error of each nuisance function in order to bound the overall gradient estimation error. To this end, we impose complexity and approximation assumptions on the corresponding function classes, which allow us to derive explicit error bounds in terms of the sample size. The assumption is stated in terms of the uniform covering number $N(u, \mathcal{C}, \|\cdot\|_\infty)$ of a function class $\mathcal{C}$ (van der Vaart & Wellner, 1996), defined as the smallest number of functions in $\mathcal{C}$ such that every $f \in \mathcal{C}$ lies within distance u of at least one of them under the supremum norm $\|\cdot\|_\infty$.

Assumption B (Nuisance-class complexity)**.** For every $\ell \in \{h, j, m\}$ and $t \in (0,1]$, the class $\mathcal{F}_\ell(t)$ is nonempty, deterministic, pointwise measurable, and uniformly bounded:

$$\sup_{f \in \mathcal{F}_\ell(t)} \sup_x \|f(x)\| \le b_\ell, \qquad \text{where} \quad (b_h, b_j, b_m) = (B, D, A). \tag{12}$$

Moreover, there exist fixed $p \in (0,2)$, $\gamma \ge 0$, and finite constants $K_\ell \ge 0$ such that for all ℓ

$$\log N\big(\epsilon b_\ell, \mathcal{F}_\ell(t), \|.\|_\infty\big) \le K_\ell t^{-\gamma} \epsilon^{-p}, \qquad \forall t, \epsilon \in (0,1]. \tag{13}$$

Assumption C (Approximation)**.** There exist $\nu > 0$ and finite $L_\ell \ge 0$ such that, for all ℓ,

$$\sup_{\omega \in \Omega} \inf_{f \in \mathcal{F}_\ell(t)} \|f - f^\star_{\ell,\omega}\|^2_{L^2(P_X)} \le L_\ell^2 t^{2\nu}, \qquad \forall t \in (0,1]. \tag{14}$$

Assumption B governs the empirical-process terms and Assumption C the approximation error. The polynomial entropy condition in Equation (13) allows nonparametric nuisance classes and also covers simpler classes. Equation (14) is an approximation condition and does not require the true nuisances to belong to the search classes for any value of the approximation scale t. The parameter ν determines the rate at which the approximation error decreases as $t \downarrow 0$. The complexity of the classes is characterized by the exponents γ and p: γ determines how the entropy grows as the approximation scale t decreases, while p governs the dependence of the entropy on the covering radius. The next assumption controls the fitting errors of the nuisance estimates in terms of a prescribed threshold.

Assumption D (Inexact least-squares fitting)**.** The nuisance estimates satisfy, almost surely and simultaneously for every $\omega \in \Omega$,

$$\widehat{R}^\ell_\omega(\widehat{f}_{\ell,\omega}) \le \inf_{f \in \mathcal{F}_\ell(t_n)} \widehat{R}^\ell_\omega(f) + \varepsilon_{\ell,n}, \qquad \ell \in \{h, j, m\}, \tag{15}$$

where $\varepsilon_{\ell,n} \ge 0$ is a prescribed fitting error on the nuisance ℓ at sample-size n.

The following result provides a bound on the expected uniform gradient estimation error $\mathfrak{s}$ arising in Proposition 2.1, in terms of the sample size and the parameters ν, γ, and p.

Theorem 3.2 (Second-order uniform hypergradient error)**.** *Write* $\varphi_\omega(h,j,m) := (\varphi_{k,\omega}(h, j_k, m))_{k=1}^d$ *for* $j = (j_k)_{k=1}^d$. *Assume A and B, and consider the gradient estimator defined in equation* (9) *using the nuisance estimates* $(\widehat{h}_\omega, \widehat{j}_\omega, \widehat{m}) = (\widehat{f}_{h,\omega}, \widehat{f}_{j,\omega}, \widehat{f}_{m,\omega})$. *Define*

$$R_{h,n} := \sup_{\omega \in \Omega} \|\widehat{h}_\omega - h^\star_{\omega,P}\|_{L^2(P_X)}, \; R_{j,n} := \sup_{\omega \in \Omega} \|\widehat{j}_\omega - j^\star_{\omega,P}\|_{L^2(P_X)^d}, \; R_{m,n} := \|\widehat{m} - m^\star_P\|_{L^2(P_X)},$$

and let $\rho_{\ell,n} := (\mathbb{E}[R^2_{\ell,n}])^{1/2}$ *for* $\ell \in \{h,j,m\}$ *and* $\delta_n := (nt_n^\gamma)^{-1/(2+p)}$. *Then there exist finite constants* C_0, *depending only on* A, B, D, L, d, q *and* $\operatorname{diam}(\Omega)$, *and* C, *depending in addition on* p, γ *and* $(K_\ell)_{\ell \in \{h,j,m\}}$, *such that, for every* $n \ge 1$ *and every deterministic* $t_n \in (0,1]$,

$$\mathbb{E}\Big[\sup_{\omega \in \Omega} \big\|\widehat{\nabla \mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla \mathcal{F}(\omega)\big\|\Big] \le C_0 n^{-1/2} + C\big\{\delta_n \textstyle\sum_\ell \rho_{\ell,n} + \delta_n^2 + \rho_{j,n}(\rho_{h,n} + \rho_{m,n})\big\}. \tag{16}$$

If, in addition, C and D hold, then $\rho^2_{\ell,n} \le C_\ell\{t_n^{2\nu} + (nt_n^\gamma)^{-2/(2+p)} + \varepsilon_{\ell,n}\}$ *for each* $\ell \in \{h,j,m\}$, *with finite* C_ℓ *independent of* n, t_n *and* $\varepsilon_{\ell,n}$. *In particular, let* $b := \nu(2+p) + \gamma$ *and, for any training sample-size* $n \ge 1$, *set the approximation scale to* $t_n = n^{-1/b}$ *with sub-optimality error such that* $\varepsilon_{\ell,n} \le c_\ell n^{-2\nu/b}$ *for fixed* $c_\ell < \infty$. *Then* $\rho_{\ell,n} = O(n^{-\nu/b})$ *and* $\delta_n = n^{-\nu/b}$, *and there exists a finite constant* C_1, *depending only on* $A, B, D, L, p, \nu, \gamma, d, q$, $\operatorname{diam}(\Omega)$, *and* $(c_\ell, K_\ell, L_\ell)_{\ell \in \{h,j,m\}}$, *such that, for every* $n \ge 1$,

$$\mathfrak{s} \le C_0 n^{-1/2} + C_1 n^{-2\nu/b}. \tag{17}$$

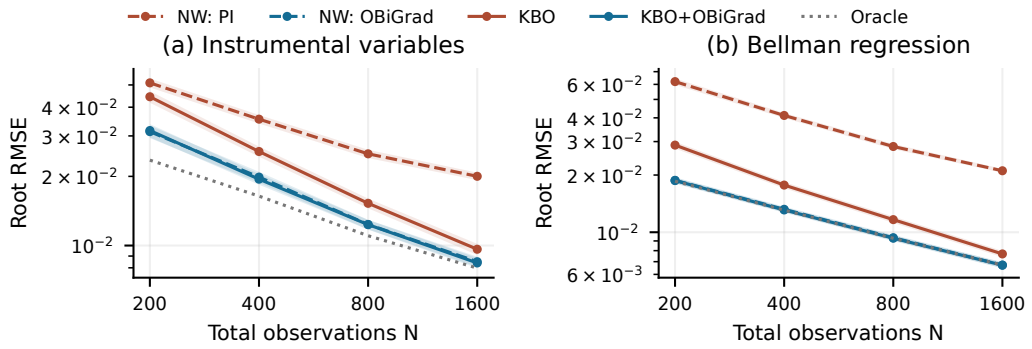


Figure 1: Root RMSE is computed over 800 replications per N in the zero-residual designs. Dashed and solid curves correspond to NW and KBO, respectively; bands show pointwise 95% Monte Carlo intervals.

Theorem 3.2, which is proven in Appendix A.7, shows that the expected uniform gradient estimation error consists of three contributions: an oracle sampling term $O(n^{-1/2})$; a nuisance-dependent empirical-process term $\delta_n \sum_\ell \rho_{\ell,n} + \delta_n^2$, controlled by localization under Assumption B; and the population bias $\rho_{j,n}(\rho_{h,n} + \rho_{m,n})$, of product order by orthogonality (Theorem 3.1). Under Assumptions C and D and the above tuning, the last two are $O(n^{-2\nu/b})$, and are dominated by the parametric term whenever $\nu(2-p) \geq \gamma$. In particular, if the target nuisance functions belong to $\mathcal{F}_\ell(t_0)$ for some $t_0 > 0$, then the classes can be redefined as $\mathcal{F}_\ell(t) = \mathcal{F}_\ell(t_0)$ for all $t \leq t_0$. The entropy bound is then uniform in t, so that $\gamma = 0$ and the condition $\nu(2-p) \geq \gamma$ is automatically satisfied. Consequently, the non-parametric nuisance estimation does not affect the overall $O(n^{-1/2})$ gradient rate. In this fixed-class regime, the corresponding plug-in upper bound has a first-order nuisance contribution $O(n^{-1/(2+p)})$. Moreover, Appendix A.8 establishes a matching worst-case $n^{-1/2}$ lower bound for hypergradient estimation over the distribution class defined there. This matches the OBiGrad rate whenever $\nu(2-p) \geq \gamma$. Appendix B gives the uniform direct plug-in and sample-complexity comparisons, while Appendix E gives fixed-parameter inference for the cross-fitted estimator.

4 Numerical Experiments

We evaluate OBiGrad on two synthetic problems with known unregularized population objectives: a triangular instrumental-variable (IV) model with a quadratic first stage, and a two-feature model of a single projected Bellman backup. These designs instantiate conditional-moment problems arising in IV estimation (Newey & Powell, 2003) and fitted Q methods (Ernst et al., 2005). We use two-fold cross-fitting (Chernozhukov et al., 2018): nuisances are fitted on one half of the data and the score in (7) is evaluated on the other; exchanging the folds and averaging the held-out gradient fields gives the estimator used below.

To isolate the effect of orthogonalization, we compare plug-in and corrected gradients for Gaussian Nadaraya–Watson regression (NW; Nadaraya, 1964; Watson, 1964) and the kernel-ridge specialization of KBO (El Khoury et al., 2025). Each pair shares observations, folds, tuning, fitted inner functions, and sensitivities; OBiGrad additionally estimates $m^\star(X) = \mathbb{E}[Y \mid X]$ with the same regression operator. Tuning uses observed inner targets within training folds, while population quantities serve only for evaluation and oracle references. After exploratory pilots, we fixed the protocol and ran two independent suites of 400 replications per setting and total sample size N. Appendix F provides the full specification and supplementary analyses.

We first solve the estimated gradient equations directly to isolate statistical error from iterative optimization error. Figure 1 shows that OBiGrad reduces root RMSE relative to the population optimizer for both learners throughout the sample-size grid. At $N = 800$, IV RMSE falls from 0.0251 to 0.0124 for NW and from 0.0153 to 0.0123 for KBO; in fitted Q-regression, it falls from 0.0283 to 0.0094 and from 0.0117 to 0.0093, respectively. The corrected estimates approach the finite-sample oracle benchmark computed with true nuisances.

To assess whether these gains translate into improved population stationarity, we run projected gradient descent from zero with step size 0.1 on $[-2, 2]^d$, holding the estimated gradient field fixed. At

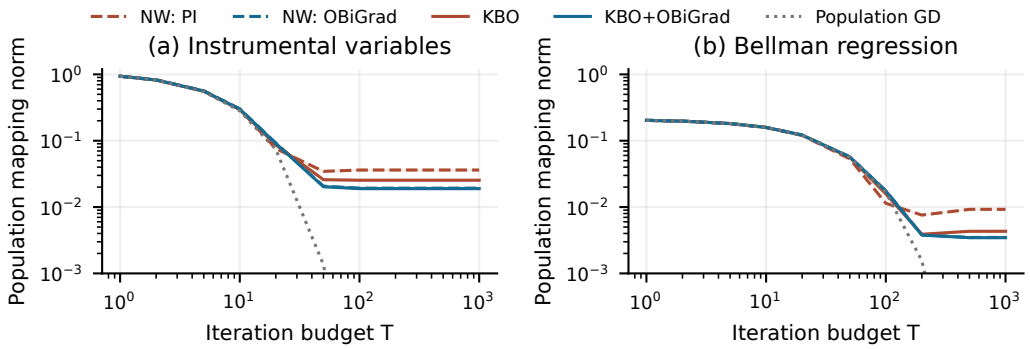


Figure 2: Population gradient-mapping norms at the returned iterates are averaged over 800 replications with $N = 400$. Bands show pointwise 95% Monte Carlo intervals. Selection uses estimated updates; population norms are evaluated afterward. Population GD eventually falls below the displayed range.

budget T, we return the iterate with the smallest estimated projected-update norm among the first T iterates, without using population information for selection. Figure 2 evaluates the population gradient-mapping norm at this iterate. Orthogonalization lowers the eventual statistical error for both learners, although improvements need not occur at every intermediate iteration. Descent with the exact population gradient provides a reference for the remaining optimization error.

The preceding designs satisfy $h^\star_{\omega^\star} = m^\star$, so a shared linear smoother can cancel regularization bias at the corrected population root. We therefore introduce a conditional residual of amplitude 0.25 orthogonal to the true sensitivity, preserving the population optimizer while breaking this equality. Table 1 shows that correction continues to improve root estimation and returned stationarity, although the KBO gains are smaller in fitted Q-regression. The IV extension now represents an approximate conditional-moment problem. Across both residual settings, all sample sizes, and both learners, the 64 prespecified paired comparisons of root squared error and returned stationarity favor correction under approximate simultaneous 95% Monte Carlo intervals. The penalty sweeps and further checks in Appendix F.1 qualify this finding: orthogonalization retains product-order bias and need not improve performance for every penalty or data-generating process.

Table 1: With a nonzero conditional residual and $N = 800$, we report root RMSE and the mean returned population mapping norm after 1000 updates. "OBi" denotes correction of the indicated learner.

Design	Learner	PI root	OBi root	PI mapping	OBi mapping
IV	NW	0.0354	0.0168	0.0375	0.0168
IV	KBO	0.0205	0.0169	0.0202	0.0168
Bellman	NW	0.0416	0.0265	0.0095	0.0067
Bellman	KBO	0.0284	0.0264	0.0070	0.0066

5 Conclusion

We studied population bilevel optimization with learned conditional moments and constructed OBiGrad, an orthogonal hypergradient estimator whose bias is a product of nuisance errors. Uniform gradient control translates this correction into sharper population-stationarity bounds. In the studied IV and fitted-Q-regression regimes, orthogonalizing a fixed nuisance learner improves stationary-point estimation and the population stationarity of returned optimization iterates, including with matched KBO fits. These gains persist beyond exact conditional fit, while regularization can still induce product-order bias. Limitations remain: our guarantees concern gradient estimation and approximate stationarity, not a full end-to-end analysis of arbitrary bilevel algorithms, and they rely on nuisance learners satisfying the complexity, approximation, and fitting conditions of Assumptions B to D. Verifying these conditions for highly adaptive neural learners and extending the method to infinite-horizon reinforcement learning, meta-learning, and other structured bilevel problems remain natural next steps.

AI USE STATEMENT

In this work, we used generative AI tools to polish exposition, to help write and refactor the code generating the synthetic designs and simulations, and to check mathematical derivations. We have reviewed all AI-assisted work: generated code was executed and checked against the analytical specification of each data-generating process, and every definition, statement, and proof was checked and written by the authors. We take responsibility for the final content of this work, including all text, claims, and artifacts produced with the aid of generative AI.

REFERENCES

András Antos, Rémi Munos, and Csaba Szepesvári. Fitted Q-iteration in continuous action-space MDPs. In *Advances in Neural Information Processing Systems (NIPS)*, 2007.

Peter L. Bartlett, Olivier Bousquet, and Shahar Mendelson. Local Rademacher Complexities. *The Annals of Statistics*, 33(4):1497–1537, 2005. doi: 10.1214/009053605000000282.

Andrew Bennett, Nathan Kallus, and Tobias Schnabel. Deep Generalized Method of Moments for Instrumental Variable Analysis. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Peter J. Bickel, Chris A. J. Klaassen, Ya'acov Ritov, and Jon A. Wellner. *Efficient and Adaptive Estimation for Semiparametric Models*. Johns Hopkins University Press, 1993.

Andrea Caponnetto and Ernesto De Vito. Optimal Rates for the Regularized Least-Squares Algorithm. *Foundations of Computational Mathematics*, 7(3):331–368, 2007.

Victor Chernozhukov, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins. Double/Debiased Machine Learning for Treatment and Structural Parameters. *The Econometrics Journal*, 21(1):C1–C68, 2018.

Victor Chernozhukov, Juan Carlos Escanciano, Hidehiko Ichimura, Whitney K Newey, and James M Robins. Locally Robust Semiparametric Estimation. *Econometrica*, 90(4):1501–1535, 2022.

Fares El Khoury, Edouard Pauwels, Samuel Vaiter, and Michael Arbel. Learning Theory for Kernel Bilevel Optimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2025.

Damien Ernst, Pierre Geurts, and Louis Wehenkel. Tree-based batch mode reinforcement learning. *Journal of Machine Learning Research (JMLR)*, 6:503–556, 2005.

Amir-massoud Farahmand, Mohammad Ghavamzadeh, Csaba Szepesvári, and Shie Mannor. Regularized fitted Q-iteration for planning in continuous-space Markovian decision problems. In *American Control Conference*, 2009.

Luisa Turrin Fernholz. *Von Mises Calculus for Statistical Functionals*. Springer, New York, 1983.

Dylan J Foster and Vasilis Syrgkanis. Orthogonal Statistical Learning. *The Annals of Statistics*, 51 (3):879–908, 2023.

Luca Franceschi, Paolo Frasconi, Saverio Salzo, Riccardo Grazzi, and Massimiliano Pontil. Bilevel Programming for Hyperparameter Optimization and Meta-Learning. In *International Conference on Machine Learning (ICML)*, 2018.

Saeed Ghadimi and Mengdi Wang. Approximation Methods for Bilevel Programming. *arXiv preprint arXiv:1802.02246*, 2018.

Frank R Hampel. The influence curve and its role in robust estimation. *Journal of the American Statistical Association*, 69(346):383–393, 1974.

Jason Hartford, Greg Lewis, Kevin Leyton-Brown, and Matt Taddy. Deep IV: A Flexible Approach for Counterfactual Prediction. In *International Conference on Machine Learning (ICML)*, 2017.

Kaiyi Ji, Junjie Yang, and Yingbin Liang. Bilevel Optimization: Convergence Analysis and Enhanced Design. In *International Conference on Machine Learning (ICML)*, 2021.

Michael R. Kosorok. *Introduction to Empirical Processes and Semiparametric Inference*. Springer, 2008.

E. A. Nadaraya. On Estimating Regression. *Theory of Probability and its Applications*, 9(1):141–142, 1964. doi: 10.1137/1109020.

Whitney K. Newey. The Asymptotic Variance of Semiparametric Estimators. *Econometrica*, 62(6): 1349–1382, 1994.

Whitney K. Newey and James L. Powell. Instrumental Variable Estimation of Nonparametric Models. *Econometrica*, 71(5):1565–1578, 2003.

Fabian Pedregosa. Hyperparameter optimization with approximate gradient. In *International Conference on Machine Learning (ICML)*, 2016.

Ieva Petrulionyte, Julien Mairal, and Michael Arbel. Functional Bilevel Optimization for Machine Learning. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2024.

Bernhard Schölkopf, Ralf Herbrich, and Alex J Smola. A Generalized Representer Theorem. In *International conference on computational learning theory*, pp. 416–426. Springer, 2001.

Zikai Shen, Nathan Kallus, Dimitri Meunier, Houssam Zenati, Arthur Gretton, and Aurélien Bibaut. Nonparametric Instrumental Variable Analysis Without Structural Equations: Debiased Inference on Functionals of Inverse Problems with No Solutions. *arXiv preprint arXiv:2604.24660*, 2026.

Rahul Singh, Maneesh Sahani, and Arthur Gretton. Kernel Instrumental Variable Regression. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Ingo Steinwart and Andreas Christmann. *Support Vector Machines*. Springer, 2008.

Anastasios A. Tsiatis. *Semiparametric Theory and Missing Data*. Springer, 2006.

Alexandre B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer, New York, 2009.

Sara A. van de Geer. *Empirical Processes in M-Estimation*. Cambridge University Press, 2000.

Lars van der Laan. A Researcher's Guide to Empirical Risk Minimization. *arXiv preprint arXiv:2602.21501*, 2026.

Lars van der Laan, Aurélien Bibaut, Nathan Kallus, and Alex Luedtke. Automatic Debiased Machine Learning for Smooth Functionals of Nonparametric M-Estimands. *arXiv preprint arXiv:2501.11868*, 2025.

Aad van der Vaart and Jon A Wellner. A local maximal inequality under uniform entropy. *Electronic Journal of Statistics*, 5(2011):192, 2011.

Aad W. van der Vaart. *Asymptotic Statistics*. Cambridge University Press, 1998.

Aad W. van der Vaart and Jon A. Wellner. *Weak Convergence and Empirical Processes: With Applications to Statistics*. Springer, 1996.

Richard von Mises. On the asymptotic distribution of differentiable statistical functions. *Annals of Mathematical Statistics*, 18(3):309–348, 1947.

Martin J. Wainwright. *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Number 48 in Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, Cambridge, 2019. ISBN 9781108498029.

Geoffrey S. Watson. Smooth Regression Analysis. *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4):359–372, 1964. URL `https://www.jstor.org/stable/25049340`.

APPENDIX

CONTENTS

A Uniform Control and Optimization **12**

A.1 Smoothness of $\mathcal{F}$ 12

A.2 Projected gradient descent with an inexact hypergradient 13

A.3 Hypergradient bias identities 14

A.4 Auxiliary empirical-process lemmas 16

A.5 Conditional maximal inequality 18

A.6 Polynomial-entropy instantiation 22

A.7 Unconditional rates under source and entropy conditions 23

A.8 Minimax optimality of OBiGrad 29

B Nuisance Learning and Plug-in Hypergradients **32**

B.1 Nuisance learning 32

B.2 Plug-in hypergradient estimation 33

B.3 Uniform plug-in maximal inequality 35

B.4 Sample complexity: OBiGrad versus plug-in 37

C Efficient Influence Function **38**

D Functional von Mises Expansion **41**

E Asymptotic Normality **44**

F Additional Experiment Details **47**

F.1 Population optimization with matched nuisance learning 48

F.2 Instrumental-variable experiments 55

F.3 Fitted Q-regression experiments 61

A UNIFORM CONTROL AND OPTIMIZATION

A.1 SMOOTHNESS OF $\mathcal{F}$

Lemma A.1 (Smoothness)**.** *Under Assumption A, $\mathcal{F}$ is $L_{\mathcal{F}}$-smooth on Ω with $L_{\mathcal{F}} \coloneqq D^2+(A+B)L$, and $0 \le \mathcal{F} \le \frac{1}{2}(A+B)^2$, so $F_{\inf} > -\infty$.*

Proof. Collect the derivative nuisances into the matrix-valued map

$$j^\star_{\omega,P}(x) \coloneqq \big(j^\star_{1,\omega,P}(x), \dots, j^\star_{d,\omega,P}(x)\big) \in \mathbb{R}^{q\times d},$$

whose k-th column is $j^\star_{k,\omega,P}(x)$ and whose block norm $\|\cdot\|_{2,d}$ is the Frobenius norm of the stack. Since the k-th entry of $j^\star_{\omega,P}(x)^\top v$ is $\langle j^\star_{k,\omega,P}(x), v\rangle$, the direct form of the hypergradient stacks to $\nabla\mathcal{F}(\omega) = \mathbb{E}_P[j^\star_{\omega,P}(X)^\top(h^\star_{\omega,P}(X) - Y)]$. Let $\omega_1, \omega_2 \in \Omega$ and abbreviate $j_i \coloneqq j^\star_{\omega_i,P}(X)$,

$h_i \coloneqq h^\star_{\omega_i,P}(X)$ and $R_i \coloneqq h_i - Y$ for $i = 1, 2$. We have

$$\begin{aligned}
\big\|\nabla\mathcal{F}(\omega_1) - \nabla\mathcal{F}(\omega_2)\big\| &= \Big\|\mathbb{E}_P\big[\, j_1^\top R_1 - j_2^\top R_2 \,\big]\Big\| \\
&= \Big\|\mathbb{E}_P\big[\, j_1^\top (R_1 - R_2) + (j_1 - j_2)^\top R_2 \,\big]\Big\| \\
&= \Big\|\mathbb{E}_P\big[\, j_1^\top (h_1 - h_2) + (j_1 - j_2)^\top R_2 \,\big]\Big\| \\
&\le \mathbb{E}_P\big[\, \big\|j_1^\top (h_1 - h_2)\big\| + \big\|(j_1 - j_2)^\top R_2\big\| \,\big] \\
&\le \mathbb{E}_P\big[\, \|j_1\|_{2,d}\, \|h_1 - h_2\| + \|j_1 - j_2\|_{2,d}\, \|R_2\| \,\big] \\
&\le D\, \mathbb{E}_P\big\|h^\star_{\omega_1,P}(X) - h^\star_{\omega_2,P}(X)\big\| \\
&\qquad + (A + B)\, \mathbb{E}_P\big\|j^\star_{\omega_1,P}(X) - j^\star_{\omega_2,P}(X)\big\|_{2,d},
\end{aligned} \tag{18}$$

where the second line adds and subtracts the mixed term $j_1^\top R_2$; the third uses $R_1 - R_2 = (h_1 - Y) - (h_2 - Y) = h_1 - h_2$, the outcome Y being independent of ω; the fourth uses Jensen's inequality and the triangle inequality; the fifth uses $\|M^\top v\| \le \|M\|_{\mathrm{op}}\|v\| \le \|M\|_{2,d}\|v\|$; and the last uses the almost sure bounds supplied by Assumption A through conditional Jensen, namely $\|j^\star_{\omega,P}(X)\|_{2,d} \le D$ and $\|R_2\| \le \|h^\star_{\omega_2,P}(X)\| + \|Y\| \le A + B$.

It remains to show that both nuisances are Lipschitz in ω. By conditional Jensen and the mean value theorem applied to $\omega \mapsto g_\omega(z)$ along the segment $[\omega_2, \omega_1] \subset \Omega$,

$$\begin{aligned}
\mathbb{E}_P\left[\big\|h^\star_{\omega_1,P}(X) - h^\star_{\omega_2,P}(X)\big\|\right] &\le \mathbb{E}_P\left[\big\|g_{\omega_1}(Z) - g_{\omega_2}(Z)\big\|\right] \\
&\le \sup_{\omega\in\Omega,\, z\in\mathcal{Z}} \|D_\omega g_\omega(z)\|_F\, \|\omega_1 - \omega_2\| \;\le\; D\|\omega_1 - \omega_2\|.
\end{aligned} \tag{19}$$

Similarly, for any unit $u \in \mathbb{R}^d$ the directional derivative of $\omega \mapsto D_\omega g_\omega(z)$ along u is the matrix with columns $\sum_j u_j \partial_{\omega_j}\partial_{\omega_k} g_\omega(z)$, $k = 1, \dots, d$, whose Frobenius norm is precisely the quantity bounded by L in Assumption A. Hence $\omega \mapsto D_\omega g_\omega(z)$ is L-Lipschitz in $\|\cdot\|_F$, uniformly in z, and conditional Jensen gives

$$\mathbb{E}_P\left[\big\|j^\star_{\omega_1,P}(X) - j^\star_{\omega_2,P}(X)\big\|_{2,d}\right] \le \mathbb{E}_P\left[\big\|D_\omega g_{\omega_1}(Z) - D_\omega g_{\omega_2}(Z)\big\|_F\right] \le L\|\omega_1 - \omega_2\|. \tag{20}$$

Substituting (19) and (20) into (18) yields

$$\big\|\nabla\mathcal{F}(\omega_1) - \nabla\mathcal{F}(\omega_2)\big\| \le \underbrace{\big(D^2 + (A + B)L\big)}_{L_{\mathcal{F}}} \|\omega_1 - \omega_2\|.$$

Finally $0 \le \mathcal{F}(\omega) = \frac{1}{2}\mathbb{E}_P\left[\|Y - h^\star_{\omega,P}(X)\|^2\right] \le \frac{1}{2}(A + B)^2$, so $F_{\inf}$ is finite. □

A.2 Projected gradient descent with an inexact hypergradient

Since Ω is compact and convex (Section 2), the Euclidean projection Proj_Ω onto Ω is well-defined and non-expansive: $\|\mathrm{Proj}_\Omega(y) - \mathrm{Proj}_\Omega(y')\| \le \|y - y'\|$ for all $y, y' \in \mathbb{R}^d$. With projection, the correct notion of gradient descent restricted to Ω replaces the unconstrained update by

$$\omega_{t+1} = \mathrm{Proj}_\Omega\big(\omega_t - \eta\widehat{\nabla\mathcal{F}}(\omega_t)\big),$$

and empirical stationarity of this projected iteration is no longer measured by the raw gradient norm, but by the *gradient mapping*. For a point $\omega \in \Omega$, a vector $g \in \mathbb{R}^d$, and step size $\eta > 0$, define

$$G_\eta(\omega; g) \coloneqq \frac{1}{\eta}\Big(\omega - \mathrm{Proj}_\Omega(\omega - \eta g)\Big),$$

and write $G_\eta(\omega) \coloneqq G_\eta(\omega; \nabla\mathcal{F}(\omega))$ for the gradient mapping evaluated at the population gradient, which is the quantity Proposition 2.1 controls; $G_\eta(\omega) = 0$ if and only if ω is a stationary point of $\mathcal{F}$ on Ω in the usual projected sense. We record two standard facts about $G_\eta(\omega; \cdot)$. First, non-expansiveness of Proj_Ω gives, for any $g, g' \in \mathbb{R}^d$,

$$\|G_\eta(\omega; g) - G_\eta(\omega; g')\| \le \|g - g'\|. \tag{21}$$

Second, writing $z = \mathrm{Proj}_\Omega(\omega - \eta g)$ and $G = G_\eta(\omega; g) = (\omega - z)/\eta$, the variational characterization of the Euclidean projection onto the convex set Ω, $\langle (\omega - \eta g) - z, u - z\rangle \le 0$ for all $u \in \Omega$, applied at $u = \omega \in \Omega$ gives $\langle \eta G - \eta g, \eta G\rangle \le 0$, i.e.

$$\langle g, G\rangle \ge \|G\|^2. \tag{22}$$

When $\Omega = \mathbb{R}^d$, Proj_Ω is the identity and $G_\eta(\omega; g) = g$, so both facts reduce to their unconstrained counterparts.

Proof of Proposition 2.1. Fix a realization of $\widehat{\nabla\mathcal{F}}$ and set $\Delta \coloneqq \sup_{\omega\in\Omega} \|\widehat{\nabla\mathcal{F}}(\omega) - \nabla\mathcal{F}(\omega)\|$. Write $e_t \coloneqq \widehat{\nabla\mathcal{F}}(\omega_t) - \nabla\mathcal{F}(\omega_t)$, so that $\|e_t\|_2 \le \Delta$, and let $\widehat{G}_t \coloneqq G_\eta(\omega_t; \widehat{\nabla\mathcal{F}}(\omega_t)) = (\omega_t - \omega_{t+1})/\eta$ denote the *empirical* gradient mapping actually used by the update, as opposed to the population gradient mapping $G_t \coloneqq G_\eta(\omega_t) = G_\eta(\omega_t; \nabla\mathcal{F}(\omega_t))$ that the proposition bounds. By the Lipschitz property (21) applied with $g = \widehat{\nabla\mathcal{F}}(\omega_t)$ and $g' = \nabla\mathcal{F}(\omega_t)$,

$$\|G_t\| \le \|\widehat{G}_t\| + \|e_t\| \le \|\widehat{G}_t\| + \Delta. \tag{23}$$

It therefore suffices to bound $\min_{0\le t<T} \|\widehat{G}_t\|$. By $L_\mathcal{F}$-smoothness and $\omega_{t+1} - \omega_t = -\eta\widehat{G}_t$,

$$\mathcal{F}(\omega_{t+1}) \le \mathcal{F}(\omega_t) - \eta\langle \nabla\mathcal{F}(\omega_t), \widehat{G}_t\rangle + \frac{L_\mathcal{F}\eta^2}{2}\|\widehat{G}_t\|^2.$$

Writing $\langle \nabla\mathcal{F}(\omega_t), \widehat{G}_t\rangle = \langle \widehat{\nabla\mathcal{F}}(\omega_t), \widehat{G}_t\rangle - \langle e_t, \widehat{G}_t\rangle$ and applying the inner-product identity (22) with $g = \widehat{\nabla\mathcal{F}}(\omega_t)$, $G = \widehat{G}_t$, together with Cauchy–Schwarz on the second term,

$$\langle \nabla\mathcal{F}(\omega_t), \widehat{G}_t\rangle \ge \|\widehat{G}_t\|^2 - \|e_t\|\,\|\widehat{G}_t\|.$$

Combining the two displays and using $\eta \le 1/L_\mathcal{F}$ so that $1 - L_\mathcal{F}\eta/2 \ge 1/2$,

$$\mathcal{F}(\omega_{t+1}) \le \mathcal{F}(\omega_t) - \eta\|\widehat{G}_t\|^2\Big(1 - \frac{L_\mathcal{F}\eta}{2}\Big) + \eta\|e_t\|\,\|\widehat{G}_t\| \le \mathcal{F}(\omega_t) - \frac{\eta}{2}\|\widehat{G}_t\|^2 + \eta\|e_t\|\,\|\widehat{G}_t\|.$$

By Young's inequality, $\eta\|e_t\|\,\|\widehat{G}_t\| \le \frac{\eta}{4}\|\widehat{G}_t\|^2 + \eta\|e_t\|^2$, so

$$\mathcal{F}(\omega_{t+1}) \le \mathcal{F}(\omega_t) - \frac{\eta}{4}\|\widehat{G}_t\|^2 + \eta\|e_t\|^2 \le \mathcal{F}(\omega_t) - \frac{\eta}{4}\|\widehat{G}_t\|^2 + \eta\Delta^2.$$

Summing over $t = 0, \dots, T-1$, dividing by $\eta T/4$, and taking the minimum over t,

$$\min_{0\le t<T} \|\widehat{G}_t\|^2 \le \frac{4\{\mathcal{F}(\omega_0) - F_{\inf}\}}{\eta T} + 4\Delta^2,$$

so that, using $\sqrt{a+b} \le \sqrt{a} + \sqrt{b}$, $\min_{0\le t<T} \|\widehat{G}_t\| \le 2\big(\{\mathcal{F}(\omega_0) - F_{\inf}\}/(\eta T)\big)^{1/2} + 2\Delta$. Combining with (23) at the minimizing step,

$$\min_{0\le t<T} \|G_t\| \le 2\left(\frac{\mathcal{F}(\omega_0) - F_{\inf}}{\eta T}\right)^{1/2} + 3\Delta.$$

Taking expectations over both samples and any fitting randomness, with $\mathbb{E}[\Delta] = \mathfrak{s} < \infty$, gives the stated bound with universal implicit constant $\max(2, 3)$. Only the first moment of Δ is needed, since the square root is taken before expectation. □

A.3 Hypergradient bias identities

Proof of Proposition 2.2. Fix $\omega \in \Omega$, $k \in \{1, \dots, d\}$ and $h, j_k \in L^2(P_X)$. Under Assumption A, $\|Y\| \le A$ almost surely, so by Cauchy–Schwarz

$$\mathbb{E}_P\big|\langle h(X) - Y,\, j_k(X)\rangle\big| \;\le\; \big(\|h\|_{L^2(P_X)} + A\big)\,\|j_k\|_{L^2(P_X)} \;<\; \infty;$$

all expectations below are therefore absolutely convergent and the tower property applies. Since h and j_k are $\sigma(X)$-measurable and $\mathbb{E}_P[Y \mid X] = m_P^\star(X)$,

$$\mathbb{E}_P\Big[\big[\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega)\big]_k\Big] = \mathbb{E}_P\big[\varphi^{\mathrm{PI}}(h, j_k)\big] = \big\langle h - m_P^\star,\; j_k\big\rangle_{L^2(P_X)}. \tag{24}$$

Evaluating Equation (24) at $h = h^\star_{\omega,P}$ and $j_k = j^\star_{k,\omega,P}$ and comparing with Equation (3),

$$\partial_{\omega_k}\mathcal{F}(\omega) = \big\langle h^\star_{\omega,P} - m^\star_P,\ j^\star_{k,\omega,P}\big\rangle_{L^2(P_X)}. \tag{25}$$

Adding and subtracting $h^\star_{\omega,P}$ in the first argument of Equation (24) and $j^\star_{k,\omega,P}$ in the second, and expanding by bilinearity of $\langle\cdot,\cdot\rangle_{L^2(P_X)}$,

$$\begin{aligned}\big\langle h - m^\star_P,\ j_k\big\rangle_{L^2(P_X)} &= \big\langle \big(h - h^\star_{\omega,P}\big) + \big(h^\star_{\omega,P} - m^\star_P\big),\ j^\star_{k,\omega,P} + \big(j_k - j^\star_{k,\omega,P}\big)\big\rangle_{L^2(P_X)}\\ &= \big\langle h^\star_{\omega,P} - m^\star_P,\ j^\star_{k,\omega,P}\big\rangle_{L^2(P_X)} + \big\langle h - h^\star_{\omega,P},\ j^\star_{k,\omega,P}\big\rangle_{L^2(P_X)}\\ &\quad + \big\langle h^\star_{\omega,P} - m^\star_P,\ j_k - j^\star_{k,\omega,P}\big\rangle_{L^2(P_X)}\\ &\quad + \big\langle h - h^\star_{\omega,P},\ j_k - j^\star_{k,\omega,P}\big\rangle_{L^2(P_X)}.\end{aligned}$$

The first of the four terms is $\partial_{\omega_k}\mathcal{F}(\omega)$ by Equation (25). Subtracting it from Equation (24),

$$\begin{aligned}\mathbb{E}_P\Big[\big[\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega)\big]_k\Big] - \partial_{\omega_k}\mathcal{F}(\omega) = \underbrace{\big\langle h - h^\star_{\omega,P},\ j^\star_{k,\omega,P}\big\rangle_{L^2(P_X)}}_{\text{(I)}}\\ + \underbrace{\big\langle h^\star_{\omega,P} - m^\star_P,\ j_k - j^\star_{k,\omega,P}\big\rangle_{L^2(P_X)}}_{\text{(II)}}\\ + \underbrace{\big\langle h - h^\star_{\omega,P},\ j_k - j^\star_{k,\omega,P}\big\rangle_{L^2(P_X)}}_{\text{(III)}},\end{aligned}$$

which is Equation (6). □

Proof of Theorem 3.1. Fix $\omega \in \Omega$, $k \in \{1,\dots,d\}$ and $h, j_k, m \in L^2(P_X)$. Expanding the second inner product in the debiased score of Equation (7),

$$\begin{aligned}\varphi_{k,\omega}(h, j_k, m) &= \big\langle g_\omega(Z) - Y,\ j_k(X)\big\rangle\\ &\quad + \big\langle \partial_{\omega_k} g_\omega(Z),\ h(X) - m(X)\big\rangle - \big\langle j_k(X),\ h(X) - m(X)\big\rangle.\end{aligned}$$

Under Assumption A, $\|g_\omega(Z) - Y\| \le A + B$ and $\|\partial_{\omega_k} g_\omega(Z)\| \le \|D_\omega g_\omega(Z)\|_F \le D$ almost surely, so in absolute value the three terms have expectations bounded by $(A+B)\|j_k\|_{L^2(P_X)}$, $D\,\|h-m\|_{L^2(P_X)}$ and $\|j_k\|_{L^2(P_X)}\|h-m\|_{L^2(P_X)}$ respectively; all expectations below are therefore absolutely convergent and the tower property applies. Since h, j_k and m are $\sigma(X)$-measurable, and $\mathbb{E}_P[g_\omega(Z) \mid X] = h^\star_{\omega,P}(X)$, $\mathbb{E}_P[Y \mid X] = m^\star_P(X)$ and $\mathbb{E}_P[\partial_{\omega_k} g_\omega(Z) \mid X] = j^\star_{k,\omega,P}(X)$ by Equation (4), taking $\mathbb{E}_P$ term by term gives

$$\begin{aligned}\mathbb{E}_P\big[\varphi_{k,\omega}(h, j_k, m)\big] &= \big\langle h^\star_{\omega,P} - m^\star_P,\ j_k\big\rangle_{L^2(P_X)} + \big\langle j^\star_{k,\omega,P},\ h - m\big\rangle_{L^2(P_X)}\\ &\quad - \big\langle j_k,\ h - m\big\rangle_{L^2(P_X)}\\ &= \big\langle h^\star_{\omega,P} - m^\star_P,\ j_k\big\rangle_{L^2(P_X)} - \big\langle j_k - j^\star_{k,\omega,P},\ h - m\big\rangle_{L^2(P_X)}.\end{aligned} \tag{26}$$

Evaluating Equation (26) at $h = h^\star_{\omega,P}$, $j_k = j^\star_{k,\omega,P}$ and $m = m^\star_P$, the second term vanishes and comparison with Equation (5) gives

$$\partial_{\omega_k}\mathcal{F}(\omega) = \big\langle h^\star_{\omega,P} - m^\star_P,\ j^\star_{k,\omega,P}\big\rangle_{L^2(P_X)}. \tag{27}$$

Subtracting Equation (27) from Equation (26) and using bilinearity in the first term,

$$\begin{aligned}&\mathbb{E}_P\Big[\big[\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)\big]_k\Big] - \partial_{\omega_k}\mathcal{F}(\omega)\\ &\qquad = \big\langle h^\star_{\omega,P} - m^\star_P,\ j_k - j^\star_{k,\omega,P}\big\rangle_{L^2(P_X)} - \big\langle j_k - j^\star_{k,\omega,P},\ h - m\big\rangle_{L^2(P_X)}\\ &\qquad = \big\langle j_k - j^\star_{k,\omega,P},\ \big(h^\star_{\omega,P} - m^\star_P\big) - \big(h - m\big)\big\rangle_{L^2(P_X)}.\end{aligned}$$

Finally $\big(h^\star_{\omega,P} - m^\star_P\big) - \big(h - m\big) = \big(m - m^\star_P\big) - \big(h - h^\star_{\omega,P}\big)$, so

$$\mathbb{E}_P\Big[\big[\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)\big]_k\Big] - \partial_{\omega_k}\mathcal{F}(\omega) = \big\langle j_k - j^\star_{k,\omega,P},\ \big(m - m^\star_P\big) - \big(h - h^\star_{\omega,P}\big)\big\rangle_{L^2(P_X)},$$

which is Equation (10). □

A.4 AUXILIARY EMPIRICAL-PROCESS LEMMAS

Throughout this subsection, write $\|f\|_{Q,2} \coloneqq (\int \|f\|^2 \, dQ)^{1/2}$ for the $L^2(Q)$ norm of a function f, and $N(\delta, \mathcal{C}, L^2(Q))$ for the δ-covering number of $\mathcal{C}$ with respect to the $L^2(Q)$ metric, *i.e.*, the minimum number of $L^2(Q)$-balls of radius δ needed to cover $\mathcal{C}$. Its logarithm $\log N(\delta, \mathcal{C}, L^2(Q))$ is called the metric entropy of $\mathcal{C}$. For a class $\mathcal{C}$ with envelope C satisfying $\sup_{f \in \mathcal{C}} \|f\| \le C$ pointwise, define

$$\bar{H}(\varepsilon; \mathcal{C}, C) \coloneqq \sup_Q \log N(\varepsilon \|C\|_{Q,2}, \mathcal{C}, L^2(Q)), \qquad 0 < \varepsilon \le 1,$$

where the supremum is over all finitely supported probability measures Q.

