Fairness in recommender systems : insights from social choice

Virginie Do

▶ **To cite this version:**

Virginie Do. Fairness in recommender systems : insights from social choice. Other [cs.OH]. Université Paris sciences et lettres, 2023. English. NNT : 2023UPSLD007 . tel-04213955

HAL Id: tel-04213955
https://theses.hal.science/tel-04213955

Submitted on 21 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PSL
UNIVERSITÉ PARIS

THÈSE DE DOCTORAT
DE L'UNIVERSITÉ PSL

Préparée à l'Université Paris-Dauphine

Fairness in Recommender Systems: Insights from Social Choice

Soutenue par
Virginie Do
Le 11 juillet 2023

Ecole doctorale n° ED 543
Ecole doctorale SDOSE

Spécialité
Informatique

Composition du jury :

Yann CHEVALEYRE Professeur, Université Paris Dauphine-PSL	*Président*
Patrick LOISEAU Chercheur, INRIA	*Rapporteur*
Julia STOYANOVICH Associate Professor, New York University	*Rapporteure*
Isabel VALERA Professor, Saarland University	*Examinatrice*
Craig BOUTILIER Chercheur, Google	*Examinateur*
Nicolas USUNIER Chercheur, Meta	*Directeur de thèse*
Jérôme LANG Directeur de Recherche, CNRS	*Directeur de thèse*
Jamal ATIF Professeur, Université Paris Dauphine-PSL	*Directeur de thèse*
Sam CORBETT-DAVIES Chercheur, Meta	*Membre invité*

Dauphine | PSL
UNIVERSITÉ PARIS

Fairness in Recommender Systems: Insights from Social Choice

Virginie Do

Dissertation submitted in partial fulfillment
of the requirements for the degree of Doctor of Philosophy

Université Paris Dauphine – PSL University
in collaboration with Meta AI
2023

Abstract

Machine learning algorithms are widely used in the recommender systems that drive newsfeeds, streaming platforms, online marketplaces and social networking services. Their main purpose is to provide users with personalized recommendations by predicting their preferences and sorting available content based on those predictions. However, by selecting content from some producers over others, recommendation algorithms decide who is visible and who is not. These decisions have real ethical and societal implications, including the potential to overlook disadvantaged social groups when suggesting profiles to employers, or the possibility of certain voices and cultures being under- or over-represented on social media. It has therefore become crucial to ensure that these automated decisions are unbiased and fair towards content producers, avoiding giving some groups an excessive advantage or disadvantage. In addition to deciding which producers are visible, recommendation algorithms also control the information and opportunities that users are exposed to, including job and housing ads. Consequently, concerns have emerged about whether these algorithms provide fair access to information and opportunities among their users.

This thesis seeks to address the limitations of current recommendation algorithms by developing fairer systems that consider the welfare of both users and content producers. However, developing fair algorithms presents several challenges, including the definition of appropriate fairness criteria and the implementation of computationally efficient ranking algorithms that satisfy these criteria. Drawing on the rich literature of social choice theory, we propose a conceptual framework to assess the fairness of ranked recommendations, relying on established concepts for fair division problems that have been relatively overlooked by the machine learning and recommender systems communities. This framework guides the development of new recommendation methods that follow the principles of fair division, and distribute exposure more equitably among content producers, without compromising the quality of recommendations for users. These methods are supported by theoretical results on the fairness properties, convergence guarantees and computational efficiency of the proposed algorithms, as well as experimental evaluations on publicly available datasets.

Résumé en français

Les algorithmes d'apprentissage automatique (*machine learning*) sont largement utilisés dans les systèmes de recommandation qui alimentent les plateformes de streaming, de commerce et les réseaux sociaux. Leur principal objectif est de fournir aux utilisateurs des recommandations personnalisées en prédisant leurs préférences et en triant les contenus disponibles en fonction de ces prédictions. Cependant, en sélectionnant le contenu de certains producteurs plutôt que d'autres, les algorithmes de recommandation décident de qui est visible ou non. Ces décisions ont de réelles implications éthiques et sociales, comme les risques d'invisibilisation de groupes minoritaires ou défavorisés dans la suggestion de profils à des employeurs, ou les problèmes de sous- ou surreprésentation de certaines opinions et cultures sur les réseaux sociaux. Il est donc devenu crucial de garantir que

