FINITE-SAMPLE GUARANTEES FOR CONVEX NEURAL PREFERENCE LEARNING

Anonymous authors
Paper under double-blind review

ABSTRACT

Pairwise preference data are commonly used to train reward models. Yet existing statistical analyses of pairwise preference learning often assume linear reward heads, leaving finite-sample guarantees for nonlinear heads less developed. We study a Bradley–Terry estimator that fits a two-layer ReLU reward head on fixed representations through a convex activation-pattern formulation. The estimator jointly fits the preference likelihood and selects among retained activation patterns using a block-sparsity penalty. Conditioning on the representations and retained patterns, we analyze squared error in reward differences on the observed comparisons. For rewards supported on at most s of P patterns in dimension d with n samples, we establish a minimax lower bound of order $s[d + \log(P/s)]/n$ under a sparse design condition. We further derive a prediction error bound of order $s(d + \log P)/n$ for the convex estimator under standard conditions. When s is known, a sharper calibration of the same estimator achieves the minimax lower bound rate under a stronger restricted eigenvalue condition. Experiments with synthetic rewards, language-model representations, and five preference datasets assess reward recovery and held-out preference prediction.

1 INTRODUCTION

Learning a reward function from pairwise comparisons is critical in preference-based language-model training. In reinforcement learning from human feedback (RLHF), pairwise human judgments are used to fit a reward model, whose outputs provide the preference signal for subsequent policy optimization (Stiennon et al., 2020; Ouyang et al., 2022; Bai et al., 2022). The statistical quality of the learned reward model therefore directly affects how faithfully the preference information is represented during post-training. Given two responses to the same prompt, a comparison provides a judgment about which response is preferred, rather than a numerical assessment of their quality. Reward modeling turns these judgments into a function whose differences explain the observed choices. Beyond fitting the preference labels, an important statistical question is how accurately a reward model can recover these latent differences from a limited number of comparisons.

A substantial body of theoretical work on preference learning assumes that the latent reward is linear in a fixed feature representation. This assumption underlies finite-sample analyses of reward estimation and policy learning (Zhu et al., 2023), as well as sparsity-aware estimation in high-dimensional feature spaces (Yao et al., 2025). Related statistical work accommodates heterogeneous feedback through a linear reward with annotator-dependent scales (Liu et al., 2025) or a low-rank bilinear reward that remains linear in prompt–response features conditional on user context (Lee et al., 2026). However, the underlying reward model need not be linear in the prescribed representation. In particular, even when the representation is extracted from a nonlinear pretrained model, there is no guarantee that a single linear head can capture the reward structure relevant to human preferences. Nonlinear reward estimation has also received recent theoretical attention. Li et al. (2024) derive nonparametric guarantees for deep ReLU reward learning within an off-policy evaluation pipeline, while Sun et al. (2025) establish asymptotic convergence guarantees for neural-network Bradley–Terry reward models on fixed representations. To the best of our knowledge, the minimax complexity of nonlinear pairwise reward-difference estimation remains uncharacterized, and no globally solvable neural estimator has been shown to attain the corresponding minimax rate.

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

Convex neural networks provide a natural framework for addressing this gap because they combine nonlinear modeling with globally solvable optimization. Earlier work studied convex neural networks with an unbounded collection of hidden units (Bengio et al., 2005), while Bach (2017) developed a statistical theory for single-hidden-layer positively homogeneous networks under non-Euclidean regularization of the output weights. Pilanci & Ergen (2020) show that weight-decayed two-layer ReLU networks admit a convex reformulation indexed by activation patterns on the observed design. Building on this framework, COALA (Feng & Pilanci, 2026) trains a convex neural head on frozen language-model representations and subsequently performs preference fine-tuning with its first-layer weights held fixed. Its optimization guarantees, however, do not characterize the statistical accuracy of the estimated reward differences or establish rate optimality. Establishing such guarantees requires identifying an appropriate measure of complexity for the nonlinear reward class. Nominal network width is unsuitable because hidden units can be duplicated and rescaled without changing the represented function. This motivates our central question: *what determines the statistical complexity of nonlinear reward-difference estimation from pairwise comparisons, and can a convex estimator attain the minimax rate?*

To address this question, we propose a one-stage Bradley–Terry estimator that fits a two-layer ReLU reward head directly on fixed representations through an activation-pattern convex formulation. The estimator jointly optimizes the retained pattern coefficients under a block-sparsity penalty, and it can be computed using a matrix-free ADMM scheme. With fixed representations and retained activation patterns, we establish the minimax rate for estimating pairwise reward differences over the class of rewards supported on at most s patterns. Under standard conditions, we show that the proposed convex estimator achieves a high-probability prediction bound, which can be sharpened through an s-dependent calibration that attains the minimax rate under a stronger restricted-eigenvalue condition, without introducing a cardinality constraint or a different penalty. We further extend the guarantee beyond exact sparse realizability by showing that the prediction error is controlled by the approximation error of the retained sparse class together with the finite-sample estimation error. The proposed method is evaluated on synthetic, semi-synthetic, and real preference data using fixed language-model representations.

2 One-stage convex neural preference learning

2.1 Preference model and reward class

We observe n independent comparisons $\mathcal{D}_n = \{(\mathscr{S}_i, \mathscr{A}_{i1}, \mathscr{A}_{i0}, Y_i)\}_{i=1}^n$, where $\mathscr{S}_i \in \mathcal{S}$ is a prompt, $\mathscr{A}_{ij} \in \mathcal{A}$ is a response, and $Y_i \in \{-1, +1\}$ indicates whether response 0 or 1 is preferred. Write $Z_{ij} = (\mathscr{S}_i, \mathscr{A}_{ij}) \in \mathcal{Z} := \mathcal{S} \times \mathcal{A}$, and let $h : \mathcal{Z} \to \mathbb{R}^d$ be a fixed representation map, such as an embedding extracted from a frozen language model. For a reward function r, define $\Delta_i r = r(Z_{i1}) - r(Z_{i0})$. We consider the following Bradley–Terry model (Bradley & Terry, 1952),

$$\mathbb{P}_{r^\star}(Y_i = y \mid Z_{i1}, Z_{i0}) = \sigma(y \Delta_i r^\star), \qquad y \in \{-1, +1\}, \tag{1}$$

where $\sigma(t) = (1 + e^{-t})^{-1}$ and $r^\star$ is the unknown reward.

We assume that preference probabilities are bounded away from zero and one: there exists a fixed $\varepsilon \in (0, 1/2)$ such that, for every comparison, $\varepsilon \le \mathbb{P}_{r^\star}(Y = 1 \mid Z_1, Z_0) \le 1 - \varepsilon$. Under equation 1, this is equivalent to $|r^\star(Z_1) - r^\star(Z_0)| \le M$, where $M = \log\{(1 - \varepsilon)/\varepsilon\}$. This standard overlap condition rules out nearly deterministic comparisons, for which binary feedback carries little information about the magnitude of the reward difference. It ensures that the logistic loss retains uniformly positive curvature over the feasible range.

We consider the reward class that consists of width-m two-layer ReLU heads $r_{\theta_1,\theta_2}(z) = \sum_{j=1}^m \theta_{2j}(\theta_{1j}^\top h(z))_+$, where $(t)_+ = \max\{t, 0\}$. A direct estimator minimizes the paired Bradley–Terry loss with weight decay:

$$(\widehat{\theta}_1, \widehat{\theta}_2) \in \underset{\substack{\theta_{1j} \in \mathbb{R}^d,\ j \in [m],\ \theta_2 \in \mathbb{R}^m, \\ \max_{i \in [n]} |\Delta_i r_{\theta_1,\theta_2}| \le M}}{\arg\min} \left\{ \frac{1}{n} \sum_{i=1}^n \ell(Y_i \Delta_i r_{\theta_1,\theta_2}) + \frac{\lambda}{2} \sum_{j=1}^m (\|\theta_{1j}\|_2^2 + \theta_{2j}^2) \right\}. \tag{2}$$

where $\ell(t) = \log(1 + e^{-t})$. This formulation is nonconvex in the neural-network parameters, making a sharp statistical analysis of the resulting estimator challenging. We exploit the positive homogeneity of ReLU and reformulate the problem as a convex optimization over activation-pattern

coefficients, as described next. The equivalence between the weight-decay penalty and the resulting group-structured penalty is established in Appendix B.

2.2 PROPOSED CONVEX ESTIMATOR

Stack the embeddings into $X \in \mathbb{R}^{2n\times d}$ with rows $X_{2i-1,\cdot} = h(Z_{i1})^\top$ and $X_{2i,\cdot} = h(Z_{i0})^\top$. Let $B \in \mathbb{R}^{n\times 2n}$ be the pairwise incidence operator, defined by $(Ba)_i = a_{2i-1} - a_{2i}$. Draw Gaussian directions $g_p \sim N(0, I_d)$ and retain P distinct masks $d_p = \mathbf{1}\{Xg_p \geq 0\}$. Write $D_p = \mathrm{diag}(d_p)$, $E_p = (2D_p - I_{2n})X$, and $\mathcal{K}_p = \{a \in \mathbb{R}^d : E_p a \geq 0\}$. For $a \in \mathcal{K}_p$, the training activations satisfy $(Xa)_+ = D_p Xa$. The sampled direction indexes a cone, over which the coefficients are optimized.

For each pattern, introduce positive- and negative-output blocks $v_p, w_p \in \mathcal{K}_p$. Define $A_p = BD_pX \in \mathbb{R}^{n\times d}$ and $A = [A_1 \cdots A_P] \in \mathbb{R}^{n\times dP}$. Collect the blocks as $q = (v_1, \ldots, v_P, w_1, \ldots, w_P) \in \mathbb{R}^{2dP}$, with feasible set $\mathcal{K}_P = \{q : v_p, w_p \in \mathcal{K}_p \text{ for all } p\}$ and penalty $\|q\|_{2,1} = \sum_{p=1}^P (\|v_p\|_2 + \|w_p\|_2)$. Define the linear pairwise-margin operator $F = [A_1, \cdots, A_P, -A_1, \cdots, -A_P] \in \mathbb{R}^{n\times 2dP}$, so that $Fq = \sum_{p=1}^P A_p(v_p - w_p)$ is the vector of fitted reward differences on the observed comparisons. ReLU homogeneity converts the balanced weight-decay penalty into $\|q\|_{2,1}$, yielding the following one-stage estimator

$$\widehat{q} \in \underset{q\in\mathcal{K}_P,\ \|Fq\|_\infty \leq M}{\arg\min} \left\{ \frac{1}{n}\sum_{i=1}^n \ell\big(Y_i(Fq)_i\big) + \lambda\|q\|_{2,1} \right\}. \tag{3}$$

With the full collection of activation patterns and sufficient hidden-layer width, this formulation is exactly equivalent to the weight-decayed neural problem (Pilanci & Ergen, 2020). For computational feasibility, we analyze the estimator conditional on a sampled set of activation patterns, yielding a restricted convex model over the retained cones. Section 3.2 discusses how the theoretical results extend to the full activation-pattern collection.

For statistical analysis, the contribution of pattern p depends on (v_p, w_p) only through $u_p = v_p - w_p$. Let $\mathcal{U}_p = \{v - w : v \in \mathcal{K}_p, w \in \mathcal{K}_p\}$ and $\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_P$. The penalty induced on u_p is

$$\gamma_p(u_p) = \inf_{v,w\in\mathcal{K}_p:\ v-w=u_p} \big\{\|v\|_2 + \|w\|_2\big\}, \qquad \Gamma(u) = \sum_{p=1}^P \gamma_p(u_p). \tag{4}$$

This induces a norm γ_p on $\mathcal{U}_p$; its basic properties are established in Appendix B.2. Since $Fq = Au$, minimizing over the cone decompositions gives the equivalent estimator

$$\widehat{u} \in \underset{u\in\mathcal{U},\ \|Au\|_\infty \leq M}{\arg\min} \left\{ \frac{1}{n}\sum_{i=1}^n \ell[Y_i(Au)_i] + \lambda\Gamma(u) \right\}, \tag{5}$$

which has the same optimal value as equation 3 with $\sum_{p=1}^P A_p\widehat{u}_p = \sum_{p=1}^P A_p(\widehat{v}_p - \widehat{w}_p) = F\widehat{q}$.

2.3 COMPUTATION

We solve equation 3 using the alternating direction method of multipliers (ADMM) (Boyd et al., 2011). Let $C = \mathrm{blkdiag}(E_1, \ldots, E_P, E_1, \ldots, E_P)$, so that $q \in \mathcal{K}_P$ is equivalent to $Cq \geq 0$. Let $\mathcal{B}_M = [-M, M]^n$, $\mathbb{R}^{4nP}_+ = \{\xi \in \mathbb{R}^{4nP} : \xi \geq 0\}$, and let $\mathcal{I}_\mathcal{C}$ be zero on $\mathcal{C}$ and $+\infty$ otherwise. Introducing auxiliary variables gives the following equivalent form of equation 3:

$$\min_{q,\eta,z,\xi} \frac{1}{n}\sum_{i=1}^n \ell(Y_i\eta_i) + \mathcal{I}_{\mathcal{B}_M}(\eta) + \lambda\|z\|_{2,1} + \mathcal{I}_{\mathbb{R}^{4nP}_+}(\xi) \qquad \text{subject to} \quad \eta = Fq,\ q = z,\ Cq = \xi. \tag{6}$$

Take the two blocks of optimization variables as q and (η, z, ξ), with scaled dual variables (α, τ, ν) and fixed positive ADMM parameters $(\rho_\eta, \rho_z, \rho_\xi)$. The pseudo-algorithm is provided in Algorithm 1, where PCG denotes the standard preconditioned conjugate gradients. For a block vector $a = (a_1, \ldots, a_{2P})$ with $a_j \in \mathbb{R}^d$, define the block soft-thresholding operator $\mathcal{S}_t(a) = [s_t(a_1), \ldots, s_t(a_{2P})]$, where $s_t(x) = (1 - t/\|x\|_2)_+ x$, with $s_t(0) = 0$.

Algorithm 1 One-stage convex neural preference learning

Require: Design X, labels Y, pattern count P, λ, M, positive ADMM parameters $(\rho_\eta, \rho_z, \rho_\xi)$.
Sample the retained patterns from X and define matrix-free operators F, C.
Initialize $q^0, \eta^0, z^0, \xi^0, \alpha^0, \tau^0, \nu^0$ to zero.
Define $H = \rho_\eta F^\top F + \rho_z I + \rho_\xi C^\top C$ and construct its Nyström preconditioner.
repeat
$b^k \leftarrow \rho_\eta F^\top(\eta^k + \alpha^k) + \rho_z(z^k - \tau^k) + \rho_\xi C^\top(\xi^k - \nu^k)$.
Solve for q^{k+1} satisfying $Hq^{k+1} = b^k$ by PCG.
$\eta_i^{k+1} \leftarrow \arg\min_{|t|\le M} \left\{ \frac{1}{n}\ell(Y_i t) + \frac{\rho_\eta}{2}\left[t - (Fq^{k+1})_i + \alpha_i^k\right]^2 \right\}$, for all i.
$z^{k+1} \leftarrow \mathcal{S}_{\lambda/\rho_z}(q^{k+1} + \tau^k)$; $\xi^{k+1} \leftarrow (Cq^{k+1} + \nu^k)_+$.
$\alpha^{k+1} \leftarrow \alpha^k + \eta^{k+1} - Fq^{k+1}$.
$\tau^{k+1} \leftarrow \tau^k + q^{k+1} - z^{k+1}$; $\nu^{k+1} \leftarrow \nu^k + Cq^{k+1} - \xi^{k+1}$.
until convergence
return $\widehat{q} = q^{k+1}$, $\widehat{u}_p = \widehat{v}_p - \widehat{w}_p$, and $\widehat{r}$.

Our optimization scheme is closely related to COALA (Feng & Pilanci, 2026) and CRONOS (Feng et al., 2024). Like COALA, we fit a convex neural head on top of fixed representations. The main difference is that COALA uses a two-stage procedure, while our method estimates the preference model jointly under the Bradley–Terry loss. We introduce an auxiliary margin variable $\eta = Fq$ so that the q-subproblem becomes quadratic. The resulting linear system is solved by Nyström-preconditioned PCG (Frangella et al., 2023). Implementation details are given in Appendix C.

3 Theoretical analysis

Throughout this section, we condition on the design X and the sampled activation patterns $D_1, \dots, D_P$. Let $\eta^\star \in \mathbb{R}^n$ denote the true pairwise margin vector on the observed comparisons, with $\eta_i^\star = r^\star(Z_{i1}) - r^\star(Z_{i0})$, and suppose that the labels are independent with $\mathbb{P}(Y_i = y \mid X) = \sigma(y\eta_i^\star)$ for $y \in \{-1, +1\}$. For the realized pattern collection, we assume that $\eta^\star$ admits a sparse representation $\eta^\star = Au^\star$ for some $u^\star \in \mathcal{U}$, and write $S = \operatorname{supp}(u^\star)$ with $|S| \le s$. Here $\eta^\star$ is determined by the underlying preference model, whereas $u^\star$, its support S, and the design conditions below are defined relative to the retained pattern collection.

