Decision-Aware Proximal Bridge Learning for Optimal Treatment Selection

Tomàs Garriga [1,2,3*†] **Alejandro Almodóvar**[4*] **Axel Brando**[2‡] **Gerard Sanz**[1]

Eduard Serrahima de Cambra[1] **Juan Parras**[4]

Abstract

Individualized treatment selection with continuous actions requires accurate causal response estimation in decision-relevant regions, rather than uniformly over the entire action space. Estimating a global causal response surface and then choosing the treatment that maximizes it can therefore be suboptimal, since standard estimation objectives allocate modeling effort according to the observed treatment distribution rather than the regions that determine the optimal decision. While decision-aware approaches have been studied in unconfounded settings, this problem remains underexplored in proximal causal inference, where proxy variables and bridge functions enable identification under suitable assumptions even in the presence of hidden confounding. Despite recent progress, proximal methods have primarily focused on treatment-effect and potential-outcome estimation rather than treatment selection and optimal decision-making. To bridge this gap, we introduce a policy-targeted weighted bridge loss that emphasizes decision-relevant treatment regions while retaining global stabilization. We prove a regret bound showing that the proposed weighted bridge loss controls treatment-selection regret through a weighted ill-posedness constant. We instantiate the framework in decision-aware variants of several proximal bridge solvers, yielding practical algorithms that alternate between weighted bridge estimation, response-surface projection, policy update, and weight refinement. Empirically, we find that decision-aware weighting reduces regret across several bridge solvers, suggesting improved treatment selection in proximal settings.

1 Introduction

Individualized treatment selection from observational data is a central goal in machine learning and causal inference. In applications such as medicine, marketing, education, and public policy, the aim is not only to estimate treatment effects, but to select an intervention with high utility for each individual. This distinction is especially important for continuous treatments, where the learner must recover enough of the response surface to recommend an optimal dose. However, most counterfactual prediction and causal-effect estimation methods optimize global accuracy over the treatment space. As emphasized by Zou et al. [47], such objectives can be poorly aligned with decision quality: a model may fit the response surface well on average while being inaccurate near the treatment optimum.

[1]Novartis
[2]Barcelona Supercomputing Center
[3]Universitat Politècnica de Catalunya
[4]Information Processing and Telecommunications Center, ETSI de Telecomunicación, Universidad Politécnica de Madrid
[‡]Present affiliation: Trustworthy and Reflective Artificial Intelligence Lab (TRAILAB).
[*]Co-first authors.
[†]Corresponding author: tomas.garriga_dicuzzo@novartis.com.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

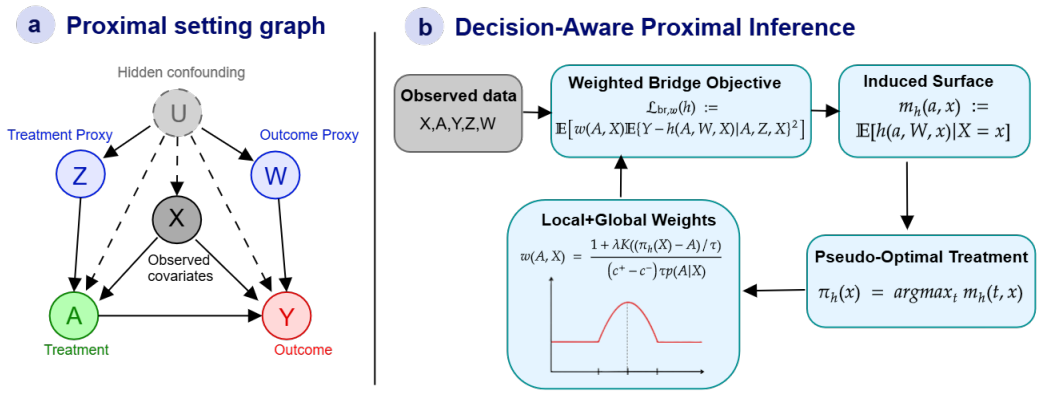


Figure 1: (a) Proximal causal graph with proxy variables. (b) Decision-aware bridge-learning pipeline with local–global weighting around the pseudo-optimal treatment.

The challenge is amplified in observational studies with unmeasured confounding. Standard individualized-treatment methods typically assume exchangeability after adjusting for observed covariates, but treatment assignment may depend on latent severity, physician judgment, patient frailty, unrecorded preferences, or institutional constraints. When such latent factors affect both treatment and outcome, ignoring them can bias both response-surface estimates and the resulting decisions.

Proximal causal inference provides a principled approach to this setting by using proxy variables for hidden confounders and identifying causal effects through bridge functions satisfying conditional moment restrictions [29, 39]. Recent work has developed kernel, moment-restriction, neural, doubly robust, and treatment-rule methods for proximal learning [27, 24, 42, 34, 37]. Yet most of these methods remain estimation-centric: they aim to recover bridge functions or dose-response surfaces accurately in a global sense.

For treatment selection, global accuracy is not the right target. Errors in clearly suboptimal treatment regions may have little effect on policy value, whereas small errors near the optimum can change the selected dose. We therefore study *policy-targeted proximal bridge learning*. Our key idea is that proximal bridge estimation is an inverse problem, and the norm used to solve it should reflect the downstream decision. We construct a weighted bridge objective that emphasizes treatment regions near a pseudo-optimal dose estimated from a pilot bridge while retaining global stabilization.

Our contributions are:[1] **(1)** we formulate individualized continuous-treatment selection under hidden confounding as a regret-targeted proximal bridge-learning problem, extending Outcome Oriented Counterfactual Prediction [47] to proximal inference through a policy-targeted weighted bridge objective; **(2)** we show that, under a weighted ill-posedness condition, a policy-targeted weighted bridge loss controls a weighted causal-surface loss and hence a treatment-selection regret surrogate; **(3)** we develop an iterative algorithm alternating between bridge estimation, response-surface projection, policy update, and weight refinement, instantiate it for several proximal bridge solvers, and demonstrate its effectiveness on synthetic and semi-synthetic benchmarks with hidden confounding and proxy variables.

2 Related Work

Decision-oriented learning and continuous treatments. A large literature studies treatment-effect estimation and policy learning from observational data, including generalized propensity-score methods for continuous treatments [19] and modern approaches to dose-response estimation and continuous-action policy optimization [22, 38]. A complementary line of work emphasizes that

[1] Code is available at https://github.com/aalmodovares/DecisionAwareProximalLearning.

accurate causal estimation need not imply accurate treatment selection. In particular, Frauen et al. [13] propose a decision-aware learning objective for conditional average treatment effect that prioritizes errors near the decision boundary in binary treatment settings. Individualized treatment-rule methods such as outcome-weighted learning directly target decision quality [46], while outcome-oriented counterfactual prediction shows that errors near decision-relevant treatments can matter more than global surface error [47]. These ideas are especially important for continuous treatments, where only a small region of the action space may determine the recommended dose. Most such methods, however, assume ignorability.

Proximal causal inference and bridge learning. Proximal causal inference uses negative-control or proxy variables to identify causal quantities under unmeasured confounding through outcome and treatment bridge functions satisfying conditional moment restrictions [29, 39, 9]. Recent machine-learning methods estimate these bridges using kernel two-stage regression and maximum moment restrictions, including kernel proxy variables (KPV) and proxy maximum moment restriction (PMMR) [27]; neural or adversarial conditional-moment objectives, including neural maximum moment restriction (NMMR) [24, 23, 16]; deep proxy representations, including deep feature proxy variables (DFPV) [43]; and kernel methods for causal response functions [38]. Related work develops continuous-treatment, doubly robust, and density-ratio-free proximal estimators [42, 6, 7], as well as implicit bridge learning through deconfounding generative models [2] and proximal treatment-rule learning under hidden confounding [34, 37]. These methods advance estimation and policy learning in proximal models, but their bridge-learning objectives remain largely estimation-centric rather than decision oriented. See Appendix A for an extended discussion.

Research gap. Existing outcome-oriented methods target decision-relevant counterfactual prediction but usually assume observed-confounder identification. Existing proximal methods handle latent confounding but usually optimize global bridge or effect estimation. Our work targets the intersection: decision-aware bridge learning for continuous treatment policies under unmeasured confounding.

3 Background and Problem Setting

3.1 Proximal Inference

We observe i.i.d. samples $O_i = (Y_i, A_i, Z_i, W_i, X_i)$, for $i = 1, \dots, n$, where $A \in \mathcal{A} = [c_-, c_+]$ is a continuous treatment, Y is the outcome, and X denotes observed pre-treatment covariates. The variable U denotes latent confounders that may affect both treatment assignment and the outcome. Proximal inference uses two additional observed variables as proxies for U: a *treatment-inducing proxy* Z and an *outcome-inducing proxy* W. Informally, Z carries information about latent factors that affect treatment assignment, while W carries information about latent factors that affect the outcome. The left-hand side of Figure 1 shows the causal graph of the proximal setting.

The proximal strategy is to replace adjustment for the unobserved U by an observable bridge equation. The bridge is not itself the causal response surface. Rather, it is a function of the observed treatment, outcome proxy, and covariates whose conditional expectation recovers the causal response after averaging over the distribution of W. This distinction is central for our objective: policy decisions will be made through the causal response surface induced by a learned bridge.

We write $Y(a)$ for the potential outcome under treatment level a. All expectations are with respect to the observed-data distribution unless potential outcomes are explicitly shown. We impose the following standard proximal inference assumptions.

Assumption 1 (Consistency and latent exchangeability)**.** *For every $a \in \mathcal{A}$, $Y(a)$ is well defined, $Y = Y(A)$ almost surely, and $Y(a) \perp\!\!\!\perp A \mid U, X$.*

Assumption 2 (Negative-control proxies)**.** *For every $a \in \mathcal{A}$, $Y(a) \perp\!\!\!\perp Z \mid A, U, X$ and $W \perp\!\!\!\perp (A, Z) \mid U, X$.*

Assumption 3 (Completeness)**.** *For any square-integrable ℓ, if $\mathbb{E}\{\ell(U) \mid A = a, Z = z, X = x\} = 0$ for all (a, z, x) in the support of (A, Z, X), then $\ell(U) = 0$ almost surely. For any square-integrable g, if $\mathbb{E}\{g(Z) \mid A = a, W = w, X = x\} = 0$ for all (a, w, x) in the support of (A, W, X), then $g(Z) = 0$ almost surely.*

Assumption 4 (Positivity)**.** *For almost every (u, x), the conditional law of A given $(U, X) = (u, x)$ has support containing $\mathcal{A}$ and admits a density $p(a \mid u, x) > 0$ on $\mathcal{A}$. Moreover, the observed density $p(a \mid x)$ is bounded away from zero on $\mathcal{A}$.*

We further assume the existence of a square-integrable outcome bridge $h_0 : \mathcal{A} \times \mathcal{W} \times \mathcal{X} \to \mathbb{R}$ satisfying the outcome bridge restriction $\mathbb{E}\{Y - h_0(A, W, X) \mid A, Z, X\} = 0$. Sufficient regularity and solvability conditions are given in Appendix C. Appendix B further discusses the assumptions.

Proposition 1 (Proximal identification)**.** *Under the bridge-existence conditions in Appendix C, for every $a \in \mathcal{A}$ and almost every x, $m_0(a, x) := \mathbb{E}\{Y(a) \mid X = x\} = \mathbb{E}\{h_0(a, W, x) \mid X = x\}$.*

3.2 Causal response surfaces, policy value and regret

Proposition 1 identifies the causal response surface $m_0(a, x)$ for the true bridge h_0. For any candidate bridge $h \in \mathcal{H}$, define its induced response surface $m_h(a, x) := \mathbb{E}\{h(a, W, x) \mid X = x\}$. A treatment policy is a measurable map $\pi : \mathcal{X} \to \mathcal{A}$. Its value is $V(\pi) := \mathbb{E}\{Y(\pi(X))\} = \mathbb{E}\{m_0(\pi(X), X)\}$. For a candidate bridge h, define the induced policy and oracle policy by

$$\pi_h(x) \in \operatorname*{argmax}_{t \in \mathcal{A}} m_h(t, x), \qquad a^\star(x) \in \operatorname*{argmax}_{t \in \mathcal{A}} m_0(t, x). \tag{1}$$

We assume throughout that larger outcomes are preferred and, for notational simplicity, that maximizers are unique; fixed tie-breaking would give the same analysis. The regret of the bridge-induced policy is

$$\mathrm{Reg}(h) := \mathbb{E}[m_0(a^\star(X), X) - m_0(\pi_h(X), X)] = V(a^\star) - V(\pi_h). \tag{2}$$

This regret is decision-specific. It does not require the entire surface m_h to be uniformly accurate; it requires sufficient accuracy at treatment levels that can change the optimizer.

3.3 Why is ordinary proximal bridge learning not decision-aligned?

Standard bridge estimators minimize a global violation of the conditional moment restriction or a kernelized maximum-moment analogue. This is natural for estimating a bridge or a full dose-response curve. It is not necessarily aligned with treatment selection. For example, a bridge may fit the moment equation well in treatment regions that are frequently observed but clearly suboptimal, while remaining inaccurate near the maximizer of $m_0(\cdot, x)$. Such localized errors can change $\pi_h(x)$ and produce large regret even when the global moment loss is small.

Our goal is therefore not to replace proximal identification. We keep the same bridge equation and the same proxy assumptions. The difference is the norm in which the bridge equation is solved. Section 4 constructs a localized-plus-global proximal risk that emphasizes treatment regions near the policy induced by a pilot bridge, while retaining a global component to stabilize the inverse problem and protect against missing the oracle region.

Our motivation is related to outcome-oriented counterfactual prediction, which shows that global counterfactual prediction error can be poorly aligned with treatment selection [47]. The challenge here is different: under unmeasured confounding, the causal surface is available only through a proximal bridge equation. Decision-targeting must therefore be imposed at the level of an inverse conditional-moment problem rather than at the level of a supervised counterfactual prediction loss.

4 Policy-Targeted Proximal Risk

This section develops the main population object of the paper: a weighted proximal bridge risk whose geometry is aligned with optimal treatment selection and that controls a regret surrogate. All the proofs are in Appendix D.

4.1 A Surface Surrogate on Regret

Define the squared surface error $G_h(t, X) := \big(m_0(t, X) - m_h(t, X)\big)^2$.

Lemma 1. *The regret satisfies* $\mathrm{Reg}(h) \le \sqrt{\mathbb{E}[G_h(\pi_h(X), X)]} + \sqrt{\mathbb{E}[G_h(a^\star(X), X)]}$.

Lemma 1 makes the target explicit. To control regret, it is sufficient to control the surface error near the learned dose $\pi_h(X)$ and the oracle dose $a^\star(X)$. The first term can be targeted using the current or pilot policy. The second term is not directly observable and motivates retaining a global component.

Assume overlap, so that $p(t \mid x) > 0$ for all $(x, t) \in \mathcal{X} \times \mathcal{A}$. Let K be a nonnegative symmetric kernel supported on $[-1, 1]$ with $\int K(u)\,du = 1$, $\int uK(u)\,du = 0$, and finite second moment $\mu_2(K) = \int u^2 K(u)\,du < \infty$. Define $K_\tau(u) = \tau^{-1}K(u/\tau)$.

The localized surface loss around the pseudo-optimal treatment is

$$A_\tau(h) := \mathbb{E}_{X,A}\left[\frac{K((\pi_h(X) - A)/\tau)}{\tau\, p(A \mid X)}\, G_h(A, X)\right]. \tag{3}$$

This approximates $G_h(\pi_h(X), X)$, the surface error at $\pi_h(X)$; the proof and a precise kernel-bias statement is given in Appendix D. The interiority condition used for this approximation avoids kernel truncation at the treatment boundaries; a support-normalized boundary extension is described in Appendix D.3.1.

Define the global surface loss

$$B(h) := \mathbb{E}_X\left[\frac{1}{c_+ - c_-}\int_{c_-}^{c_+} G_h(t, X)\,dt\right] = \mathbb{E}_{X,A}\left[\frac{G_h(A, X)}{(c_+ - c_-)p(A \mid X)}\right]. \tag{4}$$

In appendix D we show that $B(h)$ controls the oracle dose term $G_h(a^\star(X), X)$ up to a Lipschitz remainder. The bound is deliberately coarse; its role is not to sharply approximate the unknown oracle region, but to retain global stabilization.

The combined surface surrogate is $\gamma A_\tau(h) + B(h)$, where $\gamma > 0$ controls the degree of localization. Larger γ prioritizes errors near the pseudo-optimal treatment, while smaller γ emphasizes the global response surface.

Proposition 2 (Weighted surface surrogate)**.** *Let $c_\gamma := \max\{1, \gamma^{-1}\}$. Under the local kernel-approximation and global Lipschitz conditions stated in Propositions 6 and 7 (Appendix D),*

$$\mathrm{Reg}(h) \le \sqrt{2c_\gamma}\,\sqrt{\gamma A_\tau(h) + B(h)} + \sqrt{2\left(C_h\tau^2 + \frac{L}{2}(c_+ - c_-)\right)}. \tag{5}$$

Using the definitions of $A_\tau(h)$ and $B(h)$, the combined surface objective $\gamma A_\tau(h) + B(h)$ can be written as the policy-targeted surface risk:

$$\mathcal{L}^{\mathrm{surf}}_{\tau,\lambda}(h) := \mathbb{E}_{X,A}\left[\frac{1 + \lambda K((\pi_h(X) - A)/\tau)}{(c_+ - c_-)p(A \mid X)}\left(m_0(A, X) - m_h(A, X)\right)^2\right], \tag{6}$$

where $\lambda = (c_+ - c_-)\gamma/\tau$.

4.2 From weighted surface risk to weighted bridge risk

The surface risk in (6) is not directly observable because it depends on m_0. Proximal inference instead learns through the bridge moment equation. This raises two key questions: (1) can the policy-targeted weighting be justified at the bridge level, so that the weighted bridge loss controls the decision-relevant surface risk and hence regret? (2) does introducing such weights change the population bridge target?

For any strictly positive measurable weight $\omega : \mathcal{A} \times \mathcal{X} \to (0, \infty)$, define

$$\mathcal{L}_{\mathrm{surf},\omega}(h) := \mathbb{E}\left[\omega(A, X)\left(m_h(A, X) - m_0(A, X)\right)^2\right], \tag{7}$$

$$\mathcal{L}_{\mathrm{br},\omega}(h) := \mathbb{E}\left[\omega(A, X)\,\mathbb{E}\{Y - h(A, W, X) \mid A, Z, X\}^2\right]. \tag{8}$$

Let $\Delta_h(a, w, x) := h(a, w, x) - h_0(a, w, x)$ and define the outcome-bridge operator

$$(T\Delta)(a, z, x) := \mathbb{E}\{\Delta(a, W, x) \mid A = a, Z = z, X = x\}. \tag{9}$$

Proposition 3 (Weighted bridge loss controls weighted surface loss)**.** *Assume that h_0 satisfies the outcome bridge restriction. Let $\nu := P_{A,W,X}$ and $\mu := P_{W|X} \otimes P_{A|X} \otimes P_X$. Suppose there exists a finite weighted ill-posedness constant*

$$\chi_\omega := \sup_{g \in \mathcal{H}} \frac{\|\Delta_g\|_{L_2(\omega\, d\nu)}}{\|T\Delta_g\|_{L_2(\omega\, dP_{A,Z,X})}} < \infty, \tag{10}$$

and suppose the density ratio $d\mu/d\nu$ is essentially bounded by C_ρ. Then, for every $h \in \mathcal{H}$,

$$\mathcal{L}_{\mathrm{surf},\omega}(h) \le C_\rho \chi_\omega^2 \, \mathcal{L}_{\mathrm{br},\omega}(h). \tag{11}$$

Proof in Appendix D.5. The condition $\chi_\omega < \infty$ is the weighted analogue of the usual restricted ill-posedness condition for proximal inverse problems, used in works like [23, 10, 8]. It is also implied by this standard unweighted condition under bounded weights; see Appendix D.6. Appendix D.6 also provides a finite-dimensional characterization and a simple example illustrating when the weighted ill-posedness constant is finite.

The result shows that weighting the bridge loss changes the population norm in which bridge moment violations are controlled, and this norm directly controls the weighted surface target that appears in the regret surrogate.

For each $h \in \mathcal{H}$, define the self-weighted policy-targeted bridge weight

$$\omega_h(a, x) := \frac{1 + \lambda K((\pi_h(x) - a)/\tau)}{(c_+ - c_-)p(a \mid x)}, \qquad \lambda := \frac{(c_+ - c_-)\gamma}{\tau}. \tag{12}$$

Theorem 1 (Policy-targeted bridge loss controls regret)**.** *Under the conditions of Propositions 2 and 3, for every $h \in \mathcal{H}$,*

$$\mathrm{Reg}(h) \le \sqrt{2c_\gamma C_\rho}\, \chi_{\omega_h} \sqrt{\mathcal{L}_{\mathrm{br},\omega_h}(h)} + \sqrt{2\left(C_h \tau^2 + \frac{L}{2}(c_+ - c_-)\right)}. \tag{13}$$

Remark 1. Theorem 1 highlights the main conceptual distinction from outcome-oriented reweighting in unconfounded settings. Under unmeasured confounding, the learner cannot directly regress on counterfactual outcomes or on the causal response surface. It must instead solve a proximal inverse problem through the bridge moment equation. The theorem shows that solving this inverse problem in a policy-targeted weighted norm controls a regret surrogate, with sensitivity to the localized inverse problem captured by the weighted ill-posedness constant χ_{ω_h}.

