ONLINE PREFERENCE LEARNING VIA GENERALIZED LINEAR MODELS: MONOTONICITY AND CONVERGENCE OF SGD

Anonymous authors
Paper under double-blind review

ABSTRACT

Learning from pairwise comparisons underlies applications ranging from sports rankings to preference-based finetuning of language models. We cast comparison-based models as Generalized Linear Models learned online by constant-stepsize Stochastic Gradient Descent, recovering Elo as a special case. We first establish convergence guarantees for the last iterate by applying existing results for SGD on the population loss. We then study monotonicity: upon a gradient step relative to "a is preferred to b", how do the scores of a, b, and a third alternative c evolve? We characterize notions of monotonicity along a training trajectory both geometrically and algebraically, giving conditions on alternatives' feature vectors. We then derive consequences: *(i)* the score difference between the preferred and rejected alternative always moves in the correct direction, *(ii)* stronger forms of monotonicity are impossible when the number of alternatives A exceeds 2^D, where D is the number of features; and *(iii)* complementarily, these stronger forms hold almost surely as $D \gg \log A$, for independent subgaussian feature vectors. We conclude with experiments illustrating the convergence and monotonicity results.

1 INTRODUCTION

Comparison-based observations in the form "a is better than b" is arguably pervasive across several applications where absolute values are inferred from relative, or ordinal, information. Methods such as Elo rating (Élő, 1978) update chess players' ratings from game outcomes, while Reinforcement Learning from Human Feedback (RLHF) (Christiano et al., 2017; Stiennon et al., 2020) and Direct Preference Optimization (DPO) (Rafailov et al., 2023) align language models by learning from human preferences between candidate answers. Yet, despite their similarity, these methods have largely evolved independently, each with its own specialized theory.

Performing statistical inference in this setting presents significant challenges. First, accommodating both arbitrary outcome spaces—beyond binary win/loss—together with multidimensional item feature vectors requires handling a broad class of loss functions with few shared structural properties. Second, a basic requirement for any comparison-based system is *monotonicity*: an item should not become less preferred as a result of being reported as the winner of a comparison. This property has long been recognized as desirable in social choice theory, dating back to classical tournament ranking methods (Zermelo, 1929). It is naturally satisfied by covariate-free models such as Bradley–Terry–Elo (Bradley & Terry, 1952; Élő, 1978), where each item has its own parameter.

Modern systems, however, represent items through shared, low-dimensional feature vectors rather than free per-item parameters, precisely to *generalize* across items, as is the case for LLM alignment via RLHF/DPO. This creates a tension: enforcing monotonicity in its strongest, per-item sense conflicts with the very mechanism that enables generalization, since updating item a's score through shared features can unavoidably move the scores of unrelated items. Consistent with this tension, recent empirical findings show that fine-tuning on feedback "answer a beats b" can decrease the probability that a model assigns to a (Pal et al., 2024; Razin et al., 2024), eroding trust in training procedures and complicating data collection. Despite its practical importance, this tension between monotonicity and generalization remains poorly understood.

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.

Moreover, growing concerns around AI safety (Bengio et al., 2026; Jin & Lee, 2025) have made it increasingly important to establish provable guarantees for alignment procedures. This calls for a theoretical framework able to analyze convergence and monotonicity in a setting that accommodates rich outcome spaces and item representations, while capturing key features of practical learning procedures: online updates, constant learning rates, and reliance on the last iterate.

We provide such a framework by modeling pairwise comparison outcomes with a *Generalized Linear Model* (GLM) driven by a *base distribution*. Each item's feature vector is mapped linearly to a scalar score, and the resulting score difference tilts the base distribution towards the preferred item. We then study the online learning of this model via *Stochastic Gradient Descent* (SGD) with a constant step size. Specifically, our contributions are as follows:

- We formalize this construction in Section 2.1, building on the recent Generalized Bradley–Terry model (GBT) (Fageot et al., 2024; 2025), and its online learning as constant-stepsize SGD on the population risk in Section 2.2. This places ranking of sports and games (e.g., Elo) and supervised finetuning of LLMs (e.g., RLHF/DPO) as special cases of a single online GLM.
- Building on this framework, we show in Section 3 that recent results on constant-stepsize SGD directly yield sublinear or linear last-iterate convergence guarantees *on the population loss*, depending on easy-to-check properties of the base distribution (i.e., smoothness and strong convexity of the cumulant generating function). This analysis relies on existing tools; our contribution is to show they apply naturally to preference learning under our framework, not to derive new convergence techniques.
- We elucidate in Section 4 the tension between monotonicity and generalization identified above, through the following results.
  - A weaker form of monotonicity—the score difference between any two compared items moves in the direction of the observed outcome—holds universally, regardless of the feature vectors.
  - The stronger, social-choice-style notion of *gap* monotonicity (accounting for effects on uninvolved items) cannot hold for *any* feature vector once the number of items A exceeds 2^D, where D is the feature dimension—an impossibility result formalizing the tension with generalization.
  - Conversely, gap monotonicity holds almost surely for random sub-Gaussian feature vectors when $D \gg \log A$.

  These results do not resolve the tension between monotonicity and generalization in general, but pin down precisely when it leads to an outright impossibility versus when it can be avoided.
- We conduct numerical experiments in Section 5 illustrating convergence of the iterates across various settings, and the validity of our monotonicity guarantees.

We present additional related work in Appendix A.

2 Model and learning algorithm

Although our primary motivation is the online learning of preference and pairwise comparison models, most of our results and analysis can be framed more generally in the context of Generalized Linear Models (Nelder & Wedderburn, 1972). In this section, we fix notation and state the definitions and assumptions for Generalized Linear Models (GLMs) and Stochastic Gradient Descent (SGD) used throughout the paper. We also show that several common formulations of comparison-based learning arise as special cases of this general framework. This connection between GLMs and comparison-based models such as Bradley–Terry and Thurstone–Mosteller has been noted previously in the literature (Critchlow & Fligner, 1991); we present it in full generality in this section.

2.1 Generative Model

We assume random observations of the form (x, r) where $x \in \mathbb{R}^D$ is a feature vector and $r \in \mathbb{R}$ is the *outcome*, following a Generalized Linear Model in canonical form: $r \mid x$ is the member of the natural exponential family generated by a *base probability distribution* f with natural parameter $\theta = \langle \beta, x \rangle$, where $\beta \in \mathbb{R}^D$ parametrizes the linear component of the model. In other words, the conditional distribution of $r \mid x$ has density with respect to f is given by

$$\exp\left[r\langle \beta, x\rangle - \Phi_f\big(\langle \beta, x\rangle\big)\right]. \tag{1}$$

Table 1: Common base distributions f, description and properties of their corresponding CGFs Φ_f.

Name	$f(r), r \in \mathrm{dom}(f)$	$\mathrm{dom}(f)$	Parameter	$\Phi_f(\theta)$	Φ_f M-smoothness	Φ_f μ-strong convexity
Gaussian	$\frac{1}{\sqrt{2\pi\sigma_0^2}}\exp\left(-\frac{r^2}{\sigma_0^2}\right)$	$\mathbb{R}$	$\sigma_0^2 > 0$	$\frac{\sigma_0^2\theta^2}{2}$	✓ $M = \sigma_0^2$	✓ $\mu = \sigma_0^2$
Bradley–Terry	$\frac{1}{2}$	$\{-1, 1\}$	–	$\log(\cosh\theta)$	✓ $M = 1$	✗
Uniform	$\frac{1}{2}$	$(-1, 1)$	–	$\log\left(\frac{\sinh\theta}{\theta}\right)$	✓ $M = \frac{1}{3}$	✗
K-ary	$\frac{1}{2K+1}$	$\left\{\frac{r-K}{K}\right\}_{r=0}^{2K}$	$K \geq 1$	$\log\left(\frac{1}{2K+1}\frac{\sinh\left(\frac{(2K+1)\theta}{2K}\right)}{\sinh\left(\frac{\theta}{2K}\right)}\right)$	✓ $M = \frac{K+1}{6K}$	✗
Poisson ($\mathbb{N}$)	$\frac{e^{-\lambda}\lambda^r}{r!}$	$\mathbb{N}$	$\lambda > 0$	$\lambda(e^\theta - 1)$	✗	✗

In this canonical form, Φ_f is the cumulant generating function (CGF) of f:

$$\Phi_f(\theta) \triangleq \log\left(\int_{\mathbb{R}} e^{u\theta}\, df(u)\right),$$

and its derivative is the inverse link function (i.e. transfer function) of the model, mapping the natural parameter to the mean outcome (see Lemma 2):

$$\mathbb{E}[r \mid x] = \Phi_f'\big(\langle \beta, x\rangle\big). \tag{2}$$

We assume that the base measure f has finite exponential moments so that the above quantities are well-defined. The following definition summarizes the generative model. We comment on identifiability of the model in Appendix C.

Definition 1. Given, a feature distribution $\mathbb{P}_x$ over $\mathbb{R}^D$ and base probability measure f with finite exponential moments, the *generative model* with parameters $\beta \in \mathbb{R}^D$ is the probability distribution $\mathbb{P}_\beta$ over $\mathbb{R}^D \times \mathbb{R}$, whose log-density with respect to $\mathbb{P}_x \otimes f$ is given for $(x, r) \in \mathbb{R}^D \times \mathbb{R}$ by

$$\log \frac{\mathrm{d}\,\mathbb{P}_\beta}{\mathrm{d}(\mathbb{P}_x \otimes f)}(x, r) = r\langle x, \beta\rangle - \Phi_f\big(\langle x, \beta\rangle\big).$$

Example 1. The Bradley–Terry (BT) model is a popular probabilistic model of pairwise comparisons; it can be seen as a special case of Definition 1. In this model, there is a finite set of alternatives $\mathcal{A} \triangleq \{1, \ldots, A\}$ to be compared, each alternative $a \in \mathcal{A}$ having a *strength parameter* $\beta_a \in \mathbb{R}$. A pair of alternatives $(a, b) \in \mathcal{A}^2$ is randomly selected according to a distribution Q over $\mathcal{A}^2$, and their comparison yields an outcome r such that the probability of $r = 1$ (a wins against b) is $\frac{e^{\beta_a}}{e^{\beta_a}+e^{\beta_b}} = \sigma(\beta_a - \beta_b)$, where $\sigma(\beta) = (1 + e^{-\beta})^{-1}$ is the sigmoid function.

If $e_a \in \mathbb{R}^A$ denotes the indicator vector of alternative $a \in A$, then the BT model is the special case of Definition 1 where the feature vector associated with the pair (a, b) is $e_a - e_b$. The probability $\mathbb{P}_x$ is then the pushforward measure of Q through the map $(a, b) \mapsto e_a - e_b$. In the notation above, the base probability distribution f is obtained when the latent strengths β_a and β_b are equal: $f = \frac{1}{2}(\delta_1 + \delta_{-1})$. The corresponding CGF expresses as $\Phi_f(\theta) = \log\cosh(\theta)$; see Table 1. Then, the general probability (1) of $a \succ b$ given the strengths of a and b recovers the BT probability:

$$\frac{1}{2}\exp[(\beta_a - \beta_b) - \log\cosh(\beta_a - \beta_b)] = \frac{\exp[\beta_a - \beta_b]}{\exp[\beta_a - \beta_b] + \exp[\beta_b - \beta_a]} = \sigma(\beta_a - \beta_b).$$

Example 2. The Generalized Bradley–Terry (GBT) model (Fageot et al., 2024) generalizes the above example to accommodate non-binary comparisons, while still fitting Definition 1. Comparison of pairs of alternatives $(a, b) \in \mathcal{A}^2$ yields an outcome r such that the distribution of $r \mid a, b$ is the member of the natural exponential family generated by f with parameter $\beta_a - \beta_b$. Table 1 lists base distributions.

Example 3. We can further generalize Example 2 by having a feature vector $x_a \in \mathbb{R}^D$ associated with each alternative $a \in \mathcal{A}$ and resulting strength $\theta_a = \langle \beta, x_a\rangle$ for a vector of parameters $\beta \in \mathbb{R}^D$. The distribution of r is now the exponential family distribution generated by f with parameter $\theta_a - \theta_b$. This is again a special case of Definition 1, with feature vector $x = x_a - x_b$ and $\mathbb{P}_x$ the pushforward of Q through $(a, b) \mapsto x_a - x_b$.

2.2 Online learning via SGD

We now observe T i.i.d. samples $\mathcal{D} = \{(x^t, r^t)\}_{1 \leq t \leq T}$ from a model $\mathbb{P}_{\beta^\star}$ as in Definition 1 with unknown parameter $\beta^\star \in \mathcal{V}_x \triangleq \mathrm{SPAN}(\mathrm{supp}\,\mathbb{P}_x)$. The goal is to learn $\beta^\star$ in an online manner, wherein an estimate β^t is updated after receiving each sample.

Our proposed procedure is to perform a single pass of SGD on the *population loss* $\mathcal{L} : \mathbb{R}^D \to \mathbb{R}$,

$$\mathcal{L}(\beta) \triangleq \mathbb{E}_{(x,r)\sim\mathbb{P}_{\beta^\star}}\left[\ell_{x,r}(\beta)\right], \tag{Pop. loss}$$

where $\ell_{x,r} : \beta \mapsto \Phi_f\big(\langle x, \beta\rangle\big) - r\,\langle x, \beta\rangle$ denotes the *sample* loss. Starting with initialization $\beta^0 \in \mathcal{V}_x$, the resulting online update rule is given for $1 \le t \le T$ by

$$\beta^t = \beta^{t-1} - \eta\left[\Phi_f'\big(\langle x^t, \beta^{t-1}\rangle\big) - r^t\right] x^t, \tag{3}$$

where $\eta > 0$ is the learning rate. Using (2), we have

$$\Phi_f'\big(\langle x^t, \beta^{t-1}\rangle\big) - r^t = \mathbb{E}_{\beta^{t-1}}[r^t \mid x^t] - r^t.$$

Hence, the scaling factor in (3) has a natural interpretation: it quantifies the extent to which the observed outcome r^t under- or over-performed relative to the expected outcome under the current model β^{t-1}. This structure underlies our monotonicity results in Section 4.

Example 4 (Generalized Elo rating)**.** For the GBT model of Example 2, the observation at time t arises from the random selection of a pair $(a^t, b^t) \in \mathcal{A}$, and $x^t = e_{a^t} - e_{b^t}$. The update (3) thus only changes the coordinates a^t and b^t of the current estimate. For these coordinates, the updates are

$$\begin{cases} \beta^t_{a^t} = \beta^{t-1}_{a^t} + \eta\big[r^t - \Phi_f'(\beta^{t-1}_{a^t} - \beta^{t-1}_{b^t})\big] \\ \beta^t_{b^t} = \beta^{t-1}_{b^t} - \eta\big[r^t - \Phi_f'(\beta^{t-1}_{a^t} - \beta^{t-1}_{b^t})\big] \end{cases}. \tag{4}$$

The Bradley–Terry model corresponds to the case where f is uniform over $\{0, 1\}$. For this distribution $\Phi_f'(\theta) = \frac{1}{1+e^{-\theta}}$ is the standard logistic function, and (4) then becomes the update rule of the Elo rating system. This recovers the folklore observation that the Elo rating system is SGD on the Bradley–Terry model (Király & Qian, 2017, §3.1), or equivalently, on logistic regression when the feature vectors are pairwise difference of indicator vectors (Morse, 2019).

3 Convergence results

By Lemma 5 (Appx. C.2), the population loss $\mathcal{L}$ (Pop. loss) is convex on $\mathbb{R}^D$ and strictly convex on $\mathcal{V}_x$. Hence, when the observations are sampled from the GLM $\mathbb{P}_{\beta^\star}$ with $\beta^\star \in \mathcal{V}_x$, the unique minimizer of the population loss over $\mathcal{V}_x$ is precisely $\beta^\star$. Moreover, the expected negative log-likelihood of a GLM inherits many structural properties from the CGF of its base distribution, making it particularly amenable to a convergence analysis under SGD (Toulis et al., 2014; Bach, 2010).