Define $\mu_M = \inf_{|t|\le M} \sigma(t)[1 - \sigma(t)] = \sigma(M)[1 - \sigma(M)]$. With $Au^\star = \eta^\star$, our analytical target is the pairwise prediction error of the proposed estimator $\|\widehat{u} - u^\star\|_{A,n}^2 := \|A\widehat{u} - \eta^\star\|_n^2 := \frac{1}{n}\|A\widehat{u} - \eta^\star\|_2^2 = \frac{1}{n}\sum_{i=1}^n [(A\widehat{u})_i - \eta_i^\star]^2$, which measures reward-difference estimation error on the observed comparison design.

3.1 Minimax lower bound

Define the sparse bounded-margin parameter class and its induced margin class by

$$\mathcal{F}_{s,M} := \{u = (u_1, \dots, u_P) \in \mathcal{U} : |\operatorname{supp}(u)| \le s,\ \|Au\|_\infty \le M\}, \quad \mathcal{G}_{s,M} := \{Au : u \in \mathcal{F}_{s,M}\}.$$

For an estimator $\widehat{\eta}$ of the true margin vector $\eta^\star = Au^\star$, consider the minimax risk

$$\mathfrak{R}_n(\mathcal{G}_{s,M}) := \inf_{\widehat{\eta}} \sup_{\eta^\star \in \mathcal{G}_{s,M}} \mathbb{E}_{\eta^\star}\left[\|\widehat{\eta} - \eta^\star\|_n^2\right] = \inf_{\widehat{u}} \sup_{u^\star \in \mathcal{F}_{s,M}} \mathbb{E}_{u^\star}\left[\|\widehat{u} - u^\star\|_{A,n}^2\right]. \tag{7}$$

Assumption 1. *There exist constants $0 < \phi_- \le \phi_+$ such that every $u = (u_1, \dots, u_P) \in \mathcal{U}$ satisfying $|\operatorname{supp}(u)| \le 2s$ obeys $\phi_- \sum_{p=1}^P \|u_p\|_2^2 \le \|Au\|_n^2 \le \phi_+ \sum_{p=1}^P \|u_p\|_2^2$.*

Assumption 1 is standard in minimax analyses of high-dimensional statistics, where they ensure that separation in the coefficient space is preserved in prediction space while preventing the induced distributions from becoming too easily distinguishable (Raskutti et al., 2011; Lounici et al., 2011). The lower constant ϕ_- controls the separation of the resulting margin vectors, whereas ϕ_+ controls their maximal separation and hence the Kullback–Leibler divergence used in the lower-bound argument.

Theorem 1 (Minimax lower bound)**.** *Suppose Assumption 1 holds and (without loss of generality) the representations are normalized such that* $\max_{i,j} \|h(Z_{ij})\|_2 \le \sqrt{d}$*. Let* $1 \le s \le P/2$*, and suppose that* sd *and* $s\lfloor \log_2 \lfloor P/s \rfloor \rfloor$ *are sufficiently large. Then there exists a universal constant* $c > 0$ *such that, for all sufficiently large* n*, the minimax risk in equation 7 satisfies*

$$\mathfrak{R}_n(\mathcal{G}_{s,M}) \ge c \frac{s[d + \log(P/s)]}{n}.$$

The minimax risk characterizes the intrinsic statistical difficulty of the problem: it is the smallest worst-case prediction error achievable by any estimator over the class $\mathcal{G}_{s,M}$. Hence, a minimax lower bound identifies an information-theoretic limit that no convex estimator based on the fixed activation patterns can uniformly improve upon. The rate in Theorem 1 has two natural components. The term sd/n is the within-pattern estimation cost: even if the active patterns were known, one must estimate a d-dimensional coefficient for each of the s active patterns. The term $s \log(P/s)/n$ is the pattern-selection cost, reflecting the combinatorial complexity of identifying s active patterns among P candidates, since $\log \binom{P}{s} \asymp s \log(P/s)$. Thus the lower bound combines the continuous complexity of estimating the active coefficients with the discrete complexity of selecting their support.

3.2 Estimation upper bound

Assumption 2 (Block spectral normalization)**.** *For every sampled activation pattern* $p \in [P]$*, assume* $\sup_{a \in \mathcal{U}_p, \|a\|_2 = 1} n^{-1/2} \|A_p a\|_2 \le 2$.

Assumption 3 (Restricted eigenvalue)**.** *Let* $\mathbb{C}(S) = \{a \in \mathcal{U} : \Gamma(a_{S^c}) \le 3\Gamma(a_S)\}$ *for the penalty in equation 4. Assume there exists* $\kappa > 0$ *such that* $\|Aa\|_n^2 \ge \kappa \sum_{p \in S} \gamma_p(a_p)^2$ *for all* $a \in \mathbb{C}(S)$.

Assumption 2 is a standard block-normalization condition commonly used in group-Lasso analysis (Negahban et al., 2012), while Assumption 3 is a standard restricted-eigenvalue condition used in high-dimensional regression (Bickel et al., 2009; Lounici et al., 2011).

Theorem 2 (Estimation upper bound)**.** *Suppose Assumptions 2 and 3 hold. For* $\delta \in (0, 1)$*, choose* $\lambda = 4\sqrt{2} n^{-1/2} \left[d \log 5 + \log(2P/\delta) \right]^{1/2}$*. With probability at least* $1 - \delta$*, the proposed estimator achieves*

$$\|\widehat{u} - u^\star\|_{A,n}^2 = \frac{1}{n} \|A(\widehat{u} - u^\star)\|_2^2 \le \frac{288 s [d \log 5 + \log(2P/\delta)]}{\mu_M^2 \kappa n}.$$

For fixed M and κ, Theorem 2 gives $\|\widehat{u} - u^\star\|_{A,n}^2 = \mathcal{O}_{\mathbb{P}}\left(n^{-1} s(d + \log P)\right)$. The bound has the same two sources of complexity as the minimax lower bound: estimating the coefficients within the active patterns contributes sd/n, while searching over the candidate activation patterns contributes $s \log P/n$. To interpret these complexity parameters in terms of the underlying neural network, we next relate s to the effective network width and P to the number of possible activation regimes.

Remark 1. *Let* $m_{\text{eff}}(u^\star)$ *denote the minimum number of hidden neurons whose aggregated signed pattern coefficients equal* $u^\star$*. Since each hidden neuron induces exactly one activation pattern, every active pattern requires at least one neuron, so* $s \le m_{\text{eff}}(u^\star)$*. Conversely, each nonzero quotient block* $u_p^\star$ *admits a cone decomposition* $u_p^\star = v_p^\star - w_p^\star$*, where each nonzero block* $v_p^\star$ *or* $w_p^\star$ *can be represented by a single balanced ReLU neuron. Hence at most two neurons are required for each active pattern, giving* $m_{\text{eff}}(u^\star) \le 2s$*. Therefore,* $s \le m_{\text{eff}}(u^\star) \le 2s$ *implies that the pattern sparsity and effective hidden-layer width are equivalent up to a universal constant factor.*

Remark 2. *The preceding results condition on the retained activation patterns. We now relate this conditional analysis to the full activation-pattern formulation. Let* $\mathcal{P}_{\text{full}}(X) = \{D_g : g \in \mathbb{R}^d\}$ *denote the collection of all activation patterns induced by the observed design* X*, with cardinality* P_{full}*. Let* A_{full} *and* $\mathcal{U}_{\text{full}}$ *denote the corresponding full-pattern design and parameter space, define*

$$\mathcal{G}_{s,M}^{\text{full}} = \{A_{\text{full}} u : u \in \mathcal{U}_{\text{full}}, |\operatorname{supp}(u)| \le s, \|A_{\text{full}} u\|_\infty \le M\}.$$

Suppose $\eta^\star \in \mathcal{G}_{s,M}^{\text{full}}$*, and fix an* s*-sparse representation of* $\eta^\star$ *with active pattern set* S_{full}*. For a pattern* $D_p \in \mathcal{P}_{\text{full}}(X)$*, let* $\pi_p = \mathbb{P}_g(D_g = D_p \mid X)$ *denote its probability under a Gaussian direction* $g \sim N(0, I_d)$*. Let* $\widehat{\mathcal{P}}$ *denote the set of* P *distinct activation patterns observed after Gaussian direction samplings, and let* $\widehat{S}_P = \{p \in [P_{\text{full}}] : D_p \in \widehat{\mathcal{P}}\}$*. If* $\pi_\star := \min_{p \in S_{\text{full}}} \pi_p > 0$*,*

then we have $\mathbb{P}(S_{\text{full}} \subseteq \widehat{S}_P \mid X) \geq 1 - se^{-P\pi_\star}$*, which tends to one as* P *increases. Moreover, if the full-pattern design satisfies Assumptions 2 and 3, the same conditions hold for any* $\widehat{S}_P$ *containing* S_{full}*. Therefore, on this event, the conditional upper bound applies with the same design constants as in the full-pattern problem. For the full activation-pattern collection, letting* $r_X = \operatorname{rank}(X)$*, the hyperplane-arrangement bound of Pilanci & Ergen (2020) gives* $P \leq P_{\text{full}} \leq 2\sum_{k=0}^{r_X-1}\binom{2n-1}{k} \leq 2r_X\left[\frac{e(2n-1)}{r_X}\right]^{r_X}$*. Hence,* $\log P_{\text{full}} = \mathcal{O}(r_X \log(n/r_X))$*, and Theorem 2 yields* $\|\widehat{u} - u^\star\|_{A,n}^2 = \mathcal{O}_{\mathbb{P}}\left(s[d + r_X \log(n/r_X)]/n\right)$*. In summary, although the full collection may contain combinatorially many activation patterns, its size enters the estimation error only through its logarithm. At fixed* X*, if every full pattern has positive Gaussian mass, repeated Gaussian sampling eventually recovers the full pattern collection almost surely.*

Previous analysis (and later Section 3.3) assumes that the true margins admit an exact s-sparse representation over the retained patterns. This model may fail in practice either because $\eta^\star \notin \mathcal{G}_{s,M}^{\text{full}}$, so that even the full activation-pattern class does not admit an exact s-sparse representation, or $\eta^\star \in \mathcal{G}_{s,M}^{\text{full}}$ but the retained collection omits patterns needed for such a representation. The following corollary relaxes this assumption and explicitly characterize the separation between approximation and statistical estimation errors.

Corollary 1 (Approximation–estimation tradeoff)**.** *Suppose the labels follow the bounded Bradley–Terry model with true margin vector satisfying* $\|\eta^\star\|_\infty \leq M$*, without assuming* $\eta^\star \in \mathcal{G}_{s,M}$*. Denote the model approximation error* $\mathcal{E}_{\text{app}} = \inf_{u \in \mathcal{F}_{s,M}} \|Au - \eta^\star\|_n^2 = \inf_{\eta \in \mathcal{G}_{s,M}} \|\eta - \eta^\star\|_n^2$*, and let* $u^\circ \in \arg\min_{u \in \mathcal{F}_{s,M}} \|Au - \eta^\star\|_n^2$ *with* $S^\circ = \operatorname{supp}(u^\circ)$*. Suppose instead that Assumption 2 and 3 hold with* $S = S^\circ$*. For* $\delta \in (0,1)$*, choose* λ *as in Theorem 2. Then, with probability at least* $1 - \delta$*,*

$$\|A\widehat{u} - \eta^\star\|_n^2 \leq \frac{1}{2\mu_M}\mathcal{E}_{\text{app}} + \frac{576s[d\log 5 + \log(2P/\delta)]}{\mu_M^2 \kappa n}.$$

Corollary 1 shows that the prediction error is controlled by the approximation error $\mathcal{E}_{\text{app}}$ together with the same statistical estimation term of order $s(d + \log P)/n$ as in Theorem 2. Under exact realizability that $\eta^\star \in \mathcal{G}_{s,M}$, we have $\mathcal{E}_{\text{app}} = 0$, thus recovering the previous rate in Theorem 2 up to constants. No separate optimization error term appears because the statistical analysis is stated for an exact minimizer of the convex program.

3.3 Minimax-optimal estimation upper bound

Comparing the statistical estimation upper bound in Theorem 2 with the minimax lower bound in Theorem 1, the only discrepancy is the pattern-selection term: $\log P$ versus $\log(P/s)$. One standard approach to attaining a minimax-optimal rate over an s-sparse parameter class is to employ an ℓ_0-type regularizer, or an explicit cardinality constraint on the coefficients (Raskutti et al., 2011). However, such a formulation leads to a combinatorial subset-selection problem and is NP-hard in general (Natarajan, 1995; Bertsimas et al., 2016). Nonconvex penalties such as MCP can achieve minimax convergence rates in certain high-dimensional linear-regression settings (Zhang, 2010). On the other hand, convex alternatives such as SLOPE can attain minimax rates in linear regression (Bellec et al., 2018), with related extensions to grouped predictors (Brzyski et al., 2019).

We take a different approach by keeping the original convex estimator in equation 5 unchanged and ask whether a sharper calibration and analysis can recover the minimax rate. Following the ordered-score argument developed for Lasso and Group Lasso by Dedieu (2019), we assume that the sparsity level s is known and impose a stronger restricted-eigenvalue condition.

For any $a = (a_1, \ldots, a_P) \in \mathcal{U}$, let $J_s(a) \subseteq [P]$ denote the indices of the s largest values among $\gamma_1(a_1), \ldots, \gamma_P(a_P)$, with ties resolved arbitrarily, and define $\Gamma_{2,s}(a) := [\sum_{p \in J_s(a)} \gamma_p(a_p)^2]^{1/2}$.

Assumption 4. *Define the cone* $\widetilde{\mathbb{C}}(J_s) := \left\{a \in \mathcal{U} : \Gamma\left(a_{J_s(a)^c}\right) \leq 2\Gamma\left(a_{J_s(a)}\right) + \sqrt{s}\,\Gamma_{2,s}(a)\right\}$*. Assume that there exists a constant* $\widetilde{\kappa} > 0$ *such that* $\|Aa\|_n^2 \geq \widetilde{\kappa}\,\Gamma_{2,s}(a)^2$ *for all* $a \in \widetilde{\mathbb{C}}(J_s)$*.*

Theorem 3 (Minimax-optimal estimation upper bound)**.** *Suppose Assumptions 2 and 4 hold. Assume further that the sparsity level* s *is known. For* $\delta \in (0,1)$*, choose* $\lambda = 8\sqrt{[d\log 5 + \log(6P/s\delta)]/n}$*. Then, with probability at least* $1 - \delta$*,*

$$\|\widehat{u} - u^\star\|_{A,n}^2 \leq \frac{576s[d\log 5 + \log(6P/s\delta)]}{\widetilde{\kappa}\mu_M^2 n}.$$

For fixed M and $\widetilde{\kappa}$, Theorem 3 gives $\|\widehat{u} - u^\star\|_{A,n}^2 = \mathcal{O}_{\mathbb{P}}\left(s[d + \log(P/s)]/n\right)$, which attains the minimax lower bound in Theorem 1 up to universal constants. The improvement over Theorem 2 comes from controlling the ordered block scores rather than their overall maximum. Knowing s allows the stochastic error to be controlled at the scale of the sth largest block score, replacing the $\log P$ term by $\log(P/s)$. The stronger restricted-eigenvalue condition then converts this sharper stochastic bound into prediction error.

4 Experiments

We evaluate the proposed method in three settings. First, fully synthetic experiments examine the dependence of the estimation error on the quantities in our theoretical rates. Second, semi-synthetic experiments use real LLM representations with synthetic oracle reward functions to compare reward-difference estimation. Finally, experiments on real preference data evaluate held-out preference prediction.

For the semi-synthetic and the real preference data application, we use final-layer representations from Qwen3-1.7B and compare our method with COALA (Feng & Pilanci, 2026), the sparse linear reward model of Yao et al. (2025), which has provable in-sample prediction guarantees, and a nonconvex two-layer ReLU baseline obtained by directly optimizing equation 2. The nonconvex baseline assesses whether the sampled convex formulation remains competitive with direct nonconvex neural-network optimization, for which comparable statistical guarantees are unavailable. We configure COALA and the sparse linear reward model following their respective papers and GitHub repositories. Appendix J provides the configurations of our method and the nonconvex baseline.