In practice, we approximate this self-weighted objective by a lagged fixed-point procedure: a bridge fitted under the current weights induces a response surface and pseudo-optimal policy, which are then used to define the weights for the next bridge fit. Section 5 describes this practical approximation.

The preceding results explain why the weighted norm matters for regret control. It remains to check that this weighting does not change the population bridge target in the well-specified problem. The next proposition shows that, under strictly positive weights, it does not.

Proposition 4 (Population target preservation)**.** *Suppose $h_0 \in \mathcal{H}$ satisfies the outcome bridge restriction. Let $\omega : \mathcal{A} \times \mathcal{X} \to (0, \infty)$ be measurable and satisfy $\omega(A, X) > 0$ almost surely. Then h_0 is a population minimizer of $\mathcal{L}_{\mathrm{br},\omega}$. Moreover, for any $h \in \mathcal{H}$,*

$$\mathcal{L}_{\mathrm{br},\omega}(h) = 0 \quad \Longleftrightarrow \quad \mathbb{E}\{Y - h(A, W, X) \mid A, Z, X\} = 0 \quad \textit{a.s.} \tag{14}$$

Consequently, the weighted and unweighted population bridge risks have the same population minimizer.

Proof in Appendix D.7. For the policy-targeted weight in (12), the global component in the numerator, together with positivity of $p(a \mid x)$, ensures that $\omega_h(A, X) > 0$ almost surely, so the preservation result applies.

Thus, the role of the weight is not to define a different population bridge target, but to define the geometry in which deviations from the bridge equation are penalized. This distinction matters in **finite samples**, under **regularization**, and under **misspecification**, where not all bridge-moment violations can be driven to zero.

5 Algorithms

Let $\{O_i = (Y_i, A_i, Z_i, W_i, X_i)\}_{i=1}^n$ denote the observed sample, with $A_i \in \mathcal{A} = [c_-, c_+]$. Write $S_i = (A_i, W_i, X_i)$, $V_i = (A_i, Z_i, X_i)$ and $r_i(h) = Y_i - h(S_i)$. The population results above suggest

solving the proximal bridge equation in a policy-targeted weighted norm. Algorithmically, this leads to an iterative procedure that alternates between

$$\widehat{h}^{(s)} \longrightarrow \widehat{m}^{(s)} \longrightarrow \widehat{\pi}^{(s)} \longrightarrow \widehat{\omega}^{(s+1)}. \tag{15}$$

Thus the treatment rule is not read directly from the fitted bridge. Each iteration first estimates an outcome bridge, then projects the bridge into an induced causal response surface, then optimizes the treatment over this surface, and finally uses the resulting pseudo-optimal treatment to construct the next policy-targeted bridge weights.

Policy-targeted bridge fitting. Let $\widehat{p}(a \mid x)$ be an estimate of the generalized propensity score (in our experiments we estimate this conditional density using a normalizing-flow model). We initialize with the global inverse-propensity weight $\widehat{\omega}_i^{(0)} = \frac{1}{(c_+ - c_-)\widehat{p}(A_i \mid X_i)}$. This initialization makes the first bridge fit globally treatment-uniform: rather than emphasizing treatment values in proportion to how often they are observed, the inverse-density factor targets the bridge loss averaged uniformly over $\mathcal{A}$. Later iterations add the policy-localized component.

For iteration $s = 0, \ldots, S$, given weights $\widehat{\omega}^{(s)} = (\widehat{\omega}_1^{(s)}, \ldots, \widehat{\omega}_n^{(s)})$, we estimate a bridge by solving $\widehat{h}^{(s)} \in \arg\min_{h \in \mathcal{H}} \left\{ \widehat{\mathcal{R}}_{n,\widehat{\omega}^{(s)}}(h) + \eta_h \Omega_h(h) \right\}$, where $\widehat{\mathcal{R}}_{n,\widehat{\omega}}(h)$ denotes the solver-specific empirical fitting criterion and Ω_h is the corresponding regularizer. Different proximal solvers enforce the weighted bridge moment restriction through different empirical objectives, which need not be direct empirical estimates of $\mathcal{L}_{\mathrm{br},\omega}$ in (8). Their statistical properties and relationship to the ideal bridge risk are discussed in Appendix D.9.

Projection to a response surface. The fitted bridge is then projected into an induced causal response surface, $m_{\widehat{h}^{(s)}}(a, x) = \mathbb{E}\{\widehat{h}^{(s)}(a, W, x) \mid X = x\}$. In practice, we approximate this projection by evaluating the bridge on a treatment grid $\mathcal{G} = \{a_1, \ldots, a_M\} \subset \mathcal{A}$ and forming bridge pseudo-outcomes over this grid. These pseudo-outcomes are formed by cross-fitting: for each fold I_k in a partition of $\{1, \ldots, n\}$, the bridge $\widehat{h}_{-k}^{(s)}$ is trained on the observations outside I_k and evaluated on the held-out observations, giving $\widetilde{H}_{im}^{(s)} = \widehat{h}_{-k}^{(s)}(a_m, W_i, X_i)$ for $i \in I_k$ and $m = 1, \ldots, M$. We then pool these out-of-fold pseudo-outcomes across folds and regress them on (a_m, X_i) to obtain a single surface estimator $\widehat{m}^{(s)}(a, x)$. The observation-specific pseudo-optimal treatment used for weighting is $\widehat{\pi}_i^{(s)} \in \arg\max_{a \in \mathcal{G}} \widehat{m}^{(s)}(a, X_i)$.

Weight update. For $s < S$, we update the empirical weights by evaluating (12) at $(a, x) = (A_i, X_i)$, replacing p by $\widehat{p}$, and using the lagged pseudo-policy $\widehat{\pi}_i^{(s)}$. Thus the new weights combine a global inverse-density component with a localized component around the current pseudo-optimal treatment. The global component prevents the bridge fit from collapsing entirely onto the current pseudo-optimal region and stabilizes the proximal inverse problem.

Statistical guarantees. Appendix D.8 makes the statistical implications of Proposition 3 explicit, showing that, under a fixed weighting scheme and a bridge-risk oracle inequality, the induced causal surface risk is controlled by the bridge approximation and estimation errors. Appendix D.9 establishes a finite-round oracle inequality for the solver-specific population criteria $\mathcal{R}_\omega$ under adaptive weighting and a uniform deviation condition. We further provide sufficient calibration conditions under which these guarantees yield control of the ideal bridge risk and, consequently, the causal surface risk.

5.1 Solver adapters

The policy-targeted weighting scheme is a general wrapper for proximal bridge learning. In this work, we instantiate it in four representative proximal bridge solvers: **PMMR**, **NMMR**, **KPV**, and **DFPV**. The construction does not change the target outcome bridge: it still targets an outcome bridge h_0 satisfying the bridge equation. Instead, it changes the empirical norm in which violations of that equation are penalized.

The same weights $\widehat{\omega}_i$ can be used across different proximal bridge solvers. The key solver-specific step is to insert the weights at the level where each method represents the bridge moment. In PMMR and NMMR, the bridge moment is represented directly by an MMR quadratic in residual pairs. In KPV and DFPV, the bridge equation is enforced through a second-stage bridge regression after a first-stage nuisance regression. Table 1 summarizes the resulting adapters. K_V is the moment-kernel

Table 1: Solver-specific adapters for the policy-targeted bridge objective.

Solver	Baseline objective component	Policy-targeted adapter
PMMR	$r^\top K_V r$	$r^\top D_{\widehat{\omega}}^{1/2} K_V D_{\widehat{\omega}}^{1/2} r$
NMMR	$r(\theta)^\top K_V r(\theta)$	$r(\theta)^\top D_{\widehat{\omega}}^{1/2} K_V D_{\widehat{\omega}}^{1/2} r(\theta)$
KPV (DFPV)	second-stage kernel ridge	weight second-stage bridge loss by $\widehat{\omega}_i$

Gram matrix on $V = (A, Z, X)$ and $D_{\widehat{\omega}} = \mathrm{diag}(\widehat{\omega}_1, \ldots, \widehat{\omega}_n)$. For PMMR and NMMR, the adapter is equivalent to replacing each residual by $\sqrt{\widehat{\omega}_i} r_i$ inside the MMR loss. For KPV and DFPV, the weights enter as ordinary loss weights in the second-stage bridge regression; the first-stage nuisance regression is left unchanged. In the experiments we report two NMMR variants: NMMR-V uses the full V-statistic MMR objective, while NMMR-U uses the corresponding diagonal-deleted U-statistic objective. Appendix E presents the general implementation template, Appendices F and G detail the MMR-based and two-stage solver variants, respectively, and Appendix H provides additional implementation details.

6 Evaluation

Datasets and baselines. We evaluate on two proximal continuous-treatment benchmarks with hidden confounding and proxy variables. Each observation is $O = (X, Z, W, A, Y)$, where X are observed covariates, Z and W are treatment- and outcome-inducing proxies, A is a continuous treatment, and Y is the outcome. Both benchmarks have non-monotone causal response surfaces with interior, covariate-dependent optima, making regret sensitive to errors near the optimal dose. The first benchmark is a fully synthetic proximal DGP adapted from Wu et al. [42]; the second is a semi-synthetic TCGA benchmark using gene-expression features from The Cancer Genome Atlas [41]. See Appendix I. We compare PMMR, NMMR, KPV, and DFPV with their decision-aware (DA) variants. We additionally include DR-KPV and DR-PMMR, based on the doubly robust proximal estimators of Bozkurt et al. [7], adapted to conditional dose-response estimation. We also include non-proximal S-learner and weighted S-learner baselines to assess the effect of ignoring proxy-based adjustment. Base learners are tuned by held-out factual RMSE, and the selected hyperparameters are reused for their DA variants. The weighting parameters $(\tau, \lambda, n_{\mathrm{rounds}})$ are calibrated once on a synthetic validation setting and then fixed across methods and datasets; this prevents the decision-aware variants from receiving model-specific tuning based on oracle regret, which would not be available in observational applications. See Appendix J for details.

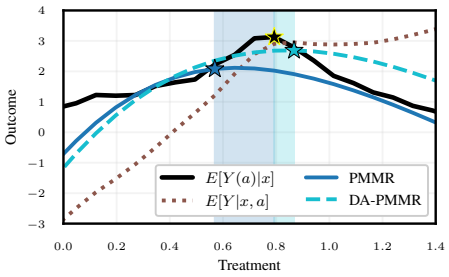
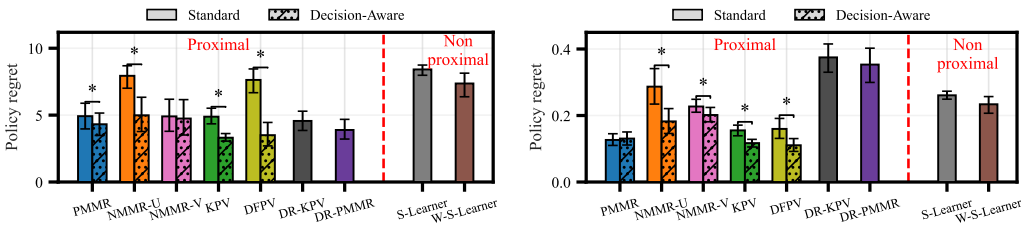


Figure 2: Single-individual example from the synthetic benchmark. DA-PMMR improves the response estimate near the oracle optimum, yielding a treatment recommendation closer to the optimal dose even though it does not uniformly improve the full response surface.

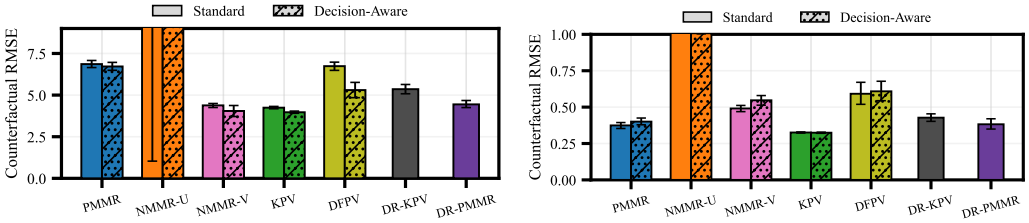
Results. Figure 3 reports policy regret on the synthetic and semi-synthetic benchmarks. Across both datasets, *decision-aware weighting generally reduces regret for proximal bridge learners*. The gains are especially visible for the solvers whose baseline response surfaces place substantial error near the treatment optimum, such as NMMR-U, KPV, and DFPV. In the synthetic benchmark, the best-performing method is DA-KPV, while in the semi-synthetic benchmark the lowest regret is obtained by DA-DFPV; Wilcoxon tests show statistical differences between DA and baseline versions for most models. Moreover, the best-performing DA variants achieve lower policy regret than both doubly robust proximal baselines across the two benchmarks. Also, we see that the non-proximal S-learner baselines are consistently worse. This supports the main hypothesis that, under hidden confounding, proxy-based bridge learning is necessary, and that reweighting the bridge loss toward decision-relevant treatment regions can further improve treatment selection.

Figure 4 reports counterfactual-estimation RMSE over the full treatment region sampled uniformly. The results are mixed: overall, we do not observe substantial differences between the decision-aware (DA) variants and their baseline counterparts. This is expected, since the proposed objective is



(a) Synthetic experiment (b) semi-synthetic experiment

Figure 3: Regret metrics in the synthetic and semi-synthetic datasets. Mean and 95% confidence intervals over 10 seeds are shown. * denotes statistical difference between baselines and DA models (paired Wilcoxon signed-rank test, $p < 0.05$).

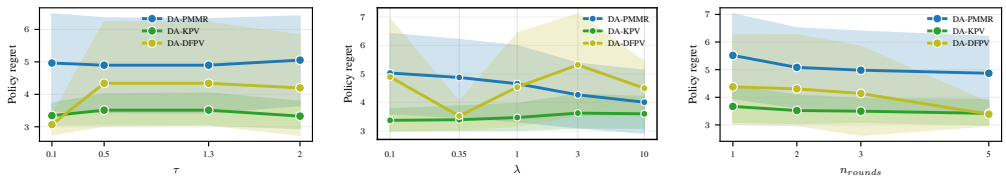


(a) Synthetic experiment (b) semi-synthetic experiment

Figure 4: Counterfactual RMSE in the synthetic and semi-synthetic datasets. Mean and 95% confidence intervals over 10 seeds are shown. The behavior of NMMR-U can be more clearly appreciated in Appendix K.1, where we provide the corresponding counterfactual-RMSE tables.

not designed to reduce global counterfactual error uniformly over the treatment space, but rather to improve the learned response surface near treatment values that affect the induced policy. The individual-level example in Figure 2 illustrates this distinction. Importantly, the DA variants *do not appear to degrade general counterfactual estimation*. The doubly robust baselines also exhibit counterfactual RMSE comparable to several of the other proximal estimators. In Appendix K.1, we provide the corresponding counterfactual-RMSE tables, where the behavior of NMMR-U can be more clearly seen. We also report factual prediction metrics, which lead to a similar conclusion.

Sensitivity to weighting hyperparameters. We vary the weighting bandwidth τ, localization strength λ, and number of reweighting rounds n_{rounds} one at a time. The results in Figure 5 show that performance is relatively stable over a nontrivial range of bandwidths τ. DA-PMMR and DA-KPV change only mildly as τ varies, while DA-DFPV is somewhat more sensitive, with stronger performance for narrower localization in this experiment. Varying λ shows the expected trade-off between global stabilization and local policy targeting. Larger values tend to improve DA-PMMR by concentrating more weight around the pseudo-optimal treatment, whereas DA-KPV remains comparatively stable and DA-DFPV exhibits a more non-monotone response. Finally, increasing



(a) Varying τ (b) Varying λ (c) Varying n_{rounds}

Figure 5: Sensitivity of weighting hyperparameters, holding the other two fixed. Mean and 95% confidence intervals over the regret in 5 seeds reported.

n_{rounds} generally reduces regret, consistent with the iterative interpretation of the method. The gains become smaller after a few rounds.

Sensitivity to hidden confounding. We vary the hidden-confounding strength on the synthetic benchmark. As shown in Figure 6, regret increases for all methods as confounding becomes stronger. The non-proximal S-learner deteriorates sharply, highlighting the importance of proxy-based adjustment. Among proximal methods, decision-aware weighting often improves regret, especially for PMMR and DFPV at moderate to high confounding levels. For this experiment, we used only 5 seeds while for the main experiments we used 10.

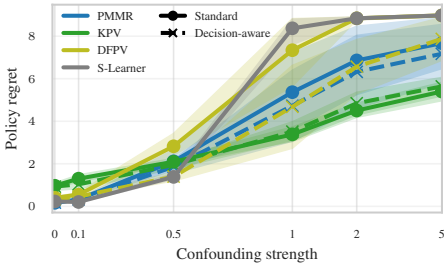


Figure 6: Regret as a function of the confounding strength for three of the DA models. Mean and 95% confidence intervals over 5 seeds reported.

Sensitivity to weakened completeness. We also assess sensitivity to completeness, a central proximal identification assumption, by perturbing only the proxy-generation mechanism in the synthetic benchmark. At degradation level α, latent values in an expanding central band are removed before entering the corresponding proxies, and independent Gaussian noise with increasing variance is subsequently added; $\alpha = 0$ recovers the original DGP. Across PMMR, KPV, and DFPV, performance generally deteriorates as α increases, consistently with the progressive loss of proxy information. Nevertheless, the decision-aware variants attain lower mean policy regret than their unweighted counterparts at every degradation level considered, while global counterfactual RMSE remains broadly comparable. Thus, decision-aware weighting can remain beneficial under moderate proxy degradation, although it cannot recover identifying information that the proxies no longer contain. Full details and results are provided in Appendix K.2.

Pure-local ablation. Finally, we compare unweighted proximal learners, pure-local weighting, and local-plus-global decision-aware weighting on the synthetic benchmark. The local-plus-global variant attains lower mean policy regret and global counterfactual RMSE than pure-local weighting for PMMR, KPV, and DFPV (Appendix K.3). These findings support the intended role of the global component in preventing exclusive focus on the current pseudo-optimal treatment region and providing global stabilization.

7 Conclusion

We introduced a decision-aware framework for proximal bridge learning in individualized continuous-treatment selection. Rather than estimating the causal response surface uniformly over the treatment space, the proposed objective weights bridge estimation toward treatment regions that are most relevant for the induced policy, while retaining a global component for stability. Our theory shows that the resulting weighted bridge loss controls a regret-oriented surface surrogate through a weighted ill-posedness condition, without changing the population bridge target. We instantiated this idea as a general wrapper around several proximal bridge solvers, including PMMR, NMMR, KPV, and DFPV. Across synthetic and semi-synthetic benchmarks with hidden confounding and proxy variables, decision-aware weighting generally reduced policy regret, often without corresponding improvements in factual prediction error. These results suggest that, in proximal causal inference, aligning bridge estimation with the downstream treatment-selection objective can improve decision quality beyond globally accurate response estimation. As with other proximal methods, the main limitation of our approach is its reliance on identification assumptions involving valid proxies and bridge existence, which may be difficult to verify in practice.

Acknowledgments

We gratefully acknowledge Novartis for sponsoring Tomàs Garriga's industrial PhD. We also thank the Government of Catalonia's Industrial PhDs Plan for funding part of this research. We acknowledge Horizon Europe Programme under the AI4DEBUNK Project (https://www.ai4debunk.eu), grant agreement num. 101135757. Alejandro Almodóvar and Juan Parras have received the funding from

the SYNTHIA project, supported by the Innovative Health Initiative Joint Undertaking (IHI JU) under grant agreement No 101172872. Funded by the European Union, the private members, and those contributing partners of the IHI JU. Views and opinions expressed are, however, those of the authors only and do not necessarily reflect those of the aforementioned funding agencies. Neither of the aforementioned parties can be held responsible for them.

The authors declare no competing interests beyond the affiliations and funding sources disclosed above.

References

[1] Takuya Akiba, Shotaro Sano, Toshihiko Yanase, Takeru Ohta, and Masanori Koyama. Optuna: A next-generation hyperparameter optimization framework. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2019.

[2] Alejandro Almodóvar, Adrián Javaloy, Juan Parras, Santiago Zazo, and Isabel Valera. DeCaFlow: A deconfounding causal generative model. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pages 136612–136666, 2025. doi: 10.52202/085713-4558.

[3] James Bergstra, Rémi Bardenet, Yoshua Bengio, and Balázs Kégl. Algorithms for hyper-parameter optimization. In *Advances in Neural Information Processing Systems*, 2011.

[4] Ioana Bica, Ahmed M. Alaa, James Jordon, and Mihaela van der Schaar. Estimating counterfactual treatment outcomes over time through adversarially balanced representations. In *Proceedings of the 8th International Conference on Learning Representations*, 2020.

[5] Ioana Bica, James Jordon, and Mihaela van der Schaar. Estimating the effects of continuous-valued interventions using generative adversarial networks. *Advances in neural information processing systems*, 2020.

[6] Bariscan Bozkurt, Ben Deaner, Dimitri Meunier, Liyuan Xu, and Arthur Gretton. Density ratio-based proxy causal learning without density ratios. In *Proceedings of The 28th International Conference on Artificial Intelligence and Statistics*, volume 258 of *Proceedings of Machine Learning Research*, pages 5095–5103. PMLR, 2025.