In this section, we focus on the case where the learning rate η is constant and leverage recent SGD convergence results (Attia et al., 2025; Cortild et al., 2026) to state non-asymptotic bounds for the iterates (3). These results are particularly well-suited to our setting, as they rely on minimal assumptions. Most notably, they do not require a uniform bound on the variance of the stochastic gradients. We first treat the case of a base distribution f whose CGG Φ_f is smooth and strongly convex in Section 3.1, the more general smooth and convex case in Section 3.2, and discuss how these results specialize to comparison-based learning in Appendix B.

A common feature of these bounds, typical for SGD with constant step size, is their decomposition as the sum of two terms: a transient "bias" term depending on the squared error of the first iterate, and a steady-state "variance" term proportional to the step size and the curvature of the loss $\mathcal{L}$ at its minimum. This curvature, denoted σ_*^2, is given by

$$\sigma_*^2 \triangleq \mathbb{E}\left[\|\nabla\ell_{x,r}(\beta^\star)\|^2\right] = \mathbb{E}\left[(\Phi_f'(\langle x, \beta^\star\rangle) - r)^2\|x\|^2\right] = \mathbb{E}\left[\Phi_f''(\langle x, \beta^\star\rangle)\|x\|^2\right] = \operatorname{Tr} d^2\mathcal{L}(\beta^\star), \tag{5}$$

where equalities follow from Lemmas 2 and 4. These bounds are also driven by a uniform bound on the smoothness constants of *each* sample loss $\ell_{x,r}$, namely (see Lemma 5(4))

$$L_\star \triangleq MR_x^2, \qquad \text{where} \qquad R_x \triangleq \sup\{\|x\| \mid x \in \operatorname{supp}\mathbb{P}_x\}. \tag{6}$$

This differs from prior SGD convergence results that would feature the (smaller) smoothness constant N of $\mathcal{L}$, at the cost of not characterizing the last iterate convergence.

We make the following blanket assumption, which holds for all comparison-based models of Examples 2 and 3 with finitely many alternatives.

Assumption 1 (On the probabilistic model)**.** *The base distribution f has finite exponential moments, and $R_x \triangleq \sup\{\|x\| \mid x \in \operatorname{supp}\mathbb{P}_x\}$ is bounded.*

3.1 SMOOTH AND STRONGLY CONVEX CGF

When the CGF Φ_f of the base distribution f is smooth and strongly convex (e.g. when f is a normal distribution), then so is the loss $\mathcal{L}$ (cf. Lemma 5). In this case, we obtain the following convergence rate on the mean squared error of the SGD iterates, where the bias term decays linearly.

Proposition 1. *Let $\mathbb{P}_{\beta^\star}$ be the GLM of Definition 1 with ground truth parameter $\beta^\star \in \mathcal{V}_x$ that satisfies Assumption 1. Assume that the CGF Φ_f is M-smooth and μ-strongly convex, and let $\nu \triangleq \mu\lambda^+_{\min}(\mathbb{E}[xx^\top])$. Then, for each $\eta \in (0, 2/L_\star)$, given T i.i.d. samples from $\mathbb{P}_{\beta^\star}$, the SGD iterate β^T (3) satisfies*

$$\mathbb{E}_{\beta^\star}\|\beta^T - \beta^\star\|^2 \leq \exp\left(-\frac{1}{2}\eta(2-\eta L_\star)\nu T\right)\|\beta_0 - \beta^\star\|^2 + \frac{8\eta\sigma_*^2}{\nu(2-\eta L_\star)^2}.$$

Proof. By Lemma 5(4), almost every sample loss $\ell_{x,r}$ is convex and $L_\star$-smooth and by Lemma 5(3) the population loss $\mathcal{L}$ is ν-strongly convex on $\mathcal{V}_x$. The iterates (3) remain in $\mathcal{V}_x$ since $g_{x,r}(\beta)$ is collinear with x and $\beta^0 \in \mathcal{V}_x$. Finally $\sigma_*^2 = \mathbb{E}\|g_{x,r}(\beta^\star)\|^2 < \infty$ by (5). The bound then follows from Attia et al. (2025, proposition 9) applied on the Hilbert space $\mathcal{V}_x$, whose hypotheses require smoothness of each $\ell_{x,r}$, and strong convexity of the population loss. □

3.2 SMOOTH AND CONVEX CGF

We now relax the assumptions on Φ_f and only require smoothness and convexity ($\mu = 0$). In this case we obtain weaker convergence bounds on the loss gap for the *average* and *last* iterate.

Proposition 2. *Let $\mathbb{P}_{\beta^\star}$ be the GLM of Definition 1 with ground truth parameter $\beta^\star \in \mathcal{V}_x$ that satisfies Assumption 1. Assume that the CGF Φ_f is M-smooth. Then, given T i.i.d. samples from $\mathbb{P}_{\beta^\star}$,*

1. *if $\eta \in (0, 1/L_\star)$, the average of the SGD iterates (3), $\bar{\beta}^T \triangleq \frac{1}{T}\sum_{t=1}^T \beta^t$, satisfies:*

$$\mathbb{E}_{\beta^\star}\left[\mathcal{L}(\bar{\beta}^T)\right] - \mathcal{L}(\beta^\star) \leq \frac{\|\beta^0 - \beta^\star\|^2}{2\eta T} + \frac{\eta\sigma_*^2}{2(1-\eta L_\star)}.$$

2. *if $\eta \in (0, 2/L_\star)$, the last iterate β^T satisfies*

$$\mathbb{E}_{\beta^\star}\left[\mathcal{L}(\beta^T)\right] - \mathcal{L}(\beta^\star) \leq \frac{12\|\beta^0 - \beta^\star\|^2}{\eta(2-L_\star\eta)T^{1-L_\star\eta/2}} + \frac{24\eta}{(2-\eta L_\star)^2}\sigma_*^2 T^{L_\star\eta/2}\log_2(T+2).$$

Proof. By Lemma 5(4), almost every sample loss $\ell_{x,r}$ is convex and $L_\star$-smooth, and $\sigma_*^2 < \infty$ by (5). The first bound then follows from Cortild et al. (2026, Theorem 4.1), using $\frac{L_\star\|\beta^0-\beta^\star\|^2}{2\eta L_\star T + 2(1-\eta L_\star)} \leq \frac{\|\beta^0-\beta^\star\|^2}{2\eta T}$, and the second from Attia et al. (2025, Theorem 2). □

Remark 1. In contrast to the previous bounds, the variance term in the last-iterate guarantee grows with T. This mirrors earlier results for decaying learning rates, which show that the last SGD iterate has a strictly worse error rate than the average iterate (Harvey et al., 2019). However, as noted in Garrigos et al. (2025, Lemma 3.2), if the learning rate is sufficiently small relative to the number of iterations, namely $\eta \leq \frac{1}{L_\star \log_2(T+2)}$, we obtain the better-behaved bound

$$\mathbb{E}_{\beta^\star}\left[\mathcal{L}(\beta^T)\right] - \mathcal{L}(\beta^\star) \leq \frac{12\|\beta^0 - \beta^\star\|^2}{\eta T} + \frac{48}{L_\star\sigma_*^2}$$

Although this constraint on η requires prior knowledge the training horizon, and thus departs from a strictly online setting, it is easy to satisfy in practice due to the slow growth of $\log_2(T+2)$.

Remark 2. For the GLM with a Poisson base distribution (i.e. Poisson regression), we have $\Phi_f''(\theta) = \lambda e^\theta$, where λ is the Poisson parameter. In particular Φ_f is neither smooth nor strongly convex and the above results do not apply. In fact, it was observed that, in this setting, SGD is highly sensitive to the choice of learning rate, leading to numerical instability in practice (Toulis et al., 2014, Lemma 4.2 & §5.1; Toulis et al., 2021, §4.1.2 & §5.2). This sensitivity can be understood by noting that, for smooth objectives, the learning rate should scale on the order of the reciprocal of the

smoothness parameter, i.e., $1/N$. In Poisson regression, however, the *local* smoothness parameter at β is $\lambda e^{\langle \beta, x \rangle}$, so that any fixed learning rate, being independent of the current iterate, may be too large by a factor that grows exponentially with $\langle \beta, x \rangle$. Implicit SGD is a natural alternative to SGD (3) for generalized linear regression; see Toulis et al. (2014; 2021) for details and references.

4 Monotonicity of SGD updates

In this section, we restrict attention to the setting of Example 3, in which outcomes arise from pairwise comparisons. At each time step t, a pair of alternatives $(a_t, b_t) \in \mathcal{A}^2$ is drawn according to the distribution Q. The outcome r^t follows the exponential-family distribution with base distribution f and natural parameter $\langle \beta^\star, x_{a_t} - x_{b_t} \rangle$, where $\beta^\star \in \mathbb{R}^D$ is the true parameter vector and $x_a \in \mathbb{R}^D$ denotes the feature vector of alternative $a \in \mathcal{A}$. Let $\theta_a^t \triangleq \langle \beta^t, x_a \rangle$ denote the estimated strength of alternative $a \in \mathcal{A}$ at time t. For any $v \in \mathbb{R}^{\mathcal{A}}$, we write $v_{a \ominus b} \triangleq v_a - v_b$, so that $\theta^t_{a \ominus b}$ denotes the *strength gap* between alternatives a and b. Taking the inner product of (3) with x_c shows that the SGD iteration induces, for each $c \in \mathcal{A}$, the strength update

$$\theta_c^t = \theta_c^{t-1} + \eta \left[r^t - \Phi_f'\big(\theta^{t-1}_{a_t \ominus b_t}\big) \right] \langle x_{a_t} - x_{b_t}, x_c \rangle. \tag{7}$$

The sign of the residual $\delta^t \triangleq r^t - \Phi_f'\big(\theta^{t-1}_{a_t \ominus b_t}\big)$ indicates whether the outcome is above or below that predicted by the current strength gap. Monotonicity requires the resulting changes in strengths or strength gaps to align with the sign of the residual. We distinguish three notions according to whether this requirement concerns an individual strength, the gap between the compared alternatives, or the gaps between a compared alternative and every other alternative. One motivation is that failures are commonly observed in the RLHF literature and generally deemed undesirable: even when a is preferred over b, the update can decrease the score of a (Chen et al., 2024; Fageot et al., 2025).

4.1 Definition and geometric conditions

It is convenient to first isolate the one-step update underlying the stochastic recursion (7). Accordingly, for an arbitrary strength vector $\theta \in \mathbb{R}^{\mathcal{A}}$ and observation tuple $(a, b, r) \in \mathcal{A}^2 \times \mathbb{R}$, the update operator $\mathcal{U}_{a,b,r} : \mathbb{R}^{\mathcal{A}} \to \mathbb{R}^{\mathcal{A}}$ is defined, for $c \in \mathcal{A}$, by

$$\big[\mathcal{U}_{a,b,r}(\theta)\big]_c \triangleq \theta_c + \eta\, \delta(\theta; a, b, r) \langle x_a - x_b, x_c \rangle, \tag{8}$$

where $\delta(\theta; a, b, r) \triangleq r - \Phi_f'(\theta_{a \ominus b})$. Monotonicity comes in three variants that we now define.

Definition 2. Let $P \subseteq \mathcal{A}^2$ be a comparison set. The update rule $\mathcal{U}$ is said to be

- *individually monotone on* P if $\delta(\theta; a, b, r)[\mathcal{U}_{a,b,r}(\theta) - \theta]_a \geq 0$;
- *weakly gap monotone on* P if $\delta(\theta; a, b, r)[\mathcal{U}_{a,b,r}(\theta) - \theta]_{a \ominus b} \geq 0$;
- *gap monotone on* P if $\delta(\theta; a, b, r)[\mathcal{U}_{a,b,r}(\theta) - \theta]_{a \ominus c} \geq 0$ for every $c \in \mathcal{A}$,

where each condition is required for every $\theta \in \mathbb{R}^{\mathcal{A}}$, every $(a, b) \in P$, and every $r \in \mathbb{R}$. When $P = \mathcal{A}^2$, we omit the qualifier "on P".

Weak gap monotonicity constrains only the gap between the two compared alternatives, whereas gap monotonicity constrains the gap between a and every alternative $c \in \mathcal{A}$, compared or not. In particular, gap monotonicity implies weak gap monotonicity, by taking $c = b$. We now show that weak gap monotonicity holds for every configuration of feature vectors, while individual and gap monotonicity are characterized by simple, easily checked geometric conditions on this configuration.

Theorem 1. *Let $P \subseteq \mathcal{A}^2$. The update rule $\mathcal{U}$ is weakly gap monotone on P. Moreover, it is*

1. *individually monotone on P iff $\langle x_a - x_b, x_a \rangle \geq 0$ for every $(a, b) \in P$;*
2. *gap monotone on P iff $\langle x_a - x_b, x_a - x_c \rangle \geq 0$ for every $(a, b) \in P$ and $c \in \mathcal{A}$.*

Theorem 1 is proved in Appendix D.1. When $P = \mathcal{A}^2$, the condition in Item 2 has a particularly simple geometric meaning: no triple of feature vectors forms an obtuse angle. The condition in Item 1 similarly constrains the angle formed by two feature vectors together with a "phantom" alternative whose feature vector is the origin of $\mathbb{R}^D$.

Remark 3. When $P = \mathcal{A}^2$, both conditions also admit a linear-algebraic formulation via the Gram matrix G of the feature vectors, that is, the matrix with entries $G_{ab} = \langle x_a, x_b \rangle$ for $(a, b) \in \mathcal{A}^2$. The condition in Item 1 becomes the entrywise dominance condition $G_{aa} \geq G_{ab}$ for $(a, b) \in \mathcal{A}^2$, and the condition in Item 2 becomes $G_{aa} + G_{bc} - G_{ac} - G_{ab} \geq 0$ for $(a, b, c) \in \mathcal{A}^3$. Both are systems of homogeneous linear inequalities, so each defines a polyhedral convex cone in $\mathbb{R}^{\mathcal{A}\times\mathcal{A}}$.

Example 5. A sufficient condition for individual monotonicity is that the feature vectors all lie on a sphere of radius ρ centered at the origin. Indeed, in this case the Cauchy–Schwarz inequality gives $\langle x_a, x_b \rangle \leq \rho^2 = \langle x_a, x_a \rangle$ for every $(a, b) \in \mathcal{A}^2$, which is precisely the condition in Item 1.

Danzer & Grünbaum (1962, Theorem II a) proved a conjecture of Erdős,[1] showing that the maximum cardinality of a collection of points in $\mathbb{R}^D$ such that no three of them form an obtuse angle is 2^D. Combined with Theorem 2, this yields the following necessary condition for gap monotonicity.

Corollary 1. *If the feature vectors are pairwise distinct and $\mathcal{U}$ is gap monotone, then $|\mathcal{A}| \leq 2^D$. This bound is attained by taking the feature vectors to be the 2^D vertices of a hyperrectangle.*

Pathwise monotonicity. The notions of monotonicity defined in Definition 2 quantify universally over the tuple (a, b, r) and the strength vector θ. We now turn to their pathwise counterparts, which put to the test only the realized triples (a_t, b_t, r^t) and the corresponding iterates θ^{t-1}. Accordingly, we say that step t of the SGD trajectory is individually monotone, weakly gap monotone, or gap monotone when the corresponding inequality of Definition 2 holds for $(\theta, a, b, r) = (\theta^{t-1}, a_t, b_t, r^t)$, and that the SGD trajectory is monotone, in any of these three senses, when all its steps is. Since the strength sequence induced by SGD satisfies $\theta^t = \mathcal{U}_{a_t,b_t,r^t}(\theta^{t-1})$, this amounts to a non-negativity constraint on suitable coordinates of the vector $\delta^t(\theta^t - \theta^{t-1})$, and is thus a property of the realized trajectory rather than of $\mathcal{U}$. As soon as f is non-degenerate (i.e. not a Dirac measure), however, the following dichotomy shows that pathwise monotonicity is no weaker than monotonicity of $\mathcal{U}$ on $\operatorname{supp} Q$: either the latter holds and every trajectory is monotone, or it fails and almost every trajectory violates monotonicity at a positive fraction of its steps. The proof is given in Appendix D.1.

