BEYOND GLOBAL CLUSTERS: OBJECTIVE-SPECIFIC USER CLUSTERING FOR MULTI-OBJECTIVE BANDITS

Anonymous authors
Paper under double-blind review

ABSTRACT

Clustering of bandits is a common approach to improving the sample efficiency of contextual bandits in recommendation: it adaptively groups users with similar preferences and pools their interaction data. Existing methods, however, group users under a single similarity structure for the whole reward, which fails in multi-objective recommendation, where users who agree on one objective may disagree on another. To fill in this gap, we propose Clustering of Multi-Objective Bandits (CMOB), a framework that maintains one user graph per objective and pools data only within objective-specific clusters, so that two users can share data for one objective while being separated for another. We instantiate the framework as COMOLB for linear rewards; since rewards are often not linear in the context, we further propose COMONB, which applies the same scheme to the gradient features of a neural network. Both algorithms separate two users for an objective only when a confidence test certifies that their preferences differ. Our proposed algorithms are proved to never separate users of the same cluster with high probability, to recover the objective-specific partitions after an identification time that depends on the horizon only logarithmically, and to attain a Pareto regret whose dominant term scales with the number of objective-specific clusters instead of the number of users. Experiments on synthetic instances show that our algorithms outperform existing clustering methods, and on semi-synthetic MovieLens data with objective-specific user clusters, our method nearly halves the regret of learning each user independently.

1 INTRODUCTION

Contextual multi-armed bandits are a standard model for sequential recommendation (Slivkins, 2019; Lattimore & Szepesvári, 2020): in every round the recommender observes a user and a set of candidate items described by contextual features, selects one item, and observes only the reward of the selected item, so that it must balance exploring uncertain items against exploiting what it has learned. Contextual bandit algorithms based on linear models (Chu et al., 2011) or neural networks (Zhou et al., 2020) typically learn a separate model for every user, which ignores the similarities between users and is sample-inefficient when the interaction data of an individual user are sparse. Clustering of bandits addresses this limitation by adaptively grouping users with similar preferences on the basis of their feedback and pooling the data within a group, so that every user benefits from the observations of the others (Gentile et al., 2014; Li et al., 2019; Yang et al., 2024).

Existing clustering methods, however, operate under a *single-faceted* premise: two users are either similar, and then share the same preferences in every respect, or they are not. Recommendation is rarely judged by a single criterion. A platform typically balances several objectives, such as the relevance of an item to the user and its novelty, and user similarity is then *multi-faceted*: two users may share a taste for action movies (relevance) yet differ in their taste for blockbusters versus independent films (novelty). Multi-objective bandits evaluate a recommendation by its Pareto suboptimality with respect to all objectives (Drugan & Nowé, 2013; Tekin & Turğay, 2018; Turğay et al., 2018; Lu et al., 2019), but they learn every user separately and have no collaborative structure. Applying a single-structure clustering method to a multi-objective problem, on the other hand, forces one global grouping onto all objectives, which causes *negative transfer* for the objective on which the pooled users disagree. Under Pareto-based selection this is particularly harmful: an inaccurate estimate in

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.

a single objective changes dominance relations and excludes otherwise competitive items from the estimated Pareto set.

To fill in this gap, we propose Clustering of Multi-Objective Bandits (CMOB), a framework in which user communities may differ across objectives. It maintains one user graph per objective and pools data only within objective-specific clusters, so that two users can share data for one objective while being separated for another. We instantiate the framework as COMOLB for linear rewards, which deletes an edge of the graph of an objective only when a confidence test certifies that the two users differ on that objective; the test bounds the noise and the regularization bias of the estimates separately, so that it separates users as early as the data permit while never separating users of the same cluster. Furthermore, since rewards are often not linear in the context, we propose COMONB-Frozen, which runs the same scheme on the frozen gradient features of a randomly initialized network and tests edges in prediction space, with exact ridge estimators that keep the confidence bounds and the analysis exact. Our contributions are the following.

- **Framework and algorithms.** CMOB models multi-faceted user similarity through objective-specific user graphs. COMOLB (Section 3.1) and COMONB-Frozen (Section 3.2) instantiate it for linear rewards and for rewards that are linear in fixed network features, with deletion tests that certify every separation.
- **Finite-time analysis** (Section 4). With high probability, neither algorithm ever separates users of the same cluster, both recover the objective-specific partitions after an explicit identification time that depends on T only logarithmically, and their Pareto regret is at most twice this time plus $\widetilde{O}(d\sqrt{m_c T})$ for COMOLB and $\widetilde{O}(\bar{\beta}_T\sqrt{m_c T G_T})$ for COMONB-Frozen, where m_c is the total number of objective-specific clusters; the dominant term thus scales with the number of clusters instead of the number of users. We further show that the Pareto regret is dominated by any single-objective regret and give a preference-conditioned selection rule with a scalarized-regret guarantee.
- **Experiments** (Section 5). On synthetic instances that satisfy our conditions, the algorithms never separate users of the same cluster, recover the partitions in almost every run, and outperform every clustering baseline with the same selection rule; whether learning the clusters also beats independent learning depends on how different the clusters of an objective are, which we characterize. On semi-synthetic MovieLens data with objective-specific clusters, our method nearly halves the regret of independent learning.

Related work. Clustering of bandits (Gentile et al., 2014; Li et al., 2016; Li & Zhang, 2018; Li et al., 2019; Liu et al., 2022; Wang et al., 2023a;b; 2025; Ban & He, 2021; Ban et al., 2024b) always uses one similarity structure for the whole reward; multi-objective bandits, which measure performance by the Pareto regret (Drugan & Nowé, 2013; Tekin & Turğay, 2018; Turğay et al., 2018; Lu et al., 2019; Xu & Klabjan, 2023; Cheng et al., 2024; Xue et al., 2024), have no collaborative structure across users; and neural bandits with NTK-based confidence bounds (Jacot et al., 2018; Zhou et al., 2020; Zhang et al., 2021) do not consider user clustering. In our setting, different coordinates of one Pareto-UCB rule are computed from different cluster structures and covariance matrices, which single-graph analyses do not cover. A detailed discussion of related work is given in Appendix F.

2 Problem Setting

Vectors are boldface lowercase and matrices boldface uppercase (feature vectors $\phi^k(\boldsymbol{x})$ in plain type); $[m] := \{1, \dots, m\}$, $[a]_+ := \max\{a, 0\}$, $\|\boldsymbol{x}\|_{\boldsymbol{M}} := \sqrt{\boldsymbol{x}^\top \boldsymbol{M}\boldsymbol{x}}$ for a positive semi-definite (PSD) matrix $\boldsymbol{M}$, $\boldsymbol{A} \succeq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is PSD, and $\lambda_{\min}(\cdot)$ and $\lambda^+_{\min}(\cdot)$ denote the smallest and the smallest positive eigenvalue.

2.1 Collaborative Multi-Objective Contextual Bandits

A population of u users $\mathcal{U} = [u]$ interacts for T rounds with a recommender run by a platform (the operator of the system), and every recommendation is evaluated under m_o objectives. Let $\mathcal{X} := \{\boldsymbol{x} \in \mathbb{R}^d : \|\boldsymbol{x}\|_2 \le 1\}$. In round t, a user $i_t \in \mathcal{U}$ arrives, the learner observes a candidate set $\mathcal{X}_t = \{\boldsymbol{x}_{t,1}, \dots, \boldsymbol{x}_{t,K}\} \subset \mathcal{X}$, selects $\boldsymbol{x}_t \in \mathcal{X}_t$ and observes the reward vector $\boldsymbol{r}_t = (r_t^1, \dots, r_t^{m_o})^\top$

with

$$r_t^k = \mu_{i_t}^k(\boldsymbol{x}_t) + \eta_t^k, \qquad k \in [m_o], \tag{1}$$

where $\mu_i^k : \mathcal{X} \to [-1, 1]$ is the unknown mean reward function of user i for objective k and η_t^k is noise. $\mathcal{H}_{t-1}$ denotes the history $(i_s, \mathcal{X}_s, \boldsymbol{x}_s, \boldsymbol{r}_s)_{s<t}$ together with the internal randomization of the learner.

(A1) Uniform arrival and i.i.d. candidate sets. In every round, i_t is drawn uniformly at random from $\mathcal{U}$ and $\boldsymbol{x}_{t,1}, \ldots, \boldsymbol{x}_{t,K}$ are drawn i.i.d. from a fixed distribution $\mathcal{D}$ on $\mathcal{X}$; i_t, $\mathcal{X}_t$ and $\mathcal{H}_{t-1}$ are mutually independent.

(A2) Conditionally sub-Gaussian noise. There is a known $R > 0$ such that, for all t, k and $\nu \in \mathbb{R}$, $\mathbb{E}[\eta_t^k \mid \mathcal{H}_{t-1}, i_t, \mathcal{X}_t, \boldsymbol{x}_t] = 0$ and $\mathbb{E}[e^{\nu\eta_t^k} \mid \mathcal{H}_{t-1}, i_t, \mathcal{X}_t, \boldsymbol{x}_t] \le e^{\nu^2 R^2/2}$; the noise terms of one round need not be independent across objectives.

Objective-specific clusters. For each objective k, the users are partitioned into $m_c^{(k)}$ unknown clusters $\mathcal{C}^{(k)} = \{C_1^{(k)}, \ldots, C_{m_c^{(k)}}^{(k)}\}$, and users in the same cluster share the mean reward function of objective k: writing $j_k(i)$ for the index of the cluster of user i under objective k and $\mu_{C_j}^k$ for the common function of $C_j^{(k)}$, $\mu_i^k = \mu_{C_{j_k(i)}}^k$; two users may share a cluster for one objective but not for another. Let $m_c := \sum_k m_c^{(k)}$.

Pareto regret. For $\boldsymbol{a}, \boldsymbol{b} \in \mathbb{R}^{m_o}$, $\boldsymbol{a}$ Pareto-dominates $\boldsymbol{b}$, written $\boldsymbol{a} \succ \boldsymbol{b}$, if $a^k \ge b^k$ for all k and $a^k > b^k$ for some k. Let $\boldsymbol{\mu}_i(\boldsymbol{x}) := (\mu_i^1(\boldsymbol{x}), \ldots, \mu_i^{m_o}(\boldsymbol{x}))^\top$ and let $\mathbf{1}$ be the all-ones vector; the per-round loss is the Pareto suboptimality gap (Drugan & Nowé, 2013; Lu et al., 2019), abbreviated PSG, the infimum of the uniform increases of the mean rewards of $\boldsymbol{x} \in \mathcal{X}_t$ that make it non-dominated within $\mathcal{X}_t$:

$$\Delta_i(\boldsymbol{x}) := \inf\big\{\zeta \ge 0 : \ \boldsymbol{\mu}_i(\boldsymbol{x}') \not\succ \boldsymbol{\mu}_i(\boldsymbol{x}) + \zeta\mathbf{1} \ \text{ for all } \boldsymbol{x}' \in \mathcal{X}_t\big\}. \tag{2}$$

Equivalently (Appendix C.1),

$$\Delta_i(\boldsymbol{x}) = \max_{\boldsymbol{x}' \in \mathcal{X}_t} \min_{k \in [m_o]} \big[\mu_i^k(\boldsymbol{x}') - \mu_i^k(\boldsymbol{x})\big]_+, \tag{3}$$

and $0 \le \Delta_i(\boldsymbol{x}) \le 2$. The learner minimizes the cumulative Pareto regret $R_T := \sum_{t=1}^T \Delta_{i_t}(\boldsymbol{x}_t) \le 2T$.

2.2 Linear Preference Models

Cluster $C_j^{(k)}$ has a preference vector $\boldsymbol{\theta}_j^k \in \mathbb{R}^d$ with $\|\boldsymbol{\theta}_j^k\|_2 \le 1$, and

$$\mu_i^k(\boldsymbol{x}) = \langle \boldsymbol{\theta}_{j_k(i)}^k, \boldsymbol{x} \rangle, \qquad i \in \mathcal{U},\ k \in [m_o],\ \boldsymbol{x} \in \mathcal{X}. \tag{4}$$

(A3) Cluster separation. There is $\gamma > 0$ such that $\|\boldsymbol{\theta}_j^k - \boldsymbol{\theta}_{j'}^k\|_2 \ge \gamma$ for every $k \in [m_o]$ and all distinct $j, j' \in [m_c^{(k)}]$.

(A4) Context regularity. The context distribution $\mathcal{D}$ satisfies $\kappa_{x,K} > 0$, where

$$\kappa_{x,K} := \inf_{\boldsymbol{v} \in \mathbb{R}^d,\ \|\boldsymbol{v}\|_2 = 1} \mathbb{E}_{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K \overset{\text{i.i.d.}}{\sim} \mathcal{D}} \Big[\min_{a \in [K]} \langle \boldsymbol{v}, \boldsymbol{x}_a \rangle^2\Big]. \tag{5}$$

For every selection rule, $\mathbb{E}[\boldsymbol{x}_t \boldsymbol{x}_t^\top \mid \mathcal{H}_{t-1}, i_t] \succeq \kappa_{x,K} \boldsymbol{I}$, so every user's design matrix grows linearly in all directions despite adaptive selection. With $\lambda_x := \lambda_{\min}(\mathbb{E}_{\mathcal{D}}[\boldsymbol{x}\boldsymbol{x}^\top])$, one has $(\lambda_x/2)\big(\lambda_x/(2 - \lambda_x)\big)^K \le \kappa_{x,K} \le \lambda_x$ (Appendix C.1.1); $\kappa_{x,K}$ can be exponentially small in K but is of order K^{-2} for spherical contexts (Section 4.1 and Appendix G.1).

2.3 Fixed-Feature Preference Models

For nonlinear rewards, each objective is represented through a feature map fixed before the interaction, e.g., the gradient features of a ReLU network h frozen at a random initialization,

$\phi^k(\boldsymbol{x}) = \nabla_{\boldsymbol{\theta}} h(\boldsymbol{x}; \boldsymbol{\theta}_0)/(\sqrt{m_{\mathrm{NN}}}\, C)$, with m_{NN} the width, $\boldsymbol{\theta}_0$ drawn independently of the interaction (all statements are conditional on it) and $C > 0$ ensuring $\|\phi^k\|_2 \le 1$; no network is trained.

(A5) Fixed features. For each $k \in [m_o]$ there is a map $\phi^k : \mathcal{X} \to \mathbb{R}^{p_k}$ with $\|\phi^k(\boldsymbol{x})\|_2 \le 1$ on $\mathcal{X}$, chosen before the interaction and shared by all users, and for every cluster $C_j^{(k)}$ there is $\boldsymbol{w}_j^k \in \mathbb{R}^{p_k}$ with $\|\boldsymbol{w}_j^k\|_2 \le B$, B known, such that

$$\mu_i^k(\boldsymbol{x}) = \langle \phi^k(\boldsymbol{x}), \boldsymbol{w}_{j_k(i)}^k \rangle, \qquad i \in \mathcal{U},\ \boldsymbol{x} \in \mathcal{X}. \tag{6}$$

The linear model is the special case $\phi^k(\boldsymbol{x}) = \boldsymbol{x}$, $B = 1$. Since $\boldsymbol{w}_j^k$ is identified only through predictions at contexts drawn from $\mathcal{D}$, separation is imposed on predictions.

(A6) Observable separation. There are $\gamma' > 0$ and $\rho_\gamma > 0$ such that $\mathbb{P}_{\boldsymbol{x}\sim\mathcal{D}}\big(|\mu_{C_j}^k(\boldsymbol{x}) - \mu_{C_{j'}}^k(\boldsymbol{x})| \ge \gamma'\big) \ge \rho_\gamma$ for every $k \in [m_o]$ and all distinct $j, j' \in [m_c^{(k)}]$.

Hence the K i.i.d. candidates of a round contain such a context with probability at least $q_\gamma := 1 - (1-\rho_\gamma)^K$. Let $\boldsymbol{\Sigma}_\phi^k := \mathbb{E}_{\boldsymbol{x}\sim\mathcal{D}}[\phi^k(\boldsymbol{x})\phi^k(\boldsymbol{x})^\top] \ne \boldsymbol{0}$, $\mathcal{S}_k := \mathrm{Range}(\boldsymbol{\Sigma}_\phi^k)$ (which contains $\phi^k(\boldsymbol{x})$ almost surely), $r_k := \dim \mathcal{S}_k$, $r := \max_k r_k$ and $\lambda_\phi := \min_k \lambda_{\min}^+(\boldsymbol{\Sigma}_\phi^k)$. The analogue of (5) is

$$\kappa_{\phi,K} := \min_{k\in[m_o]} \inf_{\boldsymbol{v}\in\mathcal{S}_k,\ \|\boldsymbol{v}\|_2=1} \mathbb{E}_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_K \overset{\text{i.i.d.}}{\sim} \mathcal{D}} \Big[\min_{a\in[K]} \langle \boldsymbol{v}, \phi^k(\boldsymbol{x}_a)\rangle^2\Big] \ \ge\ \frac{\lambda_\phi}{2}\Big(\frac{\lambda_\phi}{2-\lambda_\phi}\Big)^K > 0, \tag{7}$$

abbreviated κ (the lower bound is shown in Appendix C.1.1), and for the regularization parameter $\lambda \ge 1$ the maximal information gain is

$$G_T := \max_{k\in[m_o]} \sup_{\substack{n\le T,\ \boldsymbol{x}_1,\dots,\boldsymbol{x}_n\in\mathcal{X}:\\ \phi^k(\boldsymbol{x}_s)\in\mathcal{S}_k\ \forall s\in[n]}} \log\det\Big(\boldsymbol{I} + \frac{1}{\lambda}\sum_{s=1}^n \phi^k(\boldsymbol{x}_s)\phi^k(\boldsymbol{x}_s)^\top\Big) \ \le\ r\log\Big(1 + \frac{T}{r\lambda}\Big). \tag{8}$$

Anti-concentration of $\boldsymbol{x}$ need not transfer to $\phi^k(\boldsymbol{x})$, so $\kappa_{\phi,K}$ can be exponentially small in K even for spherical contexts, and (A5) is a realizability condition on the chosen features (Appendix G.1).

3 Learning Algorithms

Both algorithms maintain one undirected user graph $G_t^k = (\mathcal{U}, E_t^k)$ per objective $k \in [m_o]$, initialized as the complete graph. In round t, the closed neighbourhood $\widehat{C}_t^k := \{i_t\} \cup \mathcal{N}_{G_{t-1}^k}(i_t)$ of the arriving user is its estimated cluster for objective k; the algorithm pools the ridge-regression statistics of $\widehat{C}_t^k$, forms one upper confidence bound (UCB) per objective, selects an arm whose UCB vector is not Pareto-dominated, updates the statistics of i_t only, and tests the edges incident to i_t separately for each objective, so that two users can share data for one objective while being separated for another. Pooling over the neighbourhood rather than the connected component keeps a single untested cross-cluster edge from merging two whole clusters; once the graphs are correct, both coincide with the true cluster. Throughout, $\lambda \ge 1$ is the regularization and $\delta \in (0,1)$ the confidence parameter, and for $\boldsymbol{M} = \lambda\boldsymbol{I} + \sum_s \boldsymbol{v}_s\boldsymbol{v}_s^\top$ we use the noise and confidence radii

$$\rho(\boldsymbol{M}) := R\sqrt{\log\frac{\det \boldsymbol{M}}{\det(\lambda\boldsymbol{I})} + 2\log\frac{8um_o}{\delta}}, \qquad \beta(\boldsymbol{M}) := B\sqrt{\lambda} + \rho(\boldsymbol{M}), \tag{9}$$

the anytime self-normalized radius of Abbasi-Yadkori et al. (2011) at level $\delta/(8um_o)$, where R is the noise level and B bounds the true parameters ($B = 1$ in the linear model): ρ bounds the noise and $B\sqrt{\lambda}$ the regularization bias. The radii require neither T nor knowledge of $\mathcal{D}$ or of the clusters; their coverage guarantee applies to groups of users fixed in advance, and for the pooled statistics of $\widehat{C}_t^k$ it is used only after the partitions have been recovered.

3.1 Clustering of Multi-Objective Linear Bandits

COMOLB (Algorithm 1) addresses the linear model (4): $\hat{\boldsymbol{w}}_{i,t}^k = \boldsymbol{M}_{i,t}^{-1}\boldsymbol{b}_{i,t}^k$ is the ridge estimate of $\boldsymbol{\theta}_{j_k(i)}^k$ from the data of user i alone, and $\bar{\boldsymbol{w}}_t^k$ the ridge estimate from the pooled samples of $\widehat{C}_t^k$ (the

Algorithm 1 Clustering of Multi-Objective Linear Bandits (COMOLB)

Require: $\lambda \geq 1$, $\delta \in (0,1)$, noise level R, parameter bound B ($B = 1$ in the linear model); radii $\rho(\cdot)$, $\beta(\cdot)$ of (9). Initialize $\boldsymbol{M}_{i,0} = \lambda \boldsymbol{I}$, $\boldsymbol{b}_{i,0}^k = \boldsymbol{0}$, $\hat{\boldsymbol{w}}_{i,0}^k = \boldsymbol{0}$, $T_{i,0} = 0$ and complete graphs G_0^k for all $i \in \mathcal{U}$, $k \in [m_o]$.

1: **for** $t = 1, 2, \ldots, T$ **do**
2: Receive user $i_t \in \mathcal{U}$ and candidate set $\mathcal{X}_t$.
3: **for** each objective $k \in [m_o]$ **do**
4: $\widehat{C}_t^k = \{i_t\} \cup \mathcal{N}_{G_{t-1}^k}(i_t)$; $\bar{\boldsymbol{M}}_t^k = \lambda \boldsymbol{I} + \sum_{i \in \widehat{C}_t^k} (\boldsymbol{M}_{i,t-1} - \lambda \boldsymbol{I})$, $\bar{\boldsymbol{b}}_t^k = \sum_{i \in \widehat{C}_t^k} \boldsymbol{b}_{i,t-1}^k$, $\bar{\boldsymbol{w}}_t^k = (\bar{\boldsymbol{M}}_t^k)^{-1} \bar{\boldsymbol{b}}_t^k$;
5: $U_t^k(\boldsymbol{x}) = \boldsymbol{x}^\top \bar{\boldsymbol{w}}_t^k + \beta(\bar{\boldsymbol{M}}_t^k) \, \|\boldsymbol{x}\|_{(\bar{\boldsymbol{M}}_t^k)^{-1}}$ for all $\boldsymbol{x} \in \mathcal{X}_t$.
6: **end for**
7: $\tilde{\mathcal{O}}_t = \{\boldsymbol{x} \in \mathcal{X}_t : \nexists \, \boldsymbol{x}' \in \mathcal{X}_t \text{ with } \boldsymbol{U}_t(\boldsymbol{x}') \succ \boldsymbol{U}_t(\boldsymbol{x})\}$ with $\boldsymbol{U}_t = (U_t^1, \ldots, U_t^{m_o})$; select $\boldsymbol{x}_t$ uniformly at random from $\tilde{\mathcal{O}}_t$ and observe $\boldsymbol{r}_t$.
8: Update user i_t only: $\boldsymbol{M}_{i_t,t} = \boldsymbol{M}_{i_t,t-1} + \boldsymbol{x}_t \boldsymbol{x}_t^\top$, $T_{i_t,t} = T_{i_t,t-1} + 1$, $\boldsymbol{b}_{i_t,t}^k = \boldsymbol{b}_{i_t,t-1}^k + r_t^k \boldsymbol{x}_t$, $\hat{\boldsymbol{w}}_{i_t,t}^k = \boldsymbol{M}_{i_t,t}^{-1} \boldsymbol{b}_{i_t,t}^k$ for all k; all other statistics are unchanged.
9: Deletion radii $r_{i,t}$ of (10) (only $r_{i_t,t}$ changes); for each k, delete every edge $(i_t, \ell) \in E_{t-1}^k$ with $\|\hat{\boldsymbol{w}}_{i_t,t}^k - \hat{\boldsymbol{w}}_{\ell,t}^k\|_2 > r_{i_t,t} + r_{\ell,t}$ (rule (10)).
10: **end for**

regularizer is counted once). Edges of G^k are tested in parameter space: with

$$r_{i,t} := \frac{\rho(\boldsymbol{M}_{i,t})}{\sqrt{\lambda_{\min}(\boldsymbol{M}_{i,t})}} + \frac{\lambda B}{\lambda_{\min}(\boldsymbol{M}_{i,t})}, \quad \text{delete } (i_t, \ell) \iff \left\|\hat{\boldsymbol{w}}_{i_t,t}^k - \hat{\boldsymbol{w}}_{\ell,t}^k\right\|_2 > r_{i_t,t} + r_{\ell,t}. \tag{10}$$

The ridge error decomposes as $\hat{\boldsymbol{w}}_{i,t}^k - \boldsymbol{\theta}_{j_k(i)}^k = \boldsymbol{M}_{i,t}^{-1}(\boldsymbol{S} - \lambda \boldsymbol{\theta}_{j_k(i)}^k)$, with $\boldsymbol{S}$ the noise sum, and the self-normalized bound gives $\|\boldsymbol{S}\|_{\boldsymbol{M}_{i,t}^{-1}} \leq \rho(\boldsymbol{M}_{i,t})$; hence the Euclidean ball of radius $r_{i,t}$ around $\hat{\boldsymbol{w}}_{i,t}^k$ contains $\boldsymbol{\theta}_{j_k(i)}^k$, and for two users of the same cluster the strict inequality in (10) cannot hold. Bounding the bias separately makes it decay like $1/\lambda_{\min}$ instead of $1/\sqrt{\lambda_{\min}}$: since $\lambda/\lambda_{\min} \leq \sqrt{\lambda}/\sqrt{\lambda_{\min}}$, the radius is never larger than the $\beta(\boldsymbol{M}_{i,t})/\sqrt{\lambda_{\min}(\boldsymbol{M}_{i,t})}$ implied by the confidence ellipsoid, it is strictly smaller once $\lambda_{\min}(\boldsymbol{M}_{i,t}) > \lambda$, markedly so at low noise, and it is equally valid. Under (A4) the radii shrink with the visits of i, and once both are at most $\gamma/6$, (A3) gives $\|\hat{\boldsymbol{w}}_{i_t,t}^k - \hat{\boldsymbol{w}}_{\ell,t}^k\|_2 \geq \gamma - r_{i_t,t} - r_{\ell,t} > r_{i_t,t} + r_{\ell,t}$ for users of different clusters, so the edge is deleted when tested.

3.2 Clustering of Multi-Objective Neural Bandits

For nonlinear rewards we analyze COMONB-Frozen under the fixed-feature model. It runs Algorithm 1 with $\boldsymbol{x}$ replaced by $\phi^k(\boldsymbol{x})$ in all statistics and with the bound B of (A5) in (9): user i keeps $\boldsymbol{M}_{i,t}^k = \lambda \boldsymbol{I} + \sum_{s \leq t:\, i_s = i} \phi^k(\boldsymbol{x}_s) \phi^k(\boldsymbol{x}_s)^\top$, $\boldsymbol{b}_{i,t}^k = \sum_{s \leq t:\, i_s = i} r_s^k \, \phi^k(\boldsymbol{x}_s)$, $\hat{\boldsymbol{w}}_{i,t}^k = (\boldsymbol{M}_{i,t}^k)^{-1} \boldsymbol{b}_{i,t}^k$ and $\hat{f}_{i,t}^k(\boldsymbol{x}) = \phi^k(\boldsymbol{x})^\top \hat{\boldsymbol{w}}_{i,t}^k$, and the pooled statistics and UCBs are formed as before with $\phi^k(\boldsymbol{x})$ in place of $\boldsymbol{x}$ (Algorithm 2 in Appendix A). Only the deletion rule differs: if $r_k < p_k$, then $\lambda_{\min}(\boldsymbol{M}_{i,t}^k) = \lambda$ for all t, so $r_{i,t}$ in (10) would never shrink, and COMONB-Frozen instead compares user-level predictions on the current candidates: with the width $b_{i,t}^k(\boldsymbol{z}) := \rho(\boldsymbol{M}_{i,t}^k) \, \|\phi^k(\boldsymbol{z})\|_{(\boldsymbol{M}_{i,t}^k)^{-1}} + \lambda B \, \|(\boldsymbol{M}_{i,t}^k)^{-1} \phi^k(\boldsymbol{z})\|_2$, which bounds the noise and the bias parts of the prediction error separately, after the update of i_t,

$$(i_t, \ell) \text{ is deleted from } G^k \iff \exists \, \boldsymbol{z} \in \mathcal{X}_t : \left|\hat{f}_{i_t,t}^k(\boldsymbol{z}) - \hat{f}_{\ell,t}^k(\boldsymbol{z})\right| > b_{i_t,t}^k(\boldsymbol{z}) + b_{\ell,t}^k(\boldsymbol{z}). \tag{11}$$

With high probability, $|\hat{f}_{i,t}^k(\boldsymbol{z}) - \mu_i^k(\boldsymbol{z})| \leq b_{i,t}^k(\boldsymbol{z})$ simultaneously for all users, objectives, rounds and points $\boldsymbol{z}$; then (11) never fires for two users of the same cluster, and it fires for users of different clusters in a round in which one of them arrives, $\mathcal{X}_t$ contains a context that separates their clusters by γ' (A6) and both widths there are at most $\gamma'/6$. All quantities depend on the features only through the kernel $\langle \phi^k(\boldsymbol{x}), \phi^k(\boldsymbol{x}') \rangle$ (Appendix D.10).

Practical variants. Two variants, not covered by our analysis, are evaluated only as ablations: the α-scaled rule multiplies the right-hand sides of (10) and (11) by $\alpha \in (0,1)$, which voids the argument excluding false deletions, and COMONB-Adam trains per-user ReLU networks online (Appendix B.3).

Platform-specified trade-offs. The analyses use the selection step only through $\boldsymbol{x}_t \in \tilde{\mathcal{O}}_t$ and the measurability conditions (R1)–(R3) of Appendix E.3. If the platform, i.e., the operator of the recommender, supplies before the selection in round t exogenous weights $\boldsymbol{\omega}_t$ in the probability simplex with strictly positive entries (condition (E) of Appendix E.3; e.g., a deterministic schedule) and the algorithm selects $\boldsymbol{x}_t \in \arg\max_{\boldsymbol{x}\in\mathcal{X}_t} \sum_k \omega_{t,k} U_t^k(\boldsymbol{x})$, with ties broken using only the past, i_t, $\mathcal{X}_t$ and $\boldsymbol{\omega}_t$, then $\boldsymbol{x}_t \in \tilde{\mathcal{O}}_t$, because a dominating UCB vector would have a strictly larger weighted sum; hence the PSG bounds of Section 4 hold unchanged, and the scalarized regret obeys the same bounds (Appendix E.3.3).

4 Theoretical Analysis

We bound $R_T = \sum_{t=1}^T \Delta_{i_t}(\boldsymbol{x}_t)$; κ denotes $\kappa_{x,K}$ in the linear and $\kappa_{\phi,K}$ in the fixed-feature model. Both analyses follow one chain: anytime confidence bounds hold simultaneously for all groups of users fixed in advance, and a matrix lower-tail inequality for adapted sequences shows that, under any selection rule, the Gram matrix of a user with n visits, minus $\lambda \boldsymbol{I}$, has all eigenvalues on the feature subspace at least $n\kappa/2$ once n exceeds a threshold of order κ^{-1} (up to a logarithm); on these events same-cluster edges are never deleted, every tested cross-cluster edge is eventually deleted, and after recovery the Pareto step and an elliptical-potential argument yield the $\sqrt{T}$ term.

4.1 Regret of COMOLB

Suppose that (A1)–(A4) hold and that $\lambda \geq 1$ and $\delta \in (0,1)$; Appendix C gives the analysis. Define

$$\bar{\rho}_{\mathrm{lin}} := R\sqrt{d\log\Big(1+\frac{T}{d\lambda}\Big) + 2\log\frac{8um_o}{\delta}}, \qquad \bar{\beta}_{\mathrm{lin}} := \sqrt{\lambda} + \bar{\rho}_{\mathrm{lin}},$$

$$n_\gamma := \min\Big\{\Big\lceil \frac{72\,\bar{\beta}_{\mathrm{lin}}^2}{\kappa_{x,K}\,\gamma^2}\Big\rceil,\ \max\Big\{\Big\lceil \frac{288\,\bar{\rho}_{\mathrm{lin}}^2}{\kappa_{x,K}\,\gamma^2}\Big\rceil, \Big\lceil \frac{24\lambda}{\kappa_{x,K}\,\gamma}\Big\rceil\Big\}\Big\},$$

$$n_{\mathrm{lin}} := \max\Big\{\Big\lceil \frac{8}{\kappa_{x,K}}\log\frac{4udT}{\delta}\Big\rceil,\ n_\gamma\Big\}, \qquad T_{\mathrm{lin}} := \Big\lceil 16u\log\frac{4u}{\delta} + 4u\,n_{\mathrm{lin}}\Big\rceil.$$

Then, with probability at least $1-\delta$, COMOLB satisfies: (i) for every $k \in [m_o]$, no edge between two users of the same cluster of objective k is ever deleted from G^k; (ii) if $T \geq T_{\mathrm{lin}}$, then at the end of round T_{lin} every graph G^k equals the true partition $\mathcal{C}^{(k)}$, i.e., its edges are exactly the pairs of users in a common cluster of objective k, so that $\widehat{C}_t^k = C^{(k)}_{j_k(i_t)}$ for all $k \in [m_o]$ and all $t > T_{\mathrm{lin}}$; and (iii)

$$R_T \leq \min\Big\{2T,\ 2T_{\mathrm{lin}} + 2\bar{\beta}_{\mathrm{lin}}\sqrt{2m_c\,dT\log\big(1+T/(d\lambda)\big)}\Big\}. \tag{12}$$

In particular, treating λ as a constant, $2T_{\mathrm{lin}} = \widetilde{O}\big(\frac{u}{\kappa_{x,K}}\big(1+\frac{R^2 d}{\gamma^2}+\frac{1}{\gamma}\big)\big)$, and treating also R as a constant, $R_T = \widetilde{O}\big(\frac{u}{\kappa_{x,K}}\big(1+\frac{d}{\gamma^2}\big) + d\sqrt{m_c T}\big)$.

Every user must be visited n_{lin} times before the graphs are guaranteed to be correct: the first term of n_{lin} ensures $\lambda_{\min}(\boldsymbol{M}_{i,t}) \geq \kappa_{x,K} T_{i,t}/2$, and n_γ makes the deletion radius at most $\gamma/6$. Because the bias is bounded separately, the identification cost per user is $\widetilde{O}\big(\kappa_{x,K}^{-1}(1+R^2 d/\gamma^2 + 1/\gamma)\big)$, which at low noise is much smaller than the $\widetilde{O}\big(\kappa_{x,K}^{-1}(1+(1+R\sqrt{d})^2/\gamma^2)\big)$ implied by the radius $\beta/\sqrt{\lambda_{\min}}$ alone. The learning cost depends on the population only through m_c (apart from $\log(8um_o/\delta)$), and on neither K nor γ.

Spherical contexts. K enters the bound only through the identification cost; since $\kappa_{x,K}$ measures the information of the least informative candidate, it can decrease with K, but for spherical contexts the dependence is polynomial. If $d \geq 3$ and $\mathcal{D}$ is the law of $\boldsymbol{x} = \boldsymbol{g}/\|\boldsymbol{g}\|_2$ with $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$, then

$$\kappa_{x,K} \ \geq\ \frac{1}{32\,c_d^2 K^2} \ \geq\ \frac{\pi}{16(d-1)K^2}, \qquad c_d := \frac{\Gamma(d/2)}{\sqrt{\pi}\,\Gamma((d-1)/2)} \leq \sqrt{\frac{d-1}{2\pi}}, \tag{13}$$

so that, under (A1)–(A3) and treating λ and R as constants, with probability at least $1-\delta$, $R_T = \widetilde{O}\big(udK^2(1+d/\gamma^2) + d\sqrt{m_c T}\big)$ (Appendix C.5). The exact value of $\kappa_{x,K}$ satisfies

$\kappa_{x,K} \sim 1/(2c_d^2K^2)$ as $K \to \infty$, so the lower bound in (13) is tight up to a constant factor. The identification cost can dominate the learning cost at practical horizons; conversely, T_{lin} and T_{clust} below are worst-case bounds, and a given instance may recover much earlier (Section 5).

4.2 Regret of COMONB-Frozen

Suppose that (A1), (A2), (A5) and (A6) hold, that $\mathbf{\Sigma}_\phi^k \neq \mathbf{0}$ for every $k \in [m_o]$, so that $\kappa := \kappa_{\phi,K} > 0$ (Appendix C.1.1), and that $\lambda \geq 1$ and $\delta \in (0,1)$; Appendix D gives the analysis. With q_γ, r and G_T as in Section 2.3, define

$$\bar{\rho}_T := R\sqrt{G_T + 2\log\frac{8um_o}{\delta}}, \qquad \bar{\beta}_T := B\sqrt{\lambda} + \bar{\rho}_T, \qquad n_{\mathrm{spec}} := \Big\lceil \frac{8}{\kappa}\log\frac{4um_o rT}{\delta}\Big\rceil,$$

$$n_\star := \max\Bigg\{ n_{\mathrm{spec}},\ \min\Big\{ \Big\lceil \frac{72\,\bar{\beta}_T^2}{\kappa\,\gamma'^2}\Big\rceil,\ \max\Big\{ \Big\lceil \frac{288\,\bar{\rho}_T^2}{\kappa\,\gamma'^2}\Big\rceil, \Big\lceil \frac{24\lambda B}{\kappa\,\gamma'}\Big\rceil \Big\}\Big\}\Bigg\},$$

$t_{\mathrm{fit}} := \lceil 16u\log(4u/\delta) + 4u\,n_\star\rceil$ and $T_{\mathrm{clust}} := t_{\mathrm{fit}} + \lceil (u/q_\gamma)\log(4u^2m_o/\delta)\rceil$. Then, with probability at least $1-\delta$, COMONB-Frozen satisfies: (i) for every $k \in [m_o]$, no edge between two users of the same cluster of objective k is ever deleted from G^k; (ii) if $T \geq T_{\mathrm{clust}}$, then at the end of round T_{clust} every graph G^k equals the true partition $\mathcal{C}^{(k)}$, so that $\widehat{C}_t^k = C_{j_k(i_t)}^{(k)}$ for all $k \in [m_o]$ and all $t > T_{\mathrm{clust}}$; and (iii) $R_T \leq \min\{2T,\ 2T_{\mathrm{clust}} + 2\bar{\beta}_T\sqrt{2m_cTG_T}\}$.

The covering term of T_{clust} accounts for the need of a separating candidate, which accompanies an arriving endpoint of a given cross-cluster edge with probability at least q_γ/u per round. For a fixed finite-dimensional feature map, $G_T = O(\log T)$, hence $T_{\mathrm{clust}} = O(\log T)$ and the bound grows as $O(\sqrt{T}\log T)$; the analysis (Appendix D) also gives a finer bound (63) in terms of the contexts actually selected for each true cluster. K enters only the identification cost, via $\kappa_{\phi,K}$ and q_γ.

4.3 What the Pareto Regret Certifies

The PSG counts every Pareto-optimal arm as optimal. Since $\Delta_i(\boldsymbol{x}) = \max_{\boldsymbol{x}'}\min_k[\mu_i^k(\boldsymbol{x}') - \mu_i^k(\boldsymbol{x})]_+$ and a minimum is at most any of its terms, for every user i, candidate set $\mathcal{X}_t$, arm $\boldsymbol{x} \in \mathcal{X}_t$ and objective $k_0 \in [m_o]$,

$$\Delta_i(\boldsymbol{x}) \leq \max_{\boldsymbol{x}' \in \mathcal{X}_t} \mu_i^{k_0}(\boldsymbol{x}') - \mu_i^{k_0}(\boldsymbol{x}), \qquad \text{hence} \qquad R_T \leq R_T^{(k_0)} \tag{14}$$

for every sequence of selected arms, where $R_T^{(k_0)} := \sum_{t=1}^T[\max_{\boldsymbol{x}\in\mathcal{X}_t}\mu_{i_t}^{k_0}(\boldsymbol{x}) - \mu_{i_t}^{k_0}(\boldsymbol{x}_t)]$ is the regret of objective k_0 (Appendix E.2). A small PSG thus does not require learning all objectives well; the PSG bounds above should be read together with parts (i)–(ii) of the guarantees, which assert recovery for every objective. Under the preference rule of Section 3.2 with exogenous weights $\boldsymbol{\omega}_t$, the scalarized regret $R_T^{\boldsymbol{\omega}} := \sum_{t=1}^T\big[\max_{\boldsymbol{x}\in\mathcal{X}_t}\boldsymbol{\omega}_t^\top\boldsymbol{\mu}_{i_t}(\boldsymbol{x}) - \boldsymbol{\omega}_t^\top\boldsymbol{\mu}_{i_t}(\boldsymbol{x}_t)\big]$ satisfies the same bounds as R_T (Appendix E.3.3), whereas uniform sampling from $\tilde{\mathcal{O}}_t$ can incur linear scalarized regret even with known rewards (Appendix E.3.4). On every fixed history, pooling over the objective-specific cluster never yields a larger geometric factor $\|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}}$ than pooling over the user's cell in any shared partition that never mixes true clusters, although this alone does not order the regrets of two adaptive algorithms (Appendix E.4).

5 Experimental Results

Protocol and learners. All UCB learners share one implementation with $\lambda = 1$, $\delta = 0.05$, $R = 0.1$ and the parameter bound B of the environment; apart from ablations, all clustering learners use the analyzed deletion test with the same radii, all graph-based learners pool over closed neighbourhoods, and arms are drawn uniformly from the UCB Pareto set unless stated otherwise. *Ind* pools nothing, *Pool-all* pools all users, *Oracle-OS* pools within the true objective-specific clusters and *Oracle-Global* within the cells of their common refinement, the coarsest single partition that never mixes clusters of any objective. *I-CLUB* keeps one graph and deletes an edge if any objective separates the two users, *S-CLUB* keeps one graph on scalarized estimates and selects by the fixed weights $\boldsymbol{\omega} = (0.5, 0.5)$, and *Shared-LOCB* and *OS-LOCB* adapt local clustering (Ban & He, 2021) with one shared or one per-objective group structure. Every configuration is run 10 times with paired seeds (details and complete results: Appendix B; discussion: Appendix G.3).

Table 1: Linear models with a shared component ($T = 8 \times 10^4$, 10 runs; $u = 40$, $m_o = 2$, $m_c^{(k)} = 4$, $d = 10$, $K = 20$; a smaller γ means a larger shared component): cumulative regret (mean $\pm$ std) and mean recovery time of COMOLB. Bold: lowest non-oracle value of a block. Complete results: Table 4.

Learner	$\gamma = 0.2$	$\gamma = 0.5$	$\gamma = 1.0$	$\gamma = 1.4$
PSG, uniform Pareto selection				
Oracle-OS	26.4± 2.4	25.4± 3.3	22.0± 2.2	21.4± 1.6
Oracle-Global	85.8± 8.8	85.6± 6.4	83.9± 7.7	83.4± 6.6
Ind	201.2± 20.0	202.6± 20.8	**197.8**± 14.7	**199.7**± 7.1
Pool-all	246.4± 26.6	1333.3± 177.4	5083.4± 738.7	11152.7± 1163.2
COMOLB (ours)	**151.1**± 12.5	**184.6**± 15.1	270.6± 22.7	385.9± 35.6
I-CLUB (one graph)	154.0± 12.6	199.2± 14.7	296.9± 22.5	424.2± 38.1
OS-LOCB	167.5± 13.5	199.0± 16.6	289.2± 28.5	408.7± 44.6
Shared-LOCB	166.0± 14.4	218.0± 16.3	324.1± 30.0	455.7± 43.5
COMOLB recovery time (rounds)	70750	13640	5270	3620
Scalarized regret $R_T^{(0.5,0.5)}$, selection by $\boldsymbol{\omega} = (0.5, 0.5)$				
Oracle-OS	25.2± 3.1	26.5± 1.8	27.2± 1.5	26.8± 1.3
Ind	247.9± 9.6	248.0± 9.0	**249.4**± 6.0	**249.0**± 5.1
COMOLB (ours)	**160.3**± 35.0	**225.3**± 42.0	387.7± 51.6	586.1± 38.1
I-CLUB (one graph)	162.3± 35.1	235.1± 44.2	409.1± 53.4	623.2± 43.1
S-CLUB (scalarized graph)	236.2± 49.5	482.4± 69.6	705.4± 74.9	976.7± 57.8

5.1 Linear Models: When Does Learning the Clusters Pay Off?

Each of $m_o = 2$ objectives partitions $u = 40$ users into 4 balanced random clusters, independently across objectives ($d = 10$, $K = 20$); the four parameters of an objective are unit vectors exactly γ apart that share a common component of norm $\sqrt{1 - \gamma^2/2}$, so a small γ models users who agree on most of an objective. Table 1 reports the results, from which we draw four conclusions.

The structure is valuable; learning it pays off for close clusters. Pooling within the true objective-specific clusters lowers the regret of Ind by a factor of about eight to nine, whereas the common refinement, whose cells are small, gains only a factor of about 2.4, and pooling all users is worse than Ind. COMOLB never deleted a same-cluster edge, recovered the partitions in every run, and has the lowest regret among the clustering learners with Pareto selection at every γ. Relative to Ind it is 25% better for $\gamma = 0.2$, on par for $\gamma = 0.5$, and 37% and 93% worse for $\gamma = 1.0$ and 1.4: almost all of its regret is incurred before recovery, while users still pool with neighbours from other clusters, and although this phase shortens as γ grows, the wrongly pooled neighbours are then more different. Learning the clusters thus pays off when they share a large component.

Both algorithmic choices matter. Pooling over components instead of neighbourhoods raises the regret of COMOLB by 23% to 33%, and additionally using the radius $\beta/\sqrt{\lambda_{\min}}$ by 63% to 227%; the deletion scale $\alpha = 0.3$, which the analysis does not cover, has the lowest PSG of all non-oracle learners with Pareto selection but deletes about a quarter of the same-cluster edges (Table 4).

Selection by a fixed trade-off. S-CLUB targets the scalarized regret, which bounds the PSG of every learner (Appendix E.2), so its PSG is not comparable with Pareto selection. When every learner selects by $\boldsymbol{\omega}$ (Table 1, bottom), COMOLB beats I-CLUB and S-CLUB in every run, by 32% to 53% relative to S-CLUB, and the comparison with Ind follows the same pattern as above.