Lemma A.2 (Entropy of sums and differences)**.** *Let $\mathcal{C}_1$ and $\mathcal{C}_2$ be classes of measurable functions taking values in the same finite-dimensional normed space, with nonnegative envelopes C_1 and C_2. Then $C_1 + C_2$ is an envelope for $\mathcal{C}_1 \pm \mathcal{C}_2 = \{f_1 \pm f_2 : f_1 \in \mathcal{C}_1,\ f_2 \in \mathcal{C}_2\}$, and*

$$\bar{H}(\varepsilon; \mathcal{C}_1 \pm \mathcal{C}_2, C_1 + C_2) \le \bar{H}(\varepsilon/2; \mathcal{C}_1, C_1) + \bar{H}(\varepsilon/2; \mathcal{C}_2, C_2).$$

Proof. Fix Q. Let $\{f_{1,i}\}_{i=1}^{N_1}$ and $\{f_{2,j}\}_{j=1}^{N_2}$ be minimal $L^2(Q)$-covers of $\mathcal{C}_1$ and $\mathcal{C}_2$ at radii $(\varepsilon/2)\|C_1\|_{Q,2}$ and $(\varepsilon/2)\|C_2\|_{Q,2}$, respectively, so that

$$N_1 = N\big(\tfrac{\varepsilon}{2}\|C_1\|_{Q,2}, \mathcal{C}_1, L^2(Q)\big), \qquad N_2 = N\big(\tfrac{\varepsilon}{2}\|C_2\|_{Q,2}, \mathcal{C}_2, L^2(Q)\big).$$

For any $f = f_1 \pm f_2 \in \mathcal{C}_1 \pm \mathcal{C}_2$, find the closest cover elements $f_{1,i}$ and $f_{2,j}$. By the triangle inequality,

$$\|f_1 \pm f_2 - (f_{1,i} \pm f_{2,j})\|_{Q,2} \le \|f_1 - f_{1,i}\|_{Q,2} + \|f_2 - f_{2,j}\|_{Q,2} \le \frac{\varepsilon}{2}\|C_1\|_{Q,2} + \frac{\varepsilon}{2}\|C_2\|_{Q,2}.$$

Since C_1 and C_2 are nonnegative, $C_1 + C_2 \ge C_1$ and $C_1 + C_2 \ge C_2$ pointwise, so $\|C_1 + C_2\|_{Q,2} \ge \|C_1\|_{Q,2}$ and $\|C_1 + C_2\|_{Q,2} \ge \|C_2\|_{Q,2}$. Adding these two inequalities gives

$$\frac{\varepsilon}{2}\|C_1\|_{Q,2} + \frac{\varepsilon}{2}\|C_2\|_{Q,2} \le \varepsilon\|C_1 + C_2\|_{Q,2},$$

so $\{f_{1,i} \pm f_{2,j}\}$ covers $\mathcal{C}_1 \pm \mathcal{C}_2$ at radius $\varepsilon\|C_1 + C_2\|_{Q,2}$, confirming that $C_1 + C_2$ is an envelope for $\mathcal{C}_1 \pm \mathcal{C}_2$. Since this cover has at most $N_1 \cdot N_2$ elements,

$$\log N\big(\varepsilon\|C_1 + C_2\|_{Q,2}, \mathcal{C}_1 \pm \mathcal{C}_2, L^2(Q)\big) \le \log N_1 + \log N_2.$$

Taking the supremum over all finitely supported Q on both sides gives

$$\bar{H}(\varepsilon; \mathcal{C}_1 \pm \mathcal{C}_2, C_1 + C_2) \le \bar{H}(\varepsilon/2; \mathcal{C}_1, C_1) + \bar{H}(\varepsilon/2; \mathcal{C}_2, C_2).$$

□

Lemma A.3 (Entropy of bounded products)**.** *Let $\mathcal{U}$ and $\mathcal{V}$ be vector-valued function classes such that, for finite constants $\bar{U}, \bar{V}$, $\sup_{u \in \mathcal{U}} \|u(O)\| \le \bar{U}$ and $\sup_{v \in \mathcal{V}} \|v(O)\| \le \bar{V}$ for all O. Define*

$$\mathcal{U} \odot \mathcal{V} \coloneqq \{O \mapsto \langle u(O), v(O)\rangle : u \in \mathcal{U},\ v \in \mathcal{V}\}.$$

Then $\bar{U}\bar{V}$ is an envelope for $\mathcal{U} \odot \mathcal{V}$, and for every $0 < \varepsilon \le 1$,

$$\sup_Q \log N(\varepsilon\bar{U}\bar{V}, \mathcal{U} \odot \mathcal{V}, L^2(Q)) \le \sup_Q \log N\Big(\tfrac{\varepsilon\bar{U}}{2}, \mathcal{U}, L^2(Q)\Big) + \sup_Q \log N\Big(\tfrac{\varepsilon\bar{V}}{2}, \mathcal{V}, L^2(Q)\Big).$$

Proof. The envelope bound follows from Cauchy-Schwarz: for all O,

$$|\langle u(O), v(O)\rangle| \le \|u(O)\|\|v(O)\| \le \bar{U}\bar{V}.$$

For the entropy bound, fix Q and let $\{u_i\}_{i=1}^{N_1}$ and $\{v_j\}_{j=1}^{N_2}$ be minimal $L^2(Q)$-covers of $\mathcal{U}$ and $\mathcal{V}$ at radii $\varepsilon\bar{U}/2$ and $\varepsilon\bar{V}/2$ respectively. For any $u \in \mathcal{U}$ and $v \in \mathcal{V}$, find the closest cover elements u_i and v_j. By the triangle inequality and the pointwise bounds,

$$\begin{aligned}\|\langle u, v\rangle - \langle u_i, v_j\rangle\|_{Q,2} &\le \|\langle u - u_i, v\rangle\|_{Q,2} + \|\langle u_i, v - v_j\rangle\|_{Q,2} \\ &\le \bar{V}\|u - u_i\|_{Q,2} + \bar{U}\|v - v_j\|_{Q,2} \\ &\le \bar{V} \cdot \frac{\varepsilon\bar{U}}{2} + \bar{U} \cdot \frac{\varepsilon\bar{V}}{2} = \varepsilon\bar{U}\bar{V}.\end{aligned}$$

So $\{u_i \odot v_j\}$ covers $\mathcal{U} \odot \mathcal{V}$ at radius $\varepsilon\bar{U}\bar{V}$ with at most $N_1 \cdot N_2$ elements. Taking logarithms and the supremum over Q gives the result. □

Lemma A.4 (Entropy of block linear combinations)**.** *Let $\mathcal{A}$ be a class of block-vector-valued functions $a = (a_1, \dots, a_d)$ with values in $(\mathbb{R}^q)^d$, equipped with the block norm $\|a(O)\|_{2,d} := (\sum_{k=1}^d \|a_k(O)\|^2)^{1/2}$, and suppose $\sup_{a\in\mathcal{A}} \|a(O)\|_{2,d} \le \bar{A}$ for all O. Define*

$$\mathscr{S}(\mathcal{A}) := \left\{ O \mapsto \sum_{k=1}^d u_k a_k(O) : a \in \mathcal{A},\ u \in \mathbb{S}^{d-1} \right\}.$$

Then $\bar{A}$ is an envelope for $\mathscr{S}(\mathcal{A})$, and for every $0 < \varepsilon \le 1$,

$$\sup_Q \log N(\varepsilon\bar{A}, \mathscr{S}(\mathcal{A}), L^2(Q)) \le d \log(1 + 4/\varepsilon) + \sup_Q \log N(\varepsilon\bar{A}/2, \mathcal{A}, L^2(Q)).$$

Proof. The envelope bound follows from Cauchy-Schwarz: for all O and $u \in \mathbb{S}^{d-1}$,

$$\left| \sum_{k=1}^d u_k a_k(O) \right| \le \|u\|_2 \|a(O)\|_{2,d} \le \bar{A}.$$

For the entropy bound, fix Q. Let $\{a_i\}_{i=1}^{N_1}$ be a minimal $L^2(Q)$-cover of $\mathcal{A}$ at radius $\varepsilon\bar{A}/2$ in the block norm, and let $\{u_\ell\}_{\ell=1}^{N_2}$ be an $(\varepsilon/2)$-cover of $\mathbb{S}^{d-1}$ in Euclidean norm, which exists with cardinality $N_2 \le (1+4/\varepsilon)^d$ by a standard volumetric argument (van der Vaart & Wellner, 1996). For any $a \in \mathcal{A}$ and $u \in \mathbb{S}^{d-1}$, choose the closest cover elements a_i and u_ℓ. By the triangle inequality,

$$\begin{aligned}
\left\| \sum_{k=1}^d u_k a_k - \sum_{k=1}^d (u_\ell)_k (a_i)_k \right\|_{Q,2} &\le \left\| \sum_{k=1}^d u_k (a_k - (a_i)_k) \right\|_{Q,2} + \left\| \sum_{k=1}^d (u_k - (u_\ell)_k)(a_i)_k \right\|_{Q,2} \\
&\le \|u\|_2 \|a - a_i\|_{Q,2,d} + \|u - u_\ell\|_2 \bar{A} \\
&\le \frac{\varepsilon\bar{A}}{2} + \frac{\varepsilon}{2}\bar{A} = \varepsilon\bar{A}.
\end{aligned}$$

So $\{u_\ell \odot a_i\}$ covers $\mathscr{S}(\mathcal{A})$ at radius $\varepsilon\bar{A}$ with at most $N_1 \cdot N_2$ elements. Taking logarithms and the supremum over Q gives

$$\begin{aligned}
\sup_Q \log N(\varepsilon\bar{A}, \mathscr{S}(\mathcal{A}), L^2(Q)) &\le \log N_2 + \sup_Q \log N_1 \\
&\le d \log(1 + 4/\varepsilon) + \sup_Q \log N(\varepsilon\bar{A}/2, \mathcal{A}, L^2(Q)).
\end{aligned}$$

□

Let S_1 and S_2 be independent folds of n i.i.d. observations $O = (X, Y, Z) \sim P$. The nuisances are fitted on S_1 and the hypergradient is evaluated on S_2. For a class $\mathcal{G}$ of real-valued measurable functions, write $\mathrm{star}(\mathcal{G}) := \{\alpha g : \alpha \in [0,1],\ g \in \mathcal{G}\}$ for its star hull around zero. For $\delta > 0$, define the conditional local Rademacher complexity

$$\mathfrak{R}_n(\mathcal{G}, \delta \mid S_1) := \mathbb{E}_{\varepsilon, S_2} \left[\sup_{\substack{g\in\mathcal{G} \\ \|g\|_{L^2(P)} \le \delta}} \left| \frac{1}{n} \sum_{i \in S_2} \varepsilon_i g(O_i) \right| \,\middle|\, S_1 \right],$$

where $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. Rademacher random variables independent of S_2, and the expectation is over both ε and S_2 conditionally on S_1.

Lemma A.5 (Localized OSL maximal inequality)**.** *Let $\mathcal{G}$ be a pointwise measurable real-valued class with envelope $G < \infty$ and $\sup_{g\in\mathcal{G}} \|g\|_{L_2(P)} \le \rho$. Define $\overline{\mathcal{G}}^0 = \{g/G : g \in \mathcal{G}\} \cup \{0\}$. Suppose δ_n satisfies*

$$\mathfrak{R}_n\left(\mathrm{star}(\overline{\mathcal{G}}^0), \delta_n\right) \le \delta_n^2.$$

Then there exist universal constants $C, c < \infty$ such that

$$\mathbb{E}\left[\sup_{g\in\mathcal{G}} |(P_n - P)g| \right] \le C \left\{ \delta_n \rho + G\delta_n^2 + G e^{-cn\delta_n^2} \right\}.$$

The same statement holds conditionally on S_1 for an S_1-measurable class and envelope, with the expectation taken over the evaluation fold.

Proof. Apply Lemma 14 in (Foster & Syrgkanis, 2023) to $\overline{\mathcal{G}}^0$ with identity loss $L_f = f$ and comparator $f^\star = 0$. The identity loss is linear, so the lower bound on δ_n^2 in Lemma 14 is not required. Here the dimension parameter in Lemma 14 is one, because $\mathcal{G}$ is real-valued; the outer-gradient dimension has already been absorbed by scalarization over $u \in S^{d-1}$. Then, with probability at least $1 - Ce^{-cn\delta_n^2}$,

$$|(P_n - P)\bar{g}| \le C\delta_n(\|\bar{g}\|_{L_2(P)} + \delta_n), \qquad \forall \bar{g} \in \overline{\mathcal{G}}^0.$$

Multiplying by G and using $\|g\|_{L^2(P)} \le \rho$ gives

$$|(P_n - P)g| \le C\{\delta_n \|g\|_{L^2(P)} + G\delta_n^2\} \le C\{\delta_n \rho + G\delta_n^2\}$$

on the good event. On the complement, which has probability at most $Ce^{-cn\delta_n^2}$, the supremum is bounded by $2G$ since $|(P_n - P)g| \le 2\|g\|_\infty \le 2G$ for all $g \in \mathcal{G}$. Taking expectations and combining gives the result. □

We also use the following standard maximal inequality for classes with finite uniform entropy integral (van der Vaart & Wellner, 1996, Chapter 2); see also (Kosorok, 2008, Chapter 8).

Lemma A.6 (Uniform entropy maximal inequality)**.** *Let $\mathcal{F}$ be a pointwise measurable class of real-valued functions with envelope $F \in L^2(P)$. Define the uniform entropy integral*

$$J(1, \mathcal{F}, F) := \int_0^1 \sup_Q \sqrt{1 + \log N\big(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L^2(Q)\big)}\, d\varepsilon,$$

where the supremum is over all finitely supported probability measures Q. If $J(1, \mathcal{F}, F) < \infty$, then there exists a universal constant $0 < C < \infty$ such that

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} |(P_n - P)f|\right] \le \frac{C}{\sqrt{n}} J(1, \mathcal{F}, F) \|F\|_{L^2(P)}.$$

Corollary A.1 (Polynomial entropy maximal inequality)**.** *Suppose that, for some $p \in (0, 2)$ and $K < \infty$,*

$$\sup_Q \log N\big(\varepsilon \|F\|_{Q,2}, \mathcal{F}, L^2(Q)\big) \le K\varepsilon^{-p}, \qquad 0 < \varepsilon \le 1.$$

Then there exists a constant $C_p < \infty$ depending only on p such that

$$\mathbb{E}\left[\sup_{f \in \mathcal{F}} |(P_n - P)f|\right] \le \frac{C_p}{\sqrt{n}} \|F\|_{L^2(P)} \sqrt{1 + K}.$$

Proof. Since $\varepsilon \le 1$ implies $1 \le \varepsilon^{-p}$, we have $1 + K\varepsilon^{-p} \le (1 + K)\varepsilon^{-p}$, and therefore

$$J(1, \mathcal{F}, F) \le \int_0^1 \sqrt{1 + K\varepsilon^{-p}}\, d\varepsilon \le \sqrt{1 + K} \int_0^1 \varepsilon^{-p/2}\, d\varepsilon = \frac{2\sqrt{1 + K}}{2 - p} < \infty,$$

where finiteness uses $p < 2$. Applying Lemma A.6 with $C_p := 2C/(2 - p)$ gives the claim. □

A.5 Conditional maximal inequality

To control the nuisance empirical process, we require the learned nuisance paths to lie in well-controlled function classes, which we now formalize.

Assumption E (Training-conditional nuisance classes)**.** Conditionally on S_1, there are pointwise measurable random classes $\mathcal{H}_n \subset L^2(P_X)$, $\mathcal{J}_n \subset L^2(P_X)^d$, and $\mathcal{M}_n \subset L^2(P_X)$ containing the learned nuisance paths $\{\hat{h}_\omega : \omega \in \Omega\} \subseteq \mathcal{H}_n$, $\{\hat{j}_\omega : \omega \in \Omega\} \subset \mathcal{J}_n$, and $\hat{m} \in \mathcal{M}_n$ almost surely, with S_1-measurable envelopes $\bar{H}_n, \bar{J}_n, \bar{M}_n < \infty$ satisfying

$$\sup_{h \in \mathcal{H}_n} \|h(X)\| \le \bar{H}_n, \qquad \sup_{j \in \mathcal{J}_n} \sum_{k=1}^d \|j_k(X)\|^2 \le \bar{J}_n^2, \qquad \sup_{m \in \mathcal{M}_n} \|m(X)\| \le \bar{M}_n.$$

Kernel ridge regression, sieve estimators, and norm-controlled neural networks satisfy assumptions of this form under the usual envelope and entropy controls; representative sufficient conditions are collected in Appendix B.1. The entropy of both the population classes and the training-conditional nuisance classes is controlled by the following two assumptions. Both assumptions below use the following notation: for a function class $\mathcal{C}$ with envelope $E_{\mathcal{C}}$ satisfying $\sup_{f\in\mathcal{C}}\|f\| \le E_{\mathcal{C}}$ pointwise, $N(\varepsilon E_{\mathcal{C}}, \mathcal{C}, L^2(Q))$ denotes the $\varepsilon E_{\mathcal{C}}$-covering number of $\mathcal{C}$ in $L^2(Q)$, and $\sup_Q$ is taken over all finitely supported probability measures Q on $\mathcal{X}\times\mathcal{Y}\times\mathcal{Z}$.

Assumption F (Entropy of population classes)**.** Define the population function classes indexed by $\omega\in\Omega$: the structural map class $\mathcal{G}_\Omega = \{g_\omega : \omega\in\Omega\}$, its gradient class $\dot{\mathcal{G}}_\Omega = \{(\partial_{\omega_1} g_\omega, \dots, \partial_{\omega_d} g_\omega) : \omega\in\Omega\}$, the oracle inner solution class $\mathcal{H}^\star = \{h^\star_{\omega,P} : \omega\in\Omega\}$, and the oracle Jacobian class $\mathcal{J}^\star = \{j^\star_{\omega,P} : \omega\in\Omega\}$. There exist $p\in(0,2)$ and finite constants $K_{\mathcal{G}_\Omega}, K_{\dot{\mathcal{G}}_\Omega}, K_{\mathcal{H}^\star}, K_{\mathcal{J}^\star}$ such that, for every $0<\varepsilon\le 1$,

$$\sup_Q \log N(\varepsilon E_{\mathcal{C}}, \mathcal{C}, L^2(Q)) \le K_{\mathcal{C}}\varepsilon^{-p}, \qquad \mathcal{C}\in\{\mathcal{G}_\Omega, \dot{\mathcal{G}}_\Omega, \mathcal{H}^\star, \mathcal{J}^\star\},$$

with envelopes $E_{\mathcal{G}_\Omega} = E_{\mathcal{H}^\star} = B$ and $E_{\dot{\mathcal{G}}_\Omega} = E_{\mathcal{J}^\star} = D$.

In words, this bounds how complex the population function classes are allowed to be as ω ranges over Ω: the logarithm of the number of ε-balls needed to cover each class grows at most polynomially in $1/\varepsilon$, with exponent $p<2$. This is a standard complexity condition from empirical process theory, and it is mild in the sense that it is exactly what holds for the usual nonparametric smoothness classes. For example, standard s-Hölder classes on a d_X-dimensional domain satisfy such an entropy bound with $p = d_X/s$; hence the condition $p<2$ corresponds to the usual smoothness requirement $s > d_X/2$ in this case.

Assumption G (Entropy of training-conditional nuisance classes)**.** Conditionally on S_1, there are finite S_1-measurable constants $K_{\mathcal{H}_n}, K_{\mathcal{J}_n}, K_{\mathcal{M}_n}$ such that the learned nuisance classes $\mathcal{H}_n$, $\mathcal{J}_n$, $\mathcal{M}_n$ satisfy the same polynomial entropy bound with exponent $p\in(0,2)$: for every $0<\varepsilon\le 1$,

$$\sup_Q \log N(\varepsilon E_{\mathcal{C}}, \mathcal{C}, L^2(Q)) \le K_{\mathcal{C}}\varepsilon^{-p}, \qquad \mathcal{C}\in\{\mathcal{H}_n, \mathcal{J}_n, \mathcal{M}_n\},$$

with envelopes $E_{\mathcal{H}_n} = \bar{H}_n$, $E_{\mathcal{J}_n} = \bar{J}_n$, and $E_{\mathcal{M}_n} = \bar{M}_n$.

The condition $p\in(0,2)$ ensures that the corresponding entropy integral is finite, providing the uniform empirical-process control required below (van der Vaart & Wellner, 1996; Kosorok, 2008). These entropy and envelope conditions are sufficient rather than necessary. They are satisfied by a range of standard nuisance classes, including bounded Hölder classes with smoothness $s > d_X/2$, fixed-dimensional sieve classes, kernel ridge regression classes under suitable bounded-kernel assumptions, and appropriately norm-controlled neural-network classes.

Denote by $\mathcal{E}_{\mathrm{tr}}(r_h, r_j, r_m)$ the event where $\sup_{\omega\in\Omega}\|\hat{h}_\omega - h^\star_{\omega,P}\|_{L^2(P_X)} \le r_h$, $\sup_{\omega\in\Omega}\|\hat{j}_\omega - j^\star_{\omega,P}\|_{L^2(P_X)^d} \le r_j$, and $\|\hat{m} - m^\star_P\|_{L^2(P_X)} \le r_m$.

Theorem A.1 (Expected maximal inequality)**.** *Under Assumptions A and E to G, there exist constants $C>0$ depending only on p and $K>0$ depending only on the population entropy constants such that, writing $\mathfrak{s}_N := \mathbb{E}\left[\sup_{\omega\in\Omega}\|\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega)\| \,\middle|\, S_1\right]$ for the expected uniform error, on $\mathcal{E}_{\mathrm{tr}}(r_h, r_j, r_m)$,*

$$\mathfrak{s}_N \le \frac{C}{\sqrt{n}}(A+B)D\sqrt{1+K} + \mathfrak{C}_{B,n}(r_h, r_j, r_m) + r_j(r_h + r_m), \tag{28}$$

where $\mathfrak{C}_{B,n}(r_h, r_j, r_m) := \mathbb{E}\left[\sup_{f\in\mathcal{F}^{\mathrm{loc}}_{B,n}(r_h,r_j,r_m)}\left|(\mathbb{E}_{P_{n,2}} - \mathbb{E}_P)[f]\right| \,\middle|\, S_1\right]$, with localized nuisance-difference class

$$\begin{aligned}\mathcal{F}^{\mathrm{loc}}_{B,n}(r_h, r_j, r_m) := \big\{ & u^\top\left[\varphi_\omega(\cdot; h, j, m) - \varphi_\omega(\cdot; h^\star_{\omega,P}, j^\star_{\omega,P}, m^\star_P)\right] : \\ & \omega\in\Omega,\ u\in\mathbb{S}^{d-1},\ h\in\mathcal{H}_n,\ j\in\mathcal{J}_n,\ m\in\mathcal{M}_n, \\ & \|h - h^\star_{\omega,P}\|_{L^2(P_X)} \le r_h, \|j - j^\star_{\omega,P}\|_{L^2(P_X)^d} \le r_j, \|m - m^\star_P\|_{L^2(P_X)} \le r_m\big\},\end{aligned}$$

which localizes the nuisance functions to a ball of radius (r_h, r_j, r_m) around the oracle, conditionally on S_1.

Appendix A.6 bounds $\mathfrak{C}_{B,n}$ via localized Rademacher critical radii (Bartlett et al., 2005; Foster & Syrgkanis, 2023).

Remark A.1 (Only the nuisance cross-products matter). The population-bias term in (28) is $r_j(r_h + r_m)$: only products of the Jacobian error with the conditional-mean errors enter. The nuisance rates need not be balanced. An $O(n^{-1/2})$ conditional expected uniform gradient rate follows if both $r_j(r_h + r_m)$ and $\mathfrak{C}_{B,n}(r_h, r_j, r_m)$ are $O(n^{-1/2})$; making both $o(n^{-1/2})$ renders the nuisance contributions negligible relative to the oracle first-order term. Thus a slower Jacobian rate can be offset by faster conditional-mean rates. Under the common rate $r_h, r_j, r_m = O(n^{-1/(2+p)})$ in Proposition A.1, both bias cross-products are $O(n^{-2/(2+p)}) = o(n^{-1/2})$ for $p < 2$.

Proof of Theorem A.1. The proof proceeds in the following 6 steps.

Step 1: population identities. Under the quadratic specialization, the unique population inner minimizer is $h^\star_{\omega,P}(X) = \mathbb{E}_P[g_\omega(Z) \mid X]$ and $j^\star_{k,\omega,P}(X) = \mathbb{E}_P[\partial_{\omega_k} g_\omega(Z) \mid X]$ for $k = 1, \dots, d$. The outer objective and its gradient are

$$\mathcal{F}(\omega) = \frac{1}{2}\mathbb{E}_P\left[\|Y - h^\star_{\omega,P}(X)\|^2\right], \qquad \partial_{\omega_k}\mathcal{F}(\omega) = \mathbb{E}_P[\langle h^\star_{\omega,P}(X) - Y, j^\star_{k,\omega,P}(X)\rangle].$$

By the conditional Jensen inequality and Assumption A,

$$\|h^\star_{\omega,P}(X)\| \le B, \qquad \|m^\star_P(X)\| \le A, \qquad \|j^\star_{\omega,P}(X)\|_{2,d} \le D,$$

where $\|a\|_{2,d} \coloneqq (\sum_{k=1}^d \|a_k\|^2)^{1/2}$ for $a = (a_1, \dots, a_d) \in (\mathbb{R}^q)^d$.

Step 2: exact bias identity. At the oracle nuisances, $\mathbb{E}_P[\varphi_\omega(\cdot; h^\star_{\omega,P}, j^\star_{\omega,P}, m^\star_P)] = \nabla\mathcal{F}(\omega)$. Indeed, for each coordinate k,

$$\begin{aligned}\mathbb{E}_P\left[\varphi_{k,\omega}(X, Y, Z; h^\star_{\omega,P}, j^\star_{\omega,P}, m^\star_P)\right] &= \mathbb{E}_P\left[\langle g_\omega(Z) - Y, j^\star_{k,\omega,P}(X)\rangle\right] \\ &\quad + \mathbb{E}_P\left[\langle \partial_{\omega_k} g_\omega(Z) - j^\star_{k,\omega,P}(X), h^\star_{\omega,P}(X) - m^\star_P(X)\rangle\right] \\ &= \mathbb{E}_P\left[\langle h^\star_{\omega,P}(X) - Y, j^\star_{k,\omega,P}(X)\rangle\right] = \partial_{\omega_k}\mathcal{F}(\omega),\end{aligned}$$

where the second term vanishes because $j^\star_{k,\omega,P}(X) = \mathbb{E}_P[\partial_{\omega_k} g_\omega(Z) \mid X]$, so $\mathbb{E}_P[\partial_{\omega_k} g_\omega(Z) - j^\star_{k,\omega,P}(X) \mid X] = 0$ and the tower property applies.

For arbitrary $h \in L^2(P_X)$, $j \in L^2(P_X)^d$, and $m \in L^2(P_X)$, the tower property gives

$$\begin{aligned}\mathbb{E}_P[\varphi_{k,\omega}(\cdot; h, j, m)] - \partial_{\omega_k}\mathcal{F}(\omega) &= -\mathbb{E}_P[\langle j_k(X) - j^\star_{k,\omega,P}(X), h(X) - h^\star_{\omega,P}(X)\rangle] \\ &\quad + \mathbb{E}_P[\langle j_k(X) - j^\star_{k,\omega,P}(X), m(X) - m^\star_P(X)\rangle].\end{aligned}$$

By Cauchy-Schwarz, summing over k and using $\|u\| = 1$,

$$\left\|\mathbb{E}_P[\varphi_\omega(\cdot; h, j, m)] - \nabla\mathcal{F}(\omega)\right\| \le \|j - j^\star_{\omega,P}\|_{L^2(P_X)^d}\left(\|h - h^\star_{\omega,P}\|_{L^2(P_X)} + \|m - m^\star_P\|_{L^2(P_X)}\right). \tag{29}$$

Step 3: estimator decomposition. Write $\varphi^\star_\omega \coloneqq \varphi_\omega(\cdot; h^\star_{\omega,P}, j^\star_{\omega,P}, m^\star_P)$ and $\hat\varphi_\omega \coloneqq \varphi_\omega(\cdot; \hat h_\omega, \hat j_\omega, \hat m)$. Then

$$\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega) = (\mathbb{E}_{P_{n,2}} - \mathbb{E}_P)[\varphi^\star_\omega] + (\mathbb{E}_{P_{n,2}} - \mathbb{E}_P)[\hat\varphi_\omega - \varphi^\star_\omega] + \Big(\mathbb{E}_P[\hat\varphi_\omega] - \nabla\mathcal{F}(\omega)\Big).$$

We work throughout on the event $\mathcal{E}_{\mathrm{tr}}(r_h, r_j, r_m)$. On this event, the product-bias bound (29) gives

$$\sup_{\omega\in\Omega}\left\|\mathbb{E}_P[\hat\varphi_\omega] - \nabla\mathcal{F}(\omega)\right\| \le r_j(r_h + r_m).$$

Using the dual representation $\|v\| = \sup_{u\in\mathbb{S}^{d-1}} u^\top v$, where $\mathbb{S}^{d-1} = \{u \in \mathbb{R}^d : \|u\| = 1\}$, we obtain

$$\begin{aligned}\mathbf{1}_{\mathcal{E}_{\mathrm{tr}}(r_h,r_j,r_m)} \sup_{\omega\in\Omega}\|\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega)\| &\le \mathbf{1}_{\mathcal{E}_{\mathrm{tr}}(r_h,r_j,r_m)} \sup_{f\in\mathcal{F}_A}|(P_{n,2} - P)f| \\ &\quad + \mathbf{1}_{\mathcal{E}_{\mathrm{tr}}(r_h,r_j,r_m)} \sup_{f\in\mathcal{F}^{\mathrm{loc}}_{B,n}(r_h,r_j,r_m)}|(P_{n,2} - P)f| \\ &\quad + r_j(r_h + r_m),\end{aligned}$$

where

$$\mathcal{F}_A \coloneqq \left\{u^\top \left(\varphi_\omega(\cdot; h^\star_{\omega,P}, j^\star_{\omega,P}, m^\star_P) - \nabla\mathcal{F}(\omega)\right) : \omega \in \Omega,\ u \in \mathbb{S}^{d-1}\right\}$$

is the oracle score class and $\mathcal{F}^{\mathrm{loc}}_{B,n}(r_h, r_j, r_m)$ is the localized nuisance-difference class defined in Theorem A.1.

Step 4: envelopes. For the oracle score, by Cauchy-Schwarz and the bounds of Step 1,

$$\begin{aligned}\|\varphi^\star_\omega(X,Y,Z)\| &\le \|g_\omega(Z) - Y\|\|j^\star_{\omega,P}(X)\|_{2,d} + \|\partial_\omega g_\omega(Z) - j^\star_{\omega,P}(X)\|_{2,d}\|h^\star_{\omega,P}(X) - m^\star_P(X)\| \\ &\le (A+B)D + 2D(A+B) = 3(A+B)D,\end{aligned}$$

where $\|\partial_\omega g_\omega(Z) - j^\star_{\omega,P}(X)\|_{2,d} \le \|\partial_\omega g_\omega(Z)\|_{2,d} + \|j^\star_{\omega,P}(X)\|_{2,d} \le 2D$ by the triangle inequality and Assumption A. Since $\nabla\mathcal{F}(\omega) = \mathbb{E}_P[\varphi^\star_\omega]$, Jensen's inequality gives $\|\nabla\mathcal{F}(\omega)\| \le 3(A+B)D$. Therefore $F_A \coloneqq 6(A+B)D$ is an envelope for $\mathcal{F}_A$.

For the nuisance-difference class, let $\mathcal{F}_{B,n}$ denote the ambient class obtained from $\mathcal{F}^{\mathrm{loc}}_{B,n}(r_h, r_j, r_m)$ by dropping the radius constraints. Fix $\omega \in \Omega$, $u \in \mathbb{S}^{d-1}$, $h \in \mathcal{H}_n$, $j \in \mathcal{J}_n$, and $m \in \mathcal{M}_n$. Write $\Delta h = h - h^\star_{\omega,P}$, $\Delta j = j - j^\star_{\omega,P}$, and $\Delta m = m - m^\star_P$. The score difference expands as

$$\begin{aligned}\varphi_{k,\omega}(X,Y,Z;h,j,m) - \varphi^\star_{k,\omega}(X,Y,Z) &= \langle g_\omega(Z) - Y, \Delta j_k(X)\rangle \\ &\quad + \langle \partial_{\omega_k} g_\omega(Z) - j^\star_{k,\omega,P}(X), \Delta h(X) - \Delta m(X)\rangle \\ &\quad - \langle \Delta j_k(X), h(X) - m(X)\rangle.\end{aligned}$$

By Assumptions A and E,

$$\|\Delta j(X)\|_{2,d} \le \bar{J}_n + D, \qquad \|\Delta h(X) - \Delta m(X)\| \le \bar{H}_n + \bar{M}_n + A + B,$$

and

$$\|h(X) - m(X)\| \le \bar{H}_n + \bar{M}_n, \qquad \|\partial_\omega g_\omega(Z) - j^\star_\omega(X)\|_{2,d} \le 2D.$$

Thus

$$F_{B,n} \coloneqq (A+B)(\bar{J}_n + D) + 2D(\bar{H}_n + \bar{M}_n + A + B) + (\bar{J}_n + D)(\bar{H}_n + \bar{M}_n)$$

is an S_1-measurable envelope for $\mathcal{F}_{B,n}$, and hence also for $\mathcal{F}^{\mathrm{loc}}_{B,n}(r_h, r_j, r_m)$. Moreover, every $f \in \mathcal{F}^{\mathrm{loc}}_{B,n}(r_h, r_j, r_m)$ satisfies

$$\|f\|_{L^2(P)} \le (A + B + \bar{H}_n + \bar{M}_n)r_j + 2D(r_h + r_m) =: \rho_{B,n},$$

where the three terms in the score difference contribute $(A+B)r_j$, $2D(r_h+r_m)$, and $(\bar{H}_n+\bar{M}_n)r_j$ respectively.

Step 5: entropy transfer. We transfer the polynomial entropy bounds of Assumptions F and G to the score classes $\mathcal{F}_A$ and $\mathcal{F}_{B,n}$ via Lemmas A.2, A.3, and A.4. Throughout, we use the notation

$$\mathscr{S}(\mathcal{A}) \coloneqq \left\{(X,Y,Z) \mapsto \sum_{k=1}^d u_k a_k(X,Y,Z) : a = (a_1,\dots,a_d) \in \mathcal{A},\ u \in \mathbb{S}^{d-1}\right\}$$

for the scalarized class associated with a block-vector class $\mathcal{A}$.