ces décisions automatisées soient non biaisées et équitables envers les producteurs de contenu, en évitant de donner à certains groupes un avantage ou un désavantage excessif. En plus de décider quels producteurs sont visibles, les algorithmes de recommandation jouent également un rôle clé dans la décision de quels utilisateurs sont exposés à certains contenus, notamment les contenus associés à des opportunités économiques telles que les offres d'emploi et annonces immobilières. Par conséquent, des préoccupations se posent quant à l'équité d'accès à ces opportunités parmi les utilisateurs des systèmes de recommandation.

Cette thèse vise à adresser les limites des algorithmes de recommandation actuels en développant des systèmes plus équitables qui tiennent compte à la fois des utilisateurs et des producteurs de contenu. Cependant, le développement d'algorithmes équitables présente plusieurs défis, notamment la définition de critères d'équité appropriés et l'implémentation efficace d'algorithmes de *ranking* qui satisfont ces critères. En nous appuyant sur la riche littérature de la théorie du choix social, nous proposons un cadre conceptuel pour évaluer l'équité des listes ordonnées de recommandations, à partir de concepts établis pour les problèmes de partage équitable qui ont été peu étudiés en *machine learning* et en recommandation. Dans ce cadre conceptuel, nous développons de nouvelles méthodes de recommandation qui suivent les principes du partage équitable et distribuent l'exposition plus équitablement entre les producteurs de contenu, sans compromettre la qualité des recommandations pour les utilisateurs. Ces méthodes sont soutenues par des résultats théoriques sur la satisfaction de propriétés d'équité, sur les garanties de convergence et l'efficacité algorithmique des algorithmes proposés, ainsi que par des évaluations expérimentales sur des jeux de données publics.

Acknowlegements

I would like to thank the members of my jury. Thank you Patrick Loiseau and Julia Stoyanovich for taking the time to carefully review this manuscript, and thank you Craig Boutilier, Yann Chevaleyre, Isabel Valera for agreeing to join in evaluating my work. What an honor!

I would like to express my sincere gratitude to my PhD advisors, in French. Jamal, merci pour tes conseils bienveillants, pour ta confiance et pour ton soutien pour que ma thèse se passe dans de bonnes conditions, surtout pendant les confinements. Merci pour l'environnement dynamique que tu as créé au sein de l'équipe MILES et de PRAIRIE, permettant aux jeunes scientifiques de faire des recherches de qualité dans une ambiance chaleureuse.

Jérôme, merci beaucoup de m'avoir ouvert les portes du choix social et d'avoir partagé avec moi ton expertise profonde et reconnue de ce domaine. Merci d'avoir été si disponible et généreux en excellents conseils tout au long de ma thèse (même si tu penses avoir été peu impliqué, ce que je démens). Même si nos travaux sur la sélection de comité ne figurent pas au cœur de ce manuscrit, j'ai beaucoup aimé travailler avec toi sur ces projets qui me tiennent à cœur.

Nicolas, ma reconnaissance est éternelle. Merci de m'avoir offert la chance de faire ma thèse à Meta. Merci pour tout ce que tu m'as transmis sur le plan scientifique, pour ton implication sans faille tout au long de ma thèse, pour tes conseils pour m'aider à devenir une chercheuse accomplie et pour ton soutien humain. Je suis si admirative et presque désemparée face à l'étendue de tes connaissances et la finesse de ta réflexion sur tant de sujets.

I wish to thank all my collaborators, with special recognition to Sam Corbett-Davies. Sam, thank you for your fantastic mentoring, particularly during the initial stages of my PhD. Your thought leadership and profound expertise in algorithmic fairness have been very valuable to me.

I am also grateful to Piotr Skowron and Matthieu Hervouin for the very pleasant collaboration on committee elections — my first course in computational social choice was taught by Piotr. Big thanks also to Maximilian Nickel and David Liu for letting me contribute to the group-free group fairness project. Thanks to my other Meta co-authors whom I learned a lot from: Matteo Pirotta, Alessandro Lazaric, and Elvis Dohmatob. Merci également à Thierry Kirat et Olivia Tambou dont j'ai beaucoup appris sur les aspects légaux de l'intelligence artificielle, ainsi qu'à Alexis Tsoukias.