Identification. Over 20 configurations of $\gamma \in \{0.2, \ldots, 1.2\}$ and $K \in \{5, \ldots, 100\}$ with 10 runs each (Table 6), COMOLB never deleted a same-cluster edge and always recovered. The recovery time grows with $1/\gamma$ more slowly than the bound, and only by about a factor of two over the 20-fold range of K, whereas $\kappa_{x,K}$ in the bound decreases by more than two orders of magnitude: the bound is conservative mainly because $\kappa_{x,K}$ is a worst case over selection rules.

5.2 Fixed-Feature and Nonlinear Rewards

All learners of this section use the frozen gradient features of a random ReLU network unless applied to the raw contexts (Table 8). On a theory-matched instance whose rewards are exactly linear in the features of a width-8 network ($p = 57$, $u = 20$, $m_o = 2$, $m_c^{(k)} = 4$, $K = 10$, $T = 10^5$), COMONB-

Table 2: MovieLens, $T = 10^5$ rounds, 10 runs: cumulative PSG and scalarized regret $R_T^{\boldsymbol{\omega}}$ for $\boldsymbol{\omega} = (0.5, 0.5)$ (mean $\pm$ std). The clustering learners use the prediction-space test (11) with identity features. In the raw version, the oracles pool over approximate k-means partitions.

	Exact clusters		Raw embeddings	
Learner	PSG	$R_T^{(0.5,0.5)}$	PSG	$R_T^{(0.5,0.5)}$
Oracle-OS	388.2± 5.6	7067± 23	2826.9± 20.7	9686± 38
Oracle-Global	803.0± 9.0	7796± 35	2799.0± 20.6	9967± 38
Ind	2157.2± 21.6	9326± 28	1771.8± 12.1	9363± 29
Pool-all	1956.2± 20.1	8366± 38	3331.6± 22.3	10302± 46
COMONB, $\phi(\boldsymbol{x}) = \boldsymbol{x}$ (ours)	1170.3± 16.5	7652± 26	2061.3± 13.7	8627± 29
I-CLUB (one graph)	1222.4± 17.5	7827± 38	2185.8± 17.7	9301± 22
OS-LOCB	1239.9± 77.5	7723± 83	2319.7± 49.0	8950± 66
Shared-LOCB	1309.8± 90.5	7921± 116	2455.1± 42.9	9649± 65
COMONB, $\phi(\boldsymbol{x}) = \boldsymbol{x}$, by $\boldsymbol{\omega}$ (ours)	332.2± 9.0	1884± 29	612.0± 4.1	3144± 18
S-CLUB (scalarized graph)	456.8± 5.3	2714± 12	776.5± 7.2	4779± 14
Single objective 1	1382.8± 43.1	6731± 59	2486.7± 32.6	8043± 52
Single objective 2	330.3± 10.3	6811± 69	597.7± 11.5	7180± 54

Frozen deleted no same-cluster edge, recovered the partitions in 9 of 10 runs, and attained a regret 30% below Ind and 16% below Oracle-Global in every run, whereas the single-graph I-CLUB did not improve on Ind; the bound of Section 4.2 is vacuous here because the feature covariance is nearly singular. On quadratic and cosine rewards, which are linear in neither feature map, COMONB-Frozen ended 35% above and 13% below Ind, and COMOLB on the raw contexts 21% and 5% below it.

5.3 MovieLens

From MovieLens 100K (Harper & Konstan, 2015) we take the 50 most active users and the 200 most rated movies; the context of a movie consists of its SVD factors and genre vector ($d = 27$), and the two objectives, satisfaction (SVD affinity) and novelty (genre dissimilarity to the user's profile), are linear in it (Appendix B.2). In the *exact* version, users are replaced by the centres of 3 k-means clusters per objective; in the *raw* version, they keep their own parameters ($K = 10$, $T = 10^5$). The contexts are nearly degenerate, so the parameter-space test of COMOLB deletes no edge within the horizon, and all clustering learners use the prediction-space test (11) with identity features, which Section 4.2 covers on the exact version. Table 2 reports the results: in the exact version, COMONB deleted no same-cluster edge, separated the clusters of the first objective in every run, and attained a regret 46% below Ind and below every clustering baseline with the same selection rule in every run; in the raw version, where users are not exactly clustered, even the oracles are worse than Ind, and COMONB is 16% above Ind but remains the best clustering learner with Pareto selection.

What the PSG certifies. The last rows of Table 2 illustrate (14): the learner that optimizes only novelty has the lowest mean PSG in both versions although it ignores satisfaction. The preference-conditioned rule of Section 3.2 serves the trade-off the platform asks for: with $\boldsymbol{\omega} = (0.5, 0.5)$, COMONB selecting by $\boldsymbol{\omega}$ attains a scalarized regret about a third below S-CLUB in every run and less than half of that of every learner with Pareto or single-objective selection, whereas uniform Pareto selection incurs linear scalarized regret (Table 7).

6 Discussion

We proposed objective-specific clustering for multi-objective contextual bandits and analyzed COMOLB and COMONB-Frozen: with high probability both never delete a same-cluster edge, recover the objective-specific partitions after an explicit identification time, and then learn at a rate governed by the total number of objective-specific clusters. Empirically, learning the clusters pays off when the clusters of an objective share a large component, on a theory-matched fixed-feature instance and on exactly clustered MovieLens data, but not for clusters that are far apart or only approximate. **Limitations.** Dense graphs need $O(m_o u^2)$ space in the worst case, and identification remains the main limitation: the theoretical identification time is conservative by orders of magnitude, and a smaller threshold without guarantee realizes much more of the structural gain.

AI USE STATEMENT

We used generative AI tools to improve the writing of this paper. We have reviewed all AI-assisted text and take responsibility for the final content of this work.

ETHICS STATEMENT

This is an algorithmic study of multi-objective contextual bandits that uses synthetic environments and the public MovieLens 100K dataset (Harper & Konstan, 2015), only for academic evaluation under the GroupLens usage terms, and involves no human-subject experiments or new data collection. Since clustering users by their inferred preferences could expose sensitive preference structure or amplify historical biases, a deployment should protect the learned user graphs and audit the recommendations across user groups.

REPRODUCIBILITY STATEMENT

The algorithms are specified in Algorithms 1 and 2, the standing conditions in Section 2, the complete derivations in Appendices C–E, and the environments, baselines, hyperparameters and complete result tables, with the number of independent runs of every experiment, in Appendix B. Code and reproduction instructions will be released upon publication.

REFERENCES

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 24, 2011.

Peter Auer. Using confidence bounds for exploitation-exploration trade-offs. *Journal of Machine Learning Research*, 3(Nov):397–422, 2002.

Peter Auer, Chao-Kai Chiang, Ronald Ortner, and Madalina Drugan. Pareto front identification from stochastic bandit feedback. In *Proceedings of the 19th International Conference on Artificial Intelligence and Statistics*, pp. 939–947, 2016.

Yikun Ban and Jingrui He. Local clustering in contextual multi-armed bandits. In *Proceedings of the Web Conference 2021*, pp. 2335–2346, 2021.

Yikun Ban, Yuchen Yan, Arindam Banerjee, and Jingrui He. EE-net: Exploitation-exploration neural networks in contextual bandits. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://openreview.net/forum?id=X_ch3VrNSRg.

Yikun Ban, Yunzhe Qi, and Jingrui He. Neural contextual bandits for personalized recommendation. In *Companion Proceedings of the ACM on Web Conference 2024*, WWW '24, pp. 1246–1249, 2024a.

Yikun Ban, Yunzhe Qi, Tianxin Wei, Lihui Liu, and Jingrui He. Meta clustering of neural bandits. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 95–106, 2024b.

Yuan Cao and Quanquan Gu. Generalization bounds of stochastic gradient descent for wide and deep neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 32, 2019.

Zekai Chen, Po-Yu Chen, and Francois Buet-Golfouse. Online personalizing white-box LLMs generation with neural bandits. In *Proceedings of the 5th ACM International Conference on AI in Finance*, pp. 711–718, 2024.

Ji Cheng, Bo Xue, Jiaxiang Yi, and Qingfu Zhang. Hierarchize Pareto dominance in multi-objective stochastic linear bandits. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pp. 11489–11497, 2024.

Wei Chu, Lihong Li, Lev Reyzin, and Robert Schapire. Contextual bandits with linear payoff functions. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 208–214, 2011.

Zhongxiang Dai, Yao Shu, Bryan Kian Hsiang Low, and Patrick Jaillet. Sample-then-optimize batch neural Thompson sampling. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 23331–23344, 2022.

Zhongxiang Dai, Yao Shu, Arun Verma, Flint Xiaofeng Fan, Bryan Kian Hsiang Low, and Patrick Jaillet. Federated neural bandits. In *International Conference on Learning Representations (ICLR)*, 2023.

Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. In *Proceedings of the 21st Annual Conference on Learning Theory*, pp. 355–366, 2008.

Madalina M Drugan and Ann Nowé. Designing multi-objective multi-armed bandits algorithms: A study. In *International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8, 2013.

Sarah Filippi, Olivier Cappé, Aurélien Garivier, and Csaba Szepesvári. Parametric bandits: The generalized linear case. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 23, 2010.

Claudio Gentile, Shuai Li, and Giovanni Zappella. Online clustering of bandits. In *International Conference on Machine Learning (ICML)*, pp. 757–765, 2014.

Quanquan Gu, Amin Karbasi, Khashayar Khosravi, Vahab Mirrokni, and Dongruo Zhou. Batched neural bandits. *ACM/JMS Journal of Data Science*, 1(1):1–18, 2024.

F Maxwell Harper and Joseph A Konstan. The MovieLens datasets: History and context. *ACM Transactions on Interactive Intelligent Systems*, 5(4):1–19, 2015.

Jiafan He, Dongruo Zhou, Tong Zhang, and Quanquan Gu. Nearly optimal algorithms for linear contextual bandits with adversarial corruptions. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 34614–34625, 2022.

Jiachen Hu, Xiaoyu Chen, Chi Jin, Lihong Li, and Liwei Wang. Near-optimal representation learning for linear bandits and linear RL. In *International Conference on Machine Learning (ICML)*, pp. 4349–4358, 2021.

Taehyun Hwang, Kyuwook Chai, and Min-hwan Oh. Combinatorial neural bandits. In *International Conference on Machine Learning (ICML)*, pp. 14203–14236, 2023.

Arthur Jacot, Franck Gabriel, and Clément Hongler. Neural tangent kernel: Convergence and generalization in neural networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 31, 2018.

Yiling Jia, Weitong Zhang, Dongruo Zhou, Quanquan Gu, and Hongning Wang. Learning neural contextual bandits through perturbed rewards. In *International Conference on Learning Representations (ICLR)*, 2022. URL `https://openreview.net/forum?id=7inCJ3MhXt3`.

Parnian Kassraie and Andreas Krause. Neural contextual bandits without regret. In *Proceedings of The 25th International Conference on Artificial Intelligence and Statistics*, pp. 240–278. PMLR, 2022.

Parnian Kassraie, Andreas Krause, and Ilija Bogunovic. Graph neural network bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 35, pp. 34519–34531, 2022.

Cyrille Kone, Emilie Kaufmann, and Laura Richert. Adaptive algorithms for relaxed Pareto set identification. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2023.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Shuai Li and Shengyu Zhang. Online clustering of contextual cascading bandits. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 32, 2018.

Shuai Li, Alexandros Karatzoglou, and Claudio Gentile. Collaborative filtering bandits. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*, pp. 539–548, 2016.

Shuai Li, Wei Chen, Shuai Li, and Kwong-Sak Leung. Improved algorithm on online clustering of bandits. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 2923–2929, 2019.

Xutong Liu, Haoru Zhao, Tong Yu, Shuai Li, and John CS Lui. Federated online clustering of bandits. In *Proceedings of the Thirty-Eighth Conference on Uncertainty in Artificial Intelligence*, pp. 1221–1231, 2022.

Shiyin Lu, Guanghui Wang, Yao Hu, and Lijun Zhang. Multi-objective generalized linear bandits. In *International Joint Conference on Artificial Intelligence (IJCAI)*, pp. 3080–3086, 2019.

Ofir Nabati, Tom Zahavy, and Shie Mannor. Online limited memory neural-linear bandits with likelihood matching. In *International Conference on Machine Learning (ICML)*, pp. 7905–7915. PMLR, 2021.

Thanh Nguyen-Tang, Sunil Gupta, A. Tuan Nguyen, and Svetha Venkatesh. Offline neural contextual bandits: Pessimism, optimization and generalization. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://openreview.net/forum?id=sPIFuucA3F.

Yunzhe Qi, Yikun Ban, and Jingrui He. Graph neural bandits. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pp. 1920–1931, 2023a.

Yunzhe Qi, Yikun Ban, Tianxin Wei, Jiaru Zou, Huaxiu Yao, and Jingrui He. Meta-learning with neural bandit scheduler. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 63994–64028, 2023b.

Carlos Riquelme, George Tucker, and Jasper Snoek. Deep Bayesian bandits showdown: An empirical comparison of Bayesian deep networks for Thompson sampling. In *International Conference on Learning Representations (ICLR)*, 2018.

Sudeep Salgia. Provably and practically efficient neural contextual bandits. In *International Conference on Machine Learning (ICML)*, pp. 29800–29844, 2023.

Aleksandrs Slivkins. Introduction to multi-armed bandits. *Foundations and Trends® in Machine Learning*, 12(1-2):1–286, 2019.

Cem Tekin and Eralp Turğay. Multi-objective contextual multi-armed bandit with a dominant objective. *IEEE Transactions on Signal Processing*, 66(14):3799–3813, 2018.

Eralp Turğay, Doruk Oner, and Cem Tekin. Multi-objective contextual bandit problem with similarity information. In *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, pp. 1673–1681, 2018.

Michal Valko, Nathan Korda, Rémi Munos, Ilias Flaounas, and Nello Cristianini. Finite-time analysis of kernelised contextual bandits. In *Proceedings of the Twenty-Ninth Conference on Uncertainty in Artificial Intelligence*, pp. 654–663, 2013.

Kristof Van Moffaert, Kevin Van Vaerenbergh, Peter Vrancx, and Ann Nowé. Multi-objective χ-armed bandits. In *International Joint Conference on Neural Networks (IJCNN)*, pp. 2331–2338, 2014.

Zhiyong Wang, Jize Xie, Xutong Liu, Shuai Li, and John C.S. Lui. Online clustering of bandits with misspecified user models. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, 2023a.

Zhiyong Wang, Jize Xie, Tong Yu, Shuai Li, and John C.S. Lui. Online corrupted user detection and regret minimization. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 36, pp. 33262–33287, 2023b.

Zhiyong Wang, Jiahang Sun, Mingze Kong, Jize Xie, Qinghua Hu, John C.S. Lui, and Zhongxiang Dai. Online clustering of dueling bandits. In *International Conference on Machine Learning (ICML)*, 2025. URL https://openreview.net/forum?id=EscpGI2XAx.

Mengfan Xu and Diego Klabjan. Pareto regret analyses in multi-objective multi-armed bandit. In *International Conference on Machine Learning (ICML)*, pp. 38499–38517, 2023.

Pan Xu, Zheng Wen, Handong Zhao, and Quanquan Gu. Neural contextual bandits with deep representation and shallow exploration. In *International Conference on Learning Representations (ICLR)*, 2022. URL https://openreview.net/forum?id=xnYACQquaGV.

Bo Xue, Ji Cheng, Fei Liu, Yimu Wang, and Qingfu Zhang. Multiobjective Lipschitz bandits under lexicographic ordering. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 38, pp. 16238–16246, 2024.

Bo Xue, Xi Lin, Xiaoyuan Zhang, and Qingfu Zhang. Multiple trade-offs: An improved approach for lexicographic linear bandits. In *AAAI Conference on Artificial Intelligence (AAAI)*, volume 39, pp. 21850–21858, 2025.

Junwen Yang, Zixin Zhong, and Vincent YF Tan. Optimal clustering with bandit feedback. *Journal of Machine Learning Research*, 25(186):1–54, 2024.

Tom Zahavy and Shie Mannor. Deep neural linear bandits: Overcoming catastrophic forgetting through likelihood matching. *arXiv preprint arXiv:1901.08612*, 2019.

Weitong Zhang, Dongruo Zhou, Lihong Li, and Quanquan Gu. Neural Thompson sampling. In *International Conference on Learning Representations (ICLR)*, 2021.

Dongruo Zhou, Lihong Li, and Quanquan Gu. Neural contextual bandits with UCB-based exploration. In *International Conference on Machine Learning (ICML)*, volume 119, pp. 11492–11502, 2020.

Yinglun Zhu, Dongruo Zhou, Ruoxi Jiang, Quanquan Gu, Rebecca Willett, and Robert Nowak. Pure exploration in kernel and neural bandits. In *Advances in Neural Information Processing Systems (NeurIPS)*, volume 34, pp. 11618–11630, 2021.

APPENDIX CONTENTS

A Pseudocode of COMONB-Frozen **15**

B Experimental Details and Additional Results **16**
B.1 Synthetic Environments . . . 17
B.2 MovieLens Dataset and Setup . . . 18
B.3 Learners, Hyperparameters and Run Time . . . 18
B.4 Complete Linear Results . . . 19
B.5 Identification with the Analyzed Rule: Dependence on γ and K . . . 21
B.6 Platform Preferences on the Linear Instance . . . 21

C Analysis of COMOLB **22**
C.1 Auxiliary Quantities . . . 23
C.2 Matrix Lower Tail and the Selected Context . . . 25
C.3 Confidence Sets, Edge Deletion and Cluster Recovery . . . 26
C.4 The Regret Bound . . . 30
C.5 Spherical Contexts . . . 32

D Analysis of COMONB-Frozen **34**
D.1 Setting, Notation and Preliminaries . . . 34
D.2 Step 1: Anytime Confidence Bounds for Fixed Groups . . . 36
D.3 Step 2: Spectral Growth on the Effective Subspaces . . . 37
D.4 Step 3: No False Deletion . . . 38
D.5 Step 4: Deletion of Cross-Cluster Edges . . . 38
D.6 Step 5: Arrivals and Covering . . . 38
D.7 Step 6: Cluster Recovery and Timing . . . 39
D.8 Step 7: Regret Decomposition . . . 39
D.9 Sublinearity . . . 41
D.10 Kernel Form of the Estimators . . . 41
D.11 Exploration, Misspecification and Inexact Solvers . . . 42

E PSG Limitation, Platform Preferences and Information Advantage **45**
E.1 Notation and the Max–Min Form of the PSG . . . 45
E.2 The PSG Is Dominated by Any Single-Objective Regret . . . 46
E.3 Preference-Conditioned Selection and Scalarized Regret . . . 47
E.4 Information Advantage of Objective-Specific Pooling . . . 51
E.5 How These Results Are Used in the Experiments . . . 52

F Extended Related Work **53**

G Additional Notes and Discussion **54**
G.1 Notes on the Standing Conditions . . . 54
G.2 Notes on the Algorithms and the Guarantees . . . 54
G.3 Complete Discussion of the Experiments . . . 56

Algorithm 2 Clustering of Multi-Objective Neural Bandits with Frozen Features (COMONB-Frozen)

Require: Regularization $\lambda \geq 1$; confidence level $\delta \in (0,1)$; noise level R; parameter bound B; frozen feature maps $\phi^k : \mathcal{X} \to \mathbb{R}^{p_k}$ with $\|\phi^k(\boldsymbol{x})\|_2 \leq 1$, $k \in [m_o]$; radii $\rho(\cdot)$ and $\beta(\cdot)$ from (9). (Every prediction, width and log-determinant below can equivalently be computed from Gram matrices by (19).)

Initialization: $\boldsymbol{M}_{i,0}^k = \lambda \boldsymbol{I}$, $\boldsymbol{b}_{i,0}^k = \boldsymbol{0}$, $\hat{\boldsymbol{w}}_{i,0}^k = \boldsymbol{0}$, $T_{i,0} = 0$ for all $i \in \mathcal{U}$, $k \in [m_o]$; complete graphs $G_0^k = (\mathcal{U}, E_0^k)$, $k \in [m_o]$.
for $t = 1, 2, \ldots, T$ **do**
  Receive user $i_t \in \mathcal{U}$ and candidate set $\mathcal{X}_t = \{\boldsymbol{x}_{t,1}, \ldots, \boldsymbol{x}_{t,K}\}$.
  for $k \in [m_o]$ **do**
    $\widehat{C}_t^k = \{i_t\} \cup \mathcal{N}_{G_{t-1}^k}(i_t)$, the closed neighbourhood of i_t in G_{t-1}^k.
    Compute $\bar{\boldsymbol{M}}_t^k$, $\bar{\boldsymbol{b}}_t^k$ and $\bar{\boldsymbol{w}}_t^k = (\bar{\boldsymbol{M}}_t^k)^{-1}\bar{\boldsymbol{b}}_t^k$ by (16).
    Compute $U_t^k(\boldsymbol{x}) = \phi^k(\boldsymbol{x})^\top \bar{\boldsymbol{w}}_t^k + \beta(\bar{\boldsymbol{M}}_t^k)\|\phi^k(\boldsymbol{x})\|_{(\bar{\boldsymbol{M}}_t^k)^{-1}}$ for all $\boldsymbol{x} \in \mathcal{X}_t$.
  end for
  $\tilde{\mathcal{O}}_t = \big\{\boldsymbol{x} \in \mathcal{X}_t : \nexists\, \boldsymbol{x}' \in \mathcal{X}_t \text{ with } \boldsymbol{U}_t(\boldsymbol{x}') \succ \boldsymbol{U}_t(\boldsymbol{x})\big\}$, where $\boldsymbol{U}_t = (U_t^1, \ldots, U_t^{m_o})^\top$.
  Select $\boldsymbol{x}_t \sim \mathrm{Uniform}(\tilde{\mathcal{O}}_t)$ and observe $\boldsymbol{r}_t = (r_t^1, \ldots, r_t^{m_o})^\top$.
  $T_{i_t,t} = T_{i_t,t-1} + 1$ (once per round); for $i \neq i_t$, keep $T_{i,t} = T_{i,t-1}$ and all statistics of i.
  for $k \in [m_o]$ **do**
    $\boldsymbol{M}_{i_t,t}^k = \boldsymbol{M}_{i_t,t-1}^k + \phi^k(\boldsymbol{x}_t)\phi^k(\boldsymbol{x}_t)^\top$, $\boldsymbol{b}_{i_t,t}^k = \boldsymbol{b}_{i_t,t-1}^k + r_t^k \phi^k(\boldsymbol{x}_t)$, $\hat{\boldsymbol{w}}_{i_t,t}^k = (\boldsymbol{M}_{i_t,t}^k)^{-1}\boldsymbol{b}_{i_t,t}^k$.
  end for
  for $k \in [m_o]$ **do**
    $E_t^k = E_{t-1}^k$.
    for each ℓ with $(i_t, \ell) \in E_{t-1}^k$ **do**
      if $|\hat{f}_{i_t,t}^k(\boldsymbol{z}) - \hat{f}_{\ell,t}^k(\boldsymbol{z})| > b_{i_t,t}^k(\boldsymbol{z}) + b_{\ell,t}^k(\boldsymbol{z})$ for some $\boldsymbol{z} \in \mathcal{X}_t$, i.e., test (18) **then**
        Remove (i_t, ℓ) from E_t^k.
      end if
    end for
  end for
end for

A PSEUDOCODE OF COMONB-FROZEN

This appendix gives the complete pseudocode of COMONB-Frozen (Algorithm 2), the fixed-feature algorithm analyzed in Section 4.2. For every objective $k \in [m_o]$, the algorithm uses a feature map $\phi^k : \mathcal{X} \to \mathbb{R}^{p_k}$ with $\|\phi^k(\boldsymbol{x})\|_2 \leq 1$ that is fixed before the interaction and is the same for all users, typically the normalized parameter gradient of a ReLU network at a random initialization that is subsequently kept frozen (Jacot et al., 2018; Zhou et al., 2020). The network is never trained: every estimator is a ridge regression on the frozen features. COMONB-Frozen has the same structure as COMOLB (Algorithm 1) with $\boldsymbol{x}$ replaced by $\phi^k(\boldsymbol{x})$. It pools statistics over the closed neighbourhood of the arriving user in the objective-specific graph G^k, selects the arm from the approximate Pareto set of the resulting UCB vectors, and updates only the statistics of the arriving user. The edge-deletion test, however, is carried out in prediction space on the current candidate set. This matches the observable-separation condition (A6), which requires different clusters to differ in their reward functions on a set of contexts of positive probability; the parameters themselves are identified only through their projections onto the effective subspace $\mathcal{S}_k$ (Section 2.3), whose dimension r_k can be much smaller than p_k.

Ridge statistics. Fix $\lambda \geq 1$ and $\delta \in (0,1)$, let R be the noise level of (A2), and let B be the parameter bound of (A5). For a user i, an objective k and a round t, the user-level statistics, which are those of Section 3.2, are

$$\begin{aligned} \boldsymbol{M}_{i,t}^k &= \lambda \boldsymbol{I} + \sum_{s \leq t:\, i_s = i} \phi^k(\boldsymbol{x}_s)\phi^k(\boldsymbol{x}_s)^\top, & \boldsymbol{b}_{i,t}^k &= \sum_{s \leq t:\, i_s = i} r_s^k\, \phi^k(\boldsymbol{x}_s), \\ \hat{\boldsymbol{w}}_{i,t}^k &= (\boldsymbol{M}_{i,t}^k)^{-1}\boldsymbol{b}_{i,t}^k, & \hat{f}_{i,t}^k(\boldsymbol{x}) &= \phi^k(\boldsymbol{x})^\top \hat{\boldsymbol{w}}_{i,t}^k. \end{aligned} \tag{15}$$

If the same feature map is used for all objectives, $\boldsymbol{M}_{i,t}^k$ does not depend on k and is stored once. In round t, the cluster-level statistics of objective k are formed on the closed neighbourhood $\widehat{C}_t^k = \{i_t\} \cup \mathcal{N}_{G_{t-1}^k}(i_t)$ of the arriving user in the graph of the previous round, and the regularizer is counted once:

$$\bar{\boldsymbol{M}}_t^k = \lambda \boldsymbol{I} + \sum_{i \in \widehat{C}_t^k} \big(\boldsymbol{M}_{i,t-1}^k - \lambda \boldsymbol{I}\big), \qquad \bar{\boldsymbol{b}}_t^k = \sum_{i \in \widehat{C}_t^k} \boldsymbol{b}_{i,t-1}^k, \qquad \bar{\boldsymbol{w}}_t^k = (\bar{\boldsymbol{M}}_t^k)^{-1} \bar{\boldsymbol{b}}_t^k. \tag{16}$$

Confidence radii and deletion test. The UCBs use the confidence radius $\beta(\boldsymbol{M}) = B\sqrt{\lambda} + \rho(\boldsymbol{M})$ of (9), with the parameter bound B of (A5); the deletion test bounds the noise part of a prediction error by the noise radius $\rho(\boldsymbol{M})$ and its bias part separately. Both are evaluated at the current matrix, so that no horizon-dependent constant enters the algorithm. Below we write the log-determinant term as $\log\det(\lambda^{-1}\boldsymbol{M}) = \log(\det \boldsymbol{M} / \det(\lambda \boldsymbol{I}))$. The UCB of objective k in round t and the deletion width of user i are

$$\begin{aligned} U_t^k(\boldsymbol{x}) &= \phi^k(\boldsymbol{x})^\top \bar{\boldsymbol{w}}_t^k + \beta(\bar{\boldsymbol{M}}_t^k)\, \|\phi^k(\boldsymbol{x})\|_{(\bar{\boldsymbol{M}}_t^k)^{-1}}, \\ b_{i,t}^k(\boldsymbol{z}) &= \rho(\boldsymbol{M}_{i,t}^k)\, \|\phi^k(\boldsymbol{z})\|_{(\boldsymbol{M}_{i,t}^k)^{-1}} + \lambda B \, \big\|(\boldsymbol{M}_{i,t}^k)^{-1} \phi^k(\boldsymbol{z})\big\|_2; \end{aligned} \tag{17}$$

the scalar width $b_{i,t}^k(\boldsymbol{z})$ should not be confused with the vector $\boldsymbol{b}_{i,t}^k$ in (15). Since $\|(\boldsymbol{M}_{i,t}^k)^{-1}\phi^k(\boldsymbol{z})\|_2 \le \lambda^{-1/2} \|\phi^k(\boldsymbol{z})\|_{(\boldsymbol{M}_{i,t}^k)^{-1}}$, the width never exceeds $\beta(\boldsymbol{M}_{i,t}^k)\|\phi^k(\boldsymbol{z})\|_{(\boldsymbol{M}_{i,t}^k)^{-1}}$, and in well-explored directions of $\mathcal{S}_k$ its bias part is much smaller. After the statistics of i_t have been updated in round t, an edge $(i_t, \ell) \in E_{t-1}^k$ is deleted from G^k if and only if

$$\exists\, \boldsymbol{z} \in \mathcal{X}_t : \qquad \big|\hat{f}_{i_t,t}^k(\boldsymbol{z}) - \hat{f}_{\ell,t}^k(\boldsymbol{z})\big| > b_{i_t,t}^k(\boldsymbol{z}) + b_{\ell,t}^k(\boldsymbol{z}). \tag{18}$$

Only edges incident to i_t are tested, the inequality is strict, and deleted edges are never restored. The interaction counter of i_t is incremented once per round, independently of the number of objectives.

Kernel form. All quantities above depend on the features only through the kernel $\mathcal{K}^k(\boldsymbol{x}, \boldsymbol{x}') = \langle \phi^k(\boldsymbol{x}), \phi^k(\boldsymbol{x}') \rangle$. Let S be a finite set of rounds, namely the rounds $s \le t$ with $i_s = i$ for the statistics of user i, or the rounds $s \le t-1$ with $i_s \in \widehat{C}_t^k$ for the cluster statistics, and let $\boldsymbol{M}_S = \lambda \boldsymbol{I} + \sum_{s \in S} \phi^k(\boldsymbol{x}_s)\phi^k(\boldsymbol{x}_s)^\top$, so that $\boldsymbol{M}_S = \boldsymbol{M}_{i,t}^k$ or $\boldsymbol{M}_S = \bar{\boldsymbol{M}}_t^k$. With the Gram matrix $\boldsymbol{K}_S = (\mathcal{K}^k(\boldsymbol{x}_s, \boldsymbol{x}_{s'}))_{s,s' \in S}$, the vector $\boldsymbol{k}_S(\boldsymbol{x}) = (\mathcal{K}^k(\boldsymbol{x}_s, \boldsymbol{x}))_{s \in S}$ and the rewards $\boldsymbol{r}_S = (r_s^k)_{s \in S}$, Appendix D.10 shows that the ridge prediction $\hat{f}_S(\boldsymbol{x}) = \phi^k(\boldsymbol{x})^\top \boldsymbol{M}_S^{-1} \sum_{s \in S} r_s^k \phi^k(\boldsymbol{x}_s)$ and the width $s_S(\boldsymbol{x}) = \|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}_S^{-1}}$ satisfy

$$\hat{f}_S(\boldsymbol{x}) = \boldsymbol{k}_S(\boldsymbol{x})^\top (\boldsymbol{K}_S + \lambda \boldsymbol{I})^{-1} \boldsymbol{r}_S, \qquad s_S^2(\boldsymbol{x}) = \frac{1}{\lambda}\Big[\mathcal{K}^k(\boldsymbol{x}, \boldsymbol{x}) - \boldsymbol{k}_S(\boldsymbol{x})^\top (\boldsymbol{K}_S + \lambda \boldsymbol{I})^{-1} \boldsymbol{k}_S(\boldsymbol{x})\Big], \tag{19}$$

and $\log\det(\lambda^{-1}\boldsymbol{M}_S) = \log\det(\boldsymbol{I} + \lambda^{-1}\boldsymbol{K}_S)$. The bias part of the deletion width is also a kernel quantity: expanding $\boldsymbol{M}_S^{-1}\phi^k(\boldsymbol{x}) = \lambda^{-1}\big(\phi^k(\boldsymbol{x}) - \sum_{s \in S} \phi^k(\boldsymbol{x}_s)[(\boldsymbol{K}_S + \lambda \boldsymbol{I})^{-1} \boldsymbol{k}_S(\boldsymbol{x})]_s\big)$ gives $\|\boldsymbol{M}_S^{-1}\phi^k(\boldsymbol{x})\|_2^2 = \lambda^{-1}\big[s_S^2(\boldsymbol{x}) - \|(\boldsymbol{K}_S + \lambda \boldsymbol{I})^{-1}\boldsymbol{k}_S(\boldsymbol{x})\|_2^2\big]$ (67). Algorithm 2 can therefore be run with $|S| \times |S|$ Gram matrices instead of $p_k \times p_k$ matrices. The factor $1/\lambda$ in the width is needed for the kernel and primal forms to coincide.

Selection rule and practical variants. The analysis uses the selection step only through $\boldsymbol{x}_t \in \tilde{\mathcal{O}}_t$. Instead of sampling uniformly from $\tilde{\mathcal{O}}_t$, the platform may select $\boldsymbol{x}_t \in \arg\max_{\boldsymbol{x} \in \mathcal{X}_t} \sum_k \omega_{t,k} U_t^k(\boldsymbol{x})$ for exogenous weights $\boldsymbol{\omega}_t$ in the simplex (Appendix E.3.3). For strictly positive weights, this arm lies in $\tilde{\mathcal{O}}_t$, so the PSG guarantee is unchanged (Section 3.2). The appendix ablations also use two practical variants that are *not* covered by the guarantee of Section 4.2: (a) α-scaled deletion, in which the right-hand side of (18) is multiplied by some $\alpha < 1$, so that the no-false-deletion argument (Appendix D.4) no longer applies; and (b) COMONB-Adam, which replaces the ridge estimators by a network trained with a few warm-started Adam steps per round, so that neither the exact ridge solution nor the residual certificate of Appendix D.11.3 is available.

B Experimental Details and Additional Results

All learners except COMONB-Adam are implemented in one code base in which the pooling structure, the deletion test and the selection rule are the only differences. Every configuration is run 10 times,

with seeds $0, \dots, 9$. Every run draws its own environment instance (except the fixed MovieLens instances and the fixed frozen feature map) and its own arrivals, candidate sets and noise, from a random stream that does not depend on the learner, so all learners run with the same seed face the same instance, arrivals, candidate sets and noise; we use this pairing when we report in how many runs one learner beats another. Recovery is checked every 50 rounds on the crossed instances and every 100 rounds otherwise; a run has recovered at the first checkpoint from which on every graph of the learner has the target partition as its connected components (the true objective-specific partitions for COMOLB and COMONB-Frozen, their common refinement for I-CLUB and S-CLUB). For the learners with the analyzed test, which never deleted a same-cluster edge in our realizable runs, this is equivalent to the graph being equal to the target partition, as in part (ii) of the guarantees of Section 4.

B.1 Synthetic Environments

Contexts and noise. Candidates are drawn i.i.d. as $\boldsymbol{x} = \boldsymbol{g}/\|\boldsymbol{g}\|_2$ with $\boldsymbol{g} \sim \mathcal{N}(0, \boldsymbol{I}_d)$, i.e., uniformly on the unit sphere, for which $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top] = \boldsymbol{I}/d$ and $\kappa_{x,K}$ is given exactly by (45) and bounded below by (13). Rewards are the mean rewards plus independent $\mathcal{N}(0, 0.1^2)$ noise in every objective, so (A2) holds with $R = 0.1$.

Linear instances. In all linear instances, the $m_c^{(k)}$ parameters of objective k are $\boldsymbol{\theta}_c^k = \sqrt{1-\gamma^2/2}\,\boldsymbol{w}^k + (\gamma/\sqrt{2})\,\boldsymbol{v}_c^k$, where $\boldsymbol{w}^k, \boldsymbol{v}_1^k, \dots, \boldsymbol{v}_{m_c^{(k)}}^k$ are the orthonormal columns of the Q factor of a $d \times (m_c^{(k)} + 1)$ Gaussian matrix drawn in each run and for each objective. They have unit norm and pairwise distance exactly γ and share the component $\sqrt{1-\gamma^2/2}\,\boldsymbol{w}^k$, whose norm is 0.99, 0.94, 0.71 and 0.14 for $\gamma = 0.2$, 0.5, 1.0 and 1.4; $\gamma = \sqrt{2}$ would give orthonormal parameters. In the *shared-component* instance of Table 1 ($u = 40$, $m_o = 2$, $m_c^{(k)} = 4$, $d = 10$, $K = 20$, $T = 8 \times 10^4$) and in the identification sweep ($u = 20$, $m_o = 2$, $m_c^{(k)} = 4$, $d = 5$, $T = 10^5$), every objective has its own balanced random partition of the users, drawn independently across objectives and runs; in the shared-component instance, the common refinement of the two partitions has 14 to 16 cells with 2.6 users on average, against 10 users per objective-specific cluster. In the *crossed* instance ($\gamma = 0.8$, $d = 8$, $K = 10$, $T = 2000u$), user $(a, b) \in [n]^2$ belongs to cluster a for objective 1 and to cluster b for objective 2, so the common refinement is the singleton partition.

Frozen-feature instance. The feature map is $\phi(\boldsymbol{x}) = \nabla_{\boldsymbol{\theta}} h(\boldsymbol{x}; \boldsymbol{\theta}_0)/(\sqrt{m_{\mathrm{NN}}}\,C)$ for the network $h(\boldsymbol{x}) = \sum_{j=1}^{m_{\mathrm{NN}}} a_j \,\mathrm{relu}(\boldsymbol{w}_j^\top \boldsymbol{x} + c_j) + e$ of width $m_{\mathrm{NN}} = 8$ on $d = 5$ inputs, initialized with $\boldsymbol{w}_j \sim \mathcal{N}(0, 2\boldsymbol{I}_d/d)$, $a_j \sim \mathcal{N}(0, 2/m_{\mathrm{NN}})$ and $c_j = e = 0$, and frozen. With $C = \sqrt{(2\sum_j a_j^2 + \sum_j \|\boldsymbol{w}_j\|_2^2 + 1)/m_{\mathrm{NN}}}$ we have $\|\phi(\boldsymbol{x})\|_2 \le 1$ for all $\|\boldsymbol{x}\|_2 \le 1$. The map has $p = m_{\mathrm{NN}} d + 2m_{\mathrm{NN}} + 1 = 57$ coordinates; since $\mathrm{relu}(\boldsymbol{w}_j^\top \boldsymbol{x}) = \boldsymbol{w}_j^\top [a_j \mathbb{1}\{\boldsymbol{w}_j^\top \boldsymbol{x} > 0\}\boldsymbol{x}]/a_j$, the m_{NN} output-weight coordinates are linear functions of the input-weight coordinates, and Σ_ϕ has rank $r = p - m_{\mathrm{NN}} = 49$ with smallest positive eigenvalue $\lambda_\phi = 1.17 \times 10^{-7}$ (estimated from 4×10^4 contexts). The cluster parameters are $\boldsymbol{w}_j^k = B\,\boldsymbol{U}\boldsymbol{a}/\|\boldsymbol{a}\|_2$ with $B = 1$, $\boldsymbol{a} \sim \mathcal{N}(0, \boldsymbol{I}_6)$ and $\boldsymbol{U}$ the top six eigenvectors of Σ_ϕ, so that (A5) holds exactly and $|\mu| \le 0.58$ on the estimation sample. Across runs, the fraction of contexts on which two clusters of the same objective differ by at least $\gamma' = 0.1$ is at least $\rho_\gamma \in [0.09, 0.46]$, so (A6) holds with $\gamma' = 0.1$; for $\gamma' = 0.2$ it is 0 in one run. The same feature map is used in all runs; the partitions and parameters vary.

Nonlinear instances. For every objective and cluster, the quadratic rewards are $\mu(\boldsymbol{x}) = \frac{1}{2}(\langle \boldsymbol{\theta}, \boldsymbol{x}\rangle + \boldsymbol{x}^\top \boldsymbol{S}\boldsymbol{x})$ with $\|\boldsymbol{\theta}\|_2 = 1$ and $\boldsymbol{S}$ a symmetrized Gaussian matrix scaled to spectral norm 1, and the cosine rewards are $\mu(\boldsymbol{x}) = \cos(3\boldsymbol{v}^\top \boldsymbol{x} + \pi/4)$ with $\|\boldsymbol{v}\|_2 = 1$; both lie in $[-1, 1]$. The linear part and the phase break the symmetry $\mu(\boldsymbol{x}) = \mu(-\boldsymbol{x})$; without them, the best linear approximation on the symmetric sphere would be identically zero. The best linear fit explains about two thirds of the variance of the rewards. The frozen learners use a width-16 network on $d = 10$ inputs ($p = 193$); since the rewards are not realizable, $B = 1$ is a tuning constant.

B.2 MovieLens Dataset and Setup

We use MovieLens 100K (Harper & Konstan, 2015), a public benchmark distributed by GroupLens, only for academic evaluation; we refer to the GroupLens README file for the applicable usage terms. We keep the 50 users with the most ratings and the 200 most rated movies (ties broken by the lower identifier), set $R_{ij} = 1$ if user i rated movie j higher than 3 and $R_{ij} = 0$ otherwise, and compute a truncated SVD $\boldsymbol{R} \approx \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^\top$ with 10 dimensions; the rows $\boldsymbol{u}_i$ of $\boldsymbol{U}\boldsymbol{S}$ and $\boldsymbol{v}_j$ of $\boldsymbol{V}$ are the user and movie factors. Let $\hat{\boldsymbol{g}}_j$ be the normalized 19-dimensional genre indicator of movie j and $\boldsymbol{\pi}_i$ the normalized sum of $\hat{\boldsymbol{g}}_j$ over the catalogue movies with $R_{ij} = 1$. The context of movie j is $\boldsymbol{x}_j = [\boldsymbol{v}_j / v_{\max},\ \hat{\boldsymbol{g}}_j]/\sqrt{2}$, with $v_{\max} = \max_j \|\boldsymbol{v}_j\|_2$, so $\|\boldsymbol{x}_j\|_2 \le 1$. Two of the 19 genres (*unknown* and *Documentary*) occur in none of the 200 movies; their coordinates vanish for every movie and every $\boldsymbol{\pi}_i$, and we drop them, so $\boldsymbol{x}_j \in \mathbb{R}^{27}$ and the uniform distribution over the catalogue has a nonsingular second-moment matrix, with smallest eigenvalue $\lambda_x \approx 1.5 \times 10^{-3}$.

- Satisfaction: $\mu_i^1(j) = \langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle / A_1 \in [-1, 1]$ with $A_1 = \max_{i,j} |\langle \boldsymbol{u}_i, \boldsymbol{v}_j \rangle|$ computed from the user factors actually used (the cluster centres in the exact version below).
- Novelty: $\mu_i^2(j) = -\langle \boldsymbol{\pi}_i, \hat{\boldsymbol{g}}_j \rangle \in [-1, 0]$, the cosine distance $1 - \cos(\boldsymbol{\pi}_i, \hat{\boldsymbol{g}}_j)$ shifted by the constant 1, which changes neither the PSG nor any scalarized regret.

Both rewards are linear in $\boldsymbol{x}_j$. In the *exact* version, $\boldsymbol{u}_i$ is replaced by the centre of its cluster in a 3-means clustering of the user factors and $\boldsymbol{\pi}_i$ by the renormalized centre of its cluster in a 3-means clustering of the genre profiles (Lloyd's algorithm with at most 100 iterations, best of 10 random initializations by within-cluster sum of squares, fixed seed; cluster sizes $13/23/14$ and $16/21/13$), so each objective has an exact partition and the two partitions differ; the resulting parameter bound is $B = 3.62$ and the smallest separation is $\gamma = 0.30$. In the *raw* version, every user keeps its own $\boldsymbol{u}_i$ and $\boldsymbol{\pi}_i$ ($B = 3.13$); the k-means partitions are then only approximate and are used by the oracles. In every round a user is drawn uniformly at random, $K = 10$ candidates are drawn uniformly with replacement from the 200 movies, and both rewards carry $\mathcal{N}(0, 0.1^2)$ noise.

B.3 Learners, Hyperparameters and Run Time

All UCB learners use the radii $\rho(\cdot)$ and $\beta(\cdot)$ of (9) with $\lambda = 1$, $\delta = 0.05$, $R = 0.1$ and the parameter bound B of the environment. Every learner that learns a clustering uses the analyzed deletion test without scaling: rule (10) on the linear instances and for COMOLB on the raw contexts of the other instances, and the prediction-space rule (11) with the widths $b_{i,t}^k$ for all other clustering learners on the frozen-feature and nonlinear instances and on MovieLens (Section 5.3). The graph-based learners pool over the closed neighbourhood of the arriving user in the relevant graph. I-CLUB keeps one graph, deletes an edge if the test fires for any objective, and pools over the same neighbourhood for all objectives. S-CLUB keeps one graph on the estimates $\sum_k \omega_k \hat{\boldsymbol{w}}_{i,t}^k$ with uniform weights $\omega_k = 1/m_o$, tests them with the same radii (which bound the error of every convex combination of the estimates of a user, because all objectives share the matrix $\boldsymbol{M}_{i,t}$), and selects the arm with the largest UCB of the scalarized reward. The LOCB adaptations draw $2 \max\{2, \lceil \log u \rceil\}$ seed users uniformly at random in each run; each seed anchors a group that initially contains all users; a user's membership in a seed's group is tested only in the rounds in which that user arrives, against the seed's current estimate, and the user leaves the group for good when the deletion test separates the two (on any objective for Shared-LOCB, on the group's own objective for OS-LOCB). The arriving user pools over the group, among those containing it, whose seed has the most observations; a user that belongs to no group pools only its own data. Pool-all pools all users, and the oracles pool over the fixed true partitions. Selection is uniform over the UCB Pareto set except for S-CLUB and for the preference-conditioned and single-objective rules (Sections 5.1 and 5.3). The ablations in Table 4 change one element of COMOLB at a time: pooling over the connected component instead of the closed neighbourhood; the radius $\beta(\boldsymbol{M}_{i,t})/\sqrt{\lambda_{\min}(\boldsymbol{M}_{i,t})}$ implied by the confidence ellipsoid instead of (10), together with component pooling; and the practical deletion scale $\alpha = 0.3$, which multiplies the right-hand side of (10) and is not covered by the guarantee of Section 4.1.