Oracle score class $\mathcal{F}_A$. Define $\mathcal{U}_g \coloneqq \mathcal{G}_\Omega - Y$, $\mathcal{V}^\star \coloneqq \mathcal{H}^\star - m^\star_P$, $\mathscr{J}^\star \coloneqq \mathscr{S}(\mathcal{J}^\star)$, $\dot{\mathscr{G}} \coloneqq \mathscr{S}(\dot{\mathcal{G}}_\Omega)$, and $\Delta\dot{\mathscr{G}}^\star \coloneqq \dot{\mathscr{G}} - \mathscr{J}^\star$. Reading off the three terms in the oracle score, we obtain the inclusion

$$\mathcal{F}_A \subset (\mathcal{U}_g \odot \mathscr{J}^\star) + (\Delta\dot{\mathscr{G}}^\star \odot \mathcal{V}^\star) - \mathcal{C}_\Psi,$$

where $\mathcal{C}_\Psi \coloneqq \{(X,Y,Z) \mapsto u^\top\nabla\mathcal{F}(\omega) : \omega \in \Omega,\ u \in \mathbb{S}^{d-1}\}$ is a class of deterministic constant functions bounded by $3(A+B)D$. Its ε-covering number grows at most at rate $d\log(1/\varepsilon)$, which is dominated by $K\varepsilon^{-p}$ for any $p > 0$. Applying Lemmas A.2, A.3, and A.4 to each product and sum, together with Assumption F, yields a deterministic constant $K_A < \infty$ depending only on the population entropy constants such that

$$\sup_Q \log N(\varepsilon F_A, \mathcal{F}_A, L^2(Q)) \le K_A\varepsilon^{-p}, \qquad 0 < \varepsilon \le 1.$$

Nuisance-difference class $\mathcal{F}_{B,n}$. Condition on S_1 and define $\mathscr{J}_n \coloneqq \mathscr{S}(\mathcal{J}_n)$, $\Delta\mathscr{J}_n \coloneqq \mathscr{J}_n - \mathscr{J}^\star$, $\Delta\mathcal{H}_n \coloneqq \mathcal{H}_n - \mathcal{H}^\star$, $\Delta\mathcal{M}_n \coloneqq \mathcal{M}_n - m_P^\star$, $\Delta\mathcal{V}_n \coloneqq \Delta\mathcal{H}_n - \Delta\mathcal{M}_n$, and $\mathcal{W}_n \coloneqq \mathcal{H}_n - \mathcal{M}_n$. The score difference expands as

$$\begin{aligned} u^\top\big(\varphi_\omega(\cdot;h,j,m) - \varphi_\omega^\star\big) = &\left\langle g_\omega - Y, \sum_{k=1}^d u_k(j_k - j_{k,\omega}^\star)\right\rangle \\ &+ \left\langle \sum_{k=1}^d u_k(\partial_{\omega_k} g_\omega - j_{k,\omega}^\star), (h-h_\omega^\star) - (m-m_P^\star)\right\rangle \\ &- \left\langle \sum_{k=1}^d u_k(j_k - j_{k,\omega}^\star), h - m\right\rangle, \end{aligned}$$

which gives the inclusion

$$\mathcal{F}_{B,n} \subset (\mathcal{U}_g \odot \Delta\mathscr{J}_n) + (\Delta\dot{\mathscr{G}}^\star \odot \Delta\mathcal{V}_n) - (\Delta\mathscr{J}_n \odot \mathcal{W}_n).$$

Applying the same three lemmas together with Assumption G yields an S_1-measurable finite constant $K_{B,n}$ such that, conditionally on S_1,

$$\sup_Q \log N(\varepsilon F_{B,n}, \mathcal{F}_{B,n}, L^2(Q)) \le K_{B,n}\varepsilon^{-p}, \qquad 0 < \varepsilon \le 1.$$

Since $\mathcal{F}_{B,n}^{\mathrm{loc}}(r_h, r_j, r_m) \subset \mathcal{F}_{B,n}$, the same bound holds for the localized class. Finally, adding the zero function and passing to the star hull each change the covering number by at most a constant depending only on p (since $p > 0$), so the normalized class

$$\mathcal{F}_{B,n}^{\mathrm{loc},0} \coloneqq \big\{f/F_{B,n} : f \in \mathcal{F}_{B,n}^{\mathrm{loc}}(r_h, r_j, r_m)\big\} \cup \{0\}$$

satisfies the same polynomial entropy bound, up to increasing $K_{B,n}$ by a constant depending only on p.

Step 6: empirical process bounds and conclusion. By Corollary A.1 applied to $\mathcal{F}_A$ with envelope $F_A = 6(A+B)D$ and entropy constant K_A,

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}_A} |(P_{n,2} - P)f| \,\middle|\, S_1\right] \le \frac{C_p}{\sqrt{n}} F_A \sqrt{1 + K_A} = \frac{6C_p}{\sqrt{n}}(A+B)D\sqrt{1+K_A}.$$

By definition of $\mathfrak{C}_{B,n}(r_h, r_j, r_m)$,

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}_{B,n}^{\mathrm{loc}}(r_h,r_j,r_m)} |(P_{n,2} - P)f| \,\middle|\, S_1\right] = \mathfrak{C}_{B,n}(r_h, r_j, r_m).$$

Taking conditional expectations in Step 3 and combining,

$$\mathbb{E}\left[\sup_{\omega\in\Omega} \|\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega)\|_2 \,\middle|\, S_1\right] \le \frac{6C_p}{\sqrt{n}}(A+B)D\sqrt{1+K_A} + \mathfrak{C}_{B,n}(r_h, r_j, r_m) + r_j(r_h + r_m),$$

which gives the stated bound with $C = 6C_p$ and $K = K_A$. □

A.6 Polynomial-entropy instantiation

Proposition A.1 (Polynomial-entropy instantiation)**.** *Under the conditions of Theorem A.1, suppose the nuisance envelopes and entropy constants remain bounded and $r_h, r_j, r_m = O(n^{-1/(2+p)})$. Then, on $\mathcal{E}_{\mathrm{tr}}(r_h, r_j, r_m)$,*

$$\mathbb{E}\left[\sup_{\omega\in\Omega} \|\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega)\| \,\middle|\, S_1\right] = O(n^{-1/2}). \tag{30}$$

Proof. Conditionally on S_1, the localized class $\mathcal{F}^{\mathrm{loc}}_{B,n}(r_h, r_j, r_m)$ has envelope $F_{B,n}$ and $L^2(P)$ radius at most $\rho_{B,n} = (A + B + \bar{H}_n + \bar{M}_n)r_j + 2D(r_h + r_m)$. By the entropy bound established in Step 5 of the proof of Theorem A.1, the normalized class $\mathcal{F}^{\mathrm{loc},0}_{B,n}$ and its star hull have entropy of order $(1 + K_{B,n})\varepsilon^{-p}$. Conditionally on S_1, set $a_n := (1 + K_{B,n})/n$. Applying the local uniform-entropy inequality of van der Vaart & Wellner (2011, Theorem 2.1) to the signed normalized star hull, as in (42), gives, for $0 < \delta \le 1$,

$$\mathfrak{R}_n\Big(\mathrm{star}(\mathcal{F}^{\mathrm{loc},0}_{B,n}), \delta \,\Big|\, S_1\Big) \le C_p \left\{ \sqrt{a_n}\, \delta^{1-p/2} + a_n \delta^{-p} \right\}.$$

Choose $\delta_{B,n} := c_* a_n^{1/(2+p)} \wedge 1$, where c_* depends only on p. On the untruncated branch,

$$\frac{\sqrt{a_n}\, \delta_{B,n}^{1-p/2}}{\delta_{B,n}^2} = c_*^{-1-p/2}, \qquad \frac{a_n \delta_{B,n}^{-p}}{\delta_{B,n}^2} = c_*^{-2-p}.$$

Taking c_* sufficiently large therefore makes the Rademacher complexity at most $\delta^2_{B,n}$, and Lemma A.5 applies. On the truncated branch, the envelope bound $\mathfrak{C}_{B,n} \le 2F_{B,n}$ is absorbed by the $F_{B,n}\delta^2_{B,n}$ term below; bounded $K_{B,n}$ makes this branch occur only for finitely many n. Thus, after increasing constants,

$$\mathfrak{C}_{B,n}(r_h, r_j, r_m) \le C \left\{ \delta_{B,n}\rho_{B,n} + F_{B,n}\delta^2_{B,n} + F_{B,n} e^{-cn\delta^2_{B,n}} \right\}.$$

Under the assumption that $F_{B,n}$, $K_{B,n}$, $\bar{H}_n$, and $\bar{M}_n$ are $O(1)$ and $r_h, r_j, r_m = O(n^{-1/(2+p)})$, we have $\delta_{B,n} = O(n^{-1/(2+p)})$, $\rho_{B,n} = O(n^{-1/(2+p)})$, and $n\delta^2_{B,n} \asymp n^{p/(2+p)} \to \infty$, so:

- $\delta_{B,n}\rho_{B,n} = O(n^{-2/(2+p)}) = o(n^{-1/2})$ since $p < 2$,
- $F_{B,n}\delta^2_{B,n} = O(n^{-2/(2+p)}) = o(n^{-1/2})$ since $p < 2$,
- $F_{B,n} e^{-cn\delta^2_{B,n}} = o(n^{-1/2})$ since $n\delta^2_{B,n} \to \infty$.

Hence $\mathfrak{C}_{B,n}(r_h, r_j, r_m) = o(n^{-1/2})$. Since the product-bias term satisfies $r_j(r_h + r_m) = O(n^{-2/(2+p)}) = o(n^{-1/2})$, substituting into Theorem A.1 gives

$$\mathbb{E}\left[\sup_{\omega\in\Omega} \|\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega)\| \,\middle|\, S_1 \right] \le \frac{6C_p}{\sqrt{n}}(A + B)D\sqrt{1 + K_A} + o(n^{-1/2}) = O(n^{-1/2}).$$

□

A.7 Unconditional rates under source and entropy conditions

We prove Theorem 3.2. The argument uses the same three-term decomposition as the proof of Theorem A.1, but with the deterministic nuisance classes of Assumption B. Conditioning on the training fold is used only as a device inside the proof; all conclusions concern full expectations. Let S_1 and S_2 be independent folds of n i.i.d. observations $O = (X, Y, Z) \sim P$, and set $N = 2n$. The nuisances are fitted on S_1 and the hypergradient is evaluated on S_2. Write $P_{n,r}$ for the empirical measure on S_r and $Pf = \mathbb{E}_P[f(O)]$. Retain Assumption A: $\omega \mapsto g_\omega(z)$ is twice continuously differentiable for every z and

$$\|Y\| \le A \text{ a.s.}, \quad \sup_{\omega\in\Omega}\|g_\omega(z)\| \le B, \quad \sup_{\omega\in\Omega}\|\dot{g}_\omega(z)\|_{2,d} \le D, \quad \sup_{\omega\in\Omega}\|D^2_\omega g_\omega(z)\|_{\mathrm{op}} \le L \quad \forall z \in \mathcal{Z},$$

where $\dot{g}_\omega = (\partial_{\omega_1} g_\omega, \ldots, \partial_{\omega_d} g_\omega) = D_\omega g_\omega$ and $\|a\|^2_{2,d} = \sum_{k=1}^d \|a_k\|^2$ for $a \in (\mathbb{R}^q)^d$, so that $\|\cdot\|_{2,d}$ is the Frobenius norm of the stacked derivative. The nuisance targets are

$$h^\star_{\omega,P}(X) = \mathbb{E}_P[g_\omega(Z) \mid X], \qquad j^\star_{\omega,P}(X) = \mathbb{E}_P[\dot{g}_\omega(Z) \mid X], \qquad m^\star_P(X) = \mathbb{E}_P[Y \mid X].$$

The feasible set $\Omega \subset \mathbb{R}^d$ is nonempty, compact, and convex, and $\mathcal{F}(\omega) = \frac{1}{2}P[\|Y - h^\star_{\omega,P}(X)\|^2]$. In this subsection, write $L_g := D + L$ for the uniform Lipschitz constant on Ω supplied by Assumption A for the pair $(g_\omega, \dot{g}_\omega)$, using the sum of the two norms. All displayed suprema and

empirical-process quantities are assumed measurable. Constants denoted by $C, c > 0$ may change from line to line.

For $\ell \in \{h, j, m\}$, let $V_h = V_m = \mathbb{R}^q$, $V_j = (\mathbb{R}^q)^d$, $\Omega_h = \Omega_j = \Omega$, and $\Omega_m = \{0\}$. Use the Euclidean norm on V_h, V_m and the block norm on V_j. Set

$$(b_h, b_j, b_m) = (B, D, A),$$
$$(T^h_\omega, T^j_\omega, T^m_0) = (g_\omega(Z), \dot{g}_\omega(Z), Y),$$
$$(f^\star_{h,\omega}, f^\star_{j,\omega}, f^\star_{m,0}) = (h^\star_{\omega,P}, j^\star_{\omega,P}, m^\star_P).$$

For classes $\mathcal{F}_\ell(t)$ of V_ℓ-valued functions on $\mathcal{X}$, indexed by $t \in (0, 1]$, define the loss and the population and empirical risks

$$\ell^\ell_\omega(O; f) = \tfrac{1}{2}\|f(X) - T^\ell_\omega(O)\|^2, \quad \widehat{R}^\ell_\omega(f) = P_{n,1}\ell^\ell_\omega(\cdot; f), \quad R^\ell_\omega(f) = P\ell^\ell_\omega(\cdot; f). \tag{31}$$

Let U be the auxiliary fitting randomness of the estimator setup in Section 3.2, independent of all data, and write $\mathscr{T}_1 = \sigma(S_1, U)$, omitting U for deterministic fitting. As in that setup, the estimates $\widehat{f}_{\ell,\omega} \in \mathcal{F}_\ell(t_n)$ are $\mathscr{T}_1$-measurable. Put $(\widehat{f}_{h,\omega}, \widehat{f}_{j,\omega}, \widehat{f}_{m,0}) = (\widehat{h}_\omega, \widehat{j}_\omega, \widehat{m})$. The search classes are $\mathcal{H}_n = \mathcal{F}_h(t_n)$, $\mathcal{J}_n = \mathcal{F}_j(t_n)$, and $\mathcal{M}_n = \mathcal{F}_m(t_n)$. Define

$$\delta_n := (nt_n^\gamma)^{-1/(2+p)}, \qquad v_n := t_n^{2\nu} + \delta_n^2, \qquad \varepsilon_n := \max_{\ell \in \{h,j,m\}} \varepsilon_{\ell,n}. \tag{32}$$

Recall the orthogonal score $\varphi_\omega = (\varphi_{1,\omega}, \ldots, \varphi_{d,\omega})^\top$, where

$$\varphi_{k,\omega}(O; h, j, m) = \langle g_\omega(Z) - Y, j_k(X)\rangle + \langle \partial_{\omega_k} g_\omega(Z) - j_k(X), h(X) - m(X)\rangle. \tag{33}$$

Set $\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) = P_{n,2}\varphi_\omega(\cdot; \widehat{h}_\omega, \widehat{j}_\omega, \widehat{m})$ and

$$R_{\ell,n} := \sup_{\omega \in \Omega_\ell} \|\widehat{f}_{\ell,\omega} - f^\star_{\ell,\omega}\|_{L^2(P_X; V_\ell)}, \qquad \Delta_{n,2} := \sup_{\omega \in \Omega} \|\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega)\|. \tag{34}$$

Since the classes and the oracle paths are bounded by b_ℓ, each $R_{\ell,n} \le 2b_\ell$ is finite, and we write $\rho_{\ell,n} := (\mathbb{E}[R^2_{\ell,n}])^{1/2}$.

Proof of Theorem 3.2. The first bound, (16), is (35) in part (i) of Proposition A.2, which uses only Assumptions A and B and the estimator setup. The nuisance second-moment bound is (36) in part (ii), the first place where Assumptions C and D are used. Under (38), (39) gives $\rho_{\ell,n} = O(n^{-\nu/b})$, $\delta_n = n^{-\nu/b}$, and (17). □

Proposition A.2 (Unconditional source–capacity rates)**.** *(i) Under Assumptions A and B and the estimator setup above, there exist finite constants C_0, depending only on A, B, D, L_g, d and $\mathrm{diam}(\Omega)$, and C, depending in addition on p, γ and $(K_\ell)_{\ell \in \{h,j,m\}}$, such that, for every $n \ge 1$ and every deterministic $t_n \in (0, 1]$,*

$$\mathbb{E}\left[\Delta_{n,2}\right] \le C_0 n^{-1/2} + C\Big\{\delta_n \sum_{\ell \in \{h,j,m\}} \rho_{\ell,n} + \delta_n^2 + \rho_{j,n}(\rho_{h,n} + \rho_{m,n})\Big\}. \tag{35}$$

(ii) If, in addition, Assumptions C and D hold, there exist finite constants C_h, C_j, C_m, C_1, depending only on $A, B, D, L_g, p, \nu, \gamma, d, q$, $\mathrm{diam}(\Omega)$, and $(K_\ell, L_\ell)_{\ell \in \{h,j,m\}}$, such that, for every $n \ge 1$,

$$\mathbb{E}[R^2_{\ell,n}] \le C_\ell \left\{t_n^{2\nu} + (nt_n^\gamma)^{-2/(2+p)} + \varepsilon_{\ell,n}\right\}, \qquad \ell \in \{h, j, m\}, \tag{36}$$

and

$$\mathbb{E}\left[\Delta_{n,2}\right] \le C_0 n^{-1/2} + C_1 \left\{t_n^{2\nu} + (nt_n^\gamma)^{-2/(2+p)} + \varepsilon_n\right\}. \tag{37}$$

The expectation is over both folds and all fitting randomness. The same gradient bounds hold for the two-fold cross-fitted estimator in (82)*, under the same conditions in both training directions. In particular, let*

$$b := \nu(2 + p) + \gamma, \qquad t_n = n^{-1/b}, \qquad \varepsilon_{\ell,n} \le c_\ell n^{-2\nu/b} \tag{38}$$

for fixed $c_\ell < \infty$. Then, with constants also allowed to depend on $(c_\ell)_\ell$,

$$\mathbb{E}[R_{\ell,n}^2] \le C'_\ell n^{-2\nu/b}, \qquad \rho_{\ell,n} = O(n^{-\nu/b}), \qquad \delta_n = n^{-\nu/b}, \\ \mathbb{E}[\Delta_{n,2}] \le C_0 n^{-1/2} + C_2 n^{-2\nu/b}. \tag{39}$$

Thus the expected uniform gradient error is $O(n^{-1/2})$ if $\nu(2-p) \ge \gamma$. The second term in the gradient bound is $o(n^{-1/2})$ if $\nu(2-p) > \gamma$.

To prove Proposition A.2, we use the following localization bound for the training losses and the evaluation score differences. Here P_n is the empirical measure of an independent sample of size n from P.

Lemma A.7 (Localization under the entropy scale)**.** *Let $\mathcal{A}_n$ be a pointwise measurable class of real-valued functions bounded by one, with*

$$\sup_Q \log N(\epsilon, \mathcal{A}_n, L^2(Q)) \le K t_n^{-\gamma} \epsilon^{-p}, \qquad 0 < \epsilon \le 1,$$

where $0 \le K < \infty$, $p \in (0,2)$, $\gamma \ge 0$, and $t_n \in (0,1]$. Set $\delta_n = (n t_n^\gamma)^{-1/(2+p)}$. There exist constants $C, c, c_0 > 0$, depending only on p, K, such that, whenever $\delta_n \le c_0$, with probability at least $1 - Ce^{-cn\delta_n^2}$,

$$|(P_n - P)a| \le C\left\{\delta_n \|a\|_{L^2(P)} + \delta_n^2\right\} \quad \textit{simultaneously for all } a \in \mathcal{A}_n. \tag{40}$$

Moreover, for every $n \ge 1$ and $r \ge 0$,

$$\mathbb{E}\left[\sup_{\substack{a \in \mathcal{A}_n \cup \{0\} \\ \|a\|_{L^2(P)} \le r}} |(P_n - P)a|\right] \le C\{\delta_n r + \delta_n^2\}. \tag{41}$$

Proof. Let $\mathcal{A}_n^\circ = \mathrm{star}(\mathcal{A}_n \cup \{0\})$, with the star hull taken about zero. A grid for the multiplier in $[0,1]$ shows that

$$\sup_Q \log N(\epsilon, \mathcal{A}_n^\circ, L^2(Q)) \le \kappa_n \epsilon^{-p}, \qquad \kappa_n = C_p(1+K) t_n^{-\gamma}.$$

Indeed, a product cover at radii $\epsilon/2$ for the function and the multiplier gives an additional term of order $\log(1/\epsilon)$, which is absorbed by $C_p \epsilon^{-p}$. For $0 < r \le 1$, write

$$\mathfrak{R}_n(\mathcal{A}_n^\circ, r) := \mathbb{E}_{\xi, O_{1:n}}\left[\sup_{\substack{a \in \mathcal{A}_n^\circ \\ \|a\|_{L^2(P)} \le r}} \left|\frac{1}{n}\sum_{i=1}^n \xi_i a(O_i)\right|\right],$$

where ξ_i are independent Rademacher variables, independent of the observations. The local uniform-entropy inequality of van der Vaart & Wellner (2011, Theorem 2.1), applied to the mean-zero class $(\xi, O) \mapsto \xi a(O)$, yields

$$\mathfrak{R}_n(\mathcal{A}_n^\circ, r) \le C_p\left\{\sqrt{\frac{1+\kappa_n}{n}}\, r^{1-p/2} + \frac{1+\kappa_n}{n}\, r^{-p}\right\}. \tag{42}$$

Here the signed class has the same L^2 covering bounds, and its entropy integral up to radius r is at most $C_p\sqrt{1+\kappa_n}\, r^{1-p/2}$. Since $(1+\kappa_n)/n \le C\delta_n^{2+p}$, substitution of $r = c_\star \delta_n$ in Equation (42) gives

$$\frac{\mathfrak{R}_n(\mathcal{A}_n^\circ, c_\star\delta_n)}{(c_\star\delta_n)^2} \le C\left\{c_\star^{-1-p/2} + c_\star^{-2-p}\right\}.$$

Choose $c_\star$ so that the right-hand side is at most one, and choose $c_0 < 1/c_\star$. The critical-radius condition therefore holds at $c_\star\delta_n$ whenever $\delta_n \le c_0$. Apply Foster & Syrgkanis (2023, Lemma 14) to $\mathcal{A}_n \cup \{0\}$ with identity loss and comparator zero. Its linear-loss version requires no additional logarithmic lower bound on the radius. This gives (40), after absorbing $c_\star$ into the constants.

On this event the localized supremum in (41) is at most $C(\delta_n r + \delta_n^2)$; on its complement it is at most two. Also $t_n \le 1$ and $\gamma \ge 0$ imply

$$\delta_n \ge n^{-1/(2+p)}, \qquad e^{-cn\delta_n^2} \le e^{-cn^{p/(2+p)}} \le Cn^{-2/(2+p)} \le C\delta_n^2. \tag{43}$$

Taking expectations proves (41) for $\delta_n \le c_0$. If $\delta_n > c_0$, the bound follows from the envelope after increasing C, since $2 \le 2c_0^{-2}\delta_n^2$. □

Proof of Proposition A.2. The proof proceeds in eight steps. Steps 1–5 prove part (i). They use only Assumptions A and B, the class membership and measurability of the fitted nuisances, and the independence of S_2 and $\mathscr{T}_1$. Steps 6–8 prove part (ii) and the tuned rates; Assumptions C and D enter only in Step 7.

Step 1: population paths and entropy transfer. Work on a common measurable set of full P-probability on which Assumption A holds. Conditional Jensen gives

$$\|h^\star_{\omega,P}(X)\| \le B, \qquad \|j^\star_{\omega,P}(X)\|_{2,d} \le D, \qquad \|m^\star_P(X)\| \le A.$$

Choose the conditional means using a regular conditional law. Conditional expectation preserves the supremum-norm Lipschitz bound implied by Assumption A, so the oracle paths satisfy the same bound on a common set of full P_X-probability. Dominated differentiation also gives $\partial_{\omega_k} h^\star_{\omega,P} = j^\star_{k,\omega,P}$ and

$$\partial_{\omega_k}\mathcal{F}(\omega) = P\langle h^\star_{\omega,P}(X) - Y, j^\star_{k,\omega,P}(X)\rangle.$$

For each of the classes $\mathcal{G}_\Omega$, $\dot{\mathcal{G}}_\Omega$, $\mathcal{H}^\star$, and $\mathcal{J}^\star$, a Euclidean net of Ω induces a supremum-norm cover. In particular,

$$\log N(\eta, \mathcal{C}, \|\cdot\|_\infty) \le d \log\left(1 + \frac{CL_g \operatorname{diam}(\Omega)}{\eta}\right), \qquad \eta > 0. \tag{44}$$

The block norm is used for the derivative classes. Since $p > 0$, this logarithmic bound is dominated by $C\eta^{-p}$ on bounded intervals of positive radii. Thus Assumption F holds with fixed constants, without any inclusion of $\mathcal{H}^\star$ or $\mathcal{J}^\star$ in the fitted classes. The class membership of the estimator setup and Assumption B give Assumptions E and G with deterministic envelopes (B, D, A) and entropy constants $(K_\ell t_n^{-\gamma})_\ell$.

Step 2: three-term decomposition. Write $\varphi^\star_\omega := \varphi_\omega(\cdot; h^\star_{\omega,P}, j^\star_{\omega,P}, m^\star_P)$ and $\widehat{\varphi}_\omega := \varphi_\omega(\cdot; \widehat{h}_\omega, \widehat{j}_\omega, \widehat{m})$, and let $P\widehat{\varphi}_\omega$ denote the integral over a new observation with the fitted functions held fixed. By Theorem 3.1 at the true nuisances, $P\varphi^\star_\omega = \nabla\mathcal{F}(\omega)$. Since $\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) = P_{n,2}\widehat{\varphi}_\omega$,

$$\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega) = (P_{n,2} - P)\varphi^\star_\omega + (P_{n,2} - P)(\widehat{\varphi}_\omega - \varphi^\star_\omega) + \big\{P\widehat{\varphi}_\omega - \nabla\mathcal{F}(\omega)\big\}. \tag{45}$$

Taking norms and suprema over $\omega \in \Omega$ gives $\Delta_{n,2} \le \mathrm{I}_n + \mathrm{II}_n + \mathrm{III}_n$, where

$$\mathrm{I}_n := \sup_{\omega\in\Omega}\|(P_{n,2}-P)\varphi^\star_\omega\|, \quad \mathrm{II}_n := \sup_{\omega\in\Omega}\|(P_{n,2}-P)(\widehat{\varphi}_\omega-\varphi^\star_\omega)\|, \quad \mathrm{III}_n := \sup_{\omega\in\Omega}\|P\widehat{\varphi}_\omega-\nabla\mathcal{F}(\omega)\|.$$

Steps 3–5 bound the expectations of these three terms separately.

Step 3: oracle empirical-process term (Assumption A). Let $\mathcal{F}_A := \{u^\top(\varphi^\star_\omega - \nabla\mathcal{F}(\omega)) : \omega \in \Omega,\ u \in \mathbb{S}^{d-1}\}$. By the dual representation of the Euclidean norm, $\mathrm{I}_n = \sup_{f\in\mathcal{F}_A}|(P_{n,2}-P)f|$. This class does not depend on the training fold, so no conditioning is needed. By Step 1 and the argument of Step 4 in the proof of Theorem A.1, $F_A := 6(A+B)D$ is a constant envelope for $\mathcal{F}_A$. The oracle score is formed from g_ω, $\dot{g}_\omega$, $h^\star_{\omega,P}$, $j^\star_{\omega,P}$, Y, and the single function $m^\star_P$ by bounded sums and inner products. The constants $u^\top\nabla\mathcal{F}(\omega)$ and the sphere contribute entropy at most $d\log(1 + C/\epsilon)$. Therefore Lemmas A.2 to A.4, together with (44), give

$$\sup_Q \log N(\epsilon F_A, \mathcal{F}_A, L^2(Q)) \le K_A\epsilon^{-1}, \qquad 0 < \epsilon \le 1,$$

where K_A depends only on A, B, D, L_g, d, and $\operatorname{diam}(\Omega)$. Here the logarithmic entropy is dominated by the fixed exponent one rather than by p. Corollary A.1, applied with exponent one to the i.i.d. sample S_2, gives

$$\mathbb{E}[\mathrm{I}_n] \le C_0 n^{-1/2}, \qquad C_0 := 6C_{(1)}(A+B)D\sqrt{1+K_A}, \tag{46}$$

where $C_{(1)}$ is the universal constant of Corollary A.1 for exponent one. Thus C_0 depends only on A, B, D, L_g, d, and $\mathrm{diam}(\Omega)$. It does not depend on n, t_n, or the nuisance classes.

Step 4: nuisance-dependent empirical-process term (Assumptions A and B). Define the deterministic class

$$\mathcal{V}_n := \big\{u^\top\{\varphi_\omega(\cdot; h, j, m) - \varphi^\star_\omega\} : \omega \in \Omega,\ u \in \mathbb{S}^{d-1},\ h \in \mathcal{H}_n,\ j \in \mathcal{J}_n,\ m \in \mathcal{M}_n\big\} \cup \{0\}.$$

By the envelope (12) and Step 1, every candidate score and every oracle score has norm at most $3(A+B)D$, so $M_\varphi := 6(A+B)D$ is an envelope for $\mathcal{V}_n$. The bounded sum and product bounds of Lemmas A.2 to A.4, including scalarization over $u \in \mathbb{S}^{d-1}$, together with the entropy bound (13), give

$$\sup_Q \log N(\epsilon M_\varphi, \mathcal{V}_n, L^2(Q)) \le C(1 + K_h + K_j + K_m) t_n^{-\gamma} \epsilon^{-p}, \qquad 0 < \epsilon \le 1. \tag{47}$$

In this calculation the sphere contributes only $d \log(1 + C/\epsilon)$, and the oracle paths are controlled by Equation (44).

For a deterministic radius triple $r = (r_h, r_j, r_m)$, restrict $\mathcal{V}_n$ to candidates with L^2 errors at most r_h, r_j, r_m, respectively, and retain zero. Call the resulting class $\mathcal{V}_n(r)$. With $\Delta h = h - h^\star_{\omega,P}$, $\Delta j = j - j^\star_{\omega,P}$, and $\Delta m = m - m^\star_P$, the exact expansion is

$$\begin{aligned}\varphi_{k,\omega}(O; h, j, m) - \varphi^\star_{k,\omega}(O) = {}& \langle g_\omega(Z) - Y, \Delta j_k(X)\rangle \\ &+ \langle \partial_{\omega_k} g_\omega(Z) - j^\star_{k,\omega,P}(X), \Delta h(X) - \Delta m(X)\rangle \\ &- \langle \Delta j_k(X), h(X) - m(X)\rangle.\end{aligned}$$

Hence every $v \in \mathcal{V}_n(r)$ satisfies

$$\|v\|_{L^2(P)} \le 2(A+B) r_j + 2D(r_h + r_m) \le C(r_h + r_j + r_m). \tag{48}$$

Applying Equation (41) to $\mathcal{V}_n / M_\varphi$, with the independent sample S_2, yields, for every $n \ge 1$,

$$\mathbb{E}\Big[\sup_{v \in \mathcal{V}_n(r)} |(P_{n,2} - P)v|\Big] \le C\{\delta_n(r_h + r_j + r_m) + \delta_n^2\}. \tag{49}$$

The class $\mathcal{V}_n(r)$ contains the localized class of Theorem A.1, so (49) also bounds $\mathfrak{C}_{B,n}(r_h, r_j, r_m)$.

We now pass to the realized nuisance errors. Conditioning on $\mathscr{T}_1$ fixes all fitted nuisance functions, while S_2 remains an i.i.d. sample from P because it is independent of $\mathscr{T}_1$. Let $\mathcal{E}_{\mathrm{tr}}(r) := \{R_{h,n} \le r_h,\ R_{j,n} \le r_j,\ R_{m,n} \le r_m\}$, a $\mathscr{T}_1$-measurable event. On $\mathcal{E}_{\mathrm{tr}}(r)$, class membership implies that $u^\top(\widehat{\varphi}_\omega - \varphi^\star_\omega) \in \mathcal{V}_n(r)$ for every $\omega \in \Omega$ and $u \in \mathbb{S}^{d-1}$, so $\mathrm{II}_n \le \sup_{v \in \mathcal{V}_n(r)} |(P_{n,2} - P)v|$. Since $\mathcal{V}_n(r)$ is deterministic, (49) and independence give

$$\mathbb{E}[\mathrm{II}_n \mid \mathscr{T}_1] \le C\{\delta_n(r_h + r_j + r_m) + \delta_n^2\} \quad \text{on } \mathcal{E}_{\mathrm{tr}}(r). \tag{50}$$

Apply Equation (50) simultaneously to all positive rational radius triples, outside a common null set. For each training realization, the errors are finite. Take rational triples decreasing to $(R_{h,n}, R_{j,n}, R_{m,n})$ from above. Each corresponding training event holds, and the right-hand side is continuous in r. Thus

$$\mathbb{E}[\mathrm{II}_n \mid \mathscr{T}_1] \le C\{\delta_n(R_{h,n} + R_{j,n} + R_{m,n}) + \delta_n^2\} \qquad \text{almost surely.} \tag{51}$$

Taking expectations and using $\mathbb{E}[R_{\ell,n}] \le \rho_{\ell,n}$ gives

$$\mathbb{E}[\mathrm{II}_n] \le C\Big\{\delta_n \sum_{\ell \in \{h,j,m\}} \rho_{\ell,n} + \delta_n^2\Big\}. \tag{52}$$

This step uses the bounds of Assumption A and the envelope and entropy conditions of Assumption B. It uses neither approximation nor fitting optimality.

Step 5: population bias (Theorem 3.1). For fixed ω, apply Theorem 3.1 coordinatewise to the fitted candidates $(\widehat{h}_\omega, \widehat{j}_{k,\omega}, \widehat{m})$, which are fixed given $\mathscr{T}_1$. Cauchy–Schwarz in $L^2(P_X)$ and summation over k then give

$$\begin{aligned}\|P\widehat{\varphi}_\omega - \nabla\mathcal{F}(\omega)\| &= \Big(\sum_{k=1}^d \langle \widehat{j}_{k,\omega} - j^\star_{k,\omega,P}, (\widehat{m} - m^\star_P) - (\widehat{h}_\omega - h^\star_{\omega,P})\rangle^2_{L^2(P_X)}\Big)^{1/2} \\ &\le \|\widehat{j}_\omega - j^\star_{\omega,P}\|_{L^2(P_X)^d}\big(\|\widehat{h}_\omega - h^\star_{\omega,P}\|_{L^2(P_X)} + \|\widehat{m} - m^\star_P\|_{L^2(P_X)}\big).\end{aligned} \tag{53}$$

Taking the supremum over ω gives $\mathrm{III}_n \le R_{j,n}(R_{h,n} + R_{m,n})$ pointwise. By Cauchy–Schwarz for the expectation,

$$\mathbb{E}[\mathrm{III}_n] \le \mathbb{E}[R_{j,n}R_{h,n}] + \mathbb{E}[R_{j,n}R_{m,n}] \le \rho_{j,n}(\rho_{h,n} + \rho_{m,n}). \tag{54}$$

No independence among the nuisance errors is used. This term is controlled by orthogonality alone, separately from the localization argument of Step 4.

Combining (46), (52), and (54) with $\Delta_{n,2} \le \mathrm{I}_n + \mathrm{II}_n + \mathrm{III}_n$, and increasing C to at least one, proves (35). This completes part (i).

Step 6: simultaneous relative deviations of the training losses (Assumption B). For each ℓ, define

$$q^{\ell}_{\omega,f} := \ell^{\ell}_{\omega}(\cdot; f) - \ell^{\ell}_{\omega}(\cdot; f^{\star}_{\ell,\omega}), \qquad \mathcal{Q}_{\ell,n} := \{q^{\ell}_{\omega,f} : \omega \in \Omega_\ell, f \in \mathcal{F}_\ell(t_n)\} \cup \{0\}.$$

The conditional-mean identity for squared loss gives

$$Pq^{\ell}_{\omega,f} = \tfrac{1}{2}\|f - f^{\star}_{\ell,\omega}\|^2_{L^2(P_X;V_\ell)}, \qquad \|q^{\ell}_{\omega,f}\|_{L^2(P)} \le 2b_\ell \|f - f^{\star}_{\ell,\omega}\|_{L^2(P_X;V_\ell)}. \tag{55}$$

The second inequality follows pointwise from

$$q^{\ell}_{\omega,f} = \tfrac{1}{2}\langle f(X) - f^{\star}_{\ell,\omega}(X), f(X) + f^{\star}_{\ell,\omega}(X) - 2T^{\ell}_{\omega}(O)\rangle.$$

Consequently, $M_\ell := 4b^2_\ell$ is a fixed envelope for $\mathcal{Q}_{\ell,n}$. Lemmas A.2 and A.3, applied to this bounded-product representation, together with Equations (13) and (44), imply

$$\sup_{Q} \log N(\epsilon M_\ell, \mathcal{Q}_{\ell,n}, L^2(Q)) \le C(1 + K_\ell) t_n^{-\gamma} \epsilon^{-p}, \qquad 0 < \epsilon \le 1. \tag{56}$$

Here Q ranges over finitely supported laws of O; the covers for the candidate functions use its X-marginal. Applying Lemma A.7 to $\mathcal{Q}_{\ell,n}/M_\ell$ and using Equation (55) shows that, whenever $\delta_n \le c_0$, an event $\mathcal{E}_{\ell,n}$ of probability at least $1 - Ce^{-cn\delta_n^2}$ satisfies

$$|(P_{n,1} - P)q^{\ell}_{\omega,f}| \le C\left\{\delta_n \|f - f^{\star}_{\ell,\omega}\|_{L^2(P_X;V_\ell)} + \delta_n^2\right\} \tag{57}$$

simultaneously over $\omega \in \Omega_\ell$ and $f \in \mathcal{F}_\ell(t_n)$. The event depends only on S_1.

Step 7: uniform nuisance moments (Assumptions C and D). Fix ℓ and work on $\mathcal{E}_{\ell,n}$ and the almost-sure fitting event of Assumption D. For any $\omega \in \Omega_\ell$ and comparator $f \in \mathcal{F}_\ell(t_n)$, write

$$r_\omega = \|\widehat{f}_{\ell,\omega} - f^{\star}_{\ell,\omega}\|_{L^2(P_X;V_\ell)}, \qquad a_\omega(f) = \|f - f^{\star}_{\ell,\omega}\|_{L^2(P_X;V_\ell)}.$$

Equations (15), (55), and (57) give

$$\begin{aligned}
\tfrac{1}{2}r^2_\omega &= Pq^{\ell}_{\omega,\widehat{f}_{\ell,\omega}} \\
&\le P_{n,1}q^{\ell}_{\omega,\widehat{f}_{\ell,\omega}} + C\delta_n r_\omega + C\delta_n^2 \\
&\le P_{n,1}q^{\ell}_{\omega,f} + \varepsilon_{\ell,n} + C\delta_n r_\omega + C\delta_n^2 \\
&\le \tfrac{1}{2}a_\omega(f)^2 + \varepsilon_{\ell,n} + C\delta_n\{r_\omega + a_\omega(f)\} + C\delta_n^2.
\end{aligned}$$

Using $C\delta_n r_\omega \le \frac{1}{4}r^2_\omega + C'\delta_n^2$ and $C\delta_n a_\omega(f) \le \frac{1}{2}a_\omega(f)^2 + C'\delta_n^2$, then taking the infimum over f and the supremum over ω, yields

$$R^2_{\ell,n} \le C\left\{L^2_\ell t^{2\nu}_n + \delta_n^2 + \varepsilon_{\ell,n}\right\} \quad \text{on } \mathcal{E}_{\ell,n}. \tag{58}$$

This uses Equation (14) directly. No population minimizer in $\mathcal{F}_\ell(t_n)$ is needed: take arbitrarily accurate comparators before passing to the infimum. On the complement, $R_{\ell,n} \le 2b_\ell$ by the envelope and conditional Jensen. Taking expectations in Equation (58) therefore gives

$$\mathbb{E}[R^2_{\ell,n}] \le C\left\{t^{2\nu}_n + \delta_n^2 + \varepsilon_{\ell,n} + e^{-cn\delta_n^2}\right\} \le C\{v_n + \varepsilon_{\ell,n}\},$$

where the last inequality uses Equation (43). If $\delta_n > c_0$, the same conclusion follows from $R_{\ell,n} \le 2b_\ell$ after increasing C. This proves Equation (36) for every n. Since the deviation event is simultaneous over the fitted classes, the conclusion includes all auxiliary fitting randomness.

Set $s_{\ell,n} \coloneqq (v_n+\varepsilon_{\ell,n})^{1/2}$. Then Equation (36) gives $\rho_{\ell,n} \le C s_{\ell,n}$. Substituting into (35), and using $\delta_n \le \sqrt{v_n+\varepsilon_n}$ and $s_{\ell,n} \le \sqrt{v_n+\varepsilon_n}$, bounds the braces by $C(v_n+\varepsilon_n)$. This proves Equation (37) with expectation over all randomness.