Je remercie Meta pour les conditions extraordinaires dans lesquelles j'ai pu réaliser ma thèse. Je remercie mes collègues de Meta. Thank you Levent Sagun for the numerous enriching discussions. Merci à Antoine Bordes d'avoir rendu possible la thèse à FAIR. Merci à Jérémy Rapin pour ton mentorat. Merci à toute l'équipe de doctorants et ex-doctorants de FAIR Paris. Je remercie particulièrement mes thésard·es parallèles Lina, Charlotte, Guillaume ; les nouveaux Wes, Timothée D., Pierre F., Badr, Jean-Baptiste ; et les anciens Baptiste, Louis, Hubert et Léonard.

Je remercie mes collègues de Dauphine. Ceux qui m'ont accueillie : Laurent, Rafael, Alexandre

A., Florian Y. ; les grimpeurs Lucas et Alexandre V. ; et le Lang's gang : Tahar et Théo.

Je voudrais remercier chaleureusement mes ami·es qui ont continuellement égayé mes 3 années de thèse – elles et ils se reconnaîtront. Merci beaucoup à ma famille pour son soutien dans cette étape curieuse et imprévue dans mon parcours qu'est le doctorat. Enfin, merci à Clément d'avoir vécu au jour le jour l'aventure de la thèse avec moi, d'en avoir partagé les petits succès comme les moments difficiles, tout en menant brillamment la tienne.

List of publications

The main material of this thesis appeared in the following publications:
[Do et al., 2021c] **Virginie Do**, Sam Corbett-Davies, Jamal Atif, and Nicolas Usunier. Two-sided fairness in rankings via Lorenz dominance. Advances in Neural Information Processing Systems, 34, 2021.
[Do and Usunier, 2022] **Virginie Do** and Nicolas Usunier. Optimizing generalized Gini indices for fairness in rankings. In Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval, SIGIR '22, page 737–747, 2022.
[Do et al., 2023] **Virginie Do**, Elvis Dohmatob, Matteo Pirotta, Alessandro Lazaric, and Nicolas Usunier. Contextual bandits with concave rewards, and an application to fair ranking. In The Eleventh International Conference on Learning Representations, 2023
[Do et al., 2022a] **Virginie Do**, Sam Corbett-Davies, Jamal Atif, and Nicolas Usunier. Online certification of preference-based fairness for personalized recommender systems. In Proceedings of the AAAI Conference on Artificial Intelligence, volume 36, pages 6532–6540, 2022.

Contributions were made to the following publications, which are also discussed in the thesis:
[Liu et al., 2023] David Liu, **Virginie Do**, Nicolas Usunier, and Maximilian Nickel. Group fairness without demographics using social networks. In 2023 ACM Conference on Fairness, Accountability, and Transparency, 2023.
[Usunier et al., 2022] Nicolas Usunier, **Virginie Do**, and Elvis Dohmatob. Fast online ranking with fairness of exposure. In 2022 ACM Conference on Fairness, Accountability, and Transparency, pages 2157–2167, 2022.
[Do et al., 2021a] **Virginie Do**, Jamal Atif, Jérôme Lang, and Nicolas Usunier. Online selection of diverse committees. In Zhi-Hua Zhou, editor, Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21, pages 154–160, 2021.

Contributions were made to the following works, which are not discussed in this thesis:
[Do et al., 2022b] **Virginie Do**, Matthieu Hervouin, Jérôme Lang, and Piotr Skowron. Online approval committee elections. In Lud De Raedt, editor, Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, pages 251–257, 2022.

Contents

Chapter 1

Introduction

Contents

1.1 The societal impacts of recommender systems

Recommender systems are an integral part of modern digital platforms, serving up to billions of users worldwide. These systems are present in online marketplaces, streaming services, content sharing platforms, and online social media. They play a crucial role in organizing the vast amount of available information by providing personalized recommendations to users for a variety of purposes, such as browsing news articles, finding products, jobs, housing, or people to connect with.