Proposition 3. *Let f be non-degenerate. For each monotonicity notion in Definition 2:*

1. *if $\mathcal{U}$ is monotone on $\operatorname{supp} Q$, then every realization of the SGD trajectory is monotone;*
2. *otherwise, almost surely, monotonicity fails along the trajectory at a positive asymptotic fraction of steps, and in particular at infinitely many steps.*

Example 6 (Class-based features)**.** For GBT (Example 2), individual and gap monotonicity clearly hold when $x = I$. We show in Appendix D.2 that gap and individual monotonicity in fact extends to a broader class of features, obtained by grouping alternatives according to one or several categorical features (e.g. gender, country of origin, age group). This provides a non-trivial example for which gap and individual monotonicity are systematically satisfied.

4.2 Asymptotic monotonicity for sub-Gaussian random features

Recall from Corollary 1 that for pairwise distinct feature vectors, gap monotonicity of $\mathcal{U}$ requires $|\mathcal{A}| \leq 2^D$. Moreover, when $|\mathcal{A}| = 2^D$, the only configurations that induce gap monotonicity are the vertices of a hyperrectangle.[2] This could suggest that monotonicity-inducing configurations are rather rare. The following theorem shows otherwise: a random configuration of well-behaved,[3] pairwise independent sub-Gaussian feature vectors induces both individual and gap monotonicity with high probability, as soon as D exceeds a constant multiple of $\log|\mathcal{A}|$. Up to constants, the dimension needed for monotonicity to be typical thus matches that needed for it to be possible at all.

Theorem 2. *Let $(x_a)_{a\in\mathcal{A}}$ be pairwise independent, centered random feature vectors in $\mathbb{R}^D$. Assume that one of the following three conditions holds for every $a \in \mathcal{A}$, and let ρ be as indicated.*

[1] An elementary exposition can be found in (Aigner & Ziegler, 2018, Chapter 17).

[2] While not stated explicitly by Danzer & Grünbaum (1962), this is immediate from their results: a set of 2^D points in $\mathbb{R}^D$ with no obtuse angle is antipodal (Theorem I a), hence is the vertex set $\{v + \sum_{i\in I} u_i \mid I \subseteq [D]\}$ of a parallelotope (Theorem II bβ). Non-obtuseness of the angles $(v+u_i, v, v+u_j)$ and $(v, v+u_i, v+u_i+u_j)$ then forces $\langle u_i, u_j \rangle = 0$ for $i \neq j$; hence the edges are pairwise orthogonal.

[3] The theorem is stated for several common cases; Remark 4 discusses more general conditions under which the same conclusion holds.

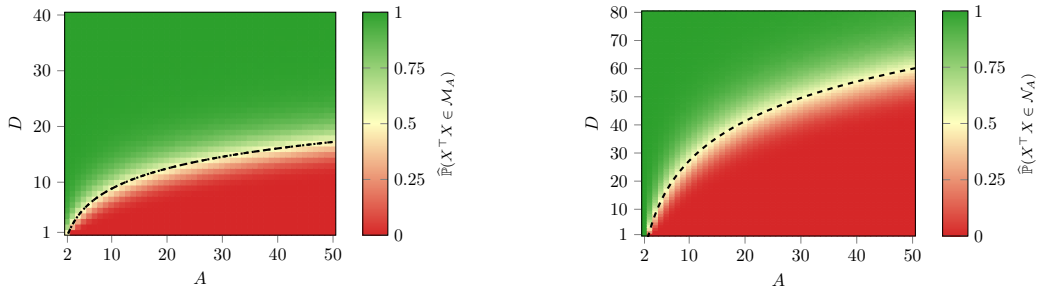


Figure 1: Probability that a family of A independent standard normal feature vectors in dimension D induce individual monotonicity (left) and gap monotonicity (right). The probabilities are estimated via Monte Carlo sampling (10^4 draws per cell). The dashed lines show the pairs (A, D) such that $A^{c_1} e^{c_2 D} = 0.5$, with $c_1 = -1.14578$, $c_2 = 0.21996$ (left), and $c_1 = -0.719920$, $c_2 = 0.035410$ (right).

1. *the coordinates of x_a are independent, with variance at least σ^2 and sub-Gaussian norm at most K; here $\rho \triangleq \sigma^2/K^2$;*
2. *$x_a \sim \mathcal{N}(0, \Sigma)$ for a positive definite $\Sigma \in \mathbb{R}^{D\times D}$ with condition number $\kappa(\Sigma)$; here $\rho \triangleq 1/\kappa(\Sigma)$;*
3. *$\|x_a\| = \sqrt{D}$ almost surely and the sub-Gaussian norm of x_a is at most K; here $\rho \triangleq 1/K^2$.*

Then there exists some universal constant $c > 0$ such that, with $A \triangleq |\mathcal{A}|$,

$$\mathbb{P}(\mathcal{U} \text{ is individually and gap monotone}) \geq 1 - 2A^2 e^{-c\rho^2 D}.$$

Thus, if $c\rho^2 D - 2\log A \to \infty$, in particular if ρ is bounded away from 0 and $D = \omega(\log A)$, then

$$\lim_{A\to\infty} \mathbb{P}(\mathcal{U} \text{ is individually and gap monotone}) = 1.$$

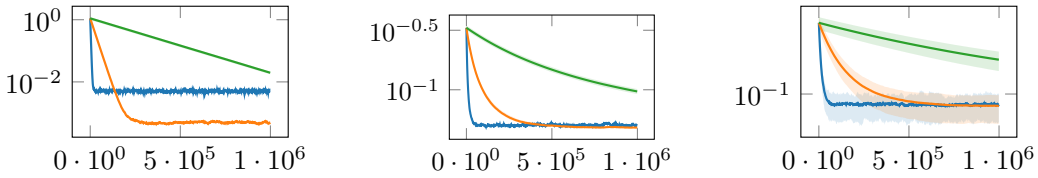
Theorem 2 follows from standard concentration of measure arguments: in high dimension, independent random vectors are nearly orthogonal and, under our assumptions, have squared norms of order D, so with high probability every $\|x_a\|^2$ dominates all cross inner products $\langle x_a, x_b\rangle$. The configuration is then close to a scaled orthonormal family, in which every triangle is nearly equilateral, hence non-obtuse. The proof is given in Appendix D.3, and Fig. 1 illustrates the $D = \Theta(\log A)$ scaling.

5 Experiments

In this Section, we provide numerical experiments that illustrate our theoretical framework. We observe sublinear and linear convergence to a neighborhood of the minimizer as guaranteed by Propositions 1 and 2, and we illustrate how the convergence rate depends on the pairing distribution.

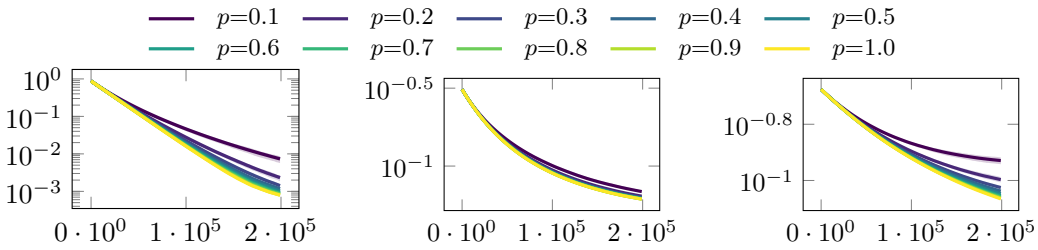
Convergence for varying base distributions. We run SGD (3) on the GBT comparison model of Example 2 with $|\mathcal{A}| = 100$ alternatives and uniform pair sampling. We consider three base distributions f (Gaussian, Bradley-Terry, Uniform) and an identity embedding. For each, we run SGD with three constant step sizes $\eta \in \{10^{-4}, 10^{-3}, 10^{-2}\}$ for $T = 10^6$ iterations, starting from $\beta^0 = 0$, and report the last-iterate suboptimality $\mathcal{L}(\beta^T) - \mathcal{L}(\beta^\star)$ and error $\|\beta^T - \beta^\star\|$ over 10 repetitions. Each repetition draws a ground truth $\beta^\star \sim \mathcal{N}(0, I)$, and observations are then sampled *i.i.d.* from $\mathbb{P}_{\beta^\star}$. Throughout, $\mathcal{L}$ denotes the population risk under uniform pair sampling, which shares its minimizer with the risk under any connected Q, and since scores are identified up to a common shift, distances are computed after projecting onto $\mathbf{1}^\perp$.

Across all three base distributions (Figs. 2 and 4), the curves exhibit the bias/variance decomposition predicted by Propositions 1 and 2: an initial transient phase in which the iterates approach $\beta^\star$, followed by a stationary regime fluctuating around a noise floor whose magnitude grows with η. The largest step size ($\eta = 10^{-2}$) reaches the steady state in a few hundred iterations but plateaus at a comparatively high error, while the smallest step size ($\eta = 10^{-4}$) is still in the transient phase at $T = 10^6$ but is on track for a smaller asymptotic error. The Gaussian panel illustrates the linear bias decay of Proposition 1, whereas the bounded-support laws decay sublinearly, in line with Proposition 2.



(a) Gaussian (b) Bradley–Terry (c) Uniform

Figure 2: Last-iterate suboptimality $\mathcal{L}(\beta^T) - \mathcal{L}(\beta^\star)$ of SGD with constant step size $\eta \in \{10^{-4}, 10^{-3}, 10^{-2}\}$, for three base distributions. GBT model with $|\mathcal{A}| = 100$ alternatives, uniform pair sampling, 10 repetitions (solid curve is mean; fill between shows first and third quartiles).



(a) Gaussian (b) Bradley-Terry (c) Uniform

Figure 3: Last-iterate suboptimality $\mathcal{L}(\beta^T) - \mathcal{L}(\beta^\star)$ of SGD with constant step size, as the pair distribution Q ranges over Erdős-Rényi graphs $G(|\mathcal{A}|, p)$ for $p \in \{0.1, \ldots, 1.0\}$, for three base distributions. GBT model with $|\mathcal{A}| = 100$ alternatives, step size $\eta = 10^{-3}$, 100 repetitions (solid curve is mean; fill between shows first and third quartiles).

Dependence on the comparison graph. The convergence bounds derived in Sections 3.1 and 3.2 depend on the pair sampling distribution Q only through the spectrum of its Laplacian L_Q: the rate is governed by the algebraic connectivity $\lambda^+_{\min}(L_Q)$, while the asymptotic floor scales with $\lambda_{\max}(L_Q)$ (cf. the discussion after Proposition 1). To probe how this prediction plays out in practice, we sample pair distributions from an Erdős-Rényi model $G(|\mathcal{A}|, p)$ with edge probability $p \in \{0.1, 0.2, \ldots, 1.0\}$, and sample pairs uniformly from the edges of the realized graph (all sampled graphs were connected). The case $p = 1$ recovers the uniform pair sampling of the previous experiment, while smaller p produces sparser graphs with smaller algebraic connectivity. The ground truth $\beta^\star \sim \mathcal{N}(0, I)$ is drawn once and shared by all runs; for each p we run $2 \cdot 10^5$ SGD iterations on 100 independent graphs.

Figure 3 reports the last-iterate suboptimality $\mathcal{L}(\beta^T) - \mathcal{L}(\beta^\star)$ as a function of the iteration count, for the same three base distributions and a fixed step size $\eta = 10^{-3}$. Curves are ordered monotonically by p, in line with theory: sparser graphs (smaller p) yield slower decay, with the gap most pronounced for $p \leq 0.2$ and shrinking as p approaches 1. The effect is dramatic in the Gaussian panel, where the curves spread over almost an order of magnitude—reflecting the linear bias decay of Proposition 1 and its dependence on the algebraic connectivity $\lambda^+_{\min}(L_Q)$—and is more modest, though still visible, for the bounded-support Bradley-Terry and Uniform laws, in line with the slower sublinear regime of Proposition 2. The qualitative behavior is consistent across base distributions, suggesting that the dependence on the pair distribution is largely decoupled from the choice of base measure, in line with the analysis.

AI USE STATEMENT

In this work, we used generative AI tools to formulate mathematical claims, assist in the writing of proofs, implement methods, get assistance with translation, support qualitative and thematic data analysis, and provide feedback on the experiments, namely checking the consistency between the experimental code and its description in the paper. We have not used generative AI tools to provide critical ingredients for proving mathematical claims, generate synthetic data sets, help develop theoretical models or conceptual frameworks, propose or refine hypotheses, design the research methodology or experiments, or interpret results. Cleaning and reformatting datasets are not applicable to this work. Additionally, we used generative AI tools to create or modify scientific figures or images, create or edit software code, draft parts of a research paper, summarize or analyse existing literature, discover research topics or identify gaps, and propose a title or keywords for a research paper. We have reviewed all AI-assisted work. In particular, the code provided in the supplementary material was verified by checking that it regenerates the data used in the paper's figures exactly. We take responsibility for the final content of this work, including text, claims or artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work is theoretical and its experiments use only synthetic data; it involves no human subjects, collects no personal data, and releases no dataset or trained model. Comparison-based rating systems such as Elo are used to rank people (players, workers, students, content creators) and to fine-tune language models from human preferences. Their behavior thus affects their final users. Our monotonicity results identify when a comparison-based update may move a score against the observed outcome, e.g. lowering an item's score after it is preferred, which is a potential source of unfair or manipulable rankings, and we view characterizing this behavior as a contribution to the reliability of such systems. These guarantees hold under the assumptions stated with each result (Propositions 1 and 2 and Section 4). They should not be read as fairness or safety guarantees for deployed rating or alignment pipelines. We are not aware of any conflict of interest.

REPRODUCIBILITY STATEMENT

All theoretical results are stated together with their assumptions in the main text (Assumption 1, Propositions 1 and 2, and Section 4); complete proofs are given in Appendix B for the convergence results and in Appendix D for the monotonicity results. The numerical experiments use synthetic data only; their setup (comparison model, base distributions, pair distributions, step sizes, iteration budgets and number of repetitions) is described in Section 5 and in the captions of Figures 1 to 3, 4 and 5. The supplementary material contains an anonymized Julia package with one script per experiment, which regenerates the data behind every figure, and a LaTeX file that renders the figures from these data. All random number generators are explicitly seeded and package versions are pinned, so that the provided environment reproduces the data used in the paper exactly; the full set of experiments runs in about 20 minutes on a laptop CPU.

REFERENCES

Arpit Agarwal, Prathamesh Patil, and Shivani Agarwal. Accelerated Spectral Ranking. In *Proceedings of the 35th International Conference on Machine Learning*, pp. 70–79. PMLR, July 2018.

Martin Aigner and Günter M. Ziegler. *Proofs from THE BOOK*. Springer, Berlin, Heidelberg, 6 edition, 2018. ISBN 978-3-662-57264-1. doi: 10.1007/978-3-662-57265-8.

Amit Attia, Matan Schliserman, Uri Sherman, and Tomer Koren. Fast last-iterate convergence of SGD in the smooth interpolation regime. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, October 2025. URL https://openreview.net/forum?id=ifGwVxiigF.

Hossein Azari Soufiani, David Parkes, and Lirong Xia. A statistical decision-theoretic framework for social choice. In *Advances in Neural Information Processing Systems*, volume 27. Curran Associates, Inc., 2014. URL https://proceedings.neurips.cc/paper_files/paper/2014/hash/405bfe28e212b7e1d100f013aae9b2d1-Abstract.html.