COMONB-Adam. COMONB-Adam is a trained-network variant of COMONB-Frozen that is not covered by our analysis; we describe what it does. For every objective, each user has one ReLU network with two hidden layers of width 32 (d–32–32–1). To decide, it copies the arriving user's

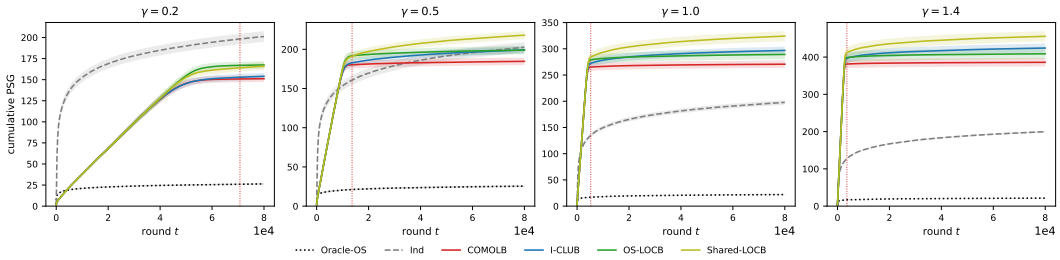

Figure 1: Shared-component instance: mean cumulative PSG over 10 runs with uniform Pareto selection (shaded: $\pm$ one standard error); the dotted vertical line marks the mean recovery time of COMOLB.

network for every objective, runs 10 Adam steps (learning rate 10^{-3}, weight decay 10^{-5}) on all data of the user's connected component, and selects uniformly from the Pareto set of the predicted means, without an exploration bonus. After observing the rewards, it runs 5 Adam steps of the user's networks on the new sample. After a warm-up of $5u$ rounds, it deletes the edge (i_t, ℓ) of objective k if the two users' predictions at the chosen context differ by more than $0.5\,(f(T_{i_t}) + f(T_\ell))$ with $f(n) = (2\sqrt{\lambda} + 2)/\sqrt{2n}$. Its cost per round grows with the size of the component, so it was run for 5000 rounds.

Run time. Table 3 reports the mean wall-clock time per round (selection and update), measured on an Intel Core i7-13700 (24 hardware threads) while about 20 runs were executed in parallel; only relative values within a column are meaningful. The per-round cost of the graph-based learners is dominated by aggregating the statistics of the arriving user's neighbourhood and, for the prediction-space test, by evaluating the widths of all its neighbours on the K candidates.

Table 3: Mean wall-clock time per round in milliseconds (mean over 10 runs). Linear: shared-component instance with $\gamma = 0.5$; frozen: fixed-feature instance ($p = 57$); quadratic: width-16 frozen features ($p = 193$), COMONB-Adam for 5000 rounds; MovieLens: exact version, where the clustering learners use the prediction-space test.

Learner	Linear, $\gamma = 0.5$	Frozen ($p = 57$)	Quadratic ($p = 193$)	MovieLens exact
Ind	0.30	0.38	1.22	0.29
Oracle-OS	0.39	0.57	3.49	0.42
COMOLB / COMONB-Frozen	0.53	1.50	13.94	6.08
I-CLUB	0.42	0.62	6.78	3.06
S-CLUB	0.49	–	–	3.88
OS-LOCB	0.51	–	–	1.34
Shared-LOCB	0.37	–	–	0.81
COMONB-Adam	–	–	39.59	–

B.4 Complete Linear Results

Table 4 completes Table 1 with the PSG of S-CLUB, the recovery times of the graph-based learners and three ablations of COMOLB, and Figure 1 shows the regret curves; Table 5 reports the crossed instance.

Theoretical and measured identification times. The identification time T_{lin} of Section 4.1, evaluated with the exact $\kappa_{x,K}$ of the spherical contexts, is 1.5×10^9, 2.4×10^8, 6.0×10^7 and 3.4×10^7 rounds for the shared-component instance with $\gamma = 0.2$, 0.5, 1.0 and 1.4 (measured: 7.1×10^4, 1.4×10^4, 5.3×10^3 and 3.6×10^3), 1.6×10^7 for the crossed instance with $u = 36$ (measured: 4605), and 1.1×10^6 and 2.8×10^8 for the identification sweep with $\gamma = 0.8$ and $K = 5$ or $K = 100$ (measured: 1420 and 2440). The bound correctly predicts the qualitative roles of γ and u but is conservative by factors between about 6×10^2 and 1.2×10^5, increasingly so with K, because $\kappa_{x,K}$ is a worst case over selection rules, whereas the measured spectral growth of the selected contexts decreases much more slowly with K (Table 6).

Table 4: Shared-component instance, complete results ($T = 8 \times 10^4$, 10 runs): final cumulative PSG (mean $\pm$ std) and, in brackets, the mean recovery time in rounds of the graph-based learners; a superscript gives the number of recovered runs if not all recovered. For I-CLUB and S-CLUB the target partition is the common refinement. †S-CLUB selects by $\boldsymbol{\omega} = (0.5, 0.5)$, so its PSG is not comparable with that of the learners with uniform Pareto selection (Appendix E.2); Table 1 compares it by its scalarized regret. The last three rows change one element of COMOLB at a time; the deletion scale $\alpha = 0.3$ is not covered by the guarantee of Section 4.1. With $\alpha = 0.3$, COMOLB deleted 82 to 98 same-cluster edges per run on average; no other graph-based learner deleted one.

Learner	$\gamma = 0.2$	$\gamma = 0.5$	$\gamma = 1.0$	$\gamma = 1.4$
Oracle-OS	26.4± 2.4	25.4± 3.3	22.0± 2.2	21.4± 1.6
Oracle-Global	85.8± 8.8	85.6± 6.4	83.9± 7.7	83.4± 6.6
Ind	201.2± 20.0	202.6± 20.8	197.8± 14.7	199.7± 7.1
Pool-all	246.4± 26.6	1333.3± 177.4	5083.4± 738.7	11152.7± 1163.2
COMOLB	151.1± 12.5 [70750]	184.6± 15.1 [13640]	270.6± 22.7 [5270]	385.9± 35.6 [3620]
I-CLUB	154.0± 12.6 [66400]	199.2± 14.7 [13310]	296.9± 22.5 [5180]	424.2± 38.1 [3470]
S-CLUB†	15.6± 2.5 [–]	70.2± 6.5 [57500^7]	220.4± 20.7 [20770]	444.5± 27.9 [11890]
OS-LOCB	167.5± 13.5	199.0± 16.6	289.2± 28.5	408.7± 44.6
Shared-LOCB	166.0± 14.4	218.0± 16.3	324.1± 30.0	455.7± 43.5
COMOLB, comp. pooling	199.7± 15.8 [69130]	227.9± 18.3 [13620]	343.3± 34.3 [5410]	512.5± 52.6 [3760]
COMOLB, $\beta/\sqrt{\lambda_{\min}}$, comp. pooling	246.4± 26.6 [–]	603.8± 56.6 [37260]	652.5± 71.6 [10390]	836.4± 89.6 [6150]
COMOLB, $\alpha = 0.3$	31.6± 2.2 [11671^7]	52.0± 3.1 [3675^8]	79.6± 6.1 [1720]	103.0± 8.7 [1278^9]

Table 5: Crossed instance ($\gamma = 0.8$, $d = 8$, $K = 10$, 10 runs): final cumulative PSG (mean $\pm$ std) at the horizon $T = 2000u$ given in the header and, in brackets, the mean recovery time in rounds of the graph-based learners. The common refinement is the singleton partition, so Oracle-Global coincides with Ind. †As in Table 4.

Learner	$u = 4$ $T = 8000$	$u = 9$ $T = 18000$	$u = 16$ $T = 32000$	$u = 36$ $T = 72000$
Oracle-OS	6.3± 1.0	8.3± 0.9	11.7± 1.6	18.7± 2.2
Oracle-Global	12.0± 2.1	26.2± 2.7	45.1± 6.2	107.1± 12.5
Ind	12.0± 2.1	26.2± 2.7	45.1± 6.2	107.1± 12.5
Pool-all	195.4± 26.8	573.2± 70.9	1118.6± 137.3	2740.7± 349.9
COMOLB	12.9± 1.5 [415]	30.1± 2.9 [1020]	57.3± 5.0 [1895]	142.0± 10.7 [4605]
I-CLUB	14.8± 1.8 [420]	35.8± 3.4 [990]	68.9± 5.4 [1820]	171.7± 11.8 [4425]
S-CLUB†	6.8± 0.8 [1090]	17.0± 2.2 [3555]	35.7± 4.6 [9485]	88.9± 11.2 [27510]
OS-LOCB	13.0± 1.5	31.0± 3.5	60.5± 6.9	153.0± 12.6
Shared-LOCB	14.6± 1.6	36.2± 3.3	69.1± 5.8	171.8± 12.7

Table 6: Identification with the analyzed rule: mean recovery time in rounds over 10 runs ($u = 20$, $d = 5$, $m_o = 2$, $m_c^{(k)} = 4$, $T = 10^5$); every run recovered. The last two rows compare the measured spectral growth $\min_i \lambda_{\min}(\boldsymbol{M}_{i,T} - \lambda \boldsymbol{I})/T_{i,T}$ at $\gamma = 0.8$ with the exact $\kappa_{x,K}$ of (45); uniform selection would give $1/d = 0.2$.

	$K = 5$	$K = 10$	$K = 20$	$K = 50$	$K = 100$
$\gamma = 0.2$	10530	12100	13970	18630	22600
$\gamma = 0.4$	3480	3880	4290	5570	7070
$\gamma = 0.8$	1420	1560	1760	2080	2440
$\gamma = 1.2$	980	990	1130	1320	1430
Measured spectral growth	0.166	0.140	0.113	0.083	0.065
$\kappa_{x,K}$ (exact)	0.0224	0.0069	0.0019	0.00034	0.000086

B.5 Identification with the Analyzed Rule: Dependence on γ and K

The guarantee of Section 4.1 places γ and K only in the identification term, through n_γ, which scales like $R^2 d/\gamma^2 + 1/\gamma$ up to logarithmic factors, and through $\kappa_{x,K}$, which is $\Theta(K^{-2})$ for spherical contexts (see (13) and (45)). We run COMOLB on linear instances whose cluster parameters are exactly γ apart ($u = 20$, $d = 5$, $m_o = 2$, $m_c^{(k)} = 4$, $T = 10^5$, 10 runs) for $\gamma \in \{0.2, 0.4, 0.8, 1.2\}$ and $K \in \{5, 10, 20, 50, 100\}$ (Table 6). No same-cluster edge was deleted in any of the 200 runs, and every run recovered the partitions within the horizon. The recovery time grows like γ^{-a}, with fitted exponents a between 1.33 and 1.55 across K, i.e., less steeply than n_γ in the bound, which at these parameters is dominated by its $1/\gamma^2$ noise term. Over the 20-fold range of K the mean recovery time grows by a factor between 1.46 and 2.15, whereas $\kappa_{x,K}$ decreases by a factor of about 260. The measured spectral growth of the selected contexts, $\min_i \lambda_{\min}(\boldsymbol{M}_{i,T} - \lambda \boldsymbol{I})/T_{i,T}$, lies between $\kappa_{x,K}$ and the value $1/d = 0.2$ of uniform selection: the UCB rule selects contexts that are far better spread than the worst case the bound must cover. In every cell, more than 82% of the regret is incurred before recovery on average.

B.6 Platform Preferences on the Linear Instance

Table 7 reports the preference experiment on the shared-component instance with $\gamma = 0.5$ ($T = 8 \times 10^4$, 10 runs). All rules use the COMOLB estimators and graphs with the analyzed test; only the selection step differs, so the guarantee of Appendix E.3.3 applies to the preference-conditioned rules. All rules recovered the partitions in every run (after 1.4×10^4 to 1.5×10^4 rounds on average) without false deletions. Each preference-conditioned rule attains the lowest scalarized regret for its own weight, e.g., 225 for $\boldsymbol{\omega} = (0.5, 0.5)$, against 7811 for uniform Pareto selection and 8731 and 8776 for the single-objective rules, and each single-objective rule is almost as good as the preference rule with the extreme weight on its objective (315 against 314 and 316 against 313). Uniform Pareto selection, which suffices for the PSG guarantee, incurs a scalarized regret that grows linearly with T, about 1910 per quarter of the horizon for $\boldsymbol{\omega} = (0.5, 0.5)$, whereas the preference-conditioned rule adds at most about 1 per quarter after the first, as the counterexample of Appendix E.3.4 allows. The PSG orders the rules differently: it is lowest for the balanced weights (38.5) and highest for uniform Pareto selection (184.6).

Table 7: Selection rules on the shared-component instance with $\gamma = 0.5$: final PSG (mean $\pm$ std) and scalarized regret $R_T^{\boldsymbol{\omega}}$ (mean over 10 runs) for five weights; bold: lowest scalarized regret of each column.

Selection rule	PSG	$(0.99, 0.01)$	$(0.75, 0.25)$	$(0.5, 0.5)$	$(0.25, 0.75)$	$(0.01, 0.99)$
Uniform Pareto	184.6± 15.1	20277	11597	7811	11572	20239
Preference $(0.99, 0.01)$	146.3± 10.5	**314**	1762	8525	22836	41630
Preference $(0.75, 0.25)$	56.5± 8.8	2241	**252**	3436	14168	29526
Preference $(0.5, 0.5)$	38.5± 6.1	12634	3982	**225**	4016	12711
Preference $(0.25, 0.75)$	57.1± 6.5	29541	14185	3444	**250**	2241
Preference $(0.01, 0.99)$	146.4± 10.5	41713	22904	8566	1776	**313**
Single objective 1	154.1± 9.9	315	1863	8731	23146	42040
Single objective 2	155.0± 10.4	42130	23220	8776	1880	316

C ANALYSIS OF COMOLB

This appendix proves the COMOLB guarantee of Section 4.1 and the spherical bound (13). Appendix C.1 gives the equivalent form of the PSG, defines the spectral parameter and bounds it from below. Appendix C.2 proves a lower-tail inequality for sums of adapted random matrices and shows that, under any selection rule, the selected context is at least as informative as the least informative candidate. These results are stated for a general bounded feature map: the linear model corresponds to the identity map, and Appendix D reuses them for fixed features. Appendix C.3 proves the confidence bounds, the absence of false deletions and the recovery of the objective-specific partitions, Appendix C.4 completes the proof of the guarantee, and Appendix C.5 treats spherical contexts.

Standing conventions. Throughout, $\lambda \geq 1$, $\delta \in (0,1)$, and every context satisfies $\|\boldsymbol{x}\|_2 \leq 1$. The selection rule may use internal randomization (for instance, uniform sampling from the approximate Pareto set). Let $\mathcal{H}_t$ be the σ-field generated by $(i_s, \mathcal{X}_s, \boldsymbol{x}_s, \boldsymbol{r}_s)_{s\leq t}$ together with the internal randomization of the selection rule in rounds $1,\dots,t$, and let $\mathcal{H}_0$ be trivial. We further use the σ-fields

$$\mathcal{H}_t^- := \sigma(\mathcal{H}_{t-1}, i_t), \qquad \mathcal{H}_{t-1}^+ := \sigma(\mathcal{H}_{t-1}, i_t, \mathcal{X}_t, \boldsymbol{x}_t).$$

Thus $\mathcal{H}_t^-$ is the information available before the candidate set of round t is revealed, and $\mathcal{H}_{t-1}^+$ is the information available after the action of round t has been chosen and before its reward is observed. We read (A1) as stating that i_t is uniform on $\mathcal{U}$ and independent of $\mathcal{H}_{t-1}$; in particular, $i_1, i_2, \dots$ are i.i.d. The candidate set $\mathcal{X}_t = \{\boldsymbol{x}_{t,1},\dots,\boldsymbol{x}_{t,K}\}$ consists of K i.i.d. draws from $\mathcal{D}$ and is independent of $\mathcal{H}_t^- = \sigma(\mathcal{H}_{t-1}, i_t)$; since $\mathcal{H}_{t-1}$ contains the internal randomization of rounds $1,\dots,t-1$, $\mathcal{X}_t$ is in particular independent of that randomization. (A2) states that, for all $k \in [m_o]$, $t \geq 1$ and $\nu \in \mathbb{R}$,

$$\mathbb{E}\big[\eta_t^k \,\big|\, \mathcal{H}_{t-1}^+\big] = 0, \qquad \mathbb{E}\big[\exp(\nu\eta_t^k) \,\big|\, \mathcal{H}_{t-1}^+\big] \leq \exp\big(\nu^2 R^2/2\big); \tag{20}$$

no independence between the noise variables of different objectives is assumed. Appendix C.2 uses only $\boldsymbol{x}_t \in \mathcal{X}_t$. Appendix C.3 uses in addition that $\boldsymbol{x}_t$ is chosen without using the rewards of round t or later, so that (20) holds, and Appendix C.4 uses in addition $\boldsymbol{x}_t \in \tilde{\mathcal{O}}_t$.

Quantities maintained by COMOLB. In the linear model $\mu_i^k(\boldsymbol{x}) = \langle \boldsymbol{\theta}^k_{j_k(i)}, \boldsymbol{x}\rangle$ with $\|\boldsymbol{\theta}_j^k\|_2 \leq 1$, so that $|\mu_i^k(\boldsymbol{x})| \leq 1$. For a user $i \in \mathcal{U}$, an objective $k \in [m_o]$ and $t \geq 0$, Algorithm 1 maintains

$$T_{i,t} = \sum_{s\leq t} \mathbb{I}\{i_s = i\}, \quad \boldsymbol{M}_{i,t} = \lambda \boldsymbol{I} + \sum_{s\leq t:\, i_s = i} \boldsymbol{x}_s \boldsymbol{x}_s^\top, \quad \boldsymbol{b}_{i,t}^k = \sum_{s\leq t:\, i_s = i} r_s^k \boldsymbol{x}_s, \quad \hat{\boldsymbol{w}}_{i,t}^k = \boldsymbol{M}_{i,t}^{-1}\boldsymbol{b}_{i,t}^k. \tag{21}$$

The statistics of a user who does not arrive in round t do not change in that round, the counter $T_{i_t,t}$ is incremented once per round (not once per objective), and the matrix $\boldsymbol{M}_{i,t}$ is shared by all objectives. The noise and confidence radii are

$$\rho(\boldsymbol{M}) := R\sqrt{\log\det(\lambda^{-1}\boldsymbol{M}) + 2\log(8um_o/\delta)}, \qquad \beta(\boldsymbol{M}) := \sqrt{\lambda} + \rho(\boldsymbol{M}), \tag{22}$$

that is, the radii of the fixed-feature model with $B = 1$. Let G_0^k be the complete graph on $\mathcal{U}$ and let $G_t^k = (\mathcal{U}, E_t^k)$ denote the graph of objective k at the end of round t. In round t, for every k, COMOLB forms the closed neighbourhood $\widehat{C}_t^k := \{i_t\} \cup \mathcal{N}_{G_{t-1}^k}(i_t)$ of i_t and the cluster statistics

$$\bar{\boldsymbol{M}}_t^k = \lambda \boldsymbol{I} + \sum_{i\in\widehat{C}_t^k} \big(\boldsymbol{M}_{i,t-1} - \lambda \boldsymbol{I}\big), \qquad \bar{\boldsymbol{b}}_t^k = \sum_{i\in\widehat{C}_t^k} \boldsymbol{b}_{i,t-1}^k, \qquad \bar{\boldsymbol{w}}_t^k = (\bar{\boldsymbol{M}}_t^k)^{-1}\bar{\boldsymbol{b}}_t^k, \tag{23}$$

in which the regularizer is counted once. It computes $U_t^k(\boldsymbol{x}) = \langle \bar{\boldsymbol{w}}_t^k, \boldsymbol{x}\rangle + \beta(\bar{\boldsymbol{M}}_t^k)\|\boldsymbol{x}\|_{(\bar{\boldsymbol{M}}_t^k)^{-1}}$, writes $\boldsymbol{U}_t(\boldsymbol{x}) = (U_t^1(\boldsymbol{x}),\dots,U_t^{m_o}(\boldsymbol{x}))$, and selects $\boldsymbol{x}_t$ in the approximate Pareto set $\tilde{\mathcal{O}}_t = \{\boldsymbol{x} \in \mathcal{X}_t : \nexists\, \boldsymbol{x}' \in \mathcal{X}_t \text{ with } \boldsymbol{U}_t(\boldsymbol{x}') \succ \boldsymbol{U}_t(\boldsymbol{x})\}$, which is nonempty because $\mathcal{X}_t$ is finite. After observing $\boldsymbol{r}_t$ it updates the statistics of i_t and then, for every k and every neighbor ℓ of i_t in G_{t-1}^k, deletes the edge (i_t, ℓ) from G^k if and only if

$$\big\|\hat{\boldsymbol{w}}_{i_t,t}^k - \hat{\boldsymbol{w}}_{\ell,t}^k\big\|_2 > r_{i_t,t} + r_{\ell,t}, \qquad r_{i,t} := \frac{\rho(\boldsymbol{M}_{i,t})}{\sqrt{\lambda_{\min}(\boldsymbol{M}_{i,t})}} + \frac{\lambda}{\lambda_{\min}(\boldsymbol{M}_{i,t})}. \tag{24}$$

Edges are never added. We write $\Delta_t := \Delta_{i_t}(\boldsymbol{x}_t)$ and $s_t^k := \|\boldsymbol{x}_t\|_{(\bar{\boldsymbol{M}}_t^k)^{-1}}$.

Two elementary inequalities. (I1) For a positive definite matrix $\boldsymbol{M}$ and a vector $\boldsymbol{z}$, $\boldsymbol{z}^\top \boldsymbol{M}\boldsymbol{z} \geq \lambda_{\min}(\boldsymbol{M})\|\boldsymbol{z}\|_2^2$, hence

$$\|\boldsymbol{z}\|_2 \leq \frac{\|\boldsymbol{z}\|_{\boldsymbol{M}}}{\sqrt{\lambda_{\min}(\boldsymbol{M})}}. \tag{25}$$

(I2) If $\boldsymbol{M} = \lambda \boldsymbol{I} + \sum_{s=1}^{n} \boldsymbol{z}_s \boldsymbol{z}_s^\top \in \mathbb{R}^{p\times p}$ with $\|\boldsymbol{z}_s\|_2 \leq 1$, then the eigenvalues $\nu_1, \ldots, \nu_p > 0$ of $\lambda^{-1}\boldsymbol{M}$ satisfy $\sum_q \nu_q = \operatorname{tr}(\lambda^{-1}\boldsymbol{M}) = p + \lambda^{-1}\sum_s \|\boldsymbol{z}_s\|_2^2 \leq p + n/\lambda$, and the inequality of arithmetic and geometric means gives

$$\log\det(\lambda^{-1}\boldsymbol{M}) = \sum_{q=1}^{p} \log \nu_q \leq p \log\Big(1 + \frac{n}{p\lambda}\Big). \tag{26}$$

For $t \leq T$, every user matrix $\boldsymbol{M}_{i,t}$ and every cluster matrix $\bar{\boldsymbol{M}}_t^k$ is $\lambda\boldsymbol{I}$ plus at most T terms $\boldsymbol{x}_s\boldsymbol{x}_s^\top$ with $\|\boldsymbol{x}_s\|_2 \leq 1$, so (26) with $p = d$ and $n \leq T$ yields

$$\rho(\boldsymbol{M}_{i,t}) \leq \bar{\rho}_{\mathrm{lin}}, \quad \beta(\boldsymbol{M}_{i,t}) \leq \bar{\beta}_{\mathrm{lin}} \quad \text{and} \quad \beta(\bar{\boldsymbol{M}}_t^k) \leq \bar{\beta}_{\mathrm{lin}} \qquad \text{for all } i \in \mathcal{U},\ k \in [m_o],\ t \leq T, \tag{27}$$

where $\bar{\rho}_{\mathrm{lin}}$ and $\bar{\beta}_{\mathrm{lin}}$ are defined in (28) below.

Guarantee to be established. Let $\kappa_{x,K}$ be the spectral parameter defined in (34) below and let

$$\bar{\rho}_{\mathrm{lin}} := R\sqrt{d\log\Big(1 + \frac{T}{d\lambda}\Big) + 2\log\frac{8um_o}{\delta}}, \qquad \bar{\beta}_{\mathrm{lin}} := \sqrt{\lambda} + \bar{\rho}_{\mathrm{lin}}, \tag{28}$$

$$n_\gamma := \min\Big\{\Big\lceil \frac{72\,\bar{\beta}_{\mathrm{lin}}^2}{\kappa_{x,K}\gamma^2}\Big\rceil,\ \max\Big\{\Big\lceil \frac{288\,\bar{\rho}_{\mathrm{lin}}^2}{\kappa_{x,K}\gamma^2}\Big\rceil, \Big\lceil \frac{24\lambda}{\kappa_{x,K}\gamma}\Big\rceil\Big\}\Big\},$$

$$n_{\mathrm{lin}} := \max\Big\{\Big\lceil \frac{8}{\kappa_{x,K}}\log\frac{4udT}{\delta}\Big\rceil,\ n_\gamma\Big\}, \tag{29}$$

$$T_{\mathrm{lin}} := \Big\lceil 16u\log\frac{4u}{\delta} + 4u\,n_{\mathrm{lin}}\Big\rceil. \tag{30}$$

The guarantee of Section 4.1 asserts that, under (A1), (A3) and (A4) and the noise condition (A2), i.e., (20), with probability at least $1-\delta$: (i) no edge between two users of the same true cluster is ever deleted; (ii) if $T \geq T_{\mathrm{lin}}$, then at the end of round T_{lin} every graph G^k equals the true partition $\mathcal{C}^{(k)}$, in the sense that its edges are exactly the pairs of distinct users in a common cluster of $\mathcal{C}^{(k)}$; (iii) the regret satisfies

$$R_T \leq \min\Big\{2T,\ 2T_{\mathrm{lin}} + 2\bar{\beta}_{\mathrm{lin}}\sqrt{2m_c\,d\,T\log\Big(1 + \frac{T}{d\lambda}\Big)}\Big\}. \tag{31}$$

Probability budget. The proof uses three events: the confidence event $\mathcal{E}_{\mathrm{conf}}$ of Appendix C.3.1, the spectral event $\mathcal{E}_{\mathrm{spec}}$ and the arrival event $\mathcal{E}_{\mathrm{arr}}$ of Appendix C.3.3. Each fails with probability at most $\delta/4$, so their intersection has probability at least $1 - 3\delta/4 \geq 1 - \delta$. The analysis of COMONB-Frozen in Appendix D uses analogous events together with a fourth, covering event, which the linear analysis does not need.

C.1 Auxiliary Quantities

Equivalent form of the PSG. For every user i, round t and $\boldsymbol{x} \in \mathcal{X}_t$,

$$\Delta_i(\boldsymbol{x}) = \max_{\boldsymbol{x}' \in \mathcal{X}_t} \min_{k \in [m_o]} \big[\mu_i^k(\boldsymbol{x}') - \mu_i^k(\boldsymbol{x})\big]_+, \qquad [a]_+ := \max\{a, 0\}. \tag{32}$$

Consequently $0 \leq \Delta_i(\boldsymbol{x}) \leq 2$, because every difference $\mu_i^k(\boldsymbol{x}') - \mu_i^k(\boldsymbol{x})$ is at most $|\mu_i^k(\boldsymbol{x}')| + |\mu_i^k(\boldsymbol{x})| \leq 2$.

To verify (32), for $\boldsymbol{x}' \in \mathcal{X}_t$ let $g(\boldsymbol{x}') := \min_k[\mu_i^k(\boldsymbol{x}') - \mu_i^k(\boldsymbol{x})]$, and let $Z(\boldsymbol{x}')$ be the set of $\zeta \geq 0$ such that $\boldsymbol{\mu}_i(\boldsymbol{x}')$ does not Pareto-dominate $\boldsymbol{\mu}_i(\boldsymbol{x}) + \zeta\mathbf{1}$. If $\zeta > g(\boldsymbol{x}')$, then $\mu_i^k(\boldsymbol{x}') < \mu_i^k(\boldsymbol{x}) + \zeta$ for an index k attaining the minimum, so $\boldsymbol{\mu}_i(\boldsymbol{x}')$ does not dominate $\boldsymbol{\mu}_i(\boldsymbol{x}) + \zeta\mathbf{1}$ and $\zeta \in Z(\boldsymbol{x}')$. If $0 \leq \zeta < g(\boldsymbol{x}')$, then $\mu_i^k(\boldsymbol{x}') > \mu_i^k(\boldsymbol{x}) + \zeta$ for every k, so $\boldsymbol{\mu}_i(\boldsymbol{x}')$ dominates $\boldsymbol{\mu}_i(\boldsymbol{x}) + \zeta\mathbf{1}$ and

$\zeta \notin Z(\boldsymbol{x}')$. Hence $([g(\boldsymbol{x}')]_+, \infty) \subseteq Z(\boldsymbol{x}') \subseteq [[g(\boldsymbol{x}')]_+, \infty)$: if $g(\boldsymbol{x}') < 0$ the first case applies to every $\zeta \geq 0$, and if $g(\boldsymbol{x}') \geq 0$ this follows from the two cases. By (2), $\Delta_i(\boldsymbol{x}) = \inf \bigcap_{\boldsymbol{x}' \in \mathcal{X}_t} Z(\boldsymbol{x}')$. The finite intersection contains $(\max_{\boldsymbol{x}'}[g(\boldsymbol{x}')]_+, \infty)$ and is contained in $[\max_{\boldsymbol{x}'}[g(\boldsymbol{x}')]_+, \infty)$, so its infimum is $\max_{\boldsymbol{x}'}[g(\boldsymbol{x}')]_+$. Finally, $[\min_k a_k]_+ = \min_k [a_k]_+$ because $a \mapsto [a]_+$ is nondecreasing, which gives (32).

Spectral parameter. Let $\phi : \mathbb{R}^d \to \mathbb{R}^p$ be a measurable map with $\|\phi(\boldsymbol{x})\|_2 \leq 1$ on the support of $\mathcal{D}$, and let

$$\boldsymbol{\Sigma}_\phi := \mathbb{E}_{\boldsymbol{x} \sim \mathcal{D}}\big[\phi(\boldsymbol{x})\phi(\boldsymbol{x})^\top\big], \qquad \mathcal{S}_\phi := \mathrm{Range}(\boldsymbol{\Sigma}_\phi), \qquad \lambda_\phi^+ := \lambda_{\min}^+(\boldsymbol{\Sigma}_\phi),$$

the smallest positive eigenvalue of $\boldsymbol{\Sigma}_\phi$ (we assume $\boldsymbol{\Sigma}_\phi \neq \mathbf{0}$). For $\boldsymbol{v} \perp \mathcal{S}_\phi$ we have $\mathbb{E}\langle \boldsymbol{v}, \phi(\boldsymbol{x})\rangle^2 = \boldsymbol{v}^\top \boldsymbol{\Sigma}_\phi \boldsymbol{v} = 0$; applying this to a basis of $\mathcal{S}_\phi^\perp$ shows that $\phi(\boldsymbol{x}) \in \mathcal{S}_\phi$ almost surely. We define

$$\kappa_K(\phi) := \inf_{\boldsymbol{v} \in \mathcal{S}_\phi,\ \|\boldsymbol{v}\|_2 = 1} \mathbb{E}_{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K \overset{\text{i.i.d.}}{\sim} \mathcal{D}}\Big[\min_{1 \leq a \leq K} \langle \boldsymbol{v}, \phi(\boldsymbol{x}_a)\rangle^2\Big]. \tag{33}$$

It is the second moment, in the worst direction, of the least informative of K fresh candidates. For the maps ϕ^k of the fixed-feature model (A5) we write $\boldsymbol{\Sigma}_\phi^k := \boldsymbol{\Sigma}_{\phi^k}$ and $\mathcal{S}_k := \mathcal{S}_{\phi^k}$, so that $\kappa_{\phi,K} = \min_k \kappa_K(\phi^k)$ is the quantity in (7). The spectral parameter of the linear model is

$$\kappa_{x,K} := \inf_{\|\boldsymbol{v}\|_2 = 1} \mathbb{E}_{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K \overset{\text{i.i.d.}}{\sim} \mathcal{D}}\Big[\min_{1 \leq a \leq K} \langle \boldsymbol{v}, \boldsymbol{x}_a\rangle^2\Big], \tag{34}$$

where the infimum runs over all unit vectors of $\mathbb{R}^d$, and (A4) requires $\kappa_{x,K} > 0$. If $\kappa_{x,K} > 0$, then $\mathbb{E}\langle \boldsymbol{v}, \boldsymbol{x}\rangle^2 \geq \kappa_{x,K} > 0$ for every unit $\boldsymbol{v}$, so $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top]$ is positive definite; hence, for ϕ the identity, $\mathcal{S}_\phi = \mathbb{R}^d$ and $\kappa_K(\phi) = \kappa_{x,K}$. Since $\|\phi\|_2 \leq 1$, we have $\kappa_K(\phi) \leq 1$; in particular $\kappa_{x,K} \leq 1$.

C.1.1 Lower bound on the spectral parameter

For every unit vector $\boldsymbol{v} \in \mathcal{S}_\phi$,

$$\mathbb{E}\Big[\min_{1 \leq a \leq K} \langle \boldsymbol{v}, \phi(\boldsymbol{x}_a)\rangle^2\Big] \geq \frac{\lambda_\phi^+}{2}\Big(\frac{\lambda_\phi^+}{2 - \lambda_\phi^+}\Big)^K, \qquad \text{hence} \qquad \kappa_K(\phi) \geq \frac{\lambda_\phi^+}{2}\Big(\frac{\lambda_\phi^+}{2 - \lambda_\phi^+}\Big)^K > 0.$$

In particular: (a) if $\lambda_x := \lambda_{\min}(\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top]) > 0$, then $\kappa_{x,K} \geq (\lambda_x/2)\big(\lambda_x/(2 - \lambda_x)\big)^K > 0$; (b) if $\lambda_\phi := \min_k \lambda_{\min}^+(\boldsymbol{\Sigma}_\phi^k) > 0$, then $\kappa_{\phi,K} \geq (\lambda_\phi/2)\big(\lambda_\phi/(2 - \lambda_\phi)\big)^K > 0$.

To see this, fix a unit vector $\boldsymbol{v} \in \mathcal{S}_\phi$ and let $Z := \langle \boldsymbol{v}, \phi(\boldsymbol{x})\rangle^2$ with $\boldsymbol{x} \sim \mathcal{D}$. Then $Z \in [0, 1]$ by the Cauchy–Schwarz inequality, and $\mathbb{E}Z = \boldsymbol{v}^\top \boldsymbol{\Sigma}_\phi \boldsymbol{v} \geq \lambda_\phi^+$, because the restriction of $\boldsymbol{\Sigma}_\phi$ to its range has all eigenvalues at least λ_ϕ^+. In particular $\lambda_\phi^+ \leq 1$. Let $p := \mathbb{P}(Z \geq \lambda_\phi^+/2)$. Since $Z < \lambda_\phi^+/2$ on the complement of this event and $Z \leq 1$ everywhere,

$$\lambda_\phi^+ \leq \mathbb{E}Z \leq \frac{\lambda_\phi^+}{2}(1 - p) + p, \qquad \text{i.e.} \qquad p\Big(1 - \frac{\lambda_\phi^+}{2}\Big) \geq \frac{\lambda_\phi^+}{2}, \qquad \text{i.e.} \qquad p \geq \frac{\lambda_\phi^+}{2 - \lambda_\phi^+}.$$

Let $Z_a := \langle \boldsymbol{v}, \phi(\boldsymbol{x}_a)\rangle^2$. The variables $Z_1, \ldots, Z_K$ are independent copies of Z, so $\mathbb{P}(\min_a Z_a \geq \lambda_\phi^+/2) = p^K$, and since $\min_a Z_a \geq 0$,

$$\mathbb{E}\Big[\min_a Z_a\Big] \geq \frac{\lambda_\phi^+}{2}\,\mathbb{P}\Big(\min_a Z_a \geq \frac{\lambda_\phi^+}{2}\Big) \geq \frac{\lambda_\phi^+}{2}\Big(\frac{\lambda_\phi^+}{2 - \lambda_\phi^+}\Big)^K.$$

Taking the infimum over $\boldsymbol{v}$ proves the bound on $\kappa_K(\phi)$. For (a), take ϕ equal to the identity: then $\boldsymbol{\Sigma}_\phi = \mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top]$ is positive definite, $\mathcal{S}_\phi = \mathbb{R}^d$ and $\lambda_\phi^+ = \lambda_x$. For (b), the function $h(\ell) := (\ell/2)(\ell/(2 - \ell))^K$ is increasing on $(0, 1]$, and $\lambda_\phi \leq \lambda_{\min}^+(\boldsymbol{\Sigma}_\phi^k) \leq 1$ for every k; hence $\kappa_K(\phi^k) \geq h(\lambda_{\min}^+(\boldsymbol{\Sigma}_\phi^k)) \geq h(\lambda_\phi)$ for every k, and the bound follows by taking the minimum over k.

This lower bound decays exponentially in K. It is a sufficient condition for $\kappa_K(\phi) > 0$, not an estimate of its size; Appendix C.5 shows that $\kappa_{x,K}$ decays only as K^{-2} for spherical contexts.

C.2 Matrix Lower Tail and the Selected Context

For a subspace $S \subseteq \mathbb{R}^p$ with orthogonal projector $\boldsymbol{P}_S$ and a symmetric matrix $\boldsymbol{A}$, we write $\lambda_{\min}(\boldsymbol{A}|_S) := \min\{\boldsymbol{v}^\top \boldsymbol{A}\boldsymbol{v} : \boldsymbol{v} \in S,\ \|\boldsymbol{v}\|_2 = 1\}$; for $S = \mathbb{R}^p$ this is $\lambda_{\min}(\boldsymbol{A})$. We use two standard results from matrix analysis. First, the Golden–Thompson inequality: $\operatorname{tr} e^{\boldsymbol{A}+\boldsymbol{B}} \le \operatorname{tr}\big(e^{\boldsymbol{A}} e^{\boldsymbol{B}}\big)$ for all symmetric $\boldsymbol{A}, \boldsymbol{B}$. Second, if $\boldsymbol{A} \succeq \boldsymbol{0}$ and $\boldsymbol{B}$ is symmetric with $\boldsymbol{B} \preceq c\boldsymbol{I}$, then $\operatorname{tr}(\boldsymbol{A}\boldsymbol{B}) \le c \operatorname{tr}(\boldsymbol{A})$, because $\operatorname{tr}(\boldsymbol{A}(c\boldsymbol{I} - \boldsymbol{B})) = \operatorname{tr}(\boldsymbol{A}^{1/2}(c\boldsymbol{I} - \boldsymbol{B})\boldsymbol{A}^{1/2}) \ge 0$. Conditional expectation is linear and preserves the Loewner order.

C.2.1 Matrix lower tail for adapted sequences

Let $S \subseteq \mathbb{R}^p$ be a subspace of dimension $r_S \ge 1$, let $(\mathcal{F}_n)_{n\ge 0}$ be a filtration, and let $(\boldsymbol{Y}_n)_{n\ge 1}$ be symmetric random $p \times p$ matrices such that, for every $n \ge 1$, $\boldsymbol{Y}_n$ is $\mathcal{F}_n$-measurable, $\boldsymbol{Y}_n = \boldsymbol{P}_S \boldsymbol{Y}_n \boldsymbol{P}_S$ and $\boldsymbol{0} \preceq \boldsymbol{Y}_n \preceq \boldsymbol{P}_S$ almost surely, and

$$\mathbb{E}\big[\boldsymbol{Y}_n \,\big|\, \mathcal{F}_{n-1}\big] \succeq \kappa\, \boldsymbol{P}_S \quad \text{almost surely}$$

for a constant $\kappa > 0$. Then, for every $n \ge 1$,

$$\mathbb{P}\Big(\lambda_{\min}\Big(\sum_{m=1}^{n} \boldsymbol{Y}_m\Big|_S\Big) \le \frac{n\kappa}{2}\Big) \le r_S\, e^{-n\kappa/8}.$$

All inequalities between random matrices in the following argument hold almost surely; every step below is either pointwise or an inequality between conditional expectations, so the almost-sure hypotheses suffice.

Reduction to $\mathbb{R}^{r_S}$. Let $\boldsymbol{Q} \in \mathbb{R}^{p\times r_S}$ have orthonormal columns spanning S, so that $\boldsymbol{Q}^\top \boldsymbol{Q} = \boldsymbol{I}_{r_S}$ and $\boldsymbol{Q}\boldsymbol{Q}^\top = \boldsymbol{P}_S$, and let $\tilde{\boldsymbol{Y}}_n := \boldsymbol{Q}^\top \boldsymbol{Y}_n \boldsymbol{Q} \in \mathbb{R}^{r_S \times r_S}$. For $\boldsymbol{w} \in \mathbb{R}^{r_S}$ we have $\boldsymbol{w}^\top \tilde{\boldsymbol{Y}}_n \boldsymbol{w} = (\boldsymbol{Q}\boldsymbol{w})^\top \boldsymbol{Y}_n (\boldsymbol{Q}\boldsymbol{w}) \in [0, \|\boldsymbol{w}\|_2^2]$, because $\boldsymbol{0} \preceq \boldsymbol{Y}_n \preceq \boldsymbol{P}_S$ and $(\boldsymbol{Q}\boldsymbol{w})^\top \boldsymbol{P}_S (\boldsymbol{Q}\boldsymbol{w}) = \|\boldsymbol{Q}\boldsymbol{w}\|_2^2 = \|\boldsymbol{w}\|_2^2$; hence $\boldsymbol{0} \preceq \tilde{\boldsymbol{Y}}_n \preceq \boldsymbol{I}_{r_S}$. Moreover $\mathbb{E}[\tilde{\boldsymbol{Y}}_n | \mathcal{F}_{n-1}] = \boldsymbol{Q}^\top \mathbb{E}[\boldsymbol{Y}_n | \mathcal{F}_{n-1}] \boldsymbol{Q} \succeq \kappa \boldsymbol{Q}^\top \boldsymbol{P}_S \boldsymbol{Q} = \kappa \boldsymbol{I}_{r_S}$. Every unit vector of S is of the form $\boldsymbol{Q}\boldsymbol{w}$ with $\boldsymbol{w} \in \mathbb{R}^{r_S}$ a unit vector, so $\lambda_{\min}(\sum_{m\le n} \boldsymbol{Y}_m|_S) = \lambda_{\min}(\boldsymbol{Z}_n)$ with $\boldsymbol{Z}_n := \sum_{m\le n} \tilde{\boldsymbol{Y}}_m$ and $\boldsymbol{Z}_0 := \boldsymbol{0}$. Fix $\theta > 0$.

Chord bound. The function $y \mapsto e^{-\theta y}$ is convex, so on $[0,1]$ it lies below its chord: $e^{-\theta y} \le 1 - (1 - e^{-\theta})y$ for $y \in [0,1]$. Applying this inequality to the eigenvalues of $\tilde{\boldsymbol{Y}}_n$, which lie in $[0,1]$, in its spectral decomposition gives $e^{-\theta \tilde{\boldsymbol{Y}}_n} \preceq \boldsymbol{I}_{r_S} - (1 - e^{-\theta})\tilde{\boldsymbol{Y}}_n$.

Conditional moment generating function. Taking conditional expectations and using $\mathbb{E}[\tilde{\boldsymbol{Y}}_n | \mathcal{F}_{n-1}] \succeq \kappa \boldsymbol{I}_{r_S}$ together with $1 - a \le e^{-a}$,

$$\mathbb{E}\big[e^{-\theta \tilde{\boldsymbol{Y}}_n} \,\big|\, \mathcal{F}_{n-1}\big] \preceq \boldsymbol{I}_{r_S} - (1-e^{-\theta})\, \mathbb{E}\big[\tilde{\boldsymbol{Y}}_n \,\big|\, \mathcal{F}_{n-1}\big] \preceq \big(1 - (1-e^{-\theta})\kappa\big)\boldsymbol{I}_{r_S} \preceq e^{-(1-e^{-\theta})\kappa}\, \boldsymbol{I}_{r_S}. \tag{35}$$

Iteration. By the Golden–Thompson inequality, $\operatorname{tr} e^{-\theta \boldsymbol{Z}_n} = \operatorname{tr} e^{-\theta \boldsymbol{Z}_{n-1} - \theta \tilde{\boldsymbol{Y}}_n} \le \operatorname{tr}\big(e^{-\theta \boldsymbol{Z}_{n-1}} e^{-\theta \tilde{\boldsymbol{Y}}_n}\big)$. The matrix $e^{-\theta \boldsymbol{Z}_{n-1}}$ is positive definite and $\mathcal{F}_{n-1}$-measurable, all matrices involved are bounded, and the trace is linear; hence, by (35) and the trace inequality stated above,

$$\mathbb{E}\big[\operatorname{tr} e^{-\theta \boldsymbol{Z}_n} \,\big|\, \mathcal{F}_{n-1}\big] \le \operatorname{tr}\Big(e^{-\theta \boldsymbol{Z}_{n-1}}\, \mathbb{E}\big[e^{-\theta \tilde{\boldsymbol{Y}}_n} \,\big|\, \mathcal{F}_{n-1}\big]\Big) \le e^{-(1-e^{-\theta})\kappa} \operatorname{tr} e^{-\theta \boldsymbol{Z}_{n-1}}.$$

Taking expectations and iterating down to $\operatorname{tr} e^{-\theta \boldsymbol{Z}_0} = \operatorname{tr} \boldsymbol{I}_{r_S} = r_S$ yields $\mathbb{E} \operatorname{tr} e^{-\theta \boldsymbol{Z}_n} \le r_S\, e^{-n\kappa(1-e^{-\theta})}$.

Markov's inequality. All eigenvalues of $e^{-\theta \boldsymbol{Z}_n}$ are positive, so $\operatorname{tr} e^{-\theta \boldsymbol{Z}_n} \ge \lambda_{\max}(e^{-\theta \boldsymbol{Z}_n}) = e^{-\theta \lambda_{\min}(\boldsymbol{Z}_n)}$. Therefore $\lambda_{\min}(\boldsymbol{Z}_n) \le n\kappa/2$ implies $\operatorname{tr} e^{-\theta \boldsymbol{Z}_n} \ge e^{-\theta n\kappa/2}$, and Markov's inequality gives

$$\mathbb{P}\Big(\lambda_{\min}(\boldsymbol{Z}_n) \le \frac{n\kappa}{2}\Big) \le e^{\theta n\kappa/2}\, \mathbb{E} \operatorname{tr} e^{-\theta \boldsymbol{Z}_n} \le r_S \exp\Big(n\kappa\Big(\frac{\theta}{2} - (1 - e^{-\theta})\Big)\Big).$$

With $\theta = \log 2$ the exponent equals $n\kappa\,(\frac{1}{2}\log 2 - \frac{1}{2}) = -n\kappa\,\frac{1-\log 2}{2}$, and $\frac{1-\log 2}{2} \ge \frac{1}{8}$ because $\log 2 \le \frac{3}{4}$ (numerically, $\frac{1}{2}\log 2 - \frac{1}{2} \approx 0.3466 - 0.5 = -0.1534$). Hence the probability is at most $r_S\, e^{-n\kappa/8}$.