In the era of machine learning and its increasing adoption in many applications that affect our daily lives, recommender systems stand out as one of the most successful applications of machine learning algorithms. Machine learning have been instrumental in leveraging the vast amounts of data available on online platforms to personalize user experience and facilitate the discovery of new and relevant items. These algorithms analyze statistical patterns in users' past browsing behavior, interactions with items, expressed preferences, and other characteristics to predict their future interests. These predictions enable the retrieval of items to recommend with the aim of maximizing user engagement, such as increasing the number of clicks, likes, reshares, or time spent on the platform. Machine learning offers the promise of highly tailored recommendations that reflect individual tastes and preferences, leading to higher user satisfaction and increased platform usage.

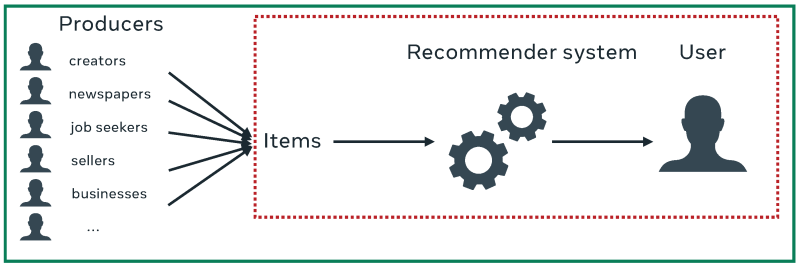


Figure 1.1: The recommendation ecosystem. The traditional view of recommender systems is user-centric: recommender systems are designed to find the most relevant items for the user (red dots). Modern recommender systems should also account for their impact on the people who produce the available items (green box).

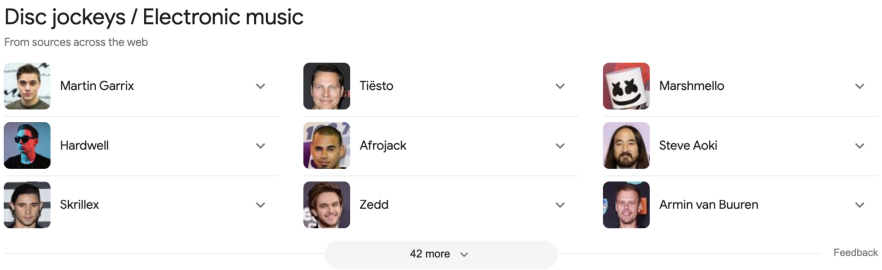


Figure 1.2: Results from a Google search with query "best electronic music DJs" which only exhibits male artists (screenshot from February 2023).

However, beyond the promise of increased user engagement, recommendation algorithms have profound social consequences. With the power to decide who is visible and who is not, these algorithms have a significant impact on item producers (Figure 1.1). For example, news outlets rely on their exposure on newsfeeds to generate revenue from readers, while creators on content sharing platforms and artists on streaming platforms rely on viewers and listeners to remain sustainable. Similarly, the attractiveness of businesses such as restaurants and shops largely depends on their exposure to potential customers in Google Maps' local recommendations. The success of a job seeker on job search platforms such as LinkedIn depends on which recruiter gets to see their resume, and the effectiveness of a dating application also depends on which users someone's profile is being recommended to.

By determining which item producers are visible or not, recommender systems make decisions that pose real ethical and social concerns. These include the risks of overlooking or disadvantaging job seekers from underrepresented groups [Geyik et al., 2019], amplifying racial biases in dating applications [Hutson et al., 2018] and overrepresenting demographic, cultural, or political groups on social media and search results. For example, research has shown that women are systematically underrepresented in search results for a variety of occupations [Kay et al., 2015]. We provide a new example of this in Figure 1.2, where the search results for the term "DJ" predominantly show male DJs. Other research from Twitter showed that their recommendation algorithm favored content from rightwing politicians and news outlets over leftwing content [Huszár et al., 2022], a finding that received large press coverage[1]. Recommender systems also have the potential to disproportionately favor established creators and artists on content sharing platforms, leading to the marginalization and eventual decline of smaller ones who do not receive enough exposure to succeed [Mehrotra et al., 2018]. To mitigate the potential negative impact of recommender systems on item producers,

[1] see e.g., The Guardian https://www.theguardian.com/technology/2021/oct/22/twitter-admits-bias-in-algorithm-for-rightwing-politicians-and-news-outlets

it is crucial to carefully evaluate their societal implications and ensure that they do not unfairly disadvantage any groups.