Francis Bach. Self-concordant analysis for logistic regression. *Electronic Journal of Statistics*, 4: 384–414, 2010. ISSN 1935-7524, 1935-7524. doi: 10.1214/09-EJS521.

Francis Bach. Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression. *The Journal of Machine Learning Research*, 15(1):595–627, 2014.

Gilles Bareilles, Julien Fageot, Lê-Nguyên Hoang, Peva Blanchard, Wassim Bouaziz, Sébastien Rouault, and El-Mahdi El-Mhamdi. On monotonicity in AI alignment. *arXiv preprint arXiv:2506.08998*, 2025.

Ole E. Barndorff-Nielsen. *Information and Exponential Families: In Statistical Theory*. John Wiley & Sons, 2014. ISBN 978-1-118-85750-2. doi: 10.1002/9781118857281.

Yoshua Bengio, Stephen Clare, Carina Prunkl, Maksym Andriushchenko, Ben Bucknall, Malcolm Murray, Rishi Bommasani, Stephen Casper, Tom Davidson, Raymond Douglas, David Duvenaud, Philip Fox, Usman Gohar, Rose Hadshar, Anson Ho, Tiancheng Hu, Cameron Jones, Sayash Kapoor, Atoosa Kasirzadeh, Sam Manning, Nestor Maslej, Vasilios Mavroudis, Conor McGlynn, Richard Moulange, Jessica Newman, Kwan Yee Ng, Patricia Paskov, Shalaleh Rismani, Girish Sastry, Elizabeth Seger, Scott Singer, Charlotte Stix, Lucia Velasco, Nicole Wheeler, Daron Acemoglu, Vincent Conitzer, Thomas G. Dietterich, Fredrik Heintz, Geoffrey Hinton, Nick Jennings, Susan Leavy, Teresa Ludermir, Vidushi Marda, Helen Margetts, John McDermid, Jane Munga, Arvind Narayanan, Alondra Nelson, Clara Neppel, Sarvapali D. Ramchurn, Stuart Russell, Marietje Schaake, Bernhard Schölkopf, Alvaro Soto, Lee Tiedrich, Gaël Varoquaux, Andrew Yao, Ya-Qin Zhang, Leandro Angelo Aguirre, Olubunmi Ajala, Fahad Albalawi, Noora AlMalek, Christian Busch, Jonathan Collas, André Carlos Ponce de Leon Ferreira de Carvalho, Amandeep Gill, Ahmet Halit Hatip, Juha Heikkilä, Chris Johnson, Gill Jolly, Ziv Katzir, Mary N. Kerema, Hiroaki Kitano, Antonio Krüger, Kyoung Mu Lee, José Ramón López Portillo, Aoife McLysaght, Oleksii Molchanovskyi, Andrea Monti, Mona Nemer, Nuria Oliver, Raquel Pezoa, Audrey Plonk, Balaraman Ravindran, Hammam Riza, Crystal Rugege, Haroon Sheikh, Denise Wong, Yi Zeng, Liming Zhu, Daniel Privitera, and Sören Mindermann. International ai safety report 2026, 2026. URL https://arxiv.org/abs/2602.21012.

Vladimir I. Bogachev. *Measure Theory*, volume 2. Springer, Berlin, Heidelberg, 2007. ISBN 978-3-540-34513-8. doi: 10.1007/978-3-540-34514-5.

Léon Bottou, Frank E Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *SIAM review*, 60(2):223–311, 2018.

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

Chris Burges, Tal Shaked, Erin Renshaw, Ari Lazier, Matt Deeds, Nicole Hamilton, and Greg Hullender. Learning to rank using gradient descent. In *Proceedings of the 22nd International Conference on Machine Learning - ICML '05*, pp. 89–96. ACM Press, 2005. ISBN 978-1-59593-180-1. doi: 10.1145/1102351.1102363. URL http://portal.acm.org/citation.cfm?doid=1102351.1102363.

François Caron and Arnaud Doucet. Efficient Bayesian Inference for Generalized Bradley–Terry Models. *Journal of Computational and Graphical Statistics*, 21(1):174–196, January 2012. ISSN 1061-8600, 1537-2715. doi: 10.1080/10618600.2012.638220.

Angelica Chen, Sadhika Malladi, Lily H Zhang, Xinyi Chen, Qiuyi Zhang, Rajesh Ranganath, and Kyunghyun Cho. Preference learning algorithms do not learn preference rankings. *Advances in Neural Information Processing Systems*, 37:101928–101968, 2024.

Paul F. Christiano, Jan Leike, Tom B. Brown, Miljan Martic, Shane Legg, and Dario Amodei. Deep reinforcement learning from human preferences. In Isabelle Guyon, Ulrike von Luxburg, Samy Bengio, Hanna M. Wallach, Rob Fergus, S. V. N. Vishwanathan, and Roman Garnett (eds.), *Advances in Neural Information Processing Systems 30: Annual Conference on Neural Information Processing Systems 2017, December 4-9, 2017, Long Beach, CA, USA*, pp. 4299–4307, 2017. URL https://proceedings.neurips.cc/paper/2017/hash/d5e2c0adad503c91f91df240d0cd4e49-Abstract.html.

Sylvain Le Corff, Matthieu Lerasle, and Elodie Vernet. A Bayesian nonparametric approach for generalized Bradley–Terry models in random environment. *arXiv preprint arXiv:1808.08104*, 2018.

Daniel Cortild, Lucas Ketels, Juan Peypouquet, and Guillaume Garrigos. Bias-optimal bounds for SGD: A computer-aided Lyapunov analysis, January 2026.

Rémi Coulom. Whole-history rating: A Bayesian rating system for players of time-varying strength. In H. Jaap van den Herik, Xinhe Xu, Zongmin Ma, and Mark H. M. Winands (eds.), *Computers and Games*, number 5131 in Lecture Notes in Computer Science, pp. 113–124, Berlin, Heidelberg, 2008. Springer. ISBN 978-3-540-87608-3. doi: 10.1007/978-3-540-87608-3_11.

Douglas E. Critchlow and Michael A. Fligner. Paired comparison, triple comparison, and ranking experiments as generalized linear models, and their implementation on GLIM. *Psychometrika*, 56(3):517–533, September 1991. ISSN 1860-0980. doi: 10.1007/BF02294488.

L. Danzer and B. Grünbaum. Über zwei Probleme bezüglich konvexer Körper von P. Erdös und von V. L. Klee. *Mathematische Zeitschrift*, 79:95–99, 1962. ISSN 0025-5874; 1432-1823.

Roger R Davidson. On extending the Bradley–Terry model to accommodate ties in paired comparison experiments. *Journal of the American Statistical Association*, 65(329):317–328, 1970.

Roland Diel, Sylvain Le Corff, and Matthieu Lerasle. Learning the distribution of latent variables in paired comparison models with round-robin scheduling. *Bernoulli*, 26(4):2670–2698, November 2020. ISSN 1350-7265. doi: 10.3150/20-BEJ1203.

Julien Fageot, Sadegh Farhadkhani, Lê-Nguyên Hoang, and Oscar Villemaud. Generalized Bradley–Terry Models for Score Estimation from Paired Comparisons. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(18):20379–20386, March 2024. ISSN 2374-3468. doi: 10.1609/aaai.v38i18.30020.

Julien Fageot, Peva Blanchard, Gilles Bareilles, and Lê-Nguyên Hoang. Generalizing while preserving monotonicity in comparison-based preference learning models. In *Advances in Neural Information Processing Systems*, volume 38, pp. 84380–84405. Curran Associates, Inc., October 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/79b90b4c2ee23cc35fdd8de2969dc4e8-Paper-Conference.pdf.

Jianqing Fan, Jikai Hou, and Mengxin Yu. Uncertainty quantification of MLE for entity ranking with covariates. *Journal of Machine Learning Research*, 25(358):1–83, 2024. ISSN 1533-7928. URL http://jmlr.org/papers/v25/23-0554.html.

Shuxing Fang, Ruijian Han, Yuanhang Luo, and Yiming Xu. Recent advances in the Bradley–Terry Model: Theory, algorithms, and applications, January 2026.

Peter C. Fishburn. Paradoxes of preferential voting. *Mathematics Magazine*, 56(4):207–214, 1983.

Rachel Freedman, Jana Schaich Borg, Walter Sinnott-Armstrong, John P. Dickerson, and Vincent Conitzer. Adapting a kidney exchange algorithm to align with human values. *Artificial Intelligence*, 283, 2020.

Alfred Galichon. *Discrete Choice Models: Mathematical Methods, Econometrics, and Data Science*. Princeton University Press, Princeton, 2026. ISBN 978-0-691-22180-9.

Guillaume Garrigos and Robert M. Gower. Handbook of Convergence Theorems for (Stochastic) Gradient Methods, March 2024.

Guillaume Garrigos, Daniel Cortild, Lucas Ketels, and Juan Peypouquet. Last-Iterate Complexity of SGD for Convex and Smooth Stochastic Problems, July 2025.

Chris Godsil and Gordon Royle. *Algebraic Graph Theory*. Number 207 in Graduate Texts in Mathematics. Springer, New York, NY, 2001. ISBN 978-0-387-95220-8. doi: 10.1007/978-1-4613-0163-9.

Ian Hamilton, Nick Tawn, and David Firth. The many routes to the ubiquitous Bradley-Terry model, August 2025.

Ruijian Han, Yiming Xu, and Kani Chen. A General Pairwise Comparison Model for Extremely Sparse Networks. *Journal of the American Statistical Association*, 118(544):2422–2432, October 2023. ISSN 0162-1459, 1537-274X. doi: 10.1080/01621459.2022.2053137.

Nicholas J. A. Harvey, Christopher Liaw, Yaniv Plan, and Sikander Randhawa. Tight analyses for non-smooth stochastic gradient descent. In Alina Beygelzimer and Daniel Hsu (eds.), *Proceedings of the Thirty-Second Conference on Learning Theory*, volume 99 of *Proceedings of Machine Learning Research*, pp. 1579–1613. PMLR, 25–28 Jun 2019. URL https://proceedings.mlr.press/v99/harvey19a.html.

Julien Hendrickx, Alex Olshevsky, and Venkatesh Saligrama. Minimax rate for learning from pairwise comparisons in the BTL model. In Hal Daumé III and Aarti Singh (eds.), *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *Proceedings of Machine Learning Research*, pp. 4193–4202. PMLR, 2020. URL https://proceedings.mlr.press/v119/hendrickx20a.html.

Ralf Herbrich, Tom Minka, and Thore Graepel. Trueskill: A Bayesian skill rating system. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 19, pp. 569–576, 2006.

David R. Hunter. MM algorithms for generalized Bradley–Terry models. *Annals of Statistics*, 32 (1):384–406, 2004. doi: 10.1214/aos/1079120141.

Ming Jin and Hyunin Lee. Position: AI safety must embrace an antifragile perspective. In *Forty-second International Conference on Machine Learning Position Paper Track*, 2025. URL https://openreview.net/forum?id=WpePuya3Ki.

Franz J Király and Zhaozhi Qian. Modelling competitive sports: Bradley-Terry-Élő models for supervised and on-line learning of paired competition outcomes. *arXiv preprint arXiv:1701.08055*, 2017.

Erich Leo Lehmann and George Casella. *Theory of Point Estimation*. Springer Texts in Statistics. Springer, 2 edition, 1998. ISBN 978-0-387-98502-2. doi: 10.1007/b98854.

Wanshan Li, Shamindra Shrotriya, and Alessandro Rinaldo. ℓ^∞-bounds of the MLE in the BTL model under general comparison graphs. In James Cussens and Kun Zhang (eds.), *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, volume 180 of *Proceedings of Machine Learning Research*, pp. 1178–1187. PMLR, 2022. URL https://proceedings.mlr.press/v180/li22g.html.

Robert Duncan Luce. *Individual Choice Behavior: A Theoretical Analysis*. Wiley & Sons, New York, 1959. ISBN 978-0-486-44136-8. URL https://archive.org/details/dli.scoerat.3346ndividualchoicebehavior.

Lucas Maystre and Matthias Grossglauser. Fast and Accurate Inference of Plackett– Luce Models. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.

Steven Morse. Elo as a statistical learning model, May 2019. URL https://stmorse.github.io/journal/Elo.html.

Hervé Moulin. *Axioms of Cooperative Decision Making*. Econometric Society Monographs. Cambridge University Press, 1988.

Sahand Negahban, Sewoong Oh, and Devavrat Shah. Rank Centrality: Ranking from Pairwise Comparisons. *Operations Research*, 65(1):266–287, February 2017. ISSN 0030-364X, 1526-5463. doi: 10.1287/opre.2016.1534.

John Ashworth Nelder and Robert William Maclagan Wedderburn. Generalized linear models. *Journal of the Royal Statistical Society*, 135(3):370–384, May 1972. ISSN 0035-9238. doi: 10.2307/2344614.

M. E. J. Newman. Efficient Computation of Rankings from Pairwise Comparisons. *Journal of Machine Learning Research*, 24(238):1–25, 2023. ISSN 1533-7928.

Ritesh Noothigattu, Dominik Peters, and Ariel Procaccia. Axioms for learning from pairwise comparisons. In *Advances in Neural Information Processing Systems*, volume 33, pp. 17745–17754. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/hash/cdaa9b682e10c291d3bbadca4c96f5de-Abstract.html.

Sam Olesker-Taylor and Luca Zanetti. An analysis of Elo rating systems via Markov chains, 2024. URL https://arxiv.org/abs/2406.05869.

Arka Pal, Deep Karkhanis, Samuel Dooley, Manley Roberts, Siddartha Naidu, and Colin White. Smaug: Fixing failure modes of preference optimisation with dpo-positive. *CoRR*, abs/2402.13228, 2024. doi: 10.48550/ARXIV.2402.13228. URL https://doi.org/10.48550/arXiv.2402.13228.

Robin Lyman Plackett. The analysis of permutations. *Journal of the Royal Statistical Society Series C: Applied Statistics*, 24(2):193–202, June 1975. ISSN 0035-9254. doi: 10.2307/2346567.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. In A. Oh, T. Naumann, A. Globerson, K. Saenko, M. Hardt, and S. Levine (eds.), *Advances in Neural Information Processing Systems*, volume 36, pp. 53728–53741. Curran Associates, Inc., 2023. URL https://proceedings.neurips.cc/paper_files/paper/2023/file/a85b405ed65c6477a4fe8302b5e06ce7-Paper-Conference.pdf.

Poduri Venkateswara Rao and Lawrence Lee Kupper. Ties in paired-comparison experiments: A generalization of the Bradley–Terry model. *Journal of the American Statistical Association*, 62 (317):194–204, 1967.

Noam Razin, Sadhika Malladi, Adithya Bhaskar, Danqi Chen, Sanjeev Arora, and Boris Hanin. Unintentional unalignment: Likelihood displacement in direct preference optimization. *CoRR*, abs/2410.08847, 2024. doi: 10.48550/ARXIV.2410.08847. URL https://doi.org/10.48550/arXiv.2410.08847.

Nihar B. Shah, Sivaraman Balakrishnan, Joseph Bradley, Abhay Parekh, Kannan Ramchandran, and Martin J. Wainwright. Estimation from pairwise comparisons: Sharp minimax bounds with topology dependence. *Journal of Machine Learning Research*, 17(58):1–47, 2016. URL http://jmlr.org/papers/v17/15-189.html.

Nihar B. Shah, Sivaraman Balakrishnan, Adityanand Guntuboyina, and Martin J. Wainwright. Stochastically transitive models for pairwise comparisons: Statistical and computational issues. *IEEE Transactions on Information Theory*, 63(2):934–959, February 2017. ISSN 0018-9448, 1557-9654. doi: 10.1109/TIT.2016.2634418. URL http://ieeexplore.ieee.org/document/7763782/.