C.2.2 THE SELECTED CONTEXT DOMINATES THE WORST CANDIDATE

Let ϕ be as in Appendix C.1, and suppose that in round t the context $\boldsymbol{x}_t \in \mathcal{X}_t$ is chosen by an arbitrary rule, possibly randomized and possibly depending on $\mathcal{H}_{t-1}$, i_t and $\mathcal{X}_t$. Then, for every unit vector $\boldsymbol{v} \in \mathcal{S}_\phi$, almost surely

$$\mathbb{E}\big[\langle \boldsymbol{v}, \phi(\boldsymbol{x}_t)\rangle^2 \,\big|\, \mathcal{H}_t^-\big] \geq \mathbb{E}_{\boldsymbol{x}_1,\dots,\boldsymbol{x}_K \overset{\text{i.i.d.}}{\sim} \mathcal{D}}\Big[\min_{1\leq a\leq K} \langle \boldsymbol{v}, \phi(\boldsymbol{x}_a)\rangle^2\Big] \geq \kappa_K(\phi),$$

and $\mathbb{E}[\phi(\boldsymbol{x}_t)\phi(\boldsymbol{x}_t)^\top \mid \mathcal{H}_t^-] \succeq \kappa_K(\phi)\, \boldsymbol{P}_{\mathcal{S}_\phi}$ almost surely. In particular, in the linear model $\mathbb{E}[\boldsymbol{x}_t\boldsymbol{x}_t^\top \mid \mathcal{H}_t^-] \succeq \kappa_{x,K}\boldsymbol{I}$ almost surely.

To see this, fix a unit vector $\boldsymbol{v} \in \mathcal{S}_\phi$. Since $\boldsymbol{x}_t$ is one of the candidates, pointwise

$$\langle \boldsymbol{v}, \phi(\boldsymbol{x}_t)\rangle^2 \geq m_{\boldsymbol{v}}(\mathcal{X}_t) := \min_{1\leq a\leq K} \langle \boldsymbol{v}, \phi(\boldsymbol{x}_{t,a})\rangle^2,$$

whatever the selection rule. The right-hand side is a function of $\mathcal{X}_t$ alone, and, by (A1) as read in the standing conventions at the beginning of this appendix, $\mathcal{X}_t$ consists of K i.i.d. draws from $\mathcal{D}$ and is independent of $\mathcal{H}_t^-$; hence $\mathbb{E}[m_{\boldsymbol{v}}(\mathcal{X}_t)|\mathcal{H}_t^-] = \mathbb{E}[m_{\boldsymbol{v}}(\mathcal{X}_t)]$ almost surely, which is the middle term of the first display of this subsubsection and is at least $\kappa_K(\phi)$ by (33). Monotonicity of conditional expectation proves the first inequality. For the matrix inequality, let $\boldsymbol{\Xi}_t$ be a version of $\mathbb{E}[\phi(\boldsymbol{x}_t)\phi(\boldsymbol{x}_t)^\top|\mathcal{H}_t^-]$. Each $\phi(\boldsymbol{x}_{t,a})$ lies in $\mathcal{S}_\phi$ almost surely, hence so does $\phi(\boldsymbol{x}_t)$. For a fixed $\boldsymbol{w} \in \mathbb{R}^p$ write $\boldsymbol{v} := \boldsymbol{P}_{\mathcal{S}_\phi}\boldsymbol{w}$; then $\langle \boldsymbol{w}, \phi(\boldsymbol{x}_t)\rangle = \langle \boldsymbol{v}, \phi(\boldsymbol{x}_t)\rangle$ almost surely, and the first inequality applied to $\boldsymbol{v}/\|\boldsymbol{v}\|_2$ (the case $\boldsymbol{v} = \boldsymbol{0}$ is trivial) gives $\boldsymbol{w}^\top\boldsymbol{\Xi}_t\boldsymbol{w} = \mathbb{E}[\langle \boldsymbol{v}, \phi(\boldsymbol{x}_t)\rangle^2|\mathcal{H}_t^-] \geq \kappa_K(\phi)\|\boldsymbol{v}\|_2^2 = \kappa_K(\phi)\, \boldsymbol{w}^\top\boldsymbol{P}_{\mathcal{S}_\phi}\boldsymbol{w}$ almost surely. This holds simultaneously for all $\boldsymbol{w}$ in a countable dense subset of $\mathbb{R}^p$ outside a null set, and both sides are continuous in $\boldsymbol{w}$; hence $\boldsymbol{\Xi}_t \succeq \kappa_K(\phi)\boldsymbol{P}_{\mathcal{S}_\phi}$ almost surely. The linear case is ϕ equal to the identity, for which $\mathcal{S}_\phi = \mathbb{R}^d$ under (A4).

C.3 CONFIDENCE SETS, EDGE DELETION AND CLUSTER RECOVERY

Fixed groups. For each objective k let

$$\mathcal{A}^{(k)} := \big\{\{i\} : \, i \in \mathcal{U}\big\} \cup \mathcal{C}^{(k)}$$

be the family consisting of all single users and all true clusters of objective k. Every $A \in \mathcal{A}^{(k)}$ is contained in a single true cluster $C_j^{(k)}$, and all users in A share the mean reward function of objective k. These families are determined by the true partitions before the interaction starts; they do not depend on the graphs built by the algorithm. Their total size satisfies $\sum_k |\mathcal{A}^{(k)}| \leq um_o + m_c \leq 2um_o$, because $m_c^{(k)} \leq u$ for every k; the first inequality can be strict, because a true cluster that consists of a single user coincides with the corresponding single-user group.

Since the argument is the same for every fixed feature map, we state the confidence bounds for general fixed features: for each k, a map $\phi^k : \mathcal{X} \to \mathbb{R}^{p_k}$ with $\|\phi^k(\boldsymbol{x})\|_2 \leq 1$ is fixed before the interaction, and $\mu_i^k(\boldsymbol{x}) = \langle \phi^k(\boldsymbol{x}), \boldsymbol{w}_{j_k(i)}^k\rangle$ with $\|\boldsymbol{w}_j^k\|_2 \leq B$, as in (A5). The linear model is the special case $\phi^k =$ identity, $p_k = d$, $\boldsymbol{w}_j^k = \boldsymbol{\theta}_j^k$ and $B = 1$; Appendix D uses the same argument for COMONB-Frozen. For $A \in \mathcal{A}^{(k)}$ contained in $C_j^{(k)}$ we write $\boldsymbol{w}_A^k := \boldsymbol{w}_j^k$ and $\mu_A^k := \mu_i^k$ for any $i \in A$, and define the group statistics

$$\boldsymbol{M}_{A,t}^k := \lambda\boldsymbol{I} + \sum_{s\leq t:\, i_s\in A} \phi^k(\boldsymbol{x}_s)\phi^k(\boldsymbol{x}_s)^\top, \qquad \boldsymbol{b}_{A,t}^k := \sum_{s\leq t:\, i_s\in A} r_s^k\, \phi^k(\boldsymbol{x}_s),$$
$$\hat{\boldsymbol{w}}_{A,t}^k := (\boldsymbol{M}_{A,t}^k)^{-1}\boldsymbol{b}_{A,t}^k, \tag{36}$$

together with the noise sum $\boldsymbol{S}_{A,t}^k := \sum_{s\leq t:\, i_s\in A} \eta_s^k\, \phi^k(\boldsymbol{x}_s)$ and the radii $\rho(\boldsymbol{M}) := R\sqrt{\log\det(\lambda^{-1}\boldsymbol{M}) + 2\log(8um_o/\delta)}$ and $\beta(\boldsymbol{M}) := B\sqrt{\lambda} + \rho(\boldsymbol{M})$, which reduce to (22) for $B = 1$.

C.3.1 ANYTIME CONFIDENCE FOR USERS AND TRUE CLUSTERS

Let $\mathcal{E}_{\text{conf}}$ be the event that

$$\big\|\boldsymbol{S}_{A,t}^k\big\|_{(\boldsymbol{M}_{A,t}^k)^{-1}} \leq \rho(\boldsymbol{M}_{A,t}^k) \qquad \text{for all } k \in [m_o],\ A \in \mathcal{A}^{(k)} \text{ and } 0 \leq t \leq T.$$

Then $\mathbb{P}(\mathcal{E}_{\mathrm{conf}}) \geq 1 - \delta/4$ for every selection rule that chooses $\boldsymbol{x}_t \in \mathcal{X}_t$ without using the rewards of round t or later, so that (20) holds. On $\mathcal{E}_{\mathrm{conf}}$, for all k, $A \in \mathcal{A}^{(k)}$, $0 \leq t \leq T$ and simultaneously for all contexts $\boldsymbol{x}$, with $\boldsymbol{M} = \boldsymbol{M}_{A,t}^k$,

$$\left\|\hat{\boldsymbol{w}}_{A,t}^k - \boldsymbol{w}_A^k\right\|_{\boldsymbol{M}} \leq \beta(\boldsymbol{M}), \qquad \left|\phi^k(\boldsymbol{x})^\top \hat{\boldsymbol{w}}_{A,t}^k - \mu_A^k(\boldsymbol{x})\right| \leq \beta(\boldsymbol{M}) \left\|\phi^k(\boldsymbol{x})\right\|_{\boldsymbol{M}^{-1}}, \tag{37}$$

$$\begin{aligned} \left\|\hat{\boldsymbol{w}}_{A,t}^k - \boldsymbol{w}_A^k\right\|_2 &\leq \frac{\rho(\boldsymbol{M})}{\sqrt{\lambda_{\min}(\boldsymbol{M})}} + \frac{\lambda B}{\lambda_{\min}(\boldsymbol{M})}, \\ \left|\phi^k(\boldsymbol{x})^\top \hat{\boldsymbol{w}}_{A,t}^k - \mu_A^k(\boldsymbol{x})\right| &\leq \rho(\boldsymbol{M}) \left\|\phi^k(\boldsymbol{x})\right\|_{\boldsymbol{M}^{-1}} + \lambda B \left\|\boldsymbol{M}^{-1}\phi^k(\boldsymbol{x})\right\|_2. \end{aligned} \tag{38}$$

To show this, fix k and $A \in \mathcal{A}^{(k)}$, and abbreviate $\boldsymbol{M} = \boldsymbol{M}_{A,t}^k$, $\boldsymbol{w} = \boldsymbol{w}_A^k$ and $\boldsymbol{\phi}_s = \phi^k(\boldsymbol{x}_s)$. For every round s with $i_s \in A$ we have $r_s^k = \langle \boldsymbol{\phi}_s, \boldsymbol{w} \rangle + \eta_s^k$, because all users of A share $\boldsymbol{w}$ for objective k. Hence $\boldsymbol{b}_{A,t}^k = (\boldsymbol{M} - \lambda \boldsymbol{I})\boldsymbol{w} + \boldsymbol{S}_t$ with $\boldsymbol{S}_t := \boldsymbol{S}_{A,t}^k$, and

$$\hat{\boldsymbol{w}}_{A,t}^k - \boldsymbol{w} = \boldsymbol{M}^{-1}\big((\boldsymbol{M} - \lambda \boldsymbol{I})\boldsymbol{w} + \boldsymbol{S}_t\big) - \boldsymbol{w} = -\lambda \boldsymbol{M}^{-1}\boldsymbol{w} + \boldsymbol{M}^{-1}\boldsymbol{S}_t.$$

We first show that the four bounds follow from $\|\boldsymbol{S}_t\|_{\boldsymbol{M}^{-1}} \leq \rho(\boldsymbol{M})$. Using $\|\boldsymbol{M}^{-1}\boldsymbol{z}\|_{\boldsymbol{M}} = \|\boldsymbol{z}\|_{\boldsymbol{M}^{-1}}$ and $\boldsymbol{M}^{-1} \preceq \lambda^{-1}\boldsymbol{I}$,

$$\left\|\hat{\boldsymbol{w}}_{A,t}^k - \boldsymbol{w}\right\|_{\boldsymbol{M}} \leq \lambda \|\boldsymbol{w}\|_{\boldsymbol{M}^{-1}} + \|\boldsymbol{S}_t\|_{\boldsymbol{M}^{-1}} \leq \sqrt{\lambda}\,\|\boldsymbol{w}\|_2 + \|\boldsymbol{S}_t\|_{\boldsymbol{M}^{-1}} \leq B\sqrt{\lambda} + \rho(\boldsymbol{M}) = \beta(\boldsymbol{M}),$$

and the Cauchy–Schwarz inequality $|\langle \boldsymbol{z}, \boldsymbol{y} \rangle| \leq \|\boldsymbol{z}\|_{\boldsymbol{M}} \|\boldsymbol{y}\|_{\boldsymbol{M}^{-1}}$ with $\boldsymbol{z} = \hat{\boldsymbol{w}}_{A,t}^k - \boldsymbol{w}$ and $\boldsymbol{y} = \phi^k(\boldsymbol{x})$ gives the second bound in (37). For the Euclidean bound, $\|\boldsymbol{M}^{-1}\boldsymbol{S}_t\|_2 = \|\boldsymbol{M}^{-1/2}(\boldsymbol{M}^{-1/2}\boldsymbol{S}_t)\|_2 \leq \|\boldsymbol{S}_t\|_{\boldsymbol{M}^{-1}} / \sqrt{\lambda_{\min}(\boldsymbol{M})}$ and $\|\lambda \boldsymbol{M}^{-1}\boldsymbol{w}\|_2 \leq \lambda \|\boldsymbol{w}\|_2 / \lambda_{\min}(\boldsymbol{M}) \leq \lambda B / \lambda_{\min}(\boldsymbol{M})$, so the triangle inequality gives the first bound in (38). For the second, $|\phi^k(\boldsymbol{x})^\top \boldsymbol{M}^{-1}\boldsymbol{S}_t| \leq \|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}} \|\boldsymbol{S}_t\|_{\boldsymbol{M}^{-1}}$ by the Cauchy–Schwarz inequality for $\langle \cdot, \cdot \rangle_{\boldsymbol{M}^{-1}}$, and $|\lambda \phi^k(\boldsymbol{x})^\top \boldsymbol{M}^{-1}\boldsymbol{w}| \leq \lambda \|\boldsymbol{M}^{-1}\phi^k(\boldsymbol{x})\|_2 \|\boldsymbol{w}\|_2$. These bounds hold for every $\boldsymbol{x}$ at once, so no union bound over candidates is needed.

To control the noise term we apply the self-normalized bound of Abbasi-Yadkori et al. (2011) with $\boldsymbol{V} = \lambda \boldsymbol{I}$, the filtration

$$\mathcal{F}_s := \mathcal{H}_s^+ \quad \text{for } 0 \leq s < T, \qquad \mathcal{F}_s := \mathcal{H}_T \quad \text{for } s \geq T,$$

the vectors $\boldsymbol{q}_s := \mathbb{I}\{i_s \in A\}\boldsymbol{\phi}_s$ and the scalar noise η_s^k for $1 \leq s \leq T$, padded with $\boldsymbol{q}_s := \boldsymbol{0}$ and $\eta_s^k := 0$ for $s > T$; in this way only rounds $1, \ldots, T$ enter. Its hypotheses hold. First, $(\mathcal{F}_s)_{s \geq 0}$ is a filtration, because $\mathcal{H}_s^+ \subseteq \mathcal{H}_{s+1} \subseteq \mathcal{H}_{s+1}^+$ for $0 \leq s < T - 1$ and $\mathcal{H}_{T-1}^+ \subseteq \mathcal{H}_T$. Second, for $1 \leq s \leq T$, the vector $\boldsymbol{q}_s$ is $\mathcal{H}_{s-1}^+$-measurable, i.e., $\mathcal{F}_{s-1}$-measurable; $\eta_s^k = r_s^k - \mu_{i_s}^k(\boldsymbol{x}_s)$ is $\mathcal{H}_s$-measurable and hence $\mathcal{F}_s$-measurable, because $\mathcal{H}_s \subseteq \mathcal{F}_s$; and η_s^k is conditionally R-sub-Gaussian given $\mathcal{F}_{s-1} = \mathcal{H}_{s-1}^+$ by (20). Third, for $s > T$ the constants $\boldsymbol{q}_s = \boldsymbol{0}$ and $\eta_s^k = 0$ satisfy the same conditions trivially. Since $\lambda \boldsymbol{I} + \sum_{s \leq t} \boldsymbol{q}_s \boldsymbol{q}_s^\top = \boldsymbol{M}_{A,t}^k$ and $\sum_{s \leq t} \eta_s^k \boldsymbol{q}_s = \boldsymbol{S}_t$ for $0 \leq t \leq T$, this bound gives, with probability at least $1 - \delta_A$ and simultaneously for all $0 \leq t \leq T$,

$$\|\boldsymbol{S}_t\|^2_{(\boldsymbol{M}_{A,t}^k)^{-1}} \leq 2R^2 \log \frac{\det(\boldsymbol{M}_{A,t}^k)^{1/2} \det(\lambda \boldsymbol{I})^{-1/2}}{\delta_A} = R^2 \Big(\log \det\big(\lambda^{-1}\boldsymbol{M}_{A,t}^k\big) + 2 \log \frac{1}{\delta_A} \Big).$$

With $\delta_A := \delta / (8um_o)$ the right-hand side is exactly $\rho(\boldsymbol{M}_{A,t}^k)^2$, so the bound defining $\mathcal{E}_{\mathrm{conf}}$ holds for this pair (k, A) with probability at least $1 - \delta/(8um_o)$. Only the noise of objective k enters, so the dependence between objectives is irrelevant. A union bound over the at most $um_o + m_c \leq 2um_o$ pairs (k, A) gives $\mathbb{P}(\mathcal{E}_{\mathrm{conf}}^c) \leq 2um_o \cdot \delta/(8um_o) = \delta/4$. The selection rule enters only through the measurability of $\boldsymbol{q}_s$ and through (20), so the bound holds for every rule that chooses $\boldsymbol{x}_t \in \mathcal{X}_t$ without using the rewards of round t or later.

Linear specialization and the order of events. In the linear model the group matrix does not depend on k, and we write $\boldsymbol{M}_{A,t} := \lambda \boldsymbol{I} + \sum_{s \leq t:\, i_s \in A} \boldsymbol{x}_s \boldsymbol{x}_s^\top$. For $A = \{i\}$ these are the user statistics of (21): $\boldsymbol{M}_{\{i\},t} = \boldsymbol{M}_{i,t}$ and $\hat{\boldsymbol{w}}_{\{i\},t}^k = \hat{\boldsymbol{w}}_{i,t}^k$. For a true cluster $C = C_j^{(k)}$, regrouping the sums in (23) by rounds shows that

$$\widehat{C}_t^k = C \quad \Longrightarrow \quad \bar{\boldsymbol{M}}_t^k = \boldsymbol{M}_{C,t-1}, \qquad \bar{\boldsymbol{b}}_t^k = \boldsymbol{b}_{C,t-1}^k, \qquad \bar{\boldsymbol{w}}_t^k = \hat{\boldsymbol{w}}_{C,t-1}^k. \tag{39}$$

Indeed, $\sum_{i\in C}(\boldsymbol{M}_{i,t-1}-\lambda\boldsymbol{I})=\sum_{s\le t-1:\, i_s\in C}\boldsymbol{x}_s\boldsymbol{x}_s^\top$ and $\sum_{i\in C}\boldsymbol{b}^k_{i,t-1}=\sum_{s\le t-1:\, i_s\in C} r^k_s\boldsymbol{x}_s$. The logical order of the argument matters here. The confidence bounds above concern the fixed groups in $\mathcal{A}^{(k)}$. The recovery of the partitions (Appendix C.3.3) is a separate event. Only on the intersection of the two events do we use (39) to transfer the confidence bound of the true cluster to the statistics computed by the algorithm. At no point is a concentration inequality applied to the data-dependent set $\widehat{C}^k_t$, and at no point do we condition on correct clustering.

C.3.2 No false deletion

On $\mathcal{E}_{\mathrm{conf}}$, for every round $1\le t\le T$, every objective k and all users i,ℓ with $j_k(i)=j_k(\ell)$,

$$\big\|\hat{\boldsymbol{w}}^k_{i,t}-\hat{\boldsymbol{w}}^k_{\ell,t}\big\|_2\le r_{i,t}+r_{\ell,t}.$$

Consequently, on $\mathcal{E}_{\mathrm{conf}}$ COMOLB never deletes an edge between two users of the same true cluster: for all $0\le t\le T$ and all k, G^k_t contains every such edge, and therefore $\widehat{C}^k_t\supseteq C^{(k)}_{j_k(i_t)}$ for all $1\le t\le T$.

To see this, let $j:=j_k(i)=j_k(\ell)$. By the Euclidean bound (38) of Appendix C.3.1 applied to the group $\{i\}$, with $B=1$,

$$\big\|\hat{\boldsymbol{w}}^k_{i,t}-\boldsymbol{\theta}^k_j\big\|_2\le\frac{\rho(\boldsymbol{M}_{i,t})}{\sqrt{\lambda_{\min}(\boldsymbol{M}_{i,t})}}+\frac{\lambda}{\lambda_{\min}(\boldsymbol{M}_{i,t})}=r_{i,t},$$

and likewise $\|\hat{\boldsymbol{w}}^k_{\ell,t}-\boldsymbol{\theta}^k_j\|_2\le r_{\ell,t}$. The triangle inequality gives the displayed bound. The bound holds in every round, including round 1 and for users without observations (then $\boldsymbol{M}_{i,t}=\lambda\boldsymbol{I}$ and $\hat{\boldsymbol{w}}^k_{i,t}=\boldsymbol{0}$). The deletion rule (24) requires a strict inequality in the opposite direction, so the test never fires for a same-cluster edge. Since G^k_0 is complete and edges are only removed, induction over t shows that every same-cluster edge belongs to G^k_t for all $t\le T$. In particular, every user of the true cluster $C^{(k)}_{j_k(i_t)}$ other than i_t is a neighbour of i_t in G^k_{t-1}, so the cluster is contained in the closed neighbourhood $\widehat{C}^k_t$.

C.3.3 Cluster recovery for COMOLB

For $i\in\mathcal{U}$ and $n\ge 1$ let $\tau_{i,n}$ denote the round of the n-th arrival of user i, and set $\tau_{i,0}:=0$. To have $\tau_{i,n}$ defined for every n, we let the protocol run beyond round T. For $t>T$, users and candidate sets are drawn as before, so that the independence properties stated in the standing conventions continue to hold, $\boldsymbol{x}_t\in\mathcal{X}_t$ is chosen by an arbitrary rule (for instance $\boldsymbol{x}_t=\boldsymbol{x}_{t,1}$), and rewards are ignored; the σ-fields $\mathcal{H}_t$ and $\mathcal{H}^-_t$ are defined as before, without rewards. This extension does not change rounds $1,\dots,T$. Since the arrivals are independent and uniform, $\tau_{i,n}<\infty$ almost surely, and $\{\tau_{i,n}=t\}$ is determined by $i_1,\dots,i_t$, so that $\{\tau_{i,n}=t\}\in\mathcal{H}^-_t$.

Suppose that $\kappa_{x,K}>0$, i.e., that (A4) holds, and let

$$\mathcal{E}_{\mathrm{spec}}:=\Big\{\lambda_{\min}(\boldsymbol{M}_{i,t})\ge\lambda+\frac{\kappa_{x,K}\,T_{i,t}}{2}\ \text{ for all } i\in\mathcal{U}\text{ and all } t\le T\text{ with } T_{i,t}\ge n_{\mathrm{lin}}\Big\}.$$

(a) $\mathbb{P}(\mathcal{E}_{\mathrm{spec}})\ge 1-\delta/4$. On $\mathcal{E}_{\mathrm{spec}}$, for all $i\in\mathcal{U}$ and $t\le T$ with $T_{i,t}\ge n_{\mathrm{lin}}$,

$$r_{i,t}\le\min\left\{\bar\beta_{\mathrm{lin}}\sqrt{\frac{2}{\kappa_{x,K}\,T_{i,t}}},\ \bar\rho_{\mathrm{lin}}\sqrt{\frac{2}{\kappa_{x,K}\,T_{i,t}}}+\frac{2\lambda}{\kappa_{x,K}\,T_{i,t}}\right\}\le\frac{\gamma}{6}.$$

(b) On $\mathcal{E}_{\mathrm{conf}}\cap\mathcal{E}_{\mathrm{spec}}$, let $k\in[m_o]$, $1\le t\le T$ and users i,ℓ with $j_k(i)\ne j_k(\ell)$. If $(i,\ell)\in E^k_{t-1}$, $i_t\in\{i,\ell\}$ and $T_{i,t},T_{\ell,t}\ge n_{\mathrm{lin}}$, then COMOLB deletes (i,ℓ) from G^k in round t. Consequently, $(i,\ell)\notin E^k_t$ for every t with $\max\{\tau_{i,n_{\mathrm{lin}}},\tau_{\ell,n_{\mathrm{lin}}}\}\le t\le T$.

(c) Suppose $T\ge T_{\mathrm{lin}}$ and let $\mathcal{E}_{\mathrm{arr}}:=\{T_{i,T_{\mathrm{lin}}}\ge n_{\mathrm{lin}}\text{ for all } i\in\mathcal{U}\}$. Then $\mathbb{P}(\mathcal{E}_{\mathrm{arr}})\ge 1-\delta/4$. On $\mathcal{E}_{\mathrm{conf}}\cap\mathcal{E}_{\mathrm{spec}}\cap\mathcal{E}_{\mathrm{arr}}$, for every k the edges of $G^k_{T_{\mathrm{lin}}}$ are exactly the pairs of distinct users in a common cluster of $\mathcal{C}^{(k)}$. Moreover $G^k_t=G^k_{T_{\mathrm{lin}}}$ for all $T_{\mathrm{lin}}\le t\le T$, and $\widehat{C}^k_t=C^{(k)}_{j_k(i_t)}$ for all $T_{\mathrm{lin}}<t\le T$.

We verify the three parts in turn.

(a) Spectral growth. Fix a user i. For $n \geq 1$ let $\boldsymbol{Y}_n^{(i)} := \boldsymbol{x}_{\tau_{i,n}} \boldsymbol{x}_{\tau_{i,n}}^\top$ and $\mathcal{F}_n^{(i)} := \mathcal{H}_{\tau_{i,n}}$, the σ-field of the stopping time $\tau_{i,n}$, and let $\mathcal{F}_0^{(i)} := \mathcal{H}_0$. Since $\tau_{i,n-1} < \tau_{i,n}$, $(\mathcal{F}_n^{(i)})_{n\geq 0}$ is a filtration; $\boldsymbol{Y}_n^{(i)}$ is $\mathcal{F}_n^{(i)}$-measurable because $(\boldsymbol{x}_t)$ is adapted to $(\mathcal{H}_t)$; and $\mathbf{0} \preceq \boldsymbol{Y}_n^{(i)} \preceq \boldsymbol{I}$ because $\|\boldsymbol{x}\|_2 \leq 1$. We show that $\mathbb{E}[\boldsymbol{Y}_n^{(i)} | \mathcal{F}_{n-1}^{(i)}] \succeq \kappa_{x,K} \boldsymbol{I}$ almost surely. Let $\boldsymbol{v}$ be a unit vector and $A \in \mathcal{F}_{n-1}^{(i)}$. For every $t \geq 1$, $A \cap \{\tau_{i,n} = t\} = (A \cap \{\tau_{i,n-1} \leq t-1\}) \cap \{\tau_{i,n} = t\}$. The first set belongs to $\mathcal{H}_{t-1}$ by the defining property of $\mathcal{H}_{\tau_{i,n-1}}$, and the second belongs to $\mathcal{H}_t^-$, so $A \cap \{\tau_{i,n} = t\} \in \mathcal{H}_t^-$. Appendix C.2.2 (with ϕ the identity; the extended rounds use an arbitrary selection rule, which it allows) and $\tau_{i,n} < \infty$ almost surely give

$$\begin{aligned}\mathbb{E}\big[\mathbb{I}_A \langle \boldsymbol{v}, \boldsymbol{x}_{\tau_{i,n}} \rangle^2\big] &= \sum_{t\geq 1} \mathbb{E}\Big[\mathbb{I}_{A\cap\{\tau_{i,n}=t\}}\, \mathbb{E}\big[\langle \boldsymbol{v}, \boldsymbol{x}_t\rangle^2 \,\big|\, \mathcal{H}_t^-\big]\Big] \\ &\geq \kappa_{x,K} \sum_{t\geq 1} \mathbb{P}\big(A \cap \{\tau_{i,n} = t\}\big) = \kappa_{x,K}\, \mathbb{P}(A).\end{aligned}$$

Hence $\mathbb{E}[\langle \boldsymbol{v}, \boldsymbol{x}_{\tau_{i,n}}\rangle^2 | \mathcal{F}_{n-1}^{(i)}] \geq \kappa_{x,K}$ almost surely for each $\boldsymbol{v}$, and the countable-dense-set argument of Appendix C.2.2 gives the matrix inequality. Appendix C.2.1 with $S = \mathbb{R}^d$ (so that $r_S = d$), $\mathcal{F}_n = \mathcal{F}_n^{(i)}$, $\boldsymbol{Y}_n = \boldsymbol{Y}_n^{(i)}$ and $\kappa = \kappa_{x,K}$ therefore gives, for every $n \geq 1$,

$$\mathbb{P}\left(\lambda_{\min}\Big(\sum_{m=1}^{n} \boldsymbol{Y}_m^{(i)}\Big) \leq \frac{n\kappa_{x,K}}{2}\right) \leq d\, e^{-n\kappa_{x,K}/8}.$$

If $t \leq T$ and $T_{i,t} = n \geq 1$, the rounds $s \leq t$ with $i_s = i$ are exactly $\tau_{i,1}, \ldots, \tau_{i,n}$, so $\boldsymbol{M}_{i,t} = \lambda \boldsymbol{I} + \sum_{m=1}^{n} \boldsymbol{Y}_m^{(i)}$ and $\lambda_{\min}(\boldsymbol{M}_{i,t}) = \lambda + \lambda_{\min}(\sum_{m\leq n} \boldsymbol{Y}_m^{(i)})$. Moreover $T_{i,t} \leq t \leq T$. Hence $\mathcal{E}_{\mathrm{spec}}^c$ is contained in the union, over $i \in \mathcal{U}$ and $n_{\mathrm{lin}} \leq n \leq T$, of the events in the last display, and by the first term in (29),

$$\mathbb{P}(\mathcal{E}_{\mathrm{spec}}^c) \leq u\, T\, d\, e^{-n_{\mathrm{lin}}\kappa_{x,K}/8} \leq u\, T\, d \cdot \frac{\delta}{4udT} = \frac{\delta}{4}.$$

Because the matrix $\boldsymbol{M}_{i,t}$ is shared by all objectives in the linear model, no union bound over objectives is needed. On $\mathcal{E}_{\mathrm{spec}}$, if $T_{i,t} \geq n_{\mathrm{lin}}$ then $\lambda_{\min}(\boldsymbol{M}_{i,t}) \geq \kappa_{x,K} T_{i,t}/2$, and (27) gives

$$r_{i,t} = \frac{\rho(\boldsymbol{M}_{i,t})}{\sqrt{\lambda_{\min}(\boldsymbol{M}_{i,t})}} + \frac{\lambda}{\lambda_{\min}(\boldsymbol{M}_{i,t})} \leq \bar{\rho}_{\mathrm{lin}} \sqrt{\frac{2}{\kappa_{x,K}\, T_{i,t}}} + \frac{2\lambda}{\kappa_{x,K}\, T_{i,t}},$$

and, since $\lambda/\lambda_{\min}(\boldsymbol{M}_{i,t}) \leq \sqrt{\lambda}/\sqrt{\lambda_{\min}(\boldsymbol{M}_{i,t})}$, also $r_{i,t} \leq \beta(\boldsymbol{M}_{i,t})/\sqrt{\lambda_{\min}(\boldsymbol{M}_{i,t})} \leq \bar{\beta}_{\mathrm{lin}}\sqrt{2/(\kappa_{x,K} T_{i,t})}$. Since $T_{i,t} \geq n_{\mathrm{lin}} \geq n_\gamma$, one of the two terms defining n_γ in (29) is at most $T_{i,t}$. If $T_{i,t} \geq 72\bar{\beta}_{\mathrm{lin}}^2/(\kappa_{x,K}\gamma^2)$, the second bound gives $r_{i,t} \leq \bar{\beta}_{\mathrm{lin}}\sqrt{2\gamma^2/(72\bar{\beta}_{\mathrm{lin}}^2)} = \gamma/6$. Otherwise $T_{i,t} \geq 288\bar{\rho}_{\mathrm{lin}}^2/(\kappa_{x,K}\gamma^2)$ and $T_{i,t} \geq 24\lambda/(\kappa_{x,K}\gamma)$, and the first bound gives

$$r_{i,t} \leq \bar{\rho}_{\mathrm{lin}} \sqrt{\frac{2\kappa_{x,K}\gamma^2}{288\, \kappa_{x,K}\, \bar{\rho}_{\mathrm{lin}}^2}} + \frac{2\lambda\, \kappa_{x,K}\gamma}{24\lambda\, \kappa_{x,K}} = \frac{\gamma}{12} + \frac{\gamma}{12} = \frac{\gamma}{6}.$$

(b) Deletion of cross-cluster edges. Suppose, without loss of generality, that $i_t = i$, so that ℓ is a neighbor of i_t in G_{t-1}^k and the edge is tested in round t after the update of i. Let $j := j_k(i)$ and $j' := j_k(\ell) \neq j$. By the argument of Appendix C.3.2, on $\mathcal{E}_{\mathrm{conf}}$ we have $\|\hat{\boldsymbol{w}}_{i,t}^k - \boldsymbol{\theta}_j^k\|_2 \leq r_{i,t}$ and $\|\hat{\boldsymbol{w}}_{\ell,t}^k - \boldsymbol{\theta}_{j'}^k\|_2 \leq r_{\ell,t}$; by (a), $r_{i,t}, r_{\ell,t} \leq \gamma/6$. (A3) and the triangle inequality give

$$\big\|\hat{\boldsymbol{w}}_{i,t}^k - \hat{\boldsymbol{w}}_{\ell,t}^k\big\|_2 \geq \big\|\boldsymbol{\theta}_j^k - \boldsymbol{\theta}_{j'}^k\big\|_2 - r_{i,t} - r_{\ell,t} \geq \gamma - \frac{\gamma}{3} = \frac{2\gamma}{3} > \frac{\gamma}{3} \geq r_{i,t} + r_{\ell,t},$$

so the strict test (24) fires and the edge is deleted. No separating context is needed. For the second statement, let $t^* := \max\{\tau_{i,n_{\mathrm{lin}}}, \tau_{\ell,n_{\mathrm{lin}}}\}$ and suppose $t^* \leq T$. Only one user arrives per round, so $\tau_{i,n_{\mathrm{lin}}} \neq \tau_{\ell,n_{\mathrm{lin}}}$, and the user attaining the maximum arrives in round t^*. After the update of round t^* both users have at least n_{lin} visits. If $(i,\ell) \in E_{t^*-1}^k$, the first statement shows that it is

deleted in round t^*; otherwise it was deleted earlier. Since edges are never added, $(i,\ell) \notin E_t^k$ for all $t^* \le t \le T$.

(c) Arrivals and recovery. Fix a user i. The count $T_{i,T_{\mathrm{lin}}}$ is a sum of T_{lin} independent Bernoulli$(1/u)$ variables, with mean $\mu := T_{\mathrm{lin}}/u \ge 16\log(4u/\delta) + 4n_{\mathrm{lin}}$ by (30). The multiplicative Chernoff bound $\mathbb{P}(N \le (1-\varepsilon)\mu) \le e^{-\varepsilon^2\mu/2}$ for a sum N of independent Bernoulli variables with mean μ, applied with $\varepsilon = 1/2$, gives $\mathbb{P}(T_{i,T_{\mathrm{lin}}} \le \mu/2) \le e^{-\mu/8}$. Since $\mu/2 \ge 2n_{\mathrm{lin}} \ge n_{\mathrm{lin}}$, we have $\{T_{i,T_{\mathrm{lin}}} < n_{\mathrm{lin}}\} \subseteq \{T_{i,T_{\mathrm{lin}}} \le \mu/2\}$, and hence

$$\mathbb{P}\big(T_{i,T_{\mathrm{lin}}} < n_{\mathrm{lin}}\big) \le e^{-\mu/8} \le e^{-2\log(4u/\delta)} = \Big(\frac{\delta}{4u}\Big)^2 \le \frac{\delta}{4u}.$$

A union bound over the u users gives $\mathbb{P}(\mathcal{E}_{\mathrm{arr}}^c) \le \delta/4$. On $\mathcal{E}_{\mathrm{arr}}$, $\tau_{i,n_{\mathrm{lin}}} \le T_{\mathrm{lin}} \le T$ for every user i, so by (b) every cross-cluster edge is absent from $G^k_{T_{\mathrm{lin}}}$, and by Appendix C.3.2 every same-cluster edge is present in every G^k_t, $t \le T$. Hence the edges of $G^k_{T_{\mathrm{lin}}}$ are exactly the same-cluster pairs. After round T_{lin} no cross-cluster edge remains to be deleted and no same-cluster edge is ever deleted, so $G^k_t = G^k_{T_{\mathrm{lin}}}$ for $T_{\mathrm{lin}} \le t \le T$. In $G^k_{T_{\mathrm{lin}}}$ the closed neighbourhood of every user is its true cluster. Since round t uses G^k_{t-1}, we obtain $\widehat{C}^k_t = C^{(k)}_{j_k(i_t)}$ for every $T_{\mathrm{lin}} < t \le T$: the partitions are recovered at the end of round T_{lin}, and the correct cluster statistics are used from round $T_{\mathrm{lin}}+1$ on.

C.4 The Regret Bound

Let $\mathcal{E} := \mathcal{E}_{\mathrm{conf}} \cap \mathcal{E}_{\mathrm{spec}} \cap \mathcal{E}_{\mathrm{arr}}$, where we set $\mathcal{E}_{\mathrm{arr}} := \Omega$ if $T < T_{\mathrm{lin}}$. By Appendices C.3.1 and C.3.3, $\mathbb{P}(\mathcal{E}^c) \le 3\delta/4 \le \delta$. We work on $\mathcal{E}$ for the rest of the proof. Conclusion (i) is shown in Appendix C.3.2, and conclusion (ii) is part (c) of Appendix C.3.3. It remains to prove (iii).

Step 1: trivial bound and rounds before recovery. By (32), $\Delta_t \le 2$ in every round, so $R_T \le 2T$. If $T < T_{\mathrm{lin}}$, then $R_T \le 2T < 2T_{\mathrm{lin}}$, and both terms of the minimum in (31) are at least R_T. Assume from now on that $T \ge T_{\mathrm{lin}}$. The rounds $t \le T_{\mathrm{lin}}$ contribute at most $2T_{\mathrm{lin}}$ to the regret.

Step 2: Pareto step. Fix a round t with $T_{\mathrm{lin}} < t \le T$ and write $C^k := C^{(k)}_{j_k(i_t)}$. By Appendix C.3.3(c), $\widehat{C}^k_t = C^k$ for every k, so by (39) the upper confidence bound of objective k is computed from the statistics of the fixed true cluster C^k at index $t-1$: $\bar{\boldsymbol{w}}^k_t = \hat{\boldsymbol{w}}^k_{C^k,t-1}$ and $\bar{\boldsymbol{M}}^k_t = \boldsymbol{M}_{C^k,t-1}$. Appendix C.3.1, applied to the group $C^k \in \mathcal{A}^{(k)}$ at index $t-1$, and (27) give, for all $\boldsymbol{x} \in \mathcal{X}_t$ and all k,

$$\mu^k_{i_t}(\boldsymbol{x}) \le U^k_t(\boldsymbol{x}) \le \mu^k_{i_t}(\boldsymbol{x}) + 2\beta(\bar{\boldsymbol{M}}^k_t)\|\boldsymbol{x}\|_{(\bar{\boldsymbol{M}}^k_t)^{-1}} \le \mu^k_{i_t}(\boldsymbol{x}) + 2\bar{\beta}_{\mathrm{lin}}\|\boldsymbol{x}\|_{(\bar{\boldsymbol{M}}^k_t)^{-1}}. \tag{40}$$

Fix a competitor $\boldsymbol{x}' \in \mathcal{X}_t$. Since $\boldsymbol{x}_t \in \tilde{\mathcal{O}}_t$, the vector $\boldsymbol{U}_t(\boldsymbol{x}')$ does not Pareto-dominate $\boldsymbol{U}_t(\boldsymbol{x}_t)$. Hence there is an objective $k^* = k^*(\boldsymbol{x}')$ with $U^{k^*}_t(\boldsymbol{x}_t) \ge U^{k^*}_t(\boldsymbol{x}')$; otherwise $U^k_t(\boldsymbol{x}') > U^k_t(\boldsymbol{x}_t)$ for all k, which is Pareto dominance. Using (40) twice,

$$\begin{aligned}\min_{k\in[m_o]}\big[\mu^k_{i_t}(\boldsymbol{x}') - \mu^k_{i_t}(\boldsymbol{x}_t)\big] &\le \mu^{k^*}_{i_t}(\boldsymbol{x}') - \mu^{k^*}_{i_t}(\boldsymbol{x}_t) \le U^{k^*}_t(\boldsymbol{x}') - \mu^{k^*}_{i_t}(\boldsymbol{x}_t)\\ &\le U^{k^*}_t(\boldsymbol{x}_t) - \mu^{k^*}_{i_t}(\boldsymbol{x}_t) \le 2\bar{\beta}_{\mathrm{lin}}\, s^{k^*}_t.\end{aligned}$$

The right-hand side is nonnegative and at most $2\bar{\beta}_{\mathrm{lin}}\max_k s^k_t$, so the same bound holds for the positive part. Taking the maximum over $\boldsymbol{x}' \in \mathcal{X}_t$ in (32) yields

$$\Delta_t \le 2\bar{\beta}_{\mathrm{lin}} \max_{k\in[m_o]} s^k_t \qquad \text{for all } T_{\mathrm{lin}} < t \le T. \tag{41}$$

Step 3: elliptical potential per true cluster. Fix an objective k and a true cluster $C := C^{(k)}_j$, and let $t_1 < \cdots < t_N$ be the rounds $t \in \{T_{\mathrm{lin}}+1, \dots, T\}$ with $i_t \in C$. For $m \in [N]$, Step 2 gives $s^k_{t_m} = \|\boldsymbol{x}_{t_m}\|_{\boldsymbol{M}^{-1}_{C,t_m-1}}$, and

$$\boldsymbol{M}_{C,t_m-1} = \underbrace{\lambda\boldsymbol{I} + \sum_{m'<m} \boldsymbol{x}_{t_{m'}}\boldsymbol{x}_{t_{m'}}^\top}_{=:\boldsymbol{V}_m} + \sum_{s\le T_{\mathrm{lin}}:\, i_s\in C} \boldsymbol{x}_s\boldsymbol{x}_s^\top \succeq \boldsymbol{V}_m,$$

because the rounds $s \in (T_{\text{lin}}, t_m)$ with $i_s \in C$ are exactly $t_1, \dots, t_{m-1}$. The cluster matrix also contains the samples of C collected before recovery. These samples only help: matrix inversion reverses the Loewner order on positive definite matrices, so $\boldsymbol{M}_{C,t_m-1}^{-1} \preceq \boldsymbol{V}_m^{-1}$ and

$$\big(s_{t_m}^k\big)^2 \le z_m := \|\boldsymbol{x}_{t_m}\|_{\boldsymbol{V}_m^{-1}}^2 \le \frac{\|\boldsymbol{x}_{t_m}\|_2^2}{\lambda} \le 1,$$

where the last inequality uses $\boldsymbol{V}_m \succeq \lambda \boldsymbol{I}$, $\lambda \ge 1$ and $\|\boldsymbol{x}\|_2 \le 1$. By the rank-one determinant identity, $\det \boldsymbol{V}_{m+1} = \det(\boldsymbol{V}_m + \boldsymbol{x}_{t_m}\boldsymbol{x}_{t_m}^\top) = \det(\boldsymbol{V}_m)(1 + z_m)$. Moreover $z \le 2\log(1+z)$ for $z \in [0,1]$, because $\log(1+z)$ is concave and therefore $\log(1+z) \ge z \log 2 \ge z/2$ on $[0,1]$. Hence, with (26) (the matrix $\boldsymbol{V}_{N+1}$ is $\lambda \boldsymbol{I}$ plus $N \le T$ rank-one terms),

$$\sum_{m=1}^{N} \big(s_{t_m}^k\big)^2 \le \sum_{m=1}^{N} z_m \le 2 \sum_{m=1}^{N} \log(1+z_m) = 2 \log\det\big(\lambda^{-1}\boldsymbol{V}_{N+1}\big) \le 2d \log\Big(1 + \frac{T}{d\lambda}\Big).$$

This is the elliptical potential argument of Abbasi-Yadkori et al. (2011), applied separately to each true cluster. By the choice of $t_1, \dots, t_N$, the left-hand side of the last display equals $\sum_{t=T_{\text{lin}}+1}^{T} \mathbb{I}\{i_t \in C_j^{(k)}\}(s_t^k)^2$. For each k and t, the user i_t belongs to exactly one true cluster of objective k, so $\max_k (s_t^k)^2 \le \sum_k (s_t^k)^2 = \sum_k \sum_j \mathbb{I}\{i_t \in C_j^{(k)}\}(s_t^k)^2$. Summing over t and applying the last display to each of the $m_c = \sum_k m_c^{(k)}$ pairs (k,j) gives

$$\sum_{t=T_{\text{lin}}+1}^{T} \max_{k \in [m_o]} \big(s_t^k\big)^2 \le \sum_{k=1}^{m_o} \sum_{j=1}^{m_c^{(k)}} \sum_{t=T_{\text{lin}}+1}^{T} \mathbb{I}\big\{i_t \in C_j^{(k)}\big\}\big(s_t^k\big)^2 \le 2 m_c\, d \log\Big(1 + \frac{T}{d\lambda}\Big). \tag{42}$$

Step 4: conclusion. By (41), the Cauchy–Schwarz inequality and (42),

$$\sum_{t=T_{\text{lin}}+1}^{T} \Delta_t \le 2\bar{\beta}_{\text{lin}} \sqrt{T \sum_{t=T_{\text{lin}}+1}^{T} \max_k \big(s_t^k\big)^2} \le 2\bar{\beta}_{\text{lin}} \sqrt{2 m_c\, d\, T \log\Big(1 + \frac{T}{d\lambda}\Big)}.$$

Adding the contribution $2T_{\text{lin}}$ of the first T_{lin} rounds and combining with $R_T \le 2T$ proves (31) on $\mathcal{E}$.