[7] Bariscan Bozkurt, Houssam Zenati, Dimitri Meunier, Liyuan Xu, and Arthur Gretton. Density ratio-free doubly robust proxy causal learning. In *Advances in Neural Information Processing Systems*, 2025.

[8] Xiaohong Chen and Demian Pouzo. Estimation of nonparametric conditional moment models with possibly nonsmooth generalized residuals. *Econometrica*, 80(1):277–321, 2012.

[9] Yifan Cui, Hongming Pu, Xu Shi, Wang Miao, and Eric J. Tchetgen Tchetgen. Semiparametric proximal causal inference. *Journal of the American Statistical Association*, 119(546):1348–1359, 2024.

[10] Nishanth Dikkala, Greg Lewis, Lester W. Mackey, and Vasilis Syrgkanis. Minimax estimation of conditional moment models. In *Advances in Neural Information Processing Systems*, volume 33, pages 12248–12262, 2020.

[11] Adam N. Elmachtoub and Paul Grigas. Smart "predict, then optimize". *Management Science*, 68(1):9–26, 2022.

[12] Carlos Fernández-Loría and Foster Provost. Causal decision making and causal effect estimation are not the same... and why it matters. *INFORMS Journal on Data Science*, 1(1):4–16, 2022.

[13] Dennis Frauen, Valentyn Melnychuk, Jonas Schweisthal, Mihaela van der Schaar, and Stefan Feuerriegel. Treatment effect estimation for optimal decision-making. In D. Belgrave, C. Zhang, H. Lin, R. Pascanu, P. Koniusz, M. Ghassemi, and N. Chen, editors, *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pages 133883–133918. Curran Associates, Inc., 2025.

[14] Tomas Garriga, Gerard Sanz, Eduard Serrahima de Cambra, and Axel Brando. CEPAE: Conditional entropy-penalized autoencoders for time series counterfactuals. *Transactions on Machine Learning Research*, 2026. ISSN 2835-8856.

[15] Tomàs Garriga, Valentyn Melnychuk, Konstantin Hess, Eduard Serrahima de Cambra, Axel Brando, Gerard Sanz, and Stefan Feuerriegel. Orthogen: A generative orthogonal learner for time-varying treatments, 2026. URL https://arxiv.org/abs/2610.10210.

[16] Amiremad Ghassami, Andrew Ying, Ilya Shpitser, and Eric Tchetgen Tchetgen. Minimax kernel machine learning for a class of doubly robust functionals with application to proximal causal inference. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, volume 151 of *Proceedings of Machine Learning Research*, pages 7210–7239. PMLR, 28–30 Mar 2022.

[17] Miguel A. Hernán and James M. Robins. *Causal Inference: What If*. Chapman & Hall/CRC, 2020.

[18] Konstantin Hess, Dennis Frauen, Valentyn Melnychuk, and Stefan Feuerriegel. IGC-net for conditional average potential outcome estimation over time. In *The Fourteenth International Conference on Learning Representations*, 2026.

[19] Keisuke Hirano and Guido W. Imbens. The propensity score with continuous treatments. In Andrew Gelman and Xiao-Li Meng, editors, *Applied Bayesian Modeling and Causal Inference from Incomplete-Data Perspectives*, pages 73–84. John Wiley & Sons, Hoboken, NJ, 2004.

[20] Guido W. Imbens and Donald B. Rubin. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction*. Cambridge University Press, 2015.

[21] Fredrik D. Johansson, Uri Shalit, and David Sontag. Learning representations for counterfactual inference. In *Proceedings of the 33rd International Conference on Machine Learning*, pages 3020–3029, 2016.

[22] Nathan Kallus and Angela Zhou. Policy evaluation and optimization with continuous treatments. In Amos Storkey and Fernando Perez-Cruz, editors, *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pages 1243–1251. PMLR, 2018.

[23] Nathan Kallus, Xiaojie Mao, and Masatoshi Uehara. Causal inference under unmeasured confounding with negative controls: A minimax learning approach, 2021. URL https://arxiv.org/abs/2103.14029.

[24] Benjamin Kompa, David Bellamy, Tom Kolokotrones, James M. Robins, and Andrew Beam. Deep learning methods for proximal inference via maximum moment restriction. In *Advances in Neural Information Processing Systems*, volume 35, 2022.

[25] Rui Li, Stephanie Hu, Mingyu Lu, Yuria Utsumi, Prithwish Chakraborty, Daby M Sow, Piyush Madan, Jun Li, Mohamed Ghalwash, Zach Shahn, et al. G-net: a recurrent network approach to g-computation for counterfactual prediction under a dynamic treatment regime. In *Machine Learning for Health*, pages 282–299. PMLR, 2021.

[26] Jayanta Mandi, James Kotary, Senne Berden, Maxime Mulamba, Victor Bucarey, Tias Guns, and Ferdinando Fioretto. Decision-focused learning: Foundations, state of the art, benchmark and future opportunities. *Journal of Artificial Intelligence Research*, 80, 2024.

[27] Afsaneh Mastouri, Yuchen Zhu, Limor Gultchin, Anna Korba, Ricardo Silva, Matt Kusner, Arthur Gretton, and Krikamol Muandet. Proximal causal learning with kernels: Two-stage estimation and moment restriction. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 7512–7523. PMLR, 18–24 Jul 2021.

[28] Valentyn Melnychuk, Dennis Frauen, and Stefan Feuerriegel. Causal transformer for estimating counterfactual outcomes. In *Proceedings of the 39th International Conference on Machine Learning*, pages 15293–15329, 2022.

[29] Wang Miao, Zhi Geng, and Eric J. Tchetgen Tchetgen. Identifying causal effects with proxy variables of an unmeasured confounder. *Biometrika*, 105(4):987–993, December 2018.

[30] Wang Miao, Wenjie Hu, Elizabeth L. Ogburn, and Xiao-Hua Zhou. Identifying effects of multiple treatments in the presence of unmeasured confounding. *Journal of the American Statistical Association*, 118(543):1953–1967, 2023.

[31] Wang Miao, Xu Shi, Yilin Li, and Eric J. Tchetgen Tchetgen. A confounding bridge approach for double negative control inference on causal effects. *Statistical Theory and Related Fields*, 8 (4):262–273, October 2024.

[32] Álvaro Parafita and Jordi Vitrià. Estimand-agnostic causal query estimation with deep causal graphs. *IEEE Access*, 10:71370–71386, 2022.

[33] Judea Pearl. *Causality: Models, Reasoning, and Inference*. Cambridge University Press, 2 edition, 2009.

[34] Zhengling Qi, Rui Miao, and Xiaoke Zhang. Proximal learning for individualized treatment regimes under unmeasured confounding. *Journal of the American Statistical Association*, 119 (546):915–928, 2024.

[35] Donald B. Rubin. Estimating causal effects of treatments in randomized and nonrandomized studies. *Journal of Educational Psychology*, 66(5):688–701, 1974.

[36] Uri Shalit, Fredrik D. Johansson, and David Sontag. Estimating individual treatment effect: Generalization bounds and algorithms. In *Proceedings of the 34th International Conference on Machine Learning*, pages 3076–3085, 2017.

[37] Tao Shen and Yifan Cui. Optimal treatment regimes for proximal causal learning. In *Advances in Neural Information Processing Systems*, volume 36, pages 47735–47748, 2023.

[38] Rahul Singh, Liyuan Xu, and Arthur Gretton. Kernel methods for causal functions: Dose, heterogeneous and incremental response curves. *Biometrika*, 111(2):497–516, 2024.

[39] Eric J. Tchetgen Tchetgen, Andrew Ying, Yifan Cui, Xu Shi, and Wang Miao. An introduction to proximal causal inference. *Statistical Science*, 39(3):375–390, 2024.

[40] Stefan Wager and Susan Athey. Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association*, 113(523):1228–1242, 2018.

[41] John N. Weinstein, Eric A. Collisson, Gordon B. Mills, Kenna R. Mills Shaw, Brad A. Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander, Joshua M. Stuart, and Cancer Genome Atlas Research Network. The cancer genome atlas pan-cancer analysis project. *Nature Genetics*, 45(10):1113–1120, 2013.

[42] Yong Wu, Yanwei Fu, Shouyan Wang, and Xinwei Sun. Doubly robust proximal causal learning for continuous treatments. In *International Conference on Learning Representations*, 2024.

[43] Liyuan Xu, Heishiro Kanagawa, and Arthur Gretton. Deep proxy causal learning and its application to confounded bandit policy evaluation. In *Advances in Neural Information Processing Systems*, volume 34, pages 26264–26275, 2021.

[44] Andrew Ying, Wang Miao, Xu Shi, and Eric J. Tchetgen Tchetgen. Proximal causal inference for complex longitudinal studies. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 85(3):684–704, 2023.

[45] Jeffrey Zhang, Wei Li, Wang Miao, and Eric J. Tchetgen Tchetgen. Proximal causal inference without uniqueness assumptions. *Statistics & Probability Letters*, 198:109836, 2023.

[46] Yingqi Zhao, Donglin Zeng, A. John Rush, and Michael R. Kosorok. Estimating individualized treatment rules using outcome weighted learning. *Journal of the American Statistical Association*, 107(499):1106–1118, 2012.

[47] Hao Zou, Bo Li, Jiangang Han, Shuiping Chen, Xuetao Ding, and Peng Cui. Counterfactual prediction for outcome-oriented treatments. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 27693–27706. PMLR, 17–23 Jul 2022.

Appendix Index

A Extended Related Work **17**

B Standard proximal inference assumptions **18**

C Sufficient conditions for outcome-bridge existence **18**

D Proofs for the Policy-Targeted Risk **20**

D.1 Population objects and notation 20

D.2 Regret decomposition 21

D.3 Kernel approximation and global surface bounds 21

D.4 Proof of the weighted surface surrogate 23

D.5 Bridge-to-surface control 24

D.6 The weighted ill-posedness condition 24

D.7 Proof of Population Target Preservation 26

D.8 Fixed-weight statistical transfer to surface risk 26

D.9 A finite-round oracle inequality for adaptive weights 27

E Practical Implementation Template for Policy-Targeted Proximal Solvers **28**

E.1 Policy-targeted weights 28

E.2 Two ways in which the weights enter 29

E.3 Why the weighted objective still targets the bridge equation 29

F MMR-Based Solvers: PMMR and NMMR **29**

F.1 Weighted maximum moment restriction 30

F.2 Policy-targeted PMMR 30

F.3 Policy-targeted NMMR 31

G Two-Stage Solvers: KPV and DFPV **31**

G.1 Policy-targeted KPV 31

G.2 Policy-targeted DFPV 33

H Implementation Details and Pitfalls **34**

I Dataset Details **34**

I.1 Synthetic proximal continuous-treatment benchmark 34

I.2 Semi-synthetic TCGA proximal benchmark 36

I.3 Monte Carlo construction of ground-truth causal curves 39

J Hyperparameter Selection **40**

J.1 Stage 1: Base model selection using factual validation error 40

J.2 Stage 2: Shared calibration of decision-aware weighting 41

J.3 Sensitivity Studies 41

K Additional Results **41**

K.1 Global RMSE 41

K.2 Sensitivity to Weakened Completeness 42

K.3 Pure-Local versus Local–Global Weighting 44

A Extended Related Work

Causal inference and representation learning. Causal inference provides formal tools for reasoning about interventions and counterfactuals from randomized or observational data, with foundations in potential outcomes, graphical models, and semiparametric estimation [35, 33, 20, 17]. Modern machine-learning approaches build on these ideas by learning representations, nuisance functions, or policies for heterogeneous treatment-effect estimation, decision making and explainability [21, 36, 40, 32]. Related ideas appear in several domains, including longitudinal and time-series settings [4, 28, 25, 14, 15, 18]. Our setting differs from this line of work because the relevant dose-response is continuous and hidden confounding is addressed through negative-control proxies rather than by assuming that all confounders are observed.

Decision-oriented causal learning and continuous treatments. A large literature studies individualized treatment rules, treatment assignment, and policy learning from observational or logged data. Classical individualized-treatment-rule methods often reduce value maximization to a weighted classification or regression problem; for example, outcome-weighted learning estimates treatment rules by weighting classification decisions according to observed outcomes and treatment propensities [46]. More broadly, work on causal decision making has emphasized that optimizing a treatment rule is not the same statistical objective as estimating treatment effects or potential-outcome surfaces uniformly well [12]. This distinction is especially important for continuous treatments, where the learner must recommend a dose or intensity rather than choose among a small number of actions. Early work on generalized propensity scores formalized dose-response estimation for continuous exposures [19], while later work developed off-policy evaluation and optimization methods for continuous actions using kernelized inverse-propensity and doubly robust estimators [22]. These approaches motivate our continuous-treatment setting and our focus on regret, but they typically rely on measured-covariate exchangeability rather than proxy-based identification under latent confounding.

Outcome-oriented and decision-focused counterfactual prediction. Our motivation is closely related to the broader decision-focused learning principle that prediction errors should be weighted by their downstream decision consequences. In operations and machine learning, predict-then-optimize methods such as SPO train predictors using losses that measure the quality of the induced decision rather than only pointwise prediction error [11]; recent surveys describe decision-focused learning as an end-to-end approach for aligning predictive models with constrained optimization objectives [26]. In causal inference, Zou et al. [47] make an analogous point for counterfactual prediction: global accuracy over the treatment space may be poorly aligned with treatment-selection quality, and estimators should emphasize regions that affect the recommended action. Our work adopts this outcome-oriented perspective, but differs in the source of identification. Existing outcome-oriented counterfactual prediction methods generally assume that confounding is removed by observed covariates, whereas our setting requires proximal bridge functions and proxy variables to handle unmeasured confounding.

Proximal causal inference. Proximal causal inference addresses latent confounding by using observed proxies, often formalized as treatment-inducing proxies Z and outcome-inducing proxies W, to solve bridge-function moment equations. Early nonparametric identification results with proxy variables were established by Miao et al. [29], while the double-negative-control bridge formulation was developed by Miao et al. [31]. Tchetgen Tchetgen et al. [39] placed these ideas in a potential-outcome framework and derived the proximal g-formula, showing how causal effects can be identified even when exchangeability with respect to measured covariates fails. Subsequent work has expanded the framework to semiparametric and doubly robust inference [9, 16], multiple-treatment settings [30], longitudinal studies [44], and settings where bridge solutions may not be unique [45]. This literature provides the identification foundation for our method. The key difference is that most proximal estimands and estimators are designed for effect or response-surface recovery, whereas we target the downstream continuous-treatment decision induced by the learned bridge.

Machine learning for bridge estimation. A central computational challenge in proximal inference is that bridge functions solve ill-posed conditional moment or Fredholm integral equations. Kernel methods by Mastouri et al. [27] formulate proximal causal learning through two-stage regression and maximum moment restriction in reproducing-kernel Hilbert spaces. Minimax approaches estimate bridge functions through adversarial conditional-moment objectives [23, 16], and Kompa et al. [24]

scale this idea with neural maximum moment restriction. For continuous treatments, Wu et al. [42] propose a kernel-smoothed doubly robust proximal estimator, addressing the difficulty that binary-treatment proximal doubly robust estimators do not transfer directly to continuous actions because exact treatment matching has probability zero. Almodóvar et al. [2] propose a deconfounding normalizing flow that, applied to proximal settings, implicitly solves the bridge equation. These works improve the statistical and computational toolkit for estimating bridges and causal response functions. Our objective is complementary: rather than changing only the bridge estimator class or the moment-restriction solver, we change the loss being optimized so that bridge fitting is concentrated in treatment regions that matter for the induced policy, while retaining a global component for stabilization.

Proximal treatment-rule learning. The closest proximal decision-making literature studies optimal treatment regimes under unmeasured confounding. Qi et al. [34] develop proximal learning procedures for individualized treatment regimes by combining bridge-based identification with classification-style policy learning, and Shen and Cui [37] characterize optimal treatment regimes using outcome and treatment confounding bridges. These works show that proximal identification can support decision making under latent confounding. They primarily focus on identifying and optimizing treatment-regime values over specified policy classes, often in discrete-action settings. Our focus is different but complementary: we study continuous treatment selection and ask how the proximal bridge-learning objective itself should be shaped when the downstream criterion is dose-selection regret. In this sense, our method connects proximal policy learning with outcome-oriented counterfactual prediction by introducing a localized, policy-aware bridge loss.

B Standard proximal inference assumptions

In this appendix, we provide additional discussion of Assumptions 1–4, which are stated in the main text.

Assumption 1: Consistency and latent exchangeability. Assumption 1 says that treatment is as-if randomized after conditioning on the observed covariates X and the latent confounders U. Since U is unobserved, this condition cannot be used directly for ordinary covariate adjustment.

Assumption 2: Negative-control proxies. Assumption 2 requires that, for every $a \in \mathcal{A}$, $Y(a) \perp\!\!\!\perp Z \mid A, U, X$ and $W \perp\!\!\!\perp (A, Z) \mid U, X$.

The first relation says that, after conditioning on treatment, observed covariates, and latent confounders, the treatment-inducing proxy Z has no residual association with the potential outcome. The second says that, after conditioning on (U, X), the outcome-inducing proxy W has no residual association with treatment assignment or with Z. Thus, the proxies are useful because they carry information about latent confounding while satisfying exclusion restrictions that make proximal identification possible.

Assumption 3: Completeness. Assumption 3 is a proxy-relevance condition. It rules out nonzero latent or proxy signals that are invisible after conditioning on the observed proxy system. In discrete settings, it requires the proxies to contain enough variation to distinguish the relevant latent confounding states; in continuous settings, it is the nonparametric analogue of an injectivity condition for conditional-expectation operators.

Assumption 4: Positivity. Assumption 4 ensures that every treatment level in the target action space is observable with positive probability, both within latent strata and after marginalizing over the latent confounders.

C Sufficient conditions for outcome-bridge existence

This appendix records one set of sufficient conditions under which the outcome bridge used in the main text exists. The main results of the paper are developed conditional on a valid bridge equation; the conditions below are included to connect this bridge equation to the standard proximal identification framework.

Throughout this appendix, assume that Assumptions 1– 4 hold. For each (a, x), define the latent outcome regression

$$q_0(a, u, x) := \mathbb{E}\{Y \mid A = a, U = u, X = x\}. \tag{16}$$

By Assumption 2,

$$\mathbb{E}\{Y \mid A = a, Z = z, X = x\} = \mathbb{E}\{q_0(a, U, x) \mid A = a, Z = z, X = x\}. \tag{17}$$

We impose the following additional regularity and solvability conditions.

Assumption 5 (Regular conditional laws and square integrability)**.** *The variables (Y, A, Z, W, X, U) admit regular conditional distributions. For every $a \in \mathcal{A}$, the regression $q_0(a, U, X)$ is square-integrable. Moreover, the conditional laws of W given (U, X) and of U given (A, Z, X) are such that the conditional expectations appearing below are well defined as maps between the corresponding L_2 spaces.*

Assumption 6 (Latent bridge solvability)**.** *For every $a \in \mathcal{A}$, there exists a square-integrable function*

$$h_0(a, \cdot, \cdot) : \mathcal{W} \times \mathcal{X} \to \mathbb{R} \tag{18}$$

such that, for almost every (u, x),

$$q_0(a, u, x) = \mathbb{E}\{h_0(a, W, x) \mid U = u, X = x\}. \tag{19}$$

Assumption 6 is a range condition for the conditional-expectation operator that maps functions of (W, X) to functions of (U, X). It says that the latent outcome regression $q_0(a, U, X)$ can be represented by averaging a function of the outcome proxy W over the conditional distribution of W given (U, X).

Proposition 5 (Outcome-bridge existence)**.** *Under Assumptions 1– 4 and 5–6, the function h_0 in Assumption 6 satisfies the observed outcome-bridge equation*

$$\mathbb{E}\{Y - h_0(A, W, X) \mid A, Z, X\} = 0. \tag{20}$$

Equivalently, for every (a, z, x) in the support of (A, Z, X),

$$\mathbb{E}\{Y \mid A = a, Z = z, X = x\} = \mathbb{E}\{h_0(a, W, x) \mid A = a, Z = z, X = x\}. \tag{21}$$

Proof. Fix (a, z, x) in the support of (A, Z, X). By iterated expectation and the first negative-control relation,

$$\begin{aligned}\mathbb{E}\{Y \mid A = a, Z = z, X = x\} &= \mathbb{E}[\mathbb{E}\{Y \mid A = a, Z = z, U, X = x\} \mid A = a, Z = z, X = x] \\ &= \mathbb{E}[\mathbb{E}\{Y \mid A = a, U, X = x\} \mid A = a, Z = z, X = x] \\ &= \mathbb{E}\{q_0(a, U, x) \mid A = a, Z = z, X = x\}.\end{aligned} \tag{22}$$

By Assumption 6,

$$q_0(a, U, x) = \mathbb{E}\{h_0(a, W, x) \mid U, X = x\}. \tag{23}$$

Therefore,

$$\mathbb{E}\{Y \mid A = a, Z = z, X = x\} = \mathbb{E}[\mathbb{E}\{h_0(a, W, x) \mid U, X = x\} \mid A = a, Z = z, X = x]. \tag{24}$$

By the second negative-control relation, $W \perp\!\!\!\perp (A, Z) \mid U, X$, so

$$\mathbb{E}\{h_0(a, W, x) \mid U, A = a, Z = z, X = x\} = \mathbb{E}\{h_0(a, W, x) \mid U, X = x\}. \tag{25}$$

Applying iterated expectation again gives

$$\begin{aligned}\mathbb{E}\{Y \mid A = a, Z = z, X = x\} &= \mathbb{E}[\mathbb{E}\{h_0(a, W, x) \mid U, A = a, Z = z, X = x\} \mid A = a, Z = z, X = x] \\ &= \mathbb{E}\{h_0(a, W, x) \mid A = a, Z = z, X = x\}.\end{aligned} \tag{26}$$

Hence

$$\mathbb{E}\{Y - h_0(A, W, X) \mid A = a, Z = z, X = x\} = 0, \tag{27}$$

and the claim follows. □

Remark 2 (Observed Fredholm form)**.** The conclusion of Proposition 5 can also be written as the observed Fredholm equation

$$\mathbb{E}\{Y \mid A = a, Z = z, X = x\} = \int h_0(a, w, x)\, dP(w \mid A = a, Z = z, X = x). \tag{28}$$

Thus the outcome bridge is a solution to a first-kind integral equation. The range condition in Assumption 6 is one way to ensure that this integral equation has a square-integrable solution.