For two-fold cross-fitting, write $\widehat{\nabla\mathcal{F}}_2^{\mathrm{OBi}} = \widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}$ for the estimate trained on S_1, and let $\widehat{\nabla\mathcal{F}}_1^{\mathrm{OBi}}$ denote the estimate trained on S_2. Pointwise in both samples,

$$\sup_{\omega\in\Omega}\left\|\frac{\widehat{\nabla\mathcal{F}}_1^{\mathrm{OBi}}(\omega)+\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)}{2}-\nabla\mathcal{F}(\omega)\right\| \le \frac{1}{2}\sum_{r=1}^{2}\sup_{\omega\in\Omega}\|\widehat{\nabla\mathcal{F}}_r^{\mathrm{OBi}}(\omega)-\nabla\mathcal{F}(\omega)\|.$$

Apply Equations (35) and (37) separately in each training direction and take expectations. Independence of the two resulting estimates is not required.

Step 8: balancing approximation and estimation. Under Equation (38),

$$t_n^{2\nu}=n^{-2\nu/b}, \qquad (nt_n^{\gamma})^{-2/(2+p)}=n^{-\frac{2}{2+p}(1-\gamma/b)}=n^{-2\nu/b}, \qquad \varepsilon_n \le \Big(\max_{\ell} c_\ell\Big)n^{-2\nu/b}.$$

Since $1-\gamma/b=\nu(2+p)/b$, the middle identity is $\delta_n^2=n^{-2\nu/b}$, that is, $\delta_n=n^{-\nu/b}$. Substituting into Equations (36) and (37) gives Equation (39), including $\rho_{\ell,n}\le (C'_\ell)^{1/2}n^{-\nu/b}$. Finally,

$$\frac{2\nu}{\nu(2+p)+\gamma}\ge\frac{1}{2} \quad\Longleftrightarrow\quad \nu(2-p)\ge\gamma,$$

with strict inequalities equivalent as well. This proves the stated root-n conclusions. □

A.8 Minimax optimality of OBiGrad

Section 3 established a finite-sample bound on $\mathfrak{s} = \mathbb{E}[\sup_{\omega\in\Omega}\|\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)-\nabla\mathcal{F}(\omega)\|]$ for the OBiGrad estimator and showed that this error attains the parametric rate $O(n^{-1/2})$ whenever the entropy–approximation exponents of the nuisance classes in Assumptions B and C satisfy a suitable condition. We are now interested in establishing a matching lower bound. To avoid dealing with the degenerate case when the population hypergradient $\nabla\mathcal{F}(\omega)$ is identically 0, we make the following non-degeneracy assumption on the Jacobian $j^\star_{\omega,P}(x)=(j^\star_{1,\omega,P}(x),\dots,j^\star_{d,\omega,P}(x))$.

Assumption H (Gradient non-degeneracy)**.** There exists a parameter $\omega_0\in\Omega$ for which the matrix $M_0 \coloneqq \mathbb{E}\big[j^\star_{\omega_0,P}(X)^\top j^\star_{\omega_0,P}(X)\big]$ has strictly positive largest eigenvalue, i.e. $\lambda_0 \coloneqq \lambda_{\max}(M_0)>0$.

Since P varies in what follows, we write $\mathcal{F}_P$ and $\nabla\mathcal{F}_P$ for the value function (FBO) and its hypergradient formed under the law P. The lower bound is to be compared with the upper bound of Theorem 3.2, so the two must range over the same laws. Theorem 3.2 operates at the single approximation scale $t_n=n^{-1/b}$, so it is the approximation condition (14) at that scale that must hold. Writing $\mathcal{P}_Y(P_{XZ})$ for the set of laws P of (X,Y,Z) with $(X,Z)\sim P_{XZ}$ and $\|Y\|\le A$ almost surely, we therefore restrict attention to

$$\mathcal{P}_Y^{\mathrm{reg}}(P_{XZ};n) \coloneqq \Big\{P\in\mathcal{P}_Y(P_{XZ}) : \inf_{f\in\mathcal{F}_m(t_n)}\|f-m_P^\star\|^2_{L^2(P_X)}\le L_m^2\,t_n^{2\nu}\Big\}. \tag{59}$$

Condition (14) is imposed for each $\ell\in\{h,j,m\}$. Its instances for $\ell\in\{h,j\}$ concern $h^\star_{\omega,P}$ and $j^\star_{\omega,P}$, which depend on P only through P_{XZ}. Since P_{XZ} is held fixed, these hold for every member of $\mathcal{P}_Y(P_{XZ})$ or for none, and we assume they hold for the given P_{XZ}. The outcome regression $m_P^\star$ is therefore the only nuisance that varies over the class, and (59) is the one remaining instance, $\ell=m$. Consequently Theorem 3.2 applies to every $P\in\mathcal{P}_Y^{\mathrm{reg}}(P_{XZ};n)$.

Proposition A.3 (Lower bound)**.** *Fix a law P_{XZ} over (X,Z) on $\mathcal{X}\times\mathcal{Z}$. Let g be a structural map such that Assumption A holds for g and Assumption H holds at some $\omega_0\in\Omega$ for the pair (g,P_{XZ}), and assume $0\in\mathcal{F}_m(t)$ for every $t\in(0,1]$ and $L_m\ge AD/\sqrt{\lambda_0}$. Then, for any integer $n\ge D^2/\lambda_0$, it holds that:*

$$\inf_{\widehat\Psi}\sup_{P\in\mathcal{P}_Y^{\mathrm{reg}}(P_{XZ};n)}\mathbb{E}_P\big[\|\widehat\Psi-\nabla\mathcal{F}_P(\omega_0)\|\big] \;\ge\; c_0\,A\,\sqrt{\lambda_0}\,n^{-1/2}, \qquad c_0\coloneqq\frac{1-\sqrt{1/2}}{4},$$

where the infimum is over all estimators $\widehat\Psi$ measurable with respect to n i.i.d. draws from P. Since $\lambda_0\le D^2$, the threshold satisfies $D^2/\lambda_0\ge 1$.

Proposition A.3 shows that there exists a worst-case problem for which no estimator of the hypergradient at a non-degenerate point ω_0, whose existence is guaranteed by Assumption H, can attain a rate faster than the parametric rate. The proof reduces the problem to Bernoulli mean estimation through a one-dimensional submodel and applies a standard Le Cam two-point argument (Tsybakov, 2009, Chapter 2). The same lower bound holds for the uniform risk $\sup_{\omega\in\Omega} \|\widehat{\Psi}_\omega - \nabla\mathcal{F}_P(\omega)\|$ of any estimator of the full target $\nabla\mathcal{F}_P(\omega)$, since this uniform risk dominates the fixed-ω_0 risk considered in Proposition A.3. Combined with Theorem 3.2, which yields an upper bound $O(n^{-1/2})$ on the gradient estimation error whenever $\nu(2-p) \geq \gamma$, Proposition A.3 shows that this rate matches the lower bound. Thus, in this regime, OBiGrad is minimax rate-optimal for hypergradient estimation over $\mathcal{P}_Y^{\mathrm{reg}}(P_{XZ}; n)$.

Proof of Proposition A.3. Step 1 (a perturbation direction and a family of laws P_b). Let $u^\star \in \mathbb{R}^d$ be a unit eigenvector of M_0 associated with λ_0, and set

$$\zeta(x) \coloneqq j^\star_{\omega_0,P_{XZ}}(x)\, u^\star \in \mathbb{R}^q, \qquad \varrho(x) \coloneqq \|\zeta(x)\|, \qquad e(x) \coloneqq \begin{cases} \varrho(x)^{-1}\zeta(x), & \zeta(x) \neq 0, \\ e_1, & \text{otherwise}, \end{cases} \tag{60}$$

so that $\langle e(x), \zeta(x)\rangle = \varrho(x)$ and $\|e(x)\| = 1$ for every x; all three are measurable functions of x alone. By the choice of $u^\star$, and by conditional Jensen together with Assumption A,

$$\|\varrho\|^2_{L^2(P_X)} = \mathbb{E}\|j^\star_{\omega_0,P_{XZ}}(X)u^\star\|^2 = (u^\star)^\top M_0 u^\star = \lambda_0, \qquad |\varrho(x)| \leq \|j^\star_{\omega_0,P_{XZ}}(x)\|_{2,d} \leq D. \tag{61}$$

This is the only place Assumption H is used, and it is what makes the gap in Step 2 nonzero.

For a measurable $b : \mathcal{X} \to [-1,1]$, define P_b to be the law under which $(X,Z) \sim P_{XZ}$ and, conditionally on (X,Z), $Y = A\,\xi\, e(X)$ for a random sign $\xi \in \{-1,+1\}$ with

$$\mathbb{P}(\xi = 1 \mid X, Z) = \frac{1 + b(X)}{2}.$$

Two deterministic facts about Y follow immediately. First, since $|\xi| = 1$ and $\|e(X)\| = 1$, $\|Y\| = A|\xi|\,\|e(X)\| = A$ for every realization. Second, $\mathbb{E}[\xi \mid X, Z] = 1 \cdot \frac{1+b(X)}{2} + (-1)\cdot\frac{1-b(X)}{2} = b(X)$ already depends on X alone (not on Z), so the tower rule collapses immediately and $m^\star_{P_b}(X) \coloneqq \mathbb{E}[Y \mid X] = A\,b(X)\,e(X)$. Because the (X,Z)-law is unchanged, the nuisances $h^\star_{\omega,P_b}(X) = \mathbb{E}[g_\omega(Z) \mid X]$ and $j^\star_{k,\omega,P_b}(X) = \mathbb{E}[\partial_{\omega_k} g_\omega(Z) \mid X]$ are identical, as functions of X, to those of P_{XZ}, for *every* choice of b; in particular ζ, ϱ and e of (60) are the same under every P_b. The two conditions defining $\mathcal{P}_Y(P_{XZ})$ are therefore both satisfied, so $P_b \in \mathcal{P}_Y(P_{XZ})$.

It remains to check membership in $\mathcal{P}_Y^{\mathrm{reg}}(P_{XZ}; n)$ for the two members used below. Using $\varrho(x)e(x) = \zeta(x)$, these have $m^\star_{P_{b_0}} = 0$ and $m^\star_{P_{b_1}} = (A\tau/D)\,\zeta$ with τ as in (62). As $0 \in \mathcal{F}_m(t_n)$ the first is immediate, and for the second, comparing with $f = 0$ and using $\|\zeta\|_{L^2(P_X)} = \sqrt{\lambda_0}$ from (61),

$$\inf_{f\in\mathcal{F}_m(t_n)} \|f - m^\star_{P_{b_1}}\|^2_{L^2(P_X)} \leq \|m^\star_{P_{b_1}}\|^2_{L^2(P_X)} = \frac{A^2\tau^2\lambda_0}{D^2} \leq \frac{A^2D^2}{\lambda_0}\, n^{-1} \leq L_m^2\, t_n^{2\nu},$$

the last two steps by $\tau \leq D/(\sqrt{\lambda_0}\sqrt{n})$, by $2\nu/b = 2\nu/(\nu(2+p)+\gamma) \leq 1$ so that $n^{-1} \leq n^{-2\nu/b} = t_n^{2\nu}$, and by $A^2D^2/\lambda_0 \leq L_m^2$. Comparing with $f = 0$ suffices precisely because the perturbation shrinks at least as fast as $n^{-1/2}$; no approximation power of $\mathcal{F}_m$ beyond containing the origin is required.

Step 2 (the target under P_b). By the direct form $\partial_{\omega_k}\mathcal{F}_P(\omega) = \mathbb{E}_P[\langle h^\star_{\omega,P}(X) - Y, j^\star_{k,\omega,P}(X)\rangle]$: since $h^\star_{\omega_0,P_b}$ and $j^\star_{k,\omega_0,P_b}$ are both already functions of X alone, conditioning on X first replaces Y by its conditional mean $m^\star_{P_b}(X) = Ab(X)e(X)$ from Step 1, and contracting with $u^\star$ and using $\langle e(X), \zeta(X)\rangle = \varrho(X)$,

$$\begin{aligned} \big\langle u^\star, \nabla\mathcal{F}_{P_b}(\omega_0)\big\rangle &= \mathbb{E}\big[\big\langle h^\star_{\omega_0,P_{XZ}}(X) - m^\star_{P_b}(X),\, \zeta(X)\big\rangle\big] \\ &= \underbrace{\mathbb{E}\big[\langle h^\star_{\omega_0,P_{XZ}}(X), \zeta(X)\rangle\big]}_{\text{independent of } b} - A\,\mathbb{E}\big[b(X)\varrho(X)\big]. \end{aligned}$$

The first term does not depend on b at all, so it will cancel exactly once two members of the family are subtracted below; only the second term remains with b. The pointwise bound $|\varrho(x)| \le D$ recorded in (61) is what keeps the perturbation constructed next inside the legal range $[-1, 1]$.

Now pick the two specific members of the family P_b in Step 1: $b_0 := 0$ and

$$b_1 := \frac{\tau\,\varrho}{D}, \qquad \tau := \min\Big\{1,\ \frac{D}{\sqrt{\lambda_0}\,\sqrt{n}}\Big\}, \tag{62}$$

the perturbation size τ being distinct from the approximation scale t of Assumption B. Admissibility of b_1 is immediate from (61): $|b_1(x)| = \tau|\varrho(x)|/D \le \tau \le 1$. Since the b-independent term cancels between b_1 and b_0, the two targets differ, in the direction $u^\star$, by

$$\Delta := \big|\langle u^\star, \nabla\mathcal{F}_{P_{b_1}}(\omega_0) - \nabla\mathcal{F}_{P_{b_0}}(\omega_0)\rangle\big| = A\,\mathbb{E}[b_1(X)\varrho(X)] = \frac{A\tau\,\|\varrho\|^2_{L^2(P_X)}}{D} = \frac{A\tau\lambda_0}{D}, \tag{63}$$

the signed difference being $-A\,\mathbb{E}[b_1\varrho]$. Substituting τ, and pulling the positive constant $\sqrt{\lambda_0}/D$ through the minimum (valid since $c\cdot\min\{a,b\} = \min\{ca, cb\}$ for $c > 0$), gives

$$\Delta = A\sqrt{\lambda_0}\,\min\Big\{n^{-1/2},\ \frac{\sqrt{\lambda_0}}{D}\Big\},$$

so that $\Delta = A\sqrt{\lambda_0}\,n^{-1/2}$ for every $n \ge D^2/\lambda_0$.

Step 3 (the divergence between the n-fold products). Since P_{b_1} and P_{b_0} share the identical (X, Z)-law P_{XZ} (Step 1), the chain rule for KL leaves only the conditional piece, and since $Y \mid X, Z$ is, under either law, a relabeling of $\mathrm{Bern}(\frac{1+b(X)}{2})$ depending on X alone, this reduces to a Bernoulli comparison. KL also tensorizes over the n i.i.d. draws, giving

$$\mathrm{KL}\big(P_{b_1}^{\otimes n}\,\big\|\,P_{b_0}^{\otimes n}\big) = n\,\mathbb{E}_X\Big[\mathrm{KL}\Big(\mathrm{Bern}\big(\tfrac{1+b_1(X)}{2}\big)\,\Big\|\,\mathrm{Bern}\big(\tfrac{1}{2}\big)\Big)\Big].$$

Bounding each Bernoulli KL by its chi-squared divergence ($\mathrm{KL} \le \chi^2$) and computing that divergence directly,

$$\chi^2\Big(\mathrm{Bern}\big(\tfrac{1+b}{2}\big)\,\Big\|\,\mathrm{Bern}\big(\tfrac{1}{2}\big)\Big) = b^2,$$

then substituting $b_1 = \tau\varrho/D$, gives

$$\mathrm{KL}\big(P_{b_1}^{\otimes n}\,\big\|\,P_{b_0}^{\otimes n}\big) \le n\,\mathbb{E}[b_1(X)^2] = \frac{n\,\tau^2\lambda_0}{D^2} \ \le\ 1, \tag{64}$$

where the last step holds by exactly how τ was chosen in Step 2, in either branch of the minimum. So no matter how large n is, $P_{b_1}^{\otimes n}$ and $P_{b_0}^{\otimes n}$ stay within KL divergence 1 of each other even though their targets differ by the non-vanishing gap Δ from Step 2.

Step 4. Write $\widehat{\Psi}_{u^\star} := \langle u^\star, \widehat{\Psi}\rangle$ and $\theta_i := \langle u^\star, \nabla\mathcal{F}_{P_{b_i}}(\omega_0)\rangle$, so that $|\theta_1 - \theta_0| = \Delta$ by (63). We show

$$\max_{i\in\{0,1\}}\ \mathbb{P}_{b_i}^{\otimes n}\big(|\widehat{\Psi}_{u^\star} - \theta_i| \ge \Delta/2\big) \ \ge\ \tfrac{1}{2}\big(1 - \sqrt{1/2}\big). \tag{65}$$

Define the auxiliary set $\mathcal{A} := \{y \in \mathbb{R} : |y - \theta_0| < \Delta/2\}$, the symbol being distinct from the constant A bounding $\|Y\|$. For $y \in \mathcal{A}$, the triangle inequality together with $|\theta_0 - \theta_1| = \Delta$ (Step 2) gives

$$|y - \theta_1| \ \ge\ \Delta - |y - \theta_0| \ >\ \Delta - \Delta/2 \ =\ \Delta/2,$$

so $\mathcal{A} \subseteq \{y \in \mathbb{R} : |y - \theta_1| \ge \Delta/2\}$. Since $\widehat{\Psi}_{u^\star}$ is a measurable function of the sample, the total variation distance between its laws under $\mathbb{P}_{b_0}^{\otimes n}$ and $\mathbb{P}_{b_1}^{\otimes n}$ is at most $\mathrm{TV}(P_{b_0}^{\otimes n}, P_{b_1}^{\otimes n})$. Evaluating at the set $\mathcal{A}$, and applying Pinsker's inequality to Step 3's bound (64) — legitimate in this order because TV is symmetric while KL is not — gives

$$\mathbb{P}_{b_0}^{\otimes n}(\widehat{\Psi}_{u^\star} \in \mathcal{A}) - \mathbb{P}_{b_1}^{\otimes n}(\widehat{\Psi}_{u^\star} \in \mathcal{A}) \ \le\ \mathrm{TV}\big(P_{b_1}^{\otimes n}, P_{b_0}^{\otimes n}\big) \ \le\ \sqrt{\tfrac{1}{2}\,\mathrm{KL}\big(P_{b_1}^{\otimes n}\|P_{b_0}^{\otimes n}\big)} \ \le\ \sqrt{1/2}. \tag{66}$$

By definition of $\mathcal{A}$, $\mathbb{P}_{b_0}^{\otimes n}(\widehat{\Psi}_{u^\star} \in \mathcal{A}) = 1 - \mathbb{P}_{b_0}^{\otimes n}(|\widehat{\Psi}_{u^\star} - \theta_0| \ge \Delta/2)$, and using $\mathcal{A} \subseteq \{y : |y - \theta_1| \ge \Delta/2\}$, $\mathbb{P}_{b_1}^{\otimes n}(\widehat{\Psi}_{u^\star} \in \mathcal{A}) \le \mathbb{P}_{b_1}^{\otimes n}(|\widehat{\Psi}_{u^\star} - \theta_1| \ge \Delta/2)$. Substituting both into (66) and rearranging gives

$$\mathbb{P}_{b_0}^{\otimes n}\big(|\widehat{\Psi}_{u^\star} - \theta_0| \ge \Delta/2\big) + \mathbb{P}_{b_1}^{\otimes n}\big(|\widehat{\Psi}_{u^\star} - \theta_1| \ge \Delta/2\big) \ \ge\ 1 - \sqrt{1/2},$$

and since $\max(x, y) \geq (x + y)/2$ for any $x, y \geq 0$, (65) follows.

Step 5. Let $i^\star$ denote an index attaining the maximum in (65). Since $\|u^\star\| = 1$, Cauchy–Schwarz gives $\|\widehat{\Psi} - \nabla\mathcal{F}_P(\omega_0)\| \geq |\widehat{\Psi}_{u^\star} - \langle u^\star, \nabla\mathcal{F}_P(\omega_0)\rangle|$, so by Markov's inequality applied at level $\Delta/2$ under $\mathbb{P}_{b_{i^\star}}^{\otimes n}$,

$$\mathbb{E}_{P_{b_{i^\star}}}\left[\|\widehat{\Psi} - \nabla\mathcal{F}_{P_{b_{i^\star}}}(\omega_0)\|\right] \;\geq\; \frac{\Delta}{2} \cdot \mathbb{P}_{b_{i^\star}}^{\otimes n}\big(|\widehat{\Psi}_{u^\star} - \theta_{i^\star}| \geq \Delta/2\big) \;\geq\; \frac{\Delta}{2} \cdot \frac{1 - \sqrt{1/2}}{2} \;=\; c_0\Delta.$$

Since $P_{b_0}, P_{b_1} \in \mathcal{P}_Y^{\mathrm{reg}}(P_{XZ}; n)$ (Step 1):

$$\sup_{P \in \mathcal{P}_Y^{\mathrm{reg}}(P_{XZ};n)} \mathbb{E}_P\left[\|\widehat{\Psi} - \nabla\mathcal{F}_P(\omega_0)\|\right] \;\geq\; \max_{i\in\{0,1\}} \mathbb{E}_{P_{b_i}}\left[\|\widehat{\Psi} - \nabla\mathcal{F}_{P_{b_i}}(\omega_0)\|\right] \;\geq\; c_0\Delta.$$

Substituting Δ from Step 2 gives the stated bound, for every estimator $\widehat{\Psi}$ and hence for the infimum. The uniform-risk version follows since $\sup_\omega \|\widehat{\Psi}_\omega - \nabla\mathcal{F}_P(\omega)\| \geq \|\widehat{\Psi}_{\omega_0} - \nabla\mathcal{F}_P(\omega_0)\|$ pointwise. □

B Nuisance Learning and Plug-in Hypergradients

B.1 Nuisance learning

This appendix records standard sufficient conditions under which the nuisance assumptions used in Section 3.2 can be verified. We recall:

$$h^\star_{\omega,P}(X) = \mathbb{E}_P[g_\omega(Z) \mid X], \qquad j^\star_{k,\omega,P}(X) = \mathbb{E}_P[\partial_{\omega_k} g_\omega(Z) \mid X], \qquad m^\star_P(X) = \mathbb{E}_P[Y \mid X].$$

Thus nuisance learning reduces to standard regression. For example, define

$$R^h_\omega(h) = \frac{1}{2}\mathbb{E}_P[\|h(X) - g_\omega(Z)\|^2], \qquad R^j_\omega(j) = \frac{1}{2}\mathbb{E}_P[\|j(X) - \partial_\omega g_\omega(Z)\|^2_{2,d}],$$

and $R^m(m) = \frac{1}{2}\mathbb{E}_P\|m(X) - Y\|^2$. Their minimizers are respectively $h^\star_{\omega,P}$, $j^\star_{\omega,P}$, and $m^\star_P$.

A generic verification route is the following. Let $\mathcal{H}_n, \mathcal{J}_n, \mathcal{M}_n$ be the classes used to estimate $h^\star_{\omega,P}$, $j^\star_{\omega,P}$, and $m^\star_P$. Let $h^\circ_{n,\omega}$, $j^\circ_{n,\omega}$, and m°_n denote the corresponding population risk minimizers over these classes. For squared loss,

$$R^h_\omega(h) - R^h_\omega(h^\star_\omega) = \frac{1}{2}\|h - h^\star_\omega\|^2_{L^2(P_X)},$$

and similarly for j and m. Hence an oracle inequality such as $R^h_\omega(\hat{h}_\omega) - R^h_\omega(h^\circ_{n,\omega}) = O_p(\delta^2_{h,n})$ implies

$$\|\hat{h}_\omega - h^\star_\omega\|_{L^2(P_X)} = O_p(a_{h,n} + \delta_{h,n}), \qquad a_{h,n} := \|h^\circ_{n,\omega} - h^\star_\omega\|_{L^2(P_X)}.$$

Uniformly over ω, the same argument gives

$$\sup_{\omega\in\Omega} \|\hat{h}_\omega - h^\star_\omega\|_{L^2(P_X)} = O_p(a_{h,n} + \delta_{h,n})$$

whenever the oracle inequality is uniform in ω. Analogous bounds hold for $\hat{j}_\omega$ and $\hat{m}$. For the expected uniform guarantees under Assumptions A to D, (36) supplies the required uniform second-moment bounds. Such oracle inequalities are standard consequences of empirical-process bounds for least-squares estimators over entropy- or Rademacher-controlled classes (van der Vaart & Wellner, 1996; van de Geer, 2000; Kosorok, 2008; Wainwright, 2019).

The rate δ_n depends on the size of the regression class. A common formulation uses a localized critical radius. If δ_n satisfies a local Rademacher complexity inequality of the form

$$\mathfrak{R}_n\big(\mathcal{F}_n \cap \{f : \|f - f^\circ_n\|_{L^2(P_X)} \leq \delta_n\}\big) \lesssim \delta^2_n,$$

then least-squares ERM over $\mathcal{F}_n$ typically satisfies $\|\hat{f} - f^\star\|_{L^2(P_X)} = O_p(a_n + \delta_n)$, up to standard boundedness and measurability conditions (Bartlett et al., 2005; Wainwright, 2019; van der Laan, 2026). For classes with metric entropy $\log N(\varepsilon, \mathcal{F}_n, L_2) \lesssim \varepsilon^{-p}$, $p \in (0, 2)$, the corresponding nonparametric norm rate is typically $n^{-1/(2+p)}$ modulo approximation error. Since $p < 2$ gives

$1/(2+p) > 1/4$, such rates are compatible with the second-order conditions in Assumption I, provided the approximation terms are negligible.

Fixed-ω inference in Appendix E uses Assumption I: each nuisance error is $o_p(N^{-1/4})$, so the relevant cross-products and squared errors are $o_p(N^{-1/2})$. In particular, the mixed-bias bound is controlled by $\|\hat{j}_\omega - j^\star_{\omega,P}\|_{L^2(P_X)^d}(\|\hat{h}_\omega - h^\star_{\omega,P}\|_{L^2(P_X)} + \|\hat{m} - m^\star_P\|_{L^2(P_X)})$. Standard least-squares, entropy, and Rademacher arguments can verify these rates under stronger versions of the regression conditions above (Chernozhukov et al., 2018; Foster & Syrgkanis, 2023; van der Laan et al., 2025). The empirical remainder is handled by the linear-plus-product decomposition in Remark E.1.

For the uniform result in Section 3, one also needs complexity control of the score classes generated by the learned nuisance paths. Assumptions E and G give a sufficient condition: conditionally on the training fold, the learned nuisances lie in bounded classes with polynomial entropy. For kernel ridge regression with bounded kernel $K(x,x) \le \kappa^2$ and ridge parameter $\lambda_n > 0$, comparison with the zero function gives $\lambda_n \|\hat{f}\|^2_{\mathcal{H}_K} \le C^2$ when the regression target is bounded by C. Hence $\|\hat{f}\|_{\mathcal{H}_K} \le C/\sqrt{\lambda_n}$ and $\|\hat{f}\|_\infty \le \kappa C/\sqrt{\lambda_n}$. Entropy or Rademacher bounds for the induced RKHS ball then yield the required class-complexity control (Schölkopf et al., 2001; Steinwart & Christmann, 2008; Caponnetto & De Vito, 2007). For neural-network classes with bounded depth, width, activation Lipschitz constants, and weight norms, standard covering-number bounds give entropy of order $s_n \log(C_n/\varepsilon)$, where s_n is the number of effective parameters and C_n depends on the norm constraints. If the architecture grows with n, these constants grow accordingly and must be small enough for the localized critical-radius terms in Theorem A.1 to vanish.

For nonquadratic inner problems, additional Riesz or adjoint correction nuisances may appear. The same verification principle applies: one needs oracle inequalities for the additional nuisance learners and product-rate conditions matching the corresponding von Mises remainder (van der Laan et al., 2025).

B.2 Plug-in hypergradient estimation

This section records what the direct plug-in hypergradient estimates in the quadratic specialization, and why it should not be expected to deliver semiparametric-efficient inference. Functional implicit differentiation and KBO both lead to plug-in hypergradients in this broad sense: one estimates the inner solution, or an adjoint/sensitivity representation of it, and substitutes the estimate into a gradient formula (Petrulionyte et al., 2024; El Khoury et al., 2025). This is natural for optimization. It is not, by itself, an orthogonal statistical estimator.

For clarity, we analyze the sample-split plug-in estimator. Let $\hat{h}_\omega$ and $\hat{j}_\omega = (\hat{j}_{1,\omega}, \dots, \hat{j}_{d,\omega})$ be trained on S_1 and evaluated on the independent fold S_2:

$$[\widehat{\nabla \mathcal{F}}^{\mathrm{PI}}(\omega)]_k := \mathbb{E}_{P_{n,2}}\big[\langle \hat{h}_\omega(X) - Y, \hat{j}_{k,\omega}(X)\rangle\big], \qquad k = 1, \dots, d. \tag{67}$$

The full-sample version replaces $P_{n,2}$ by P_n, but then the empirical-process term involves a function estimated on the same sample on which it is averaged. The sample-split version isolates the statistical issue cleanly.

Exact plug-in bias. For arbitrary $h, v \in L^2(P_X)$, define

$$\varphi^{\mathrm{PI}}_{k,\omega}(X, Y, Z; h, v) := \langle h(X) - Y, v(X)\rangle.$$

Since $h(X)$ and $v(X)$ are functions of X,

$$\mathbb{E}_P\big[\varphi^{\mathrm{PI}}_{k,\omega}(\cdot; h, v)\big] = \langle h - m^\star_P, v\rangle_{L^2(P_X)}.$$

Also,

$$\partial_{\omega_k} \mathcal{F}(\omega) = \langle h^\star_{\omega,P} - m^\star_P, j^\star_{k,\omega,P}\rangle_{L^2(P_X)}.$$

Therefore

$$\begin{aligned}
&\mathbb{E}_P\big[\varphi^{\mathrm{PI}}_{k,\omega}(\cdot; h, v)\big] - \partial_{\omega_k}\mathcal{F}(\omega) \\
&\quad = \langle h - h^\star_{\omega,P}, j^\star_{k,\omega,P}\rangle_{L^2(P_X)} + \langle h^\star_{\omega,P} - m^\star_P, v - j^\star_{k,\omega,P}\rangle_{L^2(P_X)} + \langle h - h^\star_{\omega,P}, v - j^\star_{k,\omega,P}\rangle_{L^2(P_X)}.
\end{aligned} \tag{68}$$

The first two terms are first-order nuisance errors. Thus the plain plug-in hypergradient is generally not Neyman-orthogonal with respect to $h^\star_\omega$ or $j^\star_{k,\omega}$.

In particular, if $\|j^\star_{k,\omega,P}\|_{L^2(P_X)} < \infty$ and $\|h^\star_{\omega,P} - m^\star_P\|_{L^2(P_X)} < \infty$, then

$$\begin{aligned}\left|\mathbb{E}_P[\varphi^{\mathrm{PI}}_{k,\omega}(\cdot; h, v)] - \partial_{\omega_k}\mathcal{F}(\omega)\right| \le{}& \|j^\star_{k,\omega,P}\|_{L^2(P_X)}\|h - h^\star_{\omega,P}\|_{L^2(P_X)} \\ &+ \|h^\star_{\omega,P} - m^\star_P\|_{L^2(P_X)}\|v - j^\star_{k,\omega,P}\|_{L^2(P_X)} \\ &+ \|h - h^\star_{\omega,P}\|_{L^2(P_X)}\|v - j^\star_{k,\omega,P}\|_{L^2(P_X)}.\end{aligned} \tag{69}$$

Proposition B.1 (Rate of the sample-split plug-in estimator)**.** *Fix $\omega \in \Omega$ and $k \in \{1, \dots, d\}$. Suppose $\hat{h}_\omega$ and $\hat{j}_{k,\omega}$ are trained on S_1, and suppose*

$$\mathbb{E}_P\left[\varphi^{\mathrm{PI}}_{k,\omega}(\cdot; \hat{h}_\omega, \hat{j}_{k,\omega})^2\right] = O_p(1).$$

If $\|\hat{h}_\omega - h^\star_{\omega,P}\|_{L^2(P_X)} = O_p(n^{-\alpha})$ and $\|\hat{j}_{k,\omega} - j^\star_{k,\omega,P}\|_{L^2(P_X)} = O_p(n^{-\beta})$, then

$$[\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega)]_k - \partial_{\omega_k}\mathcal{F}(\omega) = O_p\left(n^{-1/2} + n^{-\alpha} + n^{-\beta} + n^{-(\alpha+\beta)}\right).$$

Proof. Conditionally on S_1, the function $\varphi^{\mathrm{PI}}_{k,\omega}(\cdot; \hat{h}_\omega, \hat{j}_{k,\omega})$ is fixed, and the evaluation observations in S_2 are independent. Hence

$$(\mathbb{E}_{P_{n,2}} - \mathbb{E}_P)[\varphi^{\mathrm{PI}}_{k,\omega}(\cdot; \hat{h}_\omega, \hat{j}_{k,\omega})] = O_p(n^{-1/2}).$$

The population bias is bounded by (69). Substituting the assumed nuisance rates gives the result. □

This proposition should be read literally. For the direct plug-in estimator, the generic bias is not a product of nuisance errors. Under the score-moment condition above, a sufficient condition for root-n consistency is that the sum of the bias terms in (68) is $O_p(n^{-1/2})$. The centered oracle-score expansion below instead requires this sum to be $o_p(n^{-1/2})$, together with score stability. Bounding the first-order terms separately gives sufficient nuisance rates of essentially parametric order, but cancellations may relax this requirement. A product-rate condition such as $\alpha + \beta > 1/2$ is not enough by itself unless additional structure controls the linear terms.

Asymptotic distribution and efficiency. If the plug-in bias in (68) is $o_p(n^{-1/2})$ and

$$\left\|\varphi^{\mathrm{PI}}_{k,\omega}(\cdot; \hat{h}_\omega, \hat{j}_{k,\omega}) - \varphi^{\mathrm{PI}}_{k,\omega}(\cdot; h^\star_{\omega,P}, j^\star_{k,\omega,P})\right\|_{L^2(P)} = o_p(1),$$

then the sample-split plug-in estimator admits an asymptotic linear representation with oracle-score summand

$$\varphi^{\mathrm{PI}}_{k,\omega}(\cdot; h^\star_{\omega,P}, j^\star_{k,\omega,P}) - \partial_{\omega_k}\mathcal{F}(\omega).$$

Consequently,

$$\sqrt{n}\left([\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega)]_k - \partial_{\omega_k}\mathcal{F}(\omega)\right) \xrightarrow{d} \mathcal{N}\left(0, \mathrm{Var}_P\left[\langle h^\star_{\omega,P}(X) - Y, j^\star_{k,\omega,P}(X)\rangle\right]\right).$$

The oracle-score summand generally differs from the canonical gradient in Theorem C.1, which includes additional correction terms. Its asymptotic variance need not equal $\Sigma_{P,\omega,kk}$, the semiparametric efficiency bound.

Contrast with the debiased estimator. In the same quadratic specialization, the debiased score used in Section 3.2 is

$$\varphi_{k,\omega}(X, Y, Z; h, j, m) = \langle g_\omega(Z) - Y, j_k(X)\rangle + \langle \partial_{\omega_k} g_\omega(Z) - j_k(X), h(X) - m(X)\rangle.$$

For arbitrary h, j, m, its population bias satisfies

$$\mathbb{E}_P\left[\varphi_{k,\omega}(\cdot; h, j, m)\right] - \partial_{\omega_k}\mathcal{F}(\omega) = -\langle j_k - j^\star_{k,\omega,P}, h - h^\star_{\omega,P}\rangle_{L^2(P_X)} + \langle j_k - j^\star_{k,\omega,P}, m - m^\star_P\rangle_{L^2(P_X)}. \tag{70}$$

Thus

$$\left|\mathbb{E}_P\left[\varphi_{k,\omega}(\cdot; h, j, m)\right] - \partial_{\omega_k}\mathcal{F}(\omega)\right| \le \|j_k - j^\star_{k,\omega,P}\|_{L^2(P_X)}\left(\|h - h^\star_{\omega,P}\|_{L^2(P_X)} + \|m - m^\star_P\|_{L^2(P_X)}\right).$$

The first-order terms present in the plug-in bias (68) are absent. This is the practical content of orthogonality: the estimator can be root-n and asymptotically normal under product-rate nuisance conditions, such as

$$\|\hat{j}_{k,\omega} - j^\star_{k,\omega}\|_{L^2(P_X)} \left(\|\hat{h}_\omega - h^\star_{\omega,P}\|_{L^2(P_X)} + \|\hat{m} - m^\star_P\|_{L^2(P_X)} \right) = o_p(n^{-1/2}),$$

rather than requiring each nuisance to be estimated at a parametric rate. Rates at the $n^{-1/4}$ boundary are not enough as big-O statements; the product must be little-$o(n^{-1/2})$ for the centered CLT in Theorem E.1.

This is the core distinction. For the plug-in hypergradient, first-order nuisance errors can obstruct root-n inference. The debiased estimator changes the score so that those first-order terms cancel, leaving only second-order nuisance products and the empirical average of the efficient influence function.

B.3 Uniform plug-in maximal inequality

Proposition B.1 controls the plug-in estimator at a single fixed ω. We now give its uniform-over-Ω analogue, paralleling Theorem A.1. Stack the coordinate-wise plug-in bias identity (68) into the vector

$$\varphi^{\mathrm{PI}}_\omega(X, Y; h, j) \coloneqq \big(\langle h(X) - Y, j_1(X)\rangle, \ldots, \langle h(X) - Y, j_d(X)\rangle\big)^\top,$$

using a shared nuisance pair $(h, j) = (h, (j_1, \ldots, j_d)) \in L^2(P_X) \times L^2(P_X)^d$, with $\hat{h}_\omega, \hat{j}_\omega$ trained on S_1 and evaluated on S_2, as in Section 3.2. Define the one-fold plug-in estimator

$$\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega) \coloneqq \mathbb{E}_{P_{n,2}}[\,\varphi^{\mathrm{PI}}_\omega(\cdot; \hat{h}_\omega, \hat{j}_\omega)].$$

Since $\varphi^{\mathrm{PI}}_\omega$ does not depend on g_ω, $\partial_\omega g_\omega$, or the regression nuisance m, the following result needs only the restrictions of Assumptions E to G to the classes $\mathcal{H}_n, \mathcal{J}_n, \mathcal{H}^\star, \mathcal{J}^\star$ (dropping $\mathcal{G}_\Omega, \dot{\mathcal{G}}_\Omega, \mathcal{M}_n$, which the plug-in score does not use).

Theorem B.1 (Uniform plug-in maximal inequality)**.** *Under Assumption A and the restrictions of Assumptions E to G described above, there exist constants $C^{\mathrm{PI}} > 0$ depending only on p and $K^{\mathrm{PI}}_A > 0$ depending only on the population entropy constants of $\mathcal{H}^\star, \mathcal{J}^\star$ such that, on the event $\sup_{\omega\in\Omega} \|\hat{h}_\omega - h^\star_{\omega,P}\|_{L^2(P_X)} \le r_h$, $\sup_{\omega\in\Omega} \|\hat{j}_\omega - j^\star_{\omega,P}\|_{L^2(P_X)^d} \le r_j$,*

$$\begin{aligned}
&\mathbb{E}\left[\sup_{\omega\in\Omega} \|\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega) - \nabla\mathcal{F}(\omega)\| \,\middle|\, S_1 \right] \\
&\le \frac{C^{\mathrm{PI}}}{\sqrt{n}}(A+B)D\sqrt{1+K^{\mathrm{PI}}_A} + \mathfrak{C}^{\mathrm{PI}}_{B,n}(r_h, r_j) + Dr_h + (A+B)r_j + r_h r_j, \qquad (71)
\end{aligned}$$

where $\mathfrak{C}^{\mathrm{PI}}_{B,n}(r_h, r_j)$ is the localized nuisance-empirical-process complexity of the plug-in nuisance-difference class.

Proof. We follow the six-step structure of the proof of Theorem A.1 (Section A.5), noting exactly where the plug-in score departs from the debiased score.

Step 1 (population identities) is unchanged: $\|h^\star_{\omega,P}(X)\| \le B$, $\|Y\| \le A$, $\|j^\star_{\omega,P}(X)\|_{2,d} \le D$.