Order of the bound. Since $\lceil y \rceil \le y + 1$, a maximum is at most a sum, and $(a+b)^2 \le 2a^2 + 2b^2$,

$$T_{\text{lin}} \le 16u \log\frac{4u}{\delta} + 4u\, n_{\text{lin}} + 1, \qquad n_{\text{lin}} \le \frac{8}{\kappa_{x,K}} \log\frac{4udT}{\delta} + \frac{288\,\bar{\rho}_{\text{lin}}^2}{\kappa_{x,K}\gamma^2} + \frac{24\lambda}{\kappa_{x,K}\gamma} + 1,$$

$$\bar{\rho}_{\text{lin}}^2 = R^2\Big(d \log\Big(1 + \frac{T}{d\lambda}\Big) + 2\log\frac{8um_o}{\delta}\Big), \qquad \bar{\beta}_{\text{lin}}^2 \le 2\lambda + 2\bar{\rho}_{\text{lin}}^2.$$

Treating λ as a constant and hiding factors polylogarithmic in T, u, m_o, d and $1/\delta$, we have $\bar{\rho}_{\text{lin}}^2 = \widetilde{O}(R^2 d)$; since $\kappa_{x,K} \le 1$, the terms $16u\log(4u/\delta)$, $4u$ and 1 are $\widetilde{O}(u/\kappa_{x,K})$, and hence $2T_{\text{lin}} = \widetilde{O}\big((u/\kappa_{x,K})(1 + R^2 d/\gamma^2 + 1/\gamma)\big)$. Treating also R as a constant, $\bar{\beta}_{\text{lin}}^2 = \widetilde{O}(d)$, $2T_{\text{lin}} = \widetilde{O}\big((u/\kappa_{x,K})(1 + d/\gamma^2)\big)$ and $2\bar{\beta}_{\text{lin}}\sqrt{2m_c\, d\, T \log(1 + T/(d\lambda))} = \widetilde{O}(d\sqrt{m_c T})$, so that, on the event $\mathcal{E}$ of the proof and hence with probability at least $1 - \delta$,

$$R_T = \widetilde{O}\Big(\frac{u}{\kappa_{x,K}}\Big(1 + \frac{d}{\gamma^2}\Big) + d\sqrt{m_c T}\Big). \tag{43}$$

The first term does not grow with T except through logarithmic factors, but it can be large. It scales with $1/\kappa_{x,K}$, and in general Appendix C.1.1 only guarantees an exponentially small $\kappa_{x,K}$; for instance, if $\mathcal{D}$ is uniform on $\{\pm \boldsymbol{e}_1, \dots, \pm \boldsymbol{e}_d\}$ (Appendix G.1), then $\kappa_{x,K} \le d^{-K}$, since for $\boldsymbol{v} = \boldsymbol{e}_1$ the minimum over the K candidates is nonzero only if all candidates lie in $\{\pm \boldsymbol{e}_1\}$. The first term is a worst-case bound: T_{lin} is a time after which the analysis guarantees recovery, not a prediction of when the edges are actually deleted, and this term need not be reached within finite experimental horizons. For spherical contexts, (13) replaces $1/\kappa_{x,K}$ by $O(dK^2)$.

C.5 Spherical Contexts

Let $\mathcal{D}$ be the law of $\boldsymbol{x} = \boldsymbol{g}/\|\boldsymbol{g}\|_2$ with $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ and $d \geq 3$, i.e., the uniform distribution on the unit sphere; in particular $\|\boldsymbol{x}\|_2 = 1$. Let

$$c_d := \frac{\Gamma(d/2)}{\sqrt{\pi}\,\Gamma((d-1)/2)}$$

and let $I_s(\cdot,\cdot)$ denote the regularized incomplete beta function. We show the following bounds, which contain (13):

$$\kappa_{x,K} \geq \frac{1}{32\,c_d^2 K^2} \geq \frac{\pi}{16(d-1)K^2}, \qquad c_d \leq \sqrt{\frac{d-1}{2\pi}}, \tag{44}$$

$$\kappa_{x,K} = \int_0^1 \Big[1 - I_s\Big(\frac{1}{2}, \frac{d-1}{2}\Big)\Big]^K ds, \qquad \kappa_{x,K} \sim \frac{1}{2c_d^2 K^2} \quad \text{as } K \to \infty \text{ for fixed } d, \tag{45}$$

and, consequently, under the remaining conditions of Section 4.1, with probability at least $1-\delta$,

$$R_T = \widetilde{O}\Big(u\,d\,K^2\Big(1 + \frac{d}{\gamma^2}\Big) + d\sqrt{m_c T}\Big). \tag{46}$$

To avoid a clash with the threshold a used below, we index the K candidates by $b \in [K]$ in this subsection.

Step 1: law of $\langle \boldsymbol{v}, \boldsymbol{x}\rangle$. For an orthogonal matrix $\boldsymbol{O}$, $\boldsymbol{Og} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ and $\|\boldsymbol{Og}\|_2 = \|\boldsymbol{g}\|_2$, so $\boldsymbol{Ox}$ has the same law as $\boldsymbol{x}$. Given a unit vector $\boldsymbol{v}$, choose $\boldsymbol{O}$ with $\boldsymbol{Ov} = \boldsymbol{e}_1$; then $\langle \boldsymbol{v}, \boldsymbol{x}\rangle = \langle \boldsymbol{e}_1, \boldsymbol{Ox}\rangle$ has the law of $Y := x_1$ for every unit $\boldsymbol{v}$. The variable Y is symmetric because g_1 is. Moreover $Y^2 = g_1^2/(g_1^2 + \sum_{q\geq 2} g_q^2)$, where $g_1^2 \sim \chi_1^2$ and $\sum_{q\geq 2} g_q^2 \sim \chi_{d-1}^2$ are independent; hence $Y^2 \sim \mathrm{Beta}(\frac{1}{2}, \frac{d-1}{2})$. For $y \in (0,1)$, differentiating $\mathbb{P}(|Y| \leq y) = \mathbb{P}(Y^2 \leq y^2)$ gives the density $2y \cdot (y^2)^{-1/2}(1-y^2)^{(d-3)/2}/B(\frac{1}{2}, \frac{d-1}{2})$ of $|Y|$. By symmetry, Y has the density

$$f_d(y) = c_d\,(1-y^2)^{(d-3)/2}, \qquad y \in (-1,1),$$

because $1/B(\frac{1}{2}, \frac{d-1}{2}) = \Gamma(\frac{d}{2})/(\Gamma(\frac{1}{2})\Gamma(\frac{d-1}{2})) = c_d$ and $\Gamma(\frac{1}{2}) = \sqrt{\pi}$. By the same symmetries (sign changes and permutations of coordinates), $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top]$ is a multiple of $\boldsymbol{I}_d$ with trace $\mathbb{E}\|\boldsymbol{x}\|_2^2 = 1$, so $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top] = \boldsymbol{I}_d/d$ and $\lambda_x = 1/d$. Appendix C.1.1(a) therefore only gives $\kappa_{x,K} \geq \frac{1}{2d}(2d-1)^{-K}$, which is exponentially small in K; the following steps give a polynomial bound.

Step 2: bounded density. For $d \geq 3$ the exponent $(d-3)/2$ is nonnegative, so $f_d(y) \leq c_d$ for all $y \in (-1,1)$. Consequently, for every unit $\boldsymbol{v}$ and every $a \geq 0$, $\mathbb{P}(|\langle \boldsymbol{v}, \boldsymbol{x}\rangle| \leq a) = \int_{-a}^{a} f_d(y)\,dy \leq 2ac_d$ (the integrand vanishes outside $(-1,1)$).

Step 3: union bound. Let $a := 1/(4c_dK)$ and let $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_K$ be independent with law $\mathcal{D}$. By Step 2 and a union bound over the K candidates,

$$\mathbb{P}\Big(\min_{1\leq b\leq K} |\langle \boldsymbol{v}, \boldsymbol{x}_b\rangle| \leq a\Big) \leq K \cdot 2ac_d = \frac{1}{2}.$$

Step 4: expectation of the minimum. Since $\min_b \langle \boldsymbol{v}, \boldsymbol{x}_b\rangle^2 \geq 0$, and $\min_b \langle \boldsymbol{v}, \boldsymbol{x}_b\rangle^2 > a^2$ with probability at least $1/2$ by Step 3,

$$\mathbb{E}\Big[\min_{1\leq b\leq K} \langle \boldsymbol{v}, \boldsymbol{x}_b\rangle^2\Big] \geq a^2\,\mathbb{P}\Big(\min_b \langle \boldsymbol{v}, \boldsymbol{x}_b\rangle^2 > a^2\Big) \geq \frac{a^2}{2} = \frac{1}{32\,c_d^2 K^2}.$$

The unit vector $\boldsymbol{v}$ was arbitrary, so $\kappa_{x,K} \geq 1/(32c_d^2K^2)$.

Step 5: Gautschi-type bound on c_d. For $y > 0$, the Cauchy–Schwarz inequality gives

$$\Gamma\Big(y + \frac{1}{2}\Big) = \int_0^\infty \big(t^{y-1}e^{-t}\big)^{1/2}\big(t^y e^{-t}\big)^{1/2} dt \leq \Gamma(y)^{1/2}\,\Gamma(y+1)^{1/2} = \sqrt{y}\,\Gamma(y),$$

which is the log-convexity of Γ combined with $\Gamma(y+1) = y\Gamma(y)$; equivalently, it is the lower half $\sqrt{y} \leq \Gamma(y+1)/\Gamma(y+\frac{1}{2})$ of Gautschi's inequality. With $y = (d-1)/2$ this reads $\Gamma(d/2) \leq$

$\sqrt{(d-1)/2}\,\Gamma((d-1)/2)$, i.e., $c_d \le \sqrt{(d-1)/(2\pi)}$. Hence $1/(32c_d^2K^2) \ge 2\pi/(32(d-1)K^2) = \pi/(16(d-1)K^2)$, which completes the proof of (44).

Step 6: exact value. Fix a unit $\boldsymbol{v}$ and let $Z_b := \langle \boldsymbol{v}, \boldsymbol{x}_b\rangle^2$. By Step 1, $Z_1, \dots, Z_K$ are independent with law $\mathrm{Beta}(\frac{1}{2}, \frac{d-1}{2})$, whose distribution function is $F(s) := I_s(\frac{1}{2}, \frac{d-1}{2})$. Since $0 \le \min_b Z_b \le 1$,

$$\mathbb{E}\Big[\min_b Z_b\Big] = \int_0^1 \mathbb{P}\Big(\min_b Z_b > s\Big)\,ds = \int_0^1 \big(1 - F(s)\big)^K\,ds.$$

This value does not depend on $\boldsymbol{v}$, so it equals the infimum $\kappa_{x,K}$, which proves the first part of (45).

Step 7: asymptotics. Substituting $s = w/K^2$,

$$K^2\kappa_{x,K} = \int_0^\infty \mathbb{I}\{w \le K^2\}\Big(1 - F\big(w/K^2\big)\Big)^K\,dw.$$

By Step 1, $F(s) = \mathbb{P}(|Y| \le \sqrt{s}) = 2\int_0^{\sqrt{s}} f_d(y)\,dy$. Since f_d is continuous at 0 with $f_d(0) = c_d$, we have $F(s)/\sqrt{s} \to 2c_d$ as $s \to 0$. For fixed $w > 0$, therefore, $K\,F(w/K^2) = \sqrt{w}\,F(w/K^2)/\sqrt{w/K^2} \to 2c_d\sqrt{w}$ and $F(w/K^2) \to 0$. Since $\log(1-z) = -z(1+o(1))$ as $z \to 0$, this gives

$$\big(1 - F(w/K^2)\big)^K = \exp\big(K\log(1 - F(w/K^2))\big) \to e^{-2c_d\sqrt{w}}.$$

For domination, note that $(1-y^2)^{(d-3)/2} \ge 2^{-(d-3)/2}$ for $|y| \le 1/\sqrt{2}$. Hence $F(s) \ge 2c_d 2^{-(d-3)/2}\sqrt{s} \ge c_*\sqrt{s}$ for $s \le 1/2$, where $c_* := 2c_d 2^{-(d-2)/2}$. For $s \in (1/2, 1]$ we have $F(s) \ge F(1/2) \ge c_* \ge c_*\sqrt{s}$. Hence, for $w \le K^2$, $(1 - F(w/K^2))^K \le \exp(-K\,F(w/K^2)) \le e^{-c_*\sqrt{w}}$, which is integrable on $[0,\infty)$. By dominated convergence and the substitution $w = y^2$,

$$K^2\kappa_{x,K} \;\longrightarrow\; \int_0^\infty e^{-2c_d\sqrt{w}}\,dw = \int_0^\infty 2y\,e^{-2c_d y}\,dy = \frac{2}{(2c_d)^2} = \frac{1}{2c_d^2},$$

which is the second part of (45). In particular, the lower bound (44) has the correct order K^{-2} in K, and for large K it is smaller than $\kappa_{x,K}$ by a factor of about 16.

Step 8: resulting order. The contexts satisfy $\|\boldsymbol{x}\|_2 \le 1$ and, by (44), $\kappa_{x,K} > 0$, so (A4) holds and the guarantee of Section 4.1 applies. Moreover n_{lin} and T_{lin} are nonincreasing functions of $\kappa_{x,K}$. By (44), $1/\kappa_{x,K} \le 16(d-1)K^2/\pi \le 16dK^2/\pi$, and substituting this into (43) gives

$$\frac{u}{\kappa_{x,K}}\Big(1 + \frac{d}{\gamma^2}\Big) \le \frac{16}{\pi}\,u\,d\,K^2\Big(1 + \frac{d}{\gamma^2}\Big),$$

which yields (46) on the event $\mathcal{E}$ of Appendix C.4, whose probability is at least $1 - \delta$.

D ANALYSIS OF COMONB-FROZEN

This appendix proves the guarantee of Section 4.2 for COMONB-Frozen (Algorithm 2). The argument uses no property of neural networks. It relies only on uniform arrival with i.i.d. candidate sets (A1), the noise condition (A2), the fixed-feature model (A5), observable separation (A6), the standing condition $\mathbf{\Sigma}_\phi^k \neq \mathbf{0}$ for every k of Section 2.3 (which gives $\kappa > 0$), and $\lambda \geq 1$. If the features are built from a random initialization, the analysis is conditional on that initialization, which is drawn before and independently of the interaction; the feature-dependent quantities $\mathcal{S}_k$, r, λ_ϕ, κ, G_T and B then refer to the realized features. The argument follows the same chain as the analysis of COMOLB in Appendix C, except that edges are deleted in prediction space. The spectral results that we use, Appendices C.1.1, C.2.1 and C.2.2, are stated and proved in Appendix C; here they are only invoked, on the effective subspaces $\mathcal{S}_k$.

Conclusions and probability allocation. Recall the constants of Section 4.2:

$$\begin{aligned}
&\bar{\rho}_T = R\sqrt{G_T + 2\log(8um_o/\delta)}, \quad \bar{\beta}_T = B\sqrt{\lambda} + \bar{\rho}_T, \quad n_{\mathrm{spec}} = \Big\lceil \frac{8}{\kappa} \log \frac{4um_o rT}{\delta} \Big\rceil, \\
&n_\star = \max\Big\{ n_{\mathrm{spec}},\ \min\Big\{ \Big\lceil \frac{72\,\bar{\beta}_T^2}{\kappa\,\gamma'^2} \Big\rceil,\ \max\Big\{ \Big\lceil \frac{288\,\bar{\rho}_T^2}{\kappa\,\gamma'^2} \Big\rceil, \Big\lceil \frac{24\lambda B}{\kappa\,\gamma'} \Big\rceil \Big\} \Big\} \Big\}, \\
&t_{\mathrm{fit}} = \Big\lceil 16u \log \frac{4u}{\delta} + 4un_\star \Big\rceil, \qquad T_{\mathrm{clust}} = t_{\mathrm{fit}} + \Big\lceil \frac{u}{q_\gamma} \log \frac{4u^2 m_o}{\delta} \Big\rceil.
\end{aligned} \tag{47}$$

We construct an event $\mathcal{E}$ with $\mathbb{P}(\mathcal{E}) \geq 1 - \delta$ on which (i) no edge between two users of the same cluster of any objective is ever deleted; (ii) if $T \geq T_{\mathrm{clust}}$, then at the end of round T_{clust}, and in every later round, the edges of every graph G^k are exactly the pairs of users in a common true cluster, so that the closed neighbourhood of every user is its true cluster; and (iii) $R_T \leq \min\{2T,\ 2T_{\mathrm{clust}} + 2\bar{\beta}_T\sqrt{2m_c T G_T}\}$. The event $\mathcal{E}$ is the intersection of a probability-one event Ω_0 and four events, each of failure probability at most $\delta/4$: the confidence event $\mathcal{E}_{\mathrm{conf}}$ (Appendix D.2), the spectral event $\mathcal{E}_{\mathrm{spec}}$ (Appendix D.3), the arrival event $\mathcal{E}_{\mathrm{arr}}$ (Appendix D.6.1) and the covering event $\mathcal{E}_{\mathrm{cov}}$ (Appendix D.6.2). A union bound gives $\mathbb{P}(\mathcal{E}^c) \leq 4 \cdot \delta/4 = \delta$. The order of the arguments matters. The confidence event is established for *fixed* groups of users, namely single users and true clusters, which do not depend on the data. The clusters estimated by the algorithm are random and enter only through the recovery argument, which is carried out on the intersection of the events. At no point do we condition on correct clustering and then apply a concentration inequality.

D.1 SETTING, NOTATION AND PRELIMINARIES

Model and filtrations. By (A5), $\mu_i^k(\boldsymbol{x}) = \langle \phi^k(\boldsymbol{x}), \boldsymbol{w}_{j_k(i)}^k \rangle$ with $\|\boldsymbol{w}_j^k\|_2 \leq B$ and $\|\phi^k(\boldsymbol{x})\|_2 \leq 1$; we write $\mu_{C_j}^k(\boldsymbol{x}) = \langle \phi^k(\boldsymbol{x}), \boldsymbol{w}_j^k \rangle$ for the common reward function of $C_j^{(k)}$. Let $\mathcal{H}_t$ be the σ-field of the history up to the end of round t: the users, candidate sets, selected contexts and rewards of rounds $1, \ldots, t$, together with the internal randomization of the algorithm in these rounds, as in Section 2.1; $\mathcal{H}_0$ is trivial. As in Appendix C, let

$$\mathcal{H}_t^- := \sigma(\mathcal{H}_{t-1}, i_t), \qquad \mathcal{H}_{t-1}^+ := \sigma(\mathcal{H}_{t-1}, i_t, \mathcal{X}_t, \boldsymbol{x}_t),$$

so that $\mathcal{H}_t^-$ is the information available before the candidates of round t are revealed, and $\mathcal{H}_{t-1}^+$ is the information available when the reward of round t is generated. Since $\mathcal{H}_{t-1}^+ \subseteq \mathcal{H}_t \subseteq \mathcal{H}_t^+$, $(\mathcal{H}_t^+)_{t \geq 0}$ is a filtration. By (A2), conditionally on $\mathcal{H}_{t-1}^+$, each η_t^k has mean zero and is R-sub-Gaussian; the noise terms of different objectives may be dependent, and we only use this marginal property for each k. By (A1), i_t, $\mathcal{X}_t$ and $\mathcal{H}_{t-1}$ are mutually independent; in particular, $\mathcal{X}_t$ is independent of $\mathcal{H}_t^-$, and the pairs $(i_t, \mathcal{X}_t)$, $t \geq 1$, are i.i.d.

Algorithm 2 does not use T except as its stopping time. To define arrival times for all users, we extend the interaction beyond round T as in Appendix C, choosing the arbitrary selection rule allowed there to be uniform: for $t > T$, users and candidate sets are drawn as before, $\boldsymbol{x}_t$ is drawn uniformly at random from $\mathcal{X}_t$, independently of everything else (the uniform choice is used only in Appendix D.11.1), and no rewards are generated; $\mathcal{H}_t$, $\mathcal{H}_t^-$ and $\mathcal{H}_{t-1}^+$ are defined as before, without rewards, and $T_{i,t}$ denotes the number of arrivals of user i in rounds $1, \ldots, t$ also for $t > T$. This

extension does not change rounds $1, \dots, T$, all conclusions concern only these rounds, and, by uniform arrival, every user arrives infinitely often almost surely. Appendix C.2.2 allows an arbitrary selection rule and therefore applies to the extended rounds as well. Rewards and noise enter the analysis only through rounds $t \le T$ (Appendix D.2).

Effective subspaces. Let $\boldsymbol{\Sigma}_\phi^k$, $\mathcal{S}_k$, r_k, r and λ_ϕ be as in Section 2.3, and let $\boldsymbol{P}_{\mathcal{S}_k}$ be the orthogonal projection onto $\mathcal{S}_k$. Since $\boldsymbol{v}^\top \boldsymbol{\Sigma}_\phi^k \boldsymbol{v} = \mathbb{E}\langle \boldsymbol{v}, \phi^k(\boldsymbol{x})\rangle^2 = 0$ for every $\boldsymbol{v} \perp \mathcal{S}_k$, applying this to a basis of $\mathcal{S}_k^\perp$ shows that $\phi^k(\boldsymbol{x}) \in \mathcal{S}_k$ for $\mathcal{D}$-almost every $\boldsymbol{x}$. As countably many candidates are drawn from $\mathcal{D}$, the event

$$\Omega_0 := \big\{ \phi^k(\boldsymbol{z}) \in \mathcal{S}_k \text{ for all } k \in [m_o],\ t \ge 1,\ \boldsymbol{z} \in \mathcal{X}_t \big\}$$

has probability one. On Ω_0, every selected and every candidate feature vector lies in the corresponding $\mathcal{S}_k$; all events below are intersected with Ω_0 without further mention. As stated with (8), the supremum defining G_T ranges over contexts whose features lie in $\mathcal{S}_k$, which on Ω_0 includes all contexts that occur. For a symmetric matrix $\boldsymbol{M}$ and a subspace $\mathcal{S}$, write $\lambda_{\min}(\boldsymbol{M}|_{\mathcal{S}}) := \min\{\boldsymbol{v}^\top \boldsymbol{M} \boldsymbol{v} : \boldsymbol{v} \in \mathcal{S},\ \|\boldsymbol{v}\|_2 = 1\}$.

Block structure. Let $\boldsymbol{M} = \lambda \boldsymbol{I} + \sum_{s=1}^n \boldsymbol{\phi}_s \boldsymbol{\phi}_s^\top$ with $\boldsymbol{\phi}_1, \dots, \boldsymbol{\phi}_n \in \mathcal{S}_k$, and let $\boldsymbol{A} = \boldsymbol{M} - \lambda \boldsymbol{I}$. For $\boldsymbol{y} \in \mathcal{S}_k^\perp$, $\boldsymbol{A}\boldsymbol{y} = \sum_s \boldsymbol{\phi}_s \langle \boldsymbol{\phi}_s, \boldsymbol{y}\rangle = \boldsymbol{0}$, and $\boldsymbol{A}$ maps $\mathcal{S}_k$ into $\mathcal{S}_k$. Hence $\boldsymbol{M}$ leaves $\mathcal{S}_k$ and $\mathcal{S}_k^\perp$ invariant, acts as $\lambda \boldsymbol{I}$ on $\mathcal{S}_k^\perp$, and $\boldsymbol{M}^{-1}$ has the same block structure. Its block on $\mathcal{S}_k$ is the inverse of the block of $\boldsymbol{M}$ on $\mathcal{S}_k$, whose eigenvalues are at least $\lambda_{\min}(\boldsymbol{M}|_{\mathcal{S}_k})$. Therefore

$$\|\boldsymbol{v}\|_{\boldsymbol{M}^{-1}}^2 \le \frac{\|\boldsymbol{v}\|_2^2}{\lambda_{\min}(\boldsymbol{M}|_{\mathcal{S}_k})} \qquad \text{for all } \boldsymbol{v} \in \mathcal{S}_k. \tag{48}$$

Size of κ. Keeping only $a = 1$ inside the minimum in (7) gives $\kappa \le \min_k \inf\{\boldsymbol{v}^\top \boldsymbol{\Sigma}_\phi^k \boldsymbol{v} : \boldsymbol{v} \in \mathcal{S}_k,\ \|\boldsymbol{v}\|_2 = 1\} = \min_k \lambda_{\min}^+(\boldsymbol{\Sigma}_\phi^k) = \lambda_\phi$, and $\lambda_\phi \le 1$ because $\lambda_{\min}^+(\boldsymbol{\Sigma}_\phi^k) \le \operatorname{tr}(\boldsymbol{\Sigma}_\phi^k) = \mathbb{E}\|\phi^k(\boldsymbol{x})\|_2^2 \le 1$. Appendix C.1.1(b) gives the lower bound in

$$\lambda_\phi \ \ge\ \kappa \ \ge\ \frac{\lambda_\phi}{2}\Big(\frac{\lambda_\phi}{2 - \lambda_\phi}\Big)^K \ >\ 0. \tag{49}$$

Group statistics. For an objective k and a set of users A contained in a single true cluster $C_j^{(k)}$, write $\boldsymbol{w}_A^k := \boldsymbol{w}_j^k$ and $\mu_A^k := \mu_{C_j}^k$, and define, for $0 \le t \le T$,

$$\begin{aligned} \boldsymbol{M}_{A,t}^k &= \lambda \boldsymbol{I} + \sum_{s \le t:\, i_s \in A} \phi^k(\boldsymbol{x}_s) \phi^k(\boldsymbol{x}_s)^\top, \qquad \boldsymbol{b}_{A,t}^k = \sum_{s \le t:\, i_s \in A} r_s^k\, \phi^k(\boldsymbol{x}_s), \\ \hat{\boldsymbol{w}}_{A,t}^k &= (\boldsymbol{M}_{A,t}^k)^{-1} \boldsymbol{b}_{A,t}^k, \qquad \hat{f}_{A,t}^k(\boldsymbol{x}) = \phi^k(\boldsymbol{x})^\top \hat{\boldsymbol{w}}_{A,t}^k. \end{aligned} \tag{50}$$

For $A = \{i\}$, these are the user statistics (15) maintained by Algorithm 2, because the statistics of a user change only in the rounds in which the user arrives. For $A = C_j^{(k)}$, they are the statistics of the true cluster; the algorithm does not compute them, but they are used in the analysis. Since every round belongs to exactly one user, $\sum_{i \in C} (\boldsymbol{M}_{i,t-1}^k - \lambda \boldsymbol{I}) = \boldsymbol{M}_{C,t-1}^k - \lambda \boldsymbol{I}$ and $\sum_{i \in C} \boldsymbol{b}_{i,t-1}^k = \boldsymbol{b}_{C,t-1}^k$ for every set of users C. Hence, by (16),

$$\widehat{C}_t^k = C_j^{(k)} \quad \Longrightarrow \quad \bar{\boldsymbol{M}}_t^k = \boldsymbol{M}_{C_j^{(k)},t-1}^k, \qquad \bar{\boldsymbol{w}}_t^k = \hat{\boldsymbol{w}}_{C_j^{(k)},t-1}^k. \tag{51}$$

Log-determinants. Every matrix $\boldsymbol{M}_{A,t}^k$ with $t \le T$ has the form $\lambda \boldsymbol{I} + \sum_{s=1}^n \phi^k(\boldsymbol{x}_s) \phi^k(\boldsymbol{x}_s)^\top$ with $n \le T$ observed contexts, whose features lie in $\mathcal{S}_k$ on Ω_0, so $\log\det(\lambda^{-1} \boldsymbol{M}_{A,t}^k) \le G_T$ by (8). Since $\rho(\cdot)$ and $\beta(\cdot)$ are increasing in the log-determinant,

$$\rho(\boldsymbol{M}_{i,t}^k) \le \bar{\rho}_T, \quad \beta(\boldsymbol{M}_{i,t}^k) \le \bar{\beta}_T \quad \text{and} \quad \beta(\boldsymbol{M}_{C_j^{(k)},t}^k) \le \bar{\beta}_T \qquad \text{for all } i,\ k,\ j \text{ and } 0 \le t \le T. \tag{52}$$

We also use the log-determinant of the contexts actually selected for the true cluster $C_j^{(k)}$ up to round T,

$$l_{j,T}^k := \log\det\big(\lambda^{-1} \boldsymbol{M}_{C_j^{(k)},T}^k\big) \ \le\ G_T \ \le\ r \log\Big(1 + \frac{T}{r\lambda}\Big). \tag{53}$$

The last inequality in (53) is part of (8). For completeness: if $n \le T$ and $\phi^k(\boldsymbol{x}_s) \in \mathcal{S}_k$ for all $s \le n$, then the eigenvalues of $\sum_{s\le n} \phi^k(\boldsymbol{x}_s)\phi^k(\boldsymbol{x}_s)^\top$ are nonnegative, at most r_k of them, say $\nu_1, \dots, \nu_{r_k}$, are nonzero, and their sum is its trace, at most $n \le T$. Concavity of the logarithm gives $\log\det(\boldsymbol{I} + \lambda^{-1}\sum_{s\le n} \phi^k(\boldsymbol{x}_s)\phi^k(\boldsymbol{x}_s)^\top) = \sum_{q=1}^{r_k} \log(1+\nu_q/\lambda) \le r_k \log(1+T/(r_k\lambda))$, and $a \mapsto a\log(1+c/a)$ is nondecreasing on $(0,\infty)$ because its derivative $\log(1+c/a) - c/(a+c)$ is nonnegative. The dimension p_k of the features appears nowhere in the bounds below; only r, κ and G_T do.

Finally, $n_\star \ge n_{\mathrm{spec}}$ and, by construction of $n_\star$, either $n_\star \ge 72\bar\beta_T^2/(\kappa\gamma'^2)$ or both $n_\star \ge 288\bar\rho_T^2/(\kappa\gamma'^2)$ and $n_\star \ge 24\lambda B/(\kappa\gamma')$. Hence, for all $n \ge n_\star$,

$$\min\Big\{\bar\beta_T\sqrt{\frac{2}{\kappa n}},\ \bar\rho_T\sqrt{\frac{2}{\kappa n}} + \frac{2\lambda B}{\kappa n}\Big\} \ \le\ \frac{\gamma'}{6}, \tag{54}$$

because the first term is then at most $\sqrt{2\gamma'^2/72} = \gamma'/6$ or the second at most $\sqrt{2\gamma'^2/288} + 2\gamma'/24 = \gamma'/12 + \gamma'/12$.

D.2 Step 1: Anytime Confidence Bounds for Fixed Groups

Let Γ be the set of pairs $(\{i\}, k)$ with $i \in \mathcal{U}$, $k \in [m_o]$, and the pairs $(C_j^{(k)}, k)$ with $k \in [m_o]$, $j \in [m_c^{(k)}]$. With probability at least $1-\delta/4$, simultaneously for all $(A,k) \in \Gamma$, all rounds $0 \le t \le T$ and all contexts $\boldsymbol{x} \in \mathcal{X}$,

$$\big|\hat f_{A,t}^k(\boldsymbol{x}) - \mu_A^k(\boldsymbol{x})\big| \ \le\ \beta(\boldsymbol{M}_{A,t}^k)\,\|\phi^k(\boldsymbol{x})\|_{(\boldsymbol{M}_{A,t}^k)^{-1}} \tag{55}$$

and

$$\big|\hat f_{A,t}^k(\boldsymbol{x}) - \mu_A^k(\boldsymbol{x})\big| \ \le\ \rho(\boldsymbol{M}_{A,t}^k)\,\|\phi^k(\boldsymbol{x})\|_{(\boldsymbol{M}_{A,t}^k)^{-1}} + \lambda B\,\big\|(\boldsymbol{M}_{A,t}^k)^{-1}\phi^k(\boldsymbol{x})\big\|_2. \tag{56}$$

We denote this event by $\mathcal{E}_{\mathrm{conf}}$.

To see this, fix $(A,k) \in \Gamma$ and abbreviate $\boldsymbol{M}_t = \boldsymbol{M}_{A,t}^k$, $\boldsymbol{w} = \boldsymbol{w}_A^k$. For every round $s \le T$, let $\boldsymbol{g}_s := \mathbb{I}\{i_s \in A\}\,\phi^k(\boldsymbol{x}_s)$, and pad the sequences after round T with $\boldsymbol{g}_s := \boldsymbol{0}$ and $\eta_s^k := 0$ for $s > T$. Let $\boldsymbol{\zeta}_t := \sum_{s\le t} \boldsymbol{g}_s\eta_s^k$. Then $\boldsymbol{M}_t = \lambda\boldsymbol{I} + \sum_{s\le t}\boldsymbol{g}_s\boldsymbol{g}_s^\top$ for $0 \le t \le T$. Since all users of A belong to the same cluster, $r_s^k = \phi^k(\boldsymbol{x}_s)^\top\boldsymbol{w} + \eta_s^k$ whenever $i_s \in A$, and hence $\boldsymbol{b}_{A,t}^k = \sum_{s\le t}\boldsymbol{g}_s(\boldsymbol{g}_s^\top\boldsymbol{w} + \eta_s^k) = (\boldsymbol{M}_t - \lambda\boldsymbol{I})\boldsymbol{w} + \boldsymbol{\zeta}_t$. This gives the exact decomposition

$$\hat{\boldsymbol{w}}_{A,t}^k - \boldsymbol{w} = -\lambda\boldsymbol{M}_t^{-1}\boldsymbol{w} + \boldsymbol{M}_t^{-1}\boldsymbol{\zeta}_t. \tag{57}$$

For any $\boldsymbol{x} \in \mathcal{X}$ and $0 \le t \le T$, the Cauchy–Schwarz inequality for the inner product $\langle \boldsymbol{a}, \boldsymbol{c}\rangle_{\boldsymbol{M}_t^{-1}} = \boldsymbol{a}^\top\boldsymbol{M}_t^{-1}\boldsymbol{c}$ yields

$$\big|\hat f_{A,t}^k(\boldsymbol{x}) - \mu_A^k(\boldsymbol{x})\big| = \big|\phi^k(\boldsymbol{x})^\top\boldsymbol{M}_t^{-1}(-\lambda\boldsymbol{w} + \boldsymbol{\zeta}_t)\big| \le \|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}_t^{-1}}\big(\lambda\|\boldsymbol{w}\|_{\boldsymbol{M}_t^{-1}} + \|\boldsymbol{\zeta}_t\|_{\boldsymbol{M}_t^{-1}}\big).$$

Since $\boldsymbol{M}_t \succeq \lambda\boldsymbol{I}$, the bias term satisfies $\lambda\|\boldsymbol{w}\|_{\boldsymbol{M}_t^{-1}} \le \lambda\|\boldsymbol{w}\|_2/\sqrt{\lambda} \le B\sqrt{\lambda}$. For the noise term, we apply the self-normalized bound of Abbasi-Yadkori et al. (2011) with the filtration $(\mathcal{H}_s^+)_{s\ge 0}$. The vector $\boldsymbol{g}_s$ is $\mathcal{H}_{s-1}^+$-measurable, because i_s and $\boldsymbol{x}_s$ are. For $s \le T$, the noise $\eta_s^k = r_s^k - \mu_{i_s}^k(\boldsymbol{x}_s)$ is $\mathcal{H}_s$-measurable, hence $\mathcal{H}_s^+$-measurable, and it is conditionally R-sub-Gaussian given $\mathcal{H}_{s-1}^+$ by (A2); for $s > T$, the padded terms satisfy the same conditions trivially. With the regularizer $\lambda\boldsymbol{I}$, this bound, which holds simultaneously for all $t \ge 0$ of the padded sequence, states in particular that, with probability at least $1-\delta_A$, simultaneously for all $0 \le t \le T$,

$$\|\boldsymbol{\zeta}_t\|_{\boldsymbol{M}_t^{-1}}^2 \le 2R^2\log\Big(\frac{\det(\boldsymbol{M}_t)^{1/2}\det(\lambda\boldsymbol{I})^{-1/2}}{\delta_A}\Big) = R^2\Big(\log\det(\lambda^{-1}\boldsymbol{M}_t) + 2\log\frac{1}{\delta_A}\Big).$$

We choose $\delta_A = \delta/(8um_o)$, so that the right-hand side of the last display is $\rho(\boldsymbol{M}_t)^2$. Combining the three displays gives (55) for the pair (A,k), simultaneously for all $0 \le t \le T$ and all $\boldsymbol{x}$. For (56), we bound the two terms of (57) separately instead: $|\phi^k(\boldsymbol{x})^\top\boldsymbol{M}_t^{-1}\boldsymbol{\zeta}_t| \le \|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}_t^{-1}}\|\boldsymbol{\zeta}_t\|_{\boldsymbol{M}_t^{-1}} \le \rho(\boldsymbol{M}_t)\|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}_t^{-1}}$, and $|\lambda\phi^k(\boldsymbol{x})^\top\boldsymbol{M}_t^{-1}\boldsymbol{w}| \le \lambda\|\boldsymbol{M}_t^{-1}\phi^k(\boldsymbol{x})\|_2\|\boldsymbol{w}\|_2 \le \lambda B\|\boldsymbol{M}_t^{-1}\phi^k(\boldsymbol{x})\|_2$. The random event involved concerns only $\boldsymbol{\zeta}_t$, so no union bound over contexts or rounds is needed. This is the confidence ellipsoid of Abbasi-Yadkori et al. (2011) with parameter-norm bound B. Finally, $|\Gamma| = um_o + m_c \le 2um_o$ because $m_c^{(k)} \le u$ for every k, and a union bound over Γ gives failure probability at most $2um_o \cdot \delta/(8um_o) = \delta/4$.

D.3 Step 2: Spectral Growth on the Effective Subspaces

With probability at least $1-\delta/4$, for all users i, objectives k and rounds $t \le T$ with $T_{i,t} \ge n_{\mathrm{spec}}$,

$$\lambda_{\min}\big(\boldsymbol{M}_{i,t}^{k}|_{\mathcal{S}_k}\big) \;\ge\; \lambda + \frac{\kappa\, T_{i,t}}{2}, \tag{58}$$

and consequently, for every context $\boldsymbol{z}$ with $\phi^k(\boldsymbol{z}) \in \mathcal{S}_k$, in particular for every candidate $\boldsymbol{z} \in \mathcal{X}_{t'}$, $t' \ge 1$,

$$\begin{aligned} \|\phi^k(\boldsymbol{z})\|_{(\boldsymbol{M}_{i,t}^{k})^{-1}} \;&\le\; \Big(\lambda + \frac{\kappa\, T_{i,t}}{2}\Big)^{-1/2} \;\le\; \sqrt{\frac{2}{\kappa\, T_{i,t}}}, \\ \big\|(\boldsymbol{M}_{i,t}^{k})^{-1}\phi^k(\boldsymbol{z})\big\|_2 \;&\le\; \Big(\lambda + \frac{\kappa\, T_{i,t}}{2}\Big)^{-1} \;\le\; \frac{2}{\kappa\, T_{i,t}}. \end{aligned} \tag{59}$$

We denote this event by $\mathcal{E}_{\mathrm{spec}}$.

Selected features dominate $\kappa \boldsymbol{P}_{\mathcal{S}_k}$. By Appendix C.2.2 with $\phi = \phi^k$ (so that $\mathcal{S}_\phi = \mathcal{S}_k$ and $\kappa_K(\phi^k) \ge \kappa$), for every objective k and every round $t \ge 1$, including the extended rounds,

$$\mathbb{E}\big[\phi^k(\boldsymbol{x}_t)\phi^k(\boldsymbol{x}_t)^\top \,\big|\, \mathcal{H}_t^-\big] \;\succeq\; \kappa\, \boldsymbol{P}_{\mathcal{S}_k} \qquad \text{almost surely.} \tag{60}$$

Transfer to the visits of a fixed user. Fix i and k, and write $\boldsymbol{P} := \boldsymbol{P}_{\mathcal{S}_k}$. For $n \ge 1$, let $\tau_{i,n}$ be the round of the n-th arrival of user i in the extended interaction ($\tau_{i,n} < \infty$ almost surely), and let $\tau_{i,0} := 0$. The event $\{\tau_{i,n} = t\}$ is determined by $i_1, \dots, i_t$ and therefore belongs to $\mathcal{H}_t^- \subseteq \mathcal{H}_t$, so $\tau_{i,n}$ is a stopping time for $(\mathcal{H}_t)_{t\ge 0}$. Let $\mathcal{F}_n^{(i)} := \mathcal{H}_{\tau_{i,n}}$ be the stopped σ-field, with $\mathcal{F}_0^{(i)} := \mathcal{H}_0$, as in Appendix C.3.3(a). Since $\tau_{i,n-1} < \tau_{i,n}$, $(\mathcal{F}_n^{(i)})_{n\ge 0}$ is a filtration. Let

$$\boldsymbol{Y}_n := \boldsymbol{P}\,\phi^k(\boldsymbol{x}_{\tau_{i,n}})\phi^k(\boldsymbol{x}_{\tau_{i,n}})^\top \boldsymbol{P}, \qquad n \ge 1.$$

Since $(\boldsymbol{x}_t)_{t\ge 1}$ is adapted to $(\mathcal{H}_t)_{t\ge 1}$, $\boldsymbol{Y}_n$ is $\mathcal{F}_n^{(i)}$-measurable. By construction $\boldsymbol{Y}_n = \boldsymbol{P}\boldsymbol{Y}_n\boldsymbol{P}$, and $\boldsymbol{0} \preceq \boldsymbol{Y}_n \preceq \boldsymbol{P}$, because $\boldsymbol{v}^\top \boldsymbol{Y}_n \boldsymbol{v} = \langle \phi^k(\boldsymbol{x}_{\tau_{i,n}}), \boldsymbol{P}\boldsymbol{v}\rangle^2 \le \|\boldsymbol{P}\boldsymbol{v}\|_2^2 = \boldsymbol{v}^\top \boldsymbol{P}\boldsymbol{v}$ for every $\boldsymbol{v}$ by the Cauchy–Schwarz inequality and $\|\phi^k\|_2 \le 1$. On Ω_0, $\boldsymbol{P}\phi^k(\boldsymbol{x}_{\tau_{i,n}}) = \phi^k(\boldsymbol{x}_{\tau_{i,n}})$, so that $\boldsymbol{Y}_n = \phi^k(\boldsymbol{x}_{\tau_{i,n}})\phi^k(\boldsymbol{x}_{\tau_{i,n}})^\top$.

We show that $\mathbb{E}[\boldsymbol{Y}_n \mid \mathcal{F}_{n-1}^{(i)}] \succeq \kappa \boldsymbol{P}$ almost surely. Let $F \in \mathcal{F}_{n-1}^{(i)}$ and $t \ge 1$. Then $F \cap \{\tau_{i,n} = t\} = (F \cap \{\tau_{i,n-1} \le t-1\}) \cap \{\tau_{i,n} = t\}$. The first set belongs to $\mathcal{H}_{t-1}$ by the defining property of the stopped σ-field $\mathcal{H}_{\tau_{i,n-1}}$ (for $n = 1$, $F \in \mathcal{H}_0 \subseteq \mathcal{H}_{t-1}$), and the second belongs to $\mathcal{H}_t^-$; hence $F \cap \{\tau_{i,n} = t\} \in \mathcal{H}_t^-$. For every fixed $\boldsymbol{v}$, (60) applied to the vector $\boldsymbol{P}\boldsymbol{v}$ gives $(\boldsymbol{P}\boldsymbol{v})^\top \mathbb{E}[\phi^k(\boldsymbol{x}_t)\phi^k(\boldsymbol{x}_t)^\top \mid \mathcal{H}_t^-]\,\boldsymbol{P}\boldsymbol{v} \ge \kappa \|\boldsymbol{P}\boldsymbol{v}\|_2^2$ almost surely. Using $\tau_{i,n} < \infty$ almost surely and $\sum_{t\ge 1} \mathbb{P}(F \cap \{\tau_{i,n} = t\}) = \mathbb{P}(F)$, we obtain

$$\mathbb{E}\big[\mathbb{I}_F\, \boldsymbol{v}^\top \boldsymbol{Y}_n \boldsymbol{v}\big] = \sum_{t\ge 1} \mathbb{E}\Big[\mathbb{I}_{F\cap\{\tau_{i,n}=t\}}\, (\boldsymbol{P}\boldsymbol{v})^\top \mathbb{E}\big[\phi^k(\boldsymbol{x}_t)\phi^k(\boldsymbol{x}_t)^\top \,\big|\, \mathcal{H}_t^-\big]\, \boldsymbol{P}\boldsymbol{v}\Big] \ge \kappa\, \|\boldsymbol{P}\boldsymbol{v}\|_2^2\, \mathbb{P}(F).$$

As $F \in \mathcal{F}_{n-1}^{(i)}$ is arbitrary, $\boldsymbol{v}^\top \mathbb{E}[\boldsymbol{Y}_n \mid \mathcal{F}_{n-1}^{(i)}]\boldsymbol{v} \ge \kappa\, \boldsymbol{v}^\top \boldsymbol{P}\boldsymbol{v}$ almost surely for each fixed $\boldsymbol{v}$. Applying this to a countable dense set of vectors and using continuity in $\boldsymbol{v}$ gives the matrix inequality.

Matrix lower tail. The filtration $(\mathcal{F}_n^{(i)})_{n\ge 0}$ and the matrices $(\boldsymbol{Y}_n)_{n\ge 1}$ therefore satisfy the hypotheses of Appendix C.2.1 with $S = \mathcal{S}_k$, whose dimension is $r_k \ge 1$. That bound gives, for every $n \ge 1$,

$$\mathbb{P}\Big(\lambda_{\min}\Big(\Big(\sum_{m\le n} \boldsymbol{Y}_m\Big)\Big|_{\mathcal{S}_k}\Big) \le \frac{n\kappa}{2}\Big) \;\le\; r_k\, e^{-n\kappa/8}.$$

A union bound over $i \in \mathcal{U}$, $k \in [m_o]$ and $n \in \{n_{\mathrm{spec}}, \dots, T\}$ bounds the total failure probability by $u m_o T\, r\, e^{-n_{\mathrm{spec}}\kappa/8} \le \delta/4$, because $r_k \le r$ and $n_{\mathrm{spec}}\kappa/8 \ge \log(4u m_o r T/\delta)$.