D Proofs for the Policy-Targeted Risk

This appendix proves the population bounds used in Section 4. The proof proceeds in three steps. First, Lemma 1 reduces policy regret to two surface errors: one at the treatment selected by the learned bridge and one at the oracle treatment. Second, Propositions 6 and 7 show how these two errors are controlled by the local-plus-global surface surrogate. Third, Proposition 3 connects the resulting weighted surface loss to the weighted proximal bridge loss.

D.1 Population objects and notation

For completeness, we restate the objects used in the proofs. Let $O = (Y, A, Z, W, X)$ denote a generic observation, with $A \in \mathcal{A} = [c_-, c_+]$. Let h_0 be an outcome bridge satisfying

$$\mathbb{E}\{Y - h_0(A, W, X) \mid A, Z, X\} = 0, \tag{29}$$

and let $h \in \mathcal{H}$ be a candidate bridge. We write

$$\Delta_h(a, w, x) := h(a, w, x) - h_0(a, w, x) \tag{30}$$

for its bridge error. The true and bridge-induced causal response surfaces are

$$m_0(a, x) := \mathbb{E}\{h_0(a, W, x) \mid X = x\}, \qquad m_h(a, x) := \mathbb{E}\{h(a, W, x) \mid X = x\}, \tag{31}$$

where the conditional expectation over W is with respect to $P_{W|X=x}$. The squared surface error is

$$G_h(t, x) := \{m_0(t, x) - m_h(t, x)\}^2. \tag{32}$$

The policy induced by h and the oracle policy are

$$\pi_h(x) \in \operatorname*{argmax}_{t \in \mathcal{A}} m_h(t, x), \qquad a^\star(x) \in \operatorname*{argmax}_{t \in \mathcal{A}} m_0(t, x), \tag{33}$$

with the same fixed tie-breaking convention as in the main text. The regret is

$$\mathrm{Reg}(h) := \mathbb{E}\{m_0(a^\star(X), X) - m_0(\pi_h(X), X)\}. \tag{34}$$

We use two population laws for (A, W, X). The first is the observed joint law

$$\nu := P_{A,W,X}. \tag{35}$$

The second is the product conditional law

$$\mu := P_{W|X} \otimes P_{A|X} \otimes P_X, \tag{36}$$

meaning that, under μ, $X \sim P_X$ and then A and W are drawn independently from their observed conditional laws $P_{A|X}$ and $P_{W|X}$. Thus μ is the law naturally associated with the surface quantity $\mathbb{E}\{h(a, W, x) \mid X = x\}$ evaluated at treatments distributed as $P_{A|X}$, while ν is the observed law under which the bridge function is learned. We assume $\mu \ll \nu$ when invoking the bridge-to-surface result and write

$$C_\rho := \left\| \frac{d\mu}{d\nu} \right\|_\infty < \infty, \tag{37}$$

where the essential supremum is taken with respect to ν. Equivalently, for all nonnegative measurable q,

$$\int q \, d\mu \le C_\rho \int q \, d\nu. \tag{38}$$

For a nonnegative measurable weight $\omega : \mathcal{A} \times \mathcal{X} \to [0, \infty)$, define weighted norms by

$$\|f\|^2_{L_2(\omega \, d\nu)} := \int \omega(a, x) f(a, w, x)^2 \, d\nu(a, w, x), \tag{39}$$

and analogously for $L_2(\omega \, d\mu)$ and $L_2(\omega \, dP_{A,Z,X})$. The outcome-bridge operator is

$$(T\Delta)(a, z, x) := \mathbb{E}\{\Delta(a, W, x) \mid A = a, Z = z, X = x\}. \tag{40}$$

The weighted surface and bridge risks are

$$\mathcal{L}_{\mathrm{surf},\omega}(h) := \mathbb{E}\big[\omega(A,X)\{m_h(A,X) - m_0(A,X)\}^2\big], \tag{41}$$

$$\mathcal{L}_{\mathrm{br},\omega}(h) := \mathbb{E}\big[\omega(A,X)\mathbb{E}\{Y - h(A,W,X) \mid A,Z,X\}^2\big]. \tag{42}$$

Finally, the weighted proximal ill-posedness constant is

$$\chi_\omega := \sup_{g\in\mathcal{H}:\, \|T\Delta_g\|_{L_2(\omega\, dP_{A,Z,X})}>0} \frac{\|\Delta_g\|_{L_2(\omega\, d\nu)}}{\|T\Delta_g\|_{L_2(\omega\, dP_{A,Z,X})}}. \tag{43}$$

The condition $\chi_\omega < \infty$ is a weighted stability condition for the proximal inverse problem: it rules out bridge-error directions that are large in the weighted bridge norm but nearly invisible through the weighted conditional moment operator.

When expectations of the form $\mathbb{E}_{X,A}$ appear in the surface bounds, A is a generic treatment draw from $P_{A|X}$ conditional on X. This notation is only a device for writing integrals over treatment values. With this convention,

$$A_\tau(h) := \mathbb{E}_{X,A}\left[\frac{K((\pi_h(X) - A)/\tau)}{\tau p(A \mid X)} G_h(A,X)\right], \tag{44}$$

$$B(h) := \mathbb{E}_X\left[\frac{1}{c_+ - c_-}\int_{c_-}^{c_+} G_h(t,X)\, dt\right]. \tag{45}$$

D.2 Regret decomposition

We first prove the basic regret decomposition. This step explains why the subsequent local and global surface bounds focus on the two treatment values $\pi_h(X)$ and $a^\star(X)$.

Proof of Lemma 1. Fix x and abbreviate $\pi_h(x)$ by π_h and $a^\star(x)$ by $a^\star$. Since π_h maximizes $m_h(\cdot,x)$ over $\mathcal{A}$,

$$m_h(\pi_h,x) \geq m_h(a^\star,x). \tag{46}$$

Therefore,

$$m_0(a^\star,x) - m_0(\pi_h,x) \leq m_0(a^\star,x) - m_h(a^\star,x) + m_h(\pi_h,x) - m_0(\pi_h,x) \tag{47}$$

$$\leq |m_0(a^\star,x) - m_h(a^\star,x)| + |m_h(\pi_h,x) - m_0(\pi_h,x)|. \tag{48}$$

Taking expectation over X and applying Cauchy–Schwarz to each absolute-error term gives

$$\mathrm{Reg}(h) \leq \sqrt{\mathbb{E}[G_h(a^\star(X),X)]} + \sqrt{\mathbb{E}[G_h(\pi_h(X),X)]}. \tag{49}$$

□

D.3 Kernel approximation and global surface bounds

The local term $A_\tau(h)$ is a kernel approximation to the squared surface error at the bridge-induced treatment $\pi_h(X)$. For the local approximation argument, we take K to be a nonnegative symmetric kernel supported on $[-1,1]$, satisfying

$$\int_{-1}^{1} K(u)\, du = 1, \qquad \int_{-1}^{1} uK(u)\, du = 0, \qquad \mu_2(K) := \int_{-1}^{1} u^2 K(u)\, du < \infty. \tag{50}$$

We write

$$K_\tau(u) := \tau^{-1} K(u/\tau). \tag{51}$$

In the implementation we use the truncated normalized Gaussian kernel

$$K_{\mathrm{TG}}(u) = \frac{\varphi(u)\mathbf{1}\{|u| \leq 1\}}{\int_{-1}^{1}\varphi(v)\, dv}, \qquad \varphi(u) = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right). \tag{52}$$

This kernel is nonnegative, symmetric, integrates to one, has zero first moment, and has finite second moment.

The following proposition makes the kernel approximation explicit. The interior condition $\pi_h(X) \in [c_- + \kappa, c_+ - \kappa]$ ensures that the support of $K_\tau(\pi_h(X) - \cdot)$ remains inside the treatment interval when $\tau < \kappa$. This avoids boundary bias and boundary-tail terms. Standard boundary kernels or boundary corrections could be used instead, but we keep the compact statement below for clarity.

Proposition 6 (Kernel approximation bias)**.** *Assume there exists $\kappa > 0$ such that*

$$\pi_h(X) \in [c_- + \kappa, c_+ - \kappa] \qquad a.s. \tag{53}$$

Assume moreover that, for almost every x, the map $t \mapsto G_h(t, x)$ is twice continuously differentiable on $[\pi_h(x) - \kappa, \pi_h(x) + \kappa]$, and that there exists an integrable envelope $M_h(X)$ such that

$$\sup_{|u| \le \kappa} \left|\partial_t^2 G_h(\pi_h(X) + u, X)\right| \le M_h(X) \qquad a.s. \tag{54}$$

If $\tau < \kappa$ and K is supported on $[-1, 1]$, then

$$|A_\tau(h) - \mathbb{E}[G_h(\pi_h(X), X)]| \le \frac{\mu_2(K)}{2}\, \tau^2\, \mathbb{E}[M_h(X)]. \tag{55}$$

Consequently,

$$\mathbb{E}[G_h(\pi_h(X), X)] \le A_\tau(h) + C_h \tau^2, \qquad C_h := \frac{\mu_2(K)}{2} \mathbb{E}[M_h(X)]. \tag{56}$$

Proof. By the definition of $A_\tau(h)$ and overlap,

$$A_\tau(h) = \mathbb{E}_X \left[\int_{c_-}^{c_+} K_\tau(\pi_h(X) - t) G_h(t, X)\, dt \right]. \tag{57}$$

The support condition on K and the requirement $\tau < \kappa$ imply that, for almost every X, the integrand is nonzero only when $t \in [\pi_h(X) - \tau, \pi_h(X) + \tau] \subset [c_-, c_+]$. Fix such an x, write $\pi = \pi_h(x)$ and $g_x(t) = G_h(t, x)$, and change variables $u = (\pi - t)/\tau$. Then

$$\int K_\tau(\pi - t) g_x(t)\, dt = \int K(u) g_x(\pi - \tau u)\, du. \tag{58}$$

For each u in the support of K, Taylor's theorem gives

$$g_x(\pi - \tau u) = g_x(\pi) - \tau u g_x'(\pi) + \frac{\tau^2 u^2}{2} g_x''(\pi - \xi_{x,u} \tau u) \tag{59}$$

for some $\xi_{x,u} \in (0, 1)$. Since K is symmetric and integrates to one, $\int u K(u)\, du = 0$, and therefore the first-order term vanishes after integration. Hence

$$\left| \int K(u) g_x(\pi - \tau u)\, du - g_x(\pi) \right| \le \frac{\tau^2}{2} \mu_2(K) \sup_{|v| \le \tau} |g_x''(\pi + v)|. \tag{60}$$

Taking expectations and using the envelope condition yields (55). □

The next proposition controls the second term in the regret decomposition, the surface error at the oracle treatment $a^\star(X)$. Because $a^\star(X)$ is not available to the learner, we use the global average surface loss $B(h)$ as a fallback. The resulting bound is intentionally coarse: it is a stabilization term rather than a sharp localization result.

Proposition 7 (Global average surface bound)**.** *Assume that, for every x, the map $t \mapsto G_h(t, x)$ is L-Lipschitz on $[c_-, c_+]$. Then*

$$\mathbb{E}[G_h(a^\star(X), X)] \le B(h) + \frac{L}{2}(c_+ - c_-). \tag{61}$$

Proof. Fix x and write $t^\star = a^\star(x)$. By the Lipschitz condition,

$$G_h(t^\star, x) \le G_h(t, x) + L|t - t^\star|, \qquad t \in [c_-, c_+]. \tag{62}$$

Equivalently,

$$G_h(t, x) \ge G_h(t^\star, x) - L|t - t^\star|. \tag{63}$$

Integrating over $[c_-, c_+]$ and dividing by the interval length gives

$$\frac{1}{c_+ - c_-} \int_{c_-}^{c_+} G_h(t, x)\, dt \ge G_h(t^\star, x) - \frac{L}{c_+ - c_-} \int_{c_-}^{c_+} |t - t^\star|\, dt. \tag{64}$$

For any $t^\star \in [c_-, c_+]$,

$$\frac{1}{c_+ - c_-} \int_{c_-}^{c_+} |t - t^\star|\, dt \le \frac{c_+ - c_-}{2}. \tag{65}$$

Rearranging and taking expectation over X proves the claim. □

D.3.1 Boundary-adaptive kernel weighting

The interiority condition in Proposition 6 avoids truncating the kernel at the treatment boundaries. A support-normalized extension allows centers at the boundary. For $\pi \in [c_-, c_+]$, define the kernel mass within the feasible interval by

$$C_\tau(\pi) := \int_{c_-}^{c_+} \frac{1}{\tau} K\left(\frac{\pi - t}{\tau}\right) dt. \tag{66}$$

For centers satisfying $C_\tau(\pi_h(x)) > 0$, the boundary-adaptive weights are

$$\omega_h^{\mathrm{bd}}(a, x) := \frac{1 + \lambda K((\pi_h(x) - a)/\tau) \,/ C_\tau(\pi_h(x))}{(c_+ - c_-) p(a \mid x)}. \tag{67}$$

When the kernel support lies inside the feasible treatment interval, $C_\tau(\pi_h(x)) = 1$, and these weights coincide with the original scheme. Near a boundary, normalization compensates for kernel mass outside the interval. The normalized kernel integrates to one on $[c_-, c_+]$, reproducing constant surface errors exactly and removing the zeroth-order truncation bias.

If, for almost every x, the map $t \mapsto G_h(t, x)$ is $L_{G,h}$-Lipschitz on $[c_-, c_+]$, then, writing $\pi = \pi_h(x)$,

$$\begin{aligned} &\left| \frac{\int_{c_-}^{c_+} K_\tau(\pi - t) G_h(t, x)\, dt}{C_\tau(\pi)} - G_h(\pi, x) \right| \\ &\qquad \le \frac{L_{G,h}}{C_\tau(\pi)} \int_{c_-}^{c_+} K_\tau(\pi - t)|t - \pi|\, dt \le L_{G,h}\tau. \end{aligned} \tag{68}$$

The last inequality uses $|t - \pi| \le \tau$ wherever the kernel is nonzero. Averaging over X therefore gives an $O(\tau)$ approximation remainder. The normalized kernel is generally asymmetric near a boundary, so the first-order term need not vanish, in contrast to the $O(\tau^2)$ interior approximation under the stronger smoothness conditions of Proposition 6. The weights remain nonnegative, retain the global component, and can be incorporated into the existing solver-specific weighting implementations.

D.4 Proof of the weighted surface surrogate

We now combine the regret decomposition with the local kernel approximation and the global average surface bound.

Proof of Proposition 2. Lemma 1 gives

$$\mathrm{Reg}(h) \le \sqrt{\mathbb{E}[G_h(\pi_h(X), X)]} + \sqrt{\mathbb{E}[G_h(a^\star(X), X)]}. \tag{69}$$

Applying Propositions 6 and 7 to the two terms yields

$$\mathrm{Reg}(h) \le \sqrt{A_\tau(h) + C_h\tau^2} + \sqrt{B(h) + \frac{L}{2}(c_+ - c_-)}. \tag{70}$$

Using $\sqrt{u} + \sqrt{v} \le \sqrt{2(u + v)}$ for $u, v \ge 0$,

$$\mathrm{Reg}(h) \le \sqrt{2\left(A_\tau(h) + B(h) + C_h\tau^2 + \frac{L}{2}(c_+ - c_-)\right)}. \tag{71}$$

Since $c_\gamma = \max\{1, \gamma^{-1}\}$,

$$A_\tau(h) + B(h) \le c_\gamma\{\gamma A_\tau(h) + B(h)\}. \tag{72}$$

Thus

$$\mathrm{Reg}(h) \le \sqrt{2c_\gamma\{\gamma A_\tau(h) + B(h)\} + 2\left(C_h\tau^2 + \frac{L}{2}(c_+ - c_-)\right)}. \tag{73}$$

Finally applying $\sqrt{u + v} \le \sqrt{u} + \sqrt{v}$ gives

$$\mathrm{Reg}(h) \le \sqrt{2c_\gamma}\sqrt{\gamma A_\tau(h) + B(h)} + \sqrt{2\left(C_h\tau^2 + \frac{L}{2}(c_+ - c_-)\right)}, \tag{74}$$

which is the desired bound. □

D.5 Bridge-to-surface control

The preceding arguments control regret by a weighted surface loss involving $m_h - m_0$. This final step shows that, under a weighted ill-posedness condition, the same weighted surface loss is controlled by the corresponding weighted proximal bridge moment loss.

Proof of Proposition 3. For every (a, x), define the surface error induced by the bridge error as

$$\bar{\Delta}_h(a, x) := m_h(a, x) - m_0(a, x) = \int \Delta_h(a, w, x)\, dP_{W|X=x}(w). \tag{75}$$

Thus

$$\mathcal{L}_{\mathrm{surf},\omega}(h) = \mathbb{E}\left[\omega(A, X)\bar{\Delta}_h(A, X)^2\right]. \tag{76}$$

By Jensen's inequality, for every (a, x),

$$\bar{\Delta}_h(a, x)^2 \le \int \Delta_h(a, w, x)^2\, dP_{W|X=x}(w). \tag{77}$$

It follows that

$$\mathcal{L}_{\mathrm{surf},\omega}(h) \le \mathbb{E}_X\left[\int \omega(a, X)\left\{\int \Delta_h(a, w, X)^2\, dP_{W|X}(w)\right\} dP_{A|X}(a \mid X)\right] \tag{78}$$

$$= \int \omega(a, x)\Delta_h(a, w, x)^2\, d\mu(w, a, x) = \|\Delta_h\|^2_{L_2(\omega\, d\mu)}. \tag{79}$$

By the bounded density-ratio assumption $d\mu/d\nu \le C_\rho$,

$$\|\Delta_h\|^2_{L_2(\omega\, d\mu)} \le C_\rho \|\Delta_h\|^2_{L_2(\omega\, d\nu)}. \tag{80}$$

Since h_0 satisfies the outcome bridge restriction,

$$\mathbb{E}[Y - h(A, W, X) \mid A, Z, X] = -\mathbb{E}[h(A, W, X) - h_0(A, W, X) \mid A, Z, X] = -T\Delta_h(A, Z, X), \tag{81}$$

and hence

$$\mathcal{L}_{\mathrm{br},\omega}(h) = \|T\Delta_h\|^2_{L_2(\omega\, dP_{A,Z,X})}. \tag{82}$$

The definition of χ_ω gives

$$\|\Delta_h\|_{L_2(\omega\, d\nu)} \le \chi_\omega \|T\Delta_h\|_{L_2(\omega\, dP_{A,Z,X})}. \tag{83}$$

Combining the last four displays proves

$$\mathcal{L}_{\mathrm{surf},\omega}(h) \le C_\rho \chi^2_\omega \mathcal{L}_{\mathrm{br},\omega}(h). \tag{84}$$

□

Proof of Theorem 1. By definition of ω_h,

$$\mathcal{L}^{\mathrm{surf}}_{\tau,\lambda}(h) = \mathcal{L}_{\mathrm{surf},\omega_h}(h). \tag{85}$$

Applying Proposition 3 with $\omega = \omega_h$ yields

$$\mathcal{L}^{\mathrm{surf}}_{\tau,\lambda}(h) \le C_\rho \chi^2_{\omega_h} \mathcal{L}_{\mathrm{br},\omega_h}(h). \tag{86}$$

Substituting this inequality into Proposition 2 proves the theorem. □

D.6 The weighted ill-posedness condition

Proposition 3 assumes the finite weighted ill-posedness constant

$$\chi_\omega := \sup_{g \in \mathcal{H}:\, \|T\Delta_g\|_{L_2(\omega\, dP_{A,Z,X})} > 0} \frac{\|\Delta_g\|_{L_2(\omega\, d\nu)}}{\|T\Delta_g\|_{L_2(\omega\, dP_{A,Z,X})}} < \infty. \tag{87}$$

This condition is the direct stability assumption needed to control the weighted surface target by the weighted bridge loss. It is also implied by the usual unweighted restricted ill-posedness condition whenever the weights are bounded away from zero and infinity.