4.1 Synthetic Data

For each comparison $i \in [n]$, generate $z_i^{(0)}, \epsilon_{i1}, \epsilon_{i0} \overset{\text{i.i.d.}}{\sim} N(0, I_d)$ and set $h(Z_{ij}) = \sqrt{\rho}\, z_i^{(0)} + \sqrt{1-\rho}\,\epsilon_{ij}$, for $j \in \{0, 1\}$, where $\rho \in [0, 1)$ controls the within-pair correlation. Thus, $\mathrm{Cov}\{h(Z_{i1}), h(Z_{i0})\} = \rho I_d$, while embeddings from different comparison pairs are independent. We use $\rho = 0.5$ and globally rescale all embeddings to satisfy $\max_{i,j} \|h(Z_{ij})\|_2 \le \sqrt{d}$. The resulting embeddings are stacked to form the design matrix $X \in \mathbb{R}^{2n \times d}$. Given the resulting design X, we draw $g_p \sim N(0, I_d)$ and retain P distinct activation patterns $D_p = \mathrm{diag}\{\mathbf{1}(Xg_p \ge 0)\}$. We randomly choose $S^\star \subset [P]$ with $|S^\star| = s$ and independent $\xi_p \sim \mathrm{Rad}(\pm 1)$. Define $u_p^\star = a\xi_p g_p / \|g_p\|_2$, for $p \in S^\star$, and $u_p^\star = 0$, for $p \notin S^\star$. Since $g_p \in \mathcal{K}_p$, this construction satisfies $u_p^\star \in \mathcal{U}_p$. The scale a is calibrated so that the oracle margins $\eta^\star = Au^\star$ satisfy $\|\eta^\star\|_\infty \le M$. Labels are then sampled independently from $\mathbb{P}(Y_i = 1 \mid X) = \sigma(\eta_i^\star)$. We evaluate $\mathrm{MSE} = \frac{1}{n}\|A\widehat{u} - Au^\star\|_2^2$. The baseline configuration is $n = 1000$, $d = 10$, $P = 20$, and $s = 5$. We vary one parameter at a time over $n \in \{200, 400, \ldots, 2000\}, d \in \{10, 20, \ldots, 100\}, P \in \{20, 30, \ldots, 110\}, s \in \{1, 2, \ldots, 10\}$. For each configuration, the design matrix X, the activation patterns, and the oracle parameter $u^\star$ are held fixed, and the reported results average over 50 independent draws of the preference labels.

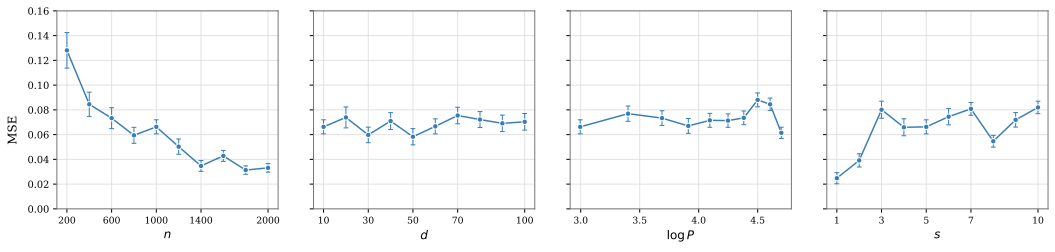


Figure 1: MSE of the estimated preference margins as a function of n, d, $\log P$, and s. Each parameter is varied separately while the others are held at their baseline values. Points and error bars denote the mean MSE and 95% Monte Carlo confidence intervals over 50 replications, respectively.

Figure 1 shows a clear decrease in MSE as n increases and an overall increase as $d, \log P$, and s grow, consistent with the qualitative dependence predicted by our theory. The dependence on n is particularly pronounced, whereas the effects of the complexity parameters are more moderate

over the ranges considered. These results provide empirical support for the main scaling behavior suggested by the theoretical bounds.

4.2 SEMI-SYNTHETIC LLM DATA

To evaluate the proposed method under more realistic representation geometry while retaining access to the ground truth, we conduct semi-synthetic experiments using the `arena-human-preference-55k` dataset, which contains 57,477 pairwise preference examples. For each prompt-response candidate in a comparison, we extract the hidden representation of the last Transformer layer of the pretrained Qwen3-1.7B model, yielding representations of dimension $d = 2048$. The extracted representations are globally rescaled according to the same normalization used in our theoretical analysis so that $\max_{i,j} \|h(Z_{ij})\|_2 \leq \sqrt{d}$. We consider four oracle reward-generating mechanisms with different degrees and forms of nonlinearity. For each oracle, the true pairwise margin is $\eta_i^\star = r^\star(Z_{i1}) - r^\star(Z_{i0})$, and preference labels are generated according to the Bradley–Terry model $\mathbb{P}(Y_i = 1 \mid Z_{i1}, Z_{i0}) = \sigma(\eta_i^\star)$. See detailed description of oracle reward generating processes in Appendix A.

We report three complementary evaluation metrics of pairwise reward-difference estimation. Let $\widehat{\eta}_i$ denote the estimated margin and $\eta_i^\star$ the corresponding oracle margin. The prediction error is $\mathrm{MSE} = \frac{1}{n}\sum_{i=1}^n (\widehat{\eta}_i - \eta_i^\star)^2$. Because our primary interest is in recovering the correct preference ordering instead of the exact magnitude of each reward difference, we additionally report the oracle direction accuracy, defined as the proportion of comparisons for which $\mathrm{sign}(\widehat{\eta}_i) = \mathrm{sign}(\eta_i^\star)$. We also report the area under the receiver operating characteristic curve (AUC), using $y_i^\star = \mathbf{1}\{\eta_i^\star > 0\}$ as the binary label and the continuous estimated margin $\widehat{\eta}_i$ as the classification score.

Table 1: Comparison of preference-learning methods under four oracle reward-generating mechanisms on `Arena` representations. Entries report mean ± sample standard deviation over 50 Monte Carlo replications. The best two performance for each DGP are shown in bold.

Oracle DGP	Method	MSE	Direction accuracy	AUC
Linear	Ours	0.0328 ± 0.0093	**78.29% ± 3.80%**	**0.8630 ± 0.0410**
	COALA	0.1091 ± 0.0074	66.31% ± 3.09%	0.7204 ± 0.0393
	Sparse Linear	**0.0303 ± 0.0106**	75.10% ± 3.56%	0.8276 ± 0.0393
	Nonconvex	**0.0237 ± 0.0074**	**79.13% ± 2.93%**	**0.8707 ± 0.0299**
Sparse ReLU	Ours	**0.0392 ± 0.0150**	**78.29% ± 4.92%**	**0.8615 ± 0.0549**
	COALA	0.1174 ± 0.0093	66.29% ± 4.50%	0.7186 ± 0.0579
	Sparse Linear	0.0412 ± 0.0152	73.40% ± 4.15%	0.8073 ± 0.0486
	Nonconvex	**0.0327 ± 0.0106**	**76.91% ± 4.95%**	**0.8451 ± 0.0593**
Interaction ReLU	Ours	**0.0620 ± 0.0154**	**73.99% ± 3.45%**	**0.8157 ± 0.0407**
	COALA	0.1366 ± 0.0116	65.24% ± 2.90%	0.7058 ± 0.0382
	Sparse Linear	0.0650 ± 0.0169	69.93% ± 3.13%	0.7663 ± 0.0388
	Nonconvex	**0.0551 ± 0.0137**	**73.30% ± 3.33%**	**0.8058 ± 0.0391**
Symmetric ReLU	Ours	**0.0555 ± 0.0172**	**76.53% ± 3.84%**	**0.8431 ± 0.0427**
	COALA	0.1303 ± 0.0122	66.65% ± 2.85%	0.7235 ± 0.0369
	Sparse Linear	0.0587 ± 0.0177	71.71% ± 4.45%	0.7868 ± 0.0532
	Nonconvex	**0.0477 ± 0.0153**	**75.63% ± 4.23%**	**0.8310 ± 0.0482**

Table 1 shows that our method achieves the highest mean direction accuracy and AUC under the three nonlinear oracles. It outperforms COALA and the sparse linear baseline on these ordering metrics under all four oracles, while also achieving lower mean MSE under the three nonlinear settings. The lower mean MSE of the nonconvex benchmark reflects the approximation error introduced by restricting our convex formulation to a finite collection of sampled activation patterns. Nevertheless, the two methods remain close in direction accuracy and AUC, supporting the practical effectiveness of our convex formulation for recovering preference orderings. The competitive performance of the sparse linear baseline may reflect its correctly specified model class under the linear oracle and the low-complexity structure of the nonlinear synthetic rewards. Since the geometry of real preference rewards is usually unknown in practice, we next compare the methods on real preference data.

4.3 REAL PREFERENCE DATA

Finally, we evaluate held-out preference prediction on five real datasets. See Appendix H for detailed datasets description. We use 30/70 train–test splits and the same Qwen3-1.7B representations and baseline configurations as in Section 4.2. Since oracle rewards are unknown, evaluation is based directly on the observed preference labels on the test set. Specifically, we compute preference accuracy, which is the proportion of comparisons satisfying $\mathrm{sign}(\widehat{\eta}_i) = Y_i$, and AUC, using $\mathbf{1}\{Y_i = 1\}$ as the binary label and $\widehat{\eta}_i$ as the score.

Table 2: Entries report mean $\pm$ sample standard deviation across the 50 train-test splits. The number in parentheses below each dataset name denotes the total number of valid preference pairs retained after excluding tied preferences (training and test combined). The best two performance for each dataset are shown in bold.

Dataset	Method	Accuracy (%)	AUC
`WebGPT` (14,346)	Ours	**60.73 ± 0.41**	**0.6521 ± 0.0039**
	COALA	58.32 ± 0.56	0.6188 ± 0.0061
	Sparse Linear	59.88 ± 0.47	0.6429 ± 0.0057
	Nonconvex	**62.75 ± 0.42**	**0.6801 ± 0.0042**
`Arena` (39,716)	Ours	**67.28 ± 0.25**	**0.7413 ± 0.0020**
	COALA	64.35 ± 0.34	0.7001 ± 0.0037
	Sparse Linear	66.40 ± 0.18	0.7299 ± 0.0021
	Nonconvex	**67.62 ± 0.19**	**0.7459 ± 0.0018**
`UltraFeedback` (55,476)	Ours	**74.16 ± 0.14**	**0.8281 ± 0.0012**
	COALA	72.33 ± 0.20	0.8054 ± 0.0018
	Sparse Linear	72.61 ± 0.14	0.8104 ± 0.0013
	Nonconvex	**74.01 ± 0.14**	**0.8265 ± 0.0012**
`HH-RLHF` (169,352)	Ours	**61.77 ± 0.14**	**0.6742 ± 0.0017**
	COALA	61.06 ± 0.14	0.6623 ± 0.0015
	Sparse Linear	60.02 ± 0.12	0.6498 ± 0.0014
	Nonconvex	**62.17 ± 0.12**	**0.6786 ± 0.0014**
`Summarize-from-Feedback` (178,944)	Ours	**65.75 ± 0.16**	**0.7205 ± 0.0018**
	COALA	64.11 ± 0.21	0.6980 ± 0.0026
	Sparse Linear	64.40 ± 0.14	0.7008 ± 0.0018
	Nonconvex	**65.84 ± 0.16**	**0.7211 ± 0.0018**

Table 2 shows that our method consistently outperforms COALA and the sparse linear baseline in both test accuracy and AUC across all five datasets. The nonconvex benchmark achieves higher means on four datasets, while our method performs best on `UltraFeedback`. This advantage may reflect a setting in which nonconvex optimization error outweighs the approximation error from restricting the convex model to sampled activation patterns. Our method remains close to the nonconvex benchmark on `Arena`, `HH-RLHF`, and `Summarize-from-Feedback`, supporting the practical effectiveness of the convex formulation. These results complement the semi-synthetic experiments by showing that nonlinear reward modeling also benefits real preference prediction.

5 CONCLUSION

We study nonlinear preference learning through a one-stage convex neural Bradley–Terry model and characterize its statistical complexity in terms of activation-pattern sparsity. For rewards supported on at most s patterns, we establish a minimax rate of order $s[d + \log(P/s)]/n$ and show that the proposed convex estimator attains this rate when s is known, while a standard calibration achieves the closely related rate $s(d + \log P)/n$. The experiments support the usefulness of the nonlinear reward class: our method performs strongly in controlled reward-recovery settings and real preference datasets. Natural directions for future work include extending the theory to out-of-sample prediction, and accommodating heterogeneous or misspecified human preferences.

AI USE STATEMENT

In this work, we used generative AI tools for providing critical ingredients for proving mathematical claims, assisting in the writing of proofs, implementing methods, assisting with translation, and cleaning and reformating dataset . We have not used generative AI tools for generating synthetic data sets, helping develop theoretical models or conceptual frameworks, formulating mathematical claims, proposing or refining hypotheses, designing or providing feedback on research methodology or experiments, supporting qualitative and thematic data analysis, or interpreting results. We have reviewed all AI-assisted work. LLM-generated code was verified and tested for correctness by 2 authors. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

REPRODUCIBILITY STATEMENT

An anonymized implementation of the proposed method and scripts for reproducing the experiments are provided in the supplementary material. All the theoretical proofs are provided in the Appendix. A complete description of all the datasets used in the experiments can be found in Appendix H.

REFERENCES

Francis Bach. Breaking the curse of dimensionality with convex neural networks. *Journal of Machine Learning Research*, 18(19):1–53, 2017.

Yuntao Bai, Andy Jones, Kamal Ndousse, Amanda Askell, Anna Chen, Nova DasSarma, Dawn Drain, Stanislav Fort, Deep Ganguli, Tom Henighan, et al. Training a helpful and harmless assistant with reinforcement learning from human feedback. *arXiv preprint arXiv:2204.05862*, 2022.

Pierre C Bellec, Guillaume Lecué, and Alexandre B Tsybakov. Slope meets lasso: improved oracle bounds and optimality. *The Annals of Statistics*, 46(6B):3603–3642, 2018.

Yoshua Bengio, Nicolas Roux, Pascal Vincent, Olivier Delalleau, and Patrice Marcotte. Convex neural networks. In Y. Weiss, B. Schölkopf, and J. Platt (eds.), *Advances in Neural Information Processing Systems*, volume 18. MIT Press, 2005. URL https://proceedings.neurips.cc/paper_files/paper/2005/file/0fc170ecbb8ff1afb2c6de48ea5343e7-Paper.pdf.

Dimitris Bertsimas, Angela King, and Rahul Mazumder. Best subset selection via a modern optimization lens. *The Annals of Statistics*, pp. 813–852, 2016.

Peter J. Bickel, Ya'acov Ritov, and Alexandre B. Tsybakov. Simultaneous analysis of Lasso and Dantzig selector. *The Annals of Statistics*, 37(4):1705 – 1732, 2009. doi: 10.1214/08-AOS620. URL https://doi.org/10.1214/08-AOS620.

Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, and Jonathan Eckstein. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning*, 3(1):1–122, 2011.

Ralph Allan Bradley and Milton E. Terry. Rank analysis of incomplete block designs: I. the method of paired comparisons. *Biometrika*, 39(3/4):324–345, 1952.

Damian Brzyski, Alexej Gossmann, Weijie Su, and Małgorzata Bogdan. Group slope–adaptive selection of groups of predictors. *Journal of the American Statistical Association*, 114(525):419–433, 2019.

Wei-Lin Chiang, Lianmin Zheng, Ying Sheng, Anastasios Nikolas Angelopoulos, Tianle Li, Dacheng Li, Hao Zhang, Banghua Zhu, Michael Jordan, Joseph E Gonzalez, et al. Chatbot arena: An open platform for evaluating llms by human preference. *arXiv preprint arXiv:2403.04132*, 2024.

Ganqu Cui, Lifan Yuan, Ning Ding, Guanming Yao, Bingxiang He, Wei Zhu, Yuan Ni, Guotong Xie, Ruobing Xie, Yankai Lin, et al. Ultrafeedback: Boosting language models with scaled ai feedback. *arXiv preprint arXiv:2310.01377*, 2023.

Antoine Dedieu. Improved error rates for sparse (group) learning with lipschitz loss functions. *arXiv preprint arXiv:1910.08880*, 2019.

Miria Feng and Mert Pilanci. Convex optimization for alignment and preference learning on a single GPU. In *Proceedings of the 43rd International Conference on Machine Learning*, Proceedings of Machine Learning Research. PMLR, 2026.

Miria Feng, Zachary Frangella, and Mert Pilanci. CRONOS: Enhancing deep learning with scalable GPU-accelerated convex neural networks. In *Advances in Neural Information Processing Systems*, volume 37, 2024.

Zachary Frangella, Joel A. Tropp, and Madeleine Udell. Randomized nyström preconditioning. *SIAM Journal on Matrix Analysis and Applications*, 44(2):718–752, 2023. doi: 10.1137/21M1466244.

Seong Jin Lee, Will Wei Sun, and Yufeng Liu. Low-rank contextual reinforcement learning from heterogeneous human feedback. *Journal of the American Statistical Association*, 121(554):1037–1050, 2026.

Zihao Li, Xiang Ji, Minshuo Chen, and Mengdi Wang. Policy evaluation for reinforcement learning from human feedback: A sample complexity analysis. In Sanjoy Dasgupta, Stephan Mandt, and Yingzhen Li (eds.), *Proceedings of The 27th International Conference on Artificial Intelligence and Statistics*, volume 238 of *Proceedings of Machine Learning Research*, pp. 2737–2745. PMLR, 02–04 May 2024. URL https://proceedings.mlr.press/v238/li24l.html.

Pangpang Liu, Junwei Lu, and Will Wei Sun. Uncertainty quantification for large language model reward learning under heterogeneous human feedback. *arXiv preprint arXiv:2512.03208*, 2025.