Conclusion. On Ω_0 and on the complement of these failure events, let $t \le T$ with $n := T_{i,t} \ge n_{\mathrm{spec}}$; note that $n \le t \le T$. The rounds $s \le t$ with $i_s = i$ are exactly $\tau_{i,1}, \dots, \tau_{i,n}$, so $\boldsymbol{M}_{i,t}^{k} = \lambda \boldsymbol{I} + \sum_{m\le n} \boldsymbol{Y}_m$ and $\lambda_{\min}(\boldsymbol{M}_{i,t}^{k}|_{\mathcal{S}_k}) = \lambda + \lambda_{\min}((\sum_{m\le n} \boldsymbol{Y}_m)|_{\mathcal{S}_k}) \ge \lambda + \kappa n/2$, which is (58). The first inequality in (59) follows from (48), $\|\phi^k(\boldsymbol{z})\|_2 \le 1$ and $\lambda + \kappa n/2 \ge \kappa n/2$; the second follows in the same way, because $(\boldsymbol{M}_{i,t}^{k})^{-1}$ maps $\mathcal{S}_k$ into itself and its block on $\mathcal{S}_k$ has operator norm at most $1/\lambda_{\min}(\boldsymbol{M}_{i,t}^{k}|_{\mathcal{S}_k})$. On Ω_0, every candidate feature lies in $\mathcal{S}_k$.

D.4 Step 3: No False Deletion

On $\mathcal{E}_{\mathrm{conf}}$, for every objective k, every round $1 \le t \le T$, all users i, ℓ with $j_k(i) = j_k(\ell)$ and every context $\boldsymbol{z} \in \mathcal{X}$,

$$\big|\hat{f}^k_{i,t}(\boldsymbol{z}) - \hat{f}^k_{\ell,t}(\boldsymbol{z})\big| \;\le\; b^k_{i,t}(\boldsymbol{z}) + b^k_{\ell,t}(\boldsymbol{z}).$$

Consequently, the strict test (18) never deletes an edge between two users of the same cluster of objective k, and for every $0 \le t \le T$ the edge set E^k_t contains all such pairs.

Indeed, since $j_k(i) = j_k(\ell)$, we have $\mu^k_i = \mu^k_\ell$. By the triangle inequality and (56) for the groups $\{i\}$ and $\{\ell\}$ at time t, whose right-hand sides are the widths $b^k_{i,t}(\boldsymbol{z})$ and $b^k_{\ell,t}(\boldsymbol{z})$ of (17),

$$\big|\hat{f}^k_{i,t}(\boldsymbol{z}) - \hat{f}^k_{\ell,t}(\boldsymbol{z})\big| \le \big|\hat{f}^k_{i,t}(\boldsymbol{z}) - \mu^k_i(\boldsymbol{z})\big| + \big|\hat{f}^k_{\ell,t}(\boldsymbol{z}) - \mu^k_\ell(\boldsymbol{z})\big| \le b^k_{i,t}(\boldsymbol{z}) + b^k_{\ell,t}(\boldsymbol{z}).$$

Hence the strict inequality in (18) fails for every $\boldsymbol{z} \in \mathcal{X}_t$. The graph G^k_0 is complete, and an edge is removed only in a round in which the test fires for it, so induction over t proves the last statement. No condition on the numbers of visits is needed; the statement holds from the first round on.

D.5 Step 4: Deletion of Cross-Cluster Edges

On $\mathcal{E}_{\mathrm{conf}} \cap \mathcal{E}_{\mathrm{spec}}$, let $t \le T$, $k \in [m_o]$ and $\ell \ne i_t$ be such that $(i_t, \ell) \in E^k_{t-1}$, $j_k(i_t) \ne j_k(\ell)$, $T_{i_t,t} \ge n_\star$ and $T_{\ell,t} \ge n_\star$. If $\mathcal{X}_t$ contains a context $\boldsymbol{z}$ with $|\mu^k_{C_{j_k(i_t)}}(\boldsymbol{z}) - \mu^k_{C_{j_k(\ell)}}(\boldsymbol{z})| \ge \gamma'$, then the edge (i_t, ℓ) is deleted in round t.

To see this, write $i = i_t$. Since $\boldsymbol{z} \in \mathcal{X}_t$, $\phi^k(\boldsymbol{z}) \in \mathcal{S}_k$. Because $T_{i,t} \ge n_\star \ge n_{\mathrm{spec}}$, the bounds (52), (59) and (54), together with $\lambda\|\boldsymbol{M}^{-1}\boldsymbol{v}\|_2 \le \sqrt{\lambda}\|\boldsymbol{v}\|_{\boldsymbol{M}^{-1}}$ for $\boldsymbol{M} \succeq \lambda \boldsymbol{I}$, give

$$\begin{aligned} b^k_{i,t}(\boldsymbol{z}) &= \rho(\boldsymbol{M}^k_{i,t})\,\|\phi^k(\boldsymbol{z})\|_{(\boldsymbol{M}^k_{i,t})^{-1}} + \lambda B\big\|(\boldsymbol{M}^k_{i,t})^{-1}\phi^k(\boldsymbol{z})\big\|_2 \\ &\le \min\Big\{\bar{\beta}_T\sqrt{\frac{2}{\kappa\, T_{i,t}}},\; \bar{\rho}_T\sqrt{\frac{2}{\kappa\, T_{i,t}}} + \frac{2\lambda B}{\kappa\, T_{i,t}}\Big\} \le \frac{\gamma'}{6}, \end{aligned}$$

and in the same way $b^k_{\ell,t}(\boldsymbol{z}) \le \gamma'/6$. By (56) for $\{i\}$ and $\{\ell\}$,

$$\big|\hat{f}^k_{i,t}(\boldsymbol{z}) - \hat{f}^k_{\ell,t}(\boldsymbol{z})\big| \ge \big|\mu^k_i(\boldsymbol{z}) - \mu^k_\ell(\boldsymbol{z})\big| - \big|\hat{f}^k_{i,t}(\boldsymbol{z}) - \mu^k_i(\boldsymbol{z})\big| - \big|\hat{f}^k_{\ell,t}(\boldsymbol{z}) - \mu^k_\ell(\boldsymbol{z})\big| \ge \gamma' - b^k_{i,t}(\boldsymbol{z}) - b^k_{\ell,t}(\boldsymbol{z}).$$

Therefore

$$\big|\hat{f}^k_{i,t}(\boldsymbol{z}) - \hat{f}^k_{\ell,t}(\boldsymbol{z})\big| - b^k_{i,t}(\boldsymbol{z}) - b^k_{\ell,t}(\boldsymbol{z}) \;\ge\; \gamma' - 2\big(b^k_{i,t}(\boldsymbol{z}) + b^k_{\ell,t}(\boldsymbol{z})\big) \;\ge\; \gamma' - \frac{2\gamma'}{3} = \frac{\gamma'}{3} > 0,$$

so the test (18) fires for the pair (i_t, ℓ), which is tested in round t because $(i_t, \ell) \in E^k_{t-1}$.

D.6 Step 5: Arrivals and Covering

The following two events depend only on the exogenous sequence $(i_t, \mathcal{X}_t)_{t \ge 1}$, which is i.i.d. and does not depend on the algorithm. The numbers of arrivals are those of the extended interaction, which coincide with the counters $T_{i,t}$ of Algorithm 2 for $t \le T$.

D.6.1 Arrivals

With probability at least $1 - \delta/4$, every user has arrived at least $n_\star$ times in the first t_{fit} rounds, that is, $T_{i,t_{\mathrm{fit}}} \ge n_\star$ for all $i \in \mathcal{U}$. We denote this event by $\mathcal{E}_{\mathrm{arr}}$.

To see this, fix i. By (A1), the indicators $\mathbb{I}\{i_t = i\}$, $t \ge 1$, are independent Bernoulli variables with mean $1/u$. Hence $N := T_{i,t_{\mathrm{fit}}}$ is binomial with mean $\mu_N = t_{\mathrm{fit}}/u \ge 16\log(4u/\delta) + 4n_\star$. The multiplicative Chernoff bound for the lower tail, $\mathbb{P}(N \le \mu_N/2) \le e^{-\mu_N/8}$, gives

$$\mathbb{P}\big(N \le \mu_N/2\big) \le \exp\big(-2\log(4u/\delta)\big) \le \frac{\delta}{4u},$$

where we used $\log(4u/\delta) \ge 0$. On the complement, $N > \mu_N/2 \ge 2n_\star \ge n_\star$. A union bound over the u users completes the proof.

D.6.2 COVERING

Let $L := T_{\mathrm{clust}} - t_{\mathrm{fit}} = \lceil (u/q_\gamma)\log(4u^2 m_o/\delta)\rceil$. For each objective k and each unordered pair $\{i,\ell\}$ of users with $j_k(i) \neq j_k(\ell)$, designate one endpoint, say i. With probability at least $1-\delta/4$, for every such k and pair, there is a round $t \in \{t_{\mathrm{fit}}+1,\dots,T_{\mathrm{clust}}\}$ with $i_t = i$ in which $\mathcal{X}_t$ contains a context $\boldsymbol{z}$ with $|\mu^k_{C_{j_k(i)}}(\boldsymbol{z}) - \mu^k_{C_{j_k(\ell)}}(\boldsymbol{z})| \geq \gamma'$. We denote this event by $\mathcal{E}_{\mathrm{cov}}$.

To see this, fix k and a pair $\{i,\ell\}$ with $j := j_k(i) \neq j' := j_k(\ell)$, and let $\mathcal{B}_t$ be the event that $i_t = i$ and $\mathcal{X}_t$ contains a context $\boldsymbol{z}$ with $|\mu^k_{C_j}(\boldsymbol{z}) - \mu^k_{C_{j'}}(\boldsymbol{z})| \geq \gamma'$. By (A6), each candidate has this property independently with probability $p \geq \rho_\gamma$, so $\mathcal{X}_t$ contains such a context with probability $1-(1-p)^K \geq q_\gamma$. Since i_t is independent of $\mathcal{X}_t$ and uniform on $\mathcal{U}$, $\mathbb{P}(\mathcal{B}_t) \geq q_\gamma/u$. The events $\mathcal{B}_t$ are independent across rounds, because each is a function of $(i_t, \mathcal{X}_t)$. Hence

$$\mathbb{P}\Big(\bigcap_{t=t_{\mathrm{fit}}+1}^{T_{\mathrm{clust}}} \mathcal{B}_t^c\Big) \leq \Big(1-\frac{q_\gamma}{u}\Big)^L \leq \exp\Big(-\frac{Lq_\gamma}{u}\Big) \leq \frac{\delta}{4u^2 m_o}.$$

There are at most $u(u-1)/2 \leq u^2$ pairs per objective and m_o objectives, so a union bound gives failure probability at most $\delta/4$.

D.7 STEP 6: CLUSTER RECOVERY AND TIMING

Let $\mathcal{E} := \Omega_0 \cap \mathcal{E}_{\mathrm{conf}} \cap \mathcal{E}_{\mathrm{spec}} \cap \mathcal{E}_{\mathrm{arr}} \cap \mathcal{E}_{\mathrm{cov}}$. By Appendices D.2, D.3, D.6.1 and D.6.2, $\mathbb{P}(\mathcal{E}) \geq 1-\delta$.

On $\mathcal{E}$: (i) for every k and every round $t \leq T$, E^k_t contains every pair of users that belong to the same cluster of objective k; (ii) if $T \geq T_{\mathrm{clust}}$, then for every k and every $t \in \{T_{\mathrm{clust}},\dots,T\}$, the edge set E^k_t consists exactly of the pairs of users in the same cluster of objective k. In particular, the closed neighbourhood of every user in G^k_t is its true cluster.

Conclusion (i) is shown in Appendix D.4. For (ii), fix k and a pair $\{i,\ell\}$ with $j_k(i) \neq j_k(\ell)$ and designated endpoint i. By $\mathcal{E}_{\mathrm{cov}}$, there is a round $t \in \{t_{\mathrm{fit}}+1,\dots,T_{\mathrm{clust}}\} \subseteq [T]$ with $i_t = i$ in which $\mathcal{X}_t$ contains a context separating $C^{(k)}_{j_k(i)}$ and $C^{(k)}_{j_k(\ell)}$ by at least γ'. If $(i,\ell) \notin E^k_{t-1}$, the edge was deleted before round t. Otherwise, the counters are nondecreasing and $t > t_{\mathrm{fit}}$, so $\mathcal{E}_{\mathrm{arr}}$ gives $T_{i,t} \geq T_{i,t_{\mathrm{fit}}} \geq n_\star$ and $T_{\ell,t} \geq T_{\ell,t_{\mathrm{fit}}} \geq n_\star$. Appendix D.5 then shows that the edge is deleted in round t. Since edges are never added, $(i,\ell) \notin E^k_{T_{\mathrm{clust}}}$. Thus $E^k_{T_{\mathrm{clust}}}$ contains no pair from different clusters and, by (i), contains all pairs from the same cluster. For $t \geq T_{\mathrm{clust}}$, $E^k_t \subseteq E^k_{T_{\mathrm{clust}}}$ because edges are only deleted, and E^k_t still contains all same-cluster pairs by (i); hence $E^k_t = E^k_{T_{\mathrm{clust}}}$. A graph whose edges are exactly the same-cluster pairs is a disjoint union of cliques on the true clusters, so the closed neighbourhood of every user is its true cluster.

Timing. In round t, the algorithm uses the graph G^k_{t-1} of the previous round. The graphs are recovered at the end of round T_{clust}, so the correct cluster statistics are used from round $T_{\mathrm{clust}}+1$ on. By statement (ii) above and (51), on $\mathcal{E}$ and for every $t \in \{T_{\mathrm{clust}}+1,\dots,T\}$ and $k \in [m_o]$,

$$\widehat{C}^k_t = C^{(k)}_{j_k(i_t)}, \qquad \bar{\boldsymbol{M}}^k_t = \boldsymbol{M}^k_{C^{(k)}_{j_k(i_t)},t-1}, \qquad \bar{\boldsymbol{w}}^k_t = \hat{\boldsymbol{w}}^k_{C^{(k)}_{j_k(i_t)},t-1}. \tag{61}$$

Conclusions (i) and (ii) of Section 4.2 follow from the two statements above and (61).

D.8 STEP 7: REGRET DECOMPOSITION

For $t \in [T]$ and $k \in [m_o]$, let $s^k_t := \|\phi^k(\boldsymbol{x}_t)\|_{(\bar{\boldsymbol{M}}^k_t)^{-1}}$ be the width of the selected context in the statistics used by the algorithm.

D.8.1 VALID UCBS AFTER RECOVERY

On $\mathcal{E}$, if $T > T_{\mathrm{clust}}$, then for all $t \in \{T_{\mathrm{clust}}+1,\dots,T\}$, $k \in [m_o]$ and $\boldsymbol{x} \in \mathcal{X}$,

$$\mu^k_{i_t}(\boldsymbol{x}) \;\leq\; U^k_t(\boldsymbol{x}) \;\leq\; \mu^k_{i_t}(\boldsymbol{x}) + 2\bar{\beta}_T\,\|\phi^k(\boldsymbol{x})\|_{(\bar{\boldsymbol{M}}^k_t)^{-1}}.$$

To see this, let $C = C^{(k)}_{j_k(i_t)}$. By (61), $U^k_t(\boldsymbol{x}) = \hat{f}^k_{C,t-1}(\boldsymbol{x}) + \beta(\boldsymbol{M}^k_{C,t-1})\|\phi^k(\boldsymbol{x})\|_{(\boldsymbol{M}^k_{C,t-1})^{-1}}$, and $\mu^k_{i_t} = \mu^k_C$. The confidence bound (55) for the fixed group (C,k) at time $t-1$ gives $|\hat{f}^k_{C,t-1}(\boldsymbol{x}) - \mu^k_C(\boldsymbol{x})| \le \beta(\boldsymbol{M}^k_{C,t-1})\|\phi^k(\boldsymbol{x})\|_{(\boldsymbol{M}^k_{C,t-1})^{-1}}$. This yields the lower bound and $U^k_t(\boldsymbol{x}) \le \mu^k_C(\boldsymbol{x}) + 2\beta(\boldsymbol{M}^k_{C,t-1})\|\phi^k(\boldsymbol{x})\|_{(\boldsymbol{M}^k_{C,t-1})^{-1}}$. Finally, $\beta(\boldsymbol{M}^k_{C,t-1}) \le \bar{\beta}_T$ by (52).

D.8.2 Pareto step

On $\mathcal{E}$, for every $t \in \{T_{\text{clust}}+1, \dots, T\}$, $\Delta_{i_t}(\boldsymbol{x}_t) \le 2\bar{\beta}_T \max_{k\in[m_o]} s^k_t$.

To see this, fix $\boldsymbol{x}' \in \mathcal{X}_t$. Since $\boldsymbol{x}_t \in \tilde{\mathcal{O}}_t$, $\boldsymbol{U}_t(\boldsymbol{x}') \not\succ \boldsymbol{U}_t(\boldsymbol{x}_t)$. Hence there is an objective $k^* = k^*(\boldsymbol{x}')$ with $U^{k^*}_t(\boldsymbol{x}_t) \ge U^{k^*}_t(\boldsymbol{x}')$: otherwise $U^k_t(\boldsymbol{x}') > U^k_t(\boldsymbol{x}_t)$ for every k, and $\boldsymbol{U}_t(\boldsymbol{x}')$ would dominate $\boldsymbol{U}_t(\boldsymbol{x}_t)$. By Appendix D.8.1, applied at $\boldsymbol{x}'$ and at $\boldsymbol{x}_t$,

$$\begin{aligned}\min_{k\in[m_o]}\big[\mu^k_{i_t}(\boldsymbol{x}') - \mu^k_{i_t}(\boldsymbol{x}_t)\big] &\le \mu^{k^*}_{i_t}(\boldsymbol{x}') - \mu^{k^*}_{i_t}(\boldsymbol{x}_t) \le U^{k^*}_t(\boldsymbol{x}') - \mu^{k^*}_{i_t}(\boldsymbol{x}_t)\\ &\le U^{k^*}_t(\boldsymbol{x}_t) - \mu^{k^*}_{i_t}(\boldsymbol{x}_t) \le 2\bar{\beta}_T\, s^{k^*}_t.\end{aligned}$$

The right-hand side is nonnegative, so $\min_k[\mu^k_{i_t}(\boldsymbol{x}') - \mu^k_{i_t}(\boldsymbol{x}_t)]_+ = [\min_k(\mu^k_{i_t}(\boldsymbol{x}') - \mu^k_{i_t}(\boldsymbol{x}_t))]_+ \le 2\bar{\beta}_T \max_k s^k_t$. Taking the maximum over $\boldsymbol{x}' \in \mathcal{X}_t$ and using the equivalent form (3) of (2) gives the bound.

D.8.3 Elliptical potential per true cluster

Fix k and j, and let $t_1 < \cdots < t_N$ be the rounds $t \le T$ with $i_t \in C^{(k)}_j$. Then

$$\sum_{n=1}^{N} \|\phi^k(\boldsymbol{x}_{t_n})\|^2_{(\boldsymbol{M}^k_{C^{(k)}_j,t_n-1})^{-1}} \;\le\; 2\, l^k_{j,T}.$$

To see this, write $C = C^{(k)}_j$, $\boldsymbol{\phi}_n = \phi^k(\boldsymbol{x}_{t_n})$ and $\boldsymbol{V}_n = \boldsymbol{M}^k_{C,t_n}$ for $n \ge 1$, with $\boldsymbol{V}_0 = \lambda \boldsymbol{I}$. Between two consecutive rounds of C, the matrix $\boldsymbol{M}^k_{C,\cdot}$ does not change, so $\boldsymbol{M}^k_{C,t_n-1} = \boldsymbol{V}_{n-1}$ and $\boldsymbol{V}_n = \boldsymbol{V}_{n-1} + \boldsymbol{\phi}_n\boldsymbol{\phi}_n^\top$. Let $a_n := \|\boldsymbol{\phi}_n\|^2_{\boldsymbol{V}_{n-1}^{-1}}$. By the rank-one determinant identity, $\det \boldsymbol{V}_n = \det \boldsymbol{V}_{n-1}\,(1+a_n)$, and hence $\sum_{n=1}^N \log(1+a_n) = \log\det(\lambda^{-1}\boldsymbol{V}_N) = l^k_{j,T}$, since $\boldsymbol{V}_N = \boldsymbol{M}^k_{C,T}$. Because $\boldsymbol{V}_{n-1} \succeq \lambda\boldsymbol{I}$, $\lambda \ge 1$ and $\|\boldsymbol{\phi}_n\|_2 \le 1$, we have $a_n \le 1$. On $[0,1]$, the concave function $a \mapsto \log(1+a)$ lies above its chord, so $\log(1+a) \ge a\log 2 \ge a/2$. Therefore $\sum_n a_n \le 2\sum_n \log(1+a_n) = 2l^k_{j,T}$ (cf. Abbasi-Yadkori et al., 2011).

D.8.4 Part (iii) of the guarantee

We work on $\mathcal{E}$. Since $|\mu^k_i| \le 1$, every PSG is at most 2, so $R_T \le 2T$ always. If $T \le T_{\text{clust}}$, then $R_T \le 2T \le 2T_{\text{clust}}$, and (iii) holds. Let $T > T_{\text{clust}}$. The regret of the rounds $t \le T_{\text{clust}}$ is at most $2T_{\text{clust}}$. By Appendix D.8.2 and the Cauchy–Schwarz inequality over the at most T remaining rounds,

$$\begin{aligned}\sum_{t=T_{\text{clust}}+1}^{T} \Delta_{i_t}(\boldsymbol{x}_t) &\le 2\bar{\beta}_T \sum_{t=T_{\text{clust}}+1}^{T} \max_k s^k_t \le 2\bar{\beta}_T \sqrt{T \sum_{t=T_{\text{clust}}+1}^{T} \max_k (s^k_t)^2}\\ &\le 2\bar{\beta}_T \sqrt{T \sum_{t=T_{\text{clust}}+1}^{T} \sum_{k=1}^{m_o} (s^k_t)^2}.\end{aligned}$$

For $t > T_{\text{clust}}$, (61) gives $s^k_t = \|\phi^k(\boldsymbol{x}_t)\|_{(\boldsymbol{M}^k_{C,t-1})^{-1}}$ with $C = C^{(k)}_{j_k(i_t)}$. This is the term of round t in the sum of Appendix D.8.3 for the pair $(k, j_k(i_t))$. Distinct pairs (t,k) correspond to distinct terms, and the terms of the rounds $t \le T_{\text{clust}}$ are nonnegative. The cluster matrices $\boldsymbol{M}^k_{C,t-1}$ also contain the samples collected before recovery, which can only decrease the widths, and these samples are

accounted for by keeping all rounds of C in Appendix D.8.3. Hence

$$\sum_{t=T_{\mathrm{clust}}+1}^{T}\sum_{k=1}^{m_o}(s_t^k)^2 \;\le\; \sum_{k=1}^{m_o}\sum_{j=1}^{m_c^{(k)}} 2\,l_{j,T}^k \;\le\; 2m_cG_T, \tag{62}$$

where the last step uses (53). Combining the three bounds,

$$R_T \;\le\; \min\Big\{2T,\; 2T_{\mathrm{clust}}+2\bar\beta_T\sqrt{2T\textstyle\sum_{k=1}^{m_o}\sum_{j=1}^{m_c^{(k)}} l_{j,T}^k}\Big\} \;\le\; \min\Big\{2T,\; 2T_{\mathrm{clust}}+2\bar\beta_T\sqrt{2m_cTG_T}\Big\}. \tag{63}$$

The second inequality is conclusion (iii). The first is the finer form mentioned in Section 4.2, in which m_cG_T is replaced by the sum of the log-determinants $l_{j,T}^k$ of the contexts actually selected for the true clusters. This completes the proof of the guarantee of Section 4.2.

D.9 Sublinearity

Since $2\bar\beta_T\sqrt{2m_cTG_T} = 2\sqrt{2}\,\sqrt{T\cdot m_c\bar\beta_T^2G_T}$, the bound (63) implies $R_T = o(T)$ whenever

$$T_{\mathrm{clust}} = o(T) \qquad \text{and} \qquad m_c\,\bar\beta_T^2\,G_T = o(T).$$

These conditions are sufficient, not necessary. They hold for fixed finite-dimensional features, that is, when the maps ϕ^k, the distribution $\mathcal{D}$, K, u, m_o, δ, λ, R and the cluster parameters do not change with T, so that r, κ, γ', q_γ and B are constants. Then $G_T \le r\log(1+T/(r\lambda)) = O(\log T)$ by (53), and $\bar\beta_T^2 \le 2B^2\lambda + 2R^2(G_T + 2\log(8um_o/\delta)) = O(\log T)$. Consequently n_{spec}, $n_\star$, t_{fit} and T_{clust} are all $O(\log T)$, and the right-hand side of (63) is $O(\sqrt{T}\log T)$ with instance-dependent constants. If the network width, and hence the feature map, changes with T, then $\boldsymbol{\Sigma}_\phi^k$, $\mathcal{S}_k$, r, λ_ϕ, κ and G_T change with T as well, as do the realizing parameters $\boldsymbol{w}_j^k$ and their bound B, and even the validity of (A5) itself; the fixed-distribution argument above then no longer applies. (The separation parameters γ', ρ_γ and q_γ are defined through the mean rewards and $\mathcal{D}$ alone and do not depend on the features.) The bound (63) then has to be evaluated for the features actually used at horizon T.

D.10 Kernel Form of the Estimators

Fix an objective k and a finite set of rounds S with contexts $(\boldsymbol{x}_s)_{s\in S}$ and rewards $\boldsymbol{r}_S = (r_s^k)_{s\in S}$. Let $\boldsymbol{\Phi}_S \in \mathbb{R}^{|S|\times p_k}$ have rows $\phi^k(\boldsymbol{x}_s)^\top$, and let $\boldsymbol{M}_S = \lambda\boldsymbol{I}_{p_k} + \boldsymbol{\Phi}_S^\top\boldsymbol{\Phi}_S$, $\hat{\boldsymbol{w}}_S = \boldsymbol{M}_S^{-1}\boldsymbol{\Phi}_S^\top\boldsymbol{r}_S$, $\hat f_S(\boldsymbol{x}) = \phi^k(\boldsymbol{x})^\top\hat{\boldsymbol{w}}_S$ and $s_S(\boldsymbol{x}) = \|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}_S^{-1}}$. Let $\boldsymbol{K}_S = \boldsymbol{\Phi}_S\boldsymbol{\Phi}_S^\top = (\mathcal{K}^k(\boldsymbol{x}_s,\boldsymbol{x}_{s'}))_{s,s'\in S}$ and $\boldsymbol{k}_S(\boldsymbol{x}) = \boldsymbol{\Phi}_S\phi^k(\boldsymbol{x}) = (\mathcal{K}^k(\boldsymbol{x}_s,\boldsymbol{x}))_{s\in S}$. Then, for every $\boldsymbol{x}\in\mathcal{X}$,

$$\hat f_S(\boldsymbol{x}) = \boldsymbol{k}_S(\boldsymbol{x})^\top(\boldsymbol{K}_S+\lambda\boldsymbol{I}_{|S|})^{-1}\boldsymbol{r}_S, \tag{64}$$

$$s_S^2(\boldsymbol{x}) = \frac{1}{\lambda}\Big[\mathcal{K}^k(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{k}_S(\boldsymbol{x})^\top(\boldsymbol{K}_S+\lambda\boldsymbol{I}_{|S|})^{-1}\boldsymbol{k}_S(\boldsymbol{x})\Big], \tag{65}$$

$$\log\det(\lambda^{-1}\boldsymbol{M}_S) = \log\det\big(\boldsymbol{I}_{|S|}+\lambda^{-1}\boldsymbol{K}_S\big), \tag{66}$$

$$\big\|\boldsymbol{M}_S^{-1}\phi^k(\boldsymbol{x})\big\|_2^2 = \frac{1}{\lambda}\Big[s_S^2(\boldsymbol{x}) - \big\|(\boldsymbol{K}_S+\lambda\boldsymbol{I}_{|S|})^{-1}\boldsymbol{k}_S(\boldsymbol{x})\big\|_2^2\Big]. \tag{67}$$

In particular, taking for S the rounds of user i up to round t gives $\boldsymbol{M}_S = \boldsymbol{M}_{i,t}^k$ and $\hat f_S = \hat f_{i,t}^k$. Taking for S the rounds $s\le t-1$ with $i_s\in\widehat{C}_t^k$ gives $\boldsymbol{M}_S = \bar{\boldsymbol{M}}_t^k$, with the regularizer counted once, and $\hat{\boldsymbol{w}}_S = \bar{\boldsymbol{w}}_t^k$. This proves the kernel form (19) and the kernel form of the bias part of the deletion width used in Appendix A.

To verify these identities, write $\boldsymbol{\Phi} = \boldsymbol{\Phi}_S$ and $n = |S|$. The push-through identity

$$(\lambda\boldsymbol{I}_{p_k}+\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}\boldsymbol{\Phi}^\top = \boldsymbol{\Phi}^\top(\lambda\boldsymbol{I}_n+\boldsymbol{\Phi}\boldsymbol{\Phi}^\top)^{-1} \tag{68}$$

follows from $\boldsymbol{\Phi}^\top(\lambda\boldsymbol{I}_n+\boldsymbol{\Phi}\boldsymbol{\Phi}^\top) = (\lambda\boldsymbol{I}_{p_k}+\boldsymbol{\Phi}^\top\boldsymbol{\Phi})\boldsymbol{\Phi}^\top$ by multiplying on the left by $(\lambda\boldsymbol{I}_{p_k}+\boldsymbol{\Phi}^\top\boldsymbol{\Phi})^{-1}$ and on the right by $(\lambda\boldsymbol{I}_n+\boldsymbol{\Phi}\boldsymbol{\Phi}^\top)^{-1}$; both matrices are positive definite because $\lambda>0$. Hence

$\hat{f}_S(\boldsymbol{x}) = \phi^k(\boldsymbol{x})^\top \boldsymbol{\Phi}^\top (\boldsymbol{K}_S + \lambda \boldsymbol{I}_n)^{-1} \boldsymbol{r}_S = \boldsymbol{k}_S(\boldsymbol{x})^\top (\boldsymbol{K}_S + \lambda \boldsymbol{I}_n)^{-1} \boldsymbol{r}_S$, which is (64). For (65), we show that

$$\boldsymbol{M}_S^{-1} = \frac{1}{\lambda}\Big[\boldsymbol{I}_{p_k} - \boldsymbol{\Phi}^\top (\lambda \boldsymbol{I}_n + \boldsymbol{\Phi}\boldsymbol{\Phi}^\top)^{-1} \boldsymbol{\Phi}\Big].$$

Indeed, multiplying the right-hand side on the left by $\boldsymbol{M}_S = \lambda \boldsymbol{I}_{p_k} + \boldsymbol{\Phi}^\top \boldsymbol{\Phi}$ and using $(\lambda \boldsymbol{I}_{p_k} + \boldsymbol{\Phi}^\top \boldsymbol{\Phi}) \boldsymbol{\Phi}^\top = \boldsymbol{\Phi}^\top (\lambda \boldsymbol{I}_n + \boldsymbol{\Phi}\boldsymbol{\Phi}^\top)$ gives $\lambda^{-1}[\lambda \boldsymbol{I}_{p_k} + \boldsymbol{\Phi}^\top \boldsymbol{\Phi} - \boldsymbol{\Phi}^\top \boldsymbol{\Phi}] = \boldsymbol{I}_{p_k}$. Therefore

$$s_S^2(\boldsymbol{x}) = \phi^k(\boldsymbol{x})^\top \boldsymbol{M}_S^{-1} \phi^k(\boldsymbol{x}) = \frac{1}{\lambda}\Big[\|\phi^k(\boldsymbol{x})\|_2^2 - (\boldsymbol{\Phi}\phi^k(\boldsymbol{x}))^\top (\lambda \boldsymbol{I}_n + \boldsymbol{\Phi}\boldsymbol{\Phi}^\top)^{-1} \boldsymbol{\Phi}\phi^k(\boldsymbol{x})\Big],$$

which is (65), because $\|\phi^k(\boldsymbol{x})\|_2^2 = \mathcal{K}^k(\boldsymbol{x}, \boldsymbol{x})$ and $\boldsymbol{\Phi}\phi^k(\boldsymbol{x}) = \boldsymbol{k}_S(\boldsymbol{x})$. The factor $1/\lambda$ is essential: without it, the kernel expression equals $\lambda s_S^2(\boldsymbol{x})$, so every width would be inflated by $\sqrt{\lambda}$. Next, (66) is Sylvester's determinant identity $\det(\boldsymbol{I}_{p_k} + \lambda^{-1} \boldsymbol{\Phi}^\top \boldsymbol{\Phi}) = \det(\boldsymbol{I}_n + \lambda^{-1} \boldsymbol{\Phi}\boldsymbol{\Phi}^\top)$. For (67), write $\boldsymbol{A} = (\lambda \boldsymbol{I}_n + \boldsymbol{\Phi}\boldsymbol{\Phi}^\top)^{-1}$ and $\boldsymbol{k} = \boldsymbol{k}_S(\boldsymbol{x})$; the expression for $\boldsymbol{M}_S^{-1}$ gives $\boldsymbol{M}_S^{-1}\phi^k(\boldsymbol{x}) = \lambda^{-1}(\phi^k(\boldsymbol{x}) - \boldsymbol{\Phi}^\top \boldsymbol{A}\boldsymbol{k})$, hence $\|\boldsymbol{M}_S^{-1}\phi^k(\boldsymbol{x})\|_2^2 = \lambda^{-2}\big(\mathcal{K}^k(\boldsymbol{x}, \boldsymbol{x}) - 2\boldsymbol{k}^\top \boldsymbol{A}\boldsymbol{k} + \boldsymbol{k}^\top \boldsymbol{A}\boldsymbol{\Phi}\boldsymbol{\Phi}^\top \boldsymbol{A}\boldsymbol{k}\big)$, and $\boldsymbol{\Phi}\boldsymbol{\Phi}^\top = \boldsymbol{A}^{-1} - \lambda \boldsymbol{I}_n$ turns the last term into $\boldsymbol{k}^\top \boldsymbol{A}\boldsymbol{k} - \lambda \|\boldsymbol{A}\boldsymbol{k}\|_2^2$; with (65) this is (67). For the two particular cases, the user statistics (15) are, by construction, the ridge statistics of the rounds of user i. For the cluster statistics, $\bar{\boldsymbol{M}}_t^k - \lambda \boldsymbol{I} = \sum_{i \in \widehat{C}_t^k} (\boldsymbol{M}_{i,t-1}^k - \lambda \boldsymbol{I}) = \boldsymbol{\Phi}_S^\top \boldsymbol{\Phi}_S$ and $\bar{\boldsymbol{b}}_t^k = \boldsymbol{\Phi}_S^\top \boldsymbol{r}_S$ when S is the set of rounds $s \le t-1$ with $i_s \in \widehat{C}_t^k$.

D.11 Exploration, Misspecification and Inexact Solvers

D.11.1 Forced exploration: a different algorithm that needs n_0

The identification cost in Section 4.2 is governed by κ. The general lower bound (49) on κ decays exponentially in K, and freezing the features does not remove this worst case, because anti-concentration of the contexts need not transfer to the features. Consider the following *different* algorithm, COMONB-Frozen-Explore, which takes an exploration length n_0 as an additional input. In every round t with $T_{i_t,t-1} < n_0$, that is, during the first n_0 visits of the arriving user, it selects $\boldsymbol{x}_t$ uniformly at random among the K candidates, independently of everything else, instead of from $\tilde{\mathcal{O}}_t$. All other steps, including the statistics and the deletion test, are those of Algorithm 2. In an exploration round, the candidates are i.i.d. from $\mathcal{D}$ and the selected index is uniform and independent of them and of $\mathcal{H}_t^-$; hence, on the $\mathcal{H}_t^-$-measurable event that round t is an exploration round, the conditional law of $\boldsymbol{x}_t$ given $\mathcal{H}_t^-$ is $\mathcal{D}$, and $\mathbb{E}[\phi^k(\boldsymbol{x}_t)\phi^k(\boldsymbol{x}_t)^\top \mid \mathcal{H}_t^-] = \boldsymbol{\Sigma}_\phi^k \succeq \lambda_\phi \boldsymbol{P}_{\mathcal{S}_k}$. The same holds in the extended rounds $t > T$, in which $\boldsymbol{x}_t$ is uniform on $\mathcal{X}_t$ by construction. Here $\boldsymbol{\Sigma}_\phi^k \succeq \lambda_\phi \boldsymbol{P}_{\mathcal{S}_k}$ because $\boldsymbol{v}^\top \boldsymbol{\Sigma}_\phi^k \boldsymbol{v} = (\boldsymbol{P}_{\mathcal{S}_k}\boldsymbol{v})^\top \boldsymbol{\Sigma}_\phi^k \boldsymbol{P}_{\mathcal{S}_k}\boldsymbol{v} \ge \lambda_{\min}^+(\boldsymbol{\Sigma}_\phi^k)\|\boldsymbol{P}_{\mathcal{S}_k}\boldsymbol{v}\|_2^2 \ge \lambda_\phi\, \boldsymbol{v}^\top \boldsymbol{P}_{\mathcal{S}_k}\boldsymbol{v}$ for every $\boldsymbol{v}$. Fix i and k, and let $\tau_{i,m}$, $\mathcal{F}_m^{(i)}$ and $\boldsymbol{Y}_m$ be as in Appendix D.3. For $m \le n_0$, the round $\tau_{i,m}$ is an exploration round or an extended round, because $T_{i,\tau_{i,m}-1} = m - 1 < n_0$; the stopping-time argument of Appendix D.3, with this bound in place of (60), gives $\mathbb{E}[\boldsymbol{Y}_m \mid \mathcal{F}_{m-1}^{(i)}] \succeq \lambda_\phi \boldsymbol{P}_{\mathcal{S}_k}$. For $m > n_0$ and $\tau_{i,m} \le T$, the round $\tau_{i,m}$ is a Pareto-UCB round, for which only the bound with κ is available, so the hypothesis of Appendix C.2.1 with λ_ϕ in place of κ need not hold for the whole sequence $(\boldsymbol{Y}_m)_{m \ge 1}$. We therefore apply Appendix C.2.1, with $S = \mathcal{S}_k$, the filtration $(\mathcal{F}_m^{(i)})_{m \ge 0}$ and λ_ϕ in place of κ, to the sequence

$$\boldsymbol{Y}'_m := \boldsymbol{Y}_m \quad (m \le n_0), \qquad \boldsymbol{Y}'_m := \lambda_\phi \boldsymbol{P}_{\mathcal{S}_k} \quad (m > n_0).$$

This sequence is adapted, satisfies $\boldsymbol{Y}'_m = \boldsymbol{P}_{\mathcal{S}_k} \boldsymbol{Y}'_m \boldsymbol{P}_{\mathcal{S}_k}$ and $\boldsymbol{0} \preceq \boldsymbol{Y}'_m \preceq \boldsymbol{P}_{\mathcal{S}_k}$ (for $m > n_0$ because $\lambda_\phi \le 1$, as shown just before (49)), and $\mathbb{E}[\boldsymbol{Y}'_m \mid \mathcal{F}_{m-1}^{(i)}] \succeq \lambda_\phi \boldsymbol{P}_{\mathcal{S}_k}$ for every $m \ge 1$. Since $\sum_{m \le n_0} \boldsymbol{Y}'_m = \sum_{m \le n_0} \boldsymbol{Y}_m$, the matrix lower tail of Appendix C.2.1 at index n_0 gives $\mathbb{P}\big(\lambda_{\min}((\sum_{m \le n_0} \boldsymbol{Y}_m)|_{\mathcal{S}_k}) \le n_0 \lambda_\phi / 2\big) \le r_k e^{-n_0 \lambda_\phi / 8}$. Suppose that

$$n_0 \;\ge\; \max\Big\{ \frac{8}{\lambda_\phi} \log \frac{4um_o r}{\delta},\; \min\Big\{ \frac{72\,\bar{\beta}_T^2}{\lambda_\phi\, \gamma'^2},\; \max\Big\{ \frac{288\,\bar{\rho}_T^2}{\lambda_\phi\, \gamma'^2},\; \frac{24\lambda B}{\lambda_\phi\, \gamma'} \Big\} \Big\} \Big\}.$$

Then $r_k e^{-n_0 \lambda_\phi / 8} \le r \cdot \delta / (4um_o r)$, and a union bound over the u users and m_o objectives gives failure probability at most $\delta/4$; the complement of these failure events replaces $\mathcal{E}_{\mathrm{spec}}$. On this complement and on Ω_0, if $t \le T$ and $T_{i,t} \ge n_0$, then $\boldsymbol{M}_{i,t}^k \succeq \lambda \boldsymbol{I} + \sum_{m \le n_0} \boldsymbol{Y}_m$, because the matrices $\boldsymbol{M}_{i,t}^k$ are nondecreasing in t; hence $\lambda_{\min}(\boldsymbol{M}_{i,t}^k|_{\mathcal{S}_k}) \ge \lambda + \lambda_\phi n_0 / 2$. By (48) and the argument of (59), every

deletion width at a candidate is then at most $\min\{\bar{\beta}_T\sqrt{2/(\lambda_\phi n_0)},\ \bar{\rho}_T\sqrt{2/(\lambda_\phi n_0)}+2\lambda B/(\lambda_\phi n_0)\} \le \gamma'/6$. The statement of Appendix D.5 therefore holds with $n_\star$ replaced by n_0. The statements of Appendices D.2 and D.4 hold for any selection rule, and those of Appendices D.6.1, D.6.2 and D.7 are unchanged with $n_\star$ replaced by n_0. On the arrival event, every user has completed its exploration phase by round t_{fit}, so all rounds after T_{clust} are Pareto-UCB rounds, which is what Appendix D.8.2 requires. Consequently, conclusions (i)–(iii) of Section 4.2 hold for COMONB-Frozen-Explore with $n_\star$ replaced by n_0 in t_{fit} and T_{clust}. In other words, κ is replaced by $\lambda_\phi \ge \kappa$ in the identification cost, and no factor exponential in K appears. The price is a change of the algorithm, which explores uniformly during the first n_0 visits of every user and incurs a PSG of up to 2 in each of these rounds, and the choice of n_0 requires knowledge of λ_ϕ, γ', $\bar{\rho}_T$ and B. The guarantees in the main text are stated for Algorithm 2, not for this variant.

D.11.2 Misspecification

Suppose that (A5) holds only approximately: $\mu_i^k(\boldsymbol{x}) = \langle \phi^k(\boldsymbol{x}), \boldsymbol{w}^k_{j_k(i)}\rangle + a^k_{j_k(i)}(\boldsymbol{x})$ with $\|\boldsymbol{w}_j^k\|_2 \le B$ and $\sup_{\boldsymbol{x}\in\mathcal{X}} |a_j^k(\boldsymbol{x})| \le \varepsilon_{\mathrm{app}}$ for all k and j, where users of the same cluster share both $\boldsymbol{w}_j^k$ and a_j^k, and $|\mu_i^k| \le 1$ as before. Fix a group $(A,k)\in\Gamma$ as in Appendix D.2 with n samples up to time $t \le T$, and write $\boldsymbol{M} = \boldsymbol{M}^k_{A,t}$, $s_n(\boldsymbol{x}) = \|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}}$, $\boldsymbol{\Phi}\in\mathbb{R}^{n\times p_k}$ for the matrix of sample features and $\boldsymbol{a} = (a_A^k(\boldsymbol{x}_s))_s \in \mathbb{R}^n$ for the approximation errors at the samples, where $a_A^k := a_j^k$ for $A \subseteq C_j^{(k)}$. With $\boldsymbol{\zeta}_t$ as in Appendix D.2, now $\boldsymbol{b}^k_{A,t} = (\boldsymbol{M}-\lambda\boldsymbol{I})\boldsymbol{w}_A^k + \boldsymbol{\Phi}^\top\boldsymbol{a} + \boldsymbol{\zeta}_t$, so the decomposition (57) acquires an additional term:

$$\hat{\boldsymbol{w}}^k_{A,t} - \boldsymbol{w}_A^k = -\lambda\boldsymbol{M}^{-1}\boldsymbol{w}_A^k + \boldsymbol{M}^{-1}\boldsymbol{\zeta}_t + \boldsymbol{M}^{-1}\boldsymbol{\Phi}^\top\boldsymbol{a}.$$

By (68), $\|\boldsymbol{\Phi}^\top\boldsymbol{a}\|^2_{\boldsymbol{M}^{-1}} = \boldsymbol{a}^\top\boldsymbol{\Phi}\boldsymbol{\Phi}^\top(\lambda\boldsymbol{I}+\boldsymbol{\Phi}\boldsymbol{\Phi}^\top)^{-1}\boldsymbol{a} \le \|\boldsymbol{a}\|_2^2 \le n\varepsilon_{\mathrm{app}}^2$, because the eigenvalues of $\boldsymbol{\Phi}\boldsymbol{\Phi}^\top(\lambda\boldsymbol{I}+\boldsymbol{\Phi}\boldsymbol{\Phi}^\top)^{-1}$ lie in $[0,1)$. The noise $\eta_s^k = r_s^k - \mu^k_{i_s}(\boldsymbol{x}_s)$ is defined relative to the true mean reward, so the self-normalized bound $\|\boldsymbol{\zeta}_t\|_{\boldsymbol{M}^{-1}} \le R\sqrt{\log\det(\lambda^{-1}\boldsymbol{M}) + 2\log(8um_o/\delta)}$ in the argument of Appendix D.2 does not use realizability and holds, simultaneously for all groups in Γ and all $0\le t\le T$, with probability at least $1-\delta/4$; the bias term is again at most $\lambda\|\boldsymbol{w}_A^k\|_{\boldsymbol{M}^{-1}} \le B\sqrt{\lambda}$. Since $\mu_A^k(\boldsymbol{x}) = \phi^k(\boldsymbol{x})^\top\boldsymbol{w}_A^k + a_A^k(\boldsymbol{x})$, the Cauchy–Schwarz argument of Appendix D.2 gives, on the event of this self-normalized bound,

$$\left|\hat{f}^k_{A,t}(\boldsymbol{x}) - \mu_A^k(\boldsymbol{x})\right| \ \le\ \left(\beta(\boldsymbol{M}) + \varepsilon_{\mathrm{app}}\sqrt{n}\right) s_n(\boldsymbol{x}) + \varepsilon_{\mathrm{app}}.$$

The approximation error therefore does not simply add $\varepsilon_{\mathrm{app}}$ to the well-specified bound. It also enters the self-normalized term, where it grows like $\varepsilon_{\mathrm{app}}\sqrt{n}$ and is multiplied by the width $s_n(\boldsymbol{x})$. Under misspecification, the radii of Algorithm 2 no longer guarantee valid UCBs or the absence of false deletions, and the guarantee of Section 4.2 does not apply. The quadratic and cosine environments, whose rewards are not linear in the frozen features, are therefore tests under misspecification and are not covered by our guarantees. The MovieLens rewards are linear in the context; with identity features, the exact version satisfies the conditions of Section 4.2 and deviates from those of Section 4.1 only through parameter norms above 1 ($B = 3.62$), and the raw version ($B = 3.13$) is only approximately clustered.