Dianiel A. Spielman. Spectral and algebraic graph theory. Book draft, April 2025. URL http://cs-www.cs.yale.edu/homes/spielman/sagt/.

Nisan Stiennon, Long Ouyang, Jeffrey Wu, Daniel M. Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul F. Christiano. Learning to summarize with human feedback. In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (eds.), *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL https://proceedings.neurips.cc/paper/2020/hash/1f89885d556929e98d3ef9b86448f951-Abstract.html.

Leszek Szczecinski. G-Elo: generalization of the Elo algorithm by modeling the discretized margin of victory. *Journal of Quantitative Analysis in Sports*, 18(1):1–14, 2022.

Louis L Thurstone. A law of comparative judgment. *Psychological Review*, 34(4):273–286, 1927.

Panagiotis Toulis, Edoardo Airoldi, and Jason Rennie. Statistical analysis of stochastic gradient methods for generalized linear models. In *International Conference on Machine Learning*, pp. 667–675. PMLR, 2014.

Panos Toulis, Thibaut Horel, and Edoardo M. Airoldi. The proximal Robbins–Monro method. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 83(1):188–212, February 2021. ISSN 1369-7412. doi: 10.1111/rssb.12405.

Roman Vershynin. *High-Dimensional Probability: An Introduction with Applications in Data Science*, volume 47 of *Cambridge Series in Statistical and Probabilistic Mathematics*. Cambridge University Press, Cambridge, 2 edition, 2026. ISBN 978-1009-49064-1. doi: 10.1017/9781009490672.

Milan Vojnovic, Se-Young Yun, and Kaifang Zhou. Convergence Rates of Gradient Descent and MM Algorithms for Bradley-Terry Models. In *Proceedings of the Twenty Third International Conference on Artificial Intelligence and Statistics*, pp. 1254–1264. PMLR, June 2020.

Milan Vojnović, Se-Young Yun, and Kaifang Zhou. Accelerated MM Algorithms for Inference of Ranking Scores from Comparison Data. *Operations Research*, 71(4):1318–1342, July 2023. ISSN 0030-364X. doi: 10.1287/opre.2022.2264.

David Williams. *Probability with Martingales*. Cambridge University Press, Cambridge, February 1991. ISBN 978-0-511-81365-8. doi: 10.1017/CBO9780511813658.

Ting Yan. Inference in a generalized Bradley–Terry model for paired comparisons with covariates and a growing number of subjects, July 2025. URL http://arxiv.org/abs/2507.22472.

E. Zermelo. Die Berechnung der Turnier-Ergebnisse als ein Maximumproblem der Wahrscheinlichkeitsrechnung. *Mathematische Zeitschrift*, 29(1):436–460, December 1929. ISSN 1432-1823. doi: 10.1007/BF01180541.

Arpad Élő. *The rating of chessplayers, past and present*. Arco Pub., 1978.

Supplementary material

Outline of the appendix

A Additional related work . . . 16

B Convergence for comparison-based learning . . . 18

C Modeling and convergence . . . 18

C.1 Generative model . . . 19

C.2 Population loss . . . 20

C.3 Comparison-based models . . . 22

D Proofs on Monotonicity . . . 23

D.1 Geometric conditions . . . 23

D.2 Monotonicity for categorical features . . . 24

D.3 Asymptotic monotonicity . . . 25

E Complementary figures for the numerical experiments . . . 26

A Additional related work

Pairwise comparison models. Pairwise observations of the form "a is better than b" are pervasive in settings where absolute values must be inferred from relative or ordinal information. The Bradley–Terry (BT) model (Bradley & Terry, 1952) is the canonical model for binary comparisons. It was already foreshadowed by Zermelo's tournament model (Zermelo, 1929) and is closely related to Thurstone's law of comparative judgment (Thurstone, 1927).

Several extensions enrich the outcome space: ties (Davidson, 1970; Rao & Kupper, 1967), binary, ternary, or countably infinite outcomes (Király & Qian, 2017), and k-ary outcomes (Szczecinski, 2022). The Plackett–Luce model (Luce, 1959; Plackett, 1975) and, more broadly, random-utility models extend the construction to rankings and multiple alternatives (Galichon, 2026). The Generalized Bradley–Terry (GBT) model (Fageot et al., 2024) unifies these extensions within a single construction, replacing the logistic base law with a general base distribution.

The connection with generalized linear models (GLMs) was made explicit early on by Critchlow & Fligner (1991), who cast both BT and Thurstone–Mosteller as GLMs and also treat ties in this framework. More recent reviews reiterate this connection (Hamilton et al., 2025; Fang et al., 2026). The general pairwise-comparison structure of Han et al. (2023) is close to a GLM as well, although the authors do not use that term explicitly.

Covariates and features of the alternatives—or of the pair as a whole—have been incorporated in several models (Király & Qian, 2017; Yan, 2025; Fan et al., 2024). Fageot et al. (2025) extend GBT to this setting, replacing free per-item parameters with linear scores from shared feature vectors. The generative model underlying our online learning framework builds on this feature-based GBT construction. This shift from free per-item parameters to shared features is precisely what enables generalization across items, at the cost of the tension with monotonicity that we study in Section 4. Our focus differs from Fageot et al. (2025), however: they study batch/offline inference, whereas we study online learning from a stream of comparisons.

Offline inference for BT and variants. The dominant setting for fitting BT models is offline maximum likelihood estimation (MLE). The iterative scheme of Zermelo (1929), reinterpreted as a minorization-maximization (MM) method by Hunter (2004) and revisited by Newman (2023), forms the algorithmic backbone of this literature, with accelerations via spectral methods (Negahban et al., 2017; Maystre & Grossglauser, 2015) and faster iterative schemes (Vojnovic et al., 2020; Vojnović

et al., 2023; Agarwal et al., 2018). Bayesian approaches (Caron & Doucet, 2012; Corff et al., 2018; Diel et al., 2020) and consistency guarantees for the MLE (Shah et al., 2016; Han et al., 2023; Li et al., 2022) have also been studied, as have gradient-based methods on vanilla BT (Hendrickx et al., 2020) and stochastically transitive models (Shah et al., 2017). All these methods assume a static environment and repeated access to stored data; none provide last-iterate, constant-stepsize guarantees in an online stream.

Online inference for BT and variants. The Elo rating system (Élő, 1978), designed for chess, is the prototypical online ranking method. Király & Qian (2017) identify Elo as precisely SGD on the BT log-likelihood with constant stepsize—or equivalently, SGD on logistic regression when features are pairwise differences of indicator vectors (Morse, 2019). Coulom (2008) derives Elo-like updates as an online Bayesian approximation, and TrueSkill (Herbrich et al., 2006) extends this to team settings. Olesker-Taylor & Zanetti (2024) analyses Elo via Markov chains—a perspective complementary to ours, but restricted to binary outcomes and without convergence rate or monotonicity guarantees. Toulis et al. (2014) compares SGD with implicit SGD in the GLM setting. Learning-to-rank methods such as RankNet (Burges et al., 2005) share the online spirit but go beyond GLMs with nonlinear features, though they lack theoretical guarantees. We generalize Elo to account for a wide range of base distributions and arbitrary covariates, establishing last-iterate convergence and monotonicity for the first time, and thereby making the tension between monotonicity and generalization concrete and quantifiable.

SGD with constant stepsize. With decaying stepsize, SGD is well understood (Bottou et al., 2018). With constant stepsize, SGD converges to a neighbourhood of the optimum, exhibiting a bias-variance tradeoff controlled by the stepsize. For GLMs, the results of Bach (2014) guarantee convergence with constant stepsize, for the *average* iterate and assuming bounded features. Last-iterate guarantees are significantly harder. Harvey et al. (2019) provides tight SGD analyses, showing that the last-iterate converges slower than the average one (with decaying stepsize). Attia et al. (2025) and Garrigos & Gower (2024) establish last-iterate convergence for smooth convex problems, with minimal variance assumptions. We build on the latter to obtain constant-stepsize, last-iterate convergence guarantees for SGD on the proposed model.

Monotonicity and convergence. Monotonicity is a notion distinct from convergence, and largely independent of it. Convergence guarantees remain crucial nonetheless: they ensure that iterates stabilize and have a meaningful asymptotic behavior, so that monotonicity of a diverging trajectory would otherwise be an uninformative property.

Monotonicity as a social choice axiom. Monotonicity's origins as a normative requirement for comparison-based aggregation trace back to early tournament-ranking ideas (Zermelo, 1929), and to the subsequent social-choice literature on paradoxes and Condorcet procedures (Fishburn, 1983; Moulin, 1988). In probabilistic preference models, Azari Soufiani et al. (2014) study monotonicity for Condorcet-style noise models and Bayesian estimators. Noothigattu et al. (2020) establish it for random-utility models estimated by MLE, including Thurstone–Mosteller and BT, under strict log-concavity conditions. The normative motivation is concrete in applications such as the kidney-exchange prioritization system studied by Freedman et al. (2020), where pairwise comparisons elicited from crowdworkers inform high-stakes decisions.

All of these results concern covariate-free models, where each item carries its own free parameter and monotonicity is comparatively easy to guarantee.

Monotonicity with covariates. More recent work extends monotonicity guarantees to the feature-based GBT model, considering MAP estimators (a regularized form of MLE) in an offline setting (Fageot et al., 2024; 2025)—precisely the setting where shared features can, in principle, generate the tension with monotonicity that motivates this paper. Bareilles et al. (2025) instead focus on local monotonicity properties, in both static and dynamic settings, for more general models where scores depend nonlinearly on the features.

All these results, however, characterize the estimator obtained after fitting on a fixed dataset, leaving open the question of whether monotonicity holds along the trajectory of an online gradient method.

Monotonicity failures in preference-based fine-tuning. This question is far from academic: empirical work on preference-based fine-tuning of language models has documented recurring, though not systematic, monotonicity failures under RLHF/DPO-style updates, precisely in the regime where

a shared, low-dimensional feature representation is used to generalize across a large space of possible completions. Pal et al. (2024) report a "squeezing" effect, where training on a pair (a, b) can push probability mass towards unrelated tokens or sequences c rather than the preferred completion a, occasionally decreasing the probability of a itself. Razin et al. (2024) similarly document "likelihood displacement," where such training shifts probability mass toward completions c unrelated to a, sometimes reversing the intended preference; Chen et al. (2024) report related effects. In our terminology, these are empirical instances of individual and gap monotonicity failing, as shared, low-dimensional feature vectors let a gradient step on the pair (a, b) perturb the score of an uninvolved c.

Positioning. Existing results thus characterize batch estimators, not the iterates of constant-stepsize SGD. We fill this gap by analyzing monotonicity throughout the online trajectory: weak gap monotonicity holds universally, whereas the stronger notions of gap and individual monotonicity depend on the relation between feature dimension and number of alternatives—equivalently, on the relevant conditions on feature geometry (in particular, D versus A).

B Convergence for comparison-based learning

In this Section, we discuss how the quantities featured in the convergence bounds of Section 3 specialize in the case of comparison-based models.

Strong convexity. As seen in Proposition 1, the strong-convexity constant ν of the loss $\mathcal{L}$ depends on the spectrum of the second-moment matrix $\mathbb{E}[xx^\top]$: $\nu = \mu\lambda^+_{\min}(\mathbb{E}[xx^\top])$. For the comparison model of Example 3, this matrix is (cf. Lemma 7)

$$\mathbb{E}[xx^\top] = XL_QX^\top,$$

where X is the $D \times |\mathcal{A}|$ matrix whose columns are the vectors $(x_a)_{a\in\mathcal{A}}$ and L_Q is the Laplacian of Q's pairing graph (Definitions 3 and 4). Using Lemma 7 we obtain the following bound:

$$\nu \geq \mu\lambda^+_{\min}(L_Q)\sigma^+_{\min}(X)^2,$$

which requires[4] $\ker L_Q \subseteq \ker X$ and with $\sigma_{\max}$ and $\sigma^+_{\min}$ the largest and smallest nonzero singular values, respectively. In particular, the convergence guarantees in Sections 3.1 and 3.2 remain valid with ν replaced by this estimate.

Smoothness. The per-sample constant $L_\star = MR_x^2$ of (6), featured in Propositions 1 and 2, admits the analogous estimate $L_\star \leq M \max_{(a,b)\in\operatorname{supp} Q} \|x_{a\ominus b}\|^2$ and depends on the pairing distribution only through its support.

These bounds are generically not tight, however, as they fail to capture the interaction between the feature matrix X and the pairing distribution Q (for example, when the random draw of a pair depends on the alternatives' features). In addition, (5) implies

$$\mu \operatorname{Tr}\mathbb{E}[xx^\top] \leq \sigma_*^2 \leq M \operatorname{Tr}\mathbb{E}[xx^\top]$$

showing that σ_*^2 also depends on the spectrum of $\mathbb{E}[xx^\top]$.

Although one expects SGD to converge faster for well-connected pairing distributions (e.g. as quantified by the algebraic connectivity $\lambda^+_{\min}(L_Q)$), the preceding discussion shows that this dependency is hard to quantify analytically. We investigate this empirically in Section 5.

C Modeling and convergence

This section collects elementary properties of the Generalized Linear Model introduced in Definition 1 (Appendix C.1) and of its expected log-likelihood (Appendix C.2). The second-moment matrix $\mathbb{E}[xx^\top]$ of the feature distribution plays a central role throughout. In the specific case of pairwise comparison models, we show in Appendix C.3 that it can be expressed in terms of the

[4]This condition amounts to saying that the feature vectors are normalized such that each feature is zero sum across the connected components of the pairing graph G_Q (cf. Definition 3). In particular, when G_Q is connected, this condition is satisfied as soon as as the features are mean-normalized or standardized.

Laplacian of the pairing distribution. None of the results in this section are intended to be novel. We include them for the sake of self-containment and in cases where we were unable to locate a reference formulated in our exact setting.

A note on identifiability. Definition 1 defines a parametric family $\{\mathbb{P}_\beta\}_{\beta\in\mathbb{R}^D}$ of probability distributions. This family might however be unidentifiable, (i.e. $\beta \mapsto \mathbb{P}_\beta$ is non-injective). For Generalized Linear Models, it is common to assume that $\mathbb{P}_x$ is non-degenerate, in the sense that $\operatorname{supp}(\mathbb{P}_x)$ spans $\mathbb{R}^D$, which implies identifiability. As we will see below, there are common situations where this is too restrictive an assumption. More generally, Lemma 3 shows that

$$\mathbb{P}_{\beta_1} = \mathbb{P}_{\beta_2} \iff \beta_1 - \beta_2 \in \mathcal{V}_x^\perp,$$

where $\mathcal{V}_x \triangleq \textsc{span}(\operatorname{supp}\mathbb{P}_x)$. In other words, we can only hope to learn β "up to" an additive shift in $\mathcal{V}_x^\perp$, that is, the equivalence class of β in $\mathbb{R}^D/\mathcal{V}_x^\perp \cong \mathcal{V}_x$. We therefore restrict ourselves to the identifiable family $\{\mathbb{P}_\beta\}_{\beta\in\mathcal{V}_x}$ in the remainder of this paper.

Example 7. For the GBT model of Example 2, the space $\mathcal{V}_x^\perp$ is, by Lemma 6, spanned by the indicator vectors of G_Q's connected components, where G_Q is the pairing graph of Q (cf. Definition 3). In other words, $\mathcal{V}_x$ is the set of vectors such that the sum of their coordinates on each connected component of G_Q vanishes.

In Example 3, $\mathbb{P}_x$ is the pushforward of Q through the map $(a,b) \mapsto x_a - x_b$, hence

$$\mathcal{V}_x = \textsc{span}\{x_a - x_b \mid (a,b) \in \operatorname{supp} Q\}.$$

This space $\mathcal{V}_x$ is equivalently characterized by Lemma 7.

C.1 Generative model

The following lemma relates the vector span of the support[5] of a multivariate random variable to its second-moment matrix. It is common knowledge but we were not able to find an easily citable source and include a proof for completeness.