Proposition 8 (Bounded weights imply weighted stability)**.** *Assume that the global restricted ill-posedness constant*

$$\chi_{\mathrm{glob}} := \sup_{g\in\mathcal{H}:\, \|T\Delta_g\|_{L_2(dP_{A,Z,X})}>0} \frac{\|\Delta_g\|_{L_2(d\nu)}}{\|T\Delta_g\|_{L_2(dP_{A,Z,X})}} \tag{88}$$

is finite. Suppose further that, for some constants $0 < \omega_{\min} \le \omega_{\max} < \infty$,

$$\omega_{\min} \le \omega(a,x) \le \omega_{\max} \tag{89}$$

almost surely under the relevant laws. Then

$$\chi_\omega \le \sqrt{\frac{\omega_{\max}}{\omega_{\min}}}\, \chi_{\mathrm{glob}}. \tag{90}$$

Equivalently,

$$\chi_\omega^2 \le \frac{\omega_{\max}}{\omega_{\min}}\, \chi_{\mathrm{glob}}^2. \tag{91}$$

Proof. For every $g \in \mathcal{H}$,

$$\|\Delta_g\|_{L_2(\omega\, d\nu)} \le \sqrt{\omega_{\max}}\, \|\Delta_g\|_{L_2(d\nu)}. \tag{92}$$

Similarly,

$$\|T\Delta_g\|_{L_2(\omega\, dP_{A,Z,X})} \ge \sqrt{\omega_{\min}}\, \|T\Delta_g\|_{L_2(dP_{A,Z,X})}. \tag{93}$$

Therefore, whenever the denominator is nonzero,

$$\frac{\|\Delta_g\|_{L_2(\omega\, d\nu)}}{\|T\Delta_g\|_{L_2(\omega\, dP_{A,Z,X})}} \le \sqrt{\frac{\omega_{\max}}{\omega_{\min}}}\, \frac{\|\Delta_g\|_{L_2(d\nu)}}{\|T\Delta_g\|_{L_2(dP_{A,Z,X})}}. \tag{94}$$

Taking the supremum over $g \in \mathcal{H}$ gives the result. □

Proposition 8 shows that the weighted stability condition is not a qualitatively new type of assumption when bounded or clipped weights are used. It follows from the standard global restricted ill-posedness condition by equivalence of weighted and unweighted L_2 norms.

Finite-dimensional characterization. Let $S_{\mathrm{br}} = (A, W, X)$ and $V_{\mathrm{br}} = (A, Z, X)$. Consider a correctly specified fixed-representation linear bridge class

$$h_\beta(S_{\mathrm{br}}) = \beta^\top \phi(S_{\mathrm{br}}), \qquad \beta \in \mathbb{R}^d, \qquad h_0 = h_{\beta_0}. \tag{95}$$

Then $\Delta_h(S_{\mathrm{br}}) = b^\top \phi(S_{\mathrm{br}})$, where $b = \beta - \beta_0$. Define

$$\bar{\phi}(V_{\mathrm{br}}) := \mathbb{E}[\phi(S_{\mathrm{br}}) \mid V_{\mathrm{br}}], \tag{96}$$

and the weighted matrices

$$\begin{aligned} G_\omega &:= \mathbb{E}\big[\omega(A,X)\phi(S_{\mathrm{br}})\phi(S_{\mathrm{br}})^\top\big], \\ Q_\omega &:= \mathbb{E}\big[\omega(A,X)\bar{\phi}(V_{\mathrm{br}})\bar{\phi}(V_{\mathrm{br}})^\top\big]. \end{aligned} \tag{97}$$

Assume these weighted second moments are finite. On the effective feature span, after discarding directions with $b^\top G_\omega b = 0$, G_ω is positive definite. If Q_ω is also positive definite on this span, the restricted constant satisfies

$$\chi_\omega^2 = \sup_{b:\, b^\top Q_\omega b>0} \frac{b^\top G_\omega b}{b^\top Q_\omega b} = \Big[\lambda_{\min}\Big(G_\omega^{-1/2} Q_\omega G_\omega^{-1/2}\Big)\Big]^{-1}, \tag{98}$$

where the inverses and eigenvalue are evaluated on that span. Thus, the constant is finite when conditional expectation does not annihilate any nonzero bridge-feature direction in the considered bridge-error class. This is a restricted finite-dimensional stability condition, not a claim that the unrestricted proximal inverse problem is well posed. The construction includes fixed-representation second-stage linear bridge estimation, such as DFPV conditional on its learned features.

Scalar Gaussian example. Suppress A and X, and let W and Z be standardized jointly Gaussian variables with correlation $\rho \neq 0$. For the feature $\phi = W$, we have $\mathbb{E}[W \mid Z] = \rho Z$. Under a constant positive weight ω,

$$G_\omega = \omega, \qquad Q_\omega = \omega\rho^2, \qquad \chi_\omega = \frac{1}{|\rho|} < \infty. \tag{99}$$

The constant increases as proxy informativeness decreases, approaching infinity as $|\rho| \to 0$.

D.7 Proof of Population Target Preservation

Let

$$S := (A, W, X), \qquad V := (A, Z, X), \qquad r_h := Y - h(S). \tag{100}$$

Then

$$\mathcal{L}_{\mathrm{br},\omega}(h) = \mathbb{E}\big[\omega(A, X)\{\mathbb{E}[r_h \mid V]\}^2\big]. \tag{101}$$

Since h_0 satisfies the outcome bridge restriction,

$$\mathbb{E}[r_{h_0} \mid V] = 0 \quad \text{a.s.} \tag{102}$$

and hence $\mathcal{L}_{\mathrm{br},\omega}(h_0) = 0$. Because $\mathcal{L}_{\mathrm{br},\omega}(h) \geq 0$ for all $h \in \mathcal{H}$, this shows that h_0 is a population minimizer.

Now fix any $h \in \mathcal{H}$. If $\mathcal{L}_{\mathrm{br},\omega}(h) = 0$, then the nonnegative random variable

$$\omega(A, X)\{\mathbb{E}[r_h \mid V]\}^2 \tag{103}$$

has expectation zero, and therefore equals zero almost surely. Since $\omega(A, X) > 0$ almost surely, it follows that

$$\mathbb{E}[r_h \mid V] = 0 \quad \text{a.s.} \tag{104}$$

Equivalently,

$$\mathbb{E}\{Y - h(A, W, X) \mid A, Z, X\} = 0 \quad \text{a.s.} \tag{105}$$

The converse implication is immediate: if the conditional moment restriction holds, then $\mathcal{L}_{\mathrm{br},\omega}(h) = 0$. Therefore,

$$\mathcal{L}_{\mathrm{br},\omega}(h) = 0 \quad \Longleftrightarrow \quad \mathbb{E}\{Y - h(A, W, X) \mid A, Z, X\} = 0 \quad \text{a.s.} \tag{106}$$

The same equivalence holds for the unweighted bridge risk, corresponding to $\omega \equiv 1$. Hence the weighted and unweighted population bridge risks have the same zero-risk solution.

D.8 Fixed-weight statistical transfer to surface risk

Proposition 3 is a population inequality, but it also transfers statistical guarantees for a weighted proximal bridge estimator to its induced causal response surface. The next result makes this implication explicit without imposing a particular bridge-learning algorithm.

Corollary 1 (Oracle transfer from bridge risk to surface risk)**.** *Let $\omega : \mathcal{A} \times \mathcal{X} \to (0, \infty)$ be a fixed measurable weight satisfying the conditions of Proposition 3, with $C_\rho < \infty$ and $\chi_\omega < \infty$. Suppose an estimator $\widehat{h}_n \in \mathcal{H}$ satisfies, with probability at least $1 - \delta$,*

$$\mathcal{L}_{\mathrm{br},\omega}(\widehat{h}_n) \leq \inf_{h \in \mathcal{H}} \mathcal{L}_{\mathrm{br},\omega}(h) + r_n(\delta), \tag{107}$$

where $r_n(\delta) \geq 0$. Then, on the same event,

$$\mathcal{L}_{\mathrm{surf},\omega}(\widehat{h}_n) \leq C_\rho \chi_\omega^2 \left\{ \inf_{h \in \mathcal{H}} \mathcal{L}_{\mathrm{br},\omega}(h) + r_n(\delta) \right\}. \tag{108}$$

In particular, if $h_0 \in \mathcal{H}$, the approximation term vanishes. If $\mathcal{L}_{\mathrm{br},\omega}(\widehat{h}_n) = o_p(1)$ and $C_\rho \chi_\omega^2 < \infty$ is fixed, then $\mathcal{L}_{\mathrm{surf},\omega}(\widehat{h}_n) = o_p(1)$.

Proof. On the event in (107), apply Proposition 3 to $\widehat{h}_n$:

$$\mathcal{L}_{\mathrm{surf},\omega}(\widehat{h}_n) \leq C_\rho \chi_\omega^2 \mathcal{L}_{\mathrm{br},\omega}(\widehat{h}_n) \leq C_\rho \chi_\omega^2 \left\{ \inf_{h \in \mathcal{H}} \mathcal{L}_{\mathrm{br},\omega}(h) + r_n(\delta) \right\}. \tag{109}$$

When $h_0 \in \mathcal{H}$, the outcome bridge restriction implies $\mathcal{L}_{\mathrm{br},\omega}(h_0) = 0$. The consistency statement follows from the same inequality. □

The approximation term in (108) need not vanish under bridge-class misspecification. Thus the corollary transfers a solver-specific bridge-risk guarantee; it does not assert that every proximal solver is consistent under arbitrary misspecification. Nor does it require recovering the bridge itself in an unweighted norm.

For completeness, when the fixed weight is bounded as $0 < \omega_{\min} \le \omega(A, X) \le \omega_{\max} < \infty$, the nonnegativity of the squared conditional moment gives, for every $h \in \mathcal{H}$,

$$\omega_{\min}\mathcal{L}_{\mathrm{br},1}(h) \le \mathcal{L}_{\mathrm{br},\omega}(h) \le \omega_{\max}\mathcal{L}_{\mathrm{br},1}(h), \tag{110}$$

where $\mathcal{L}_{\mathrm{br},1}$ denotes the unweighted bridge risk. Together with Proposition 8, this shows that bounded positive reweighting preserves the population zero-risk target and finite restricted stability, up to weight-dependent constants. For fixed bounded weights, standard empirical-process or U/V-process arguments can also retain the sample-size order of a base learner's bound under the corresponding envelope and complexity conditions. Such statistical bounds remain solver-specific: the actual MMR and two-stage empirical objectives require their own concentration and calibration arguments, rather than being automatically identified with $\mathcal{L}_{\mathrm{br},\omega}$.

D.9 A finite-round oracle inequality for adaptive weights

The preceding corollary applies to a fixed weight. To account for weights selected from the data, we consider a deterministic class containing the weights that the learning procedure may use. This leads to a conditional finite-round guarantee for the solver-specific objective, without assuming that the chosen weights are independent of the training sample.

For a given proximal solver, let $\mathcal{R}_\omega(h)$ denote the population counterpart of the empirical fitting criterion $\widehat{\mathcal{R}}_{n,\omega}(h)$ introduced in Section 5. In general, $\mathcal{R}_\omega$ need not coincide with the ideal conditional-moment loss $\mathcal{L}_{\mathrm{br},\omega}$: PMMR and NMMR use kernelized moment criteria, whereas KPV and DFPV use two-stage fitting objectives. Under suitable conditions, these criteria target the same population bridge equation, but quantitative control of $\mathcal{L}_{\mathrm{br},\omega}$ requires additional solver-specific calibration conditions, stated below.

Proposition 9 (Finite-round adaptive oracle inequality)**.** *Fix a finite number of rounds S and let Ω_S be a deterministic class of positive measurable weights containing all weights $\widehat{\omega}^{(s)}$ reachable in the first $S+1$ bridge fits, $s = 0, \dots, S$. Suppose that, with probability at least $1-\delta$,*

$$\sup_{\substack{h\in\mathcal{H}\\ \omega\in\Omega_S}} \left|\widehat{\mathcal{R}}_{n,\omega}(h) - \mathcal{R}_\omega(h)\right| \le \epsilon_n(\delta). \tag{111}$$

Assume also that, for each round $s = 0, \dots, S$, the fitted bridge $\widehat{h}^{(s)} \in \mathcal{H}$ obeys

$$\widehat{\mathcal{R}}_{n,\widehat{\omega}^{(s)}}(\widehat{h}^{(s)}) \le \inf_{h\in\mathcal{H}} \widehat{\mathcal{R}}_{n,\widehat{\omega}^{(s)}}(h) + \xi_s, \tag{112}$$

where $\xi_s \ge 0$. Then, simultaneously for all $s = 0, \dots, S$,

$$\mathcal{R}_{\widehat{\omega}^{(s)}}(\widehat{h}^{(s)}) \le \inf_{h\in\mathcal{H}} \mathcal{R}_{\widehat{\omega}^{(s)}}(h) + 2\epsilon_n(\delta) + \xi_s \tag{113}$$

with probability at least $1-\delta$, provided (112) holds on the same event.

Proof. Work on the event (111). Since $\widehat{\omega}^{(s)} \in \Omega_S$, the deviation bound holds at the possibly data-dependent weight $\widehat{\omega}^{(s)}$. For any $s = 0, \dots, S$,

$$\begin{aligned}
\mathcal{R}_{\widehat{\omega}^{(s)}}(\widehat{h}^{(s)}) &\le \widehat{\mathcal{R}}_{n,\widehat{\omega}^{(s)}}(\widehat{h}^{(s)}) + \epsilon_n(\delta) && (114)\\
&\le \inf_{h\in\mathcal{H}} \widehat{\mathcal{R}}_{n,\widehat{\omega}^{(s)}}(h) + \xi_s + \epsilon_n(\delta) && (115)\\
&\le \inf_{h\in\mathcal{H}} \mathcal{R}_{\widehat{\omega}^{(s)}}(h) + 2\epsilon_n(\delta) + \xi_s. && (116)
\end{aligned}$$

The uniform event applies to all weights in Ω_S, so the conclusion holds simultaneously over the finite rounds without requiring a separate independence assumption for each selected weight. □

The inequality (113) controls the solver-specific population fitting criterion. A corresponding bridge-risk and surface-risk guarantee follows if the solver criterion is calibrated to the conditional moment restriction. Specifically, suppose there exist $C_{\mathrm{cal}} < \infty$ and nonnegative terms b_ω such that, uniformly over $h \in \mathcal{H}$ and $\omega \in \Omega_S$,

$$\mathcal{L}_{\mathrm{br},\omega}(h) \le b_\omega + C_{\mathrm{cal}} \left\{ \mathcal{R}_\omega(h) - \inf_{g \in \mathcal{H}} \mathcal{R}_\omega(g) \right\}. \tag{117}$$

Assume also that the conditions of Proposition 3 hold for every $\omega \in \Omega_S$. Combining (113), (117), and Proposition 3 yields, on the same event,

$$\mathcal{L}_{\mathrm{surf},\widehat{\omega}^{(s)}}(\widehat{h}^{(s)}) \le C_\rho \chi^2_{\widehat{\omega}^{(s)}} \left\{ b_{\widehat{\omega}^{(s)}} + C_{\mathrm{cal}} \big(2\epsilon_n(\delta) + \xi_s \big) \right\}, \qquad s = 0, \dots, S. \tag{118}$$

The terms b_ω allow for bridge-class approximation and imperfect calibration of the solver objective. For the ideal choice $\mathcal{R}_\omega = \mathcal{L}_{\mathrm{br},\omega}$, one may take $C_{\mathrm{cal}} = 1$ and $b_\omega = \inf_{h \in \mathcal{H}} \mathcal{L}_{\mathrm{br},\omega}(h)$. Under correct specification this term is zero. If, in addition, $\sup_{\omega \in \Omega_S} \chi_\omega \le \bar{\chi} < \infty$, $\sup_{\omega \in \Omega_S} b_\omega = o(1)$, and $\epsilon_n(\delta) + \max_{0 \le s \le S} \xi_s = o(1)$ along a sequence of high-probability events with probability tending to one, then the weighted surface risks in (118) converge to zero uniformly over the finite rounds. Uniform boundedness $0 < \omega_{\min} \le \omega \le \omega_{\max} < \infty$ over Ω_S and Proposition 8 provide one sufficient route to a uniform bound on χ_ω.

The uniform deviation condition (111) accounts for the complexity of the weight class, including its dependence on pilot policies and propensity-density estimates. It is therefore stronger than a deviation bound for a single fixed weight. Its validity, together with (117), must be checked for the chosen solver and function classes; no universal rate or improvement over an unweighted estimator is implied by the finite-round inequality.

E Practical Implementation Template for Policy-Targeted Proximal Solvers

This appendix collects the implementation details for the proximal bridge solvers used in the experiments. The goal is to separate the common policy-targeted weighting template from the solver-specific algebra. Throughout this appendix, write

$$S := (A, W, X), \qquad V := (A, Z, X), \tag{119}$$

and, for a candidate outcome bridge h, define the residual

$$r_h := Y - h(S). \tag{120}$$

The outcome bridge equation is the conditional moment restriction

$$\mathbb{E}[r_h \mid V] = 0. \tag{121}$$

All solvers below estimate this same bridge equation. They differ only in how they approximate or penalize the conditional moment restriction.

E.1 Policy-targeted weights

At iteration s, the bridge fit is projected into a response-surface estimate using bridge pseudo-outcomes over a treatment grid $\mathcal{G} = \{a_1, \dots, a_M\} \subset \mathcal{A}$. To construct the localization weights, these pseudo-outcomes are formed by cross-fitting. For each fold I_k in a partition of $\{1, \dots, n\}$, the bridge $\widehat{h}^{(s)}_{-k}$ is trained on the observations outside I_k using the current weights and evaluated on the held-out observations, giving

$$\widetilde{H}^{(s)}_{im} = \widehat{h}^{(s)}_{-k}(a_m, W_i, X_i), \qquad i \in I_k, \quad m = 1, \dots, M. \tag{122}$$

We pool these out-of-fold pseudo-outcomes across folds and regress them on (a_m, X_i) to obtain a single response-surface estimate $\widehat{m}^{(s)}$. The observation-specific pseudo-optimal treatment used for weighting is

$$\widehat{\pi}^{(s)}_i \in \operatorname*{argmax}_{a \in \mathcal{G}} \widehat{m}^{(s)}(a, X_i), \tag{123}$$

or the corresponding maximizer over $\mathcal{A}$ when continuous optimization is feasible.

The next bridge fit uses the fixed weights

$$\widehat{\omega}_i^{(s+1)} = \frac{1 + \lambda\, K\Big((\widehat{\pi}_i^{(s)} - A_i)/\tau\Big)}{(c_+ - c_-)\widehat{p}(A_i \mid X_i)}, \tag{124}$$

where $\widehat{p}(a \mid x)$ is an estimate of the generalized propensity score, i.e., the conditional treatment density for the continuous treatment A. In our experiments, this density is estimated using a conditional normalizing-flow model. The function K is the localization kernel, τ is the bandwidth, and $\lambda \geq 0$ controls the strength of localization. The constant 1 in the numerator is the global stabilizing component.

After the final weighting iteration, the solver is refit on the full sample using the last cross-fitted weights.

E.2 Two ways in which the weights enter

The same policy-targeted weights are used for every solver, but they enter the numerical objective in two different ways. For maximum-moment solvers, the weights multiply pairwise residual products. For two-stage regression solvers, the weights are applied in the second-stage bridge regression.

Table 2 summarizes the solver-specific implementations.

Table 2: Solver-specific implementation of the policy-targeted bridge weights.

Solver	Baseline object	Policy-targeted replacement
PMMR	$r^\top K_V r$	$r^\top D_\omega^{1/2} K_V D_\omega^{1/2} r$
NMMR	$r(\theta)^\top K_V r(\theta)$	$r(\theta)^\top D_\omega^{1/2} K_V D_\omega^{1/2} r(\theta)$
KPV	second-stage kernel ridge	weighted second-stage ridge with D_ω
DFPV	second-stage feature ridge	weighted second-stage ridge with D_ω

Here K_V is the moment-kernel Gram matrix on $V = (A, Z, X)$ and

$$D_\omega := \mathrm{diag}(\widehat{\omega}_1, \ldots, \widehat{\omega}_n). \tag{125}$$

For PMMR and NMMR, the algebra is equivalent to replacing each residual by $\sqrt{\widehat{\omega}_i} r_i$. For KPV and DFPV, the actual sample weight in the second-stage weighted regression is $\widehat{\omega}_i$, not $\sqrt{\widehat{\omega}_i}$.

E.3 Why the weighted objective still targets the bridge equation

The policy-targeted objective does not change the population bridge equation when the weights are strictly positive. It changes the geometry in which violations of the bridge equation are penalized. Indeed, for any positive measurable weight $\omega(V)$,

$$\omega(V)\{\mathbb{E}[r_h \mid V]\}^2 = 0 \quad \Longleftrightarrow \quad \mathbb{E}[r_h \mid V] = 0. \tag{126}$$

Equivalently,

$$\sqrt{\omega(V)}\,\mathbb{E}[r_h \mid V] = \mathbb{E}[\sqrt{\omega(V)}\, r_h \mid V], \tag{127}$$

because $\omega(V)$ is measurable with respect to V. Thus, in a well-specified population problem, the weighted and unweighted conditional moment restrictions have the same bridge solutions. The weights matter in finite samples, under regularization, and under misspecification, where they determine which moment violations receive more emphasis.

F MMR-Based Solvers: PMMR and NMMR

PMMR and NMMR both estimate the outcome bridge by penalizing a kernel maximum moment restriction. The difference is the bridge class: PMMR uses an RKHS bridge, whereas NMMR uses a neural bridge. This section first derives the common weighted MMR criterion and then gives the PMMR and NMMR specializations.