D.11.3 Inexact ridge solvers

Algorithm 2 uses the exact ridge solutions $\hat{\boldsymbol{w}} = \boldsymbol{M}^{-1}\boldsymbol{b}$. The ridge objective $\mathcal{L}(\boldsymbol{w}) = \frac{1}{2}\sum_s(\phi^k(\boldsymbol{x}_s)^\top\boldsymbol{w} - r_s^k)^2 + \frac{\lambda}{2}\|\boldsymbol{w}\|_2^2$ is λ-strongly convex with gradient $\nabla\mathcal{L}(\boldsymbol{w}) = \boldsymbol{M}\boldsymbol{w} - \boldsymbol{b}$. Suppose that an iterative solver returns $\tilde{\boldsymbol{w}}$ with residual $\boldsymbol{e} = \boldsymbol{M}\tilde{\boldsymbol{w}} - \boldsymbol{b}$, which can be computed at run time. Then $\tilde{\boldsymbol{w}} - \hat{\boldsymbol{w}} = \boldsymbol{M}^{-1}\boldsymbol{e}$ and

$$\|\tilde{\boldsymbol{w}} - \hat{\boldsymbol{w}}\|_{\boldsymbol{M}} = \|\boldsymbol{e}\|_{\boldsymbol{M}^{-1}} \le \frac{\|\boldsymbol{e}\|_2}{\sqrt{\lambda}},$$

because $\boldsymbol{M} \succeq \lambda\boldsymbol{I}$. If $\|\boldsymbol{e}\|_2 \le \varepsilon_{\mathrm{solve}}\sqrt{\lambda}$, then $|\phi^k(\boldsymbol{x})^\top\tilde{\boldsymbol{w}} - \phi^k(\boldsymbol{x})^\top\hat{\boldsymbol{w}}| \le \|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}}\|\tilde{\boldsymbol{w}}-\hat{\boldsymbol{w}}\|_{\boldsymbol{M}} \le \varepsilon_{\mathrm{solve}}\|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}}$ for all $\boldsymbol{x}$. Suppose that every user-level and cluster-level solve satisfies this residual certificate with respect to its own system matrix $\boldsymbol{M}$ and right-hand side $\boldsymbol{b}$, that $\beta(\boldsymbol{M})$ is replaced by $\beta(\boldsymbol{M}) + \varepsilon_{\mathrm{solve}}$ in the UCBs, and that $\rho(\boldsymbol{M})$ is replaced by $\rho(\boldsymbol{M}) + \varepsilon_{\mathrm{solve}}$ in the deletion widths. Consider an approximate solve whose system matrix is $\boldsymbol{M} = \boldsymbol{M}^k_{A,t}$ for a fixed group

$(A,k) \in \Gamma$: a single user, or a true cluster $C_j^{(k)}$, which by (51) is the case for the cluster-level solve of round t whenever $\widehat{C}_t^k = C_j^{(k)}$. For such a solve, the triangle inequality, (55) and (56) give, on $\mathcal{E}_{\mathrm{conf}}$ and for all $\boldsymbol{x}$,

$$\left|\phi^k(\boldsymbol{x})^\top \tilde{\boldsymbol{w}} - \mu_A^k(\boldsymbol{x})\right| \;\le\; \min\Big\{ \big(\beta(\boldsymbol{M}) + \varepsilon_{\mathrm{solve}}\big)\, \|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}}, \\ \big(\rho(\boldsymbol{M}) + \varepsilon_{\mathrm{solve}}\big)\, \|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}} + \lambda B \|\boldsymbol{M}^{-1}\phi^k(\boldsymbol{x})\|_2 \Big\}.$$

Appendices D.4 and D.5 use the second form and Appendix D.8.1 the first, in each case only for single users and, after recovery, for true clusters. The proof therefore goes through with $\bar{\rho}_T$ and $\bar{\beta}_T$ replaced by $\bar{\rho}_T + \varepsilon_{\mathrm{solve}}$ and $\bar{\beta}_T + \varepsilon_{\mathrm{solve}}$ in $n_\star$, and hence in t_{fit} and T_{clust}, and $\bar{\beta}_T$ by $\bar{\beta}_T + \varepsilon_{\mathrm{solve}}$ in the regret bound; the cluster-level solves before recovery are not used by the analysis, which bounds the regret of the rounds $t \le T_{\mathrm{clust}}$ by $2T_{\mathrm{clust}}$. Unlike a uniform bound on the prediction error of a trained network, this certificate can be verified during the run.

E PSG LIMITATION, PLATFORM PREFERENCES AND INFORMATION ADVANTAGE

This appendix collects the results of the main text that concern what the Pareto suboptimality gap (PSG) measures and how the learned per-objective models can serve trade-offs chosen by the platform. Appendix E.1 fixes notation and records the max–min form of the PSG. Appendix E.2 proves (14): the PSG regret is at most the regret of any single objective, so a small PSG does not require learning all objectives well. Appendix E.3 treats preference-conditioned selection: it proves that, for strictly positive weights, the selected arm lies in the UCB Pareto set (the statement used in Section 3.2), proves the scalarized-regret guarantee (Appendix E.3.3) for both the linear and the fixed-feature algorithm together with the event on which it holds, and shows that uniform sampling from the Pareto set can have linear scalarized regret even when the rewards are known. Appendix E.4 proves the information advantage of objective-specific pooling and states its limitation. Appendix E.5 explains how these results are used in the experiments.

E.1 NOTATION AND THE MAX–MIN FORM OF THE PSG

For $a \in \mathbb{R}$ we write $[a]_+ := \max\{a, 0\}$. For symmetric matrices, $\boldsymbol{A} \succeq \boldsymbol{B}$ (equivalently, $\boldsymbol{B} \preceq \boldsymbol{A}$) denotes the Loewner order, i.e., $\boldsymbol{A} - \boldsymbol{B}$ is positive semi-definite; it should not be confused with Pareto dominance $\boldsymbol{a} \succ \boldsymbol{b}$ of reward vectors, used in (2). We write $\boldsymbol{\mu}_i(\boldsymbol{x}) = (\mu_i^1(\boldsymbol{x}), \dots, \mu_i^{m_o}(\boldsymbol{x}))^\top$ and $\boldsymbol{U}_t(\boldsymbol{x}) = (U_t^1(\boldsymbol{x}), \dots, U_t^{m_o}(\boldsymbol{x}))^\top$, where

$$U_t^k(\boldsymbol{x}) = \phi^k(\boldsymbol{x})^\top \bar{\boldsymbol{w}}_t^k + \beta(\bar{\boldsymbol{M}}_t^k)\, \|\phi^k(\boldsymbol{x})\|_{(\bar{\boldsymbol{M}}_t^k)^{-1}}$$

is the upper confidence bound of objective k at round t, and $\tilde{\mathcal{O}}_t = \{\boldsymbol{x} \in \mathcal{X}_t : \nexists\, \boldsymbol{x}' \in \mathcal{X}_t \text{ with } \boldsymbol{U}_t(\boldsymbol{x}') \succ \boldsymbol{U}_t(\boldsymbol{x})\}$ is the UCB Pareto set. In the linear case (COMOLB) the feature map is the identity and $B = 1$ in $\beta(\cdot)$; Appendix C writes $\boldsymbol{\theta}$ for $\boldsymbol{w}$. We use the width

$$s_t^k := \|\phi^k(\boldsymbol{x}_t)\|_{(\bar{\boldsymbol{M}}_t^k)^{-1}}, \qquad k \in [m_o],$$

of the selected context, and we write $C^k(i) := C^{(k)}_{j_k(i)}$ for the true cluster of user i for objective k. All users of a true cluster $C = C_j^{(k)}$ share the mean reward function of objective k, which we denote by μ_C^k; for a single user, $\mu_{\{i\}}^k = \mu_i^k$.

Group statistics. For a set of users $A \subseteq \mathcal{U}$, an objective k and $t \geq 0$, let

$$\begin{gathered} \boldsymbol{M}_{A,t}^k := \lambda \boldsymbol{I} + \sum_{s \leq t:\, i_s \in A} \phi^k(\boldsymbol{x}_s)\phi^k(\boldsymbol{x}_s)^\top, \qquad \boldsymbol{b}_{A,t}^k := \sum_{s \leq t:\, i_s \in A} r_s^k\, \phi^k(\boldsymbol{x}_s), \\ \hat{\boldsymbol{w}}_{A,t}^k := (\boldsymbol{M}_{A,t}^k)^{-1} \boldsymbol{b}_{A,t}^k, \qquad \hat{f}_{A,t}^k(\boldsymbol{x}) := \phi^k(\boldsymbol{x})^\top \hat{\boldsymbol{w}}_{A,t}^k. \end{gathered} \tag{69}$$

For $A = \{i\}$ we write $\boldsymbol{M}_{i,t}^k := \boldsymbol{M}_{\{i\},t}^k$, $\boldsymbol{b}_{i,t}^k := \boldsymbol{b}_{\{i\},t}^k$ and $\hat{\boldsymbol{w}}_{i,t}^k := \hat{\boldsymbol{w}}_{\{i\},t}^k$; these are the per-user statistics of the algorithms, where the superscript k on $\boldsymbol{M}$ indicates the feature map ϕ^k and can be dropped when ϕ^k does not depend on k (in particular in the linear case). Since $\boldsymbol{M}_{i,t-1}^k - \lambda \boldsymbol{I} = \sum_{s \leq t-1:\, i_s = i} \phi^k(\boldsymbol{x}_s)\phi^k(\boldsymbol{x}_s)^\top$ and every round contributes to the statistics of exactly one user, the cluster statistics of the algorithms satisfy

$$\bar{\boldsymbol{M}}_t^k = \boldsymbol{M}^k_{\widehat{C}_t^k,\, t-1}, \qquad \bar{\boldsymbol{w}}_t^k = \hat{\boldsymbol{w}}^k_{\widehat{C}_t^k,\, t-1}, \tag{70}$$

where $\widehat{C}_t^k = \{i_t\} \cup \mathcal{N}_{G_{t-1}^k}(i_t)$ is the closed neighbourhood of the arriving user.

Max–min form of the PSG. For every user i, every round t and every $\boldsymbol{x} \in \mathcal{X}_t$,

$$\Delta_i(\boldsymbol{x}) = \max_{\boldsymbol{x}' \in \mathcal{X}_t} \min_{k \in [m_o]} \big[\mu_i^k(\boldsymbol{x}') - \mu_i^k(\boldsymbol{x})\big]_+ = \max_{\boldsymbol{x}' \in \mathcal{X}_t} \min_{k \in [m_o]} \big[\mu_i^k(\boldsymbol{x}') - \mu_i^k(\boldsymbol{x})\big]. \tag{71}$$

Moreover $\Delta_i(\boldsymbol{x}) \leq 2$, and $\Delta_i(\boldsymbol{x}) = 0$ whenever no $\boldsymbol{x}' \in \mathcal{X}_t$ satisfies $\boldsymbol{\mu}_i(\boldsymbol{x}') \succ \boldsymbol{\mu}_i(\boldsymbol{x})$.

The first identity in (71) is (3), proved in Appendix C.1; we include a short proof of both identities for completeness.

Fix i, t and $\boldsymbol{x} \in \mathcal{X}_t$. For $\boldsymbol{x}' \in \mathcal{X}_t$ let $d_k(\boldsymbol{x}') := \mu_i^k(\boldsymbol{x}') - \mu_i^k(\boldsymbol{x})$, $g(\boldsymbol{x}') := \min_k d_k(\boldsymbol{x}')$, and $g^\star := \max_{\boldsymbol{x}' \in \mathcal{X}_t} g(\boldsymbol{x}')$, which exists because $\mathcal{X}_t$ is finite. Since $\boldsymbol{x} \in \mathcal{X}_t$ and $g(\boldsymbol{x}) = 0$, we have $g^\star \geq 0$. Let Z be the set of admissible $\zeta \geq 0$ in (2), so that $\Delta_i(\boldsymbol{x}) = \inf Z$.

First, every $\zeta > g^\star$ belongs to Z. Indeed, for every $\boldsymbol{x}' \in \mathcal{X}_t$ we have $\zeta > g(\boldsymbol{x}')$, so there is an objective k with $\mu_i^k(\boldsymbol{x}) + \zeta > \mu_i^k(\boldsymbol{x}')$; hence $\boldsymbol{\mu}_i(\boldsymbol{x}')$ does not weakly exceed $\boldsymbol{\mu}_i(\boldsymbol{x}) + \zeta\mathbf{1}$ in every coordinate, and in particular $\boldsymbol{\mu}_i(\boldsymbol{x}') \not\succ \boldsymbol{\mu}_i(\boldsymbol{x}) + \zeta\mathbf{1}$. Therefore $\inf Z \leq g^\star$.

Second, if $g^\star > 0$, no $\zeta \in [0, g^\star)$ belongs to Z. Indeed, let $\boldsymbol{x}''$ attain $g^\star$; then $\mu_i^k(\boldsymbol{x}'') - \mu_i^k(\boldsymbol{x}) \geq g^\star > \zeta$ for every k, so $\boldsymbol{\mu}_i(\boldsymbol{x}'') \succ \boldsymbol{\mu}_i(\boldsymbol{x}) + \zeta\mathbf{1}$. Together with $Z \subseteq [0, \infty)$ this gives $\inf Z \geq g^\star$, which is trivial if $g^\star = 0$.

Hence $\Delta_i(\boldsymbol{x}) = g^\star$, which is the second identity in (71). Since $g^\star \geq 0$, $g^\star = [g^\star]_+$, and since $a \mapsto [a]_+$ is nondecreasing, it commutes with maxima and minima over finite sets, so $[g^\star]_+ = \max_{\boldsymbol{x}'} \min_k [d_k(\boldsymbol{x}')]_+$; this is the first identity. Because $|\mu_i^k| \leq 1$, every $d_k(\boldsymbol{x}')$ is at most 2, so $\Delta_i(\boldsymbol{x}) \leq 2$. Finally, if $\Delta_i(\boldsymbol{x}) > 0$, then $g^\star > 0$ and the maximizer $\boldsymbol{x}''$ satisfies $\mu_i^k(\boldsymbol{x}'') > \mu_i^k(\boldsymbol{x})$ for all k, so $\boldsymbol{\mu}_i(\boldsymbol{x}'') \succ \boldsymbol{\mu}_i(\boldsymbol{x})$; contrapositively, $\Delta_i(\boldsymbol{x}) = 0$ when no arm of $\mathcal{X}_t$ Pareto-dominates $\boldsymbol{x}$.

E.2 The PSG Is Dominated by Any Single-Objective Regret

For $k_0 \in [m_o]$, the single-objective regret of objective k_0 is

$$R_T^{(k_0)} := \sum_{t=1}^{T} \Big[\max_{\boldsymbol{x} \in \mathcal{X}_t} \mu_{i_t}^{k_0}(\boldsymbol{x}) - \mu_{i_t}^{k_0}(\boldsymbol{x}_t) \Big],$$

and we write $R_T^{\mathrm{PSG}} := R_T = \sum_{t=1}^{T} \Delta_{i_t}(\boldsymbol{x}_t)$ for the PSG regret when the distinction matters.

In the notation of this appendix, (14) reads as follows. *Fix any objective $k_0 \in [m_o]$. For every user i, every round t and every $\boldsymbol{x} \in \mathcal{X}_t$,*

$$\Delta_i(\boldsymbol{x}) \leq \max_{\boldsymbol{x}' \in \mathcal{X}_t} \mu_i^{k_0}(\boldsymbol{x}') - \mu_i^{k_0}(\boldsymbol{x}). \tag{72}$$

Consequently, for every policy and every realization of the interaction, $R_T^{\mathrm{PSG}} \leq R_T^{(k_0)}$.

Indeed, since $\boldsymbol{x} \in \mathcal{X}_t$, Appendix E.1 gives $\Delta_i(\boldsymbol{x}) = \max_{\boldsymbol{x}' \in \mathcal{X}_t} \min_k [\mu_i^k(\boldsymbol{x}') - \mu_i^k(\boldsymbol{x})]$. For every $\boldsymbol{x}'$, the minimum over k is at most the term with $k = k_0$, so

$$\Delta_i(\boldsymbol{x}) \leq \max_{\boldsymbol{x}' \in \mathcal{X}_t} \big[\mu_i^{k_0}(\boldsymbol{x}') - \mu_i^{k_0}(\boldsymbol{x}) \big] = \max_{\boldsymbol{x}' \in \mathcal{X}_t} \mu_i^{k_0}(\boldsymbol{x}') - \mu_i^{k_0}(\boldsymbol{x}),$$

which is (72). Applying (72) with $i = i_t$ and $\boldsymbol{x} = \boldsymbol{x}_t \in \mathcal{X}_t$ in every round and summing over $t \in [T]$ yields $R_T^{\mathrm{PSG}} \leq R_T^{(k_0)}$.

Discussion. The bound (14) is a pathwise statement: it uses neither the reward model (linear or fixed-feature) nor the arrival and noise conditions, and it holds for every policy. Since k_0 is arbitrary, $R_T^{\mathrm{PSG}} \leq \min_{k_0 \in [m_o]} R_T^{(k_0)}$. The same argument, with the minimum over k bounded by the convex combination $\sum_k \omega_k [\mu_i^k(\boldsymbol{x}') - \mu_i^k(\boldsymbol{x})]$, gives $R_T^{\mathrm{PSG}} \leq R_T^{\boldsymbol{\omega}}$ for every fixed $\boldsymbol{\omega}$ in the probability simplex, with $R_T^{\boldsymbol{\omega}}$ the scalarized regret of Appendix E.3; a learner that selects by a fixed trade-off, such as S-CLUB or the preference-conditioned rule, thus has a PSG at most its scalarized regret for that trade-off. In particular, a learner that optimizes a single objective k_0 and ignores the other $m_o - 1$ reward streams inherits every regret guarantee it has for objective k_0 as a PSG guarantee. A small PSG therefore does not require learning all objectives well, and it does not certify that the selected arms are good for any objective other than the one being optimized. For the same reason, the PSG bounds of Sections 4.1 and 4.2, which grow with the total number $m_c = \sum_k m_c^{(k)}$ of objective-specific clusters because COMOLB and COMONB-Frozen learn every objective, should not be read as the intrinsic cost of attaining a small PSG. Criteria that involve every objective are needed to assess whether a method serves the trade-offs between objectives; the scalarized regret of the next subsection is such a criterion.

E.3 Preference-Conditioned Selection and Scalarized Regret

Preference schedules and the preference rule. Let $\mathcal{W}_{m_o} := \{\boldsymbol{\omega} \in \mathbb{R}^{m_o} : \omega_k \geq 0 \,\forall k,\ \sum_k \omega_k = 1\}$ be the probability simplex. A preference schedule is a sequence $(\boldsymbol{\omega}_t)_{t\geq 1}$ with $\boldsymbol{\omega}_t \in \mathcal{W}_{m_o}$, where $\boldsymbol{\omega}_t$ is revealed to the learner in round t before the arm is selected. The preference rule selects

$$\boldsymbol{x}_t \in \operatorname*{arg\,max}_{\boldsymbol{x}\in\mathcal{X}_t} \boldsymbol{\omega}_t^\top \boldsymbol{U}_t(\boldsymbol{x}) = \operatorname*{arg\,max}_{\boldsymbol{x}\in\mathcal{X}_t} \sum_{k=1}^{m_o} \omega_{t,k}\, U_t^k(\boldsymbol{x}), \tag{73}$$

with ties broken by any rule that depends only on the past, on i_t, $\mathcal{X}_t$ and $\boldsymbol{\omega}_t$, and on internal randomization of the learner (for instance, uniformly at random among the maximizers); as in Section 2, this internal randomization is part of the history. Rule (73) replaces only the selection step (uniform sampling from $\tilde{\mathcal{O}}_t$); the per-user statistics, the cluster statistics, the UCBs, the edge-deletion tests and the user counter (incremented once per round) are computed exactly as in COMOLB and COMONB-Frozen. The scalarized gap and the scalarized regret are

$$\Delta_t^{\boldsymbol{\omega}} := \max_{\boldsymbol{x}\in\mathcal{X}_t} \boldsymbol{\omega}_t^\top \boldsymbol{\mu}_{i_t}(\boldsymbol{x}) - \boldsymbol{\omega}_t^\top \boldsymbol{\mu}_{i_t}(\boldsymbol{x}_t), \qquad R_T^{\boldsymbol{\omega}} := \sum_{t=1}^{T} \Delta_t^{\boldsymbol{\omega}}. \tag{74}$$

Since $|\boldsymbol{\omega}^\top \boldsymbol{\mu}_i(\boldsymbol{x})| \leq \sum_k \omega_k |\mu_i^k(\boldsymbol{x})| \leq 1$ for every $\boldsymbol{\omega} \in \mathcal{W}_{m_o}$, we have $0 \leq \Delta_t^{\boldsymbol{\omega}} \leq 2$.

For $t \geq 0$, let $\mathcal{H}_t^{\boldsymbol{\omega}}$ denote the history $\mathcal{H}_t$ of Section 2 (the users, candidate sets, selected contexts and rewards of rounds $1, \ldots, t$, together with the internal randomization of the learner in these rounds), augmented by the weights $\boldsymbol{\omega}_1, \ldots, \boldsymbol{\omega}_t$.

Condition (E): exogenous preference schedule. A preference schedule is *exogenous* if the probabilistic conditions of the setting remain valid when the weights are added to the history, namely, for every round $t \geq 1$: (a) conditionally on $\mathcal{H}_{t-1}^{\boldsymbol{\omega}}$, the user i_t is uniformly distributed on $\mathcal{U}$, the K candidates of $\mathcal{X}_t$ are i.i.d. with distribution $\mathcal{D}$, and i_t and $\mathcal{X}_t$ are independent; equivalently, i_t, $\mathcal{X}_t$ and $\mathcal{H}_{t-1}^{\boldsymbol{\omega}}$ are mutually independent, with these laws for i_t and $\mathcal{X}_t$; (b) conditionally on $(\mathcal{H}_{t-1}^{\boldsymbol{\omega}}, i_t, \mathcal{X}_t, \boldsymbol{\omega}_t, \boldsymbol{x}_t)$, each noise variable η_t^k has mean zero and is R-sub-Gaussian in the sense of (A2).

Every deterministic schedule is exogenous, and so is every schedule in which $\boldsymbol{\omega}_t$ is a measurable function of $\mathcal{H}_{t-1}$, i_t and $\mathcal{X}_t$: in both cases the weights are functions of variables that already appear in the conditioning, $\mathcal{H}_{t-1}^{\boldsymbol{\omega}}$ generates the same σ-field as $\mathcal{H}_{t-1}$, and (a) and (b) are (A1) and (A2) for the learner that computes the weights itself. Condition (E) is the meaning of "exogenous" in Section 3.2 and Appendix E.3.3. It allows $\boldsymbol{\omega}_t$ to depend on the past, on i_t and on $\mathcal{X}_t$. It excludes, for instance, weights that depend on the noise of round t, which may violate (b), and weights that depend on the users or candidate sets of later rounds: if $\boldsymbol{\omega}_t$ depended on $\mathcal{X}_{t+1}$, then $\mathcal{H}_t^{\boldsymbol{\omega}}$ would carry information about $\mathcal{X}_{t+1}$, contrary to (a) in round $t+1$. The latter case would not be excluded by merely requiring that the weights not depend on the rewards of round t or later, and the arguments below need (a) in every round.

Role of the selection rule in the analyses of Sections 4.1 and 4.2. The selection step enters these proofs in only three ways.

- (R1) *Measurability.* The self-normalized bound (Abbasi-Yadkori et al., 2011) behind the confidence event requires a filtration $(\mathcal{F}_t)_{t\geq 0}$ such that i_t and $\phi^k(\boldsymbol{x}_t)$ are $\mathcal{F}_{t-1}$-measurable, η_t^k is $\mathcal{F}_t$-measurable, and η_t^k is conditionally zero-mean and R-sub-Gaussian given $\mathcal{F}_{t-1}$.
- (R2) $\boldsymbol{x}_t \in \mathcal{X}_t$. The spectral lower bound behind the spectral event uses only that the selected context is one of the candidates and that $\mathcal{X}_t$ is independent of the information available before it is revealed: for every $\boldsymbol{v}$, $\langle \boldsymbol{v}, \phi^k(\boldsymbol{x}_t)\rangle^2 \geq \min_{a\leq K} \langle \boldsymbol{v}, \phi^k(\boldsymbol{x}_{t,a})\rangle^2$ whatever the rule, and the conditional expectation of the right-hand side given the past and i_t is at least $\kappa\|\boldsymbol{v}\|^2$ for $\boldsymbol{v} \in \mathcal{S}_k$ (Appendix C.2.2). The arrival and covering events involve only the arrivals and the candidate sets. Finally, the arguments that derive conclusions (i) and (ii) of the guarantees on the high-probability event are deterministic and involve only the per-user statistics and the deletion tests, which are computed in the same way whatever the selection rule.

(R3) $\boldsymbol{x}_t \in \tilde{\mathcal{O}}_t$. This is used only in the Pareto step, which bounds $\Delta_{i_t}(\boldsymbol{x}_t)$ by $2\bar{\beta}\max_k s_t^k$ after recovery ($\bar{\beta} = \bar{\beta}_T$ or $\bar{\beta}_{\mathrm{lin}}$). The elliptical-potential argument that sums these widths holds for every sequence of selected contexts.

Consequently, for any selection rule satisfying (R1) and (R2), the events $\mathcal{E}^{\mathrm{lin}}$ and $\mathcal{E}^{\mathrm{nb}}$ defined below have probability at least $1-\delta$ and conclusions (i) and (ii) of the respective guarantee hold on them; if in addition $\boldsymbol{x}_t \in \tilde{\mathcal{O}}_t$ in every round, conclusion (iii) holds as well. For rule (73) with an exogenous schedule, we run the proofs of Appendices C and D with the histories $\mathcal{H}_t$ replaced by $\mathcal{H}_t^{\boldsymbol{\omega}}$: the information before the candidates of round t are revealed is $\sigma(\mathcal{H}_{t-1}^{\boldsymbol{\omega}}, i_t)$, and the information when the reward of round t is generated is $\mathcal{F}_{t-1} := \sigma(\mathcal{H}_{t-1}^{\boldsymbol{\omega}}, i_t, \mathcal{X}_t, \boldsymbol{\omega}_t, \boldsymbol{x}_t)$. Then (R1) holds for this filtration: i_t and $\phi^k(\boldsymbol{x}_t)$ are $\mathcal{F}_{t-1}$-measurable because i_t and $\boldsymbol{x}_t$ are; $\eta_t^k = r_t^k - \mu_{i_t}^k(\boldsymbol{x}_t)$ is $\mathcal{H}_t^{\boldsymbol{\omega}}$-measurable and $\mathcal{H}_t^{\boldsymbol{\omega}} \subseteq \mathcal{F}_t$; $\mathcal{F}_{t-1} \subseteq \mathcal{H}_t^{\boldsymbol{\omega}}$, so $(\mathcal{F}_t)$ is increasing; and the conditional noise property is condition (E)(b). The independence used in (R2) is condition (E)(a), which also implies that the pairs $(i_t, \mathcal{X}_t)$, $t \geq 1$, are i.i.d., because $(i_s, \mathcal{X}_s)_{s<t}$ is $\mathcal{H}_{t-1}^{\boldsymbol{\omega}}$-measurable; this is all that the arrival and covering events use. Finally, $\boldsymbol{x}_t \in \mathcal{X}_t$ holds by construction.

E.3.1 Strictly positive weights select from the UCB Pareto set

Fix a round t and a weight vector $\boldsymbol{\omega} \in \mathbb{R}^{m_o}$ with $\omega_k > 0$ for all $k \in [m_o]$. Then every maximizer of $\boldsymbol{x} \mapsto \boldsymbol{\omega}^\top \boldsymbol{U}_t(\boldsymbol{x})$ over $\mathcal{X}_t$ belongs to $\tilde{\mathcal{O}}_t$.

To see this, let $\boldsymbol{x}_t$ be a maximizer and suppose, for contradiction, that some $\boldsymbol{x}' \in \mathcal{X}_t$ satisfies $\boldsymbol{U}_t(\boldsymbol{x}') \succ \boldsymbol{U}_t(\boldsymbol{x}_t)$, that is, $U_t^k(\boldsymbol{x}') \geq U_t^k(\boldsymbol{x}_t)$ for all k and $U_t^j(\boldsymbol{x}') > U_t^j(\boldsymbol{x}_t)$ for some j. Since all weights are nonnegative and $\omega_j > 0$,

$$\boldsymbol{\omega}^\top \boldsymbol{U}_t(\boldsymbol{x}') - \boldsymbol{\omega}^\top \boldsymbol{U}_t(\boldsymbol{x}_t) = \sum_{k=1}^{m_o} \omega_k \big(U_t^k(\boldsymbol{x}') - U_t^k(\boldsymbol{x}_t)\big) \geq \omega_j \big(U_t^j(\boldsymbol{x}') - U_t^j(\boldsymbol{x}_t)\big) > 0,$$

which contradicts the maximality of $\boldsymbol{x}_t$. Hence no arm of $\mathcal{X}_t$ Pareto-dominates $\boldsymbol{x}_t$ in UCB, i.e., $\boldsymbol{x}_t \in \tilde{\mathcal{O}}_t$.

This membership property is the one used in Section 3.2. It concerns only the selection step and holds for any vector of UCBs. Combined with (R1)–(R3), it shows: *if the schedule is exogenous and $\omega_{t,k} > 0$ for all t and k, then COMOLB and COMONB-Frozen with rule* (73) *satisfy conclusions (i)–(iii) of Sections 4.1 and 4.2, respectively, on the events $\mathcal{E}^{\mathrm{lin}}$ and $\mathcal{E}^{\mathrm{nb}}$ defined below, each of probability at least $1-\delta$*; in particular the PSG bounds are unchanged. Strict positivity cannot be dropped: for $m_o = 2$, $\boldsymbol{\omega} = (1,0)^\top$ and two arms with $\boldsymbol{U}_t(\boldsymbol{x}_a) = (1,0)^\top$ and $\boldsymbol{U}_t(\boldsymbol{x}_b) = (1,1)^\top$, both arms maximize $\boldsymbol{\omega}^\top \boldsymbol{U}_t$, but $\boldsymbol{x}_a \notin \tilde{\mathcal{O}}_t$. The scalarized-regret guarantee of Appendix E.3.3 does not need this membership and holds for all weights in $\mathcal{W}_{m_o}$.

E.3.2 The high-probability events

We now make explicit the events on which the guarantees of Sections 4.1 and 4.2 and of Appendix E.3.3 hold. They are the events constructed in Appendices C and D, evaluated on the trajectory generated by whichever selection rule is used. Recall the constants n_{spec}, $n_\star$, t_{fit}, T_{clust}, $\bar{\beta}_T$, G_T of Section 4.2 and n_{lin}, T_{lin}, $\bar{\beta}_{\mathrm{lin}}$ of Section 4.1. For fixed features, $\mathcal{E}^{\mathrm{nb}} := \Omega_0 \cap \mathcal{E}_{\mathrm{conf}} \cap \mathcal{E}_{\mathrm{spec}} \cap \mathcal{E}_{\mathrm{arr}} \cap \mathcal{E}_{\mathrm{cov}}$ is the event $\mathcal{E}$ of Appendix D.7, where the numbers of arrivals and the rounds after T refer to the interaction extended beyond round T as in Appendix D.1, which does not change the first T rounds:

- $\Omega_0 := \{\phi^k(\boldsymbol{z}) \in \mathcal{S}_k$ for all $k \in [m_o]$, $t \geq 1$ and $\boldsymbol{z} \in \mathcal{X}_t\}$ has probability one, because $\phi^k(\boldsymbol{x}) \in \mathcal{S}_k$ for $\mathcal{D}$-almost every $\boldsymbol{x}$ and countably many candidates are drawn;
- $\mathcal{E}_{\mathrm{conf}}$ (confidence, Appendix D.2): for every $k \in [m_o]$, every group $A \in \{\{i\} : i \in \mathcal{U}\} \cup \mathcal{C}^{(k)}$ (single users and true clusters of objective k), every $0 \leq t \leq T$ and every context $\boldsymbol{x} \in \mathcal{X}$, both
$$\big|\hat{f}_{A,t}^k(\boldsymbol{x}) - \mu_A^k(\boldsymbol{x})\big| \leq \beta(\boldsymbol{M}_{A,t}^k)\, \|\phi^k(\boldsymbol{x})\|_{(\boldsymbol{M}_{A,t}^k)^{-1}} \tag{75}$$
and the split bound (56) hold; both follow from the noise-sum bound $\|\sum_{s \leq t:\, i_s \in A} \eta_s^k \phi^k(\boldsymbol{x}_s)\|_{(\boldsymbol{M}_{A,t}^k)^{-1}} \leq \rho(\boldsymbol{M}_{A,t}^k)$ established in Appendix D.2;

- $\mathcal{E}_{\mathrm{spec}}$ (spectral, Appendix D.3): for every user i, every k and every $t \le T$ with $T_{i,t} \ge n_{\mathrm{spec}}$, $\boldsymbol{v}^\top \boldsymbol{M}^k_{i,t} \boldsymbol{v} \ge \lambda + \kappa T_{i,t}/2$ for all unit vectors $\boldsymbol{v} \in \mathcal{S}_k$;
- $\mathcal{E}_{\mathrm{arr}}$ (arrival, Appendix D.6.1): $T_{i,t_{\mathrm{fit}}} \ge n_\star$ for every user i;
- $\mathcal{E}_{\mathrm{cov}}$ (covering, Appendix D.6.2): for every k and every unordered pair $\{i,\ell\}$ of users with $j_k(i) \ne j_k(\ell)$ and designated endpoint i, there is a round $t \in \{t_{\mathrm{fit}}+1,\dots,T_{\mathrm{clust}}\}$ with $i_t = i$ such that $\mathcal{X}_t$ contains a context $\boldsymbol{z}$ with $|\mu^k_i(\boldsymbol{z}) - \mu^k_\ell(\boldsymbol{z})| \ge \gamma'$.

For the linear model, $\mathcal{E}^{\mathrm{lin}} := \mathcal{E}^{\mathrm{lin}}_{\mathrm{conf}} \cap \mathcal{E}^{\mathrm{lin}}_{\mathrm{spec}} \cap \mathcal{E}^{\mathrm{lin}}_{\mathrm{arr}}$ is the event $\mathcal{E}$ constructed in Appendix C.4, where the superscript "lin" distinguishes the events denoted $\mathcal{E}_{\mathrm{conf}}$, $\mathcal{E}_{\mathrm{spec}}$ and $\mathcal{E}_{\mathrm{arr}}$ in Appendix C from those above:

- $\mathcal{E}^{\mathrm{lin}}_{\mathrm{conf}}$ (Appendix C.3.1): the noise-sum bound of Appendix C.3.1 for all k, all single users and true clusters A of objective k, and all $0 \le t \le T$, with $B = 1$; by (37), it implies $\|\hat{\boldsymbol{w}}^k_{A,t} - \boldsymbol{\theta}^k_A\|_{\boldsymbol{M}_{A,t}} \le \beta(\boldsymbol{M}_{A,t})$, where $\boldsymbol{\theta}^k_A$ is the common parameter of the group, and hence (75) with the identity feature map for all $\boldsymbol{x} \in \mathbb{R}^d$ and $0 \le t \le T$;
- $\mathcal{E}^{\mathrm{lin}}_{\mathrm{spec}}$ (Appendix C.3.3(a)): $\lambda_{\min}(\boldsymbol{M}_{i,t}) \ge \lambda + \kappa_{x,K} T_{i,t}/2$ for every user i and every $t \le T$ with $T_{i,t} \ge n_{\mathrm{lin}}$;
- $\mathcal{E}^{\mathrm{lin}}_{\mathrm{arr}}$ (Appendix C.3.3(c)): if $T \ge T_{\mathrm{lin}}$, then $T_{i,T_{\mathrm{lin}}} \ge n_{\mathrm{lin}}$ for every user i; if $T < T_{\mathrm{lin}}$, it is the sure event.

The cited appendices show that each of the four events of $\mathcal{E}^{\mathrm{nb}}$ other than Ω_0 and each of the three events of $\mathcal{E}^{\mathrm{lin}}$ fails with probability at most $\delta/4$, so that $\mathbb{P}(\mathcal{E}^{\mathrm{nb}}) \ge 1-\delta$ and $\mathbb{P}(\mathcal{E}^{\mathrm{lin}}) \ge 1 - 3\delta/4 \ge 1-\delta$. Appendix D.7 and Appendices C.3.2 and C.3.3 show that conclusions (i) and (ii) of the respective guarantee hold on these events. By (R1)–(R2), both statements remain true for rule (73) with an exogenous schedule.

E.3.3 Scalarized-regret guarantee for platform preferences

Let $(\boldsymbol{\omega}_t)_{t\ge1}$ be a preference schedule with $\boldsymbol{\omega}_t \in \mathcal{W}_{m_o}$ that is exogenous in the sense of condition (E), and let the arm be selected by rule (73), all other steps of the algorithm being unchanged.

(a) (Fixed features.) Under the conditions of Section 4.2, $\mathbb{P}(\mathcal{E}^{\mathrm{nb}}) \ge 1-\delta$, and on $\mathcal{E}^{\mathrm{nb}}$ COMONB-Frozen with rule (73) satisfies

$$R^{\boldsymbol{\omega}}_T \le \min\Big\{2T,\ 2T_{\mathrm{clust}} + 2\bar{\beta}_T\sqrt{2m_c T G_T}\Big\}. \tag{76}$$

(b) (Linear.) Under the conditions of Section 4.1, $\mathbb{P}(\mathcal{E}^{\mathrm{lin}}) \ge 1-\delta$, and on $\mathcal{E}^{\mathrm{lin}}$ COMOLB with rule (73) satisfies

$$R^{\boldsymbol{\omega}}_T \le \min\Big\{2T,\ 2T_{\mathrm{lin}} + 2\bar{\beta}_{\mathrm{lin}}\sqrt{2m_c\, d\, T \log\big(1 + T/(d\lambda)\big)}\Big\}. \tag{77}$$

Here $\mathcal{E}^{\mathrm{nb}}$ and $\mathcal{E}^{\mathrm{lin}}$ are evaluated on the trajectory generated by rule (73). The weights need not be strictly positive and may vary with t; the bounds do not depend on the schedule. The term $2T$ in the minimum is the trivial bound $R^{\boldsymbol{\omega}}_T \le 2T$.

We prove part (a) and indicate at the end the changes for part (b).

Step 1 (the event). By the discussion following (R1)–(R3), $\mathbb{P}(\mathcal{E}^{\mathrm{nb}}) \ge 1-\delta$ for the trajectory generated by rule (73), and conclusions (i) and (ii) of Section 4.2 hold on $\mathcal{E}^{\mathrm{nb}}$. Fix an outcome in $\mathcal{E}^{\mathrm{nb}}$ for the rest of the proof; all statements below are deterministic consequences of this choice.

Step 2 (rounds before recovery). Since $0 \le \Delta^{\boldsymbol{\omega}}_t \le 2$ in every round, $R^{\boldsymbol{\omega}}_T \le 2T$, and the rounds $t \le \min\{T, T_{\mathrm{clust}}\}$ contribute at most $2T_{\mathrm{clust}}$. If $T \le T_{\mathrm{clust}}$, (76) follows. Assume from now on that $T > T_{\mathrm{clust}}$.

Step 3 (cluster statistics after recovery). By conclusion (ii) (Appendix D.7), at the end of round T_{clust} every graph G^k has edge set $\{(i,i') :\ j_k(i) = j_k(i')\}$. Edges are never added, and by conclusion (i) no edge of this set is ever deleted; hence G^k_{t-1} is this graph for every $t \ge T_{\mathrm{clust}}+1$. In it, the closed neighbourhood of i_t is its true cluster, so $\widehat{C}^k_t = C^k(i_t)$ and, by (70),

$$\bar{\boldsymbol{M}}^k_t = \boldsymbol{M}^k_{C^k(i_t),\,t-1}, \qquad \bar{\boldsymbol{w}}^k_t = \hat{\boldsymbol{w}}^k_{C^k(i_t),\,t-1}, \qquad t \ge T_{\mathrm{clust}}+1,\ k \in [m_o].$$

Step 4 (optimism and width). Let $t \geq T_{\text{clust}} + 1$ and $k \in [m_o]$. The group $A = C^k(i_t)$ is a true cluster of objective k, and $\mu_A^k = \mu_{i_t}^k$ because $i_t \in A$. Applying (75) to A at time $t-1$ and using Step 3, for every $\boldsymbol{x} \in \mathcal{X}_t$,

$$\mu_{i_t}^k(\boldsymbol{x}) \leq U_t^k(\boldsymbol{x}) \leq \mu_{i_t}^k(\boldsymbol{x}) + 2\beta(\bar{\boldsymbol{M}}_t^k)\, \|\phi^k(\boldsymbol{x})\|_{(\bar{\boldsymbol{M}}_t^k)^{-1}}. \tag{78}$$

Moreover, $\bar{\boldsymbol{M}}_t^k/\lambda = \boldsymbol{I} + \frac{1}{\lambda}\sum_{s \leq t-1:\, i_s \in A} \phi^k(\boldsymbol{x}_s)\phi^k(\boldsymbol{x}_s)^\top$ contains at most $t-1 \leq T$ rank-one terms, whose features lie in $\mathcal{S}_k$ on Ω_0; hence $\log\det(\bar{\boldsymbol{M}}_t^k/\lambda) \leq G_T$ by (8). Since $\beta(\cdot)$ is nondecreasing in this log-determinant, $\beta(\bar{\boldsymbol{M}}_t^k) \leq \bar{\beta}_T$.

Step 5 (per-round scalarized gap). Let $t \geq T_{\text{clust}} + 1$ and let $\boldsymbol{x}_t^\star \in \arg\max_{\boldsymbol{x} \in \mathcal{X}_t} \boldsymbol{\omega}_t^\top \boldsymbol{\mu}_{i_t}(\boldsymbol{x})$. Using, in this order, the lower bound in (78) with $\omega_{t,k} \geq 0$, the maximality of $\boldsymbol{x}_t$ in (73), the upper bound in (78) and Step 4,

$$\begin{aligned}\Delta_t^{\boldsymbol{\omega}} &= \boldsymbol{\omega}_t^\top \boldsymbol{\mu}_{i_t}(\boldsymbol{x}_t^\star) - \boldsymbol{\omega}_t^\top \boldsymbol{\mu}_{i_t}(\boldsymbol{x}_t) \leq \boldsymbol{\omega}_t^\top \boldsymbol{U}_t(\boldsymbol{x}_t^\star) - \boldsymbol{\omega}_t^\top \boldsymbol{\mu}_{i_t}(\boldsymbol{x}_t) \\ &\leq \sum_{k=1}^{m_o} \omega_{t,k}\big(U_t^k(\boldsymbol{x}_t) - \mu_{i_t}^k(\boldsymbol{x}_t)\big) \leq 2\bar{\beta}_T \sum_{k=1}^{m_o} \omega_{t,k}\, s_t^k.\end{aligned}$$

Step 6 (summation). By the Cauchy–Schwarz inequality over the at most T rounds $t \in \{T_{\text{clust}} + 1, \ldots, T\}$, and then by Jensen's inequality for the convex function $y \mapsto y^2$ and the probability vector $\boldsymbol{\omega}_t$, together with $\omega_{t,k} \leq 1$,

$$\begin{aligned}\sum_{t=T_{\text{clust}}+1}^{T} \Delta_t^{\boldsymbol{\omega}} &\leq 2\bar{\beta}_T \sqrt{T \sum_{t=T_{\text{clust}}+1}^{T} \Big(\sum_k \omega_{t,k} s_t^k\Big)^2} \leq 2\bar{\beta}_T \sqrt{T \sum_{t=T_{\text{clust}}+1}^{T} \sum_k \omega_{t,k} (s_t^k)^2} \\ &\leq 2\bar{\beta}_T \sqrt{T \sum_{k=1}^{m_o} \sum_{t=T_{\text{clust}}+1}^{T} (s_t^k)^2}.\end{aligned}$$

Step 7 (elliptical potential per true cluster). Fix k and a true cluster $C = C_j^{(k)}$, and let $t_1 < \cdots < t_N$ be the rounds $t \in \{T_{\text{clust}} + 1, \ldots, T\}$ with $i_t \in C$. Put $\boldsymbol{V}_0 := \boldsymbol{M}_{C,T_{\text{clust}}}^k$ and $\boldsymbol{V}_n := \boldsymbol{V}_{n-1} + \phi^k(\boldsymbol{x}_{t_n})\phi^k(\boldsymbol{x}_{t_n})^\top$. Between T_{clust} and $t_n - 1$ the users of C arrive exactly in the rounds $t_1, \ldots, t_{n-1}$, so Step 3 gives $\bar{\boldsymbol{M}}_{t_n}^k = \boldsymbol{M}_{C,t_n-1}^k = \boldsymbol{V}_{n-1}$ and hence $s_{t_n}^k = \|\phi^k(\boldsymbol{x}_{t_n})\|_{\boldsymbol{V}_{n-1}^{-1}}$; also $\boldsymbol{V}_N = \boldsymbol{M}_{C,T}^k$. By the rank-one determinant identity, $\det \boldsymbol{V}_n = \det \boldsymbol{V}_{n-1}\,(1 + (s_{t_n}^k)^2)$. Because $\boldsymbol{V}_{n-1} \succeq \lambda \boldsymbol{I}$, $\lambda \geq 1$ and $\|\phi^k\| \leq 1$, we have $(s_{t_n}^k)^2 \leq 1/\lambda \leq 1$, and $\log(1+y) \geq y/2$ for $y \in [0,1]$ (the difference vanishes at $y = 0$ and has derivative $1/(1+y) - 1/2 \geq 0$ on $[0,1]$). Therefore

$$\sum_{n=1}^{N} (s_{t_n}^k)^2 \leq 2\sum_{n=1}^{N} \log\big(1 + (s_{t_n}^k)^2\big) = 2\log\frac{\det \boldsymbol{V}_N}{\det \boldsymbol{V}_0} \leq 2\log\det(\boldsymbol{M}_{C,T}^k/\lambda) = 2\, l_{j,T}^k,$$

where the last equality is (53) (with $C = C_j^{(k)}$) and the last inequality uses $\det(\boldsymbol{V}_0/\lambda) \geq 1$, which holds because $\boldsymbol{V}_0 \succeq \lambda \boldsymbol{I}$; in particular, the pre-recovery samples contained in $\boldsymbol{V}_0$ do not increase the bound. Since $\boldsymbol{M}_{C,T}^k$ contains at most T rank-one terms, whose features lie in $\mathcal{S}_k$ on Ω_0, we have $l_{j,T}^k \leq G_T$ by (8). Every round $t > T_{\text{clust}}$ belongs, for each k, to exactly one true cluster of objective k (the one containing i_t), so summing over $j \in [m_c^{(k)}]$ and $k \in [m_o]$ gives

$$\sum_{k=1}^{m_o} \sum_{t=T_{\text{clust}}+1}^{T} (s_t^k)^2 \leq 2\sum_{k=1}^{m_o} \sum_{j=1}^{m_c^{(k)}} l_{j,T}^k \leq 2m_c G_T.$$

Step 8 (conclusion of part (a)). Steps 2, 6 and 7 give $R_T^{\boldsymbol{\omega}} \leq 2T_{\text{clust}} + 2\bar{\beta}_T\sqrt{2m_c T G_T}$, and together with $R_T^{\boldsymbol{\omega}} \leq 2T$ this is (76). As in Section 4.2, the argument also yields the finer bound in which $m_c G_T$ is replaced by $\sum_k \sum_j l_{j,T}^k$.