Lemma 1. *Let x be a random variable over $\mathbb{R}^D$ with distribution $\mathbb{P}_x$ and define $\mathcal{V}_x \triangleq \textsc{span}(\operatorname{supp}\mathbb{P}_x)$. If x has finite second-order moments, then $\mathcal{V}_x^\perp = \ker \mathbb{E}[xx^\top]$.*

Proof. Recall that for a positive semi-definite (PSD) matrix $M \in \mathbb{R}^{D\times D}$, the spectral theorem implies that $\ker M = \{u \in \mathbb{R}^D \mid u^\top M u = 0\}$. Thus, since $\mathbb{E}[xx^\top]$ is PSD, we have

$$\begin{aligned} u \in \ker \mathbb{E}[xx^\top] &\iff \mathbb{E}[u^\top xx^\top u] = 0 \iff \mathbb{E}[\langle u, x\rangle^2] = 0 \\ &\iff \langle u, x\rangle = 0, \text{ a.s.} \\ &\iff \operatorname{supp}\mathbb{P}_x \subseteq u^\perp \\ &\iff \textsc{span}(\operatorname{supp}\mathbb{P}_x) \subseteq u^\perp \\ &\iff u \in \textsc{span}(\operatorname{supp}\mathbb{P}_x)^\perp, \end{aligned}$$

where the last equivalence follows from taking orthogonal complements on both sides of the inclusion, and the fact that in finite dimension the double orthogonal complement is the original subspace. □

We recall the classical relationship between an exponential family's cumulants and the CGF of its base measure.

Lemma 2. *For (x, r) distributed as $\mathbb{P}_\beta$, the nth cumulant of $r|x$ is $\Phi_f^{(n)}(\langle x, \beta\rangle)$, $n \geq 1$. In particular,*

$$\begin{aligned} \mathbb{E}[r \mid x] &= \Phi_f'(\langle x, \beta\rangle) \\ \mathbb{E}[(r - \mathbb{E}[r \mid x])^2 \mid x] &= \Phi_f''(\langle x, \beta\rangle) \\ \mathbb{E}[(r - \mathbb{E}[r \mid x])^3 \mid x] &= \Phi_f'''(\langle x, \beta\rangle). \end{aligned}$$

[5] We adopt the usual definition of the support of a Borel measure on a topological space: the set of points such that all their open neighborhoods have positive measure. This is a closed set, and when the measure is inner regular—in particular when the space is a finite-dimensional vector space—the support has full measure, that is, its complement is a null set (Bogachev, 2007, Prop. 7.2.9 p. 77).

Proof. By definition of the generalized linear model $\mathbb{P}_\beta$, the conditional distribution $r|x$ is the member of the canonical exponential family generated by f with parameter $\langle x, \beta\rangle$. By a standard property of exponential families (see e.g. Lehmann & Casella (1998, Theorem 5.10)), its CGF is $K : s \mapsto \Phi_f(s + \langle x, \beta\rangle) - \Phi_f(\langle x, \beta\rangle)$. Therefore, the cumulants of $r|x$ are given by the derivatives of K at 0, or equivalently the derivatives of Φ_f at $\langle x, \beta\rangle$. The lemma's identities follow since the first three cumulants of a random variable are its mean, variance and third central moment, respectively. □

Lemma 3. *Let $\{\mathbb{P}_\beta\}_{\beta\in\mathbb{R}^D}$ be the family of Definition 1 with base distribution f and feature distribution $\mathbb{P}_x$. If $\operatorname{supp} f$ contains at least two points, we have for all $\beta_1, \beta_2 \in \mathbb{R}^D$, $\mathbb{P}_{\beta_1} = \mathbb{P}_{\beta_2} \iff \beta_1 - \beta_2 \in \mathcal{V}_x^\perp$, with $\mathcal{V}_x \triangleq \operatorname{SPAN}(\operatorname{supp} \mathbb{P}_x)$.*

Proof. Fix $\beta_1, \beta_2 \in \mathbb{R}^D$. By definition of $\{\mathbb{P}_\beta\}$, we have $\mathbb{P}_{\beta_1} = \mathbb{P}_{\beta_2}$ iff the following holds almost surely

$$r\langle x, \beta_1 - \beta_2\rangle = \Phi_f\big(\langle x, \beta_1\rangle\big) - \Phi_f\big(\langle x, \beta_2\rangle\big),$$

where (x, r) is distributed according to $\mathbb{P}_x \otimes f$. In particular, on the event $E \triangleq \{\langle x, \beta_1 - \beta_2\rangle \neq 0\}$, we have

$$r = \frac{\Phi_f\big(\langle x, \beta_1\rangle\big) - \Phi_f\big(\langle x, \beta_2\rangle\big)}{\langle x, \beta_1 - \beta_2\rangle} \triangleq \ell(x).$$

Since $\operatorname{supp} f$ has at least two points, there exists $a \in \mathbb{R}$ such that $\mathbb{P}[r \leq a] > 0$ and $\mathbb{P}[r > a] > 0$. Then,

$$\mathbb{P}[r \leq a \wedge E] = \mathbb{P}[(r \leq a) \wedge (\ell(x) \leq a) \wedge E] = \mathbb{P}[\ell(x) \leq a \wedge E],$$

where both equalities use that $r = \ell(x)$ on E. Moreover, x and r are independent and E is in the σ-algebra generated by x, hence we also have

$$\mathbb{P}[(r \leq a) \wedge (\ell(x) \leq a) \wedge E] = \mathbb{P}[r \leq a]\,\mathbb{P}[\ell(x) \leq a \wedge E].$$

Combining the previous two equations, we obtain

$$\mathbb{P}[\ell(x) \leq a \wedge E] = \mathbb{P}[r \leq a]\,\mathbb{P}[\ell(x) \leq a \wedge E].$$

Since $0 < \mathbb{P}[r \leq a] < 1$, this implies $\mathbb{P}[\ell(x) \leq a \wedge E] = 0$. The same argument, replacing "$\leq a$" with "$> a$" yields $\mathbb{P}[\ell(x) > a \wedge E] = 0$. Hence,

$$\mathbb{P}[E] = \mathbb{P}[\ell(x) \leq a \wedge E] + \mathbb{P}[\ell(x) > a \wedge E] = 0.$$

In other words, $\langle x, \beta_1 - \beta_2\rangle = 0$ almost surely. The proof of Lemma 1 shows that this is equivalent to $\beta_1 - \beta_2 \in \mathcal{V}_x^\perp$. □

C.2 Population loss

In this section, we collect regularity properties of the population loss $\mathcal{L}$. Starting with differentiability, recall that the cumulant generating function Φ_f is infinitely differentiable on its domain (Barndorff-Nielsen, 2014, Corollary 7.1). The same holds for $\mathcal{L}$ as stated next.

Lemma 4. *Under Assumption 1, the population loss $\mathcal{L}$ is infinitely differentiable. For $n \geq 1$, the nth derivative at $\beta \in \mathbb{R}^D$ is the n-linear form $d^n\mathcal{L}_\beta : (\mathbb{R}^D)^n \to \mathbb{R}$ given by*

$$d^n\mathcal{L}_\beta = \mathbb{E}\left[\Phi_f^{(n)}(\langle x, \beta\rangle)x^{\otimes n}\right] - \mathbf{1}\{n = 1\}\,\mathbb{E}[rx], \tag{9}$$

where we canonically identify the nth order tensor $x^{\otimes n}$ with the n-linear form $(d_i)_{i\in[n]} \mapsto \prod_{i=1}^n \langle x, d_i\rangle$.

Proof. For each $(x, r) \in \mathbb{R}^D \times \mathbb{R}$, denote by $h^{x,r} : \beta \mapsto \Phi_f(\langle x, \beta\rangle) - r\langle x, \beta\rangle$ the integrand in the definition of $\mathcal{L}$. Since the function Φ_f is infinitely differentiable, so is $h^{x,r}$, and for $n \geq 1$ its nth derivative at $\beta \in \mathbb{R}^D$ is given by

$$d^n(h^{x,r})_\beta = \Phi_f^{(n)}(\langle x, \beta\rangle)x^{\otimes n} - \mathbf{1}\{n = 1\}rx. \tag{10}$$

By a standard corollary of the dominated convergence theorem, swapping the expectation and differentiation operations is allowed if we can show that all the partial derivatives of $h^{x,r}$ are locally

upper-bounded by an integrable function independent of β. This in turn guarantees that all the partial derivatives of $\mathcal{L}$ (of all orders) exist which in turn implies that they are all continuous, and hence that $\mathcal{L}$ is infinitely differentiable with derivatives given by (9).

From (10), we get that, for all multi-index $\alpha \in \mathbb{N}^D$ with $|\alpha| = n$, all $(x, r) \in \mathbb{R}^D \times \mathbb{R}$, and all $\beta \in \mathbb{R}^D$, the nth-order partial derivative $\partial^\alpha h^{x,r}$ exists and is continuous at β, with

$$\partial^\alpha h^{x,r}(\beta) = \Phi_f^{(n)}(\langle x, \beta \rangle) x^\alpha - \mathbf{1}\{n = 1\} r x^\alpha.$$

Now, fix $\beta_0 \in \mathbb{R}^D$ and a compact neighborhood K of β_0. Since $\Phi_f^{(n)}$ is continuous and $\operatorname{supp} \mathbb{P}_x$ is bounded, there exists a constant R such that for all $(x, \beta) \in \operatorname{supp} \mathbb{P}_x \times K$, $|\Phi_f^{(n)}(\langle x, \beta \rangle)| \leq R$. Hence, for all $(x, \beta) \in \operatorname{supp} \mathbb{P}_x \times K$ and $r \in \mathbb{R}$,

$$|\partial^\alpha h^{x,r}(\beta)| \leq \big(R + \mathbf{1}\{n = 1\} r\big) \|x\|^n \triangleq m(x, r).$$

The upper-bound m is $\mathbb{P}_{\beta^\star}$ integrable and independent of β. Hence, the partial derivative $\partial^\alpha \mathcal{L}(\beta_0)$ exists and is given by

$$\partial^\alpha \mathcal{L}(\beta_0) = \mathbb{E}\left[\Phi_f^{(n)}(\langle x, \beta_0 \rangle) x^\alpha - \mathbf{1}\{n = 1\} r x^\alpha\right].$$

Since β_0, n and α were arbitrary, this concludes the proof by the argument above. □

The function Φ_f is also strictly convex (Barndorff-Nielsen, 2014, Theorem 7.1). The following lemma shows that the same holds for $\mathcal{L}$ over $\mathcal{V}_x$ and that more generally, $\mathcal{L}$ inherits the convexity guarantees of Φ_f.

Lemma 5 (Convexity of the loss)**.** *Under Assumption 1, the population loss $\mathcal{L}$ (Pop. loss) satisfies the following properties:*

1. *It is convex on $\mathbb{R}^D$ and strictly convex on $\mathcal{V}_x$.*
2. *If Φ_f is M-smooth, then $\mathcal{L}$ is $M\lambda_{\max}(\mathbb{E}[xx^\top])$-smooth on $\mathbb{R}^D$.*
3. *If Φ_f is μ-strongly convex, then $\mathcal{L}$ is $\mu\lambda^+_{\min}(\mathbb{E}[xx^\top])$-strongly convex on $\mathcal{V}_x$.*
4. *If Φ_f is M-smooth, then almost every sample loss $\ell_{x,r} : \beta \mapsto \Phi_f(\langle x, \beta \rangle) - r\langle x, \beta \rangle$ is convex and $L_\star$-smooth on $\mathbb{R}^D$, with $L_\star \triangleq MR_x^2$.*

Proof. First, observe that $\mathcal{L}(\beta) = \mathbb{E}[\Phi_f(\langle x, \beta \rangle)] - \mathbb{E}[r\langle x, \beta \rangle]$ and that $\beta \mapsto \mathbb{E}[r\langle x, \beta \rangle]$ is linear. Hence, it suffices to prove the properties for the function $h : \beta \mapsto \mathbb{E}[\Phi_f(\langle x, \beta \rangle)]$.

1. Convexity (resp. strict convexity) of h is equivalent to monotonicity (resp. strict monotonicity) of its gradient. Using Lemma 4, we compute for $\beta_1, \beta_2 \in \mathbb{R}^D$,
$$\langle \nabla h(\beta_1) - \nabla h(\beta_2), \beta_1 - \beta_2 \rangle = \mathbb{E}\left[\big(\Phi_f'(\langle x, \beta_1 \rangle) - \Phi_f'(\langle x, \beta_2 \rangle)\big)\langle x, \beta_1 - \beta_2 \rangle\right]. \tag{11}$$
The integrand has the form $(\Phi_f'(s) - \Phi_f'(t))(s - t)$ which is non-negative since Φ_f is convex, hence h is convex on $\mathbb{R}^D$.
For the strict convexity, assume that $\beta_1, \beta_2 \in \mathcal{V}_x$ and that the right-hand side in (11) is zero. This implies that the integrand in (11) is almost surely zero, but by monotonicity of Φ_f', this implies that $\langle x, \beta_1 - \beta_2 \rangle = 0$ a.s., or equivalently that $(\beta_1 - \beta_2) \in \mathcal{V}_x^\perp$. Since we also have $(\beta_1 - \beta_2) \in \mathcal{V}_x$, this implies $\beta_1 = \beta_2$ proving the strict monotonicity of ∇h on $\mathcal{V}_x$.
2. By Lemma 4, the second-order total derivative of h is:
$$d^2 h(\beta) = \mathbb{E}\left[\Phi_f''(\langle x, \beta \rangle) x x^\top\right].$$
The conclusion follows since the smoothness of h is equivalent to an upper bound on the eigenvalues of $d^2 h$ and $\Phi_f''(\beta) \leq M$ by M-smoothness of Φ_f'.
3. The proof of this property is identical to the one of 2., lower bounding the eigenvalues of $d^2 h$ using that $\Phi_f''(\beta) \geq \mu$ by strong convexity of Φ_f'. Note that, by Lemma 1, $\mathcal{V}_x = (\ker \mathbb{E}[xx^\top])^\perp$, thus only the smallest nonzero eigenvalue of $\mathbb{E}[xx^\top]$ appears when restricting h to $\mathcal{V}_x$.
4. The hessian of $\ell_{x,r}$ is $d^2 \ell_{x,r}(\beta) = \Phi_f''(\langle x, \beta \rangle)\, xx^\top$. Hence, $0 \preceq d^2 \ell_{x,r}(\beta) \preceq M\|x\|^2 I \preceq MR_x^2 I$ for almost every (x, r), which is the announced convexity and $L_\star$-smoothness. □

C.3 COMPARISON-BASED MODELS

We now focus on the setting of Examples 2 and 3, where the feature vector x arises from pairwise comparisons among a finite set $\mathcal{A}$ of alternatives. In both examples, a random pair of alternatives $(a, b) \in \mathcal{A}^2$ is drawn according to a *pairing distribution* Q over $\mathcal{A}^2$. The goal of this section is to relate, for this setting, the second-moment matrix $\mathbb{E}[xx^\top]$ to the Laplacian matrix of a graph we now define, along with related notions.

Definition 3. Let $\mathcal{A}$ be a finite set and $Q : \mathcal{A}^2 \to [0, 1]$ be a probability distribution. The *pairing graph of* Q is the undirected weighted graph G_Q with vertex set $\mathcal{A}$ and edge weight $w_{ab} = Q(a, b) + Q(b, a)$ for each $\{a, b\} \subseteq \mathcal{A}$.

The *underlying graph* of G_Q is the (unweighted) undirected graph with vertex set $\mathcal{A}$ and edge $\{a, b\}$ for each $\{a, b\} \subseteq \mathcal{A}$ with $w_{a,b} > 0$. By a *connected component* of G_Q, we mean a connected component of its underlying graph, that is, an equivalence class for the reflexive and transitive closure of the relation defined by its edges. Finally, we denote by $\mathcal{C}_Q$ the set of connected components of G_Q, which form a partition of $\mathcal{A}$.