On the side of users, recommender systems are traditionally designed to provide them with the most relevant items, a goal which seemingly aligns with their interests. However, concerns have been raised about the impact of recommendation algorithms on users in recent years. Audits of recommender systems have exposed disparities in the content delivered to various social groups of users. For instance, Datta et al. [2015] found that equally qualified women received fewer online ads for high-paying jobs than men. To prevent the risk of unfair delivery of opportunities across users, significant efforts have been made to audit recommender systems for unintended biases or discrimination against their users. These efforts call for the development of new recommendation algorithms that provide fair access to information and opportunities to their users.

Given the real-world impacts of recommender systems on their users and item producers, fairness in recommender systems has become a central topic in machine learning and information retrieval research. Fairness in recommender systems can be examined from at least two different sides: the item side and the user side. On the item side, the goal is to provide item producers a fair share of exposure in the recommendations. On the user side, it is necessary to ensure that recommender systems do not create or amplify unintended biases and provide recommendations that benefit all users. There is a growing demand for recommender systems that simultaneously achieve both goals, in order to sustain a healthy recommendation ecosystem that serves the interests of all their stakeholders [Patro et al., 2020, Abdollahpouri et al., 2020]. The societal impact of recommender systems is significant, and ensuring fairness for both users and item producers is crucial to avoid perpetuating or amplifying existing biases and inequalities.

Fairness in recommender systems is a focal point in a broader and active debate on the societal impacts of machine learning algorithms. As machine learning algorithms continue to gain traction in our daily lives, there has been growing public concern about the potential of machine learning models to introduce biases and discrimination in algorithmic decisions [Buolamwini and Gebru, 2018, Barocas and Selbst, 2016]. As a result, fairness has become a central topic in machine learning research, particularly in the context of classification and supervised learning [Barocas et al., 2019]. With the potential for algorithms to perpetuate biases and discrimination in decision-making, researchers have proposed a range of fairness metrics and methods to address these concerns in various supervised learning tasks, including recidivism prediction, hiring, and credit scoring. These methods aim to ensure that the algorithms do not perpetuate unfair practices, such as differences in treatment or outcomes based on gender, race, or other protected characteristics. In this chapter, we will delve into the key role of fairness in recommender systems within the expansive and constantly evolving field of fair machine learning, and we will present our contributions to this critical area.

1.2 Fairness concerns in recommender systems

1.2.1 Sources of unfairness in recommender systems

Overview of recommender systems. The task of a recommender system is to provide each of its users with a ranked list of items, which are selected from a large pool of candidate items (e.g., videos) provided by producers (e.g., video creators). The recommender system evaluates the quality of the rankings with ground-truth relevance scores, which measure the value of an item to a user. At a high level, recommendation algorithms rely on two steps to generate ranked recommendations:

1. **Learning:** Estimate the value of every item for each user. This is done with a machine learning model that learns from past interactions of users with items, item features (e.g., category,

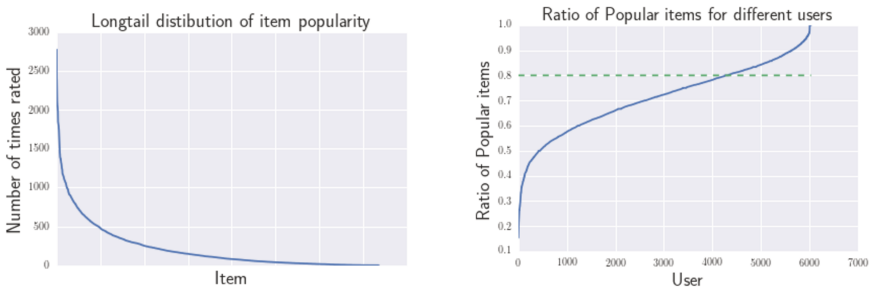


Figure 1.3: Popularity bias in the MovieLens-1m dataset [Harper and Konstan, 2015], which includes ratings of movies from users. Figures from [Abdollahpouri et al., 2019b].

publication date), and user features (e.g., age, country).