Karim Lounici, Massimiliano Pontil, Sara van de Geer, and Alexandre B. Tsybakov. Oracle inequalities and optimal inference under group sparsity. *The Annals of Statistics*, 39(4):2164 – 2204, 2011. doi: 10.1214/11-AOS896. URL https://doi.org/10.1214/11-AOS896.

Reiichiro Nakano, Jacob Hilton, Suchir Balaji, Jeff Wu, Long Ouyang, Christina Kim, Christopher Hesse, Shantanu Jain, Vineet Kosaraju, William Saunders, et al. Webgpt: Browser-assisted question-answering with human feedback. *arXiv preprint arXiv:2112.09332*, 2021.

Balas Kausik Natarajan. Sparse approximate solutions to linear systems. *SIAM Journal on Computing*, 24(2):227–234, 1995.

Sahand N. Negahban, Pradeep Ravikumar, Martin J. Wainwright, and Bin Yu. A Unified Framework for High-Dimensional Analysis of M-Estimators with Decomposable Regularizers. *Statistical Science*, 27(4):538 – 557, 2012. doi: 10.1214/12-STS400. URL https://doi.org/10.1214/12-STS400.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul F Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (eds.), *Advances in Neural Information Processing Systems*, volume 35, pp. 27730–27744. Curran Associates, Inc., 2022. doi: 10.52202/068431-2011. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/b1efde53be364a73914f58805a001731-Paper-Conference.pdf.

Mert Pilanci and Tolga Ergen. Neural networks are convex regularizers: Exact polynomial-time convex optimization formulations for two-layer networks. In *Proceedings of the 37th International Conference on Machine Learning*, pp. 7695–7705. PMLR, 2020.

Garvesh Raskutti, Martin J Wainwright, and Bin Yu. Minimax rates of estimation for high-dimensional linear regression over ℓ_q-balls. *IEEE Transactions on Information Theory*, 57(10): 6976–6994, 2011.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F Christiano. Learning to summarize with human feedback. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (eds.), *Advances in Neural Information Processing Systems*, volume 33, pp. 3008–3021. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/1f89885d556929e98d3ef9b86448f951-Paper.pdf.

Hao Sun, Yunyi Shen, and Jean-Francois Ton. Rethinking reward modeling in preference-based large language model alignment. In Y. Yue, A. Garg, N. Peng, F. Sha, and R. Yu (eds.), *International Conference on Learning Representations*, volume 2025, pp. 46885–46922, 2025. URL https://proceedings.iclr.cc/paper_files/paper/2025/file/7423902b5534e2b267438c85444a54b1-Paper-Conference.pdf.

Lewis Tunstall, Edward Beeching, Nathan Lambert, Nazneen Rajani, Kashif Rasul, Younes Belkada, Shengyi Huang, Leandro Von Werra, Clémentine Fourrier, Nathan Habib, et al. Zephyr: Direct distillation of lm alignment. *arXiv preprint arXiv:2310.16944*, 2023.

Roman Vershynin. *High-dimensional probability: An introduction with applications in data science*, volume 47. Cambridge University Press, 2018.

Yunzhen Yao, Lie He, and Michael Gastpar. Leveraging sparsity for sample-efficient preference learning: A theoretical perspective. In *Proceedings of the 42nd International Conference on Machine Learning*, volume 267 of *Proceedings of Machine Learning Research*, pp. 71739–71764. PMLR, 2025. URL https://proceedings.mlr.press/v267/yao25c.html.

Ming Yuan and Yi Lin. Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 68(1):49–67, 2006.

Cun-Hui Zhang. Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894 – 942, 2010. doi: 10.1214/09-AOS729. URL https://doi.org/10.1214/09-AOS729.

Banghua Zhu, Michael Jordan, and Jiantao Jiao. Principled reinforcement learning with human feedback from pairwise or k-wise comparisons. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett (eds.), *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pp. 43037–43067. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/zhu23f.html.

A Oracle reward generating processes for semi-synthetic experiments in Section 4.2

We present four oracle reward generating processes used in Section 4.2. For each oracle, the true pairwise margin is $\eta_i^\star = r^\star(Z_{i1}) - r^\star(Z_{i0})$, and preference labels are generated according to the Bradley–Terry model $\mathbb{P}(Y_i = 1 \mid Z_{i1}, Z_{i0}) = \sigma(\eta_i^\star)$. The overall reward scale a is calibrated using the same bounded-margin construction as in Section 4.1. Specifically, we consider:

- **Linear reward.** We draw $g \sim N(0, I_d)$ and define
$$r^\star_{\mathrm{lin}}(z) = a\frac{g^\top h(z)}{\|g\|_2}.$$
This setting provides a benchmark that the linear reward model is correctly specified.
- **Sparse ReLU reward.** We use the same sparse two-layer ReLU reward as in Section 4.1,
$$r^\star_{\mathrm{sReLU}}(z) = a\sum_{p\in S^\star}\xi_p\left(\frac{g_p^\top h(z)}{\|g_p\|_2}\right)_+,$$

where $|S^\star| = s = 5$ and $\xi_p \overset{\text{i.i.d.}}{\sim} \mathrm{Rad}(\pm 1)$. This oracle is nonlinear but consists of independently oriented ReLU ridge functions and generally retains a nonzero linear component.

- **Interaction ReLU reward.** To generate a more complex nonlinear reward, we consider

$$r^\star_{\mathrm{int}}(z) = \frac{a}{\sqrt{K}} \sum_{k=1}^{K} \xi_k \left[\left(\frac{g_{k1}^\top h(z)}{\|g_{k1}\|_2} \right)_+ + \left(\frac{g_{k2}^\top h(z)}{\|g_{k2}\|_2} \right)_+ - \left(\frac{g_{k1}^\top h(z)}{\|g_{k1}\|_2} + \frac{g_{k2}^\top h(z)}{\|g_{k2}\|_2} \right)_+ \right],$$

where $g_{k1}, g_{k2} \overset{\text{i.i.d}}{\sim} N(0, I_d), \xi_k \overset{\text{i.i.d}}{\sim} \mathrm{Rad}(\pm 1)$, and $K = 6$. Since $(t)_+ = (t + |t|)/2$, the linear components of the three ReLU units cancel exactly within each interaction block, producing a reward that is more difficult to approximate by a global linear function.

- **Symmetric ReLU reward.** We independently draw $g_k \sim N(0, I_d)$, and set

$$r^\star_{\mathrm{sym}}(z) = \frac{a}{\sqrt{s/2}} \sum_{k=1}^{s/2} \xi_k \left[\left(\frac{g_k^\top h(z)}{\|g_k\|_2} \right)_+ + \left(-\frac{g_k^\top h(z)}{\|g_k\|_2} \right)_+ \right] = \frac{a}{\sqrt{s/2}} \sum_{k=1}^{s/2} \xi_k \left| \frac{g_k^\top h(z)}{\|g_k\|_2} \right|,$$

where $\xi_k \overset{\text{i.i.d.}}{\sim} \mathrm{Rad}(\pm 1)$ and $s = 6$. This construction uses three symmetric ReLU pairs and removes the linear component through paired opposite activation directions, without introducing the cross-direction interactions present in the interaction oracle.

B Convex reformulation and additional notation

B.1 Original neural objective and weight-decay equivalence

The following argument specializes the convex reformulation of Pilanci & Ergen (2020) to the original nonconvex problem equation 2.

Positive homogeneity of ReLU implies that replacing $(\theta_{1j}, \theta_{2j})$ by $(c\theta_{1j}, \theta_{2j}/c)$, for $c > 0$, leaves the neuron unchanged. Optimizing the weight-decay cost over this rescaling gives

$$\inf_{c>0} \frac{1}{2} \left(c^2 \|\theta_{1j}\|_2^2 + \frac{\theta_{2j}^2}{c^2} \right) = |\theta_{2j}| \, \|\theta_{1j}\|_2.$$

For a nonzero neuron, write $s_j = \mathrm{sign}(\theta_{2j})$ and $\widetilde{\theta}_j = |\theta_{2j}|\theta_{1j}$. Its output is $s_j(\widetilde{\theta}_j^\top h(z))_+$ and its balanced penalty is $\|\widetilde{\theta}_j\|_2$.

For each retained pattern, let I_p^+ and I_p^- collect neurons with that training activation pattern and positive and negative output signs, respectively. Set $v_p = \sum_{j \in I_p^+} \widetilde{\theta}_j$ and $w_p = \sum_{j \in I_p^-} \widetilde{\theta}_j$. Since $\mathcal{K}_p$ is a convex cone, $v_p, w_p \in \mathcal{K}_p$. On the training design,

$$\sum_{j=1}^{m} \theta_{2j}(X\theta_{1j})_+ = \sum_{p=1}^{P} D_p X(v_p - w_p), \qquad \sum_{j=1}^{m} |\theta_{2j}| \|\theta_{1j}\|_2 \geq \sum_{p=1}^{P} (\|v_p\|_2 + \|w_p\|_2).$$

Applying B preserves all training comparison margins, including their box constraint.

Conversely, a nonzero $v_p \in \mathcal{K}_p$ is represented by the balanced neuron $\theta_1 = v_p/\sqrt{\|v_p\|_2}$ and $\theta_2 = \sqrt{\|v_p\|_2}$. A nonzero w_p is represented by $\theta_1 = w_p/\sqrt{\|w_p\|_2}$ and $\theta_2 = -\sqrt{\|w_p\|_2}$. Their weight-decay costs are exactly $\|v_p\|_2$ and $\|w_p\|_2$. Thus a cone-decomposed solution can be realized using at most $2P$ neurons, or fewer when blocks vanish. With sufficient width, the cone program is equivalent to the neural problem restricted to the retained activation patterns. Equivalence to the unrestricted neural problem requires the full pattern collection, or an additional argument showing that the retained collection is sufficient; no such sampling guarantee is used in the present statistical analysis.

B.2 The induced pattern norm

Proposition 1 (Induced pattern norm)**.** *The function γ_p in equation 4 is a norm on the linear subspace $\mathcal{U}_p = \mathcal{K}_p - \mathcal{K}_p$. The defining infimum is attained, and $\gamma_p(u_p) \geq \|u_p\|_2$ for every $u_p \in \mathcal{U}_p$.*

Proof. Because $\mathcal{K}_p$ is a convex cone, its difference set is closed under addition and real scalar multiplication, hence is a linear subspace. For a fixed $u_p \in \mathcal{U}_p$, the feasible set of decompositions is nonempty and closed. The objective $\|v\|_2 + \|w\|_2$ is coercive on this set, so a minimum is attained. Positive homogeneity follows by rescaling a feasible decomposition; symmetry follows by swapping v and w. Adding two feasible decompositions proves the triangle inequality. Finally, $\|u_p\|_2 = \|v - w\|_2 \le \|v\|_2 + \|w\|_2$ for every feasible decomposition. Taking the infimum gives the stated bound and definiteness. □

For fixed u, the loss and fitted margins in equation 3 depend on its decomposition only through $Fq = Au$. Minimizing the penalty separately over the decompositions of each u_p therefore gives $\Gamma(u)$ exactly. Conversely, the attained blockwise minima supply a feasible cone decomposition of any u. This establishes equivalence of equation 3 and equation 5 in optimal value and fitted training margins.

C Details of the matrix-free ADMM implementation

C.1 Augmented Lagrangian

For the split formulation equation 6, the scaled augmented Lagrangian is

$$\mathcal{L}_{\text{aug}} = \frac{1}{n}\sum_{i=1}^{n} \ell(Y_i\eta_i) + \mathcal{I}_{\mathcal{B}_M}(\eta) + \lambda\|z\|_{2,1} + \mathcal{I}_{\mathbb{R}_+^{4nP}}(\xi) + \frac{\rho_\eta}{2}\|\eta - Fq + \alpha\|_2^2 - \frac{\rho_\eta}{2}\|\alpha\|_2^2 + \frac{\rho_z}{2}\|q - z + \tau\|_2^2 - \frac{\rho_z}{2}\|\tau\|_2^2 + \frac{\rho_\xi}{2}\|Cq - \xi + \nu\|_2^2 - \frac{\rho_\xi}{2}\|\nu\|_2^2.$$

Each iteration minimizes this expression first over q and then jointly over (η, z, ξ), followed by the scaled dual updates in Algorithm 1. The three auxiliary updates are separable and can be computed in parallel.

C.2 Quadratic update and matrix-free implementation

Minimization over q gives the linear system $Hq^{k+1} = b^k$, with H and b^k defined in Algorithm 1. Since $H \succeq \rho_z I$ and $\rho_z > 0$, H is symmetric positive definite. We solve this system by PCG, which only requires matrix-vector products of the form

$$Hx = \rho_\eta F^\top(Fx) + \rho_z x + \rho_\xi C^\top(Cx).$$

Following CRONOS (Feng et al., 2024), these products are evaluated without explicitly forming F, C, or H. The activation-pattern matrices are applied through elementwise masking of products with X and $X^\top$. In our pairwise setting, B and $B^\top$ are implemented by pairwise differences and signed scattering, respectively.

Preconditioning. For fixed ADMM penalties, we construct a randomized Nyström approximation to

$$H_0 := \rho_\eta F^\top F + \rho_\xi C^\top C \approx U\Lambda U^\top,$$

where U has orthonormal columns and $\Lambda = \operatorname{diag}(\lambda_j)$ is nonnegative. This gives the preconditioner $P_{\text{Nys}} = \rho_z I + U\Lambda U^\top$, whose inverse action is

$$P_{\text{Nys}}^{-1} r = \frac{1}{\rho_z}\left[r - U\operatorname{diag}\left(\frac{\lambda_j}{\rho_z + \lambda_j}\right)U^\top r\right].$$

PCG is warm-started from q^k and run to the prescribed inner tolerance. The Nyström approximation is used only for preconditioning, and the PCG iterations continue to apply the original operator H.

C.3 Bounded logistic and proximal updates

For $c_i = (Fq^{k+1})_i - \alpha_i^k$, the margin update minimizes $n^{-1}\ell(Y_i\eta) + \rho_\eta(\eta - c_i)^2/2$ over $|\eta| \le M$. Set $t = Y_i\eta$ and $a_i = Y_i c_i$. The unconstrained minimizer $\widetilde{t}_i$ is the unique root of

$$g_i(t) = \rho_\eta(t - a_i) - \frac{1}{n(1 + e^t)} = 0, \qquad g_i'(t) = \rho_\eta + \frac{1}{n}\sigma(t)\sigma(-t) > 0.$$

The root lies in $[a_i, a_i + 1/(n\rho_\eta)]$ and can be found by safeguarded Newton iterations or bisection. The constrained update is

$$\eta_i^{k+1} = Y_i \,\mathrm{clip}(\widetilde{t}_i, -M, M).$$

where $\mathrm{clip}(x, a, b) = \min\{b, \max\{a, x\}\}$.

For $g \in [2P]$ and $a_g = q_g^{k+1} + \tau_g^k$, the group update is

$$z_g^{k+1} = \left(\mathcal{S}_{\lambda/\rho_z}(a)\right)_g = \left(1 - \frac{\lambda}{\rho_z \|a_g\|_2}\right)_+ a_g,$$

with zero output when $a_g = 0$. This is the standard block soft-thresholding operator associated with the group-lasso penalty (Yuan & Lin, 2006). The nonnegative-orthant projection gives $\xi^{k+1} = (Cq^{k+1} + \nu^k)_+$. The scaled dual updates are given in Algorithm 1.

C.4 Output and computational cost

Iterations terminate when the prescribed primal, dual, and feasibility tolerances are met. For the returned $\widehat{q} = (\widehat{v}_1, \ldots, \widehat{v}_P, \widehat{w}_1, \ldots, \widehat{w}_P)$, set $\widehat{u}_p = \widehat{v}_p - \widehat{w}_p$ and evaluate the recovered reward as

$$\widehat{r}(z) = \sum_{p=1}^{P} \left\{ \left(h(z)^\top \widehat{v}_p\right)_+ - \left(h(z)^\top \widehat{w}_p\right)_+ \right\}, \qquad z \in \mathcal{Z}.$$

Each application of H costs $\mathcal{O}(ndP)$. Thus J PCG iterations in a quadratic update require $\mathcal{O}(JndP)$ work in operator products, in addition to preconditioner application. Given Fq and Cq, the bounded logistic update costs $\mathcal{O}(n)$ per scalar iteration, while the group and orthant projections cost $\mathcal{O}(dP)$ and $\mathcal{O}(nP)$, respectively. Matrix-free computation avoids explicit storage of F, C, and H, although the cone variables ξ and ν still require $\mathcal{O}(nP)$ storage.