F.1 Weighted maximum moment restriction

Let k be a positive definite kernel on $\mathcal{V}$ and let $\mathcal{H}_k$ be its RKHS. The unweighted kernel maximum moment criterion is

$$\mathcal{R}_k(h) := \|\mathbb{E}[r_h k(V, \cdot)]\|^2_{\mathcal{H}_k}. \tag{128}$$

For an independent copy (Y', S', V'), with $r'_h = Y' - h(S')$, this can be written as

$$\mathcal{R}_k(h) = \mathbb{E}\big[r_h r'_h k(V, V')\big]. \tag{129}$$

Given observations $\{(y_i, s_i, v_i)\}_{i=1}^n$, define

$$r_i(h) := y_i - h(s_i), \qquad r(h) := (r_1(h), \dots, r_n(h))^\top, \tag{130}$$

and let

$$(K_V)_{ij} := k(v_i, v_j). \tag{131}$$

The empirical V-statistic is

$$\widehat{\mathcal{R}}_{k,V,n}(h) = \frac{1}{n^2} r(h)^\top K_V r(h). \tag{132}$$

For policy-targeted learning, define

$$K_{V,\widehat{\omega}} := D^{1/2}_{\widehat{\omega}} K_V D^{1/2}_{\widehat{\omega}}, \qquad D^{1/2}_{\widehat{\omega}} := \operatorname{diag}(\sqrt{\widehat{\omega}_1}, \dots, \sqrt{\widehat{\omega}_n}). \tag{133}$$

The weighted empirical MMR criterion is

$$\widehat{\mathcal{R}}_{k,V,n,\widehat{\omega}}(h) = \frac{1}{n^2} r(h)^\top K_{V,\widehat{\omega}} r(h). \tag{134}$$

Equivalently, (134) is the ordinary MMR criterion applied to the transformed residuals

$$\widetilde{r}_i(h) := \sqrt{\widehat{\omega}_i}\{y_i - h(s_i)\}. \tag{135}$$

F.2 Policy-targeted PMMR

Let ℓ be a positive definite kernel on $\mathcal{S}$ with RKHS $\mathcal{H}_\ell$. PMMR restricts $h \in \mathcal{H}_\ell$ and uses the representer form

$$h_\alpha(\cdot) = \sum_{j=1}^n \alpha_j \ell(s_j, \cdot). \tag{136}$$

Let

$$(L_S)_{ij} := \ell(s_i, s_j), \qquad y := (y_1, \dots, y_n)^\top. \tag{137}$$

Then the residual vector is

$$r(\alpha) = y - L_S \alpha. \tag{138}$$

The unweighted PMMR objective is

$$J_{\mathrm{PMMR}}(\alpha) = (y - L_S\alpha)^\top K_V (y - L_S\alpha) + \eta\, \alpha^\top L_S \alpha, \tag{139}$$

where the common factor n^{-2} is absorbed into the regularization parameter. Its stabilized solution is

$$\widehat{\alpha} = (L_S K_V L_S + \eta L_S + \delta I_n)^{-1} L_S K_V y, \tag{140}$$

with numerical jitter $\delta > 0$.

The policy-targeted PMMR objective is obtained by replacing K_V with the weighted moment Gram matrix $K_{V,\widehat{\omega}}$:

$$J_{\mathrm{PMMR},\widehat{\omega}}(\alpha) = (y - L_S\alpha)^\top K_{V,\widehat{\omega}} (y - L_S\alpha) + \eta\, \alpha^\top L_S \alpha. \tag{141}$$

Therefore

$$\widehat{\alpha}_{\widehat{\omega}} = (L_S K_{V,\widehat{\omega}} L_S + \eta L_S + \delta I_n)^{-1} L_S K_{V,\widehat{\omega}} y. \tag{142}$$

The fitted bridge is

$$\widehat{h}_{\widehat{\omega}}(a, w, x) = \sum_{j=1}^n \widehat{\alpha}_{\widehat{\omega},j} \ell\{s_j, (a, w, x)\}. \tag{143}$$

Thus policy-targeted PMMR is implemented by the single replacement

$$K_V \quad \rightsquigarrow \quad D^{1/2}_{\widehat{\omega}} K_V D^{1/2}_{\widehat{\omega}}. \tag{144}$$

F.3 Policy-targeted NMMR

NMMR uses the same conditional moment restriction, but represents the bridge by a neural network $h_\theta : \mathcal{S} \to \mathbb{R}$. Define

$$r_i(\theta) := y_i - h_\theta(s_i), \qquad r(\theta) := (r_1(\theta), \dots, r_n(\theta))^\top. \tag{145}$$

The unweighted V-statistic NMMR loss is

$$\widehat{\mathcal{R}}_{k,V,n}(\theta) = \frac{1}{n^2} r(\theta)^\top K_V r(\theta). \tag{146}$$

The corresponding U-statistic version removes diagonal terms:

$$\widehat{\mathcal{R}}_{k,U,n}(\theta) = \frac{1}{n(n-1)} r(\theta)^\top \{K_V - \operatorname{diag}(K_V)\} r(\theta). \tag{147}$$

The policy-targeted NMMR loss replaces $r_i(\theta)$ by $\sqrt{\widehat{\omega}_i} r_i(\theta)$. The weighted V-statistic objective is

$$\widehat{\mathcal{R}}_{k,V,n,\widehat{\omega}}(\theta) = \frac{1}{n^2} r(\theta)^\top D_{\widehat{\omega}}^{1/2} K_V D_{\widehat{\omega}}^{1/2} r(\theta), \tag{148}$$

and the weighted U-statistic objective is

$$\widehat{\mathcal{R}}_{k,U,n,\widehat{\omega}}(\theta) = \frac{1}{n(n-1)} r(\theta)^\top D_{\widehat{\omega}}^{1/2} \{K_V - \operatorname{diag}(K_V)\} D_{\widehat{\omega}}^{1/2} r(\theta). \tag{149}$$

The estimator minimizes one of these losses plus the usual network regularization penalty:

$$\widehat{\theta}_{\widehat{\omega}} \in \operatorname*{argmin}_\theta \left\{ \widehat{\mathcal{R}}_{k,V,n,\widehat{\omega}}(\theta) + \operatorname{pen}(\theta) \right\}, \tag{150}$$

or analogously with (149).

For mini-batch training, the same formula is applied within each batch: compute the batch-level moment kernel on the batch values of V, form the residual vector, multiply residuals by $\sqrt{\widehat{\omega}_i}$, and evaluate the quadratic form. Full-batch training is preferable when feasible because the NMMR objective is pairwise.

G Two-Stage Solvers: KPV and DFPV

KPV and DFPV estimate the bridge through two stages. The first stage estimates a nuisance conditional object, such as a conditional mean embedding or conditional feature. The second stage solves the bridge equation by regressing Y on that estimated conditional object. In the policy-targeted implementation used here, the first stage is left unweighted and the policy weights are applied only in the second-stage bridge regression.

This choice matches the role of the weights in the theory. The weights define the norm over moment-view points $V = (A, Z, X)$ in which bridge residuals are penalized. The first stage estimates the nuisance conditional object itself, whereas the second stage is the step that directly enforces the bridge equation.

G.1 Policy-targeted KPV

KPV uses kernels to estimate the conditional mean embedding of S given V [27]. Let ℓ be a positive definite kernel on $\mathcal{S}$ with RKHS $\mathcal{H}_\ell$, and let k be a positive definite kernel on $\mathcal{V}$. The bridge equation can be written as

$$\mathbb{E}[Y \mid V = v] = \mathbb{E}[h(S) \mid V = v] = \left\langle h, \mu_{S|v} \right\rangle_{\mathcal{H}_\ell}, \tag{151}$$

where

$$\mu_{S|v} := \mathbb{E}\{\ell(S, \cdot) \mid V = v\} \tag{152}$$

is the conditional mean embedding.

Split the sample into a first-stage sample $\mathcal{I}_1$ of size m_1 and a second-stage sample $\mathcal{I}_2$ of size m_2. Define

$$(K_{V,11})_{ij} := k(v_i, v_j), \quad i, j \in \mathcal{I}_1, \qquad (K_{V,12})_{ij} := k(v_i, v_j), \quad i \in \mathcal{I}_1,\ j \in \mathcal{I}_2. \tag{153}$$

With first-stage regularization $\lambda_1 > 0$, the estimated conditional embedding at a second-stage point v_j, $j \in \mathcal{I}_2$, is

$$\widehat{\mu}_{S|v_j} = \sum_{i \in \mathcal{I}_1} \widehat{\Gamma}_{ij} \ell(s_i, \cdot), \tag{154}$$

where

$$\widehat{\Gamma} := (K_{V,11} + m_1 \lambda_1 I_{m_1})^{-1} K_{V,12}. \tag{155}$$

Let

$$(L_{S,11})_{ij} := \ell(s_i, s_j), \qquad i, j \in \mathcal{I}_1, \tag{156}$$

and define the second-stage embedding Gram matrix

$$G_\mu := \left(\left\langle \widehat{\mu}_{S|v_i}, \widehat{\mu}_{S|v_j} \right\rangle_{\mathcal{H}_\ell} \right)_{i,j \in \mathcal{I}_2} = \widehat{\Gamma}^\top L_{S,11} \widehat{\Gamma}. \tag{157}$$

Writing $y_2 = (y_j : j \in \mathcal{I}_2)$, the unweighted second-stage KPV objective is

$$J_{\mathrm{KPV}}(\beta) = (y_2 - G_\mu \beta)^\top (y_2 - G_\mu \beta) + m_2 \lambda_2 \beta^\top G_\mu \beta. \tag{158}$$

The corresponding stabilized solution is

$$\widehat{\beta} = (G_\mu + m_2 \lambda_2 I_{m_2} + \delta I_{m_2})^{-1} y_2. \tag{159}$$

The fitted bridge is

$$\widehat{h}(a, w, x) = \sum_{j \in \mathcal{I}_2} \widehat{\beta}_j \widehat{\mu}_{S|v_j}(a, w, x), \tag{160}$$

where

$$\widehat{\mu}_{S|v_j}(a, w, x) = \sum_{i \in \mathcal{I}_1} \widehat{\Gamma}_{ij} \ell\{s_i, (a, w, x)\}. \tag{161}$$

For policy-targeted KPV, the first-stage embedding estimate is unchanged. The second-stage bridge regression becomes

$$\widehat{h}_{\lambda_2, \widehat{\omega}} \in \operatorname*{argmin}_{h \in \mathcal{H}_\ell} \left\{ \frac{1}{m_2} \sum_{j \in \mathcal{I}_2} \widehat{\omega}_j \left(y_j - \left\langle h, \widehat{\mu}_{S|v_j} \right\rangle_{\mathcal{H}_\ell} \right)^2 + \lambda_2 \|h\|^2_{\mathcal{H}_\ell} \right\}. \tag{162}$$

In finite-dimensional form,

$$J_{\mathrm{KPV}, \widehat{\omega}}(\beta) = (y_2 - G_\mu \beta)^\top D_{\widehat{\omega}} (y_2 - G_\mu \beta) + m_2 \lambda_2 \beta^\top G_\mu \beta, \tag{163}$$

where now

$$D_{\widehat{\omega}} := \mathrm{diag}(\widehat{\omega}_j : j \in \mathcal{I}_2). \tag{164}$$

A symmetric stabilized implementation is obtained by defining

$$G_{\mu, \widehat{\omega}} := D^{1/2}_{\widehat{\omega}} G_\mu D^{1/2}_{\widehat{\omega}} \tag{165}$$

and solving

$$\widehat{\eta}_{\widehat{\omega}} = (G_{\mu, \widehat{\omega}} + m_2 \lambda_2 I_{m_2} + \delta I_{m_2})^{-1} D^{1/2}_{\widehat{\omega}} y_2, \tag{166}$$

then setting

$$\widehat{\beta}_{\widehat{\omega}} := D^{1/2}_{\widehat{\omega}} \widehat{\eta}_{\widehat{\omega}}. \tag{167}$$

The policy-targeted KPV bridge is

$$\widehat{h}_{\widehat{\omega}}(a, w, x) = \sum_{j \in \mathcal{I}_2} \widehat{\beta}_{\widehat{\omega}, j} \widehat{\mu}_{S|v_j}(a, w, x). \tag{168}$$

When $D_{\widehat{\omega}} = I_{m_2}$, this reduces to ordinary KPV.

G.2 Policy-targeted DFPV

DFPV replaces kernel conditional embeddings by learned finite-dimensional features. Let

$$\phi_\theta : \mathcal{S} \to \mathbb{R}^d \tag{169}$$

be a learned bridge-side feature map and write

$$h_{\beta,\theta}(s) = \beta^\top \phi_\theta(s). \tag{170}$$

The bridge equation becomes

$$\mathbb{E}(Y \mid V) = \beta^\top \mathbb{E}\{\phi_\theta(S) \mid V\}. \tag{171}$$

DFPV therefore first estimates the conditional feature

$$\mu_\theta(v) := \mathbb{E}\{\phi_\theta(S) \mid V = v\}, \tag{172}$$

and then regresses Y on the estimated conditional feature.

Let I_1 and I_2 denote the first-stage and second-stage samples, with cardinalities m_1 and m_2. The first stage solves

$$\widehat{g} \in \operatorname*{argmin}_{g \in \mathcal{G}} \left\{ \frac{1}{m_1} \sum_{i \in I_1} \|\phi_\theta(s_i) - g(v_i)\|_2^2 + \lambda_1 \Omega_1(g) \right\}. \tag{173}$$

For $j \in I_2$, write

$$\widehat{\mu}_j := \widehat{g}(v_j). \tag{174}$$

Let $\widehat{M} \in \mathbb{R}^{m_2 \times d}$ be the matrix whose jth row is $\widehat{\mu}_j^\top$, and let $y_2 = (y_j : j \in I_2)$.

The unweighted second-stage ridge problem is

$$\widehat{\beta} \in \operatorname*{argmin}_{\beta \in \mathbb{R}^d} \left\{ \frac{1}{m_2} \sum_{j \in I_2} (y_j - \beta^\top \widehat{\mu}_j)^2 + \lambda_2 \|\beta\|_2^2 \right\}, \tag{175}$$

with stabilized solution

$$\widehat{\beta} = \left(\widehat{M}^\top \widehat{M} + m_2 \lambda_2 I_d + \delta I_d \right)^{-1} \widehat{M}^\top y_2. \tag{176}$$

The fitted bridge is

$$\widehat{h}(a, w, x) = \widehat{\beta}^\top \phi_\theta(a, w, x). \tag{177}$$

For policy-targeted DFPV, the first-stage feature regression is left unweighted and the second-stage ridge problem becomes

$$\widehat{\beta}_{\widehat{\omega}} \in \operatorname*{argmin}_{\beta \in \mathbb{R}^d} \left\{ \frac{1}{m_2} \sum_{j \in I_2} \widehat{\omega}_j (y_j - \beta^\top \widehat{\mu}_j)^2 + \lambda_2 \|\beta\|_2^2 \right\}. \tag{178}$$

Equivalently, with

$$D_{\widehat{\omega}} := \operatorname{diag}(\widehat{\omega}_j : j \in I_2), \tag{179}$$

we minimize

$$J_{\mathrm{DFPV},\widehat{\omega}}(\beta) = (y_2 - \widehat{M}\beta)^\top D_{\widehat{\omega}} (y_2 - \widehat{M}\beta) + m_2 \lambda_2 \|\beta\|_2^2. \tag{180}$$

The stabilized solution is

$$\widehat{\beta}_{\widehat{\omega}} = \left(\widehat{M}^\top D_{\widehat{\omega}} \widehat{M} + m_2 \lambda_2 I_d + \delta I_d \right)^{-1} \widehat{M}^\top D_{\widehat{\omega}} y_2. \tag{181}$$

The policy-targeted DFPV bridge is

$$\widehat{h}_{\widehat{\omega}}(a, w, x) = \widehat{\beta}_{\widehat{\omega}}^\top \phi_\theta(a, w, x). \tag{182}$$

When $D_{\widehat{\omega}} = I_{m_2}$, this reduces to ordinary DFPV.

H Implementation Details and Pitfalls

This section records practical details shared by the solver implementations.

Cross-fitting of bridge pseudo-outcomes. Within each policy-targeted iteration, cross-fitting is used to construct the bridge pseudo-outcomes that train the projected response surface. For each fold I_k in a partition of $\{1, \ldots, n\}$, the bridge $\widehat{h}_{-k}^{(s)}$ is trained on the observations outside I_k using the current weights and evaluated on the held-out observations over the treatment grid. The resulting out-of-fold pseudo-outcomes $\widetilde{H}_{im}^{(s)} = \widehat{h}_{-k}^{(s)}(a_m, W_i, X_i)$, for $i \in I_k$, are pooled across folds and used to train a single response-surface estimator $\widehat{m}^{(s)}$. The observation-specific pseudo-optimal treatments $\widehat{\pi}_i^{(s)}$ obtained from this surface are then used to form the next weights. This prevents each observation's bridge pseudo-outcome from being computed by a bridge fit that was trained on that observation.

Projection from bridge to response surface. After fitting a bridge, the induced causal response surface is

$$m_h(a, x) = \mathbb{E}\{h(a, W, x) \mid X = x\}. \tag{183}$$

In practice, this expectation is estimated by evaluating the fitted bridge on a treatment grid and then regressing the bridge pseudo-outcomes on (a, X). During weight construction, these pseudo-outcomes are the out-of-fold values described above. The pseudo-optimal treatment is obtained by maximizing the resulting surface estimate over the treatment grid.

Computational resources All experiments were run on CPU. The implementation supports parallel execution across independent random seeds, models, and experimental configurations, which makes the evaluation pipeline naturally scalable on multi-core machines. The computational overhead introduced by the proposed decision-aware weighting strategy is linear in the number of weighting rounds: if the base proximal learner has training cost T_{base}, then running n_{rounds} weighting rounds requires approximately $\mathcal{O}(n_{rounds}\, T_{\text{base}})$ time, up to lower-order costs for computing and normalizing the weights.

I Dataset Details

This appendix describes the synthetic and semi-synthetic benchmarks used in our experiments. Both datasets are designed to evaluate proximal learning for individualized treatment selection with continuous treatments, hidden confounding, and proxy variables. In each case, the observed data consist of

$$O = (X, Z, W, A, Y), \tag{184}$$

where X denotes observed covariates, Z is a treatment-inducing proxy, W is an outcome-inducing proxy, $A \in \mathcal{A}$ is a continuous treatment, and Y is the outcome. The latent confounders are unobserved by the learner but affect both treatment assignment and the outcome.

The key property of both benchmarks is that the conditional causal response

$$m_0(a, x) = \mathbb{E}\{Y(a) \mid X = x\} \tag{185}$$

is non-monotone in a, with an interior and covariate-dependent optimal treatment. This makes the datasets suitable for evaluating decision-aware bridge learning: accurate estimation near the individualized optimum is more important for policy regret than uniform accuracy over clearly suboptimal treatment regions. The following subsections describe the fully synthetic DGP, the TCGA semi-synthetic DGP, and the Monte Carlo procedure used to compute ground-truth causal curves and policy values.

I.1 Synthetic proximal continuous-treatment benchmark

We use a synthetic benchmark adapted from the proximal continuous-treatment graph of Wu et al. [42]. The original Wu et al. simulator is designed for proximal causal effect estimation with continuous treatments; we retain its proxy-confounding structure, but modify the outcome surface so that the conditional causal response has an interior, covariate-dependent maximizer.

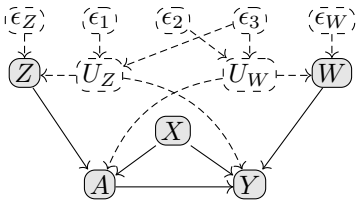


Figure 7: Synthetic proximal graph. Observed variables are shown in gray with solid nodes, while unobserved variables are shown in white with dashed nodes. The treatment-inducing proxy is Z, the outcome-inducing proxy is W, X is observed, and U_Z, U_W are latent confounding components.

The corresponding causal graph is shown in Figure 7.

Let d_X, d_Z, d_W denote the dimensions of X, Z, W, and define

$$\theta_d = (1, 1/2^2, \ldots, 1/d^2)^\top. \tag{186}$$

We sample

$$X \sim \mathcal{N}(0, \Sigma_X), \qquad (\Sigma_X)_{jk} = \begin{cases} 1, & j = k, \\ \rho_X, & |j-k| = 1, \\ 0, & \text{otherwise}, \end{cases} \tag{187}$$

and independent noises

$$\epsilon_1, \epsilon_2, \epsilon_3 \sim \mathcal{N}(0,1), \qquad V_Z \sim \mathrm{Unif}([-1,1]^{d_Z}), \qquad V_W \sim \mathrm{Unif}([-1,1]^{d_W}). \tag{188}$$

The latent variables and proxies are

$$U_Z = \epsilon_1 + \epsilon_3, \qquad U_W = \epsilon_2 + \epsilon_3, \tag{189}$$

$$Z = V_Z + c_P U_Z \mathbf{1}_{d_Z}, \qquad W = V_W + c_P U_W \mathbf{1}_{d_W}. \tag{190}$$

The treatment assignment mean is

$$\mu_A(X, Z) = \Lambda\big(3X^\top \theta_{d_X} + 3Z^\top \theta_{d_Z}\big), \qquad \Lambda(t) = 0.1 + 0.8\frac{\exp(t)}{1+\exp(t)}. \tag{191}$$

The factual treatment is

$$A = \Pi_{\mathcal{A}}\{\mu_A(X, Z) + \sigma_A U_W\}, \tag{192}$$

where $\Pi_{\mathcal{A}}$ denotes clipping to the compact treatment support $\mathcal{A} = [a_{\min}, a_{\max}]$. In our experiments, $\mathcal{A}$ is estimated once as the empirical $(q_{\min}, q_{\max})$-quantiles of the unclipped assignment distribution using n_{pilot} pilot samples.