Step 2 (exact bias identity) is given by (68), summed over coordinates via Cauchy–Schwarz exactly as in Step 2 of the proof of Theorem A.1:

$$\begin{aligned}
\big\|\mathbb{E}_P[\varphi^{\mathrm{PI}}_\omega(\cdot; h, j)] - \nabla\mathcal{F}(\omega)\big\| \le{}& D\|h - h^\star_{\omega,P}\|_{L^2(P_X)} + (A+B)\|j - j^\star_{\omega,P}\|_{L^2(P_X)^d} \\
&+ \|h - h^\star_{\omega,P}\|_{L^2(P_X)}\|j - j^\star_{\omega,P}\|_{L^2(P_X)^d}. \qquad (72)
\end{aligned}$$

Unlike the debiased bias bound (29), this is dominated by the two *first-order* terms $Dr_h + (A+B)r_j$, not by a product alone – the algebraic source of the rate difference with Theorem A.1.

Step 3 (estimator decomposition) is unchanged: writing $\varphi^{\mathrm{PI},\star}_\omega \coloneqq \varphi^{\mathrm{PI}}_\omega(\cdot; h^\star_{\omega,P}, j^\star_{\omega,P})$ and $\hat{\varphi}^{\mathrm{PI}}_\omega \coloneqq \varphi^{\mathrm{PI}}_\omega(\cdot; \hat{h}_\omega, \hat{j}_\omega)$,

$$\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega) - \nabla\mathcal{F}(\omega) = (\mathbb{E}_{P_{n,2}} - \mathbb{E}_P)[\varphi^{\mathrm{PI},\star}_\omega] + (\mathbb{E}_{P_{n,2}} - \mathbb{E}_P)[\hat{\varphi}^{\mathrm{PI}}_\omega - \varphi^{\mathrm{PI},\star}_\omega] + \big(\mathbb{E}_P[\hat{\varphi}^{\mathrm{PI}}_\omega] - \nabla\mathcal{F}(\omega)\big),$$

and the dual-representation reduction to scalar empirical processes over $(\omega, u) \in \Omega \times S^{d-1}$ proceeds identically, using (72) for the population-bias term.

Step 4 (envelopes). By Cauchy–Schwarz, $\|\varphi_\omega^{\mathrm{PI},\star}(X,Y)\| \le \|h^\star_{\omega,P}(X) - Y\| \, \|j^\star_{\omega,P}(X)\|_{2,d} \le (A+B)D$, and since $\nabla\mathcal{F}(\omega) = \mathbb{E}_P[\varphi_\omega^{\mathrm{PI},\star}]$, Jensen's inequality gives $\|\nabla\mathcal{F}(\omega)\| \le (A+B)D$; thus $F_A^{\mathrm{PI}} := 2(A+B)D$ is an envelope for the oracle score class $\mathcal{F}_A^{\mathrm{PI}} := \{u^\top(\varphi_\omega^{\mathrm{PI},\star} - \nabla\mathcal{F}(\omega)) : \omega \in \Omega, u \in S^{d-1}\}$. For the nuisance-difference class, writing $\Delta h = h - h^\star_{\omega,P}$, $\Delta j = j - j^\star_{\omega,P}$,

$$\begin{aligned}\varphi_{k,\omega}^{\mathrm{PI}}(X,Y;h,j) - \varphi_{k,\omega}^{\mathrm{PI},\star}(X,Y) =& \langle \Delta h(X), j^\star_{k,\omega,P}(X)\rangle + \langle \Delta h(X), \Delta j_k(X)\rangle \\ &+ \langle h^\star_{\omega,P}(X) - Y, \Delta j_k(X)\rangle,\end{aligned}$$

a sum of three bounded products of the same form as those already controlled in Step 5 of the proof of Theorem A.1 (there, the analogous terms bounded the second and third pieces of the debiased score difference). With $\bar{H}_n, \bar{J}_n$ the envelopes of $\mathcal{H}_n, \mathcal{J}_n$, this gives an S_1-measurable envelope $F_{B,n}^{\mathrm{PI}} := (\bar{H}_n + A)\bar{J}_n + (A+B)D$ for the plug-in nuisance-difference class $\mathcal{F}_{B,n}^{\mathrm{PI,loc}}(r_h, r_j)$, with $L^2(P)$ radius at most $\rho_{B,n}^{\mathrm{PI}} := Dr_h + \bar{H}_n r_j + (A+B)r_j$ on the event of Theorem B.1.

Step 5 (entropy transfer). The oracle class satisfies $\mathcal{F}_A^{\mathrm{PI}} \subset (\mathcal{U}_h \odot \mathscr{J}^\star) - \mathcal{C}_\Psi$, where $\mathcal{U}_h := \mathcal{H}^\star - Y$ with envelope $A + B$, and $\mathscr{J}^\star = \mathscr{S}(\mathcal{J}^\star)$ with envelope D. A single application of Lemma A.3 (in place of the two-product combination needed there) transfers the polynomial entropy bound of Assumption F on $\mathcal{H}^\star, \mathcal{J}^\star$ to $\mathcal{F}_A^{\mathrm{PI}}$ with exponent p and a constant $K_A^{\mathrm{PI}} < \infty$ depending only on $K_{\mathcal{H}^\star}, K_{\mathcal{J}^\star}$. The nuisance-difference class transfers analogously from Assumption G applied to $\mathcal{H}_n, \mathcal{J}_n$, via Lemmas A.2 and A.3 applied to the three-term expansion above, giving an S_1-measurable constant $K_{B,n}^{\mathrm{PI}} < \infty$ with the same exponent p.

Step 6 (conclusion). Corollary A.1 applied to $\mathcal{F}_A^{\mathrm{PI}}$ gives the oracle term $\frac{C^{\mathrm{PI}}}{\sqrt{n}}(A+B)D\sqrt{1 + K_A^{\mathrm{PI}}}$ with $C^{\mathrm{PI}} = C_p$; the localized-Rademacher argument of Section A.6, applied verbatim to $\mathcal{F}_{B,n}^{\mathrm{PI,loc}}(r_h, r_j)$ (which depends only on its envelope, entropy exponent p, and $L^2(P)$ radius, all of which transfer identically), defines $\mathfrak{C}_{B,n}^{\mathrm{PI}}(r_h, r_j)$; and (72) gives the remaining bias terms $Dr_h + (A+B)r_j + r_h r_j$. Combining as in Step 3 gives (71). □

Corollary B.1 (Plug-in polynomial-entropy rate)**.** *Under the conditions of Theorem B.1, suppose the nuisance envelopes and entropy constants remain bounded and $r_h, r_j = O(n^{-1/(2+p)})$. Then, on the same training event,*

$$\mathbb{E}\left[\sup_{\omega\in\Omega} \|\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega) - \nabla\mathcal{F}(\omega)\| \,\middle|\, S_1\right] = O(n^{-1/(2+p)}). \tag{73}$$

Proof. As in the proof of Proposition A.1 (Section A.6), $\mathfrak{C}_{B,n}^{\mathrm{PI}}(r_h, r_j) = o(n^{-1/2})$ under these conditions, since that argument only used the envelope, entropy exponent, and $L^2(P)$ radius of the localized class, all of which are the same order here. The first-order population-bias terms satisfy $Dr_h + (A+B)r_j + r_h r_j = O(n^{-1/(2+p)})$. Since $1/(2+p) < 1/2$, this upper bound determines the displayed rate $O(n^{-1/2} + n^{-1/(2+p)}) = O(n^{-1/(2+p)})$; it does not rule out faster decay through cancellation. □

Under the Hölder parametrization $p = d_X/s$ and the conditions of Proposition A.1 and Corollary B.1, the conditional plug-in upper-bound rate is $O(n^{-s/(2s+d_X)})$, compared with $O(n^{-1/2})$ for OBiGrad. This uniform comparison is the exact analogue, over the full outer parameter space, of the fixed-ω contrast already drawn between (69) and (70).

Plug-in rates under the main-text assumptions. Under Assumptions A to D and the tuning of Theorem 3.2, set $b = \nu(2+p) + \gamma$ and $\delta_n = (nt_n^\gamma)^{-1/(2+p)} = n^{-\nu/b}$. The nuisance errors defined in (34) satisfy $\mathbb{E}[R_{\ell,n}^2] = O(n^{-2\nu/b})$ for $\ell \in \{h, j\}$. For the fitted and oracle plug-in scores defined in the proof above, bounded-product entropy gives a deterministic nuisance-difference class with entropy of order $t_n^{-\gamma}\epsilon^{-p}$. Its localized $L^2(P)$ radius is at most $Dr_h + (A+B)r_j$, since the score difference can be written coordinatewise as $\langle h - h^\star_{\omega,P}, j_k\rangle + \langle h^\star_{\omega,P} - Y, j_k - j^\star_{k,\omega,P}\rangle$. Applying

Lemma A.7 on the independent evaluation fold and averaging over training and fitting randomness gives

$$\mathbb{E}\left[\sup_{\omega\in\Omega}\left\|(\mathbb{E}_{P_{n,2}}-\mathbb{E}_P)[\hat{\varphi}_\omega^{\mathrm{PI}}-\varphi_\omega^{\mathrm{PI},\star}]\right\|\right] \le C\{\delta_n(\mathbb{E}[R_{h,n}]+\mathbb{E}[R_{j,n}])+\delta_n^2\} = O(n^{-2\nu/b}).$$

By (72) and Cauchy–Schwarz, the expected uniform population bias is at most $D\mathbb{E}[R_{h,n}] + (A+B)\mathbb{E}[R_{j,n}]+\mathbb{E}[R_{h,n}R_{j,n}] = O(n^{-\nu/b})$. Adding the $O(n^{-1/2})$ oracle empirical-process term yields, for finite constants C_0, C_1 independent of n,

$$\mathbb{E}\left[\sup_{\omega\in\Omega}\|\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega)-\nabla\mathcal{F}(\omega)\|\right] \le C_0 n^{-1/2} + C_1 n^{-\nu/b}.$$

The expectation is over both samples and any fitting randomness.

B.4 Sample complexity: OBiGrad versus plug-in

Under Assumptions A to D and the tuning of Theorem 3.2, combine Proposition 2.1 with the unconditional gradient bounds of Theorem 3.2 and Appendix B.3. For fixed step size, $T = O(\varepsilon^{-2})$ iterations and sufficient sample sizes of orders

$$n_{\mathrm{OBiGrad}} = O\Big(\varepsilon^{-\max\{2,b/(2\nu)\}}\Big), \qquad n_{\text{plug-in}} = O\Big(\varepsilon^{-b/\nu}\Big), \qquad b=\nu(2+p)+\gamma,$$

give expected best-iterate stationarity error at most ε. Here n is the size of each fold. When $\nu(2-p) \ge \gamma$, the OBiGrad sample-size bound reduces to $O(\varepsilon^{-2})$. These are sufficient upper bounds. The following comparison retains the milder, training-conditional assumptions and applies on their training event.

The deterministic inequality in the proof of Proposition 2.1 applies to any gradient estimator. Here we take its expectation conditionally on S_1, using the training-event bounds of Theorems A.1 and B.1. For the plug-in-driven projected update $\omega_{t+1} = \mathrm{Proj}_\Omega(\omega_t - \eta\widehat{\nabla\mathcal{F}}^{\mathrm{PI}}(\omega_t))$, the corresponding expected uniform error is bounded by

$$\mathfrak{s}_N^{\mathrm{PI}} := \frac{C^{\mathrm{PI}}}{\sqrt{n}}(A+B)D\sqrt{1+K_A^{\mathrm{PI}}} + \mathfrak{C}_{B,n}^{\mathrm{PI}}(r_h,r_j) + Dr_h + (A+B)r_j + r_h r_j,$$

the bound of Theorem B.1, giving on the training event, for every $T \ge 1$,

$$\mathbb{E}\left[\min_{0\le t<T}\|G_\eta(\omega_t)\| \,\middle|\, S_1\right] \lesssim \left(\frac{\mathcal{F}(\omega_0)-F_{\inf}}{\eta T}\right)^{1/2} + \mathfrak{s}_N^{\mathrm{PI}}. \tag{74}$$

Under the conditions of Proposition A.1 and Corollary B.1 and the Hölder parametrization $p = d_X/s$, the bounds satisfy $\mathfrak{s}_N = O(n^{-1/2})$ and $\mathfrak{s}_N^{\mathrm{PI}} = O(n^{-s/(2s+d_X)})$. Applying the deterministic inequality from the proof of Proposition 2.1 conditionally on S_1 gives sufficient choices of $T = O(\varepsilon^{-2})$ for fixed step size and

$$n_{\mathrm{OBiGrad}} = O(\varepsilon^{-2}), \qquad n_{\text{plug-in}} = O\big(\varepsilon^{-(2+d_X/s)}\big),$$

up to logarithmic factors from the entropy constants and the constant $\eta \le 1/L_{\mathcal{F}}$. For $d_X/s > 0$, these upper bounds give a smaller sufficient sample-size exponent for OBiGrad; the difference between the exponents grows with the conditioning-covariate dimension d_X relative to smoothness s. The gradient error is controlled for all outer steps using the same data: because Theorems A.1 and B.1 control the gradient error uniformly over Ω from one training fold, the same event $\mathcal{E}_{\mathrm{tr}}$ covers every iterate ω_t along the (possibly adaptive) trajectory, so the statistical error does not compound with T. These are upper bounds derived from Theorems A.1 and B.1, not a matching lower bound for every possible plug-in-based procedure; we make no claim that they improve on existing sample-complexity guarantees for stochastic bilevel algorithms that redraw fresh data at every outer step, which is a different sampling regime from the fixed-sample, cross-fitted estimator analyzed here.

C Efficient Influence Function

Throughout, $L^2(P)$ denotes the space of real-valued functions of (X, Y, Z) that are square-integrable under P, equipped with inner product $\langle f, f' \rangle_{L^2(P)} = \mathbb{E}_P[ff']$ and induced norm $\|f\|^2_{L^2(P)} = \mathbb{E}_P[f^2]$; $L^1(P)$ denotes its integrable counterpart, with norm $\|f\|_{L^1(P)} = \mathbb{E}_P[|f|]$; and $L^2_0(P) \coloneqq \{S \in L^2(P) : \mathbb{E}_P[S] = 0\}$ denotes the mean-zero subspace of $L^2(P)$, the tangent space of scores of regular parametric submodels in the nonparametric model.

The efficient influence function of $\nabla\mathcal{F}(\omega)$, obtained via influence function calculus (von Mises, 1947; Hampel, 1974; Fernholz, 1983), provides a principled correction for this bias (Newey, 1994; Chernozhukov et al., 2018). We make this precise below.

The target functional $\nabla\mathcal{F}(\omega)$ depends on P through two infinite-dimensional nuisances, the inner solution $h^\star_{\omega,P}$ and its derivative $j^\star_{k,\omega,P}$, so the plug-in estimator suggested by Section 2 inherits their estimation error at first order. Correcting this bias requires understanding how the target, written locally in this appendix as $\Psi_{k,\omega}(P) \coloneqq \partial_{\omega_k}\mathcal{F}_P(\omega)$, responds to small changes in P. Since P cannot be perturbed directly, semiparametric efficiency theory instead considers every smooth one-parameter family of distributions $(P_\epsilon)_\epsilon$ passing through P at $\epsilon = 0$, called a *parametric submodel*. A canonical example is the linear submodel $\mathrm{d}P_\epsilon \coloneqq (1 + \epsilon S)\,\mathrm{d}P$, for any bounded, mean-zero function S of the data: since $\mathbb{E}_P[S] = 0$, this integrates to one for every ϵ, and for ϵ small enough that $1 + \epsilon S > 0$ everywhere, it defines a valid probability distribution with $P_0 = P$. Here S is called the *score* of this submodel: since $\mathrm{d}P_\epsilon = (1 + \epsilon S)\,\mathrm{d}P$, observations where S is positive become more likely as ϵ increases, and observations where S is negative become less likely. The *efficient influence function* $\chi_{P,\omega}$ is the function such that, for every submodel with score S,

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Psi_{k,\omega}(P_\epsilon)\bigg|_{\epsilon=0} = \langle \chi_{k,P,\omega}, S\rangle_{L^2(P)}\,.$$

This equation is what lets us fix the plug-in bias: adding $\chi_{k,P,\omega}$ to a plug-in estimate of $\Psi_{k,\omega}$ cancels its first-order sensitivity to nuisance estimation error, which is exactly the correction we build in Section 3.2. Requiring this to hold for every perturbation direction S is what makes $\Psi_{k,\omega}$ *pathwise differentiable* at P, with $\chi_{k,P,\omega}$ its efficient influence function (Bickel et al., 1993; Newey, 1994; van der Vaart, 1998; Tsiatis, 2006; Kosorok, 2008). We derive $\chi_{P,\omega}$ below by directly differentiating $\Psi_{k,\omega}$ along parametric submodels, which requires the nuisances $h^\star_{\omega,P}$ and $j^\star_{k,\omega,P}$ to vary smoothly along them. Rather than assuming this outright, the argument below shows it is already guaranteed by Assumption A: since these nuisances are themselves conditional expectations of the bounded quantities $g_\omega(Z)$ and $\partial_{\omega_k} g_\omega(Z)$ under our quadratic losses, and conditional expectations of bounded quantities vary smoothly along any regular parametric submodel, differentiability holds automatically.

The linear submodel $\mathrm{d}P_\epsilon \coloneqq (1+\epsilon S)\,\mathrm{d}P$ is defined for any bounded, mean-zero S: since $\mathbb{E}_P[S] = 0$ it integrates to one for every ϵ, and for $|\epsilon| < 1/\|S\|_\infty$, so that $1 + \epsilon S > 0$ everywhere, it is a valid probability measure with $P_0 = P$. This is a regular parametric submodel in the formal (differentiable-in-quadratic-mean) sense of the semiparametric efficiency literature (Bickel et al., 1993; van der Vaart, 1998), with score S.

Define $a^\star_{\omega,P}(X) := m^\star_P(X) - h^\star_{\omega,P}(X)$.

Theorem C.1 (Efficient influence function)**.** *Under Assumption A, the functional $\Psi_{k,\omega}$ is pathwise differentiable at P with efficient influence function*

$$\chi_{k,P,\omega}(X,Y,Z) = -\langle Y - g_\omega(Z),\, j^\star_{k,\omega,P}(X)\rangle - \langle \partial_{\omega_k} g_\omega(Z),\, a^\star_{\omega,P}(X)\rangle + \langle j^\star_{k,\omega,P}(X),\, a^\star_{\omega,P}(X)\rangle - \Psi_{k,\omega}(P). \tag{75}$$

Consequently, $\chi_{P,\omega} \coloneqq (\chi_{1,P,\omega}, \ldots, \chi_{d,P,\omega})^\top$ is the vector-valued efficient influence function for $\nabla\mathcal{F}(\omega)$ in the nonparametric model.

The proof follows below; the accompanying functional von Mises expansion in Appendix D quantifies the effect of replacing oracle nuisances by candidates. In brief, once the correction terms in (75) are subtracted from the plug-in score, the residual has no first-order dependence on the nuisance estimation errors $h - h^\star_{\omega,P}$ and $v - j^\star_{k,\omega,P}$; only cross-products of correction- and nuisance-estimation errors and a second-order remainder survive (Equations (77) and (78) in Appendix D). Equation (75) directly suggests the debiased pseudo-outcome cross-fitted in Appendix E.

Lemma C.1 (Score identity). *For every bounded, mean-zero S and every $f \in L^2(P)$,*

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathbb{E}_{P_\epsilon}[f]\Big|_{\epsilon=0} = \mathbb{E}_P[fS].$$

Proof. By definition of P_ϵ and $fS \in L^1(P)$ (Cauchy–Schwarz, as $f, S \in L^2(P)$),

$$\mathbb{E}_{P_\epsilon}[f] = \int f(1+\epsilon S)\,\mathrm{d}P = \mathbb{E}_P[f] + \epsilon\,\mathbb{E}_P[fS]$$

for every ϵ, exactly linear in ϵ; differentiating at $\epsilon = 0$ gives the claim. □

Taking $f \equiv 1$ recovers $\mathbb{E}_P[S] = 0$, consistent with S being mean-zero by construction. Lemma C.1 justifies the score identity used at the start of the proof of Theorem C.1 below, without needing to assume differentiation under the integral sign as a separate step.

Proposition C.1 (Weak differentiability of conditional expectations). *Let S be bounded and mean-zero, let $f = f(Z) \in \mathbb{R}^q$ be a random vector with $\|f(Z)\| \le M < \infty$ almost surely, and write $\mu_\epsilon(X) \coloneqq \mathbb{E}_{P_\epsilon}[f(Z) \mid X] \in \mathbb{R}^q$ and $\mu \coloneqq \mu_0$. Then $\epsilon \mapsto \mu_\epsilon$ is weakly differentiable in $L^2(P_X)$ at $\epsilon = 0$, meaning*

$$\left\langle \frac{\mu_\epsilon - \mu}{\epsilon},\, b \right\rangle_{L^2(P_X)} \xrightarrow{\epsilon\to 0} \langle \dot\mu,\, b\rangle_{L^2(P_X)} \qquad \textit{for every bounded } b \in L^2(P_X),$$

with $\dot\mu(X) = \mathbb{E}_P[(f(Z) - \mu(X))S \mid X] \in \mathbb{R}^q$.

Proof. Write $\mu(X) = \mathbb{E}_P[f(Z) \mid X]$, $a(X) \coloneqq \mathbb{E}_P[f(Z)S \mid X]$, and $c(X) \coloneqq \mathbb{E}_P[S \mid X]$; none of these depend on ϵ. By Bayes' rule, since $\mathrm{d}P_\epsilon = (1+\epsilon S)\,\mathrm{d}P$,

$$\mu_\epsilon(X) = \frac{\mathbb{E}_P[f(Z)(1+\epsilon S) \mid X]}{\mathbb{E}_P[1+\epsilon S \mid X]} = \frac{\mu(X) + \epsilon\, a(X)}{1 + \epsilon\, c(X)}$$

is (for a.e. X) an ordinary ratio of two affine functions of ϵ, with $|c(X)| \le \|S\|_\infty$, so the denominator is bounded away from zero, uniformly in X, for $|\epsilon| < 1/(2\|S\|_\infty)$. Subtracting $\mu(X)$ and putting the result over this common denominator gives

$$\mu_\epsilon(X) - \mu(X) = \frac{\epsilon\big(a(X) - \mu(X)c(X)\big)}{1 + \epsilon\, c(X)},$$

so, by the quotient rule, $\epsilon \mapsto \mu_\epsilon(X)$ is differentiable at $\epsilon = 0$ for a.e. X, with $\dot\mu(X) = a(X) - \mu(X)c(X) = \mathbb{E}_P[(f(Z) - \mu(X))S \mid X]$, using that $\mu(X)$ is X-measurable.

Since $\|f(Z)\| \le M$, we have $\|a(X)\| \le M\|S\|_\infty$ and $\|\mu(X)\| \le M$ (Jensen's inequality), so $\|a(X) - \mu(X)c(X)\| \le 2M\|S\|_\infty$, and the display above gives, still for $|\epsilon| < 1/(2\|S\|_\infty)$,

$$\left\| \frac{\mu_\epsilon(X) - \mu(X)}{\epsilon} \right\| \le \frac{2M\|S\|_\infty}{1 - |\epsilon|\|S\|_\infty} \le 4M\|S\|_\infty =: C,$$

a bound uniform in X and in ϵ near 0. Fix a bounded $b \in L^2(P_X)$. Since $\frac{\mu_\epsilon(X)-\mu(X)}{\epsilon} \to \dot\mu(X)$ pointwise as $\epsilon \to 0$ and $\big\|\langle \frac{\mu_\epsilon - \mu}{\epsilon}, b\rangle\big\| \le C\|b(X)\|$, with $C\|b\| \in L^1(P)$, dominated convergence gives

$$\left\langle \frac{\mu_\epsilon - \mu}{\epsilon},\, b \right\rangle_{L^2(P_X)} = \mathbb{E}_P\left[\left\langle \frac{\mu_\epsilon(X) - \mu(X)}{\epsilon},\, b(X)\right\rangle\right] \longrightarrow \mathbb{E}_P[\langle \dot\mu(X), b(X)\rangle] = \langle \dot\mu, b\rangle_{L^2(P_X)}.$$

□

Proposition C.2 (Differentiability of the nuisances). *Under Assumption A, along every submodel in this family, the maps $\epsilon \mapsto h^\star_{\omega,P_\epsilon}$ and $\epsilon \mapsto j^\star_{k,\omega,P_\epsilon}$ are weakly differentiable at $\epsilon = 0$ in $L^2(P_X)$: this is not an independent restriction on the submodel beyond Assumption A.*

Proof. Since $h^\star_{\omega,P}(X) = \mathbb{E}_P[g_\omega(Z) \mid X]$ and $j^\star_{k,\omega,P}(X) = \mathbb{E}_P[\partial_{\omega_k} g_\omega(Z) \mid X]$ are themselves conditional expectations of the bounded quantities $g_\omega(Z)$ and $\partial_{\omega_k} g_\omega(Z)$ under the quadratic specialization, Proposition C.1 applies with $f = g_\omega$ and $f = \partial_{\omega_k} g_\omega$, both $\mathbb{R}^q$-valued and bounded in norm under Assumption A, giving the pathwise derivatives

$$\begin{aligned} \dot{h}(X) &:= \frac{\mathrm{d}}{\mathrm{d}\epsilon} h^\star_{\omega,P_\epsilon}(X)\Big|_{\epsilon=0} = \mathbb{E}_P\big[(g_\omega(Z) - h^\star_{\omega,P}(X))S \mid X\big], \\ \dot{j}_k(X) &:= \frac{\mathrm{d}}{\mathrm{d}\epsilon} j^\star_{k,\omega,P_\epsilon}(X)\Big|_{\epsilon=0} = \mathbb{E}_P\big[(\partial_{\omega_k} g_\omega(Z) - j^\star_{k,\omega,P}(X))S \mid X\big], \end{aligned} \tag{76}$$

in the weak sense of Proposition C.1, which is the only sense in which they are used below. □

Proof of Theorem C.1. Fix $k \in \{1, \dots, d\}$ and a bounded, mean-zero S, and let $(P_\epsilon)_\epsilon$ be the corresponding linear submodel defined above. Recall from Lemma C.1 the score identity, $\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathbb{E}_{P_\epsilon}[f]|_{\epsilon=0} = \mathbb{E}_P[fS]$ for every $f \in L^2(P)$.

Our goal is to find $\chi_{k,P,\omega} \in L^2_0(P)$ such that

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Psi_{k,\omega}(P_\epsilon)\Big|_{\epsilon=0} = \langle \chi_{k,P,\omega}, S\rangle_{L^2(P)}, \qquad \forall \text{ bounded, mean-zero } S.$$

As recalled above, this already uniquely determines $\chi_{k,P,\omega}$ over the whole nonparametric tangent space $L^2_0(P)$, which is the efficient influence function.

Decomposition. Recall that $\Psi_{k,\omega}(P_\epsilon) = \mathbb{E}_{P_\epsilon}\big[\langle h^\star_{\omega,P_\epsilon}(X) - Y,\, j^\star_{k,\omega,P_\epsilon}(X)\rangle\big]$. There are three sources of ϵ-dependence: the measure P_ϵ, the inner solution $h^\star_{\omega,P_\epsilon}$, and the Jacobian $j^\star_{k,\omega,P_\epsilon}$. By the product rule,

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}\epsilon}\Psi_{k,\omega}(P_\epsilon)\Big|_{\epsilon=0} = {} & \underbrace{\frac{\mathrm{d}}{\mathrm{d}\epsilon}\mathbb{E}_{P_\epsilon}\big[\langle h^\star_{\omega,P}(X) - Y,\, j^\star_{k,\omega,P}(X)\rangle\big]\Big|_{\epsilon=0}}_{\text{(A): measure perturbation}} \\ & + \underbrace{\mathbb{E}_P\big[\langle \dot{h}(X),\, j^\star_{k,\omega,P}(X)\rangle\big]}_{\text{(B): effect on } h^\star_{\omega,P_\epsilon}} \\ & + \underbrace{\mathbb{E}_P\big[\langle h^\star_{\omega,P}(X) - Y,\, \dot{j}_k(X)\rangle\big]}_{\text{(C): effect on } j^\star_{k,\omega,P_\epsilon}}. \end{aligned}$$

Term (A): measure perturbation. By the score identity,

$$A = \mathbb{E}_P\big[\langle h^\star_{\omega,P}(X) - Y,\, j^\star_{k,\omega,P}(X)\rangle\, S\big].$$

This is the plug-in term: the direct contribution from the measure shifting.

Term (B): effect of $h^\star_{\omega,P_\epsilon}$ changing. Substituting $\dot{h}$ from (76) and applying the tower property, since $j^\star_{k,\omega,P}(X)$ is X-measurable,

$$B = \mathbb{E}_P\big[\langle \mathbb{E}_P[(g_\omega(Z) - h^\star_{\omega,P}(X))S \mid X],\, j^\star_{k,\omega,P}(X)\rangle\big] = \mathbb{E}_P\big[\langle g_\omega(Z) - h^\star_{\omega,P}(X),\, j^\star_{k,\omega,P}(X)\rangle\, S\big].$$

Term (C): effect of $j^\star_{k,\omega,P_\epsilon}$ changing. Substituting $\dot{j}_k$ from (76) and applying the tower property twice, first to replace $h^\star_{\omega,P}(X) - Y$ by its X-conditional mean $\mathbb{E}_P[h^\star_{\omega,P}(X) - Y \mid X] = h^\star_{\omega,P}(X) - m^\star_P(X) = -a^\star_{\omega,P}(X)$, and then to the remaining X-conditional term,

$$\begin{aligned} C &= \mathbb{E}_P\big[\langle h^\star_{\omega,P}(X) - Y,\, \mathbb{E}_P[(\partial_{\omega_k} g_\omega(Z) - j^\star_{k,\omega,P}(X))S \mid X]\rangle\big] \\ &= \mathbb{E}_P\big[\langle -a^\star_{\omega,P}(X),\, (\partial_{\omega_k} g_\omega(Z) - j^\star_{k,\omega,P}(X))S\rangle\big] \\ &= -\mathbb{E}_P\big[\langle \partial_{\omega_k} g_\omega(Z),\, a^\star_{\omega,P}(X)\rangle\, S\big] + \mathbb{E}_P\big[\langle j^\star_{k,\omega,P}(X),\, a^\star_{\omega,P}(X)\rangle\, S\big]. \end{aligned}$$

Combining and centering. The $h^\star_{\omega,P}(X)$ terms in A and B cancel, giving

$$A + B = \mathbb{E}_P\big[\langle g_\omega(Z) - Y,\, j^\star_{k,\omega,P}(X)\rangle\, S\big] = -\mathbb{E}_P\big[\langle Y - g_\omega(Z),\, j^\star_{k,\omega,P}(X)\rangle\, S\big].$$

Adding C,

$$\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Psi_{k,\omega}(P_\epsilon)\bigg|_{\epsilon=0} = A + B + C = \mathbb{E}_P\big[\tilde{\chi}_{k,P,\omega}\, S\big],$$

where

$$\tilde{\chi}_{k,P,\omega} = -\langle Y - g_\omega(Z),\, j^\star_{k,\omega,P}(X)\rangle - \langle \partial_{\omega_k} g_\omega(Z),\, a^\star_{\omega,P}(X)\rangle + \langle j^\star_{k,\omega,P}(X),\, a^\star_{\omega,P}(X)\rangle.$$

Since $\mathbb{E}_P[S] = 0$, we may subtract any constant from $\tilde{\chi}_{k,P,\omega}$ without changing the inner product with S. Subtracting $\Psi_{k,\omega}(P)$ centers the function and ensures $\mathbb{E}_P[\chi_{k,P,\omega}] = 0$, giving the canonical gradient $\chi_{k,P,\omega} = \tilde{\chi}_{k,P,\omega} - \Psi_{k,\omega}(P)$, matching (75). By Assumption A, $\chi_{k,P,\omega} \in L^2_0(P)$, and the identity $\frac{\mathrm{d}}{\mathrm{d}\epsilon}\Psi_{k,\omega}(P_\epsilon)\big|_{\epsilon=0} = \langle \chi_{k,P,\omega}, S\rangle_{L^2(P)}$ holds for every bounded, mean-zero S, since S was arbitrary above. This identifies $\chi_{k,P,\omega}$ as the unique element of $L^2_0(P)$ representing this derivative relative to the nonparametric tangent space, hence the efficient influence function. Stacking coordinates, $\chi_{P,\omega} = (\chi_{1,P,\omega}, \ldots, \chi_{d,P,\omega})^\top$ is the vector-valued efficient influence function for $\nabla\mathcal{F}(\omega)$. □

D Functional von Mises Expansion

This appendix proves the functional von Mises expansion underlying Theorem C.1. It quantifies the error made when the oracle nuisances are replaced by candidate functions, and shows explicitly why the corrected score (75) is orthogonal. Throughout, write $\psi_{k,P}(h, v) \coloneqq \mathbb{E}_P[\langle h(X) - Y, v(X)\rangle]$, so that $\Psi_{k,\omega}(P) \coloneqq \partial_{\omega_k}\mathcal{F}_P(\omega) = \psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P})$ by the direct form of Section 2.

Theorem D.1 (Functional von Mises expansion)**.** *Under Assumption A, for each $k = 1, \ldots, d$ there exists a neighborhood $\mathcal{N}_k$ of $(h^\star_{\omega,P}, j^\star_{k,\omega,P})$ such that, for every $(h, v) \in \mathcal{N}_k$ and every $\alpha_{1,k}, \alpha_2 \in L^2(P_X)$,*

$$\begin{aligned}
&\psi_{k,P}(h, v) - \psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P}) - \mathbb{E}_P\big[\langle h(X) - g_\omega(Z), \alpha_{1,k}(X)\rangle\big] + \mathbb{E}_P\big[\langle \partial_{\omega_k} g_\omega(Z), \alpha_2(X)\rangle\big] \\
&\quad - \mathbb{E}_P\big[\langle v(X), \alpha_2(X)\rangle\big] \\
=&\langle j^\star_{k,\omega,P} - \alpha_{1,k}, h - h^\star_{\omega,P}\rangle_{L^2(P_X)} - \langle a^\star_{\omega,P} + \alpha_2, v - j^\star_{k,\omega,P}\rangle_{L^2(P_X)} \\
&\quad + \mathrm{Rem}_k(h - h^\star_{\omega,P}, v - j^\star_{k,\omega,P}),
\end{aligned} \tag{77}$$

where

$$|\mathrm{Rem}_k(h - h^\star_{\omega,P}, v - j^\star_{k,\omega,P})| \le \frac{1}{2}\left(\|h - h^\star_{\omega,P}\|^2_{L^2(P_X)} + \|v - j^\star_{k,\omega,P}\|^2_{L^2(P_X)}\right). \tag{78}$$

Equation (77) is the key orthogonality identity: once the correction terms are subtracted from the plug-in score, the residual has no first-order dependence on the nuisance estimation errors $h - h^\star_{\omega,P}$ and $v - j^\star_{k,\omega,P}$. The only remaining terms are cross-products of the correction errors $\|j^\star_{k,\omega,P} - \alpha_{1,k}\|_{L^2(P_X)}$ and $\|a^\star_{\omega,P} + \alpha_2\|_{L^2(P_X)}$ with nuisance-estimation errors $\|h - h^\star_{\omega,P}\|_{L^2(P_X)}$ and $\|v - j^\star_{k,\omega,P}\|_{L^2(P_X)}$, plus a second-order remainder controlled by (78). The proof below establishes this expansion.

Proposition D.1. *Under the quadratic specialization, for each $k = 1, \ldots, d$, the map $(h, v) \mapsto \psi_{k,P}(h, v)$ is Fréchet differentiable on all of $L^2(P_X) \times L^2(P_X)$, with globally Lipschitz Fréchet derivative of constant* 1.

Proof. Fix $k \in \{1, \ldots, d\}$. By definition, $\psi_{k,P}(h, v) = \mathbb{E}_P[\langle h(X) - Y, v(X)\rangle]$. We can write this as

$$\psi_{k,P}(h, v) = -\mathbb{E}_P[\langle Y, v(X)\rangle] + \mathbb{E}_P[\langle h(X), v(X)\rangle],$$

which is affine in h and linear in v.

Fréchet differentiability. For any $\delta h \in L^2(P_X)$,

$$\psi_{k,P}(h+\delta h, v) - \psi_{k,P}(h,v) = \mathbb{E}_P[\langle \delta h(X), v(X)\rangle] =: \partial_1 \psi_{k,P}(h,v)[\delta h],$$

with remainder zero, so $\psi_{k,P}$ is exactly linear in h. The functional $\partial_1 \psi_{k,P}(h,v)$ is bounded since by Cauchy-Schwarz,

$$|\partial_1 \psi_{k,P}(h,v)[\delta h]| \le \|v\|_{L^2(P_X)} \|\delta h\|_{L^2(P_X)}.$$

For any $\delta v \in L^2(P_X)$,

$$\psi_{k,P}(h, v+\delta v) - \psi_{k,P}(h,v) = -\mathbb{E}_P[\langle Y - h(X), \delta v(X)\rangle] =: \partial_2 \psi_{k,P}(h,v)[\delta v],$$

with remainder zero, so $\psi_{k,P}$ is exactly linear in v. The functional $\partial_2 \psi_{k,P}(h,v)$ is bounded since by Cauchy-Schwarz,

$$|\partial_2 \psi_{k,P}(h,v)[\delta v]| \le \|Y - h(X)\|_{L^2(P)} \|\delta v\|_{L^2(P_X)}.$$

Since both partial Fréchet derivatives exist and are bounded, $\psi_{k,P}$ is Fréchet differentiable on all of $L^2(P_X) \times L^2(P_X)$.

Lipschitz derivative. For any $(h_1, v_1), (h_2, v_2) \in L^2(P_X) \times L^2(P_X)$, by Cauchy-Schwarz,

$$\|\partial_1 \psi_{k,P}(h_1,v_1) - \partial_1 \psi_{k,P}(h_2,v_2)\|_{\mathrm{op}} \le \|v_1 - v_2\|_{L^2(P_X)},$$
$$\|\partial_2 \psi_{k,P}(h_1,v_1) - \partial_2 \psi_{k,P}(h_2,v_2)\|_{\mathrm{op}} \le \|h_1 - h_2\|_{L^2(P_X)}.$$

Combining,

$$\|D\psi_{k,P}(h_1,v_1) - D\psi_{k,P}(h_2,v_2)\|_{\mathrm{op}} \le \|(h_1 - h_2, v_1 - v_2)\|_{L^2(P_X)\times L^2(P_X)},$$

so the Fréchet derivative is globally Lipschitz with constant 1. □

Lemma D.1 (Coordinatewise Taylor expansion)**.** *Fix $k \in \{1, \dots, d\}$. Under the quadratic specialization, for all $h, v \in L^2(P_X)$,*

$$\begin{aligned}\psi_{k,P}(h^\star_{\omega,P} + h, j^\star_{k,\omega,P} + v) &- \psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P}) \\ &= \partial_1 \psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P})[h] + \partial_2 \psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P})[v] + \mathrm{Rem}_k(h,v),\end{aligned}$$

where

$$|\mathrm{Rem}_k(h,v)| \le \frac{1}{2}\left(\|h\|^2_{L^2(P_X)} + \|v\|^2_{L^2(P_X)}\right).$$

Proof. Let $z^\star = (h^\star_{\omega,P}, j^\star_{k,\omega,P})$ and $r = (h, v)$. By Proposition D.1, $\psi_{k,P}$ is Fréchet differentiable on all of $L^2(P_X) \times L^2(P_X)$ with globally Lipschitz derivative of constant 1. By the integral form of Taylor's theorem,

$$\psi_{k,P}(z^\star + r) - \psi_{k,P}(z^\star) = \int_0^1 D\psi_{k,P}(z^\star + tr)[r]\,\mathrm{d}t.$$

Adding and subtracting $D\psi_{k,P}(z^\star)[r]$ gives

$$\psi_{k,P}(z^\star + r) - \psi_{k,P}(z^\star) = D\psi_{k,P}(z^\star)[r] + \int_0^1 \{D\psi_{k,P}(z^\star + tr) - D\psi_{k,P}(z^\star)\}[r]\,\mathrm{d}t,$$

where the first term equals

$$D\psi_{k,P}(z^\star)[r] = \partial_1 \psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P})[h] + \partial_2 \psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P})[v].$$

For the remainder, the Lipschitz bound from Proposition D.1 gives

$$\begin{aligned}\left|\int_0^1 \{D\psi_{k,P}(z^\star + tr) - D\psi_{k,P}(z^\star)\}[r]\,\mathrm{d}t\right| &\le \int_0^1 \|D\psi_{k,P}(z^\star + tr) - D\psi_{k,P}(z^\star)\|_{\mathrm{op}}\|r\|\,\mathrm{d}t \\ &\le \int_0^1 t\|r\|^2\,\mathrm{d}t = \frac{1}{2}\|r\|^2,\end{aligned}$$

so

$$|\mathrm{Rem}_k(h,v)| \le \frac{1}{2}\|(h,v)\|^2_{L^2(P_X)\times L^2(P_X)} \le \frac{1}{2}\left(\|h\|^2_{L^2(P_X)} + \|v\|^2_{L^2(P_X)}\right).$$

□

Proof of Theorem D.1. Fix $k \in \{1, \dots, d\}$.