D Proof of Theorem 1

D.1 Technical lemmas

We first state the three auxiliary lemmas used in the proof. For $\eta \in \mathbb{R}^n$, let $\mathbb{P}_\eta$ denote the joint law of independent labels satisfying $\mathbb{P}_\eta(Y_i = y) = \sigma(y\eta_i)$ for $y \in \{-1, +1\}$. The proof for Lemma 2 and 3 are omitted, as both results are standard in the literature

Lemma 1 (Kullback–Leibler upper bound)**.** *For every $\eta, \eta' \in \mathbb{R}^n$,*

$$D_{\mathrm{KL}}(\mathbb{P}_\eta \,\|\, \mathbb{P}_{\eta'}) \le \frac{1}{8}\|\eta - \eta'\|_2^2 = \frac{n}{8}\|\eta - \eta'\|_n^2.$$

Proof. Let $b(t) := \log(1 + e^t)$. For $y \in \{-1, +1\}$, the distribution of one label has the exponential-family representation

$$\mathbb{P}_{\eta_i}(Y_i = y) = \sigma(y\eta_i) = \exp\left\{\mathbf{1}_{\{y=1\}}\eta_i - b(\eta_i)\right\}.$$

Indeed, the right-hand side equals $e^{\eta_i}/(1 + e^{\eta_i}) = \sigma(\eta_i)$ when $y = 1$ and equals $1/(1 + e^{\eta_i}) = \sigma(-\eta_i)$ when $y = -1$. Therefore,

$$\begin{aligned}
D_{\mathrm{KL}}(\mathbb{P}_{\eta_i} \,\|\, \mathbb{P}_{\eta_i'}) &= \mathbb{E}_{\eta_i}\left[\log \frac{\mathbb{P}_{\eta_i}(Y_i)}{\mathbb{P}_{\eta_i'}(Y_i)}\right] \\
&= \mathbb{E}_{\eta_i}\left[\mathbf{1}_{\{Y_i=1\}}(\eta_i - \eta_i') + b(\eta_i') - b(\eta_i)\right] \\
&= \mathbb{P}_{\eta_i}(Y_i = 1)(\eta_i - \eta_i') + b(\eta_i') - b(\eta_i) \\
&= b'(\eta_i)(\eta_i - \eta_i') + b(\eta_i') - b(\eta_i) \\
&= b(\eta_i') - b(\eta_i) - b'(\eta_i)(\eta_i' - \eta_i) \\
&= \frac{1}{2}b''(\xi_i)(\eta_i' - \eta_i)^2 \\
&\le \frac{1}{8}(\eta_i' - \eta_i)^2,
\end{aligned}$$

where ξ_i is some point between η_i and η_i', and the last inequality holds since $b''(t) = \frac{e^t}{(1+e^t)^2} = \sigma(t)\{1-\sigma(t)\} \le \frac{1}{4}$. Finally, since the labels are independent,

$$D_{\mathrm{KL}}(\mathbb{P}_\eta \,\|\, \mathbb{P}_{\eta'}) = \mathbb{E}_\eta\left[\log \prod_{i=1}^n \frac{\mathbb{P}_{\eta_i}(Y_i)}{\mathbb{P}_{\eta_i'}(Y_i)}\right] = \sum_{i=1}^n D_{\mathrm{KL}}(\mathbb{P}_{\eta_i} \,\|\, \mathbb{P}_{\eta_i'})$$
$$\le \frac{1}{8}\sum_{i=1}^n (\eta_i' - \eta_i)^2 = \frac{1}{8}\|\eta - \eta'\|_2^2 = \frac{n}{8}\|\eta - \eta'\|_n^2,$$

which proves the claim. □

Lemma 2 (Fano's inequality)**.** *Let $\eta^{(1)}, \dots, \eta^{(N)} \in \mathcal{G}_{s,M}$, where $N \ge 2$, and suppose that a nonnegative loss d satisfies $d(\eta, \eta') \le C\{d(\eta, \eta'') + d(\eta'', \eta')\}$ for all η, η', η'' and some constant $C > 0$. If*

$$\min_{j \ne k} d(\eta^{(j)}, \eta^{(k)}) \ge \alpha, \qquad \max_{j \ne k} D_{\mathrm{KL}}(\mathbb{P}_{\eta^{(j)}} \,\|\, \mathbb{P}_{\eta^{(k)}}) \le \beta,$$

then

$$\inf_{\widehat{\eta}} \sup_{\eta^\star \in \mathcal{G}_{s,M}} \mathbb{E}_{\eta^\star}\left[d(\widehat{\eta}, \eta^\star)\right] \ge \frac{\alpha}{2C}\left(1 - \frac{\beta + \log 2}{\log N}\right).$$

In particular, $d(\eta, \eta') = \|\eta - \eta'\|_n^2$ satisfies the required inequality with $C = 2$.

Lemma 3 (Varshamov–Gilbert packing)**.** *For every integer $m \ge 8$, there exists a subset $\Omega = \{\omega^{(1)}, \dots, \omega^{(N)}\} \subseteq \{0,1\}^m$ such that $N \ge 2^{m/8}$ and $H(\omega^{(j)}, \omega^{(k)}) \ge m/8$ whenever $j \ne k$, where H denotes the Hamming distance.*

D.2 Proof of the main theorem

Proof. We construct two finite subsets of $\mathcal{G}_{s,M}$ and apply Lemma 2 to each of them.

Part 1. Packing over coefficient directions. Fix $S_0 \subseteq [P]$ with $|S_0| = s$. By Lemma 3, there exists a set $\Omega_1 \subseteq \{0,1\}^{sd}$ of cardinality $N_1 \ge 2^{sd/8}$ such that $H(w, w') \ge sd/8$ for distinct $w, w' \in \Omega_1$. Write $w = \{w_{pk} : p \in S_0,\ k \in [d]\}$. Let $\{\epsilon_{pk} : p \in S_0,\ k \in [d]\}$ be independent Rademacher random variables, and let $e_1, \dots, e_d$ be an orthonormal basis of $\mathbb{R}^d$. For fixed (w, ϵ), define

$$u_p^{(w,\epsilon)} := \begin{cases} a\sum_{k=1}^d \epsilon_{pk}(2w_{pk} - 1)e_k, & p \in S_0, \\ 0, & p \notin S_0. \end{cases}$$

By construction, $\mathrm{supp}(u^{(w,\epsilon)}) = S_0$ and $|\mathrm{supp}(u^{(w,\epsilon)})| = s$. For every $i \in [n]$ and fixed w,

$$\left(Au^{(w,\epsilon)}\right)_i = \left(\sum_{p=1}^P A_p u_p^{(w,\epsilon)}\right)_i$$
$$= \sum_{p \in S_0} (A_p)_{i\cdot} u_p^{(w,\epsilon)}$$
$$= a\sum_{p \in S_0}\sum_{k=1}^d (A_p)_{ik}(2w_{pk} - 1)\epsilon_{pk}.$$

For fixed w, each $2w_{pk} - 1$ is a deterministic element of $\{-1, +1\}$, so the variables $(2w_{pk} - 1)\epsilon_{pk}$ are again independent Rademacher random variables. Recall that the ith row of $A_p = BD_pX$ is

$$(A_p)_{i\cdot} = d_{p,2i-1}h(Z_{i1})^\top - d_{p,2i}h(Z_{i0})^\top.$$

Since $d_{p,j} \in \{0,1\}$ and $\|h(Z_{ij})\|_2 \le \sqrt{d}$, the triangle inequality gives

$$\|(A_p)_{i\cdot}\|_2 \le d_{p,2i-1}\|h(Z_{i1})\|_2 + d_{p,2i}\|h(Z_{i0})\|_2 \le 2\sqrt{d}.$$

Consequently, the sum of the squared coefficients in the preceding Rademacher sum satisfies

$$a^2 \sum_{p \in S_0}\sum_{k=1}^d (A_p)_{ik}^2 = a^2 \sum_{p \in S_0} \|(A_p)_{i\cdot}\|_2^2 \le a^2 \sum_{p \in S_0} 4d = 4sa^2 d.$$

The two-sided Hoeffding inequality for a Rademacher sum therefore gives

$$\mathbb{P}_\epsilon\left(\left|\left(Au^{(w,\epsilon)}\right)_i\right| > M\right) \le 2\exp\left\{-\frac{M^2}{2a^2\sum_{p\in S_0}\sum_{k=1}^{d}(A_p)_{ik}^2}\right\} \le 2\exp\left(-\frac{M^2}{8sa^2d}\right).$$

Taking a union bound over $i \in [n]$ yields

$$\mathbb{P}_\epsilon\left(\|Au^{(w,\epsilon)}\|_\infty > M\right) \le 2n\exp\left(-\frac{M^2}{8sa^2d}\right).$$

If

$$a^2 \le \frac{M^2}{8sd\log(8n)}, \tag{8}$$

then $\|Au^{(w,\epsilon)}\|_\infty \le M$ holds with probability at least $3/4$ for every $w \in \Omega_1$. Define $I_w(\epsilon) := \mathbf{1}_{\{\|Au^{(w,\epsilon)}\|_\infty \le M\}}$. By linearity of expectation,

$$\mathbb{E}_\epsilon\left[\sum_{w\in\Omega_1} I_w(\epsilon)\right] = \sum_{w\in\Omega_1}\mathbb{E}_\epsilon[I_w(\epsilon)] = \sum_{w\in\Omega_1}\mathbb{P}_\epsilon\left(\|Au^{(w,\epsilon)}\|_\infty \le M\right) \ge \frac{3N_1}{4}.$$

Then, there exists a realization $\epsilon^\star$ satisfying $\sum_{w\in\Omega_1} I_w(\epsilon^\star) \ge 3N_1/4$. Let

$$\widetilde{\Omega}_1 := \left\{w \in \Omega_1 : \|Au^{(w,\epsilon^\star)}\|_\infty \le M\right\}.$$

Then $|\widetilde{\Omega}_1| \ge 3N_1/4$. Fix $\epsilon^\star$ and write $u^{(w)} := u^{(w,\epsilon^\star)}$ for $w \in \widetilde{\Omega}_1$. So, the packing is deterministic and satisfies $\|Au^{(\omega)}\|_\infty \le M$ for each $\omega \in \widetilde{\Omega}_1$. Thus, the candidates $u^{(\omega)} \in \mathcal{F}_{s,M}$ and $\eta^{(\omega)} \in \mathcal{G}_{s,M}$. Moreover,

$$|\widetilde{\Omega}_1| \ge \frac{3}{4}N_1 \ge \frac{3}{4}\times 2^{sd/8}.$$

For sufficiently large sd, the right-hand side is at least $2^{sd/16}$ and $2^{sd/16} \ge 4$. Relabeling the retained cardinality as N, we thus have $N \ge 2^{sd/16} \ge 4$. Because $\widetilde{\Omega}_1 \subseteq \Omega_1$, the pairwise Hamming separation $H(w, w') \ge sd/8$ is preserved.

Fix distinct $w, w' \in \widetilde{\Omega}_1$. For $p \in S_0$,

$$u_p^{(w)} - u_p^{(w')} = a\sum_{k=1}^{d}\epsilon_{pk}^\star\{(2w_{pk}-1)-(2w'_{pk}-1)\}e_k = a\sum_{k=1}^{d}\epsilon_{pk}^\star(2w_{pk}-2w'_{pk})e_k.$$

Since $e_1, \dots, e_d$ are orthonormal, $(\epsilon_{pk}^\star)^2 = 1$, and $(w_{pk}-w'_{pk})^2 = \mathbf{1}_{\{w_{pk}\neq w'_{pk}\}}$, we obtain

$$\begin{aligned}\sum_{p=1}^{P}\|u_p^{(w)} - u_p^{(w')}\|_2^2 &= \sum_{p\in S_0}\left\|a\sum_{k=1}^{d}\epsilon_{pk}^\star(2w_{pk}-2w'_{pk})e_k\right\|_2^2 \\ &= a^2\sum_{p\in S_0}\sum_{k=1}^{d}(2w_{pk}-2w'_{pk})^2 \\ &= 4a^2\sum_{p\in S_0}\sum_{k=1}^{d}\mathbf{1}_{\{w_{pk}\neq w'_{pk}\}} \\ &= 4a^2H(w,w').\end{aligned}$$

The lower Hamming bound and the trivial upper bound $H(w, w') \le sd$ therefore give

$$\frac{a^2sd}{2} = 4a^2\frac{sd}{8} \le \sum_{p=1}^{P}\|u_p^{(w)} - u_p^{(w')}\|_2^2 \le 4a^2sd.$$

Since all candidates in this construction have the common support S_0, we have $\operatorname{supp}\{u^{(w)} - u^{(w')}\} \subseteq S_0$ and hence $|\operatorname{supp}\{u^{(w)} - u^{(w')}\}| \le s \le 2s$. Applying the lower and upper inequalities in Assumption 1 to $u^{(w)} - u^{(w')}$ yields

$$\|\eta^{(w)} - \eta^{(w')}\|_n^2 = \|A(u^{(w)} - u^{(w')})\|_n^2 \ge \phi_- \sum_{p=1}^{P} \|u_p^{(w)} - u_p^{(w')}\|_2^2 \ge \frac{\phi_- a^2 sd}{2},$$

$$\|\eta^{(w)} - \eta^{(w')}\|_n^2 \le \phi_+ \sum_{p=1}^{P} \|u_p^{(w)} - u_p^{(w')}\|_2^2 \le 4\phi_+ a^2 sd.$$

Lemma 1 then implies

$$D_{\mathrm{KL}}(\mathbb{P}_{\eta^{(w)}} \,\|\, \mathbb{P}_{\eta^{(w')}}) \le \frac{n}{8}\|\eta^{(w)} - \eta^{(w')}\|_n^2 \le \frac{n\phi_+ a^2 sd}{2}.$$

Set $\alpha_1 := \phi_- a^2 sd/2$ and $\beta_1 := n\phi_+ a^2 sd/2$. Since $\log N \ge sd \log 2/16$,

$$\frac{\beta_1}{\log N} \le \frac{n\phi_+ a^2 sd/2}{sd \log 2/16} = \frac{8n\phi_+ a^2}{\log 2}.$$

Choose

$$a^2 := \min\left\{\frac{\log 2}{32n\phi_+}, \frac{M^2}{8sd\log(8n)}\right\}.$$

The first term in this minimum ensures $\beta_1/\log N \le 1/4$, while the second term that follows from equation 8 ensures that every retained candidate satisfies the margin constraint. Furthermore, $N \ge 4$ gives $\log 2/\log N \le 1/2$. Applying Lemma 2 with $d(\eta, \eta') = \|\eta - \eta'\|_n^2$ and $C = 2$ gives

$$\begin{aligned}
\mathfrak{R}_n(\mathcal{G}_{s,M}) &\ge \frac{\alpha_1}{4}\left(1 - \frac{\beta_1}{\log N} - \frac{\log 2}{\log N}\right) \\
&\ge \frac{\alpha_1}{4}\left(1 - \frac{1}{4} - \frac{1}{2}\right) \\
&= \frac{\alpha_1}{16} \\
&= \frac{\phi_- a^2 sd}{32} \\
&= \min\left\{\frac{\phi_- sd\log 2}{1024n\phi_+}, \frac{\phi_- M^2}{256\log(8n)}\right\}.
\end{aligned}$$

Equivalently,

$$\mathfrak{R}_n(\mathcal{G}_{s,M}) \gtrsim \min\left\{\frac{\phi_-}{\phi_+}\frac{sd}{n}, \frac{\phi_- M^2}{\log n}\right\}.$$

For fixed s, M, and ϕ_+, the first term is smaller for sufficiently large n. Hence

$$\mathfrak{R}_n(\mathcal{G}_{s,M}) \gtrsim \frac{\phi_-}{\phi_+}\frac{sd}{n}.$$

Part 2. Packing over activation-pattern supports. Let $g := \lfloor P/s \rfloor$ and $L := \lfloor \log_2 g \rfloor$, so $g \ge 2$. Since $sg \le P$ and $2^L \le g$, we may choose disjoint groups $G_1, \ldots, G_s \subseteq [P]$, each containing at least 2^L patterns. For each $t \in [s]$, fix a subset $G_t' \subseteq G_t$ of cardinality 2^L and a bijection $\psi_t : \{0,1\}^L \to G_t'$. For $w = (w^{(1)}, \ldots, w^{(s)}) \in \{0,1\}^{sL}$ with $w^{(t)} \in \{0,1\}^L$, define

$$S(w) := \{\psi_1(w^{(1)}), \ldots, \psi_s(w^{(s)})\}.$$

The groups are disjoint and each $S(w)$ contains exactly one element from every G_t', so $|S(w)| = s$. Lemma 3 provides a packing $\Omega_2 \subseteq \{0,1\}^{sL}$ of cardinality $N_2 \ge 2^{sL/8}$ such that $H(w, w') \ge sL/8$ for distinct $w, w' \in \Omega_2$. Let

$$B(w, w') := \left|\left\{t \in [s] : w^{(t)} \ne w'^{(t)}\right\}\right|.$$

Decomposing the Hamming distance blockwise gives

$$H(w,w') = \sum_{t=1}^{s} H(w^{(t)}, w'^{(t)}) \le \sum_{t=1}^{s} L\mathbf{1}_{\{w^{(t)} \neq w'^{(t)}\}} = LB(w,w').$$