2. **Ranking:** Choose a top-K ranking of items for each user based on the estimated scores. This results in a personalized ranking policy, where different users are recommended different lists of items based on their predicted preferences.

Traditionally, the ranking stage simply consists in sorting items by decreasing scores for each user. When the true relevance scores are available to the recommender system, this strategy is optimal for maximizing standard ranking metrics, such as the discounted cumulated gain (DCG) [Järvelin and Kekäläinen, 2002], which measure the quality of the rankings from the users' perspective. However, it does not consider which item producers are exposed in the rankings. As we previously discussed, this is a critical fairness issue because the visibility given to item producers (or the lack thereof) has real social consequences.

The primary focus of this thesis is to address the issue of fairness in the ranking step of recommender systems. The ranking step is a crucial stage where the recommender system decides which items will be recommended to which users, once user preferences are estimated. It is a collective *decision* that impacts both users and item producers. Therefore, our goal is to ensure that the ranking step accounts for the interests of both users and item producers, and balances them in a fair manner.

Sources of unfairness. There are various ways in which the ranking step, or the combination of the learning and ranking steps, can lead to unintended and undesirable consequences. The ranking step can produce winner-take-all effects where some groups of item producers capture all the available exposure. In the traditional ranking solution that simply consists in ordering items by scores, even small differences in scores lead to large differences in exposure between item producers. This results in a long tail effect where a few popular items tend to dominate the highest positions in the rankings, leaving out a large number of less popular items with little or no exposure (Figure 1.3). This long tail effect can be problematic for small producers, as they struggle to gain any visibility or recognition, further exacerbating the power law distribution of exposure [Abdollahpouri et al., 2019b]. Moreover, systematic biases in the estimation of preferences can arise from learning stereotypes or popularity biases [Mehrotra et al., 2018]. These biases at the learning stage can be amplified by the ranking stage, where items from disadvantaged groups with systematically underestimated values are not shown to users in the end (Figure 1.2).

The combination of learning and ranking can also lead to unfair outcomes on the user side. During the estimation step, recommender systems often rely on strong modeling assumptions and multi-task learning to deal with the scarcity of per user data, with methods such as low-rank matrix factorization [Koren et al., 2009]. The limited capacity of the models or incorrect assumptions

might leave aside users with less common preference patterns. Because of this, the system may incorrectly learn stereotypical user tastes, such as gendered associations between user preferences and job categories. The ranking step then amplifies these biases by ordering items according to the estimated values, resulting in poor recommendation performance for users with non-stereotypical tastes [Ekstrand et al., 2018] or skews in the recommendation of certain content across sensitive groups [Sweeney, 2013, Imana et al., 2021]. Moreover, in the case of advertising markets, skews in ad delivery appear when the ranking decision accounts for the results of an auction in which advertisers compete for the same group of users [Ali et al., 2019]. For example, job advertisers must sometimes compete with product ads targeted at women, leading them to be shown to fewer women than men.

1.2.2 Fair recommendation *vs.* fair classification

This section aims to draw parallels between the problem of fairness in recommender systems and the more widely studied problem of fairness in classification tasks in machine learning. By making these connections, we aim to introduce the specific nuances of the former problem to readers who may already be familiar with the latter. While this serves as an introductory comparison to establish the foundation of our framework for fair recommendation, a comprehensive review of the literature on this topic is provided in Chapter 2.

Learning and decision-making in classification. Fairness in recommender systems is a critical area of research within the broader field of fairness in machine learning, which garnered significant attention in recent years. While recommender systems can be decomposed into a *learning* step and a *ranking* step, many other machine learning applications also have these *learning* and *decision-making* components [Kleinberg et al., 2018b, Kilbertus et al., 2020, Corbett-Davies et al., 2017]. The most largely studied setting is fair (binary) classification, where the goal is to predict a binary label for each individual, such as whether or not they will repay a loan, in order to assist a decision, such as whether or not to accept a loan application. Other common examples are recidivism prediction and hiring [Corbett-Davies and Goel, 2018, Barocas et al., 2019]. We discuss how fairness considerations in the learning and decision steps of classification tasks relate to the fairness considerations in the learning and ranking steps of recommender systems.