The outcome is

$$Y = g(A, X_0) + \beta_{XW}\{X^\top \theta_{d_X} + W^\top \theta_{d_W}\} + \psi(A, X, U_Z) + \epsilon_Y, \qquad \epsilon_Y \sim \mathcal{N}(0, \sigma_Y^2). \tag{193}$$

The structural response is

$$g(a, x_0) = \alpha(x_0) \exp\left[-\frac{1}{2} r(a, x_0)^2\right] \big\{1 + \kappa(x_0) \tanh\big(1.5\, r(a, x_0)\big)\big\} - \lambda_{\text{tail}} r(a, x_0)^2, \tag{194}$$

where

$$r(a, x_0) = \frac{a - c(x_0)}{s(x_0)}, \qquad c(x_0) = \frac{a_{\min} + a_{\max}}{2} + \delta_c \tanh(\eta_c x_0), \tag{195}$$

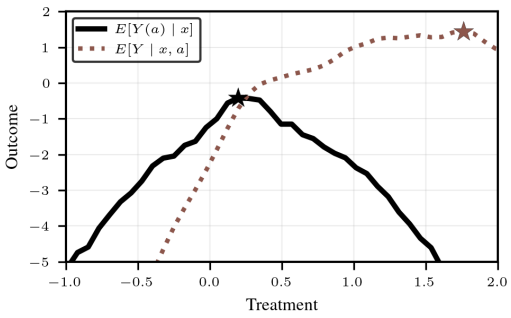
$$s(x_0) = \delta_s (a_{\max} - a_{\min})\{1 + 0.10 \tanh(0.8 x_0)\}, \tag{196}$$

$$\alpha(x_0) = 1 + \delta_\alpha \tanh(x_0), \qquad \kappa(x_0) = \delta_\kappa \tanh(x_0). \tag{197}$$

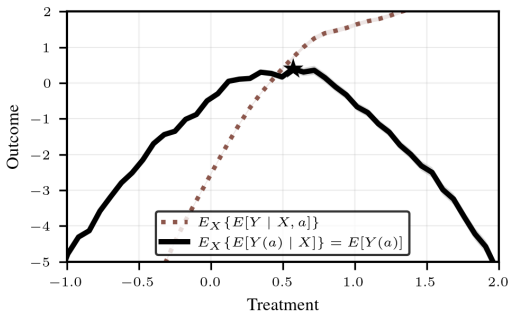
The latent outcome-confounding term is

$$\psi(a, x, U_Z) = \left[\gamma \left\{1 + \beta_A \left(\frac{a - \frac{a_{\min}+a_{\max}}{2}}{\delta_\psi (a_{\max} - a_{\min})}\right)^2\right\} + \gamma_0\right] U_Z. \tag{198}$$

Thus the latent contribution to the outcome is treatment-dependent, so failing to account for hidden confounding can distort the shape of the estimated dose-response curve.



(a) Individual-level curves for one randomly selected test point.



(b) Population-level curves.

Figure 8: Reference causal and associational curves in the synthetic benchmark. Left: for a fixed covariate value x, the solid curve shows the interventional response $\mathbb{E}\{Y(a) \mid X = x\}$, while the dotted curve shows the associational response $\mathbb{E}\{Y \mid A = a, X = x\}$. Right: the corresponding population quantities $\mathbb{E}\{Y(a)\}$ and $\mathbb{E}\{Y \mid A = a\}$ are obtained by averaging over X. Stars mark the maximizers of each curve. The separation between causal and associational curves illustrates the effect of hidden confounding; in particular, optimizing the associational curve can lead to a different treatment recommendation than optimizing the causal curve.

Under an intervention $A = a$,

$$Y(a) = g(a, X_0) + \beta_{XW}\{X^\top \theta_{d_X} + W^\top \theta_{d_W}\} + \psi(a, X, U_Z) + \epsilon_Y. \tag{199}$$

Since X is independent of (W, U_Z, ϵ_Y), and these variables are mean-zero, the conditional causal response is

$$m_0(a, x) = \mathbb{E}\{Y(a) \mid X = x\} = g(a, x_0) + \beta_{XW} x^\top \theta_{d_X}. \tag{200}$$

Consequently, the individualized optimal treatment

$$a^\star(x) \in \arg\max_{a \in \mathcal{A}} m_0(a, x) \tag{201}$$

is interior and varies with x_0. This property makes the benchmark appropriate for evaluating decision-aware bridge learning: errors near $a^\star(x)$ affect policy regret, whereas errors in clearly suboptimal regions are less consequential.

Figure 8 illustrates the effect of hidden confounding in the synthetic benchmark. For both an individual covariate value and the population average, the associational curves differ substantially from the causal curves, and their maximizers can occur at different treatment values. This confirms that selecting treatments using $\mathbb{E}[Y \mid A = a, X = x]$ or $\mathbb{E}[Y \mid A = a]$ is not equivalent to optimizing the interventional responses $\mathbb{E}\{Y(a) \mid X = x\}$ or $\mathbb{E}\{Y(a)\}$.

For reproducibility, the values used in all synthetic experiments are

$$\begin{gathered} d_X = 5, \quad d_Z = d_W = 2, \quad \rho_X = 0.5, \quad c_P = 0.35, \quad \sigma_A = 0.35, \quad \sigma_Y = 0, \\ q_{\min} = 0.001, \quad q_{\max} = 0.999, \quad n_{\text{pilot}} = 50{,}000, \quad \text{support seed} = 2718, \\ \beta_{XW} = 1.2, \quad \delta_c = 0.4 \times 0.5 = 0.2, \quad \eta_c = 2.7, \quad \delta_s = 0.1 \times 0.45 = 0.045, \\ \delta_\alpha = 0.25 + 0.1 = 0.35, \quad \delta_\kappa = 0.7, \quad \lambda_{\text{tail}} = 0.06, \\ \gamma = 0.7, \quad \beta_A = 0.25, \quad \delta_\psi = 0.18, \quad \gamma_0 = 1.2, \\ n_{\text{train}} = 3000, \quad n_{\text{test}} = 1000, \quad |\mathcal{G}_A| = 51. \end{gathered} \tag{202}$$

Here $|\mathcal{G}_A|$ denotes the number of equally spaced grid points used for policy optimization over $\mathcal{A}$.

I.2 Semi-synthetic TCGA proximal benchmark

We also evaluate on a semi-synthetic benchmark based on TCGA gene-expression covariates. The raw gene-expression matrix is taken from the TCGA benchmark used in SCIGAN [5], which is derived from the Cancer Genome Atlas pan-cancer project [41]. We use the real gene-expression features to construct observed covariates and proxy-relevant features, while the treatment, hidden

confounder, proxies, and outcome are generated from a controlled proximal structural model. This gives a benchmark with realistic covariate structure and known counterfactual outcomes.

The corresponding proximal causal graph is shown in Figure 9.

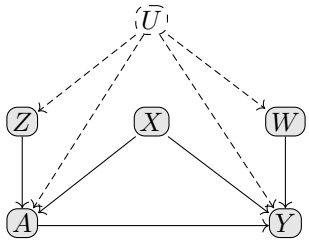


Figure 9: Semi-synthetic TCGA proximal graph. Observed variables are shown in gray, while the latent confounder is shown in white with a dashed border. The proxy Z is treatment-inducing and affects treatment assignment but not the outcome directly. The proxy W is outcome-inducing and affects the outcome but not treatment assignment directly.

Let $G^{\text{raw}} \in \mathbb{R}^{n\times p}$ denote the raw TCGA gene-expression matrix. We first standardize each gene,

$$G_{ij} = \frac{G_{ij}^{\text{raw}} - \bar{G}_j}{s_j + 10^{-8}}, \tag{203}$$

and select the $p_0 = 180$ genes with largest empirical variance. The selected genes are split into disjoint blocks used to construct observed covariates and proxy-specific features. We apply PCA to obtain

$$X_i = \text{PCA}_5(G_i^X) \in \mathbb{R}^5, \qquad \widetilde{Z}_i = \text{PCA}_3(G_i^Z) \in \mathbb{R}^3, \qquad \widetilde{W}_i = \text{PCA}_3(G_i^W) \in \mathbb{R}^3. \tag{204}$$

The variables observed by the learner are

$$O_i = (X_i, Z_i, W_i, A_i, Y_i), \tag{205}$$

where $Z_i \in \mathbb{R}^3$, $W_i \in \mathbb{R}^3$, and $A_i \in (0, 1)$.

The latent confounder is two-dimensional,

$$U_i = (U_{i1}, U_{i2})^\top, \qquad U_{i1} \sim \mathcal{N}(0, 1), \qquad U_{i2} \sim \text{Exp}(1) - 1. \tag{206}$$

The treatment and outcome proxies are generated as noisy nonlinear measurements of U_i, X_i, and the TCGA-derived proxy features:

$$Z_i = \tanh(B_Z U_i + C_Z X_i + R_Z \widetilde{Z}_i) + \epsilon_i^Z, \qquad \epsilon_i^Z \sim \mathcal{N}(0, \sigma_Z^2 I_3), \tag{207}$$

$$W_i = \tanh(B_W U_i + C_W X_i + R_W \widetilde{W}_i) + \epsilon_i^W, \qquad \epsilon_i^W \sim \mathcal{N}(0, \sigma_W^2 I_3). \tag{208}$$

Here $\sigma_Z = \sigma_W = 0.2$, and

$$B_Z = 0.8 \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1/\sqrt{2} & -1/\sqrt{2} \end{pmatrix}, \qquad B_W = 0.8 \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1/\sqrt{2} & 1/\sqrt{2} \end{pmatrix}, \tag{209}$$

$$C_Z = 0.2 \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{pmatrix}, \qquad C_W = 0.2 \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}, \tag{210}$$

$$R_Z = 0.5 I_3, \qquad R_W = 0.5 I_3. \tag{211}$$

Since both B_Z and B_W have rank two, both proxies carry information about the two-dimensional latent confounder. By construction, Z affects treatment assignment but does not enter the outcome equation directly, while W enters the outcome equation but does not enter the treatment assignment equation.

Each unit has a latent center for the dose-response curve,

$$s_i^\star = \gamma_0 + \gamma_X^\top X_i + \gamma_U^\top U_i, \qquad \bar{s}_i^\star = \frac{s_i^\star - \mu_s}{\sigma_s}, \tag{212}$$

$$a_i^\star = 0.15 + 0.70\,\sigma(\bar{s}_i^\star), \qquad \sigma(t) = \frac{1}{1+\exp(-t)}. \tag{213}$$

The constants μ_s and σ_s are the empirical mean and standard deviation of $s^\star$, estimated once from a pilot sample. The coefficients are

$$\gamma_0 = 0, \qquad \gamma_X = \frac{0.6}{\sqrt{5}}(1,-1,1,-1,1)^\top, \qquad \gamma_U = \frac{0.8}{\sqrt{2}}(1,-1)^\top. \tag{214}$$

Thus $a_i^\star \in (0.15, 0.85)$, so the relevant optima are interior to the treatment support.

Treatment assignment is generated from a beta distribution. First define

$$r_i = \eta_0 + \eta_X^\top X_i + \eta_U^\top U_i + \eta_Z^\top Z_i, \qquad \bar{r}_i = \frac{r_i - \mu_r}{\sigma_r}, \tag{215}$$

where μ_r and σ_r are pilot estimates of the mean and standard deviation of r. The assignment mean is

$$m_i = \mathrm{clip}\left\{0.3a_i^\star + 0.7\sigma(\bar{r}_i),\, 0.02,\, 0.98\right\}, \tag{216}$$

and

$$A_i \sim \mathrm{Beta}\{\phi m_i, \phi(1-m_i)\}, \qquad \phi = 20. \tag{217}$$

The assignment coefficients are

$$\eta_0 = 0, \qquad \eta_X = \frac{0.4}{\sqrt{5}}(1,1,-1,1,-1)^\top, \tag{218}$$

$$\eta_U = \frac{0.8}{\sqrt{2}}(1,1)^\top, \qquad \eta_Z = \frac{0.6}{\sqrt{3}}(1,-1,1)^\top. \tag{219}$$

Because A_i depends on U_i, the observational treatment is confounded.

The potential outcome under treatment $a \in [0,1]$ is

$$Y_i(a) = \frac{\mu_i + b_i + \lambda_W^\top W_i + \Gamma U_{i1} a X_{i0} - c_i(a - a_i^\star)^2 + \epsilon_i^Y}{4}, \qquad \epsilon_i^Y \sim \mathcal{N}(0, \sigma_Y^2), \tag{220}$$

with $\sigma_Y = 0.1$ and $\Gamma = 1.0$. The observed outcome is $Y_i = Y_i(A_i)$. The outcome components are

$$c_i = 5 + 5\,\mathrm{softplus}\left(\kappa_0 + \kappa_X^\top X_i + \kappa_U^\top U_i\right), \qquad \mathrm{softplus}(t) = \log(1+\exp(t)), \tag{221}$$

$$\mu_i = \theta_X^\top X_i + \theta_U^\top U_i, \qquad b_i = 0.5\sin(\omega_X^\top X_i) + 0.5\cos(\omega_U^\top U_i). \tag{222}$$

The corresponding coefficients are

$$\theta_X = \frac{0.25}{\sqrt{5}}(1,-1,1,1,-1)^\top, \qquad \theta_U = \frac{0.5}{\sqrt{2}}(1,-1)^\top, \tag{223}$$

$$\omega_X = \frac{0.2}{\sqrt{5}}(1,2,-1,1,-2)^\top, \qquad \omega_U = \frac{0.3}{\sqrt{2}}(1,1)^\top, \tag{224}$$

$$\kappa_0 = 0, \qquad \kappa_X = \frac{0.3}{\sqrt{5}}(1,-1,0.5,1,-0.5)^\top, \qquad \kappa_U = \frac{0.3}{\sqrt{2}}(1,-1)^\top, \tag{225}$$

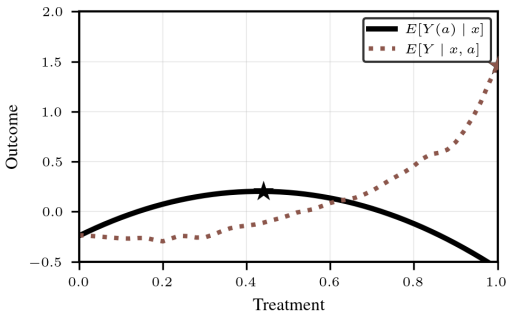
$$\lambda_W = \frac{0.5}{\sqrt{3}}(1,-1,1)^\top. \tag{226}$$

The quadratic term gives each unit a single-peaked dose-response curve centered near $a_i^\star$, while the additional interaction $(\Gamma U_{i1} a X_{i0})$ makes the latent confounding treatment-dependent. Therefore, hidden confounding can distort not only the level of the response curve, but also its shape.

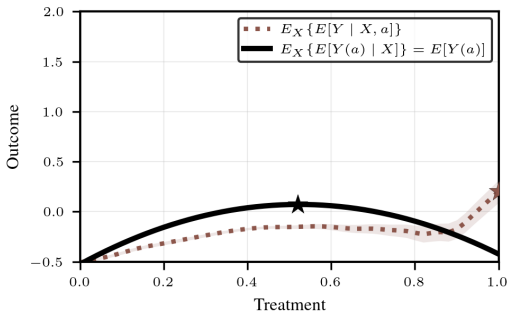
Figure 10 shows the same diagnostic comparison for the semi-synthetic TCGA benchmark. The causal curves retain an interior maximizer, while the associational curves are shifted by the confounded treatment assignment mechanism. This visualizes why the benchmark requires proxy-based adjustment: the observed conditional association does not recover the causal dose-response surface used for policy evaluation.

The causal estimand used for evaluation is the conditional response surface

$$m_0(a,x) = \mathbb{E}\{Y(a) \mid X = x\}. \tag{227}$$



(a) Individual-level curves for one randomly selected test point.



(b) Population-level curves.

Figure 10: Reference causal and associational curves in the semi-synthetic TCGA benchmark. Left: for a fixed covariate value x, the solid curve shows the interventional response $\mathbb{E}\{Y(a) \mid X = x\}$, while the dotted curve shows the associational response $\mathbb{E}\{Y \mid A = a, X = x\}$. Right: the corresponding population quantities $\mathbb{E}\{Y(a)\}$ and $\mathbb{E}\{Y \mid A = a\}$ are obtained by averaging over X. Stars mark the maximizers of the displayed curves. The discrepancy between causal and associational curves illustrates the impact of hidden confounding and shows that optimizing the associational response can lead to a different treatment recommendation than optimizing the causal response.

For this semi-synthetic DGP, $m_0(a, x)$ does not have a simple closed form because the outcome averages over U, W, proxy noise, and TCGA-derived features. We therefore evaluate $m_0(a, x)$, individualized optima, and policy values by Monte Carlo simulation from the known structural equations. The individualized policy target is

$$a^\star(x) \in \arg\max_{a \in \mathcal{G}_A} m_0(a, x), \tag{228}$$

where $\mathcal{G}_A$ is a grid of 51 equally spaced treatment values on $[0, 1]$.

For all semi-synthetic TCGA experiments we use

$$n_{\text{train}} = 3000, \qquad n_{\text{test}} = 1000, \qquad \mathcal{A} = [0, 1], \qquad |\mathcal{G}_A| = 51. \tag{229}$$

The TCGA preprocessing uses 180 top-variance genes, PCA dimensions $5, 3, 3$ for $X, \widetilde{Z}, \widetilde{W}$.

I.3 Monte Carlo construction of ground-truth causal curves

For both benchmarks, the ground-truth conditional causal response

$$m_0(a, x) = \mathbb{E}\{Y(a) \mid X = x\} \tag{230}$$

is evaluated directly from the known structural equations. This step is necessary because the effect of the latent confounders is non-additive and depends on both the treatment and covariates. Consequently, the conditional response cannot in general be read off from the deterministic part of the outcome equation by simply removing an additive confounding term.

For a fixed covariate value x and treatment level a, we use the structural intervention

$$do(A = a) \tag{231}$$

and marginalize over the remaining stochastic components of the data-generating process. That is, we write

$$m_0(a, x) = \mathbb{E}\{Y(a) \mid X = x\} = \mathbb{E}[Y \mid do(A = a), X = x], \tag{232}$$

where the expectation is taken over the latent confounders, proxy variables, and outcome noise generated under the intervention. Equivalently,

$$m_0(a, x) = \int \mathbb{E}[Y \mid do(A = a), X = x, U = u, W = w] \, dP(u, w \mid X = x). \tag{233}$$

Since the simulator specifies the full joint law, this integral can be approximated by Monte Carlo. For each evaluation covariate x and treatment value a, we repeatedly sample latent variables and outcome

proxies from their conditional distribution, set the treatment to a, and evaluate the interventional outcome equation.

Concretely, for $b = 1, \ldots, B$, we draw

$$U^{(b)}, W^{(b)} \sim P(U, W \mid X = x) \tag{234}$$

according to the corresponding data-generating process, set $A = a$, and compute

$$Y^{(b)}(a, x) = Y\Big(do(A = a), X = x, U^{(b)}, W^{(b)}\Big). \tag{235}$$

The ground-truth curve is then approximated by

$$\widehat{m}_0(a, x) = \frac{1}{B} \sum_{b=1}^{B} Y^{(b)}(a, x). \tag{236}$$

Once x is fixed, the simulator is used to generate the remaining latent and proxy variables under the intervention. We use $B = 256$ Monte Carlo draws for each pair (x, a).

Ground-truth individualized policies are computed by evaluating this Monte Carlo curve on the same treatment grid used by the learners:

$$\widehat{a}^{\star}(x) \in \arg\max_{a \in \mathcal{G}_A} \widehat{m}_0(a, x). \tag{237}$$

Policy values and regrets are then computed using these Monte Carlo estimates of the conditional causal response rather than associational quantities such as $\mathbb{E}[Y \mid A = a, X = x]$, which remain biased under hidden confounding.

J Hyperparameter Selection

This appendix describes the hyperparameter-selection protocol used in the experiments. The protocol has two stages. First, we select the base hyperparameters of each unweighted proximal learner using only factual validation error. Second, we calibrate the decision-aware weighting parameters using a limited shared calibration procedure and then keep them fixed across weighted methods and datasets. This separation avoids tuning model-specific weighting parameters using oracle quantities that would not be available in real observational applications.