Therefore $B(w,w') \ge H(w,w')/L \ge s/8$. If $w^{(t)} = w'^{(t)}$, the two supports select the same element of G'_t. If $w^{(t)} \neq w'^{(t)}$, bijectivity of ψ_t implies that the two selected elements are distinct, and this block contributes two elements to the symmetric difference. Hence, the cardinality of the symmetric difference of $S(\omega)$ and $S(\omega')$ is

$$|S(w) \triangle S(w')| = |[S(\omega)\backslash S(\omega')] \cup [S(\omega')\backslash S(\omega)]| = 2B(w,w') \ge \frac{s}{4}.$$

Let $\{\epsilon_{pk} : p \in [P],\ k \in [d]\}$ be independent Rademacher random variables, and let $e_1, \dots, e_d$ be an orthonormal basis of $\mathbb{R}^d$. For fixed (w, ϵ), define

$$u_p^{(w,\epsilon)} := b\widetilde{e}_p^{(\epsilon)} \mathbf{1}_{\{p \in S(w)\}},$$

where $\widetilde{e}_p^{(\epsilon)} = \sum_{k=1}^{d} \epsilon_{pk} e_k / \sqrt{d}$ with $\|\widetilde{e}_p^{(\epsilon)}\|_2 = 1$. For every $i \in [n]$ and fixed w,

$$\left(Au^{(w,\epsilon)}\right)_i = \sum_{p=1}^{P} (A_p)_{i\cdot} u_p^{(w,\epsilon)} = \frac{b}{\sqrt{d}} \sum_{p \in S(w)} \sum_{k=1}^{d} (A_p)_{ik} \epsilon_{pk}.$$

This is a Rademacher sum with squared coefficient sum

$$\frac{b^2}{d} \sum_{p \in S(w)} \sum_{k=1}^{d} (A_p)_{ik}^2 \le \frac{b^2}{d} \sum_{p \in S(w)} \|(A_p)_{i\cdot}\|_2^2 \le \frac{b^2}{d} \sum_{p \in S(w)} 4d = 4sb^2,$$

where the second inequality uses embedding normalization $\|(A_p)_{i\cdot}\|_2 \le 2\sqrt{d}$. Hoeffding's inequality therefore yields

$$\mathbb{P}_\epsilon \left(\left| \left(Au^{(w,\epsilon)}\right)_i \right| > M \right) \le 2\exp\left\{ -\frac{dM^2}{2b^2 \sum_{p \in S(w)} \sum_{k=1}^{d} (A_p)_{ik}^2} \right\} \le 2\exp\left(-\frac{M^2}{8sb^2} \right).$$

A union bound over the n comparisons gives

$$\mathbb{P}_\epsilon \left(\|Au^{(w,\epsilon)}\|_\infty > M \right) \le 2n\exp\left(-\frac{M^2}{8sb^2} \right).$$

If

$$b^2 \le \frac{M^2}{8s\log(8n)}, \tag{9}$$

then $\|Au^{(w,\epsilon)}\|_\infty \le M$ holds with probability at least $3/4$ for every $w \in \Omega_2$. Define $J_w(\epsilon) := \mathbf{1}_{\{\|Au^{(w,\epsilon)}\|_\infty \le M\}}$. Then

$$\mathbb{E}_\epsilon \left[\sum_{w \in \Omega_2} J_w(\epsilon) \right] = \sum_{w \in \Omega_2} \mathbb{P}_\epsilon \left(\|Au^{(w,\epsilon)}\|_\infty \le M \right) \ge \frac{3N_2}{4}.$$

Consequently, there exists a realization $\epsilon^\star$ for which $\sum_{w \in \Omega_2} J_w(\epsilon^\star) \ge 3N_2/4$. Define

$$\widetilde{\Omega}_2 := \left\{ w \in \Omega_2 : \|Au^{(w,\epsilon^\star)}\|_\infty \le M \right\}.$$

Then $|\widetilde{\Omega}_2| \ge 3N_2/4 \ge (3/4) \times 2^{sL/8}$. For sufficiently large sL, this implies $|\widetilde{\Omega}_2| \ge 2^{sL/16} \ge 4$. By fixing $\epsilon^\star$, relabelling the retained cardinality as N, and abbreviating $u^{(w)} := u^{(w,\epsilon^\star)}$ for $w \in \widetilde{\Omega}_2$, the packing is deterministic with $\|Au^{(\omega)}\|_\infty \le M$ for each $\omega \in \widetilde{\Omega}_2$. So, candidates $u^{(w)} \in \mathcal{F}_{s,M}$ and $\eta^{(w)} \in \mathcal{G}_{s,M}$ for every retained index. Since $\widetilde{\Omega}_2 \subseteq \Omega_2$, the original pairwise Hamming and support separations are preserved.

For distinct $w, w' \in \widetilde{\Omega}_2$ and every $p \in [P]$,

$$u_p^{(w)} - u_p^{(w')} = b\widetilde{e}_p^{(\epsilon^\star)} \left(\mathbf{1}_{\{p \in S(w)\}} - \mathbf{1}_{\{p \in S(w')\}} \right).$$

Using $\|\widetilde{e}_p^{(\epsilon^\star)}\|_2 = 1$ and $(\mathbf{1}_{\{p \in S(w)\}} - \mathbf{1}_{\{p \in S(w')\}})^2 = \mathbf{1}_{\{p \in S(w) \triangle S(w')\}}$, we obtain

$$\begin{aligned}
\sum_{p=1}^{P} \|u_p^{(w)} - u_p^{(w')}\|_2^2 &= b^2 \sum_{p=1}^{P} \left(\mathbf{1}_{\{p \in S(w)\}} - \mathbf{1}_{\{p \in S(w')\}} \right)^2 \\
&= b^2 |S(w) \triangle S(w')| \\
&\geq \frac{sb^2}{4}.
\end{aligned}$$

On the other hand, $|S(w) \triangle S(w')| \leq |S(w)| + |S(w')| = 2s$, so

$$\sum_{p=1}^{P} \|u_p^{(w)} - u_p^{(w')}\|_2^2 = b^2 |S(w) \triangle S(w')| \leq 2sb^2.$$

Moreover, $\operatorname{supp}\{u^{(w)} - u^{(w')}\} \subseteq S(w) \cup S(w')$, and therefore $|\operatorname{supp}\{u^{(w)} - u^{(w')}\}| \leq |S(w)| + |S(w')| = 2s$. Applying both sides of Assumption 1 gives

$$\|\eta^{(w)} - \eta^{(w')}\|_n^2 = \|A(u^{(w)} - u^{(w')})\|_n^2 \geq \phi_- \sum_{p=1}^{P} \|u_p^{(w)} - u_p^{(w')}\|_2^2 \geq \frac{\phi_- sb^2}{4},$$

$$\|\eta^{(w)} - \eta^{(w')}\|_n^2 \leq \phi_+ \sum_{p=1}^{P} \|u_p^{(w)} - u_p^{(w')}\|_2^2 \leq 2\phi_+ sb^2.$$

By Lemma 1,

$$D_{\mathrm{KL}}(\mathbb{P}_{\eta^{(w)}} \,\|\, \mathbb{P}_{\eta^{(w')}}) \leq \frac{n}{8} \|\eta^{(w)} - \eta^{(w')}\|_n^2 \leq \frac{n\phi_+ sb^2}{4}.$$

Set $\alpha_2 := \phi_- sb^2/4$ and $\beta_2 := n\phi_+ sb^2/4$. Since $\log N \geq sL \log 2/16$,

$$\frac{\beta_2}{\log N} \leq \frac{n\phi_+ sb^2/4}{sL \log 2/16} = \frac{4n\phi_+ b^2}{L \log 2}.$$

Choose

$$b^2 := \min \left\{ \frac{L \log 2}{16 n \phi_+}, \frac{M^2}{8s \log(8n)} \right\}.$$

The first term ensures $\beta_2 / \log N \leq 1/4$, and the second term that follows from equation 9 ensures the margin constraint. Since $N \geq 4$, we also have $\log 2 / \log N \leq 1/2$. Lemma 2, again with $C = 2$, therefore gives

$$\begin{aligned}
\mathfrak{R}_n(\mathcal{G}_{s,M}) &\geq \frac{\alpha_2}{4} \left(1 - \frac{\beta_2}{\log N} - \frac{\log 2}{\log N} \right) \\
&\geq \frac{\alpha_2}{4} \left(1 - \frac{1}{4} - \frac{1}{2} \right) \\
&= \frac{\alpha_2}{16} \\
&= \frac{\phi_- sb^2}{64} \\
&= \min \left\{ \frac{\phi_- sL \log 2}{1024 n \phi_+}, \frac{\phi_- M^2}{512 \log(8n)} \right\}.
\end{aligned}$$

Thus

$$\mathfrak{R}_n(\mathcal{G}_{s,M}) \gtrsim \min \left\{ \frac{\phi_-}{\phi_+} \frac{sL}{n}, \frac{\phi_- M^2}{\log n} \right\}.$$

For fixed s, M, and ϕ_+, the first term is the smaller one for all sufficiently large n. Moreover, $2^L \le g < 2^{L+1}$ and $g = \lfloor P/s \rfloor$. Since $P/s \ge 2$, these relations imply $L \asymp \log(P/s)$ up to universal numerical constants. It follows that

$$\mathfrak{R}_n(\mathcal{G}_{s,M}) \gtrsim \frac{\phi_-}{\phi_+} \frac{s \log(P/s)}{n}.$$

Let $c_1, c_2 > 0$ be the universal constants in the two packing bounds. The first construction gives $\mathfrak{R}_n(\mathcal{G}_{s,M}) \ge c_1(\phi_-/\phi_+)sd/n$, and the second gives $\mathfrak{R}_n(\mathcal{G}_{s,M}) \ge c_2(\phi_-/\phi_+)s\log(P/s)/n$. Therefore,

$$\begin{aligned}\mathfrak{R}_n(\mathcal{G}_{s,M}) &\ge \max\left\{c_1 \frac{\phi_-}{\phi_+}\frac{sd}{n}, c_2 \frac{\phi_-}{\phi_+}\frac{s\log(P/s)}{n}\right\} \\ &\ge \frac{\min\{c_1,c_2\}}{2}\frac{\phi_-}{\phi_+}\frac{s[d+\log(P/s)]}{n}.\end{aligned}$$

Taking $c = \frac{\min\{c_1,c_2\}}{2}\frac{\phi_-}{\phi_+}$ proves the theorem. □

E Proof of Theorem 2

Proof. First, we prove the restricted strong convexity of the loss function $\widehat{\mathcal{L}}_n(u) = \frac{1}{n}\sum_{i=1}^n \ell[Y_i(Au)_i]$. Let $u_t = t\widehat{u} + (1-t)u^\star$ for $t \in [0,1]$. Then $\frac{\mathrm{d}u_t}{\mathrm{d}t} = \widehat{u} - u^\star$. Since $\widehat{u}$ and $u^\star$ are both feasible, we have

$$\|Au_t\|_\infty = \|t\widehat{u} + (1-t)u^\star\|_\infty \le t\|A\widehat{u}\|_\infty + (1-t)\|Au^\star\|_\infty \le M,$$

which means u_t is also feasible. Define $g(t) = \widehat{\mathcal{L}}_n(u_t) = \frac{1}{n}\sum_{i=1}^n \ell[Y_i(Au_t)_i]]$. Then, since $Y_i^2 = 1$ and $\ell''(t) = \frac{e^t}{(1+e^t)^2} = \sigma(t)\sigma(-t)$, we have

$$\begin{aligned}g''(t) &= \frac{1}{n}\sum_{i=1}^n \sigma[Y_i(Au_t)_i]\sigma[-Y_i(Au_t)_i]\cdot Y_i^2[A(\widehat{u}-u^\star)]_i^2 \\ &= \frac{1}{n}\sum_{i=1}^n \sigma[(Au_t)_i]\{1-\sigma[(Au_t)_i]\}\cdot Y_i^2[A(\widehat{u}-u^\star)]_i^2 \\ &\ge \frac{\mu_M}{n}\sum_{i=1}^n [A(\widehat{u}-u^\star)]_i^2 \\ &= \mu_M\|A(\widehat{u}-u^\star)\|_n^2,\end{aligned}$$

where the inequality follows from the definition of μ_M and the fact that $|(Au_t)_i| \le \|Au_t\|_\infty \le M$. The integral version of Taylor's expansion gives

$$\begin{aligned}g(1) - g(0) - g'(0) &= \int_0^1 (1-t)g''(t)\mathrm{d}t \\ &\ge \int_0^1 (1-t)\mathrm{d}t \cdot \mu_M\|A(\widehat{u}-u^\star)\|_n^2 \\ &= \frac{\mu_M}{2}\|A(\widehat{u}-u^\star)\|_n^2,\end{aligned}$$

which is equivalent to

$$\widehat{\mathcal{L}}_n(\widehat{u}) - \widehat{\mathcal{L}}_n(u^\star) - \langle\nabla\widehat{\mathcal{L}}_n(u^\star), \widehat{u}-u^\star\rangle \ge \frac{\mu_M}{2}\|A(\widehat{u}-u^\star)\|_n^2. \tag{10}$$

Second, we control the score with high probability. The score of $\widehat{\mathcal{L}}_n$ at $u^\star$ at the u_p direction is

$$\begin{aligned}G_p &= \nabla_{u_p}\widehat{\mathcal{L}}_n(u^\star) \\ &= \frac{1}{n}\sum_{i=1}^n -\sigma[-Y_i(Au^\star)_i]Y_iA_{p_i} \\ &= \frac{1}{n}\sum_{i=1}^n A_{p_i}\cdot\left\{\sigma[(Au^\star)_i] - \mathbf{1}_{\{Y_i=1\}}\right\} \in \mathbb{R}^d.\end{aligned}$$

For vector $\theta \in U_p$ with $\|\theta\|_2 = 1$, we have

$$\theta^\top G_p = \frac{1}{n}\sum_{i=1}^{n}(A_p\theta)_i\zeta_i,$$

where $\zeta_i := \sigma[(Au^\star)_i] - \mathbf{1}_{\{Y_i=1\}}$. Note that $\zeta_i's$ are independent random variables that only takes value $\sigma[(Au^\star)_i]$ or $\sigma[(Au^\star)_i] - 1$, with $\mathbb{E}(\zeta_i) = 0$ under fixed design. Hoeffding inequality (Vershynin, 2018) and Assumption 2 give

$$\begin{aligned}\mathbb{P}\left(|\theta^\top G_p| > t\right) &= \mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}(A_p\theta)_i\zeta_i\right| > t\right)\\ &\le 2\exp\left\{-\frac{2n^2t^2}{\|A_p\theta\|_2^2}\right\}\\ &\le 2\exp\left\{-\frac{nt^2}{2}\right\}.\end{aligned} \tag{11}$$

For each $p \in [P]$, let $\mathcal{N}_p$ be an $\frac{1}{2}$-net of the Euclidean unit sphere in $\mathcal{U}_p$, By Proposition 4.2.13 in Vershynin (2018), we have $|\mathcal{N}_p| \le 5^d$. Let Π_p be a projection matrix onto $\mathcal{U}_p$, then $\|\Pi_p G_p\|_2 \le 2\max_{\theta\in\mathcal{N}_p}|\theta^\top\Pi_p G_p| \le 2\max_{\theta\in\mathcal{N}_p}|\theta^\top G_p|$. Therefore, by equation 11 and the union bound, we have

$$\mathbb{P}\left(\max_{p\in[P]}\|\Pi_p G_p\|_2 > 2t\right) \le 2P5^d\exp\left(-\frac{nt^2}{2}\right).$$

Let

$$t = \sqrt{\frac{2d\log 5 + 2\log(2P/\delta)}{n}},$$

then

$$\max_{p\in[P]}\|\Pi_p G_p\|_2 \le 2\sqrt{2}\sqrt{\frac{d\log 5 + \log(2P/\delta)}{n}}$$

with probability at least $1-\delta$. Denote the dual norm of $\gamma_p(G_p)$ on $\mathcal{U}_p$ as $\gamma_p^\circ(G_p) = \sup_{a\in\mathcal{U}_p,\gamma_p(a)\le 1} G_p^\top a$. Then the dual norm of $\Gamma(\nabla\widehat{\mathcal{L}}_n(u^\star))$ is

$$\begin{aligned}\Gamma^\circ\left(\nabla\widehat{\mathcal{L}}_n(u^\star)\right) &= \sup_{\Gamma(u)\le 1}\sum_{p=1}^{P}G_p^\top u_p\\ &\le \sup_{\Gamma(u)\le 1}\sum_{p=1}^{P}\gamma_p^\circ(G_p)\gamma_p(u_p)\\ &\le \sup_{\Gamma(u)\le 1}\max_{p\in[P]}\gamma_p^\circ(G_p)\sum_{p=1}^{P}\gamma_p(u_p)\\ &\le \max_{p\in[P]}\gamma_p^\circ(G_p)\\ &= \max_{p\in[P]}\sup_{\substack{a\in\mathcal{U}_p\\ \gamma_p(a)\le 1}}G_p^\top a\\ &\le \max_{p\in[P]}\sup_{\substack{a\in\mathcal{U}_p\\ \|a\|_2\le 1}}(\Pi_p G_p)^\top a\\ &= \max_{p\in[P]}\|\Pi_p G_p\|_2.\end{aligned}$$

The last inequality holds since $a \in \mathcal{U}_p$ and Π_p is a projection onto $\mathcal{U}_p$, thus $G_p^\top a = (\Pi_p G_p)^\top a$. Also, since $\gamma_p(a) \ge \|a\|_2$ for $a \in \mathcal{U}_p$, we have $\gamma_p(a) \le 1$ implies $\|a\|_2 \le 1$. Therefore, the dual norm of the score is controlled by

$$\Gamma^\circ\left(\nabla\widehat{\mathcal{L}}_n(u^\star)\right) \le \max_{p\in[P]}\|\Pi_p G_p\|_2 \le 2\sqrt{2}\sqrt{\frac{d\log 5 + \log(2P/\delta)}{n}} \le \frac{\lambda}{2}, \tag{12}$$

with probability at least $1-\delta$, where the last inequality follows the λ choice in Theorem 2.