Part (b). The linear case is the special case $\phi^k = $ identity, $B = 1$, with $\mathcal{E}^{\text{nb}}$, T_{clust} and Section 4.2 replaced by $\mathcal{E}^{\text{lin}}$, T_{lin} and Section 4.1; the event Ω_0 is not needed, because the features are the

contexts themselves and no subspace restriction enters. Steps 1–7 apply verbatim, where Step 3 uses Appendix C.3.3(c) and Step 4 uses $\mathcal{E}_{\mathrm{conf}}^{\mathrm{lin}}$. The only additional property needed is that, for $n \le T$ and $\|\boldsymbol{x}_s\| \le 1$, the arithmetic–geometric mean inequality applied to the eigenvalues gives

$$\log\det\Big(\boldsymbol{I}+\frac{1}{\lambda}\sum_{s=1}^{n}\boldsymbol{x}_s\boldsymbol{x}_s^\top\Big) \le d\log\Big(\frac{1}{d}\operatorname{tr}\Big(\boldsymbol{I}+\frac{1}{\lambda}\sum_{s=1}^{n}\boldsymbol{x}_s\boldsymbol{x}_s^\top\Big)\Big) \le d\log\Big(1+\frac{n}{d\lambda}\Big) \le d\log\Big(1+\frac{T}{d\lambda}\Big).$$

Hence $\beta(\bar{\boldsymbol{M}}_t^k) \le \bar{\beta}_{\mathrm{lin}}$ in Step 4 and $l_{j,T}^k \le d\log(1+T/(d\lambda))$ in Step 7, and Step 8 gives (77).

Only Steps 5 and 6 involve the weights; Steps 1–4 and 7 are those of the analyses in Appendices C and D. For the vertex $\boldsymbol{\omega}_t \equiv \boldsymbol{e}_{k_0}$, rule (73) selects by the UCB of objective k_0 alone and $R_T^{\boldsymbol{\omega}} = R_T^{(k_0)}$; the guarantee above then bounds the regret of objective k_0, and (14) turns this into the same bound on the PSG regret, although the selected arm need not lie in $\tilde{\mathcal{O}}_t$.

E.3.4 Uniform sampling from the Pareto set

Appendix E.3.3 concerns rule (73), not the default rule that samples uniformly from $\tilde{\mathcal{O}}_t$. The latter can have linear scalarized regret even when the rewards are known. Let $d = m_o = K = 2$, let all users form a single cluster for each objective with $\boldsymbol{\theta}^1 = \boldsymbol{e}_1$ and $\boldsymbol{\theta}^2 = \boldsymbol{e}_2$, so that every user has $\boldsymbol{\mu}_i(\boldsymbol{e}_1) = (1,0)^\top$ and $\boldsymbol{\mu}_i(\boldsymbol{e}_2) = (0,1)^\top$, let $\mathcal{D}$ be the uniform distribution on $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$, let the noise be any noise satisfying (A2), and let $\boldsymbol{\omega}_t \equiv \boldsymbol{\omega} = (0.9, 0.1)^\top$. This instance satisfies the conditions of Section 4.1: $\|\boldsymbol{x}\| \le 1$ and $\|\boldsymbol{\theta}^k\| \le 1$; cluster separation is vacuous, since each objective has a single cluster (so (A3) holds for every $\gamma > 0$); and for a unit vector $\boldsymbol{v} = (v_1, v_2)$ the two candidates are both $\boldsymbol{e}_1$, both $\boldsymbol{e}_2$, or distinct with probabilities $1/4$, $1/4$ and $1/2$, so $\mathbb{E}[\min_{a\le 2}\langle \boldsymbol{v}, \boldsymbol{x}_a\rangle^2] = \frac{1}{4}(v_1^2+v_2^2) + \frac{1}{2}\min\{v_1^2, v_2^2\} \ge \frac{1}{4}$, i.e., $\kappa_{x,2} = 1/4 > 0$ (attained at $\boldsymbol{v} = \boldsymbol{e}_1$).

Consider the idealized learner that knows $\boldsymbol{\mu}$ and in every round samples $\boldsymbol{x}_t$ uniformly from the exact Pareto set $\{\boldsymbol{x} \in \mathcal{X}_t : \nexists\, \boldsymbol{x}' \in \mathcal{X}_t,\ \boldsymbol{\mu}_{i_t}(\boldsymbol{x}') \succ \boldsymbol{\mu}_{i_t}(\boldsymbol{x})\}$, i.e., the default rule with $\boldsymbol{U}_t$ replaced by the true means (perfect estimates, zero widths). Every selected arm is Pareto optimal, so $\Delta_{i_t}(\boldsymbol{x}_t) = 0$ by Appendix E.1 and the PSG regret is zero. When both arms are present, which happens with probability $1/2$, neither dominates the other, both are sampled with probability $1/2$, and choosing $\boldsymbol{e}_2$ costs $\Delta_t^{\boldsymbol{\omega}} = 0.9 - 0.1 = 0.8$; otherwise $\Delta_t^{\boldsymbol{\omega}} = 0$. Thus, in every round in which both arms are present, the expected scalarized regret of the round is 0.4, the value quoted in Appendix G.2; averaged over the candidate sets it is 0.2. The candidate sets are i.i.d. and the sampling is independent across rounds, so $R_T^{\boldsymbol{\omega}} = 0.8\, N_T$ with $N_T \sim \mathrm{Binomial}(T, 1/4)$. Hence $\mathbb{E}[R_T^{\boldsymbol{\omega}}] = T/5$, and Hoeffding's inequality gives $\mathbb{P}(N_T \le T/8) \le \exp(-2T(1/8)^2) = \exp(-T/32)$, so $R_T^{\boldsymbol{\omega}} > T/10$ with probability at least $1 - \exp(-T/32)$. If instead $\mathcal{X}_t = \{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ in every round, as in the example of Appendix G.2, the same computation gives $R_T^{\boldsymbol{\omega}} = 0.8\, N_T'$ with $N_T' \sim \mathrm{Binomial}(T, 1/2)$, that is, 0.4 per round in expectation. This deterministic candidate set does not consist of i.i.d. draws as required by (A1), and the analogue of $\kappa_{x,2}$ computed with it is zero, because $\min_{a\le 2}\langle \boldsymbol{e}_1, \boldsymbol{x}_{t,a}\rangle^2 = 0$ in every round; this is why the candidates are randomized above for the comparison with Appendix E.3.3(b). On the randomized instance, Appendix E.3.3(b) bounds the scalarized regret of COMOLB with rule (73), with probability at least $1-\delta$ (on $\mathcal{E}^{\mathrm{lin}}$), by a sublinear function of T: for any fixed $\gamma > 0$, u and δ, both T_{lin} and $\bar{\beta}_{\mathrm{lin}}^2$ grow only logarithmically in T, so (77) is of order $\sqrt{T}\log T$. Thus no scalarized-regret guarantee can be obtained for uniform sampling from the Pareto set through estimation accuracy alone: the selection rule itself must use the preferences.

E.4 Information Advantage of Objective-Specific Pooling

For a user i, let $C^{\cap}(i) := \bigcap_{k=1}^{m_o} C^k(i)$. The sets $C^{\cap}(i)$, $i \in \mathcal{U}$, are the cells of the common refinement of the partitions $\mathcal{C}^{(1)}, \dots, \mathcal{C}^{(m_o)}$: if a user l belongs to $C^{\cap}(i) \cap C^{\cap}(i')$, then $C^k(i) = C^k(l) = C^k(i')$ for every k, so $C^{\cap}(i) = C^{\cap}(i')$. A single partition $\mathcal{P}$ of $\mathcal{U}$ that is used for all objectives never pools users from different true clusters of any objective if and only if, for every user i, its cell $\mathcal{P}(i)$ containing i satisfies $\mathcal{P}(i) \subseteq C^k(i)$ for every k, i.e., $\mathcal{P}(i) \subseteq C^{\cap}(i)$. Thus $C^{\cap}(i)$ is the largest group over which a common partition can pool the data of user i without mixing clusters.

Information advantage. Fix a round $t \in \{0, 1, \dots, T\}$ and any realization of the history $(i_s, \boldsymbol{x}_s)_{s\le t}$, generated by an arbitrary policy. For every user i and every objective k, $C^{\cap}(i) \subseteq C^k(i)$,

and the statistics (69) computed from this history satisfy

$$\boldsymbol{M}^k_{C^k(i),t} \succeq \boldsymbol{M}^k_{C^\cap(i),t}, \tag{79}$$

and hence, for every context $\boldsymbol{x}$,

$$\|\phi^k(\boldsymbol{x})\|_{(\boldsymbol{M}^k_{C^k(i),t})^{-1}} \le \|\phi^k(\boldsymbol{x})\|_{(\boldsymbol{M}^k_{C^\cap(i),t})^{-1}}. \tag{80}$$

In the linear case the feature map is the identity.

Indeed, the inclusion $C^\cap(i) \subseteq C^k(i)$ holds because $C^k(i)$ is one of the sets in the intersection; the intersection contains i and is therefore nonempty. By (69),

$$\boldsymbol{M}^k_{C^k(i),t} - \boldsymbol{M}^k_{C^\cap(i),t} = \sum_{s\le t:\ i_s\in C^k(i)\setminus C^\cap(i)} \phi^k(\boldsymbol{x}_s)\phi^k(\boldsymbol{x}_s)^\top \succeq \boldsymbol{0},$$

a sum of positive semi-definite rank-one matrices; this is (79). For (80), write $\boldsymbol{A} := \boldsymbol{M}^k_{C^k(i),t}$ and $\boldsymbol{B} := \boldsymbol{M}^k_{C^\cap(i),t}$; both satisfy $\boldsymbol{A} \succeq \boldsymbol{B} \succeq \lambda \boldsymbol{I}$, so they are positive definite and invertible. Congruence preserves the Loewner order, so $\boldsymbol{B}^{-1/2}\boldsymbol{A}\boldsymbol{B}^{-1/2} \succeq \boldsymbol{I}$; all eigenvalues of this symmetric positive definite matrix are at least 1, hence all eigenvalues of its inverse $\boldsymbol{B}^{1/2}\boldsymbol{A}^{-1}\boldsymbol{B}^{1/2}$ are at most 1, i.e., $\boldsymbol{B}^{1/2}\boldsymbol{A}^{-1}\boldsymbol{B}^{1/2} \preceq \boldsymbol{I}$. Congruence with $\boldsymbol{B}^{-1/2}$ gives $\boldsymbol{A}^{-1} \preceq \boldsymbol{B}^{-1}$, and therefore $\|\boldsymbol{v}\|^2_{\boldsymbol{A}^{-1}} = \boldsymbol{v}^\top\boldsymbol{A}^{-1}\boldsymbol{v} \le \boldsymbol{v}^\top\boldsymbol{B}^{-1}\boldsymbol{v} = \|\boldsymbol{v}\|^2_{\boldsymbol{B}^{-1}}$ for every $\boldsymbol{v}$; take $\boldsymbol{v} = \phi^k(\boldsymbol{x})$.

The same argument applies to every cell $\mathcal{P}(i) \subseteq C^\cap(i)$ of a common partition that does not mix clusters: on any fixed history, $\boldsymbol{M}^k_{\mathcal{P}(i),t} \preceq \boldsymbol{M}^k_{C^\cap(i),t} \preceq \boldsymbol{M}^k_{C^k(i),t}$. The difference $\boldsymbol{M}^k_{C^k(i),t} - \boldsymbol{M}^k_{C^\cap(i),t}$ is exactly the information contributed by users who share the cluster of i for objective k but differ from i in some other objective; objective-specific pooling uses these data for objective k, whereas a common partition cannot use them without mixing clusters.

What the comparison does not show. The comparison above involves two statistics computed from one and the same history. It does *not* prove an instance-wise ordering of the regrets of two adaptive algorithms, for example one that pools over $C^k(i)$ and one that pools over $C^\cap(i)$: the selected contexts depend on the pooled statistics, so the two algorithms generate different histories, and the comparison says nothing about statistics built from different histories. Moreover, it orders the geometric factor $\|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}}$ only. The data-dependent radius $\beta(\boldsymbol{M})$ is nondecreasing in $\boldsymbol{M}$ in the Loewner order (because $\boldsymbol{A} \succeq \boldsymbol{B}$ with $\boldsymbol{B}$ positive definite implies $\det \boldsymbol{A} \ge \det \boldsymbol{B}$), so the data-dependent widths $\beta(\boldsymbol{M})\|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}}$ are not ordered by it; what is ordered are the widths $\bar\beta\,\|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}}$ used in the analysis, in which the radius is replaced by the group-independent constant $\bar\beta_T$ or $\bar\beta_{\mathrm{lin}}$. Statements that one pooling scheme incurs less regret than the other are therefore comparisons of upper bounds or empirical findings, not consequences of this comparison.

E.5 How These Results Are Used in the Experiments

The results of this appendix determine how the experimental comparisons are read. First, by (14), a learner that selects arms by a single objective k_0, ignoring the other objectives in the selection step, has PSG regret no larger than its regret on k_0. When single-objective baselines are reported in Section 5, their PSG therefore indicates how small the PSG can be made without using the other objectives in the selection, and it says nothing about the remaining objectives. A small PSG of such a baseline is consistent with (14) and does not indicate performance on the other objectives, so a multi-objective method should not be judged against these baselines on the PSG alone; criteria that involve every objective, such as the scalarized regret, are needed for that comparison. A single-objective baseline that selects $\boldsymbol{x}_t \in \arg\max_{\boldsymbol{x}\in\mathcal{X}_t} U^{k_0}_t(\boldsymbol{x})$ is rule (73) with $\boldsymbol{\omega}_t \equiv \boldsymbol{e}_{k_0}$; when it uses the radius $\beta(\bar{\boldsymbol{M}}^k_t)$ and the unscaled deletion tests, the paragraph following the argument of Appendix E.3.3 gives its guarantee.

Second, the preference-conditioned protocol replaces only the selection step by rule (73), with weights fixed by the platform. When the UCBs use the radius $\beta(\bar{\boldsymbol{M}}^k_t)$ and the deletion tests are unscaled, as in COMOLB and COMONB-Frozen, Appendix E.3.3 bounds the scalarized regret of this protocol under the conditions of the respective guarantee, and, for strictly positive weights, Appendix E.3.1 together with (R1)–(R3) preserves the PSG guarantee. Implementations with α-scaled deletion thresholds

($\alpha < 1$) or COMONB-Adam lie outside the scope of these guarantees, and neither guarantee applies to them: for such runs, conclusions (i) and (ii) of the guarantees or the confidence bounds behind the UCBs are not proved, so the guarantee of Appendix E.3.3 motivates the protocol but does not cover the results, which are empirical. Appendix E.3.1, which concerns only the selection step, holds for every vector of UCBs; for strictly positive weights it shows that the selected arm lies in the Pareto set of the UCBs actually computed in every run, which is a membership property, not a regret bound. By Appendix E.3.4, the default uniform rule has no scalarized-regret guarantee, even with perfect estimates; the scalarized regret is therefore the criterion covered by a guarantee only for the preference-conditioned rule, and the default rule is assessed by the PSG. Third, Appendix E.4 compares pooled statistics on a common history; by Appendix E.4, the experimental comparison between pooling over objective-specific clusters and pooling over the common refinement is an empirical complement to the comparison of Appendix E.4 rather than a consequence of it.

F Extended Related Work

This appendix expands the related-work paragraph of the introduction.

Clustering of Bandits. The paradigm of clustering bandits (CB) was pioneered by Gentile et al. (2014), who introduced graph-based structures to dynamically cluster users based on observed interactions. Building on this foundation, Li et al. (2016) incorporated collaborative filtering mechanisms among items to further enhance clustering efficacy. Subsequent research has extended the CB framework to more complex environments: Li & Zhang (2018) adapted clustering to cascading bandits with partial list feedback, while Li et al. (2019) addressed the challenges posed by heterogeneous user arrival patterns. More recently, the focus has shifted towards decentralized and robust learning. Liu et al. (2022) proposed a federated learning approach to enhance privacy and communication efficiency, while Wang et al. (2023a;b) developed algorithms robust to model misspecification and adversarial data corruption. Most recently, Wang et al. (2025) expanded the scope to preference-based learning by introducing clustering algorithms for dueling bandits. Instead of a global partition, Ban & He (2021) identify overlapping local clusters around seed users, and Ban et al. (2024b) adapt a meta-learner to the dynamic cluster of each arriving user with neural reward models.

Neural Bandits. Although linear bandit algorithms have achieved both theoretical and practical success (Auer, 2002; Dani et al., 2008; Abbasi-Yadkori et al., 2011; Chu et al., 2011; Hu et al., 2021; He et al., 2022), their reliance on strict linearity limits their utility in complex real-world applications. To mitigate this, earlier works explored nonlinear frameworks like GLMs (Filippi et al., 2010) and kernel-based bandits (Valko et al., 2013), yet these approaches still depend on specific structural forms of the reward. The advent of deep learning offered a more flexible alternative. Initial empirical studies (Riquelme et al., 2018; Zahavy & Mannor, 2019) demonstrated the potential of neural networks in bandits, though they lacked formal guarantees. Following breakthroughs in neural network theory, specifically regarding the NTK regime (Jacot et al., 2018; Cao & Gu, 2019), neural contextual bandits have since gained a rigorous theoretical footing. Recent works have established provable regret bounds (Zhou et al., 2020; Zhang et al., 2021; Salgia, 2023; Xu et al., 2022; Hwang et al., 2023; Dai et al., 2023; Kassraie & Krause, 2022; Gu et al., 2024; Nabati et al., 2021; Ban et al., 2022; Jia et al., 2022; Nguyen-Tang et al., 2022; Zhu et al., 2021; Kassraie et al., 2022; Dai et al., 2022; Qi et al., 2023a;b; Wang et al., 2025), paving the way for applications in complex domains such as recommendation systems and large language models (Chen et al., 2024; Ban et al., 2024a).

Multi-Objective Bandits. The Multi-Objective Multi-Armed Bandit (MOMAB) framework was formally introduced by Drugan & Nowé (2013), who utilized Pareto dominance to define optimality and established an upper regret bound of $O(K \log T)$. This foundation was later expanded to contextual settings: Turğay et al. (2018) assumed Lipschitz continuous rewards, while Lu et al. (2019) integrated Generalized Linear Models (GLMs) into the multi-objective framework. Other significant theoretical advancements include the analysis of adversarial scenarios (Xu & Klabjan, 2023) and the study of Pareto-optimal arm identification under budget constraints (Van Moffaert et al., 2014; Auer et al., 2016; Kone et al., 2023). Beyond standard Pareto dominance, recent literature has explored lexicographic and hierarchical relationships among objectives (Cheng et al., 2024; Xue et al., 2024; 2025). However, existing contextual/structured multi-objective bandit methods typically impose

linear, GLM, Lipschitz, and scalarization, and do not model objective-dependent collaborative user clusters.

What is new relative to prior work. Our contribution lies in the new structural ingredient of **objective-specific clustering**. Prior clustering bandit methods use a *single similarity structure* over users for the whole reward, whether a global partition or local neighborhoods (Ban & He, 2021; Ban et al., 2024b); prior multi-objective bandits do not exploit objective-dependent collaborative structure; and prior NTK bandits do not analyze multi-objective user clustering. In our setting, different coordinates of the same Pareto-UCB rule are computed from different cluster structures and different covariance matrices, which creates a setting that is not covered by previous single-graph analyses. This structural change propagates through modeling, algorithm design, and regret analysis. In the experiments, the I-CLUB, S-CLUB and Shared-LOCB baselines and the Oracle-Global reference ablate it.

G Additional Notes and Discussion

This appendix collects notes on the standing conditions and on the guarantees, and the complete discussion of the experiments, which are condensed in the main text.

G.1 Notes on the Standing Conditions

Size of $\kappa_{x,K}$. Let $\lambda_x := \lambda_{\min}(\mathbb{E}_{\boldsymbol{x}\sim\mathcal{D}}[\boldsymbol{x}\boldsymbol{x}^\top])$. Keeping only $a = 1$ inside the minimum in (5) shows $\kappa_{x,K} \le \lambda_x$, and Appendix C.1.1 shows $\kappa_{x,K} \ge (\lambda_x/2)\big(\lambda_x/(2-\lambda_x)\big)^K$. Hence (A4) holds if and only if $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top]$ is nonsingular, but $\kappa_{x,K}$ can be exponentially small in K: if $d \ge 2$ and $\mathcal{D}$ is uniform on $\{\pm\boldsymbol{e}_1, \dots, \pm\boldsymbol{e}_d\}$, then $\lambda_x = 1/d$, whereas for $\boldsymbol{v} = \boldsymbol{e}_1$ the minimum in (5) is nonzero only if all K candidates lie in $\{\pm\boldsymbol{e}_1\}$, so that $\kappa_{x,K} \le d^{-K}$. For spherical contexts ($\boldsymbol{x} = \boldsymbol{g}/\|\boldsymbol{g}\|_2$ with $\boldsymbol{g} \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{I}_d)$ and $d \ge 3$), in contrast, the decay is only polynomial: $\kappa_{x,K} \ge \pi/(16(d-1)K^2)$ (see (13), shown in Appendix C.5).

Scope of the fixed-feature model. (i) Anti-concentration of the contexts does not transfer to the features. Let $\mathcal{D}$ be the spherical distribution of the previous paragraph, for which $\kappa_{x,K} \ge \pi/(16(d-1)K^2)$, and consider a single objective with the one-dimensional ReLU feature $\phi^1(\boldsymbol{x}) = \max\{x_1, 0\}$. By the symmetry of x_1 and $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top] = \boldsymbol{I}_d/d$, $\boldsymbol{\Sigma}^1_\phi = \mathbb{E}[\max\{x_1,0\}^2] = \mathbb{E}[x_1^2]/2 = 1/(2d)$, so $\mathcal{S}_1 = \mathbb{R}$ and $\lambda_\phi = 1/(2d)$. Since $\phi^1(\boldsymbol{x}) = 0$ with probability $1/2$ and $\phi^1(\boldsymbol{x})^2 \le 1$, the minimum $\min_{a\in[K]} \phi^1(\boldsymbol{x}_a)^2$ is positive only if all K candidates have a positive first coordinate, so $\kappa_{\phi,K} \le 2^{-K}$, although λ_ϕ does not depend on K. Freezing the features of a ReLU network therefore does not, in general, remove the exponential dependence on K of $\kappa_{\phi,K}$, and hence of the identification term in the bound of Section 4.2. (ii) (A5) is a realizability condition on the chosen features. The quadratic and cosine rewards of our nonlinear experiments (Section 5.2) need not lie in the span of the frozen features; those experiments are misspecification tests, not covered by the guarantee of Section 4.2. The MovieLens rewards (Section 5.3) are linear in the context: the exact version satisfies (A5) with $\phi^k(\boldsymbol{x}) = \boldsymbol{x}$ and $B = 3.62$, and departs from the conditions of Section 4.1 only through parameter norms above 1; the raw version is only approximately clustered.

G.2 Notes on the Algorithms and the Guarantees

Kernel form of COMONB-Frozen. All quantities depend on the features only through the kernel $\mathcal{K}^k(\boldsymbol{x}, \boldsymbol{x}') = \langle \phi^k(\boldsymbol{x}), \phi^k(\boldsymbol{x}')\rangle$. For the samples S of one user or of a pooled set $\widehat{C}_t^k$, with Gram matrix $\boldsymbol{K}_S$, kernel vector $\boldsymbol{k}_S(\boldsymbol{x})$ and rewards $\boldsymbol{r}_S$, the prediction is $\boldsymbol{k}_S(\boldsymbol{x})^\top(\boldsymbol{K}_S + \lambda\boldsymbol{I})^{-1}\boldsymbol{r}_S$, the squared geometric width is $s_S^2(\boldsymbol{x}) = \lambda^{-1}[\mathcal{K}^k(\boldsymbol{x},\boldsymbol{x}) - \boldsymbol{k}_S(\boldsymbol{x})^\top(\boldsymbol{K}_S + \lambda\boldsymbol{I})^{-1}\boldsymbol{k}_S(\boldsymbol{x})]$, the bias part of the deletion width uses $\|\boldsymbol{M}_S^{-1}\phi^k(\boldsymbol{x})\|_2^2 = \lambda^{-1}[s_S^2(\boldsymbol{x}) - \|(\boldsymbol{K}_S + \lambda\boldsymbol{I})^{-1}\boldsymbol{k}_S(\boldsymbol{x})\|_2^2]$, and the log-determinant in (9) is $\log\det(\boldsymbol{I} + \lambda^{-1}\boldsymbol{K}_S)$ (see (19) and Appendix D.10). The linear algebra thus involves $|S| \times |S|$ instead of $p_k \times p_k$ matrices, although each kernel evaluation still costs $O(p_k)$ for gradient features.

Platform-specified trade-offs. The paragraph on platform-specified trade-offs in Section 3.2 states, in full, the following. The analyses of both algorithms use the selection step only through $\boldsymbol{x}_t \in \tilde{\mathcal{O}}_t$ and through the requirement that $\boldsymbol{x}_t$ be chosen before the reward of round t is revealed without invalidating the arrival, candidate and noise conditions, formalized as (R1)–(R3) in Appendix E.3. Suppose the platform supplies, before the selection in round t, weights $\boldsymbol{\omega}_t$ in the probability simplex with strictly positive entries that are exogenous in the sense of condition (E) (e.g., a deterministic schedule, or $\boldsymbol{\omega}_t$ a measurable function of the past, of i_t and of $\mathcal{X}_t$), and the algorithm selects $\boldsymbol{x}_t \in \arg\max_{\boldsymbol{x}\in\mathcal{X}_t} \sum_k \omega_{t,k} U_t^k(\boldsymbol{x})$, with ties broken using only the past, i_t, $\mathcal{X}_t$ and $\boldsymbol{\omega}_t$, and with all estimators and graphs unchanged. This arm lies in $\tilde{\mathcal{O}}_t$: if $\boldsymbol{U}_t(\boldsymbol{x}') \succ \boldsymbol{U}_t(\boldsymbol{x}_t)$, then $\boldsymbol{\omega}_t^\top \boldsymbol{U}_t(\boldsymbol{x}') > \boldsymbol{\omega}_t^\top \boldsymbol{U}_t(\boldsymbol{x}_t)$ because all weights are positive, contradicting the choice of $\boldsymbol{x}_t$. Hence the PSG bounds of Sections 4.1 and 4.2 hold unchanged for every such schedule, and the same per-objective estimates can serve different or time-varying trade-offs; Appendix E.3.3 bounds the scalarized regret. Neither the PSG bound nor this rule guarantees coverage of, or balanced exposure over, the Pareto set; such requirements must be imposed through the choice of $\boldsymbol{\omega}_t$.

Sublinearity and scope of the COMONB-Frozen guarantee. Sufficient conditions for $R_T = o(T)$ are $T_{\mathrm{clust}} = o(T)$ and $m_c \bar{\beta}_T^2 G_T = o(T)$. Both hold for a fixed finite-dimensional feature map, i.e., when ϕ^k, $\mathcal{D}$, K, u, m_o, δ, λ, R, B and the cluster parameters do not change with T: then $G_T = O(\log T)$, hence $\bar{\beta}_T^2 = O(\log T)$ and $T_{\mathrm{clust}} = O(\log T)$, and the bound in (iii) grows as $O(\sqrt{T}\log T)$. If the network width is chosen as a function of T, the feature map changes with T, and with it the feature distribution, r and λ_ϕ; the fixed-distribution argument then no longer applies. The guarantee further requires rewards that are exactly linear in the frozen features and exact cluster structure. The quadratic and cosine rewards of Section 5.2 need not be linear in the frozen features, so the results obtained there test the algorithms beyond the scope of the theory. The rewards of the MovieLens environment of Section 5.3 are linear in the context, with parameter norms up to $B = 3.62$. Its exact version therefore satisfies the conditions of Section 4.2 with the identity features $\phi^k(\boldsymbol{x}) = \boldsymbol{x}$, which have $\|\phi^k(\boldsymbol{x})\|_2 \le 1$ on the catalogue, and (A6) holds because the catalogue is finite and the cluster reward functions differ; it lies outside the COMOLB guarantee as stated, which requires $B = 1$. The raw version is only approximately clustered, so neither guarantee applies to it.

The role of K in the identification cost. In both guarantees, K enters only the identification cost: through $\kappa_{x,K}$ in Section 4.1, and through $\kappa_{\phi,K}$ and q_γ in Section 4.2, where a larger K makes a separating candidate more likely. Because the analysis must hold for every selection rule, $\kappa_{x,K}$ measures the information of the least informative of the K candidates, which lower-bounds that of the selected context; it can therefore decrease with K. In general, $\kappa_{x,K} \ge (\lambda_x/2)\big(\lambda_x/(2-\lambda_x)\big)^K$ with $\lambda_x = \lambda_{\min}(\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top])$ (Appendix C.1.1), a lower bound that is exponentially small in K. For spherical contexts the dependence is polynomial, see (13): for this distribution $\mathbb{E}[\boldsymbol{x}\boldsymbol{x}^\top] = \boldsymbol{I}/d$, so the general bound evaluates to $(2d)^{-1}(2d-1)^{-K}$, whereas the exact value $\kappa_{x,K} = \int_0^1 \big[1 - I_s\big(\frac{1}{2}, \frac{d-1}{2}\big)\big]^K ds$, with I_s the regularized incomplete beta function, satisfies $\kappa_{x,K} \sim 1/(2c_d^2 K^2)$ as $K \to \infty$ for fixed d; the lower bound of (13) is therefore tight up to a constant factor (Appendix C.5). For frozen network features, $\kappa_{\phi,K}$, and hence the identification term of the upper bound in Section 4.2, can still depend exponentially on K, because anti-concentration of $\boldsymbol{x}$ need not transfer to $\phi^k(\boldsymbol{x})$ (Appendix G.1); we give no lower bound for it.

Identification at practical horizons. The identification cost is independent of T up to logarithmic factors, but it can be large. Ignoring logarithmic factors, the identification term $2T_{\mathrm{lin}}$ of (12) exceeds the learning cost $d\sqrt{m_c T}$ unless $T \gtrsim u^2(1 + d/\gamma^2)^2/(\kappa_{x,K}^2 d^2 m_c)$, which for spherical contexts reads $T \gtrsim u^2 K^4 (1 + d/\gamma^2)^2/m_c$. At practical horizons the bounds can therefore be dominated by the identification term, and the rate $d\sqrt{m_c T}$ describes only the regime after recovery. Conversely, T_{lin} and T_{clust} are worst-case upper bounds, and in a given instance the graphs may recover much earlier.

Selection rules covered by the analysis. The high-probability event used in the analyses of Sections 4.1 and 4.2 is valid for every selection rule that satisfies conditions (R1)–(R2) of Appendix E.3: $\boldsymbol{x}_t \in \mathcal{X}_t$ is determined by information available before the reward of round t is revealed, given which the noise remains conditionally zero-mean and R-sub-Gaussian, and of which future arrivals and

candidate sets remain independent. The preference rule with exogenous weights and ties broken as in Section 3.2 satisfies these conditions, and the guarantee of Appendix E.3.3 holds on the same event. This guarantee concerns the $\arg\max$ rule; uniform sampling from $\tilde{\mathcal{O}}_t$ has no such guarantee. If every candidate set consists of two arms with reward vectors $(1, 0)$ and $(0, 1)$, both are Pareto-optimal, and for $\boldsymbol{\omega}_t = (0.9, 0.1)$ uniform sampling incurs an expected scalarized regret of 0.4 per round even with known rewards, i.e., linear scalarized regret (Appendix E.3.4 gives an instance with random candidate sets that satisfies the conditions of Section 4.1). For strictly positive weights, the selected arm lies in $\tilde{\mathcal{O}}_t$, so the PSG bounds continue to hold (Section 3.2).

On the information advantage (Appendix E.4). On the same collected data, pooling within the objective-specific cluster yields a geometric factor $\|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}}$ that is never larger than the one obtained by pooling within $C^{\cap}(i)$. The same holds for the widths $\bar{\beta}\,\|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}}$ with the common radius $\bar{\beta} = \bar{\beta}_T$ or $\bar{\beta} = \bar{\beta}_{\mathrm{lin}}$ used in the analysis, but not for the data-dependent widths $\beta(\boldsymbol{M})\|\phi^k(\boldsymbol{x})\|_{\boldsymbol{M}^{-1}}$ of the algorithms, since $\beta(\boldsymbol{M})$ is nondecreasing in $\boldsymbol{M}$ in the Loewner order (Appendix E.4). That comparison involves quantities computed from one and the same history; since two adaptive algorithms collect different histories, it does not by itself imply an instance-wise ordering of their regrets.

G.3 Complete Discussion of the Experiments

This subsection discusses the results of Section 5 in more detail; the numbers are in the tables cited, and all comparisons "in every run" are paired (Appendix B).

Shared-component instance: structure. Oracle-OS lowers the regret of Ind by a factor of about eight to nine at every γ, and Oracle-Global only by a factor of about 2.4 (Table 1): since the two partitions are drawn independently, their common refinement has small cells, with fewer than three users on average against ten users per objective-specific cluster, so a shared partition that is correct for every objective captures only a small part of the value of objective-specific pooling. Pool-all ignores the clusters entirely; it is slightly worse than Ind for $\gamma = 0.2$ and far worse for large γ.

Shared-component instance: learning the clusters. COMOLB, I-CLUB and S-CLUB never deleted a same-cluster edge, and COMOLB and I-CLUB recovered their target partitions in every run (Table 4); S-CLUB, whose single graph targets the common refinement, recovered it in all runs only for $\gamma \geq 1$. Before recovery, every graph-based learner pools each user with some neighbours from other clusters, and for COMOLB this phase accounts for almost all of its regret. Its length shrinks by more than an order of magnitude as γ grows from 0.2 to 1.4, but its regret per round grows faster, because the wrongly pooled neighbours are more different, so the regret before recovery exceeds the entire regret of Ind for $\gamma \geq 1$. After recovery, COMOLB learns at the rate of Oracle-OS, an order of magnitude more slowly than Ind (Figure 1); its advantage over Ind for $\gamma = 0.2$ holds in 9 of 10 runs, for $\gamma = 0.5$ its mean regret falls below that of Ind before the midpoint of the horizon, and for $\gamma \geq 1$ Ind is better in every run. I-CLUB incurs a similar regret before recovery but afterwards pools only within the cells of the common refinement, so its regret keeps growing several times faster than that of COMOLB; the local-clustering adaptations pool only within the group of a single seed, and OS-LOCB, which keeps one structure per objective, is the closest competitor. COMOLB has a lower regret than each of I-CLUB, OS-LOCB and Shared-LOCB in at least 9 of 10 runs at every γ.

Ablations. Pooling over the connected component instead of the closed neighbourhood leaves the recovery times almost unchanged but raises the regret of COMOLB by 23% to 33%, because before recovery a single remaining cross-cluster edge merges two clusters into one component. Replacing, in addition, the radius (10) by the radius $\beta/\sqrt{\lambda_{\min}}$ implied by the confidence ellipsoid delays recovery by a factor of about two to three and raises the regret by 63% to 227%; for $\gamma = 0.2$ this variant deleted no edge within the horizon and coincides with Pool-all. The practical deletion scale $\alpha = 0.3$, which the guarantee of Section 4.1 does not cover, recovers several times faster and attains the lowest PSG of all non-oracle learners with Pareto selection at every γ, between 16% and 52% of the regret of Ind, but deletes about a quarter of the same-cluster edges per run; the clusters nevertheless remain connected in most runs. A less conservative deletion test would therefore realize much of the structural gain, which is the main open problem left by our analysis.

Table 8: Fixed-feature and nonlinear rewards, complete results: cumulative PSG (mean ± std over 10 runs). Frozen: exactly realizable in the learners' frozen features, $T = 10^5$. Quadratic and cosine: misspecified for all learners, reported at $T = 5000$ and $T = 2 \times 10^4$; COMONB-Adam was run for 5000 rounds only. All learners use the frozen features of a network of width 8 (frozen) or 16 (quadratic, cosine), except the two rows marked "raw contexts", which use the identity map on the raw contexts. The rows below the second rule use the deletion scale $\alpha = 0.3$ or a trained network and are not covered by our guarantees.

	Frozen	Quadratic		Cosine	
Learner	$T = 10^5$	$T = 5000$	$T = 2 \times 10^4$	$T = 5000$	$T = 2 \times 10^4$
Oracle-OS	233.2± 46.9	179.3± 17.2	474.5± 49.8	416.7± 25.3	1365.6± 89.8
Oracle-Global	622.1± 57.6	–	–	–	–
Ind	742.1± 70.0	325.4± 21.6	897.2± 58.2	515.3± 31.6	1720.6± 89.2
Pool-all	3665.4± 1182.1	458.0± 61.2	1728.9± 244.9	617.9± 74.0	2311.3± 296.9
COMONB-Frozen	521.6± 72.8	457.4± 61.3	1213.1± 114.8	526.4± 34.3	1505.2± 84.6
I-CLUB (one graph)	784.7± 87.1	457.8± 61.3	1243.4± 113.1	537.9± 35.1	1677.2± 91.5
COMOLB (raw contexts)	1664.5± 761.2	293.9± 29.2	706.9± 57.8	452.8± 31.7	1638.3± 108.2
Oracle-OS (raw contexts)	–	148.4± 10.2	542.8± 49.9	439.2± 37.2	1703.5± 127.3
COMONB-Frozen, $\alpha = 0.3$	249.8± 41.6	252.3± 23.0	590.9± 53.9	483.7± 32.7	1705.3± 83.0
COMONB-Adam	–	195.8± 11.2	–	444.4± 19.1	–

Selection by a fixed trade-off. In the bottom block of Table 1, every learner selects $\boldsymbol{x}_t \in \arg\max_{\boldsymbol{x} \in \mathcal{X}_t} \sum_k \omega_k U_t^k(\boldsymbol{x})$ with $\boldsymbol{\omega} = (0.5, 0.5)$, the rule of Section 3.2, and S-CLUB selects in the same way from the UCB of its scalarized estimate. COMOLB has a lower scalarized regret than I-CLUB and S-CLUB in every run at every γ, by 32% to 53% relative to S-CLUB, and relative to Ind with the same selection rule it is 35% and 9% lower for $\gamma = 0.2$ and 0.5 and higher for $\gamma \geq 1$, as with Pareto selection. With this selection rule, COMOLB and I-CLUB deleted no same-cluster edge and recovered their target partitions in every run, except two runs of COMOLB for $\gamma = 0.2$ that had not recovered at T.

Crossed instance. In the crossed instance (Table 5), the common refinement is the singleton partition, so Oracle-Global coincides with Ind, whereas Oracle-OS pools within $2n$ groups of n users and lowers the regret of Ind by a factor that grows with u. With $\gamma = 0.8$, the instance lies between the separations at which COMOLB beats Ind and loses to it on the shared-component instance. COMOLB never deleted a same-cluster edge, recovered the partitions in every run and then learned at the rate of Oracle-OS, but its regret at $T = 2000u$ is 7% to 33% above that of Ind, almost all of it incurred before recovery. It has a lower regret than I-CLUB and Shared-LOCB in every run and than OS-LOCB in most runs.

Theory-matched fixed-feature instance. The rewards are exactly linear in the frozen gradient features of a random one-hidden-layer ReLU network of width 8 on $d = 5$ inputs ($p = 57$), which all learners of this experiment use, so that (A5) holds ($u = 20$, $m_o = 2$, $m_c^{(k)} = 4$, $K = 10$, $T = 10^5$). The features span a subspace of dimension 49 whose smallest positive eigenvalue is of order 10^{-7}, so the bound of Section 4.2 on the identification time is vacuous at any practical horizon. Empirically (Table 8), COMONB-Frozen never deleted a same-cluster edge, recovered the partitions in 9 of 10 runs, and attained a regret 30% below Ind and 16% below Oracle-Global in every run; over the second half of the horizon its regret grew as slowly as that of Oracle-OS. The single-graph I-CLUB recovered the common refinement in 9 of 10 runs but pools only within its cells and ended above Ind. COMOLB applied to the raw contexts is misspecified here, deleted many same-cluster edges and ended far above Ind. With $\alpha = 0.3$, COMONB-Frozen recovered the partitions in all runs, an order of magnitude faster, with a single false deletion over all runs, and came close to Oracle-OS.

Misspecified rewards. The quadratic rewards $\mu(\boldsymbol{x}) = (\langle \boldsymbol{\theta}, \boldsymbol{x} \rangle + \boldsymbol{x}^\top \boldsymbol{S} \boldsymbol{x})/2$ and the cosine rewards $\mu(\boldsymbol{x}) = \cos(3\boldsymbol{v}^\top \boldsymbol{x} + \pi/4)$ are linear in neither feature map ($u = 20$, $m_o = 2$, $m_c^{(k)} = 4$, $d = 10$, $K = 10$, $T = 2 \times 10^4$; width-16 frozen features with $p = 193$), so none of our guarantees applies. On the quadratic rewards, COMONB-Frozen deleted no same-cluster edge but recovered the partitions in only 1 of 10 runs, so it ended 35% above Ind; with $\alpha = 0.3$ it deleted about a third of the same-cluster edges but ended 34% below Ind. On the cosine rewards, the test fired within clusters, which the theory

does not exclude for non-realizable rewards; COMONB-Frozen recovered in 3 of 10 runs and ended 13% below Ind, with a second-half regret close to that of Oracle-OS, whereas with $\alpha = 0.3$ it deleted almost all same-cluster edges and reduced to Ind. COMOLB on the raw contexts is misspecified as well and deleted same-cluster edges on both reward types, yet it attained the lowest regret among the learners with the analyzed test on the quadratic rewards and ended 5% below Ind on the cosine rewards. The trained-network COMONB-Adam, which selects greedily without an exploration bonus and is not covered by our analysis, came close to Oracle-OS on both reward types at $T = 5000$, at a per-round cost about thirty times that of Ind.

MovieLens. Table 2 gives the complete results. In the exact version the structural gap is again large: Oracle-OS attains less than a fifth of the regret of Ind and Oracle-Global less than two fifths. Learning the partitions is much harder than on the synthetic instances: the parameter norms are large relative to the separations, in particular for the second objective, whose cluster parameters are several times closer than those of the first, and the contexts are nearly degenerate, so the parameter-space radius (10) shrinks too slowly; COMOLB deleted no edge within the horizon and coincides with Pool-all, which is 9% below Ind in the exact version. The prediction-space test (11) with identity features only requires the predictions of two users to differ by more than their widths on one of the current candidates, which happens for users of different clusters of the first objective long before their parameters are resolved in every direction. COMONB deleted no same-cluster edge, cut every cross-cluster edge of the first objective in every run but only about one percent of those of the second objective; it attained a regret 46% below Ind and below I-CLUB, OS-LOCB and Shared-LOCB in every run, and over the second half of the horizon its regret was less than a third of that of Ind. In the raw version, users are not exactly clustered; the k-means partitions serve only the oracles, and deletions within a k-means cluster are not false deletions. Both oracle partitions and Pool-all are worse than Ind, and so is every clustering learner with Pareto selection; COMONB is 16% above Ind and the best of them in every run, because its test separates users whose predictions differ, including users in the same k-means cluster.

What the PSG certifies. The last rows of Table 2 illustrate (14): the learner that optimizes only novelty has the lowest mean PSG of all learners in both versions, below Oracle-OS in every run, although it ignores satisfaction entirely; the PSG therefore does not distinguish a learner that serves the trade-off from one that serves a single objective. The preference-conditioned rule of Section 3.2 addresses the question a platform actually asks. With $\boldsymbol{\omega} = (0.5, 0.5)$, COMONB selecting by $\boldsymbol{\omega}$ attains a scalarized regret about a third below S-CLUB, which selects by the same weights but pools over one graph on the scalarized estimates, in every run of both versions, and less than half of that of every learner with uniform Pareto or single-objective selection; in the exact version its scalarized regret per quarter of the horizon keeps decreasing, whereas that of S-CLUB stays almost constant after the first quarter. On the linear instance (Table 7), where all rules share the COMOLB estimators and graphs with the analyzed test, so that the guarantee of Appendix E.3.3 applies, each preference-conditioned rule attains the lowest scalarized regret for its own weight, whereas uniform sampling from the Pareto set, which suffices for the PSG guarantees, incurs a scalarized regret that grows linearly with T (Appendix B.6), as the counterexample in Appendix E.3.4 allows.