Step 1: First-order expansion of $\psi_{k,P}$. By Lemma D.1, applied at the base point $(h^\star_{\omega,P}, j^\star_{k,\omega,P})$,

$$\begin{aligned}&\psi_{k,P}(h,v) - \psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P})\\ =&\partial_1\psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P})[h - h^\star_{\omega,P}] + \partial_2\psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P})[v - j^\star_{k,\omega,P}]\\ &+ \mathrm{Rem}_k(h - h^\star_{\omega,P}, v - j^\star_{k,\omega,P}),\end{aligned}$$

where

$$|\mathrm{Rem}_k(h - h^\star_{\omega,P}, v - j^\star_{k,\omega,P})| \le \frac{1}{2}(\|h - h^\star_{\omega,P}\|^2_{L^2(P_X)} + \|v - j^\star_{k,\omega,P}\|^2_{L^2(P_X)}).$$

By the identities established above, the first-order terms equal $\langle j^\star_{k,\omega,P}, h - h^\star_{\omega,P}\rangle_{L^2(P_X)}$ and $-\langle a^\star_{\omega,P}, v - j^\star_{k,\omega,P}\rangle_{L^2(P_X)}$ respectively, so

$$\begin{aligned}&\psi_{k,P}(h,v) - \psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P})\\ &= \langle j^\star_{k,\omega,P}, h - h^\star_{\omega,P}\rangle_{L^2(P_X)} - \langle a^\star_{\omega,P}, v - j^\star_{k,\omega,P}\rangle_{L^2(P_X)} + \mathrm{Rem}_k(h - h^\star_{\omega,P}, v - j^\star_{k,\omega,P}).\end{aligned} \tag{79}$$

Step 2: Expanding the correction terms. We now express each correction term in (77) in terms of $\langle\cdot,\cdot\rangle_{L^2(P_X)}$.

First correction: $\mathbb{E}_P[\langle h(X) - g_\omega(Z), \alpha_{1,k}(X)\rangle]$. By bilinearity of the inner product, pointwise $\langle h(X) - g_\omega(Z), \alpha_{1,k}(X)\rangle = \langle h^\star_{\omega,P}(X) - g_\omega(Z), \alpha_{1,k}(X)\rangle + \langle h(X) - h^\star_{\omega,P}(X), \alpha_{1,k}(X)\rangle$, so taking expectations,

$$\mathbb{E}_P\big[\langle h(X) - g_\omega(Z), \alpha_{1,k}(X)\rangle\big] = \mathbb{E}_P\big[\langle h^\star_{\omega,P}(X) - g_\omega(Z), \alpha_{1,k}(X)\rangle\big] + \langle h - h^\star_{\omega,P}, \alpha_{1,k}\rangle_{L^2(P_X)}.$$

Since $h^\star_{\omega,P}(X) = \mathbb{E}_P[g_\omega(Z) \mid X]$, the first term vanishes by the tower property, giving

$$\mathbb{E}_P\big[\langle h(X) - g_\omega(Z), \alpha_{1,k}(X)\rangle\big] = \langle h - h^\star_{\omega,P}, \alpha_{1,k}\rangle_{L^2(P_X)}. \tag{80}$$

Second correction: $\mathbb{E}_P[\langle \partial_{\omega_k} g_\omega(Z), \alpha_2(X)\rangle]$. This does not involve h at all, as the formula shows. Differentiating the first-order condition $\mathbb{E}_P[\langle h^\star_{\omega,P}(X) - g_\omega(Z), u(X)\rangle] = 0$ (for every $u \in L^2(P_X)$) with respect to ω_k, and using $j^\star_{k,\omega,P} = \partial_{\omega_k} h^\star_{\omega,P}$, gives

$$\mathbb{E}_P\big[\langle \partial_{\omega_k} g_\omega(Z), \alpha_2(X)\rangle\big] = \langle j^\star_{k,\omega,P}, \alpha_2\rangle_{L^2(P_X)}. \tag{81}$$

Third correction: $\mathbb{E}_P[\langle v(X), \alpha_2(X)\rangle]$. This already equals $\langle v, \alpha_2\rangle_{L^2(P_X)}$ exactly, by definition of the inner product on $L^2(P_X)$.

Step 3: Assembling the expansion. Substituting (80) and (81) into $\psi_{k,P}(h,v) - \psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P}) - \mathbb{E}_P[\langle h(X) - g_\omega(Z), \alpha_{1,k}(X)\rangle] + \mathbb{E}_P[\langle \partial_{\omega_k} g_\omega(Z), \alpha_2(X)\rangle] - \mathbb{E}_P[\langle v(X), \alpha_2(X)\rangle]$ via (79), and using symmetry of $\langle\cdot,\cdot\rangle_{L^2(P_X)}$,

$$\begin{aligned}&\psi_{k,P}(h,v) - \psi_{k,P}(h^\star_{\omega,P}, j^\star_{k,\omega,P}) - \mathbb{E}_P\big[\langle h(X) - g_\omega(Z), \alpha_{1,k}(X)\rangle\big] + \mathbb{E}_P\big[\langle \partial_{\omega_k} g_\omega(Z), \alpha_2(X)\rangle\big]\\ &\quad - \mathbb{E}_P\big[\langle v(X), \alpha_2(X)\rangle\big]\\ =&\langle j^\star_{k,\omega,P}, h - h^\star_{\omega,P}\rangle_{L^2(P_X)} - \langle \alpha_{1,k}, h - h^\star_{\omega,P}\rangle_{L^2(P_X)}\\ &\quad - \langle a^\star_{\omega,P}, v - j^\star_{k,\omega,P}\rangle_{L^2(P_X)} + \langle j^\star_{k,\omega,P}, \alpha_2\rangle_{L^2(P_X)} - \langle v, \alpha_2\rangle_{L^2(P_X)} + \mathrm{Rem}_k(h - h^\star_{\omega,P}, v - j^\star_{k,\omega,P})\\ =&\langle j^\star_{k,\omega,P} - \alpha_{1,k}, h - h^\star_{\omega,P}\rangle_{L^2(P_X)} - \langle a^\star_{\omega,P} + \alpha_2, v - j^\star_{k,\omega,P}\rangle_{L^2(P_X)} + \mathrm{Rem}_k(h - h^\star_{\omega,P}, v - j^\star_{k,\omega,P}),\end{aligned}$$

where the last line combines, using symmetry of $\langle\cdot,\cdot\rangle_{L^2(P_X)}$,

$$\begin{aligned}&- \langle a^\star_{\omega,P}, v - j^\star_{k,\omega,P}\rangle_{L^2(P_X)} + \langle j^\star_{k,\omega,P}, \alpha_2\rangle_{L^2(P_X)} - \langle v, \alpha_2\rangle_{L^2(P_X)}\\ =&- \langle a^\star_{\omega,P}, v - j^\star_{k,\omega,P}\rangle_{L^2(P_X)} - \langle \alpha_2, v - j^\star_{k,\omega,P}\rangle_{L^2(P_X)}.\end{aligned}$$

□

E ASYMPTOTIC NORMALITY

For a sequence of random variables $(X_N)_{N\geq 1}$ and a deterministic sequence of positive reals $(a_N)_{N\geq 1}$, we write $X_N = o_p(a_N)$ if $X_N/a_N \to 0$ in probability as $N \to \infty$, i.e., $P(|X_N/a_N| > \varepsilon) \to 0$ for every $\varepsilon > 0$; and $X_N = O_p(a_N)$ if X_N/a_N is bounded in probability, i.e., for every $\varepsilon > 0$ there exists $M < \infty$ such that $\limsup_{N\to\infty} P(|X_N/a_N| > M) < \varepsilon$.

For any nuisance tuple η, define the centered score

$$\chi_{\eta,\omega}(X, Y, Z) \coloneqq \varphi_\omega(\eta) - \mathbb{E}_P\big[\varphi_\omega(\eta)\big].$$

Two-fold cross-fitted doubly robust estimator. We now construct a $\sqrt{N}$-consistent estimator for $\nabla\mathcal{F}(\omega)$. The construction follows the standard cross-fitting principle: estimate the nuisance functions on one fold, evaluate the orthogonal score on the held-out fold, and average (Bickel et al., 1993; Newey, 1994; Chernozhukov et al., 2018). Let S_1 and S_2 be two independent folds of i.i.d. observations from P, each of size n, and set $N = 2n$. Write $P_{n,r}$ for the empirical measure on S_r, $r = 1, 2$. For each fold r, let $\hat\eta_\omega^{(-r)} = (\hat h_\omega^{(-r)}, \hat j_\omega^{(-r)}, \hat m^{(-r)})$ denote nuisance estimates trained on the opposite fold S_{3-r}, where $\hat j_\omega^{(-r)} = (\hat j_{1,\omega}^{(-r)}, \dots, \hat j_{d,\omega}^{(-r)}) \in L^2(P_X)^d$ and $\hat m^{(-r)}$ is an estimate of $m_P^\star$ trained on S_{3-r}. For any nuisance tuple $\eta = (h, j, m)$, let $\varphi_\omega(\eta) = (\varphi_{1,\omega}(\eta), \dots, \varphi_{d,\omega}(\eta))^\top$. Consequently, the two-fold cross-fitted doubly robust estimator is

$$\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) \coloneqq \frac{1}{2}\sum_{r=1}^{2} \mathbb{E}_{P_{n,r}}\big[\varphi_\omega(\hat\eta_\omega^{(-r)})\big]. \tag{82}$$

Under the conditions below, $\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)$ is asymptotically efficient at fixed ω; we establish its $\sqrt{N}$ limit and the corresponding pointwise Wald intervals.

Since $\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)$ is an empirical average, its asymptotic distribution is obtained via a central limit theorem. For this to hold, the estimated score $\chi_{\hat\eta_\omega^{(-r)},\omega}$ must be close enough to the oracle $\chi_{P,\omega}$, and the nuisance estimation errors must vanish at a sufficient rate. The following assumption formalizes these requirements.

Assumption I (Nuisance rates)**.** For $r = 1, 2$,

$$\|\hat h_\omega^{(-r)} - h_{\omega,P}^\star\|_{L^2(P_X)} = o_p(N^{-1/4}), \qquad \|\hat j_\omega^{(-r)} - j_{\omega,P}^\star\|_{L^2(P_X)^d} = o_p(N^{-1/4}),$$
$$\|\hat m^{(-r)} - m_P^\star\|_{L^2(P_X)} = o_p(N^{-1/4}).$$

These rates can be verified by standard least-squares regression arguments under appropriate approximation and entropy or Rademacher-complexity conditions; see Appendix B.1.

Remark E.1 (Linear-plus-product score decomposition). Write $\eta_\omega^\star = (h_{\omega,P}^\star, j_{\omega,P}^\star, m_P^\star)$ and $\chi_{P,\omega} = \chi_{\eta_\omega^\star,\omega}$. For fold r, set $\Delta h_r = \hat h_\omega^{(-r)} - h_{\omega,P}^\star$, $\Delta j_r = \hat j_\omega^{(-r)} - j_{\omega,P}^\star$, $\Delta m_r = \hat m^{(-r)} - m_P^\star$, and $\Delta r_r = \Delta h_r - \Delta m_r$. Expanding the quadratic score gives exactly

$$\varphi_{k,\omega}(\hat\eta_\omega^{(-r)}) - \varphi_{k,\omega}(\eta_\omega^\star) = \underbrace{\langle g_\omega(Z) - Y - h_{\omega,P}^\star(X) + m_P^\star(X), \Delta j_{k,r}(X)\rangle}_{L_{k,r}^{(1)}}$$
$$+ \underbrace{\langle \partial_{\omega_k} g_\omega(Z) - j_{k,\omega,P}^\star(X), \Delta r_r(X)\rangle}_{L_{k,r}^{(2)}} - \underbrace{\langle \Delta j_{k,r}(X), \Delta r_r(X)\rangle}_{Q_{k,r}}.$$

The coefficients of the first two terms have conditional mean zero given X; under Assumption A, their norms are bounded by $2(A + B)$ and $2D$, respectively. Thus, holding the training fold fixed, $PL_{k,r}^{(1)} = PL_{k,r}^{(2)} = 0$ and

$$\|L_{k,r}^{(1)}\|_{L^2(P)} \leq 2(A + B)\|\Delta j_{k,r}\|_{L^2(P_X)}, \qquad \|L_{k,r}^{(2)}\|_{L^2(P)} \leq 2D\|\Delta r_r\|_{L^2(P_X)}.$$

The product term only needs an $L^1(P)$ bound:

$$P|Q_{k,r}| \leq \|\Delta j_{k,r}\|_{L^2(P_X)}\|\Delta r_r\|_{L^2(P_X)}$$
$$\leq \|\Delta j_r\|_{L^2(P_X)^d}(\|\Delta h_r\|_{L^2(P_X)} + \|\Delta m_r\|_{L^2(P_X)}) = o_p(N^{-1/2})$$

by Assumption I. This linear-plus-product decomposition suffices for the empirical remainder; it does not assert $L^2(P)$ stability of the product or of the full fitted score.

Let $\Sigma_{P,\omega} := \mathbb{E}_P[\chi_{P,\omega}\chi_{P,\omega}^\top] \in \mathbb{R}^{d\times d}$ denote the covariance matrix of the oracle efficient influence function. The following theorem provides the asymptotic normality of the cross-fitted debiased estimator.

Theorem E.1 (Asymptotic normality of the cross-fitted debiased estimator)**.** *Under Assumptions A and I,*

$$\sqrt{N}\left(\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega)\right) = \frac{1}{\sqrt{N}}\sum_{r=1}^{2}\sum_{(X_i,Y_i,Z_i)\in S_r}\chi_{P,\omega}(X_i,Y_i,Z_i) + o_p(1). \tag{83}$$

Consequently,

$$\sqrt{N}\left(\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega)\right) \xrightarrow{d} \mathcal{N}(0,\Sigma_{P,\omega}).$$

The expansion (83) separates the statistical error into three parts: the oracle empirical average, a stochastic equicontinuity term, and a population bias remainder controlled by Assumption I. The correction terms make the estimating function first-order insensitive to nuisance perturbations at the oracle nuisances, so only product-rate and quadratic remainders remain.

Proof of Theorem E.1. For $r = 1, 2$, write $\hat{\eta}_r := \hat{\eta}_\omega^{(-r)}$ for brevity. By definition of the two-fold estimator and since $\mathbb{E}_P[\varphi_\omega(\cdot;\eta_\omega^\star)] = \nabla\mathcal{F}(\omega)$,

$$\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega) = \frac{1}{2}\sum_{r=1}^{2}\left\{\mathbb{E}_{P_{n,r}}\left[\varphi_\omega(\cdot;\hat{\eta}_r)\right] - \mathbb{E}_P\left[\varphi_\omega(\cdot;\eta_\omega^\star)\right]\right\}.$$

Adding and subtracting $\mathbb{E}_P[\varphi_\omega(\cdot;\hat{\eta}_r)]$ and $\mathbb{E}_{P_{n,r}}[\varphi_\omega(\cdot;\eta_\omega^\star)]$ for each fold r yields the decomposition

$$\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega) = A_N + B_N + R_N,$$

where

$$A_N := \frac{1}{2}\sum_{r=1}^{2}\left(\mathbb{E}_{P_{n,r}} - \mathbb{E}_P\right)\left[\chi_{P,\omega}\right],$$

$$B_N := \frac{1}{2}\sum_{r=1}^{2}\left(\mathbb{E}_{P_{n,r}} - \mathbb{E}_P\right)\left[\chi_{\hat{\eta}_r,\omega} - \chi_{P,\omega}\right],$$

$$R_N := \frac{1}{2}\sum_{r=1}^{2}\left\{\mathbb{E}_P\left[\varphi_\omega(\cdot;\hat{\eta}_r)\right] - \mathbb{E}_P\left[\varphi_\omega(\cdot;\eta_\omega^\star)\right]\right\}.$$

We show that $\sqrt{N}\,A_N \xrightarrow{d} \mathcal{N}(0,\Sigma_{P,\omega})$ and $B_N, R_N = o_p(N^{-1/2})$.

Term A_N. Since the two folds are independent and each has size $n = N/2$,

$$A_N = \frac{1}{N}\sum_{r=1}^{2}\sum_{(X_i,Y_i,Z_i)\in S_r}\chi_{P,\omega}(X_i,Y_i,Z_i).$$

Under Assumption A, each term of the efficient influence function (75) is square-integrable. Indeed, the almost sure bounds $\|Y\| \le A$, $\sup_\omega \|g_\omega(Z)\| \le B$, and $\sup_{\omega\in\Omega}\sqrt{\sum_{k=1}^{d}\|\partial_{\omega_k} g_\omega(Z)\|^2} \le D$ imply that $h_{\omega,P}^\star$ and $j_{k,\omega,P}^\star$ are bounded in $L^2(P_X)$, while $a_{\omega,P}^\star = m_P^\star - h_{\omega,P}^\star \in L^2(P_X)$ follows since it is a difference of two bounded functions. Together these imply $\mathbb{E}_P\left[\|\chi_{P,\omega}\|_2^2\right] < \infty$. Therefore the multivariate central limit theorem gives

$$\sqrt{N}\,A_N = \frac{1}{\sqrt{N}}\sum_{r=1}^{2}\sum_{(X_i,Y_i,Z_i)\in S_r}\chi_{P,\omega}(X_i,Y_i,Z_i) \xrightarrow{d} \mathcal{N}(0,\Sigma_{P,\omega}).$$

Term B_N. Fix $r \in \{1,2\}$ and condition on S_{3-r} (and any independent fitting randomness), so the nuisances are fixed and the evaluation fold is independent. Since empirical centering annihilates constants,

$$(P_{n,r} - P)(\chi_{\hat{\eta}_r,\omega} - \chi_{P,\omega}) = (P_{n,r} - P)\{\varphi_\omega(\hat{\eta}_r) - \varphi_\omega(\eta^\star_\omega)\}.$$

Use Remark E.1 and put $L_{k,r} = L^{(1)}_{k,r} + L^{(2)}_{k,r}$. Its conditional mean is zero, and

$$\mathbb{E}\left[\left|\sqrt{n}(P_{n,r} - P)L_{k,r}\right|^2 \,\middle|\, S_{3-r}\right] \le PL^2_{k,r} \le C\{\|\Delta j_{k,r}\|^2_{L^2(P_X)} + \|\Delta r_r\|^2_{L^2(P_X)}\} = o_p(1).$$

Conditional Chebyshev, followed by averaging the conditional probability bound truncated at one, gives $(P_{n,r} - P)L_{k,r} = o_p(n^{-1/2})$. For the product, use $q_{k,r} := P|Q_{k,r}| = o_p(N^{-1/2})$ and

$$|(P_{n,r} - P)Q_{k,r}| \le P_{n,r}|Q_{k,r}| + q_{k,r}, \qquad \mathbb{E}[P_{n,r}|Q_{k,r}| \mid S_{3-r}] = q_{k,r}.$$

For every $\varepsilon > 0$, conditional Markov gives

$$\mathbb{P}\{\sqrt{N}P_{n,r}|Q_{k,r}| > \varepsilon\} \le \mathbb{E}\Big[\min\{1, \sqrt{N}q_{k,r}/\varepsilon\}\Big] \longrightarrow 0,$$

because the bounded integrand converges to zero in probability. Hence $(P_{n,r} - P)Q_{k,r} = o_p(N^{-1/2})$, without an $L^2(P)$ assumption on $Q_{k,r}$. As d is fixed and $N = 2n$, summing coordinates and averaging the two folds gives $B_N = o_p(N^{-1/2})$.

Term R_N. We decompose the population remainder coordinatewise. For each $r = 1, 2$ and $k = 1, \ldots, d$, Theorem D.1 gives

$$\mathbb{E}_P\big[\varphi_\omega(\cdot; \hat{\eta}_r)\big] - \mathbb{E}_P\big[\varphi_\omega(\cdot; \eta^\star_\omega)\big] = C_r + D_r + E_r,$$

where $C_r = (C_{r,1}, \ldots, C_{r,d})^\top$, $D_r = (D_{r,1}, \ldots, D_{r,d})^\top$, $E_r = (E_{r,1}, \ldots, E_{r,d})^\top$, and

$$\begin{aligned}
C_{r,k} &:= \left\langle j^\star_{k,\omega,P} - \hat{j}^{(-r)}_{k,\omega}, \hat{h}^{(-r)}_\omega - h^\star_{\omega,P} \right\rangle_{L^2(P_X)}, \\
D_{r,k} &:= -\left\langle a^\star_{\omega,P} + \hat{h}^{(-r)}_\omega - \hat{m}^{(-r)}, \hat{j}^{(-r)}_{k,\omega} - j^\star_{k,\omega,P} \right\rangle_{L^2(P_X)}, \\
E_{r,k} &:= \mathrm{Rem}_k\left(\hat{h}^{(-r)}_\omega - h^\star_{\omega,P}, \hat{j}^{(-r)}_{k,\omega} - j^\star_{k,\omega,P}\right).
\end{aligned}$$

By the Cauchy-Schwarz inequality,

$$\begin{aligned}
\|C_r\| &\le \left\|\hat{j}^{(-r)}_\omega - j^\star_{\omega,P}\right\|_{L^2(P_X)^d} \left\|\hat{h}^{(-r)}_\omega - h^\star_{\omega,P}\right\|_{L^2(P_X)} = o_p(N^{-1/2}), \\
\|D_r\| &\le \left\|\hat{h}^{(-r)}_\omega - \hat{m}^{(-r)} + a^\star_{\omega,P}\right\|_{L^2(P_X)} \left\|\hat{j}^{(-r)}_\omega - j^\star_{\omega,P}\right\|_{L^2(P_X)^d} \\
&\le \left(\left\|\hat{h}^{(-r)}_\omega - h^\star_{\omega,P}\right\|_{L^2(P_X)} + \left\|\hat{m}^{(-r)} - m^\star_P\right\|_{L^2(P_X)}\right) \left\|\hat{j}^{(-r)}_\omega - j^\star_{\omega,P}\right\|_{L^2(P_X)^d} = o_p(N^{-1/2}),
\end{aligned}$$

using $a^\star_{\omega,P} = m^\star_P - h^\star_{\omega,P}$ and the triangle inequality for the second bound, by Assumption I. For the remainder, Lemma D.1 gives

$$\|E_r\| \le \frac{1}{2}\left(\left\|\hat{h}^{(-r)}_\omega - h^\star_{\omega,P}\right\|^2_{L^2(P_X)} + \left\|\hat{j}^{(-r)}_\omega - j^\star_{\omega,P}\right\|^2_{L^2(P_X)^d}\right) = o_p(N^{-1/2}),$$

by the first two lines of Assumption I. Summing over both folds gives $R_N = o_p(N^{-1/2})$.

Conclusion. Combining the three terms,

$$\sqrt{N}\left(\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega)\right) = \frac{1}{\sqrt{N}}\sum_{r=1}^{2}\sum_{(X_i,Y_i,Z_i)\in S_r} \chi_{P,\omega}(X_i, Y_i, Z_i) + o_p(1).$$

The central limit theorem applied to A_N and Slutsky's theorem give

$$\sqrt{N}\left(\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega) - \nabla\mathcal{F}(\omega)\right) \xrightarrow{d} \mathcal{N}(0, \Sigma_{P,\omega}).$$

□

Confidence intervals for fixed-parameter gradient coordinates. The asymptotic covariance $\Sigma_{P,\omega}$ is estimated by the empirical covariance of the cross-fitted pseudo-outcomes:

$$\widehat{\Sigma}_{P,\omega} \coloneqq \frac{1}{2}\sum_{r=1}^{2} P_{n,r}\left[\left(\varphi_\omega(\hat{\eta}_\omega^{(-r)}) - \widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)\right)\left(\varphi_\omega(\hat{\eta}_\omega^{(-r)}) - \widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)\right)^\top\right].$$

Under $\widehat{\Sigma}_{P,\omega} \xrightarrow{p} \Sigma_{P,\omega}$, for a coordinate k with $\Sigma_{P,\omega,kk} > 0$, an asymptotically valid $(1-\alpha)$ Wald confidence interval for the k-th coordinate of $\nabla\mathcal{F}(\omega)$ is

$$\left[[\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)]_k \pm z_{1-\alpha/2}\sqrt{\frac{\widehat{\Sigma}_{P,\omega,kk}}{N}}\right], \tag{84}$$

where $z_{1-\alpha/2}$ is the $(1-\alpha/2)$ quantile of the standard normal distribution. These are coordinatewise, pointwise intervals at a fixed parameter; they do not provide simultaneous projected-stationarity certificates or inference for an adaptively selected optimizer.

Remark E.2. Theorem E.1 and (84) concern a fixed, non-data-dependent ω. If ω is selected using the same observations, pointwise Wald validity does not follow from the uniform estimation-error bounds of Appendix A. A simple construction selects ω on independent data and performs inference on a fresh held-out sample, treating the selected value as conditionally fixed and verifying the inference conditions there. More general post-selection inference requires additional simultaneous or selective-inference arguments not developed here.

F Additional Experiment Details

This appendix gives the data-generating processes (DGPs), implementation details, and numerical outputs behind Section 4. We first give the matched optimization experiments, then retain the supplementary fixed-gradient, inference, regularization, and Gaussian root-estimation designs. Throughout, "Plug-in" (PI) denotes the direct plug-in gradient and OBiGrad denotes its orthogonal counterpart. The original inference DGPs, evaluation points, and basis learners are preserved; KBO and KBO+OBiGrad are added on the same datasets and folds.

Computation. All experiments were run locally on a CPU-only Apple MacBook Pro with an Apple M4 chip, 10 CPU cores, 24GB memory, macOS 15.0, and arm64 Darwin kernel. No external GPU, cluster, or remote compute was used. All scripts use fixed random seeds and write configuration files alongside the results.

Reporting conventions. Gradient errors are Euclidean RMSEs of $\widehat{\nabla\mathcal{F}}^{\mathrm{OBi}}(\omega)$ relative to the unregularized population target $\nabla\mathcal{F}(\omega)$. Wald coverage is coordinate-wise and averaged across gradient coordinates, using the confidence intervals (84). KBO results are reported relative to $\nabla\mathcal{F}(\omega)$, even though fixed-$\lambda$ KBO targets a regularized gradient $\nabla\mathcal{F}_\lambda(\omega)$. In KBO tables, "KBO total" is $\|\widehat{\nabla\mathcal{F}}_\lambda^{\mathrm{KBO}}(\omega) - \nabla\mathcal{F}(\omega)\|$, "KBO estimation" is $\|\widehat{\nabla\mathcal{F}}_\lambda^{\mathrm{KBO}}(\omega) - \nabla\mathcal{F}_\lambda(\omega)\|$, and "Reg. bias" is $\|\nabla\mathcal{F}_\lambda(\omega) - \nabla\mathcal{F}(\omega)\|$, so that by the triangle inequality

$$\|\widehat{\nabla\mathcal{F}}_\lambda^{\mathrm{KBO}}(\omega) - \nabla\mathcal{F}(\omega)\| \le \|\widehat{\nabla\mathcal{F}}_\lambda^{\mathrm{KBO}}(\omega) - \nabla\mathcal{F}_\lambda(\omega)\| + \|\nabla\mathcal{F}_\lambda(\omega) - \nabla\mathcal{F}(\omega)\|.$$

Monte Carlo error bars are 95% half-widths across complete simulated datasets, using 1.96 times the Monte Carlo standard error unless a simultaneous adjustment is stated. Coordinates within a replication are not treated as independent replications. Gradient RMSE uses the Euclidean norm; "Coord. RMSE" in the inference tables is the Euclidean RMSE divided by $\sqrt{d}$. Studentized quantiles are descriptive pooled-coordinate diagnostics. Nuisance sensitivity norms use the vector L^2 norm, so the reported product is a bound diagnostic rather than the observed signed bias.

The fixed-gradient, Wald, and Gaussian Bellman-root summaries pool the original-seed rerun and an independent rerun with the master seed increased by 104729. This gives 600 IV and 400 Bellman fixed-gradient replications, 1000 IV and 400 Bellman Wald replications, and 400 Bellman-root replications per N. The matched legacy KBO and scalar-IV-root tables use their original-seed reruns (300 IV, 200 Bellman, and 300 scalar-root replications); separate independent-seed outputs are

retained with the reproducibility files. No old finite-sample "population proxy" is used as ground truth: kernel population targets are computed by converged quadrature, conditional on the fixed RFF map where applicable. The displayed KBO decomposition is a triangle bound, not an additive decomposition of RMSE.

F.1 Population optimization with matched nuisance learning

The three experiments in Section 4 examine stationary-point estimation, population stationarity of a returned optimization iterate, and robustness to nonzero conditional residuals. They use the same simulated datasets for the three analyses. The DGPs and nuisance learners below differ from the supplementary Gaussian, basis-regression designs in Appendices F.2 and F.3.

IV DGP. Let $X \sim \mathrm{Unif}[-1,1]$, $e_1(X) = \sqrt{3}X$, and $e_2(X) = \sqrt{5}(3X^2-1)/2$. Define

$$j^\star(X) = e_1(X) + \tfrac{1}{2}e_2(X), \qquad r(X) = \tfrac{1}{2}e_1(X) - e_2(X), \qquad Z = j^\star(X) + U,$$

and $Y = 0.75Z + \delta r(X) + 0.25U + \varepsilon$. Independently of X and each other, U and ε are centered uniform variables with standard deviations 1 and 0.25. We take $g_\omega(Z) = \omega Z$. Thus $h^\star_\omega = \omega j^\star$, $m^\star = 0.75 j^\star + \delta r$, and $\mathbb{E}[j^\star r] = 0$. The population gradient is $1.25(\omega - 0.75)$ for both $\delta = 0$ and $\delta = 0.25$. The first-stage R^2 is $1.25/2.25 \simeq 0.556$. The case $\delta = 0$ is a triangular IV model with endogenous treatment; $\delta = 0.25$ is an approximate conditional-moment problem, since exact conditional fit no longer holds. The learner receives (X, Z, Y), without the polynomial basis, conditional means, or population coefficients.

Fitted Q-regression DGP. Draw $S \sim \mathrm{Unif}[-\sqrt{3}, \sqrt{3}]$ and $A \sim \mathrm{Bernoulli}(0.5)$ independently, and set

$$S' = 0.7S + 0.5A + \xi, \qquad R = \sin S + 0.5A + 0.25SA + \varepsilon_R, \qquad D = 0.8(\sin S', \cos S')^\top.$$

Here ξ, ε_R, and ε_Y are independent centered uniform noises with standard deviations 0.2, 0.1, and 0.1. With $X = (S, A)$, $g_\omega = R + D^\top\omega$, and $\omega^\star = (0.65, -0.45)^\top$, generate $Y = R + D^\top\omega^\star + \delta r(X) + \varepsilon_Y$. Writing $j^\star = \mathbb{E}[D \mid X]$, the residual is

$$r(X) = \frac{q(X) - j^\star(X)^\top c}{\|q - j^{\star\top} c\|_{L^2(P_X)}}, \quad q(S,A) = \sin(2S) + 0.5(A - 0.5), \quad c = \{\mathbb{E}[j^\star j^{\star\top}]\}^{-1}\mathbb{E}[j^\star q].$$

This construction preserves $\mathbb{E}[j^\star r] = 0$ and hence the optimizer. Population projection coefficients are used by the simulator and evaluator only. The conditional sensitivity is $0.8\,\mathrm{sinc}(\sqrt{3}\,0.2)(\sin\mu, \cos\mu)^\top$, where $\mu = 0.7S + 0.5A$ and $\mathrm{sinc}(u) = \sin(u)/u$. Population integrals use 160-node Gauss–Legendre quadrature over the state and exact summation over actions. This is one Bellman-regression problem, with bounded state and noise distributions, rather than a complete fitted Q-iteration algorithm.

Matched nuisance fits. Both learners use two-fold cross-fitting on the same N complete observations. Training-fold means and standard deviations standardize the covariates. NW uses a Gaussian kernel and three-fold training-only cross-validation over bandwidths $\{0.05, 0.075, 0.1, 0.15, 0.225, 0.3375, 0.5, 0.75, 1\}$. KBO uses an exact Gaussian kernel whose bandwidth is the median positive pairwise distance among at most 256 evenly indexed standardized training covariates, and the penalty grid

$$\Lambda = \{10^{-j}, 3 \cdot 10^{-j} : j = 2, \ldots, 7\} \cup \{10^{-1}\}.$$

Cross-validation minimizes prediction error for the observed inner targets: Z in IV and (R, D) in fitted Q-regression, with scales computed on training data. It does not use Y, population gradients, or the optimizer. All chosen hyperparameters are fixed in ω and shared within a learner pair. The outcome regression uses the same selected operator and training observations. No nuisance is refitted during outer optimization.

KBO specialization and orthogonal counterpart. For the affine inner signal $g_\omega = b + D^\top\omega$, the implemented kernel system is

$$[\hat{b}, \hat{J}, \hat{m}] = K_{\mathrm{eval,tr}}(K_{\mathrm{tr,tr}} + n_{\mathrm{tr}}\lambda I)^{-1}[b_{\mathrm{tr}}, D_{\mathrm{tr}}, Y_{\mathrm{tr}}].$$

Consequently $\hat{h}_\omega = \hat{b} + \hat{J}\omega$ and $\partial_\omega \hat{h}_\omega = \hat{J}$ exactly. The existing KBO gradient in this specialization is $P_{\mathrm{eval}}[\hat{j}(\hat{h}_\omega - Y)]$, the derivative of the empirical outer squared loss. Its matched counterpart is

$$P_{\mathrm{eval}}\left[\hat{j}(g_\omega - Y) + (D - \hat{j})(\hat{h}_\omega - \hat{m})\right].$$

We implement explicit sensitivities. The same score is obtained by differentiating $P_{\mathrm{eval}}[\frac{1}{2}(Y - \hat{h}_\omega)^2 + (g_\omega - \hat{h}_\omega)(\hat{h}_\omega - \hat{m})]$, holding the regression operator fixed. The additional cost is one outcome right-hand side and its predictions using the existing factorization; no extra observations are required. This construction does not equate a general KBO adjoint with $m - h$. Numerical tests verify native-gradient equivalence, sensitivity and corrected-objective derivatives, zero-correction parity, and oracle controls.

Conditional on training data, the plug-in score targets the gradient of the fitted outer objective. For deterministic population fits at penalty λ, its target is $\nabla \mathcal{F}_\lambda$; the orthogonal target differs from $\nabla \mathcal{F}$ by

$$\mathbb{E}[(j_\lambda - j^\star)\{(m_\lambda - m^\star) - (h_\lambda - h^\star)\}].$$

Thus fixed regularization can bias both methods. The mixed identity was checked against population integration for the saved main fits, with maximum discrepancy $6.2 \cdot 10^{-16}$. For RFF fits in the supplementary experiments, the corresponding system is $(\Phi_{\mathrm{tr}}^\top \Phi_{\mathrm{tr}} + n_{\mathrm{tr}} \lambda I)^{-1} \Phi_{\mathrm{tr}}^\top$ followed by evaluation features; all equivalences remain valid for a fixed feature realization.

Protocol and Monte Carlo comparisons. The four settings are IV with $\delta = 0, 0.25$ (cases $0, 1$), and fitted Q-regression with $\delta = 0, 0.25$ (cases $2, 3$). After three exploratory pilots, we fixed the confirmation protocol and ran master seeds 2026092511 and 2026092512, each with 400 replications per case and $N \in \{200, 400, 800, 1600\}$. Pooling gives 800 replications per cell and $12{,}800$ datasets. Tables 2–5 report the complete primary grid. Parentheses give pointwise Monte Carlo 95% half-widths; RMSE uncertainty uses the delta method on squared errors. The paired comparisons in Tables 6–9 use differences on the same dataset. The declared family contains 64 contrasts: two endpoints, two learners, four cases, and four sample sizes. Approximate simultaneous intervals use the normal critical value 3.3594 (Bonferroni correction). All pooled lower limits are positive. The separate seed suites agree in direction; their adjusted intervals are positive for all 64 root comparisons and 63 of 64 stationarity comparisons across the two suites. The confirmation replication counts were not adapted to the observed differences.

Experiment 1: stationary-point estimation. We average the two held-out gradient fields before solving the affine estimating equation. Roots are evaluated against the unique unregularized population optimizer. Across the primary grid, root RMSE reductions range from 35.5% to 69.8% for NW and from 5.2% to 34.9% for KBO. There were no singular-root failures, non-positive-definite fitted quadratic objectives, or roots outside $[-2, 2]^d$ for the primary cross-validated methods. The finite-sample oracle estimator uses true nuisances and the same evaluation observations; it is distinct from exact population-gradient descent.

Experiment 2: population optimization. Projected descent starts at zero, uses $\eta = 0.1$, and runs for 1000 updates in $\Omega = [-2, 2]^d$. The fitted field remains fixed. At each budget T, we select the first minimizer

$$\hat{t}_T = \min \operatorname*{argmin}_{0 \le t < T} \|G_\eta(\omega_t; \widehat{\nabla \mathcal{F}}(\omega_t))\|.$$

We report $\|G_\eta(\omega_{\hat{t}_T})\|$, population excess objective, and distance to $\omega^\star$. Selection is feasible and does not use the smallest population norm along the trajectory. Indeed, the proof of Proposition 2.1 bounds the minimum estimated mapping; nonexpansiveness gives, with $\Delta = \sup_\Omega \|\widehat{\nabla \mathcal{F}} - \nabla \mathcal{F}\|$,

$$\|G_\eta(\omega_{\hat{t}_T})\| \le 2\sqrt{\frac{\mathcal{F}(\omega_0) - F_{\inf}}{\eta T}} + 3\Delta.$$

For the two-fold average, the uniform field error is at most the average of the two foldwise uniform errors; this step requires no independence between foldwise fields. These observations connect the reported statistic to the theory, without claiming that prediction cross-validation verifies all of its nuisance-rate and entropy assumptions. The main affine population objectives are strongly convex,

and late-iteration and root errors are closely related. Correction reduces the final statistical stationarity error, not necessarily the number of iterations needed for a fixed estimated objective. No primary trajectories failed or ended on the boundary.

Experiment 3: nonzero residual and regularization. With $\delta = 0$, a common linear population smoother satisfies $m_\lambda = h_{\omega^\star,\lambda}$, which makes the corrected population root exact despite shrinkage. This does not make its entire gradient field exact. The $\delta = 0.25$ cases remove this alignment and retain improvement under the primary cross-validated protocol (Figure 4). Figure 5 reports the full penalty sweep at $N = 800$, and Figure 6 retains $\lambda_N = 0.05N^{-0.6}$. The decreasing-penalty root RMSE is lower after correction in all 16 case/sample-size cells. These are descriptive secondary comparisons. The full grid contains a near-reversal for residual fitted Q-regression at $N = 800$, $\lambda = 10^{-7}$: KBO RMSE is 0.0263857, versus 0.0263907 after correction. We do not claim uniform superiority over penalties.