Third, we show that the error vector is in the cone $\mathbb{C}(S)$. Since $\widehat{u}$ minimizes equation 5, the basic inequality gives

$$\widehat{\mathcal{L}}_n(\widehat{u}) + \lambda\Gamma(\widehat{u}) \leq \widehat{\mathcal{L}}_n(u^\star) + \lambda\Gamma(u^\star). \tag{13}$$

Combining equation 10 and equation 13, we have

$$\frac{\mu_M}{2}\|A(\widehat{u}-u^\star)\|_n^2 \leq \lambda\left[\Gamma(u^\star) - \Gamma(\widehat{u})\right] - \langle \nabla\widehat{\mathcal{L}}_n(u^\star), \widehat{u}-u^\star\rangle. \tag{14}$$

Hölder's inequality and equation 12 give

$$\begin{aligned} -\langle \nabla\widehat{\mathcal{L}}_n(u^\star), \widehat{u}-u^\star\rangle &\leq |\langle \nabla\widehat{\mathcal{L}}_n(u^\star), \widehat{u}-u^\star\rangle| \\ &\leq \Gamma^\circ\left(\nabla\widehat{\mathcal{L}}_n(u^\star)\right)\Gamma\left(\widehat{u}-u^\star\right) \\ &\leq \frac{\lambda}{2}\left\{\Gamma\left[(\widehat{u}-u^\star)_S\right] + \Gamma\left[(\widehat{u}-u^\star)_{S^c}\right]\right\}. \end{aligned} \tag{15}$$

Since $u^\star_{S^c} = 0$, we have

$$\begin{aligned} \Gamma(u^\star) - \Gamma(\widehat{u}) &= \Gamma(u^\star) - \Gamma[u^\star + (\widehat{u}-u^\star)] \\ &= \Gamma(u^\star) - \sum_{p\in S}\gamma_p[u_p^\star + (\widehat{u}-u^\star)_p] - \Gamma[(\widehat{u}-u^\star)_{S^c}] \\ &\leq \Gamma(u^\star) - \sum_{p\in S}\left\{\gamma_p(u_p^\star) - \gamma_p[(\widehat{u}-u^\star)_p]\right\} - \Gamma[(\widehat{u}-u^\star)_{S^c}] \\ &= \Gamma\left[(\widehat{u}-u^\star)_S\right] - \Gamma\left[(\widehat{u}-u^\star)_{S^c}\right]. \end{aligned} \tag{16}$$

Combining equation 14, equation 15, and equation 16, we have

$$\frac{\mu_M}{2}\|A(\widehat{u}-u^\star)\|_n^2 + \frac{\lambda}{2}\Gamma\left[(\widehat{u}-u^\star)_{S^c}\right] \leq \frac{3\lambda}{2}\Gamma\left[(\widehat{u}-u^\star)_S\right]. \tag{17}$$

Since $\frac{\mu_M}{2}\|A(\widehat{u}-u^\star)\|_n^2 \geq 0$, we have

$$\Gamma\left[(\widehat{u}-u^\star)_{S^c}\right] \leq 3\Gamma\left[(\widehat{u}-u^\star)_S\right].$$

Therefore, the error vector $\widehat{u}-u^\star \in \mathbb{C}(S)$.

Finally, we can obtain the predictive error bound. Dropping the nonnegative off-support term in equation 17 gives

$$\mu_M\|A(\widehat{u}-u^\star)\|_n^2 \leq 3\lambda\Gamma\left[(\widehat{u}-u^\star)_S\right]. \tag{18}$$

Then, by Cauchy–Schwarz inequality and Assumption 3, we have

$$\begin{aligned} \Gamma\left[(\widehat{u}-u^\star)_S\right] &= \sum_{p\in S}\gamma_p[(\widehat{u}-u^\star)_p] \\ &\leq \sqrt{s}\left\{\sum_{p\in S}\gamma_p[(\widehat{u}-u^\star)_p]^2\right\}^{\frac{1}{2}} \\ &\leq \sqrt{\frac{s}{\kappa}}\|A(\widehat{u}-u^\star)\|_n. \end{aligned} \tag{19}$$

Combining equation 18 and equation 19 gives

$$\mu_M\|A(\widehat{u}-u^\star)\|_n^2 \leq 3\lambda\sqrt{\frac{s}{\kappa}}\|A(\widehat{u}-u^\star)\|_n.$$

If $\|A(\widehat{u}-u^\star)\|_n = 0$, the conclusion is immediate. Otherwise, we have

$$\|A(\widehat{u}-u^\star)\|_n^2 \leq \frac{9\lambda^2 s}{\mu_M^2\kappa}.$$

Substituting

$$\lambda = 4\sqrt{2}\sqrt{\frac{d\log 5 + \log(2P/\delta)}{n}},$$

gives

$$\|A(\widehat{u}-u^\star)\|_n^2 \leq \frac{288s[d\log 5 + \log(2P/\delta)]}{\mu_M^2\kappa n}.$$

□

F Proof of Corollary 1

Proof. Let $\widehat{\eta} = A\widehat{u}, \eta^\circ = Au^\circ$, and $\Delta = \widehat{u} - u^\circ$. For a margin vector η such that $\|\eta\|_\infty \le M$, define the conditional population loss

$$\mathcal{L}_n(\eta) = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n \ell(Y_i\eta_i) \,\middle|\, X, D_1, \ldots, D_P\right].$$

Let $b(t) = \log(1 + e^t)$, then

$$\mathcal{L}_n(\eta) = \frac{1}{n}\sum_{i=1}^n \{b(\eta_i) - \sigma(\eta_i^\star)\eta_i\}.$$

Define the difference in conditional population loss between η and the true margin vector $\eta^\star$ as

$$\mathscr{D}(\eta) = \mathcal{L}_n(\eta) - \mathcal{L}_n(\eta^\star).$$

Since $b'(\eta_i^\star) = \sigma(\eta_i^\star)$ and $\mu_M \le b''(t) \le 1/4$ for $|t| \le M$, Taylor's theorem gives

$$\frac{\mu_M}{2}\|\eta - \eta^\star\|_n^2 \le \mathscr{D}(\eta) \le \frac{1}{8}\|\eta - \eta^\star\|_n^2 \tag{20}$$

for all η satisfying $\|\eta\|_\infty \le M$. Recall we have defined in the proof of Theorem 2 that $\zeta_i = \sigma(\eta_i^\star) - \mathbf{1}_{\{Y_i=1\}}$. The variables ζ_i are independent, mean zero, and bounded. Define $\widetilde{G} = (\widetilde{G}_1, \ldots, \widetilde{G}_P)$ where $\widetilde{G}_p = \frac{1}{n}A_p^\top\zeta$, which is a centered analogue of the block score G_p used in the proof of Theorem 2. They coincide under exact realizability, where $\eta^\circ = \eta^\star$. Hence, by the same net and union-bound argument used for the score control in the proof of Theorem 2, we have

$$\Gamma^\circ(\widetilde{G}) \le \frac{\lambda}{2} \tag{21}$$

with probability at least $1 - \delta$. On this event, optimality of $\widehat{u}$ gives

$$\widehat{\mathcal{L}}_n(\widehat{u}) + \lambda\Gamma(\widehat{u}) \le \widehat{\mathcal{L}}_n(u^\circ) + \lambda\Gamma(u^\circ).$$

Since

$$\widehat{\mathcal{L}}_n(u) - \mathcal{L}_n(Au) = \langle\widetilde{G}, u\rangle,$$

the preceding inequality, together with equation 21 and the decomposability argument from the proof of Theorem 2, yields

$$\mathscr{D}(\widehat{\eta}) - \mathscr{D}(\eta^\circ) + \frac{\lambda}{2}\Gamma(\Delta_{(S^\circ)^c}) \le \frac{3\lambda}{2}\Gamma(\Delta_{S^\circ}). \tag{22}$$

If $\mathscr{D}(\widehat{\eta}) \le \mathscr{D}(\eta^\circ)$, the result follows immediately from equation 20 since

$$\|\widehat{\eta} - \eta^\star\|_n^2 \le \frac{2}{\mu_M}\mathscr{D}(\widehat{\eta}) \le \frac{2}{\mu_M}\mathscr{D}(\eta^\circ) \le \frac{1}{4\mu_M}\|\eta^\circ - \eta^\star\|_n^2 = \frac{\mathcal{E}_{\text{app}}}{4\mu_M}.$$

Otherwise, equation 22 implies

$$\Gamma(\Delta_{(S^\circ)^c}) \le 3\Gamma(\Delta_{S^\circ}),$$

so $\Delta \in \mathcal{C}(S^\circ)$. By the same Cauchy–Schwarz and restricted-eigenvalue argument as in the proof of Theorem 2,

$$\Gamma(\Delta_{S^\circ}) \le \sqrt{\frac{s}{\kappa}}\|A\Delta\|_n.$$

Thus,

$$\mathscr{D}(\widehat{\eta}) - \mathscr{D}(\eta^\circ) \le \frac{3\lambda}{2}\sqrt{\frac{s}{\kappa}}\|A\Delta\|_n. \tag{23}$$

Moreover, by the triangle inequality and equation 20,

$$\|A\Delta\|_n \le \|\widehat{\eta} - \eta^\star\|_n + \|\eta^\circ - \eta^\star\|_n \le \sqrt{\frac{2}{\mu_M}}\left\{\sqrt{\mathscr{D}(\widehat{\eta})} + \sqrt{\mathscr{D}(\eta^\circ)}\right\}.$$

Substituting this into equation 23 and cancelling $\sqrt{\mathscr{D}(\widehat{\eta})} + \sqrt{\mathscr{D}(\eta^\circ)}$ gives

$$\sqrt{\mathscr{D}(\widehat{\eta})} \le \sqrt{\mathscr{D}(\eta^\circ)} + \frac{3\lambda}{2}\sqrt{\frac{2s}{\kappa\mu_M}}.$$

Using equation 20 and $(a+b)^2 \le 2a^2 + 2b^2$, we obtain

$$\|\widehat{\eta} - \eta^\star\|_n^2 \le \frac{1}{2\mu_M}\|\eta^\circ - \eta^\star\|_n^2 + \frac{18s\lambda^2}{\mu_M^2\kappa}.$$

By the definition of u°,

$$\|\eta^\circ - \eta^\star\|_n^2 = \mathcal{E}_{\text{app}}.$$

Finally, substituting the choice of λ from Theorem 2 gives

$$\|\widehat{\eta} - \eta^\star\|_n^2 \le \frac{\mathcal{E}_{\text{app}}}{2\mu_M} + \frac{576s[d\log 5 + \log(2P/\delta)]}{\mu_M^2\kappa n},$$

which proves the claim. □

G Proof of Theorem 3

Proof. Let $\Xi := \widehat{u} - u^\star$ and write $J_s := J_s(\Xi)$. The restricted-strong-convexity argument used in the proof of Theorem 2, together with the basic inequality for the estimator in equation 5, yields

$$\frac{\mu_M}{2}\|A\Xi\|_n^2 \le \lambda\{\Gamma(u^\star) - \Gamma(\widehat{u})\} - \left\langle \nabla\widehat{\mathcal{L}}_n(u^\star), \Xi \right\rangle. \tag{24}$$

By the same decomposability of Γ argement as in the proof of Theorem 2, we have

$$\Gamma(u^\star) - \Gamma(\widehat{u}) = \Gamma(u^\star) - \Gamma(u^\star + \Xi) \le \Gamma(\Xi_S) - \Gamma(\Xi_{S^c}).$$

Since J_s indexes the s largest block norms of Ξ , we have $\Gamma(\Xi_{J_s}) \ge \Gamma(\Xi_S)$. Therefore,

$$\begin{aligned}\Gamma(\Xi_S) - \Gamma(\Xi_{S^c}) &= 2\Gamma(\Xi_S) - \Gamma(\Xi) \\ &\le 2\Gamma(\Xi_{J_s}) - \Gamma(\Xi) \\ &= \Gamma(\Xi_{J_s}) - \Gamma(\Xi_{J_s^c}),\end{aligned}$$

and hence

$$\Gamma(u^\star) - \Gamma(\widehat{u}) \le \Gamma(\Xi_{J_s}) - \Gamma(\Xi_{J_s^c}). \tag{25}$$

We next control the stochastic term in equation 24. For $p \in [P]$, recall that $G_p := \nabla_{u_p}\widehat{\mathcal{L}}_n(u^\star)$ and define its block score dual norm by $Z_p := \gamma_p^\circ(G_p)$. Let $Z_{(1)} \ge Z_{(2)} \ge \cdots \ge Z_{(P)}$ be the decreasing rearrangement of $Z_1, \ldots, Z_P$. Separately, let $\gamma_{(1)}(\Xi) \ge \gamma_{(2)}(\Xi) \ge \cdots \ge \gamma_{(P)}(\Xi)$ be the decreasing rearrangement of the block error norms $\gamma_1(\Xi_1), \ldots, \gamma_P(\Xi_P)$. These two orderings need not correspond to the same pattern indices. By the Hölder inequality and the rearrangement inequality,

$$\begin{aligned}\left|\left\langle \nabla\widehat{\mathcal{L}}_n(u^\star), \Xi \right\rangle\right| &= \left|\sum_{p=1}^P G_p^\top \Xi_p\right| \\ &\le \sum_{p=1}^P |G_p^\top \Xi_p| \\ &\le \sum_{p=1}^P Z_p\gamma_p(\Xi_p) \\ &\le \sum_{j=1}^P Z_{(j)}\gamma_{(j)}(\Xi) \\ &= \sum_{j=1}^s Z_{(j)}\gamma_{(j)}(\Xi) + \sum_{j=s+1}^P Z_{(j)}\gamma_{(j)}(\Xi).\end{aligned}$$

For the first term, the Cauchy–Schwarz inequality gives

$$\sum_{j=1}^{s} Z_{(j)}\gamma_{(j)}(\Xi) \le \left(\sum_{j=1}^{s} Z_{(j)}^2\right)^{1/2} \left\{\sum_{j=1}^{s} \gamma_{(j)}(\Xi)^2\right\}^{1/2} = \left(\sum_{j=1}^{s} Z_{(j)}^2\right)^{1/2} \Gamma_{2,s}(\Xi).$$

For the second term, $Z_{(j)} \le Z_{(s+1)}$ whenever $j > s$, and consequently

$$\sum_{j=s+1}^{P} Z_{(j)}\gamma_{(j)}(\Xi) \le Z_{(s+1)} \sum_{j=s+1}^{P} \gamma_{(j)}(\Xi) = Z_{(s+1)}\Gamma(\Xi_{J_s^c}).$$

Combining these two bounds yields

$$\left|\left\langle \nabla \widehat{\mathcal{L}}_n(u^\star), \Xi \right\rangle\right| \le \left(\sum_{j=1}^{s} Z_{(j)}^2\right)^{1/2} \Gamma_{2,s}(\Xi) + Z_{(s+1)}\Gamma(\Xi_{J_s^c}). \tag{26}$$

It remains to control $\left(\sum_{j=1}^{s} Z_{(j)}^2\right)^{1/2}$ and $Z_{(s+1)}$. As established in the proof of Theorem 2, for every $p \in [P]$, $\theta \in U_p$ with $\|\theta\|_2 = 1$, and $t > 0$,

$$\mathbb{P}\big(|\theta^\top G_p| > t\big) \le 2\exp\left(-\frac{nt^2}{2}\right).$$

The same $1/2$-net argument therefore gives

$$\mathbb{P}\big(\|G_p\|_2 > 2t\big) \le 2 \cdot 5^d \exp\left(-\frac{nt^2}{2}\right).$$

Proposition 1 implies that $\gamma_p(a) \ge \|a\|_2$ and hence $Z_p \le \|G_p\|_2$. Thus,

$$\mathbb{P}(Z_p > t) \le 2 \cdot 5^d \exp\left(-\frac{nt^2}{8}\right). \tag{27}$$