We also recall the definition of the Laplacian matrix of a weighted undirected graph (see e.g. (Spielman, 2025, §3.1)).

Definition 4. Let G be a weighted undirected graph with vertex set V and weight w_{uv} for each $\{u, v\} \subseteq V$. Denote by $e_u \in \mathbb{R}^{|V|}$ the indicator vector of $u \in V$. The *Laplacian* of G is the positive semidefinite matrix in $\mathbb{R}^{|V|\times|V|}$ given by

$$L_G \triangleq \sum_{\{u,v\}\subseteq V} w_{uv}(e_u - e_v)(e_u - e_v)^\top = EWE^\top, \tag{12}$$

where E is an incidence matrix of G and W is the diagonal matrix whose diagonal entries are the edge weights (in the same order as the edges are ordered in E). The diagonal entry of L_G corresponding to vertex $u \in V$ is $\sum_{v\in V} w_{uv}$ and the off-diagonal entry corresponding to $(u, v) \in V^2$ is $-w_{uv}$.

With these definitions in place, the following lemma characterizes the second-moment matrix $\mathbb{E}[xx^\top]$ for the GBT model of Example 2.

Lemma 6. *Let $\mathcal{A}$ be a finite set and Q be a probability distribution over $\mathcal{A}^2$. Denote by $e_a \in \mathbb{R}^{|\mathcal{A}|}$ the indicator vector of $a \in \mathcal{A}$ and define the random variable $x = e_a - e_b \in \mathbb{R}^{|\mathcal{A}|}$ where $(a, b) \in \mathcal{A}^2$ is a random pair distributed as Q. Then,*

1. *The second-moment matrix $\mathbb{E}[xx^\top]$ is the Laplacian of Q's pairing graph G_Q. (cf. Definition 3).*
2. *The kernel of $\mathbb{E}[xx^\top]$ is the subspace of vectors that are constant on each connected component of G_Q. It admits as an orthogonal basis the vectors $\{e_C\}_{C\in\mathcal{C}_Q}$, where e_C denotes the indicator vector of the connected component $C \in \mathcal{C}_Q$.*

Proof. 1. We have

$$\mathbb{E}[xx^\top] = \sum_{(a,b)\in\mathcal{A}^2} Q(a, b)(e_a - e_b)(e_a - e_b)^\top = \sum_{\{a,b\}\subseteq\mathcal{A}} \big(Q(a, b) + Q(b, a)\big)(e_a - e_b)(e_a - e_b)^\top$$

and the conclusion follows from the definition of G_Q and (12).

2. This is a well-known property of Laplacians, but we include a proof for completeness. By the spectral theorem $u \in \ker L_G$ iff $u^\top L_G u = 0$. By (12), this is equivalent to

$$\sum_{\{a,b\}\subseteq\mathcal{A}} w_{ab}(u_a - u_b)^2 = 0.$$

Thus, a kernel vector satisfies $u_a = u_b$ for each edge $\{a, b\}$ with $w_{ab} > 0$, that is, each edge of G_Q's underlying graph. By transitivity, u is constant on each connected component of G_Q. The conclusion follows immediately. □

The setting of Example 3 is similar, with the addition of a D-dimensional feature vector $x_a \in \mathbb{R}^D$ for each alternative $a \in \mathcal{A}$. This is closely related to the question of graph representations (a.k.a.

spring networks, see Godsil & Royle (2001, §13.3)), since x_a can be thought of as the location of alternative a in a D-dimensional space. The following lemma characterizes the second-moment matrix in this case.

Lemma 7. *Let $\mathcal{A}$ be a finite set and Q be a probability distribution over $\mathcal{A}^2$. Let $x_a \in \mathbb{R}^D$ be a feature vector for each $a \in \mathcal{A}$ and define the random variable $x = x_a - x_b \in \mathbb{R}^D$ where $(a, b) \in \mathcal{A}^2$ is a random pair distributed as Q. Then,*

$$\mathbb{E}[xx^\top] = XL_QX^\top, \tag{13}$$

where X is the $D \times |\mathcal{A}|$ matrix whose columns are the vectors $(x_a)_{a\in\mathcal{A}}$ and L_Q is the Laplacian of Q's pairing graph.

In particular, the following bound holds for the largest eigenvalue,

$$\lambda_{\max}\big(\mathbb{E}[xx^\top]\big) \le \lambda_{\max}(L_Q)\sigma_{\max}(X)^2$$

and, when $\ker L_Q \subseteq \ker X$, for the smallest nonzero eigenvalue,

$$\lambda^+_{\min}\big(\mathbb{E}[xx^\top]\big) \ge \lambda^+_{\min}(L_Q)\sigma^+_{\min}(X)^2.$$

Proof. Similar to the proof of Lemma 6 1., we have

$$\mathbb{E}[xx^\top] = \sum_{\{a,b\}\subseteq\mathcal{A}} w_{ab}(x_a - x_b)(x_a - x_b)^\top.$$

But observe that $x_a - x_b = X(e_a - e_b)$, hence we obtain (13) after using (12).

The bounds on the eigenvalues easily follow from the variational characterizations of the largest and smallest non-zero eigenvalues of a symmetric real matrix M:

$$\lambda_{\max}(M) = \max_{u\neq 0} \frac{\langle Mu, u\rangle}{\|u\|^2} \quad \text{and} \quad \lambda^+_{\min}(M) = \min\left\{ \frac{\langle Mu, u\rangle}{\|u\|^2} \;\middle|\; u \in (\ker M)^\perp \setminus \{0\} \right\}.$$

Indeed, writing $M = \mathbb{E}[xx^\top]$, we have, for each $u \in \mathbb{R}^{\mathcal{A}}$,

$$\begin{aligned}\langle Mu, u\rangle = \langle XL_QX^\top u, u\rangle &= \langle L_QX^\top u, X^\top u\rangle \\ &\le \lambda_{\max}(L_Q)\|X^\top u\|^2 \\ &\le \lambda_{\max}(L_Q)\lambda_{\max}(XX^\top)\|u\|^2 \\ &= \lambda_{\max}(L_Q)\sigma_{\max}(X)^2\|u\|^2,\end{aligned}$$

showing the upper estimate on $\lambda_{\max}(\mathbb{E}[xx^\top])$. Moreover, if $\ker L_Q \subseteq \ker X$, for $u \in (\ker M)^\perp$,

$$\begin{aligned}\langle Mu, u\rangle = \langle XL_QX^\top u, u\rangle &= \langle L_QX^\top u, X^\top u\rangle \\ &\ge \lambda^+_{\min}(L_Q)\|X^\top u\|^2 \\ &\ge \lambda^+_{\min}(L_Q)\lambda^+_{\min}(XX^\top)\|u\|^2 \\ &= \lambda^+_{\min}(L_Q)\sigma^+_{\min}(X)^2\|u\|^2,\end{aligned}$$

where the first inequality uses that $X^\top u \in \operatorname{im} X^\top = (\ker X)^\perp \subseteq (\ker L_Q)^\perp$, and the second inequality uses that, by Lemma 1, $u \in (\ker M)^\perp = \mathcal{V}_x \subseteq \operatorname{im} X = (\ker X^\top)^\perp = (\ker XX^\top)^\perp$. □

D Proofs on Monotonicity

D.1 Geometric conditions

Proof of Theorem 1. The three notions of monotonicity can all be written in the form

$$\delta(\theta; a, b, r)\big\langle \mathcal{U}_{a,b,r}(\theta) - \theta, u\big\rangle \ge 0,$$

with $u = e_a$ (individual monotonicity), $u = e_a - e_b$ (weak gap monotonicity) and $u = e_a - e_c$ (gap monotonicity). Taking a linear combination of (8) yields, for each $u \in \mathbb{R}^{\mathcal{A}}$,

$$\delta(\theta; a, b, r)\langle \mathcal{U}_{a,b,r}(\theta) - \theta, u\rangle = \eta\delta(\theta; a, b, r)^2\Big\langle x_a - x_b, \sum_{d\in\mathcal{A}} u_d x_d\Big\rangle. \tag{14}$$

With $u = e_a - e_b$, the inner product on the right-hand side becomes $\|x_a - x_b\|^2$, so weak gap monotonicity follows immediately. Taking $u = e_a$ and $u = e_a - e_c$ likewise shows the sufficiency of the conditions in Items 1 and 2.

For necessity, fix $(a, b) \in P$ and $\theta \in \mathbb{R}^{\mathcal{A}}$, and take $r = \Phi'_f(\theta_{a\ominus b})+1$, so that $\delta(\theta; a, b, r) = 1$. Since $\eta > 0$, $\langle \mathcal{U}_{a,b,r}(\theta)-\theta, u\rangle$ has the same sign as $\langle x_a-x_b, \sum_{d\in\mathcal{A}} u_d x_d\rangle$. Taking $u = e_a$ and $u = e_a-e_c$ shows that individual and gap monotonicity require $\langle x_a - x_b, x_a\rangle \geq 0$ and $\langle x_a - x_b, x_a - x_c\rangle \geq 0$, respectively. Since (a, b) and c were arbitrary, the conditions in Items 1 and 2 are necessary. □

Proof of Proposition 3. As in the proof of Theorem 1, monotonicity at step t is equivalent, after dividing by $\eta > 0$, to

$$(\delta^t)^2\Big\langle x_{a_t} - x_{b_t}, \sum_{d\in\mathcal{A}} u_d x_d\Big\rangle \geq 0, \tag{15}$$

with $u = e_{a_t}$ (individual), $u = e_{a_t} - e_{b_t}$ (weak gap), or $u = e_{a_t} - e_c$ for each $c \in \mathcal{A}$ (gap).

The first statement is immediate, since Definition 2 quantifies over all tuples (θ, a, b, r) with $(a, b) \in \operatorname{supp} Q$, among which are the ones realized along the trajectory.

For the second statement, assume that $\mathcal{U}$ fails to satisfy a monotonicity notion from Definition 2. Since weak gap monotonicity holds unconditionally by Theorem 1, the notion at hand is individual or gap monotonicity. By Theorem 1 again, there exist $(a, b) \in \operatorname{supp} Q$, together with $c \in \mathcal{A}$ in the gap case, such that $\langle x_a - x_b, x_a\rangle < 0$, respectively $\langle x_a - x_b, x_a - x_c\rangle < 0$.

Whenever step t compares the pair (a, b), that is, $(a_t, b_t) = (a, b)$, (15) is violated as soon as $\delta^t \neq 0$. In other words, monotonicity fails at step t on the event

$$E_t \triangleq \{(a_t, b_t) = (a, b)\} \cap \{\delta^t \neq 0\}.$$

Let p_{ab} denote the conditional distribution of $r|a, b$ and let $(\mathcal{F}_t)_{t\geq 1}$ be the natural filtration associated with the observations $(a_t, b_t, r^t)_{t\geq 1}$. Define $m^t \triangleq \Phi'_f(\theta^{t-1}_{a\ominus b})$, so that on the event $\{(a_t, b_t) = (a, b)\}$, $\delta^t \neq 0$ iff $r^t \neq m^t$. Then, since m^t is $\mathcal{F}_{t-1}$-measurable,

$$\mathbb{P}(E_t \mid \mathcal{F}_{t-1}) = Q(a, b)\big[1 - p_{ab}\big(\{m^t\}\big)\big] \geq Q(a, b)\big[1 - \sup_{m\in\mathbb{R}} p_{ab}\big(\{m\}\big)\big] \triangleq c_{Q,f}(a, b).$$

Being an exponential tilt of f, p_{ab} has the same null sets, hence is also not a Dirac measure. Its atoms thus all have mass less than 1, and since at most one of them can have mass greater than $1/2$, $\sup_{m\in\mathbb{R}} p_{ab}(\{m\})$ is either at most $1/2$ or attained; in both cases it is less than 1. As $(a, b) \in \operatorname{supp} Q$ gives $Q(a, b) > 0$, the lower bound $c_{Q,f}(a, b)$ is positive, hence $\sum_t \mathbb{P}(E_t \mid \mathcal{F}_{t-1}) = +\infty$ surely.

Since $E_t \in \mathcal{F}_t$ for $t \geq 1$, the conditional Borel–Cantelli lemma (Williams, 1991, §12.15 (b)) gives that $\frac{1}{n}\sum_{t\leq n} \mathbf{1}_{E_t}$ is almost surely asymptotically equivalent to $\frac{1}{n}\sum_{t\leq n} \mathbb{P}(E_t \mid \mathcal{F}_{t-1}) \geq c_{Q,f}(a, b) > 0$, so E_t occurs infinitely often. □

D.2 Monotonicity for categorical features

This subsection provides the precise statement underlying Example 6 and its proof. We consider feature vectors arising from one-hot encodings of categorical features. Consider first a single categorical feature: let $\alpha : \mathcal{A} \to [D]$ assign each alternative to one of D categories and let $(w_i)_{i\in[D]} \in \mathbb{R}^D$ be a vector of category weights. The feature vector of $a \in \mathcal{A}$ is the category-weighted one-hot encoding $w_{\alpha(a)}e_{\alpha(a)} \in \mathbb{R}^D$, where $(e_i)_{i\in[D]}$ is the canonical basis of $\mathbb{R}^D$.

More generally, consider S categorical features, with assignments $\alpha_j : \mathcal{A} \to [D_j]$ and category weights $(w^{(j)}_i)_{i\in[D_j]}$ for $j \in [S]$. Writing $x^{(j)}_a \triangleq w^{(j)}_{\alpha_j(a)}e_{\alpha_j(a)} \in \mathbb{R}^{D_j}$ for the weighted one-hot encoding of the j-th feature, the feature vector of a is obtained by stacking these S blocks vertically:

$$x_a \triangleq \big[(x^{(1)}_a)^\top, \ldots, (x^{(S)}_a)^\top\big]^\top \in \mathbb{R}^D, \qquad D \triangleq \sum_{j=1}^{S} D_j. \tag{16}$$

Proposition 4. *For the feature vectors defined in (16), $\mathcal{U}$ is individually and gap monotone.*

Proof. By Theorem 1, it suffices to show that $\langle x_a - x_b, x_a\rangle \geq 0$ and $\langle x_a - x_b, x_a - x_c\rangle \geq 0$ for all $a, b, c \in \mathcal{A}$. Since x_a is the concatenation of the blocks $x_a^{(j)}$, these inner products decompose as

$$\langle x_a - x_b, x_a\rangle = \sum_{j=1}^{S} \langle x_a^{(j)} - x_b^{(j)}, x_a^{(j)}\rangle, \qquad \langle x_a - x_b, x_a - x_c\rangle = \sum_{j=1}^{S} \langle x_a^{(j)} - x_b^{(j)}, x_a^{(j)} - x_c^{(j)}\rangle,$$

so it suffices to show that each summand is non-negative, that is, to treat the case of a single categorical feature. Two observations will be used repeatedly: alternatives in the same category have the same feature vector, and alternatives in distinct categories have orthogonal feature vectors.

For individual monotonicity, $\langle x_a - x_b, x_a\rangle$ equals 0 if $\alpha(a) = \alpha(b)$, and $\|x_a\|^2 = w^2_{\alpha(a)}$ otherwise. In both cases it is non-negative.

For gap monotonicity, $\langle x_a - x_b, x_a - x_c\rangle$ vanishes whenever $\alpha(a) = \alpha(b)$ or $\alpha(a) = \alpha(c)$, and equals $\|x_a - x_b\|^2 \geq 0$ when $\alpha(b) = \alpha(c)$. Finally, if a, b and c lie in pairwise distinct categories, their feature vectors are pairwise orthogonal, so $\langle x_a - x_b, x_a - x_c\rangle = \|x_a\|^2 = w^2_{\alpha(a)} \geq 0$.