A deterministic finite-kernel calculation isolates the bias mechanism. For $k(x, x') = e_1(x)e_1(x') + 0.25e_2(x)e_2(x')$, let $s_1 = (1 + \lambda)^{-1}$, $s_2 = 0.25/(0.25 + \lambda)$, and $w_i = 2s_i - s_i^2$. Then

$$\omega_{\mathrm{PI},\lambda} = \frac{0.75(s_1 + 0.25s_2) + 0.5\delta(s_1 - s_2)}{s_1^2 + 0.25s_2^2}, \qquad \omega_{\mathrm{OBi},\lambda} = 0.75 + \frac{0.5\delta(w_1 - w_2)}{w_1 + 0.25w_2}.$$

For nonzero δ, root displacement is first order in λ for PI and second order for OBiGrad (Figure 7). This is an analytic mechanism illustration, not a fitted convergence rate for the empirical experiments.

Additional checks and limitations. A nonlinear fitted Q-regression diagnostic uses $v(\omega) = (\omega_1, \omega_2 + \sin\omega_1)$ and transforms each fitted gradient by $Dv(\omega)^\top$. The reparameterized population Hessian has a negative eigenvalue of approximately -1.072 at $(\pi/2, 2)$; the common step satisfies $\eta L \leq 0.237$. Replaying the 6400 Bellman datasets, correction improves all 16 matched returned-stationarity comparisons and all 16 excess-objective comparisons under a separate approximate simultaneous family of 32 endpoints (Figure 8). The map is invertible, so this is not a multiple-basin benchmark or independent replication.

A Gaussian sensitivity study uses 2400 fresh datasets at $N = 200, 800$. Its separate adjusted comparisons support 15 of 16 matched root and stationarity improvements; residual Bellman KBO at $N = 800$ has a 2.1% root-RMSE reduction with an interval including zero. The Gaussian design is outside the bounded-data setting and is not pooled with the primary study. Exploratory higher-noise settings did not uniformly favor correction; all pilot configurations and outcomes are retained in the supplementary result files. Pointwise intervals support reduced uniform gradient error in 24 of 32 primary comparisons, with eight unresolved. Better root accuracy does not establish universally better field estimation or verify the general asymptotic exponents.

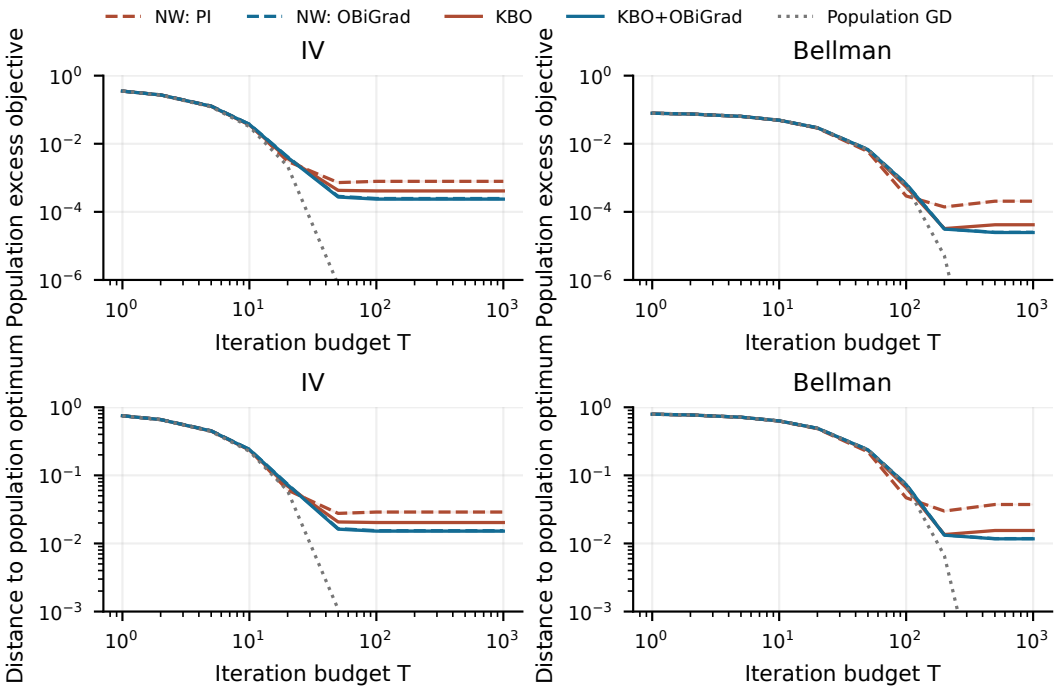


Figure 3: Population excess objective (top) and distance to the population optimizer (bottom) at the selected iterate, with $N = 400$ and $\delta = 0$. Bands give pointwise Monte Carlo 95% intervals. Population GD eventually falls below the plotted range.

Table 2: IV, $\delta = 0$: root RMSE and mean returned population mapping norm. Parentheses are pointwise 95% Monte Carlo half-widths; 800 replications per N.

N	Learner	PI root RMSE	OBi root RMSE	PI mapping	OBi mapping
200	NW	0.0511 (0.0023)	0.0313 (0.0020)	0.0516 (0.0026)	0.0297 (0.0018)
200	KBO	0.0444 (0.0026)	0.0317 (0.0021)	0.0430 (0.0024)	0.0299 (0.0018)
400	NW	0.0355 (0.0015)	0.0199 (0.0011)	0.0362 (0.0018)	0.0193 (0.0011)
400	KBO	0.0257 (0.0013)	0.0194 (0.0010)	0.0254 (0.0014)	0.0190 (0.0011)
800	NW	0.0251 (0.0010)	0.0124 (0.0006)	0.0260 (0.0012)	0.0122 (0.0007)
800	KBO	0.0153 (0.0007)	0.0123 (0.0006)	0.0153 (0.0008)	0.0122 (0.0007)
1600	NW	0.0200 (0.0007)	0.0085 (0.0004)	0.0213 (0.0009)	0.0085 (0.0004)
1600	KBO	0.0096 (0.0005)	0.0084 (0.0004)	0.0096 (0.0005)	0.0084 (0.0004)

Table 3: IV, $\delta = 0.25$: root RMSE and returned population mapping norm, with the same protocol as Table 2.

N	Learner	PI root RMSE	OBi root RMSE	PI mapping	OBi mapping
200	NW	0.0666 (0.0030)	0.0375 (0.0020)	0.0676 (0.0034)	0.0363 (0.0021)
200	KBO	0.0539 (0.0029)	0.0393 (0.0022)	0.0531 (0.0029)	0.0378 (0.0022)
400	NW	0.0464 (0.0020)	0.0239 (0.0012)	0.0476 (0.0023)	0.0236 (0.0013)
400	KBO	0.0309 (0.0015)	0.0241 (0.0012)	0.0308 (0.0016)	0.0239 (0.0013)
800	NW	0.0354 (0.0014)	0.0168 (0.0008)	0.0375 (0.0016)	0.0168 (0.0009)
800	KBO	0.0205 (0.0010)	0.0169 (0.0009)	0.0202 (0.0011)	0.0168 (0.0009)
1600	NW	0.0273 (0.0010)	0.0116 (0.0005)	0.0288 (0.0013)	0.0117 (0.0006)
1600	KBO	0.0129 (0.0006)	0.0115 (0.0005)	0.0127 (0.0007)	0.0116 (0.0006)

Table 4: Fitted Q-regression, $\delta = 0$: root RMSE and returned population mapping norm.

N	Learner	PI root RMSE	OBi root RMSE	PI mapping	OBi mapping
200	NW	0.0620 (0.0024)	0.0188 (0.0007)	0.0136 (0.0005)	0.0049 (0.0002)
200	KBO	0.0287 (0.0012)	0.0187 (0.0007)	0.0068 (0.0003)	0.0049 (0.0002)
400	NW	0.0412 (0.0012)	0.0132 (0.0004)	0.0092 (0.0003)	0.0035 (0.0001)
400	KBO	0.0177 (0.0007)	0.0131 (0.0004)	0.0043 (0.0002)	0.0035 (0.0001)
800	NW	0.0283 (0.0009)	0.0094 (0.0003)	0.0063 (0.0002)	0.0025 (0.0001)
800	KBO	0.0117 (0.0005)	0.0093 (0.0003)	0.0029 (0.0001)	0.0025 (0.0001)
1600	NW	0.0211 (0.0005)	0.0067 (0.0002)	0.0048 (0.0001)	0.0018 (0.0001)
1600	KBO	0.0077 (0.0003)	0.0067 (0.0002)	0.0019 (0.0001)	0.0018 (0.0001)

Table 5: Fitted Q-regression, $\delta = 0.25$: root RMSE and returned population mapping norm.

N	Learner	PI root RMSE	OBi root RMSE	PI mapping	OBi mapping
200	NW	0.0847 (0.0034)	0.0516 (0.0020)	0.0191 (0.0007)	0.0129 (0.0005)
200	KBO	0.0593 (0.0025)	0.0515 (0.0020)	0.0143 (0.0005)	0.0129 (0.0005)
400	NW	0.0590 (0.0022)	0.0380 (0.0014)	0.0135 (0.0005)	0.0097 (0.0003)
400	KBO	0.0412 (0.0016)	0.0379 (0.0014)	0.0102 (0.0004)	0.0096 (0.0003)
800	NW	0.0416 (0.0016)	0.0265 (0.0010)	0.0095 (0.0003)	0.0067 (0.0002)
800	KBO	0.0284 (0.0012)	0.0264 (0.0010)	0.0070 (0.0003)	0.0066 (0.0002)
1600	NW	0.0299 (0.0010)	0.0185 (0.0007)	0.0068 (0.0002)	0.0047 (0.0002)
1600	KBO	0.0195 (0.0007)	0.0185 (0.0007)	0.0049 (0.0002)	0.0047 (0.0002)

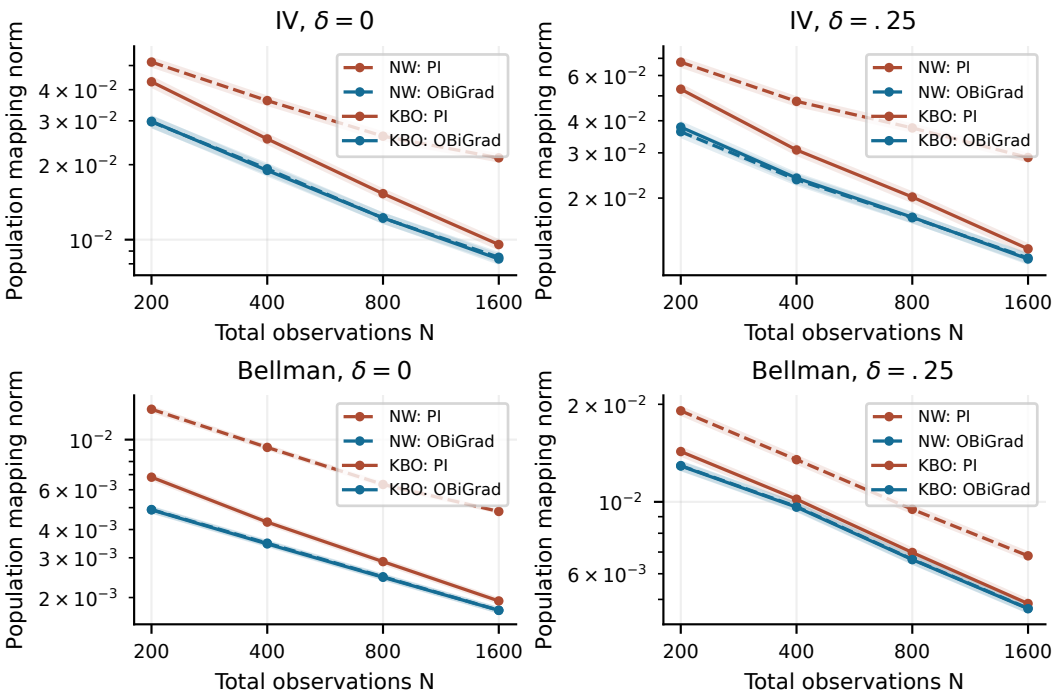


Figure 4: Returned population stationarity after 1000 updates across the complete primary grid. Bands give pointwise Monte Carlo 95% intervals. Residual cases are shown alongside aligned cases for comparison.

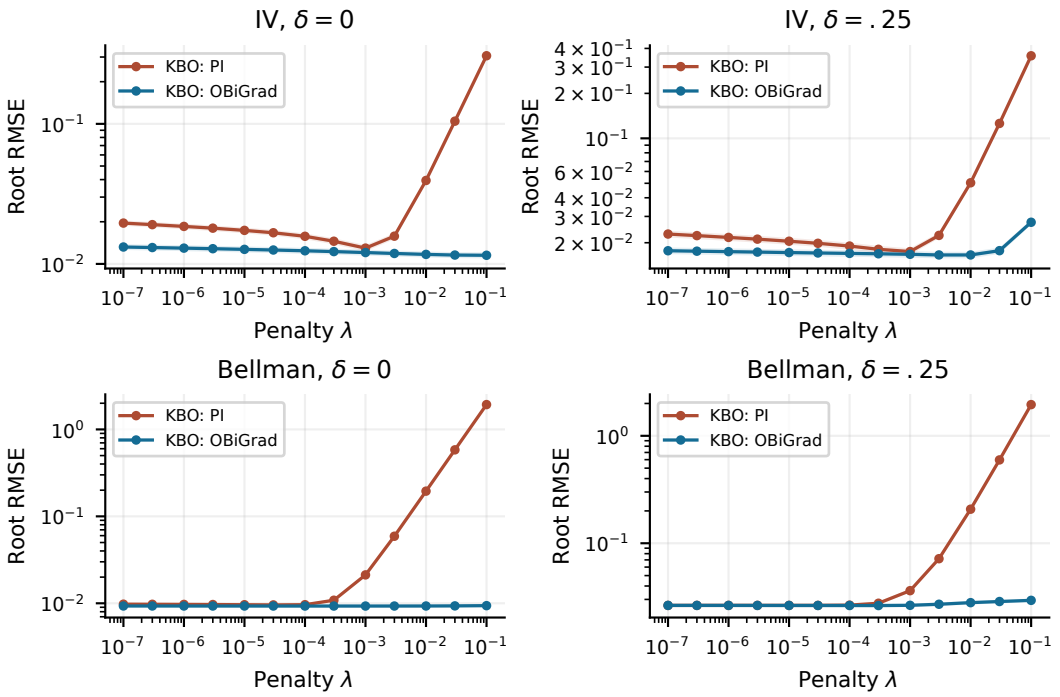


Figure 5: Matched KBO root RMSE over the complete penalty grid, at $N = 800$. Both methods use identical nuisance fits at each penalty. These curves are secondary diagnostics; the main comparison uses training-only cross-validation.

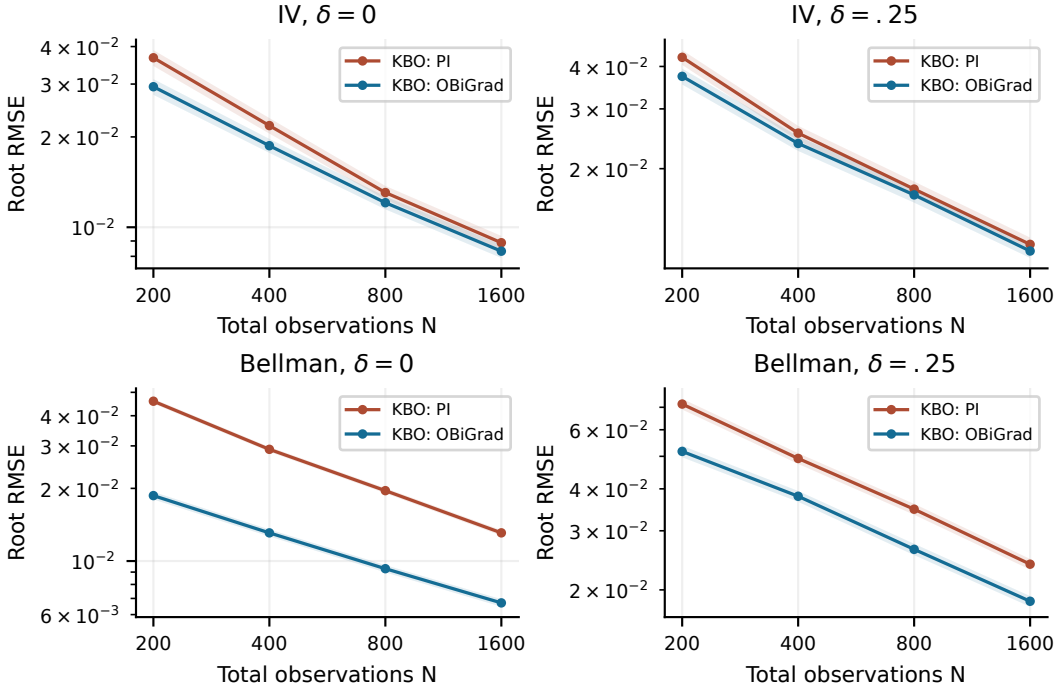


Figure 6: Matched KBO root RMSE with decreasing penalty $\lambda_N = 0.05N^{-0.6}$. Bands are pointwise Monte Carlo 95% intervals.

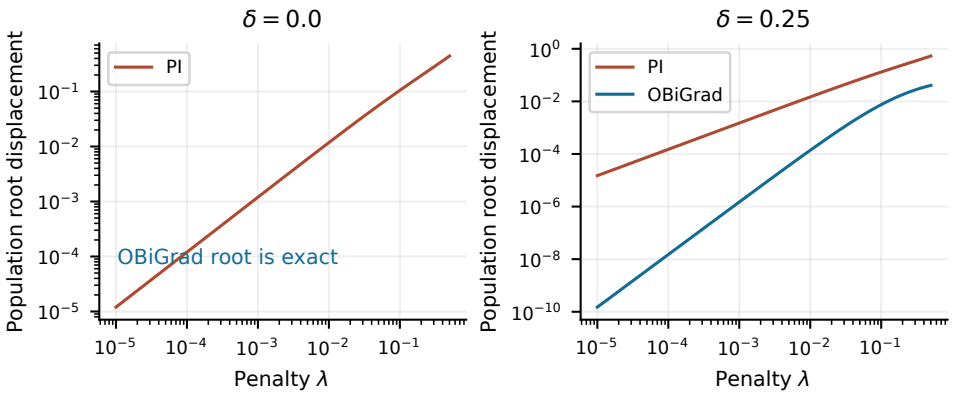


Figure 7: Analytic finite-kernel control: population root displacement under regularization. This deterministic calculation illustrates the first-order and product-order bias mechanisms.

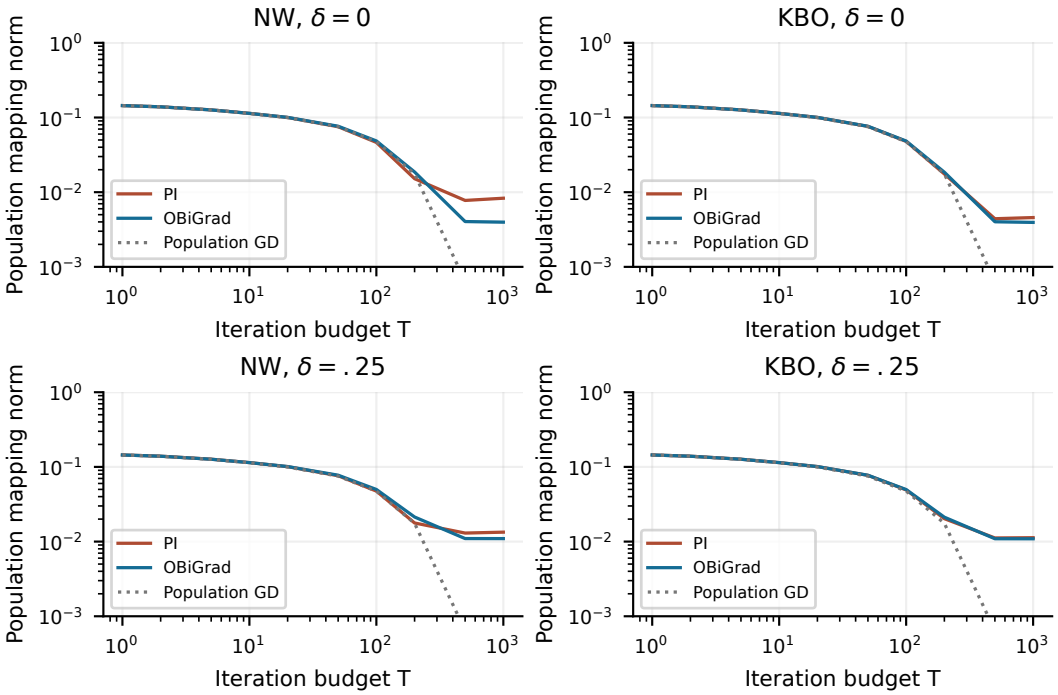


Figure 8: Nonlinear Bellman reparameterization: returned population mapping norm versus iteration budget, reusing the main Bellman samples with $N = 400$. Population truth is used only for evaluation.

Table 6: Paired primary contrasts, IV with $\delta = 0$. Positive differences favor correction. Intervals adjust over all 64 root-MSE and stationarity contrasts.

Case	N	Learner	Δ MSE	Adjusted interval	Δ mapping	Adjusted interval
0	200	NW	0.001626	[0.001232, 0.002021]	0.02192	[0.01706, 0.02678]
0	200	KBO	0.000970	[0.000662, 0.001279]	0.01317	[0.00963, 0.01671]
0	400	NW	0.000865	[0.000678, 0.001052]	0.01690	[0.01343, 0.02038]
0	400	KBO	0.000280	[0.000197, 0.000363]	0.00641	[0.00448, 0.00834]
0	800	NW	0.000476	[0.000384, 0.000568]	0.01381	[0.01137, 0.01626]
0	800	KBO	0.000081	[0.000053, 0.000109]	0.00307	[0.00201, 0.00414]
0	1600	NW	0.000328	[0.000275, 0.000381]	0.01280	[0.01101, 0.01459]
0	1600	KBO	0.000022	[0.000011, 0.000033]	0.00120	[0.00058, 0.00181]

Table 7: Paired primary contrasts, IV with $\delta = 0.25$; same adjustment as Table 6.

Case	N	Learner	Δ MSE	Adjusted interval	Δ mapping	Adjusted interval
1	200	NW	0.003034	[0.002378, 0.003690]	0.03137	[0.02540, 0.03733]
1	200	KBO	0.001359	[0.000928, 0.001791]	0.01528	[0.01109, 0.01947]
1	400	NW	0.001581	[0.001270, 0.001892]	0.02401	[0.01976, 0.02826]
1	400	KBO	0.000375	[0.000247, 0.000503]	0.00686	[0.00456, 0.00915]
1	800	NW	0.000973	[0.000808, 0.001138]	0.02073	[0.01768, 0.02378]
1	800	KBO	0.000132	[0.000082, 0.000181]	0.00335	[0.00204, 0.00466]
1	1600	NW	0.000610	[0.000515, 0.000705]	0.01708	[0.01478, 0.01939]
1	1600	KBO	0.000033	[0.000017, 0.000050]	0.00110	[0.00039, 0.00181]

Table 8: Paired primary contrasts, fitted Q-regression with $\delta = 0$.

Case	N	Learner	Δ MSE	Adjusted interval	Δ mapping	Adjusted interval
2	200	NW	0.003498	[0.002987, 0.004009]	0.00873	[0.00790, 0.00955]
2	200	KBO	0.000476	[0.000366, 0.000587]	0.00193	[0.00154, 0.00233]
2	400	NW	0.001521	[0.001346, 0.001696]	0.00576	[0.00526, 0.00626]
2	400	KBO	0.000142	[0.000105, 0.000179]	0.00086	[0.00064, 0.00108]
2	800	NW	0.000711	[0.000626, 0.000797]	0.00386	[0.00353, 0.00419]
2	800	KBO	0.000049	[0.000034, 0.000064]	0.00042	[0.00029, 0.00055]
2	1600	NW	0.000399	[0.000364, 0.000434]	0.00305	[0.00284, 0.00326]
2	1600	KBO	0.000014	[0.000009, 0.000019]	0.00018	[0.00011, 0.00025]

Table 9: Paired primary contrasts, fitted Q-regression with $\delta = 0.25$.

Case	N	Learner	Δ MSE	Adjusted interval	Δ mapping	Adjusted interval
3	200	NW	0.004507	[0.003655, 0.005360]	0.00616	[0.00509, 0.00723]
3	200	KBO	0.000871	[0.000588, 0.001154]	0.00140	[0.00095, 0.00185]
3	400	NW	0.002035	[0.001661, 0.002410]	0.00382	[0.00311, 0.00453]
3	400	KBO	0.000255	[0.000145, 0.000365]	0.00056	[0.00028, 0.00083]
3	800	NW	0.001033	[0.000838, 0.001228]	0.00283	[0.00233, 0.00333]
3	800	KBO	0.000113	[0.000066, 0.000160]	0.00034	[0.00019, 0.00050]
3	1600	NW	0.000549	[0.000459, 0.000640]	0.00212	[0.00177, 0.00247]
3	1600	KBO	0.000038	[0.000020, 0.000057]	0.00017	[0.00008, 0.00027]

F.2 Instrumental-variable experiments

These supplementary Gaussian designs (including Gaussian noise and, in some fitted-Q cases, unbounded features) fall outside the bounded-data conditions of Assumption A. They are robustness illustrations rather than literal instantiations of the finite-sample theory.

IV DGP. The IV experiments use a two-stage conditional-projection design, following the IV example of Section 2. Here X is the instrument, Z is the treatment, and the lower-level problem projects $g_\omega(Z)$ onto functions of X. We draw $X \sim \mathcal{N}(0, I_p)$ with $p = 3$, set

$$Z = 2\sum_{j=1}^{p} X_j + \eta, \qquad \eta \sim \mathcal{N}(0, \sigma_Z^2), \qquad \sigma_Z^2 = 0.1,$$

and define $\phi_\ell(z) = \sin(z + \ell)$, $\ell = 1, \dots, d$, with $d = 4$ and $g_\omega(Z) = \omega^\top \phi(Z)$. The structural parameter is $\omega^\star = (1, 2, 3, 4)^\top / \sqrt{30}$. For the gradient estimation and inference experiments, $Y = \omega^{\star\top}\phi(Z) + \varepsilon_Y$ with $\varepsilon_Y \sim \mathcal{N}(0, 0.25^2)$, which isolates gradient estimation and calibration. For the KBO regularization experiment, we use $Y = \omega^{\star\top}\phi(Z) + 0.5\eta$, which preserves $\mathbb{E}_P[Y \mid X]$ and hence $\nabla\mathcal{F}(\omega)$, but introduces correlation between the outcome noise and Z. We evaluate the population gradient at the fixed non-stationary point

$$\omega_0 = \omega^\star + 0.35\frac{(1, 1/3, -1/3, -1)^\top}{\|(1, 1/3, -1/3, -1)\|},$$

chosen away from $\omega^\star$ to measure gradient-estimation error at a non-stationary point.

The following analytic expressions are used only to compute the ground-truth population target $\nabla\mathcal{F}(\omega)$ for evaluation; they are not used by the estimator:

$$j_\ell^\star(X) = \mathbb{E}_P[\phi_\ell(Z) \mid X] = \exp(-\sigma_Z^2/2)\sin\left(2\sum_{j=1}^{p} X_j + \ell\right),$$

with $h_\omega^\star(X) = j^\star(X)^\top\omega$ and $m_P^\star(X) = j^\star(X)^\top\omega^\star$. Hence

$$\nabla\mathcal{F}(\omega) = A(\omega - \omega^\star), \qquad A_{k\ell} = \frac{\exp(-\sigma_Z^2)}{2}\left\{\cos(k-\ell) - \exp(-8p)\cos(k+\ell)\right\},$$

where the term $\exp(-8p)$ is negligible for $p = 3$. Feasible IV nuisance learners use ridge regression on Fourier features of $\sum_j X_j$, with frequencies $1, \ldots, 8$, intercept included, and ridge parameter 10^{-6}. Two-fold cross-fitting is used for the fixed-gradient, inference, and root experiments. The KBO experiment reports both matched two-fold estimates and the original one-direction split.

IV Experiment 1: fixed-gradient estimation. This experiment estimates $\nabla\mathcal{F}(\omega_0)$ for $N \in \{200, 400, 800, 1600, 3200\}$ over 600 replications, comparing the plug-in hypergradient, OBiGrad, and the oracle DR benchmark. The nuisance product proxy

$$\|\hat{j}_\omega - j_{\omega,P}^\star\|_{L^2(P_X)^d}\left(\|\hat{h}_\omega - h_{\omega,P}^\star\|_{L^2(P_X)} + \|\hat{m} - m_P^\star\|_{L^2(P_X)}\right)$$

is reported as the product-rate diagnostic corresponding to the exact mixed-bias bound in (10). Tables 10–11 and Figure 9 show that OBiGrad improves over the plug-in hypergradient at small N and approaches the oracle DR benchmark as the nuisance product decreases with N.

Table 10: Fixed-ω_0 IV gradient estimation. Parentheses report Monte Carlo 95% error bars for RMSE.

N	PI	OBiGrad	Oracle	OBi coverage	Product
200	0.0449 (0.0027)	0.0399 (0.0027)	0.0364 (0.0016)	0.947 (0.012)	0.0517
400	0.0257 (0.0010)	0.0253 (0.0010)	0.0245 (0.0010)	0.959 (0.011)	0.0217
800	0.0194 (0.0008)	0.0192 (0.0008)	0.0191 (0.0008)	0.956 (0.010)	0.0103
1600	0.0137 (0.0005)	0.0136 (0.0005)	0.0135 (0.0005)	0.949 (0.011)	0.0050
3200	0.0098 (0.0004)	0.0098 (0.0004)	0.0098 (0.0004)	0.941 (0.013)	0.0025

Table 11: Nuisance-learning diagnostics for the fixed-ω_0 IV experiment.

N	$\|\hat{h} - h^\star\|$	$\|\hat{j} - j^\star\|$	$\|\hat{m} - m^\star\|$	Product
200	0.0884	0.2053	0.1573	0.0517
400	0.0579	0.1329	0.1039	0.0217
800	0.0398	0.0917	0.0719	0.0103
1600	0.0276	0.0643	0.0495	0.0050
3200	0.0197	0.0454	0.0354	0.0025

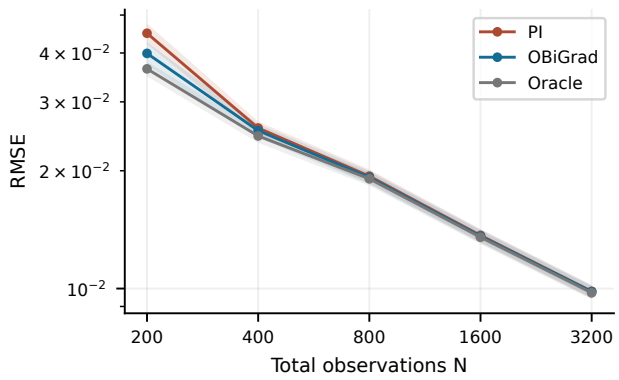


Figure 9: IV fixed-gradient estimation. OBiGrad tracks the oracle DR benchmark and improves over PI at smaller sample sizes.

IV Experiment 2: Wald calibration. This experiment uses the same vector sine-IV design and forms coordinate-wise 95% Wald confidence intervals (84) using the empirical covariance estimator. Results are averaged over 1000 replications. Tables 12–13 and Figures 10–11 report coverage, interval length, RMSE, oracle coverage (using the true nuisances $\eta_\omega^\star$), and studentized QQ diagnostics. OBiGrad coverage is approximately nominal across sample sizes, and the QQ diagnostic at $N = 3200$ is consistent with a Gaussian approximation. The implemented PI diagnostic intervals are longer and generally conservative, rather than systematically shorter and under-covering. These are marginal intervals for gradient coordinates at a fixed parameter, not confidence regions for a fitted stationary point. The four sine features span only two functions, so no nonsingular four-dimensional joint-inference claim is made.

Table 12: OBiGrad Wald calibration on the IV design.

N	Coverage	Length	Coord. RMSE	Product	Oracle cov.
200	0.942 (0.010)	0.0763 (0.0003)	0.0203 (0.0009)	0.0521	0.950 (0.009)
400	0.953 (0.009)	0.0531 (0.0001)	0.0132 (0.0004)	0.0218	0.959 (0.008)
800	0.954 (0.008)	0.0373 (4.2e-05)	0.0094 (0.0003)	0.0103	0.953 (0.008)
1600	0.952 (0.008)	0.0263 (2.1e-05)	0.0067 (0.0002)	0.0050	0.952 (0.008)
3200	0.948 (0.009)	0.0186 (1e-05)	0.0048 (0.0002)	0.0025	0.948 (0.010)

Table 13: Studentized-error diagnostics for the vector sine-IV inference experiment.

N	Mean	SD	2.5%	Median	97.5%	PI coverage
200	0.015	1.028	-1.983	0.011	2.045	0.940 (0.010)
400	0.005	0.978	-1.921	0.001	1.958	0.966 (0.007)
800	0.005	0.989	-1.880	-0.018	1.980	0.965 (0.007)
1600	0.027	0.999	-1.940	0.032	1.945	0.967 (0.007)
3200	-0.022	1.011	-2.015	-0.031	1.934	0.963 (0.008)

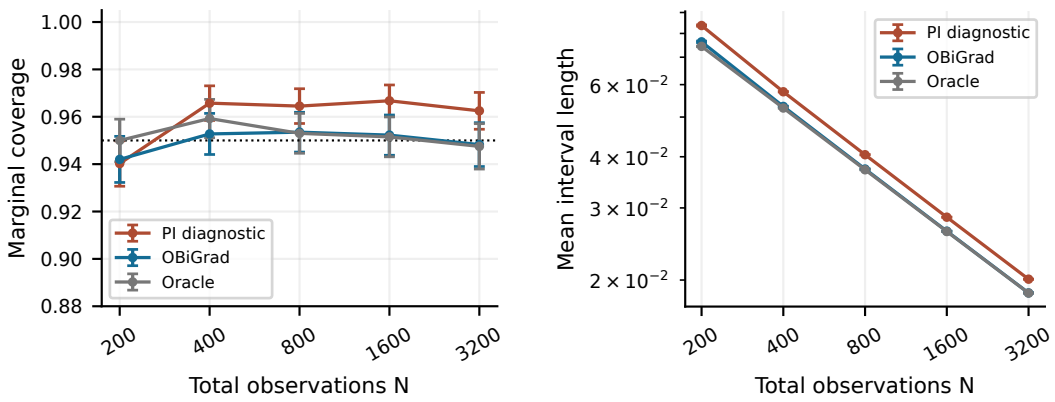


Figure 10: IV Wald diagnostics. Left: coordinate-wise coverage of nominal 95% intervals. Right: average interval length.

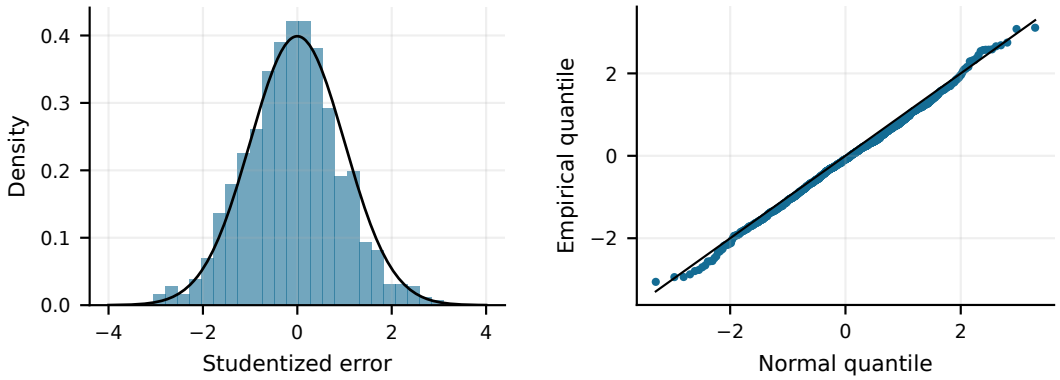


Figure 11: IV studentized OBiGrad errors for coordinate 0 at $N = 3200$.

Matched kernel Wald control. We also evaluated KBO and KBO+OBiGrad using the same IV inference samples and folds, the original Gaussian kernel on $X \in \mathbb{R}^3$ with bandwidth 0.5, and either $\lambda = 0.01$ or $\lambda_N = 0.05N^{-0.6}$. Both retain substantial calibration error in this design, so we do not use them as evidence of valid inference. Fixed regularization need not satisfy the nuisance-accuracy conditions of Theorem E.1, and the decreasing schedule does not establish these conditions at the displayed sample sizes. The complete coverage, length, and score diagnostics are retained in the supplementary result files; the Bellman extension below reports the same diagnostic protocol.

IV Experiment 3: KBO regularization bias. This experiment uses 600 training and 600 evaluation observations per direction ($N = 1200$ total), 300 replications, and the original Gaussian kernel $\exp(-\|x - x'\|^2/(2\sigma^2))$ with $\sigma = 0.5$. We retain the nine-penalty grid $\{10^{-5}, 3 \cdot 10^{-5}, 10^{-4}, 3 \cdot 10^{-4}, 10^{-3}, 3 \cdot 10^{-3}, 10^{-2}, 3 \cdot 10^{-2}, 10^{-1}\}$. Tables 14–15 compare the existing basis pair and the matched kernel pair using two-fold averages. Quadrature gives a distinct population target for each gradient construction. Small penalties can increase finite-sample error even as population bias decreases: at $\lambda = 10^{-5}$, KBO RMSE is 0.0768, versus 0.0846 for KBO+OBiGrad. Correction is beneficial at larger penalties, but does not uniformly improve this design. Figure 12 displays total error and population bias. The original one-direction protocol is retained in Table 16; within each protocol all methods use the same evaluation observations.

Table 14: Matched two-fold IV gradient RMSE at $N = 1200$. Parentheses report Monte Carlo 95% half-widths.

λ	PI	OBiGrad	KBO	KBO+OBiGrad	Oracle
1e-05	0.0137 (0.0008)	0.0134 (0.0008)	0.0768 (0.0015)	0.0846 (0.0019)	0.0133 (0.0008)
3e-05	0.0137 (0.0008)	0.0134 (0.0008)	0.0530 (0.0014)	0.0621 (0.0015)	0.0133 (0.0008)
0.0001	0.0137 (0.0008)	0.0134 (0.0008)	0.0478 (0.0013)	0.0508 (0.0014)	0.0133 (0.0008)
0.0003	0.0137 (0.0008)	0.0134 (0.0008)	0.0608 (0.0013)	0.0488 (0.0013)	0.0133 (0.0008)
0.001	0.0137 (0.0008)	0.0134 (0.0008)	0.0993 (0.0011)	0.0586 (0.0012)	0.0133 (0.0008)
0.003	0.0137 (0.0008)	0.0134 (0.0008)	0.1589 (0.0010)	0.0895 (0.0012)	0.0133 (0.0008)
0.01	0.0137 (0.0008)	0.0134 (0.0008)	0.2379 (0.0009)	0.1617 (0.0010)	0.0133 (0.0008)
0.03	0.0137 (0.0008)	0.0134 (0.0008)	0.3024 (0.0006)	0.2492 (0.0008)	0.0133 (0.0008)
0.1	0.0137 (0.0008)	0.0134 (0.0008)	0.3500 (0.0004)	0.3256 (0.0004)	0.0133 (0.0008)

Table 15: IV kernel population-bias norms and estimation RMSE around each method's own population target.