We first control $Z_{(s+1)}$. Define $N(t) := \sum_{p=1}^{P} \mathbf{1}_{\{Z_p > t\}}$. Then $\{Z_{(s+1)} > t\} = \{N(t) \ge s+1\}$, and

$$\mathbb{E}[N(t)] = \sum_{p=1}^{P} \mathbb{P}(Z_p > t) \le 2P5^d \exp\left(-\frac{nt^2}{8}\right).$$

By Markov's inequality,

$$\mathbb{P}(N(t) \ge s+1) \le \frac{\mathbb{E}[N(t)]}{s+1} \le \frac{2P5^d \exp\big(-nt^2/8\big)}{s+1},$$

and therefore

$$\mathbb{P}(Z_{(s+1)} > t) \le \frac{2P5^d \exp\big(-nt^2/8\big)}{s+1}.$$

Taking

$$t = \sqrt{\frac{8d\log 5 + 8\log\frac{4P}{(s+1)\delta}}{n}}$$

gives

$$\mathbb{P}\left[Z_{(s+1)} \le \sqrt{\frac{8d\log 5 + 8\log\frac{4P}{(s+1)\delta}}{n}}\right] \ge 1 - \frac{\delta}{2},$$

so, with probability at least $1 - \delta/2$,

$$Z_{(s+1)} \le 2\sqrt{2}\sqrt{\frac{d\log 5 + \log\frac{4P}{(s+1)\delta}}{n}}.$$

We next control $\left(\sum_{j=1}^{s} Z_{(j)}^2\right)^{1/2}$. For $0 < \alpha < n/8$ and nonnegative random variable Z, by equation 27 and tail integral formula,

$$\begin{aligned}
\mathbb{E}\exp\{\alpha Z^2\} &= \int_0^\infty \mathbb{P}(\exp\{\alpha Z^2\} > u)\mathrm{d}u \\
&= 1 + \int_1^\infty \mathbb{P}(\exp\{\alpha Z^2\} > u)\mathrm{d}u \\
&= 1 + \int_0^\infty 2\alpha t \exp\{\alpha t^2\}\mathbb{P}(Z > t)\mathrm{d}t \\
&\le 1 + 4\alpha 5^d \int_0^\infty t \exp\left\{\left(\alpha - \frac{n}{8}\right)t^2\right\}\mathrm{d}t \\
&= 1 + \frac{2\alpha 5^d}{\frac{n}{8} - \alpha}.
\end{aligned}$$

With $\alpha = \frac{n}{16}$, we have

$$\mathbb{E}\exp\left\{\frac{nZ_p^2}{16}\right\} \le 1 + 2\cdot 5^d \le 3\cdot 5^d.$$

By Jensen's inequality,

$$\exp\left\{\frac{1}{s}\sum_{j=1}^{s}\alpha Z_{(j)}^2\right\} \le \frac{1}{s}\sum_{j=1}^{s}\exp\left\{\alpha Z_{(j)}^2\right\} \le \frac{1}{s}\sum_{p=1}^{P}\exp\left\{\alpha Z_p^2\right\}.$$

Taking expectation on both sides and using $\alpha = n/16$ gives

$$\mathbb{E}\exp\left\{\frac{n}{16s}\sum_{j=1}^{s} Z_{(j)}^2\right\} \le \frac{1}{s}\sum_{p=1}^{P}\mathbb{E}\exp\left\{\frac{n}{16}Z_p^2\right\} \le \frac{3P5^d}{s}.$$

Then, by Markov's inequality,

$$\mathbb{P}\left(\exp\left\{\frac{n}{16s}\sum_{j=1}^{s} Z_{(j)}^2\right\} > \frac{6P5^d}{s\delta}\right) \le \frac{\delta}{2}.$$

So, with probability at least $1-\delta/2$, we have

$$\sum_{j=1}^{s} Z_{(j)}^2 \le \frac{16s(d\log 5 + \log\frac{6P}{s\delta})}{n} \qquad \text{and} \qquad \left(\sum_{j=1}^{s} Z_{(j)}^2\right)^{1/2} \le 4\sqrt{\frac{s(d\log 5 + \log\frac{6P}{s\delta})}{n}}$$

Since the bound for $Z_{(s+1)}$ and $(\sum_{j=1}^{s} Z_{(j)}^2)^{1/2}$ each holds with probability at least $1-\delta/2$, taking a union bound gives both two bounds hold simultaneously with probability at least $1-\delta$. For the stated choice of λ in Theorem 3, these stochastic bounds imply $Z_{(s+1)} \le \lambda/2$ and $(\sum_{j=1}^{s} Z_{(j)}^2)^{1/2} \le \sqrt{s}\lambda/2$. Substitution into equation 26 gives

$$\left|\left\langle \nabla\widehat{\mathcal{L}}_n(u^\star), \Xi\right\rangle\right| \le \frac{\lambda}{2}\left\{\sqrt{s}\,\Gamma_{2,s}(\Xi) + \Gamma(\Xi_{J_s^c})\right\}. \tag{28}$$

Combining equation 24, equation 25, and equation 28, we obtain

$$\frac{\mu_M}{2}\|A\Xi\|_n^2 \le \lambda\Gamma(\Xi_{J_s}) - \frac{\lambda}{2}\Gamma(\Xi_{J_s^c}) + \frac{\lambda\sqrt{s}}{2}\Gamma_{2,s}(\Xi).$$

Equivalently,

$$\mu_M\|A\Xi\|_n^2 + \lambda\Gamma(\Xi_{J_s^c}) \le 2\lambda\Gamma(\Xi_{J_s}) + \lambda\sqrt{s}\,\Gamma_{2,s}(\Xi). \tag{29}$$

Dropping the first nonnegative term on the left-hand side of equation 29 gives $\Gamma(\Xi_{J_s^c}) \le 2\Gamma(\Xi_{J_s}) + \sqrt{s}\,\Gamma_{2,s}(\Xi)$, so $\Xi \in \widetilde{\mathbb{C}}(J_s)$. Moreover, the Cauchy–Schwarz inequality yields $\Gamma(\Xi_{J_s}) \le$

$\sqrt{s}\,\Gamma_{2,s}(\Xi)$. Dropping the nonnegative off-support term in equation 29 and applying Assumption 4, we obtain

$$\mu_M \|A\Xi\|_n^2 \le 3\lambda\sqrt{s}\,\Gamma_{2,s}(\Xi) \le 3\lambda\sqrt{\frac{s}{\widetilde{\kappa}}}\|A\Xi\|_n.$$

If $\|A\Xi\|_n = 0$, the conclusion is immediate. Otherwise, division by $\|A\Xi\|_n$ gives

$$\|A\Xi\|_n \le \frac{3\lambda}{\mu_M}\sqrt{\frac{s}{\widetilde{\kappa}}},$$

and therefore

$$\|\widehat{u} - u^\star\|_{A,n}^2 = \|A\Xi\|_n^2 \le \frac{9s\lambda^2}{\widetilde{\kappa}\mu_M^2} = \frac{576s[d\log 5 + \log\frac{6P}{s\delta}]}{\widetilde{\kappa}\mu_M^2 n}.$$

□

H Preference Datasets

We consider five publicly available preference datasets covering diverse language-model tasks and feedback sources:

- **WebGPT.** The WebGPT comparison dataset (Nakano et al., 2021) was collected for training reward models for long-form question answering with web browsing. Each example contains a question and two model-generated answers, together with human preference judgments indicating which answer is preferred. The answers were generated by models trained to browse the web and provide supporting references, making the dataset primarily focused on informative and factual long-form responses.
- **Chatbot Arena.** We use the `arena-human-preference-55k` dataset (Chiang et al., 2024), collected through Chatbot Arena. Each comparison consists of two anonymous LLM responses to the same real-world user prompt, with the user selecting response A, response B, a tie, or a tie in which both responses are considered poor. The dataset contains responses from more than 70 language models and covers a broad range of open-ended user queries.
- **UltraFeedback.** We use the binarized version of UltraFeedback (Cui et al., 2023; Tunstall et al., 2023). The original UltraFeedback dataset contains approximately 64,000 prompts, each paired with four responses generated by a variety of language models and evaluated by GPT-4 along several quality dimensions. In the binarized version, the response with the highest overall score is treated as the preferred response, while one of the remaining responses is selected as the rejected response. Unlike the other datasets considered here, its preference labels are therefore generated from AI feedback rather than direct human comparisons.
- **HH-RLHF.** The Anthropic Helpful–Harmless RLHF dataset (Bai et al., 2022) consists of paired assistant responses with human preference labels collected for training helpful and harmless dialogue agents. Each example provides a `chosen` and a `rejected` continuation of the same dialogue context. The dataset combines comparisons targeting helpfulness with comparisons designed to encourage harmless behavior, providing preference data over conversational assistant responses.
- **Summarize-from-Feedback.** The Summarize-from-Feedback dataset (Stiennon et al., 2020) was introduced for learning summarization models from human preferences. In its comparison subset, human annotators are shown two candidate summaries of the same source text and asked to select the better summary. The reward-model training data are primarily based on Reddit posts from the TL;DR dataset, with additional evaluation data derived from news summarization tasks.

I Table of Notations

J Table of hyperparameters

Table 3: Summary of the main notation.

Notation	Description
$n,\ d,\ P,\ s$	Number of comparisons, representation dimension, number of retained activation patterns, and pattern sparsity level, respectively.
$Z_{ij},\ Y_i$	Prompt–response pair $Z_{ij} = (\mathscr{S}_i, \mathscr{A}_{ij})$ and preference label $Y_i \in \{-1, +1\}$, where $j \in \{0, 1\}$.
$h,\ X$	Fixed representation map $h : \mathcal{Z} \to \mathbb{R}^d$ and stacked design matrix $X \in \mathbb{R}^{2n\times d}$.
$r^\star,\ \widehat{r},\ \Delta_i r$	True reward, estimated reward, and pairwise reward difference $\Delta_i r = r(Z_{i1}) - r(Z_{i0})$.
$\sigma(t),\ \ell(t)$	Sigmoid function $(1 + e^{-t})^{-1}$ and logistic loss $\log\big(1 + e^{-t}\big)$.
$\varepsilon,\ M$	Overlap floor and margin bound, with $M = \log\{(1 - \varepsilon)/\varepsilon\}$.
$\lambda,\ \delta$	Regularization level and failure probability in the high-probability bounds.
$g_p,\ D_p$	Sampled direction $g_p \sim \mathcal{N}(0, I_d)$ and activation-pattern matrix $D_p = \mathrm{diag}\{\mathbf{1}(Xg_p \geq 0)\}$.
$E_p,\ \mathcal{K}_p$	Cone matrix $E_p = (2D_p - I_{2n})X$ and activation cone $\mathcal{K}_p = \{a \in \mathbb{R}^d : E_p a \geq 0\}$.
$v_p,\ w_p,\ u_p$	Positive- and negative-output coefficient blocks $v_p, w_p \in \mathcal{K}_p$ and signed block $u_p = v_p - w_p$.
$q,\ u$	Stacked cone coefficients $q = (v_1, \ldots, v_P, w_1, \ldots, w_P)$ and signed coefficients $u = (u_1, \ldots, u_P)$.
η	Pairwise-margin vector $\eta = Fq = Au$.
$\mathcal{U}_p,\ \mathcal{U}$	Signed block space $\mathcal{U}_p = \{v - w : v \in \mathcal{K}_p, w \in \mathcal{K}_p\}$ and product space $\mathcal{U} = \mathcal{U}_1 \times \cdots \times \mathcal{U}_P$.
B	Pairwise incidence operator, $(Ba)_i = a_{2i-1} - a_{2i}$.
$A_p,\ A$	Pattern-specific margin operator $A_p = BD_pX$ and $A = [A_1\ \cdots\ A_P]$, so Au is the margin vector.
$F,\ C$	Lifted margin operator $F = [A, -A]$ and cone-constraint operator $C = \mathrm{blkdiag}(E_1, \ldots, E_P, E_1, \ldots, E_P)$; $Fq = Au$ and feasibility requires $Cq \geq 0$.
$\gamma_p,\ \Gamma$	Induced cone-decomposition norm $\gamma_p(u_p) = \inf_{v,w\in\mathcal{K}_p:\, v-w=u_p}(\|v\|_2 + \|w\|_2)$ and penalty $\Gamma(u) = \sum_{p=1}^P \gamma_p(u_p)$.
$u^\star,\ \widehat{u},\ S$	True and estimated signed parameters, and active-pattern support $S = \mathrm{supp}(u^\star) = \{p : u_p^\star \neq 0\}$, with $\lvert S\rvert \leq s$.
$\|a\|_n,\ \|u\|_{A,n}$	Empirical norm $\|a\|_n = n^{-1/2}\|a\|_2$ and prediction seminorm $\|u\|_{A,n} = \|Au\|_n$.
μ_M	Minimum logistic curvature on $[-M, M]$: $\mu_M = \sigma(M)[1 - \sigma(M)]$.
$\phi_-,\ \phi_+$	Lower and upper sparse comparison constants in Assumption 1.
$\kappa,\ \widetilde{\kappa}$	Restricted-eigenvalue constants in Assumptions 3 and 4, respectively.

Table 4: Main hyperparameters and solver settings of the proposed method in the three experimental settings. Shared settings are listed once; settings that vary across experiments are listed below them. Iteration counts are upper limits, with earlier termination upon convergence.

Shared setting	**Value**		
Confidence parameter δ	0.05		
ADMM penalties (ρ_η, ρ_ξ)	$(0.05, 0.05)$		
ADMM absolute / relative tolerance	10^{-4} / 10^{-3}		
Logistic proximal absolute / relative tolerance	10^{-7} / 10^{-7}		
Maximum logistic proximal iterations	60		
Experiment-specific setting	**Synthetic**	**Semi-synthetic**	**Real preference**
Regularization multiplier c	5×10^{-3}	5×10^{-5}	5×10^{-5}
Regularization rule	$\lambda_{\mathrm{thm3}},\ s_\lambda = s$	$\lambda_{\mathrm{thm3}},\ s_\lambda = 5$	λ_{thm2}
Overlap floor ϵ	0.1	0.1	0.001
Margin bound M	$\log 9$	$\log 9$	$\log 999$
ADMM penalty ρ_z	0.05	20	20
Retained activation patterns P	20; sweep below	20	20
Maximum ADMM iterations	20,000	100,000	100,000
Maximum CG iterations per ADMM step	100	2,000	2,000
CG relative tolerance	$10^{-3} \to 10^{-5}$	10^{-5} (fixed)	10^{-5} (fixed)
CG preconditioner	Nyström	Identity	Identity
Nyström rank / oversampling	20 / 5	—	—
Train–test split	—	—	$30/70$
Real preference dataset	$\boldsymbol{n_{\mathrm{train}}}$	$\boldsymbol{\lambda}$	
WebGPT	4,303	$2.478000234 \times 10^{-4}$	
Arena	11,914	$1.489217799 \times 10^{-4}$	
UltraFeedback	16,642	$1.260039783 \times 10^{-4}$	
HH-RLHF	50,805	$7.211635839 \times 10^{-5}$	
Summarize-from-Feedback	53,683	$7.015661512 \times 10^{-5}$	

The regularization rules (with natural logarithms) are

$$\lambda_{\mathrm{thm3}} = 8c\sqrt{\frac{d\log 5 + \log(6P/(s_\lambda\delta))}{n}}, \qquad \lambda_{\mathrm{thm2}} = 4\sqrt{2}\,c\sqrt{\frac{d\log 5 + \log(2P/\delta)}{n_{\mathrm{train}}}}.$$

Table 5: Main hyperparameters of the nonconvex two-layer ReLU benchmark in the two LLM-based experiments. All fits run for exactly 200 epochs and use the final model, without early stopping.

Shared setting	**Value**	
Network dimensions	$2048 \to 40 \to 1$; ReLU hidden layer	
Optimizer	Adam	
Learning rate	10^{-3} (constant)	
Adam parameters (β_1, β_2)	$(0.9, 0.999)$	
Adam numerical stabilizer	10^{-8}	
Batch size	1,024 preference pairs	
Training epochs	200 (fixed)	
Model selection	Final epoch	
Additional optimizer weight decay	0	
Regularization multiplier c	10^{-2}	
Experiment-specific setting	**Semi-synthetic**	**Real preference**
Regularization rule	$\lambda_{\mathrm{thm3}},\ s_\lambda = 5$	λ_{thm2}
Overlap floor ϵ	0.1	0.001
Margin bound M	$\log 9$	$\log 999$
Regularization parameter λ	$1.9175706603 \times 10^{-2}$	Dataset-specific; below
Real preference dataset	n_{train}	λ
WebGPT	4,303	$4.9560004688 \times 10^{-2}$
Arena	11,914	$2.9784355984 \times 10^{-2}$
UltraFeedback	16,642	$2.5200795666 \times 10^{-2}$
HH-RLHF	50,805	$1.4423271677 \times 10^{-2}$
Summarize-from-Feedback	53,683	$1.4031323025 \times 10^{-2}$

The regularization rules are those in Table 4, evaluated with $c = 0.01$, $d = 2048$, $P = 20$, and $\delta = 0.05$; $n = 57{,}477$ in the semi-synthetic experiments.