J.1 Stage 1: Base model selection using factual validation error

For each dataset and each model class, we select the unweighted base learner by minimizing factual prediction error on held-out validation data. For a fitted bridge or response predictor $\widehat{f}_\lambda$, with hyperparameters λ, the validation objective is

$$\mathrm{RMSE}_{\mathrm{factual}}(\lambda) = \left[\frac{1}{n_{\mathrm{val}}} \sum_{i \in \mathcal{I}_{\mathrm{val}}} \Big\{ Y_i - \widehat{f}_\lambda(A_i, W_i, X_i) \Big\}^2 \right]^{1/2}. \tag{238}$$

For methods whose factual prediction is obtained through a bridge-induced surface rather than a direct regression function, $\widehat{f}_\lambda$ denotes the corresponding factual prediction used by that estimator. This objective uses only observed factual outcomes and does not require counterfactual or oracle information.

The search was performed with Bayesian optimization using a tree-structured Parzen estimator sampler [3], with a maximum budget of 100 trials per model class and dataset, using Optuna [1]. This stage includes only unweighted learners; decision-aware weighting parameters are calibrated separately in Stage J.2. The search spaces were as follows.

For PMMR, we searched the ridge regularization parameter over $\{10^{-4}, 5 \cdot 10^{-4}, 10^{-3}, 5 \cdot 10^{-3}, 10^{-2}, 5 \cdot 10^{-2}, 10^{-1}, 5 \cdot 10^{-1}, 1\}$. The bandwidth scale for the bridge input (A, W, X) was searched over $\{0.3, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.25, 2.5, 3.0, 3.5, 4.0\}$, and the bandwidth scale for the conditioning input (A, Z, X) was searched over $\{0.3, 0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.25, 2.5, 2.75, 3.0\}$.

For KPV, the first-stage regularization parameter was searched over $\{10^{-4}, 3 \cdot 10^{-4}, 10^{-3}, 3 \cdot 10^{-3}, 10^{-2}, 3 \cdot 10^{-2}, 10^{-1}\}$, and the second-stage regularization parameter over $\{10^{-3}, 3 \cdot 10^{-3}, 10^{-2}, 3 \cdot 10^{-2}, 10^{-1}, 3 \cdot 10^{-1}, 1\}$. The kernel variance scale was searched over $\{0.5, 0.75, 1.0, 1.25, 1.5, 1.75, 2.0, 2.5, 3.0\}$.

For DFPV, the first-stage and second-stage penalties were both searched over $\{10^{-3}, 3 \cdot 10^{-3}, 10^{-2}, 3 \cdot 10^{-2}, 10^{-1}, 3 \cdot 10^{-1}, 1\}$. The hidden-layer architecture was selected from $\{[32], [64], [32, 32], [64, 64], [128, 128]\}$, the learned feature dimension from $\{4, 8, 16\}$, the activation function from $\{\mathrm{ReLU}, \mathrm{ELU}, \tanh\}$, the learning rate from $\{3 \cdot 10^{-4}, 10^{-3}, 3 \cdot 10^{-3}\}$, and the weight decay from $\{0, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}\}$.

For NMMR-U, we searched the activation function over $\{\mathrm{ReLU}, \mathrm{ELU}, \mathrm{SiLU}\}$, whether to use a product kernel over $\{\mathrm{false}, \mathrm{true}\}$, the bandwidth scale for (A, Z, X) over $\{0.3, 0.5, 0.75, 1.0, 1.5, 2.0, 3.0\}$, the hidden-layer architecture over $\{[32, 32], [64, 64], [64, 64, 64]\}$, and the learning rate over $\{3 \cdot 10^{-5}, 10^{-4}, 3 \cdot 10^{-4}, 10^{-3}\}$. NMMR-V used the same search family as NMMR-U, with the only difference being the use of the V-statistic objective rather than the U-statistic objective.

J.2 Stage 2: Shared calibration of decision-aware weighting

After selecting the unweighted base hyperparameters, we calibrate the decision-aware weighting parameters. The weighted bridge objective uses a local-global weight controlled by a localization strength λ, a bandwidth τ, and the number of alternating reweighting rounds n_{rounds}. These parameters determine how strongly the bridge loss is concentrated near the current pseudo-optimal treatment while retaining global stabilization.

In real observational settings, the target policy regret is not observable, since it depends on the unknown conditional causal response $m_0(a, x)$. A full model-specific regret-based tuning procedure would therefore use information unavailable in practice. To avoid this, we use a conservative shared calibration protocol. We tune $(\lambda, \tau, n_{rounds})$ once on a synthetic calibration setting where ground-truth causal curves are available, using a single base model and a single random seed. The selected weighting parameters are then fixed and reused for all weighted methods and datasets.

The weighting search space is

$$\lambda \sim \mathrm{LogUniform}(0.2, 20.0), \qquad \tau \sim \mathrm{LogUniform}(0.02, 2.0), \qquad n_{rounds} \in \{2, 3, 4, 5\}. \tag{239}$$

This protocol deliberately prevents each weighted model from receiving separate access to oracle regret tuning. As a result, performance differences among weighted methods are not driven by unequal model-specific optimization over non-observable objectives.

J.3 Sensitivity Studies

To assess sensitivity to the weighting parameters, we also performed ablation studies in Section 6 over the localization strength, bandwidth, number of reweighting rounds, and confounding intensity. These studies show that performance is stable over a nontrivial neighborhood of the selected weighting parameters. This supports the use of a shared weighting protocol: the decision-aware gains are not attributable to a single fragile hyperparameter choice, and the calibrated parameters transfer across model classes and datasets in our experiments.

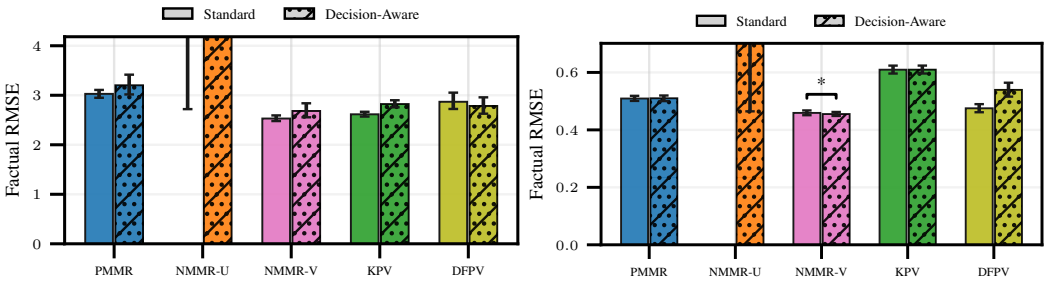
K Additional Results

K.1 Global RMSE

Table 3 shows the same counterfactual RMSE results as Figure 4 for the non doubly robust methods, but NMMR-U results are better appreciated here. Figure 11 shows factual RMSE for the baseline and DA proximal solvers and for the synthetic and semi-synthetic datasets. The conclusions are similar as with the counterfactual RMSE results.

Table 3: Counterfactual RMSE across methods and training modes for the synthetic and semi-synthetic settings. Values are reported as mean_{std} over 10 runs. Lower is better.

	Synthetic					Semi-synthetic				
	DFPV	KPV	NMMR-U	NMMR-V	PMMR	DFPV	KPV	NMMR-U	NMMR-V	PMMR
Standard	$6.59_{0.51}$	$4.25_{0.16}$	$12.16_{22.52}$	$4.38_{0.20}$	$6.85_{0.42}$	$0.59_{0.13}$	$0.33_{0.01}$	$3.17_{6.01}$	$0.49_{0.04}$	$0.37_{0.03}$
DA	$5.30_{0.80}$	$3.97_{0.11}$	$32.94_{31.28}$	$4.05_{0.57}$	$6.70_{0.54}$	$0.61_{0.11}$	$0.32_{0.01}$	$1.01_{0.01}$	$0.55_{0.06}$	$0.40_{0.04}$



(a) Synthetic experiment

(b) semi-synthetic experiment

Figure 11: Factual RMSE in the synthetic and semi-synthetic datasets. Mean and 95% confidence intervals over 10 seeds are shown. * denotes statistical difference between baselines and DA models.

K.2 Sensitivity to Weakened Completeness

Motivation. Assumption 3 requires the proxy system to contain sufficient information to distinguish the latent signals relevant to proximal identification. This condition may become weak or fail when the observed proxies are noisy or coarsened measurements of the latent confounders. We therefore conduct a controlled sensitivity analysis that progressively degrades the information carried by both proxy channels while leaving the remaining structural equations of the synthetic DGP unchanged. The purpose is to examine how the proximal bridge estimators and their decision-aware variants deteriorate along the same degradation path, rather than to claim identification once completeness fundamentally fails.

Perturbed proxy mechanism. Recall that the proxies in the original synthetic benchmark are generated as

$$Z = V_Z + c_P U_Z \mathbf{1}_{d_Z}, \qquad W = V_W + c_P U_W \mathbf{1}_{d_W}, \tag{240}$$

where U_Z and U_W are correlated latent confounding components, V_Z and V_W are proxy-specific noise variables, and c_P controls the strength of the latent signal in the proxies.

Let $\alpha \in [0, 1]$ denote the degradation level. We define the width of the collapsed central band and the scale of the additional noise as

$$b_\alpha = \alpha c_{\max}, \qquad \sigma_\alpha = \alpha \sigma_{\max}, \tag{241}$$

where $c_{\max}$ and $\sigma_{\max}$ are fixed constants controlling, respectively, the maximum collapse width and the maximum additional-noise scale. For a latent scalar u, define the central-band dropout transformation

$$\mathcal{C}_\alpha(u) = u\mathbf{1}\{|u| > b_\alpha\} = \begin{cases} 0, & |u| \le b_\alpha, \\ u, & |u| > b_\alpha. \end{cases} \tag{242}$$

The perturbed proxies are then generated as

$$\begin{aligned} Z_\alpha &= V_Z + c_P \mathcal{C}_\alpha(U_Z)\mathbf{1}_{d_Z} + \varepsilon_{Z,\alpha}, \\ W_\alpha &= V_W + c_P \mathcal{C}_\alpha(U_W)\mathbf{1}_{d_W} + \varepsilon_{W,\alpha}, \end{aligned} \tag{243}$$

where

$$\varepsilon_{Z,\alpha} \sim \mathcal{N}(0, \sigma_\alpha^2 I_{d_Z}), \qquad \varepsilon_{W,\alpha} \sim \mathcal{N}(0, \sigma_\alpha^2 I_{d_W}), \tag{244}$$

independently of one another and of the remaining variables.

Table 4: Policy regret under progressive degradation of proxy informativeness. Values are means ± standard deviations over five seeds. Lower is better.

	PMMR		KPV		DFPV	
α	Standard	DA	Standard	DA	Standard	DA
0.00	5.37 ± 2.09	4.73 ± 1.87	4.29 ± 0.33	3.30 ± 0.44	5.83 ± 2.49	3.25 ± 1.24
0.25	5.23 ± 2.16	4.56 ± 1.85	4.33 ± 0.37	3.32 ± 0.45	7.41 ± 2.07	3.18 ± 1.95
0.50	6.32 ± 1.96	5.65 ± 2.04	4.53 ± 0.48	3.79 ± 0.86	8.83 ± 0.06	4.75 ± 2.99
0.75	6.79 ± 1.64	6.19 ± 2.16	5.11 ± 0.66	4.44 ± 1.27	8.26 ± 0.83	5.11 ± 2.10
0.90	7.77 ± 0.75	7.46 ± 0.95	5.87 ± 1.11	4.89 ± 1.65	7.89 ± 1.03	5.73 ± 2.42

Table 5: Global counterfactual RMSE under progressive degradation of proxy informativeness. Values are means ± standard deviations over five seeds. Lower is better.

	PMMR		KPV		DFPV	
α	Standard	DA	Standard	DA	Standard	DA
0.00	6.72 ± 0.13	6.53 ± 0.10	3.98 ± 0.13	3.94 ± 0.15	6.12 ± 0.41	5.49 ± 0.37
0.25	6.63 ± 0.27	6.37 ± 0.13	4.15 ± 0.10	4.07 ± 0.10	6.50 ± 0.93	5.11 ± 0.84
0.50	6.77 ± 0.22	6.44 ± 0.12	4.28 ± 0.17	4.32 ± 0.17	6.65 ± 0.86	6.43 ± 0.44
0.75	7.20 ± 0.76	6.92 ± 0.75	4.43 ± 0.07	4.44 ± 0.16	6.79 ± 1.06	6.26 ± 1.06
0.90	7.29 ± 0.47	7.02 ± 0.54	4.50 ± 0.15	4.57 ± 0.20	6.48 ± 0.86	6.06 ± 0.47

At $\alpha = 0$, we have $b_0 = \sigma_0 = 0$ and $\mathcal{C}_0(u) = u$ almost surely because the latent variables are continuous. Hence,

$$Z_0 = V_Z + c_P U_Z \mathbf{1}_{d_Z}, \qquad W_0 = V_W + c_P U_W \mathbf{1}_{d_W}, \tag{245}$$

so the original synthetic DGP is recovered exactly.

For $\alpha > 0$, every latent value in the interval $[-b_\alpha, b_\alpha]$ has the same direct contribution to its corresponding proxy. Consequently, distinct latent states within this interval become indistinguishable through the direct proxy signal. The expanding dead zone therefore introduces structured information loss, while the additional Gaussian noise prevents the coarsening transformation from acting as a benign denoising operation. The same symmetric transformation is applied to both latent proxy channels, but their added noise terms remain independent.

The generation of X, the joint distribution and correlation structure of (U_Z, U_W), and the structural treatment and outcome equations are otherwise unchanged. In particular, the treatment-response function and the procedure used to calculate the ground-truth causal response remain fixed. Thus, the only modified structural equations are those producing Z and W. Because these proxies subsequently enter the treatment and outcome equations, their perturbation can naturally change the resulting observed-data distribution even though the functional forms of those mechanisms are not altered.

This construction directly reduces the information supplied by the proxies and makes the associated proximal inverse problem increasingly difficult. Nevertheless, it should be interpreted as a controlled stress test that moves the DGP away from the completeness conditions. The non-injectivity of the direct latent-to-proxy mapping does not, by itself, formally characterize the relevant conditional-expectation operators for every $\alpha > 0$.

Experimental protocol. We consider

$$\alpha \in \{0, 0.25, 0.50, 0.75, 0.90\}. \tag{246}$$

At every degradation level, we evaluate PMMR, KPV, and DFPV together with their corresponding decision-aware variants. All remaining experimental settings follow the main synthetic benchmark. We report policy regret and global counterfactual RMSE, averaged over five random seeds. The estimates at $\alpha = 0$ need not exactly match the main synthetic results because the latter are averaged over ten seeds.

Results. Tables 4 and 5 show an overall deterioration as the proxy perturbation becomes stronger, although finite-sample variability makes the trajectories non-monotone for some solvers. Trends are better seen in Figure 12. Comparing the undegraded setting with $\alpha = 0.90$, mean regret increases

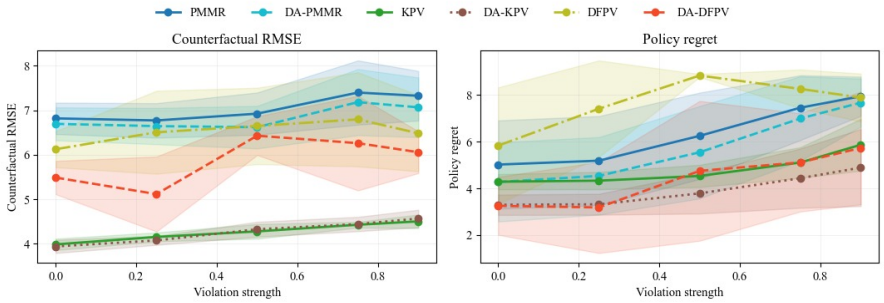


Figure 12: Counterfactual RMSE and policy regret varying the violation strenght factor, α. 95 % CIs over 10 seeds.

from 5.37 to 7.77 for PMMR, from 4.29 to 5.87 for KPV, and from 5.83 to 7.89 for DFPV. The corresponding decision-aware variants also deteriorate, showing that policy-targeted weighting does not remove the consequences of losing proxy information.

Nevertheless, the decision-aware variant achieves lower mean regret than its unweighted counterpart for all three solvers at every degradation level considered. Under the strongest perturbation, mean regret decreases from 7.77 to 7.46 for PMMR, from 5.87 to 4.89 for KPV, and from 7.89 to 5.73 for DFPV. The advantage becomes relatively small for PMMR under severe degradation, whereas clearer improvements remain for KPV and DFPV.

Global counterfactual RMSE remains broadly comparable between the weighted and unweighted estimators. DA-PMMR and DA-DFPV attain lower mean RMSE at every degradation level. For KPV, the differences are small: DA-KPV performs slightly better at $\alpha \in \{0, 0.25\}$ and slightly worse at the three stronger degradation levels. The policy-regret improvements therefore do not appear to be obtained through a large systematic sacrifice in accuracy over the full treatment region.

Overall, these results show that decision-aware weighting can remain useful when the proxies are moderately degraded and that this behavior is not specific to a single proximal bridge solver. At the same time, the deterioration of both the standard and decision-aware estimators emphasizes the limitation of this robustness: decision-aware weighting can prioritize the information that remains in decision-relevant regions, but it cannot reconstruct latent information removed from the proxy system or restore identification after a fundamental failure of completeness.

K.3 Pure-Local versus Local–Global Weighting

Experimental protocol. On the synthetic benchmark, we compare unweighted proximal learners, pure-local weighting, and the proposed local-plus-global decision-aware weighting for PMMR, KPV, and DFPV over ten random seeds. The pure-local variant uses the same hyperparameters as the corresponding decision-aware variant, except that the global component, the constant 1 in the weight

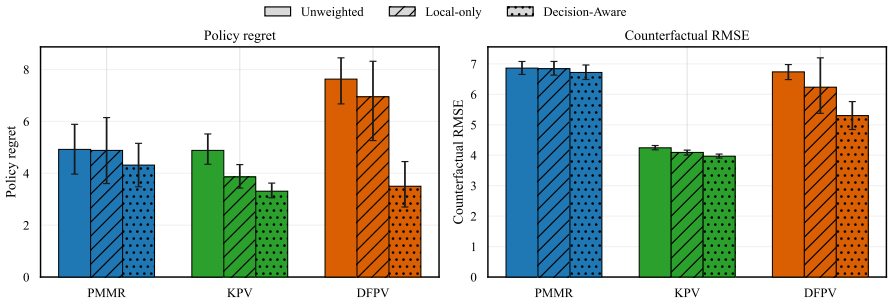


Figure 13: Policy regret and counterfactual RMSE comparing unweighted, pure-local and decision-aware weighting. 95 % confidence intervals over 5 seeds. Lower is better.

Table 6: Policy regret for unweighted, pure-local, and local-plus-global decision-aware weighting on the synthetic benchmark over ten seeds. Entries give means with uncertainty intervals in brackets. Lower is better.

Learner	Unweighted	Pure-local	Decision-Aware
PMMR	4.944 [3.867, 6.025]	4.855 [3.339, 6.360]	4.377 [3.445, 5.288]
KPV	4.867 [4.240, 5.493]	3.845 [3.292, 4.398]	3.302 [2.947, 3.657]
DFPV	7.524 [6.475, 8.572]	6.894 [5.000, 8.787]	3.493 [2.415, 4.571]

Table 7: Global counterfactual RMSE for unweighted, pure-local, and local-plus-global decision-aware weighting on the synthetic benchmark over ten seeds. Entries give means with uncertainty intervals in brackets. Lower is better.

Learner	Unweighted	Pure-local	Decision-Aware
PMMR	6.862 [6.615, 7.109]	6.842 [6.561, 7.123]	6.719 [6.442, 6.997]
KPV	4.244 [4.164, 4.323]	4.093 [3.993, 4.194]	3.960 [3.890, 4.048]
DFPV	6.738 [6.455, 7.021]	6.236 [5.119, 7.353]	5.301 [4.732, 5.870]

numerator, is removed. Tables 6 and 7 and Figure 13 report means with uncertainty intervals in brackets.

Results. Compared with pure-local weighting, local-plus-global weighting reduces mean regret by approximately 9.8%, 14.1%, and 49.3% for PMMR, KPV, and DFPV, respectively, and has lower mean global counterfactual RMSE for all three solvers. Pure-local weighting may assign zero weight away from the kernel neighborhood, whereas the global component maintains positive weighting throughout the treatment space. These findings are consistent with the intended global stabilization role.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: The claims in the abstract and introduction properly reflect the scope.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: The paper discusses the limitations, see Section 7.

   Guidelines:

   - The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

Justification: The paper provides the assumptions for all the theoretical statements.

Guidelines:

- The answer [N/A] means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: The paper discloses all the information for reproducibility. The code is provided in a public repository.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [Yes]

Justification: The paper provides access to the code and enough instructions.

Guidelines:

- The answer [N/A] means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (https://neurips.cc/public/guides/CodeSubmissionPolicy) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: The paper specifies all the settings, see Appendices H and J.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: The paper provides error bars for the results. It can be checked in all tables and figures (e.g. 3).

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.

- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
- If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: The paper provides information on the computer resources. See Appendix H

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: The paper conforms with the Neurips code of ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [Yes]

    Justification: The paper discusses the positive impact of optimal decision-making.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.

- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [Yes]

    Justification: The paper describes proper safeguards.

    Guidelines:

    - The answer [N/A] means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: The creators of datasets are properly credited.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.

- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [Yes]

    Justification: New assets are properly documented.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

    Justification: There are no experiments involving humans.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [N/A]

    Justification: There are no experiments involving humans.

    Guidelines:

    - The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
    - We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
    - For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The usage of LLMs is standard.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.