Both conditions hold for every triple, hence $\mathcal{U}$ is individually and gap monotone on $P = \mathcal{A}^2$. □

D.3 Asymptotic monotonicity

We first reduce Theorem 2 to a lower-tail bound on $\|x_a\|^2$ (Lemma 8), then prove that bound in each case. For a random vector $x \in \mathbb{R}^D$, let $\|x\|_{\psi_2} \triangleq \sup_{\|v\|=1} \|\langle x, v\rangle\|_{\psi_2}$ denote its sub-Gaussian norm (Vershynin, 2026, Def. 3.4.1). We will use repeatedly that every real random variable satisfies $\mathbb{E}[X^2] \leq \|X\|^2_{\psi_2}$, since $2 \geq \mathbb{E}[e^{X^2/s^2}] \geq 1 + \mathbb{E}[X^2]/s^2$ for $s = \|X\|_{\psi_2}$; hence, summing over coordinates, $\mathbb{E}[\|x\|^2] \leq \|x\|^2_{\psi_2} D$. Throughout this subsection, we write $A \triangleq |\mathcal{A}|$, and c denotes a positive universal constant whose value may change from one occurrence to the next.

Lemma 8. *Let $(x_a)_{a\in\mathcal{A}}$ be pairwise independent, centered random feature vectors in $\mathbb{R}^D$ with $\|x_a\|_{\psi_2} \leq K$ for every $a \in \mathcal{A}$. Then, for every $\rho > 0$,*

$$\mathbb{P}(\mathcal{U} \text{ is individually and gap monotone}) \geq 1 - \sum_{a\in\mathcal{A}} \mathbb{P}\big(\|x_a\|^2 < \rho K^2 D\big) - A^2 e^{-c\rho^2 D}.$$

Proof. Let $m \triangleq \rho K^2 D$ and let $\mathcal{E}$ be the event that $\|x_a\|^2 \geq m$ for every $a \in \mathcal{A}$ and $|\langle x_a, x_b\rangle| \leq m/3$ for all distinct $a, b \in \mathcal{A}$. On $\mathcal{E}$, both conditions of Theorem 1 hold. Indeed, for $a \neq b$, $\langle x_a - x_b, x_a\rangle = \|x_a\|^2 - \langle x_a, x_b\rangle \geq \frac{2}{3}m \geq 0$, and for pairwise distinct a, b, c,

$$\langle x_a - x_b, x_a - x_c\rangle = \|x_a\|^2 - \langle x_a, x_b\rangle - \langle x_a, x_c\rangle + \langle x_b, x_c\rangle \geq m - 3 \cdot \frac{m}{3} = 0.$$

In the remaining cases, both conditions hold for any feature vectors: $\langle x_a - x_b, x_a\rangle = 0$ if $a = b$, while $\langle x_a - x_b, x_a - x_c\rangle$ vanishes if $a \in \{b, c\}$ and equals $\|x_a - x_b\|^2$ if $b = c$.

It remains to lower bound $\mathbb{P}(\mathcal{E})$. We can assume that $A \geq 2$ and that $\rho \leq 2$. Indeed, for $A = 1$ both conditions of Theorem 1 hold trivially, while for $\rho \geq 2$ the claim is vacuous: Markov's inequality and $\mathbb{E}[\|x_a\|^2] \leq K^2 D$ give $\mathbb{P}(\|x_a\|^2 < \rho K^2 D) \geq 1 - 1/\rho \geq 1/2$, so that the sum in the claim is at least $A/2 \geq 1$. For distinct $a, b \in \mathcal{A}$, since x_a and x_b are independent, a standard concentration argument[6] gives, for every $t \geq 0$,

$$\mathbb{P}(|\langle x_a, x_b\rangle| > t) \leq 2\exp\left(-c\min\left(\frac{t^2}{K^4 D}, \frac{t}{K^2}\right)\right).$$

Taking $t = m/3$ and using $\rho \leq 2$, the minimum equals $\rho^2 D/9$, so a union bound over the $A(A-1)/2$ unordered pairs of distinct alternatives shows that the probability that $|\langle x_a, x_b\rangle| > m/3$ for some $a \neq b$ is at most $A^2 e^{-c\rho^2 D}$. A final application of the union bound gives

$$\mathbb{P}(\mathcal{E}^c) \leq \sum_{a\in\mathcal{A}} \mathbb{P}\big(\|x_a\|^2 < \rho K^2 D\big) + A^2 e^{-c\rho^2 D}.$$ □

[6] (Vershynin, 2026, Lemmas 6.2.3 and 6.2.4) imply that $\langle x_a, x_b\rangle$ is subexponential: $\mathbb{E}[e^{\lambda\langle x_a, x_b\rangle}] \leq e^{C\lambda^2 K^4 D}$ for $|\lambda| \leq 1/(CK^2)$, and the tail bound follows as in the proof of Bernstein's inequality (Thm 2.9.1).

Proof of Theorem 2. In each of the three cases of the theorem, we prove below a lower-tail bound of the form $\mathbb{P}(\|x_a\|^2 < \rho K^2 D) \le e^{-c\rho^2 D}$ for some ρ equal to that of the theorem up to a universal factor and some $K \ge \max_a \|x_a\|_{\psi_2}$. The first claim of the theorem then follows from Lemma 8. The second claim is immediate since $A^2 e^{-c\rho^2 D} \to 0$ iff $c\rho^2 D - 2\log A \to \infty$.

Cases 1 and 2. We treat these two cases together by considering the more general setting $x_a = \Sigma^{1/2} z_a$ for a positive definite Σ, and a centered z_a with independent coordinates, of variance at least 1, and with $\max_a \|z_a\|_{\psi_2} \le L$. Let $\lambda_{\min}$ and $\lambda_{\max}$ denote the extreme eigenvalues of Σ and let $\kappa \triangleq \lambda_{\max}/\lambda_{\min}$. The inequality $\|z_{a,i}\|_{\psi_2} \le \|z_a\|_{\psi_2} \le L$ for every i implies $L \ge 1$, since $\mathbb{E}[z_{a,i}^2] \le \|z_{a,i}\|_{\psi_2}^2$ and the coordinates of z_a have variance at least 1.

Note that $\|x_a\|^2 = z_a^\top \Sigma z_a$ and, by independence, $\mathbb{E}[z_a^\top \Sigma z_a] = \sum_i \Sigma_{ii}\, \mathbb{E}[z_{a,i}^2] \ge \operatorname{Tr}\Sigma$. Hence,

$$\begin{aligned}
\mathbb{P}(\|x_a\|^2 < \operatorname{Tr}\Sigma/2) &\le \mathbb{P}\big(z_a^\top \Sigma z_a \le \mathbb{E}[z_a^\top \Sigma z_a] - \operatorname{Tr}\Sigma/2\big) \\
&\le \exp\left(-c \min\left\{ \frac{(\operatorname{Tr}\Sigma)^2}{4L^4 \|\Sigma\|_F^2}, \frac{\operatorname{Tr}\Sigma}{2L^2 \|\Sigma\|_{\mathrm{op}}} \right\}\right) \\
&\le \exp\left(-cD \min\left\{ \frac{1}{4L^4\kappa^2}, \frac{1}{2L^2\kappa} \right\}\right) = \exp\left(-\frac{cD}{4L^4\kappa^2}\right),
\end{aligned}$$

where the second inequality uses a one-sided Hanson–Wright inequality at deviation $\operatorname{Tr}\Sigma/2$ (Vershynin, 2026, Thm. 6.2.2), the third inequality uses $\|\Sigma\|_F^2 \le D\lambda_{\max}^2$, $\|\Sigma\|_{\mathrm{op}} = \lambda_{\max}$, and $\operatorname{Tr}\Sigma \ge D\lambda_{\min}$, and the last equality uses $L^2\kappa \ge 1$.

Since $\langle x_a, v\rangle = \langle z_a, \Sigma^{1/2} v\rangle$ for $v \in \mathbb{R}^D$, we have $\max_a \|x_a\|_{\psi_2} \le L\|\Sigma^{1/2}\|_{\mathrm{op}} = L\sqrt{\lambda_{\max}}$. Thus, for $\rho \triangleq 1/(2L^2\kappa)$ and $K \triangleq L\sqrt{\lambda_{\max}}$, we have $\rho K^2 D = \lambda_{\min} D/2$, and the above inequality yields

$$\mathbb{P}(\|x_a\|^2 < \rho K^2 D) \le \mathbb{P}(\|x_a\|^2 < \operatorname{Tr}\Sigma/2) \le e^{-c\rho^2 D}.$$

In Case 1, take $\Sigma = \sigma^2 I_D$ and $z_a = x_a/\sigma$, whose coordinates are independent, of variance at least 1, and satisfy $\|z_a\|_{\psi_2} \le CK/\sigma$ for a universal constant C (Vershynin, 2026, Lemma 3.4.2), so that $\kappa = 1$ and ρ is a universal multiple of σ^2/K^2. In Case 2, take $z_a \sim \mathcal{N}(0, I_D)$, whose sub-Gaussian norm is a universal constant, so that ρ is a universal multiple of $1/\kappa$.

Case 3. For $\rho = 1/K^2$, we have $\mathbb{P}(\|x_a\|^2 < \rho K^2 D) = 0$ since $\|x_a\| = \sqrt{D}$ almost surely. □

Remark 4. The proof shows that Theorem 2 holds in the following anisotropic sub-Gaussian situation subsuming Cases 1 and 2: $x_a = \Sigma^{1/2} z_a$, where z_a has centered, independent coordinates, of variance at least 1, and $\|z_a\|_{\psi_2} \le K$, with $\rho = \big(K^2 \kappa(\Sigma)\big)^{-1}$.

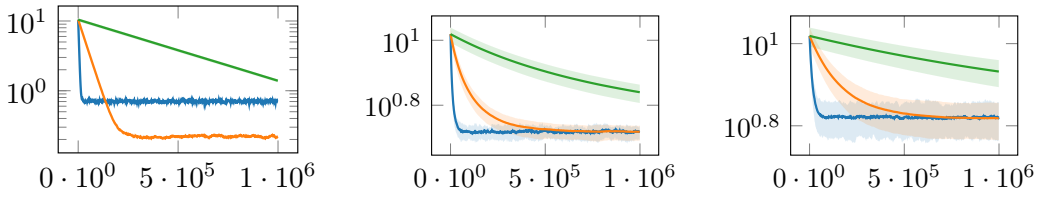
More generally, Lemma 8 shows that Theorem 2 holds for pairwise independent, centered x_a as soon as $\|x_a\|_{\psi_2} \le K$ and there exists ρ such that $\mathbb{P}(\|x_a\|^2 < \rho K^2 D) \le e^{-c\rho^2 D}$ for every $a \in \mathcal{A}$. Such a lower-tail bound is available whenever $\|x_a\|$ is sub-Gaussian around a value of order $K\sqrt{D}$, as in all cases of Theorem 2, but also in weaker regimes such as $x_a = R_a u_a$, where u_a is as in Case 3 and R_a is an independent scaling with values in $[\alpha, \beta]$, for some $0 < \alpha \le \beta$; here $\rho \triangleq \alpha^2/(\beta^2 K^2)$.

E Complementary figures for the numerical experiments

In this Section, we provide Figs. 4 to 6. The two first figures are complementary to Figs. 2 and 3, and are discussed already in the experimental Section 5.

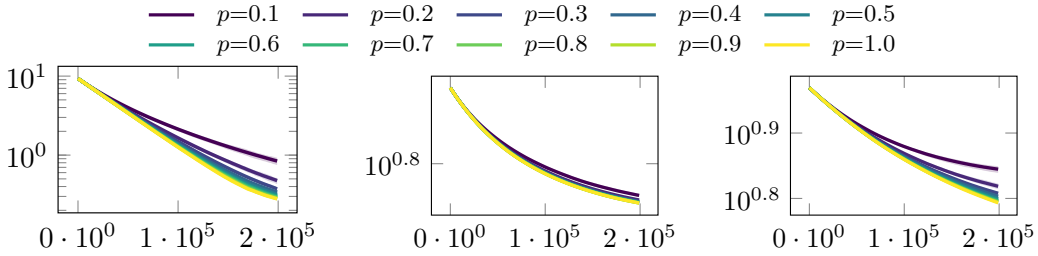
The third one is a complement to Fig. 1. It shows transition region for covariance matrix with worsening condition number. The condition number of the covariance matrix seems to control both the with and height of the region where the transition between the non-monotonous to the monotonous regime occurs.

.



(a) Gaussian (b) Bradley–Terry (c) Uniform

Figure 4: Last-iterate distance to optimum $\|\beta^T - \beta^\star\|$ of SGD with constant step size $\eta \in \{10^{-4}, 10^{-3}, 10^{-2}\}$, for three base distributions. Same setting as Fig. 2.



(a) Gaussian (b) Bradley-Terry (c) Uniform

Figure 5: Last-iterate distance to minimizer $\|\beta^T - \beta^\star\|$ of SGD as the pair distribution Q ranges over Erdős-Rényi graphs $G(|\mathcal{A}|, p)$ for $p \in \{0.1, \ldots, 1.0\}$. Same setting as Fig. 3.

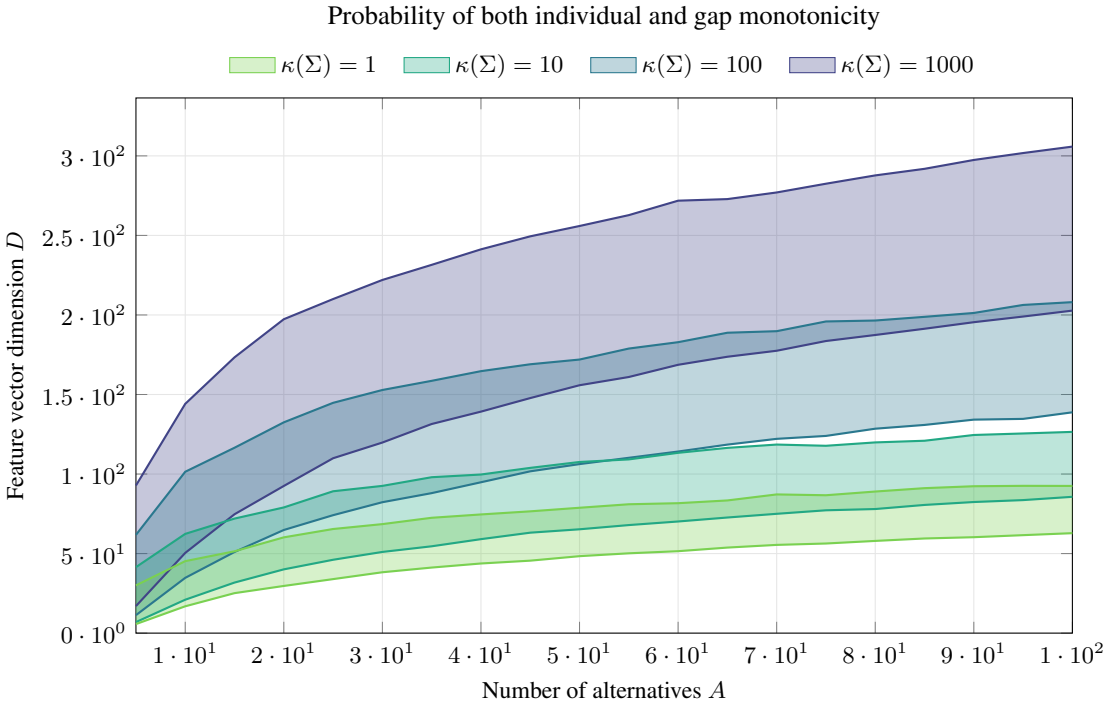


Figure 6: Transition band for feature vectors drawn from a non isotropic gaussian of covariance matrix Σ, with $\kappa(\Sigma)$ the condition number of the covariance matrix. The band show the region where the probability of individual *and* gap monotonicity is between 0.05 and 0.95