λ	KBO bias	KBO estimation	OBi bias	OBi estimation
1e-05	0.0034	0.0778 (0.0015)	0.0005	0.0842 (0.0019)
3e-05	0.0080	0.0530 (0.0013)	0.0013	0.0608 (0.0015)
0.0001	0.0196	0.0406 (0.0012)	0.0039	0.0471 (0.0014)
0.0003	0.0418	0.0339 (0.0012)	0.0103	0.0391 (0.0013)
0.001	0.0875	0.0271 (0.0011)	0.0288	0.0314 (0.0012)
0.003	0.1517	0.0209 (0.0009)	0.0682	0.0242 (0.0011)
0.01	0.2346	0.0142 (0.0007)	0.1497	0.0161 (0.0009)
0.03	0.3012	0.0087 (0.0005)	0.2444	0.0096 (0.0006)
0.1	0.3498	0.0043 (0.0003)	0.3248	0.0046 (0.0003)

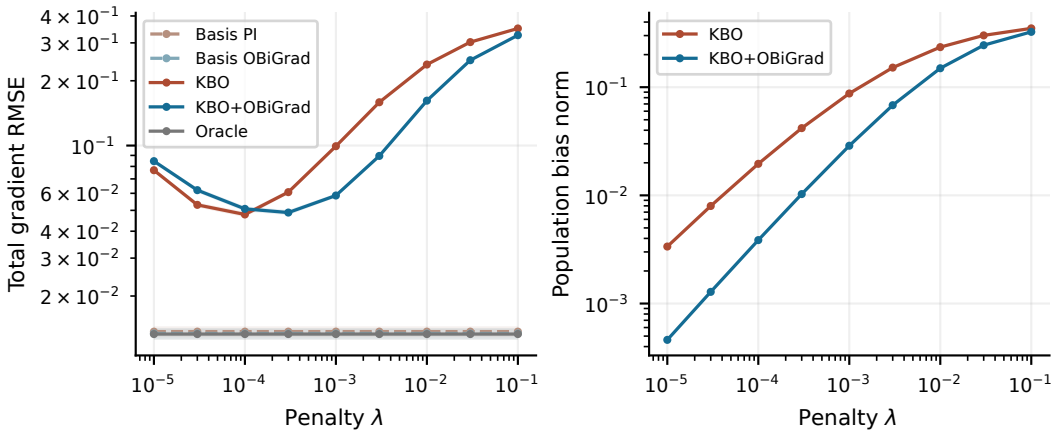


Figure 12: IV kernel diagnostics: matched total gradient RMSE (left) and population-bias norms (right).

Table 16: IV: original one-direction evaluation, with matched observations for all five estimators. Parentheses give Monte Carlo 95% half-widths.

λ	PI	OBiGrad	KBO	KBO+OBiGrad	Oracle
1e-05	0.0257 (0.0018)	0.0191 (0.0011)	0.0801 (0.0023)	0.0862 (0.0025)	0.0191 (0.0011)
3e-05	0.0257 (0.0018)	0.0191 (0.0011)	0.0567 (0.0022)	0.0634 (0.0020)	0.0191 (0.0011)
0.0001	0.0257 (0.0018)	0.0191 (0.0011)	0.0510 (0.0024)	0.0521 (0.0018)	0.0191 (0.0011)
0.0003	0.0257 (0.0018)	0.0191 (0.0011)	0.0624 (0.0024)	0.0499 (0.0017)	0.0191 (0.0011)
0.001	0.0257 (0.0018)	0.0191 (0.0011)	0.0995 (0.0021)	0.0591 (0.0016)	0.0191 (0.0011)
0.003	0.0257 (0.0018)	0.0191 (0.0011)	0.1584 (0.0018)	0.0894 (0.0014)	0.0191 (0.0011)
0.01	0.0257 (0.0018)	0.0191 (0.0011)	0.2373 (0.0013)	0.1615 (0.0011)	0.0191 (0.0011)
0.03	0.0257 (0.0018)	0.0191 (0.0011)	0.3021 (0.0009)	0.2491 (0.0008)	0.0191 (0.0011)
0.1	0.0257 (0.0018)	0.0191 (0.0011)	0.3498 (0.0004)	0.3256 (0.0004)	0.0191 (0.0011)

IV Experiment 4: root estimation. The root experiment uses a scalar IV design. We draw $X \sim \mathcal{N}(0, I_p)$ with $p = 3$, set $Z = 2\sum_{j=1}^{p} X_j + \eta$ with $\eta \sim \mathcal{N}(0, 0.1)$, and generate

$$Y = \omega^\star Z + 0.5\eta + \varepsilon_Y, \qquad \varepsilon_Y \sim \mathcal{N}(0, 0.1^2), \qquad \omega^\star = 2.$$

Here $j^\star(X) = \mathbb{E}_P[Z \mid X] = 2\sum_j X_j$, $m_P^\star(X) = \mathbb{E}_P[Y \mid X] = \omega^\star j^\star(X)$, and $\nabla\mathcal{F}(\omega) = 4p(\omega - \omega^\star)$, so the population stationary point is exactly $\omega^\star = 2$. We compare PI, OBiGrad, the oracle DR benchmark, and matched KBO/KBO+OBiGrad with either $\lambda = 10^{-2}$ or $\lambda_N = 0.05N^{-0.6}$. Every method now uses the same total N observations and two-fold evaluation, with 300 replications per N; the KBO bandwidth is 2. Population roots use converged quadrature. At $\lambda = 0.01$, the KBO population root is 2.23507, while the matched corrected population root is 2 because the outcome and inner conditional means align at the true root. Tables 17–18 and Figures 13–14 show that OBiGrad tracks the oracle DR benchmark, while fixed-λ KBO remains biased toward its regularized stationary point.

Table 17: Root-estimation RMSE. Parentheses report Monte Carlo 95% error bars.

N	PI	OBiGrad	Oracle	KBO fixed	OBi fixed	KBO decay	OBi decay
100	0.0107 (0.0012)	0.0056 (0.0005)	0.0056 (0.0005)	0.3146 (0.0065)	0.0059 (0.0005)	0.1903 (0.0064)	0.0060 (0.0006)
200	0.0056 (0.0004)	0.0038 (0.0003)	0.0038 (0.0003)	0.2730 (0.0038)	0.0039 (0.0003)	0.1265 (0.0035)	0.0040 (0.0003)
400	0.0035 (0.0003)	0.0029 (0.0002)	0.0029 (0.0002)	0.2572 (0.0023)	0.0030 (0.0002)	0.0891 (0.0019)	0.0030 (0.0002)
800	0.0023 (0.0002)	0.0019 (0.0001)	0.0019 (0.0001)	0.2452 (0.0013)	0.0020 (0.0001)	0.0599 (0.0009)	0.0020 (0.0001)
1600	0.0015 (0.0001)	0.0014 (0.0001)	0.0014 (0.0001)	0.2401 (0.0008)	0.0014 (0.0001)	0.0407 (0.0006)	0.0014 (0.0001)

Table 18: KBO population roots in the scalar root experiment.

λ	KBO root	KBO bias	OBi root	OBi bias
0.000598	2.03267	0.03267	2.00000	0.00000
0.000906	2.04458	0.04458	2.00000	0.00000
0.001373	2.05964	0.05964	2.00000	0.00000
0.002081	2.07836	0.07836	2.00000	0.00000
0.003155	2.10220	0.10220	2.00000	0.00000
0.010000	2.23507	0.23507	2.00000	0.00000

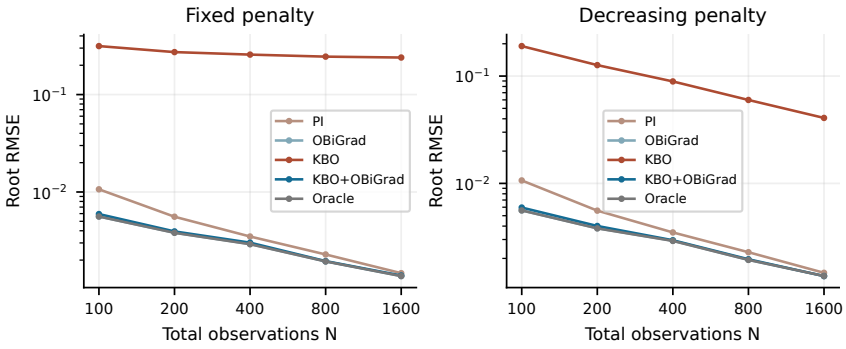


Figure 13: IV root estimation RMSE.

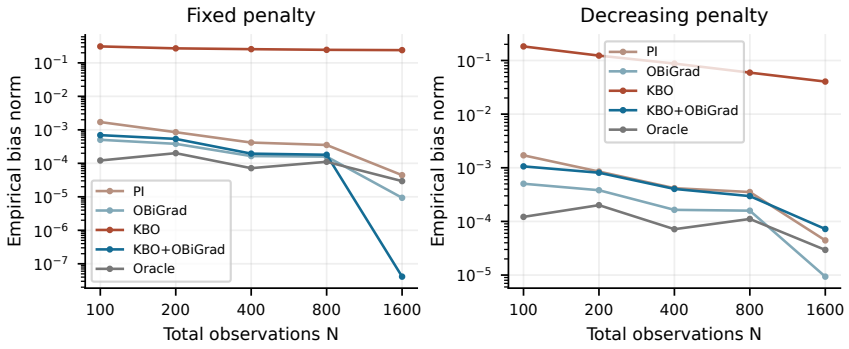


Figure 14: IV empirical root-bias norms with matched total sample sizes, under fixed and decreasing penalties.

F.3 Fitted Q-regression experiments

These supplementary Gaussian designs (including Gaussian noise and, in some fitted-Q cases, unbounded features) fall outside the bounded-data conditions of Assumption A. They are robustness illustrations rather than literal instantiations of the finite-sample theory.

Fitted Q-regression DGP. The fitted Q-regression experiments use a projected Bellman-backup design. Let $X = (S, A)$ and $Z = (S, A, R, S')$. The state is $S \sim \mathcal{N}(0, 1)$, the action is binary $A \sim \text{Bernoulli}(0.5)$, and the next state follows

$$S' = \rho S + \tau A + \xi, \qquad \xi \sim \mathcal{N}(0, \sigma_S^2),$$

with $\rho = 0.7$, $\tau = 0.5$, and $\sigma_S = 0.2$. The reward is

$$R = \sin(S) + 0.5A + 0.25SA + \varepsilon_R, \qquad \varepsilon_R \sim \mathcal{N}(0, 0.1^2).$$

For a continuation-value model $V_\omega(s') = \omega^\top \phi(s')$, the Bellman target is $g_\omega(Z) = R + \gamma V_\omega(S')$, and the lower-level loss is the quadratic Bellman-regression loss

$$\frac{1}{2}\mathbb{E}_P\big[(h(S, A) - R - \gamma V_\omega(S'))^2\big],$$

whose population solution is the projected Bellman backup

$$h^\star_{\omega,P}(S, A) = \mathbb{E}_P[R + \gamma V_\omega(S') \mid S, A],$$

as in the FQE example of Section 2.

For the gradient-estimation and KBO experiments, we use

$$\phi(s') = (\sin s', \cos s', s', (s')^2)^\top, \qquad \omega^\star = (0.55, -0.35, 0.25, 0.15)^\top,$$

$\gamma = 0.8$, and

$$Y = R + \gamma V_{\omega^\star}(S') + \varepsilon_Y, \qquad \varepsilon_Y \sim \mathcal{N}(0, 0.25^2).$$

The evaluation point is

$$\omega_0 = \omega^\star + 0.35\frac{(1, -0.5, 0.35, -0.25)^\top}{\|(1, -0.5, 0.35, -0.25)\|}.$$

For the Wald experiment, we use a state-dependent behavior policy $P(A = 1 \mid S) = \sigma(0.5S)$ and $\gamma = 0.9$, set $\omega^\star = (1, -0.6, 0.35, 0.15)^\top / \|(1, -0.6, 0.35, 0.15)\|$, and evaluate at $\omega_0 = \omega^\star + 0.35(0.6, 0.2, -0.5, 0.3)^\top / \|(0.6, 0.2, -0.5, 0.3)\|$. For the root experiment, we use the two-dimensional feature map $\phi(s') = (\sin s', \cos s')^\top$ with $\omega^\star = (0.65, -0.45)^\top$, which gives a well-conditioned population stationary point; the outcome-noise standard deviation is 0.1 in this root experiment.

The analytic nuisances are available because $S' \mid S, A$ is Gaussian. Writing $\mu(S, A) = \rho S + \tau A$,

$$\mathbb{E}_P[\sin(S') \mid S, A] = e^{-\sigma_S^2/2}\sin\{\mu(S, A)\}, \qquad \mathbb{E}_P[\cos(S') \mid S, A] = e^{-\sigma_S^2/2}\cos\{\mu(S, A)\},$$

and $\mathbb{E}_P[S' \mid S, A] = \mu(S, A)$, $\mathbb{E}_P[(S')^2 \mid S, A] = \mu(S, A)^2 + \sigma_S^2$. Hence

$$j^\star_{\omega,P}(S, A) = \mathbb{E}_P[\gamma\phi(S') \mid S, A],$$

$$h^\star_{\omega,P}(S, A) = \mathbb{E}_P[R \mid S, A] + j^\star_{\omega,P}(S, A)^\top\omega, \qquad m^\star_P(S, A) = \mathbb{E}_P[R \mid S, A] + j^\star_{\omega,P}(S, A)^\top\omega^\star.$$

The default feasible nuisance learner uses a 38-column basis of the raw observed covariates (S, A), without centering or standardization:

$$\begin{aligned} b(S, A) = \big(&1, S, A, SA, S^2, AS^2, S^3, AS^3, \\ &\{\sin(fS), \cos(fS), A\sin(fS), A\cos(fS)\}_{f\in\{0.5,0.75,1,1.5,2\}}, \\ &\{e^{-(S-c)^2/(2\cdot 0.8^2)}, Ae^{-(S-c)^2/(2\cdot 0.8^2)}\}_{c\in\{-2,-1,0,1,2\}}\big)^\top. \end{aligned}$$

On each training fold, with design matrix B and matrix T of observed regression targets, the coefficients solve $(B^\top B + \alpha\,\mathrm{diag}(0, 1, \ldots, 1))\widehat{\beta} = B^\top T$. The intercept is unpenalized; the loss uses a sum of squares, so α is not multiplied by the fold size. The fixed penalties are $\alpha = 0.5$ for fixed-gradient estimation and the basis learner in the regularization comparison, $\alpha = 0.05$ for Wald calibration, and $\alpha = 1$ for root estimation, with no tuning of these basis-ridge penalties. Oracle conditional expectations and transition parameters are used only for evaluation and for the oracle DR benchmark.

Fitted Q-regression Experiment 1: fixed-gradient estimation. This experiment estimates $\nabla\mathcal{F}(\omega_0)$ for $N \in \{200, 400, 800, 1600, 3200\}$ over 400 replications, comparing the plug-in hypergradient, OBiGrad, and the oracle DR benchmark. The nuisance product proxy $\|\hat{j}_\omega - j^\star_{\omega,P}\|_{L^2(P_X)^d}(\|\hat{h}_\omega - h^\star_{\omega,P}\|_{L^2(P_X)} + \|\hat{m} - m^\star_P\|_{L^2(P_X)})$ is the product-rate diagnostic corresponding to the exact mixed-bias bound in (10). Tables 19–20 and Figure 15 show that OBiGrad substantially improves over the plug-in hypergradient at small N and approaches the oracle DR benchmark as N grows.

Table 19: Fitted Q-regression gradient estimation. Parentheses report Monte Carlo 95% error bars.

N	PI	OBiGrad	Oracle	OBi coverage	Product
200	0.0812 (0.0112)	0.0458 (0.0045)	0.0321 (0.0015)	0.941 (0.013)	0.0832
400	0.0477 (0.0111)	0.0258 (0.0020)	0.0218 (0.0010)	0.951 (0.013)	0.0363
800	0.0215 (0.0022)	0.0167 (0.0007)	0.0160 (0.0006)	0.948 (0.013)	0.0146
1600	0.0126 (0.0006)	0.0115 (0.0005)	0.0114 (0.0005)	0.953 (0.012)	0.0069
3200	0.0081 (0.0003)	0.0079 (0.0003)	0.0078 (0.0003)	0.954 (0.012)	0.0035

Table 20: Nuisance-learning diagnostics for fitted Q-regression gradient estimation.

N	$\|\hat{h} - h^\star\|$	$\|\hat{j} - j^\star\|$	$\|\hat{m} - m^\star\|$	Product
200	0.1265	0.2519	0.1707	0.0832
400	0.0746	0.1668	0.1120	0.0363
800	0.0461	0.1160	0.0715	0.0146
1600	0.0302	0.0833	0.0494	0.0069
3200	0.0212	0.0620	0.0339	0.0035

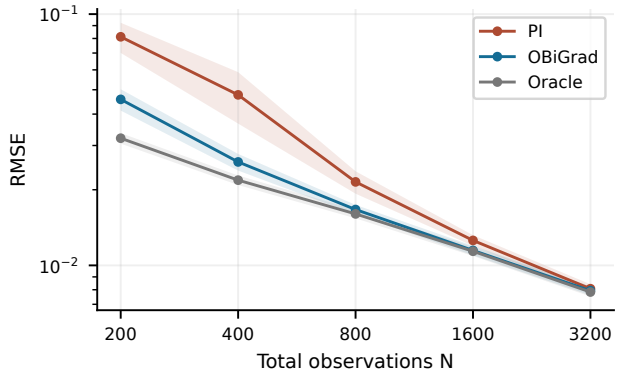


Figure 15: Fitted Q-regression fixed-gradient estimation. OBiGrad improves over PI at small sample sizes and approaches the oracle DR benchmark as N grows.

Fitted Q-regression Experiment 2: Wald calibration. This experiment forms coordinate-wise 95% Wald confidence intervals (84) using the empirical covariance estimator. Results are averaged over 400 replications. Tables 21–22 and Figures 16–17 report coverage, interval length, RMSE, studentized quantiles, and QQ diagnostics. OBiGrad intervals are shorter than the PI diagnostic intervals, with broadly near-nominal coverage but visible finite-sample departures. At $N = 3200$, pooled coverage is 0.9338 with Monte Carlo half-width 0.0154, versus 0.9806 for PI. The oracle coverage is 0.9381, so the departures cannot all be assigned to nuisance learning. The QQ plot describes the displayed coordinate; it does not validate inference after adaptive optimization.

Table 21: Fitted Q-regression Wald calibration.

N	Coverage	Length	Coord. RMSE	Product	Oracle cov.
200	0.943 (0.013)	0.1171 (0.0129)	0.0584 (0.0124)	0.2105	0.945 (0.013)
400	0.946 (0.013)	0.0653 (0.0069)	0.0330 (0.0123)	0.0885	0.950 (0.013)
800	0.952 (0.012)	0.0380 (0.0011)	0.0113 (0.0009)	0.0360	0.947 (0.013)
1600	0.943 (0.013)	0.0255 (0.0004)	0.0076 (0.0004)	0.0155	0.948 (0.012)
3200	0.934 (0.015)	0.0175 (0.0001)	0.0049 (0.0002)	0.0068	0.938 (0.014)

Table 22: Studentized-error diagnostics for fitted Q-regression Wald intervals.

N	Mean	SD	2.5%	Median	97.5%	PI coverage
200	-0.005	1.117	-2.079	0.034	1.928	0.961 (0.012)
400	-0.031	1.077	-2.022	-0.032	1.944	0.972 (0.010)
800	0.011	1.016	-1.977	0.055	1.926	0.979 (0.008)
1600	0.018	1.037	-1.962	0.017	2.058	0.968 (0.010)
3200	-0.078	1.040	-2.174	-0.082	1.961	0.981 (0.007)

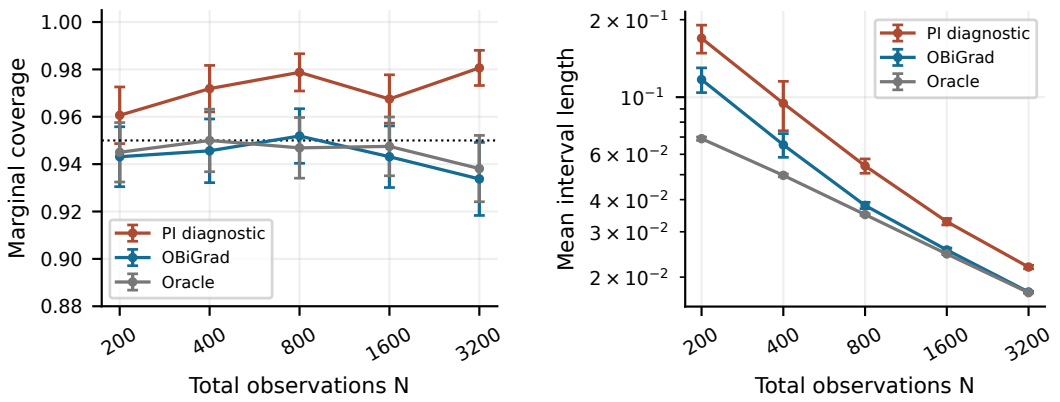


Figure 16: Fitted Q-regression Wald diagnostics. Left: coordinate-wise coverage of nominal 95% intervals. Right: average interval length.

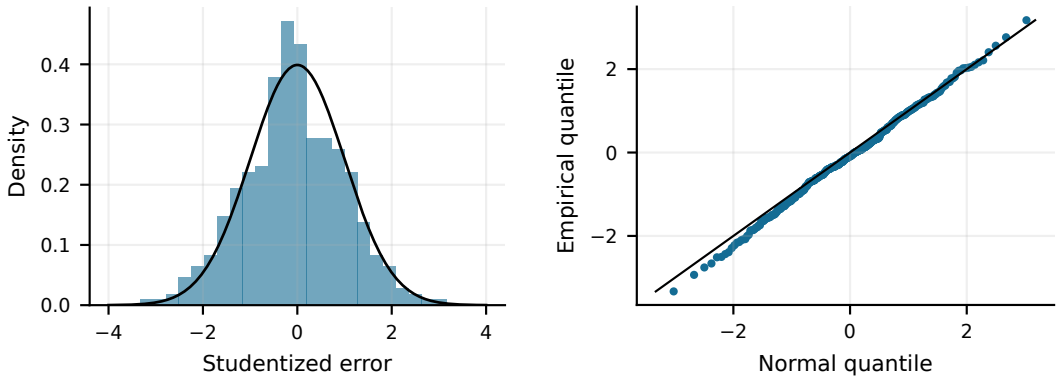


Figure 17: Fitted Q-regression studentized OBiGrad errors for coordinate 0 at $N = 3200$.

Matched kernel Wald extension. We add KBO and KBO+OBiGrad on exactly the same datasets, folds, evaluation parameter, and population target as the preceding Wald experiment. The two seed suites give 400 replications per N. The kernel specification is the Bellman KBO map with 256 RFFs and kernel $\exp(-0.35\|x - x'\|^2)$; each seed suite retains its own fixed map. We use fixed $\lambda = 0.01$ and decreasing $\lambda_N = 0.05N^{-0.6}$, without selecting penalties using coverage or population truth. Each coordinate interval is the score mean plus or minus 1.96 times its empirical score standard deviation divided by $\sqrt{N}$. KBO uses its plug-in scores and KBO+OBiGrad its corrected scores. The KBO intervals are diagnostic: this score covariance alone is not a general variance formula accounting for learned plug-in nuisances. Orthogonal intervals also require sufficiently accurate nuisance fits for the asymptotic result to apply; a fixed penalty can leave bias.

Table 23 and Figure 18 show improved coverage after correction, especially with decreasing regularization. At $N = 3200$, decreasing-penalty coverage is 0.8613 for KBO and 0.9319 for KBO+OBiGrad. With fixed $\lambda = 0.01$, coverage is 0.0025 and 0.8413, respectively. The residual fixed-penalty undercoverage is consistent with a remaining product bias; correction does not eliminate that bias entirely.

Table 23: Fitted Q-regression kernel Wald extension: marginal coverage and mean interval length. Parentheses give 95% Monte Carlo half-widths across complete datasets. "OBi" denotes KBO+OBiGrad.

N	Penalty	KBO cov.	OBi cov.	KBO length	OBi length
200	Fixed	0.293 (0.015)	0.907 (0.016)	0.0792 (0.0007)	0.0601 (0.0007)
400	Fixed	0.195 (0.011)	0.913 (0.016)	0.0558 (0.0003)	0.0431 (0.0003)
800	Fixed	0.128 (0.012)	0.908 (0.015)	0.0389 (0.0002)	0.0306 (0.0002)
1600	Fixed	0.037 (0.009)	0.879 (0.017)	0.0276 (0.0001)	0.0216 (0.0001)
3200	Fixed	0.003 (0.002)	0.841 (0.018)	0.0194 (4.7e-05)	0.0153 (4.3e-05)
200	Decreasing	0.851 (0.020)	0.924 (0.014)	0.0788 (0.0008)	0.0645 (0.0008)
400	Decreasing	0.863 (0.019)	0.937 (0.015)	0.0556 (0.0004)	0.0469 (0.0004)
800	Decreasing	0.873 (0.018)	0.941 (0.013)	0.0389 (0.0002)	0.0337 (0.0002)
1600	Decreasing	0.873 (0.017)	0.941 (0.013)	0.0277 (0.0001)	0.0240 (0.0001)
3200	Decreasing	0.861 (0.018)	0.932 (0.015)	0.0196 (0.0001)	0.0170 (0.0001)

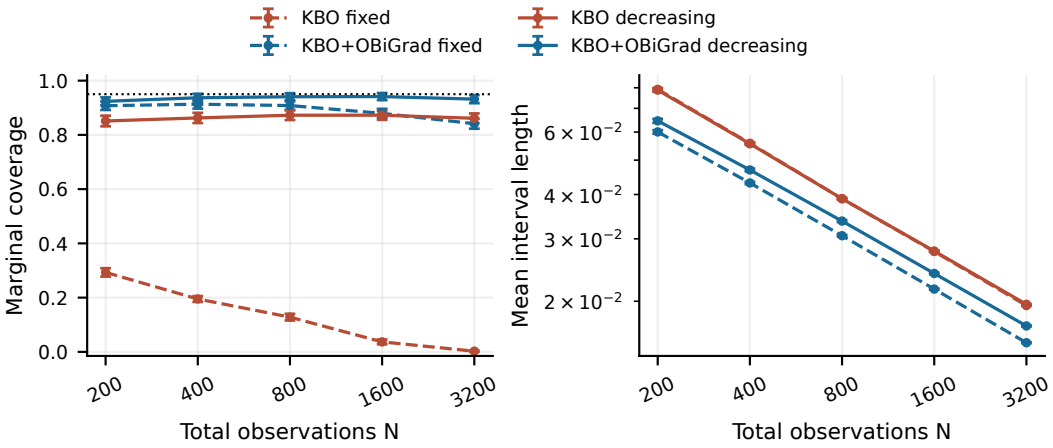


Figure 18: Fitted Q-regression kernel Wald diagnostics: coverage (left) and interval length (right), for fixed and decreasing regularization. Both methods use identical kernel nuisance fits.

Fitted Q-regression Experiment 3: root estimation. The root experiment uses the two-feature continuation model $V_\omega(S') = \omega^\top(\sin S', \cos S')^\top$ with $\omega^\star = (0.65, -0.45)^\top$. The outcome is $Y = R + \gamma V_{\omega^\star}(S') + \varepsilon_Y$, so the unregularized population stationary point is exactly $\omega^\star$. We compare the plug-in hypergradient, OBiGrad, the oracle DR benchmark, and the oracle plug-in root. The root equations are affine in ω, so they are solved directly rather than by gradient descent. Tables 24–25 and Figures 19–20 show that OBiGrad tracks the oracle DR root and improves over the plug-in hypergradient for moderate and large sample sizes.

Table 24: Fitted Q-regression root-estimation RMSE. Parentheses report Monte Carlo 95% error bars.

N	PI	OBiGrad	Oracle DR	Oracle PI
200	0.0707 (0.0090)	0.0248 (0.0028)	0.0188 (0.0010)	0.0346 (0.0018)
400	0.0316 (0.0108)	0.0172 (0.0040)	0.0133 (0.0007)	0.0216 (0.0011)
800	0.0127 (0.0011)	0.0100 (0.0005)	0.0097 (0.0005)	0.0165 (0.0008)
1600	0.0072 (0.0004)	0.0066 (0.0003)	0.0066 (0.0003)	0.0110 (0.0006)
3200	0.0049 (0.0002)	0.0048 (0.0002)	0.0048 (0.0002)	0.0082 (0.0004)

Table 25: Additional fitted Q-regression root-estimation diagnostics.

N	Method	Bias norm	Mean norm	Median norm	90% norm
200	PI	0.0379	0.0500	0.0318	0.1112
200	OBiGrad	0.0014	0.0203	0.0175	0.0351
200	Oracle DR	0.0007	0.0164	0.0147	0.0292
200	Oracle PI	0.0009	0.0305	0.0279	0.0534
400	PI	0.0096	0.0204	0.0151	0.0375
400	OBiGrad	0.0001	0.0131	0.0110	0.0231
400	Oracle DR	0.0006	0.0117	0.0107	0.0201
400	Oracle PI	0.0009	0.0190	0.0172	0.0349
800	PI	0.0024	0.0106	0.0095	0.0180
800	OBiGrad	0.0007	0.0088	0.0084	0.0150
800	Oracle DR	0.0006	0.0086	0.0081	0.0146
800	Oracle PI	0.0003	0.0144	0.0134	0.0249
1600	PI	0.0003	0.0063	0.0059	0.0111
1600	OBiGrad	0.0003	0.0059	0.0056	0.0101
1600	Oracle DR	0.0003	0.0058	0.0055	0.0100
1600	Oracle PI	0.0004	0.0096	0.0087	0.0171
3200	PI	0.0004	0.0043	0.0041	0.0075
3200	OBiGrad	0.0003	0.0042	0.0039	0.0074
3200	Oracle DR	0.0003	0.0042	0.0040	0.0073
3200	Oracle PI	0.0005	0.0073	0.0070	0.0125

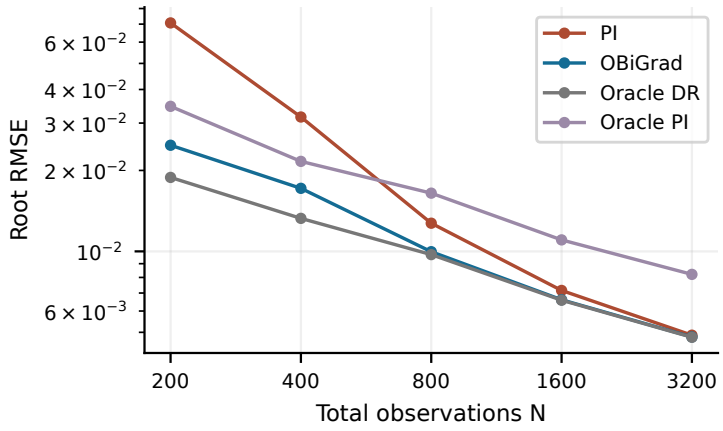


Figure 19: Fitted Q-regression root-estimation RMSE. OBiGrad tracks the oracle DR benchmark and improves over PI.

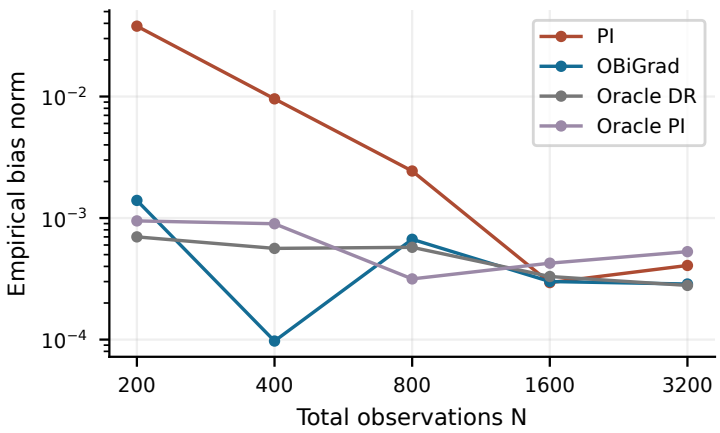


Figure 20: Fitted Q-regression root-estimation bias.

Fitted Q-regression Experiment 4: KBO regularization bias. This experiment uses $N = 1200$ total observations and 200 replications. KBO uses a fixed 256-feature RFF approximation to $k(x, x') = \exp(-0.35\|x - x'\|^2)$, with the same nine penalties as IV Experiment 3. The number 0.35 is the kernel exponent coefficient, not the Gaussian bandwidth. Tables 26–27 use matched two-fold evaluation for all methods and population targets obtained by quadrature for the saved feature map. The one-direction protocol is shown separately in Figure 22 and Table 28. Orthogonalization reduces population bias and improves total error at many penalties, but comparisons at the smallest penalties are close. At $\lambda = 10^{-4}$, matched KBO RMSE is 0.01353 and existing basis OBiGrad RMSE is 0.01317; in the independent seed/RFF suite they are 0.01359 and 0.01364. Thus a claim that the basis estimator dominates KBO at every penalty is not supported. The primary ablation compares each learner with its own orthogonal counterpart.

Table 26: Matched two-fold fitted Q-regression gradient RMSE at $N = 1200$. Parentheses report Monte Carlo 95% half-widths.

λ	PI	OBiGrad	KBO	KBO+OBiGrad	Oracle
1e-05	0.0148 (0.0011)	0.0132 (0.0008)	0.0135 (0.0008)	0.0133 (0.0009)	0.0131 (0.0009)
3e-05	0.0148 (0.0011)	0.0132 (0.0008)	0.0134 (0.0008)	0.0132 (0.0009)	0.0131 (0.0009)
0.0001	0.0148 (0.0011)	0.0132 (0.0008)	0.0135 (0.0008)	0.0131 (0.0008)	0.0131 (0.0009)
0.0003	0.0148 (0.0011)	0.0132 (0.0008)	0.0142 (0.0009)	0.0130 (0.0008)	0.0131 (0.0009)
0.001	0.0148 (0.0011)	0.0132 (0.0008)	0.0187 (0.0010)	0.0130 (0.0008)	0.0131 (0.0009)
0.003	0.0148 (0.0011)	0.0132 (0.0008)	0.0345 (0.0011)	0.0130 (0.0008)	0.0131 (0.0009)
0.01	0.0148 (0.0011)	0.0132 (0.0008)	0.0811 (0.0011)	0.0133 (0.0009)	0.0131 (0.0009)
0.03	0.0148 (0.0011)	0.0132 (0.0008)	0.1681 (0.0011)	0.0149 (0.0009)	0.0131 (0.0009)
0.1	0.0148 (0.0011)	0.0132 (0.0008)	0.3017 (0.0013)	0.0246 (0.0009)	0.0131 (0.0009)

Table 27: Fitted Q-regression kernel population-bias norms and estimation RMSE around each method's own population target.

λ	KBO bias	KBO estimation	OBi bias	OBi estimation
1e-05	0.0002	0.0135 (0.0008)	2e-05	0.0133 (0.0009)
3e-05	0.0006	0.0134 (0.0008)	0.0001	0.0132 (0.0009)
0.0001	0.0016	0.0134 (0.0008)	0.0001	0.0131 (0.0008)
0.0003	0.0043	0.0133 (0.0008)	0.0003	0.0130 (0.0008)
0.001	0.0126	0.0130 (0.0008)	0.0007	0.0130 (0.0008)
0.003	0.0314	0.0125 (0.0008)	0.0016	0.0129 (0.0008)
0.01	0.0795	0.0116 (0.0007)	0.0041	0.0127 (0.0008)
0.03	0.1673	0.0111 (0.0007)	0.0087	0.0121 (0.0008)
0.1	0.3015	0.0132 (0.0011)	0.0219	0.0108 (0.0007)

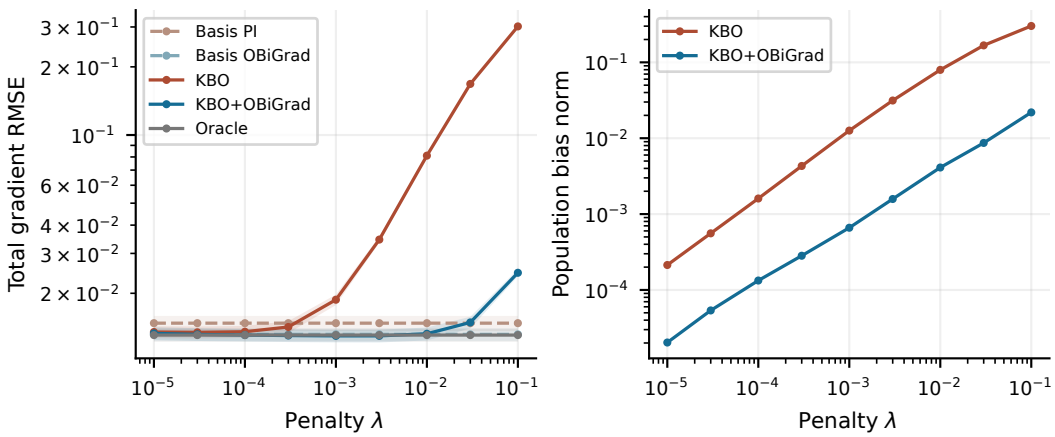


Figure 21: Fitted Q-regression kernel diagnostics: matched total gradient RMSE (left) and population-bias norms (right).

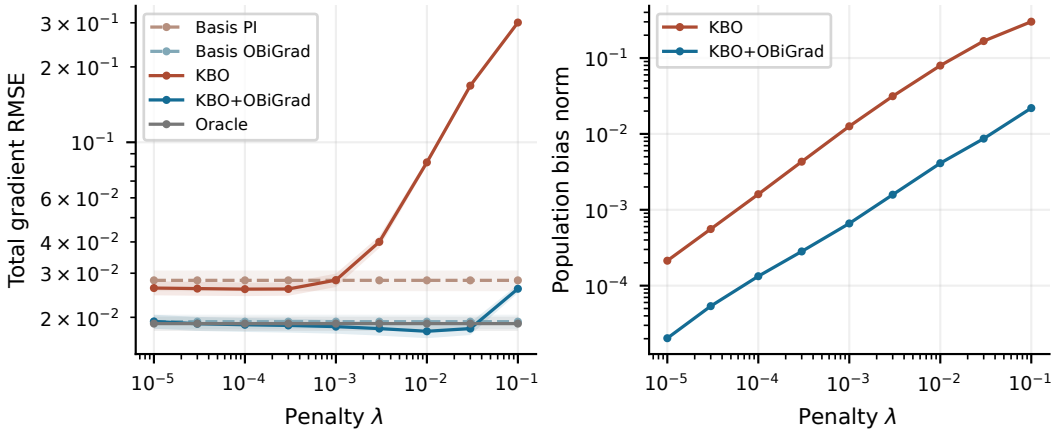


Figure 22: Fitted Q-regression under the original one-direction split: 600 training and 600 evaluation observations, shared by all methods.

Table 28: Fitted Q-regression: original one-direction evaluation, with matched observations for all five estimators. Parentheses give Monte Carlo 95% half-widths.

λ	PI	OBiGrad	KBO	KBO+OBiGrad	Oracle
1e-05	0.0281 (0.0027)	0.0192 (0.0013)	0.0262 (0.0017)	0.0192 (0.0012)	0.0189 (0.0012)
3e-05	0.0281 (0.0027)	0.0192 (0.0013)	0.0261 (0.0017)	0.0189 (0.0012)	0.0189 (0.0012)
0.0001	0.0281 (0.0027)	0.0192 (0.0013)	0.0259 (0.0016)	0.0187 (0.0012)	0.0189 (0.0012)
0.0003	0.0281 (0.0027)	0.0192 (0.0013)	0.0260 (0.0016)	0.0186 (0.0012)	0.0189 (0.0012)
0.001	0.0281 (0.0027)	0.0192 (0.0013)	0.0282 (0.0017)	0.0184 (0.0012)	0.0189 (0.0012)
0.003	0.0281 (0.0027)	0.0192 (0.0013)	0.0400 (0.0022)	0.0180 (0.0011)	0.0189 (0.0012)
0.01	0.0281 (0.0027)	0.0192 (0.0013)	0.0831 (0.0024)	0.0176 (0.0011)	0.0189 (0.0012)
0.03	0.0281 (0.0027)	0.0192 (0.0013)	0.1683 (0.0025)	0.0180 (0.0010)	0.0189 (0.0012)
0.1	0.0281 (0.0027)	0.0192 (0.0013)	0.3007 (0.0024)	0.0260 (0.0010)	0.0189 (0.0012)