Let us consider a classical example in the fair machine learning literature, where a lender uses an algorithm to determine whether or not to approve a loan application [Hardt et al., 2016a, Liu et al., 2018]. In the *learning* step, a supervised learning algorithm produces a score for each individual by estimating the probability that they belong to the positive class (i.e., the probability that they repay the loan). This score is predicted by a probabilistic classifier that is trained on historical data. Unfairness can arise in the learning step when the data used to train the model is not representative of the population to which it is applied. The resulting model may not perform well on unseen data that comes from a different population, or it may learn problematic associations between sensitive attributes and outcomes. In the lending example, if the training data contains a majority of unsuccessful loan applications from people of a certain race or socioeconomic background, the resulting model may produce estimates that are biased against those groups. This can lead to unfair outcomes where certain groups are systematically denied access to loans because of a systematic underestimation of their creditworthiness. As discussed in the previous section, learning algorithms aimed at predicting items' values in recommender systems can also overestimate the value of popular items because of the lack of user feedback for less popular items in historical data.

In the *decision-making* step, individuals are classified as positive or negative based on their

predicted scores. In the lending example, the decision to accept a loan application is based on whether the applicant is predicted as creditworthy, which is done by applying a threshold to the estimated probability of repayment. The decision threshold can have significant fairness implications, as it determines which individuals are deemed eligible for certain life opportunities or services. In particular, when choosing group-specific (or group-agnostic) threshold policies, the resulting distribution of positive outcomes may or may not lead to welfare gains for disadvantaged groups [Kleinberg et al., 2018b, Corbett-Davies et al., 2017].

In recommender systems, ranking algorithms also make a decision on who receives positive outcomes. The decision is more complex than binary classification thresholds in at least two ways. First, it consists in producing one ranking of items for each user, instead of a simple threshold per user. Second, it involves making complex trade-offs between the interests of various stakeholders who value the recommendations differently: users seek rankings that best match their preferences, while items seek high exposure – therefore, the notion of positive outcome is not absolute.

In this thesis, we focus on the fairness of the decision that occurs at the ranking stage of recommender systems, more precisely on the social planning problem that consists in choosing a trade-off between the utilities of users and items (we later clarify the definitions of utilities in Section 1.3). This is a similar stance to Kleinberg et al. [2018b] who claim that fairness considerations should affect how the social planner uses the learned scores to make a decision, rather than the choice of learning algorithm, in the context of binary decision problems (i.e., college admissions).

Fairness criteria in classification. Fairness criteria have been proposed for both the learning and decision steps. The fairness of scores produced in the learning stage has been intensely studied in classification. Criteria include *calibration* between groups and *parity*[2] of predicted scores [Kleinberg et al., 2016, Pleiss et al., 2017]. In the lending example, parity requires that the average credit score is the same for all groups, while calibration requires that the probability of repaying a loan for a given credit score is the same for all groups. In the fair recommendation literature, a few criteria for the fairness of scores have been proposed [Yao and Huang, 2017, Islam et al., 2021], but several authors highlighted the insuffiency of considering scores in isolation from the final decision, i.e. the rankings [Beutel et al., 2019a, Singh and Joachims, 2018]. In particular, calibration of scores does not trivially extends to the setting of recommender systems [Steck, 2018], because the impact of an item's score is only meaningful in comparison to the scores of other items [Beutel et al., 2019a].

A broad class of fairness criteria in the decision step of classification tasks aim at equalizing outcomes across sensitive groups. *Demographic parity* requires equal probability of positive outcomes across sensitive groups [Feldman et al., 2015, Zliobaite, 2015] and *equality of opportunity* [Hardt et al., 2016b] (or equality of error rates [Zafar et al., 2017a, 2019]) aims at equalizing the probabilities of positive outcomes for the positive class across groups. Geyik et al. [2019] propose a mapping of demographic parity and equality of opportunity to the ranking setting. When items are partitioned into sensitive groups, demographic parity requires that groups of items receive equal exposure in the rankings, while equality of opportunity is similar to a popular merit-based criterion for rankings that we present in Section 1.3.5.

Corbett-Davies et al. [2017], Hu and Chen [2020] insist on the cost for social welfare of seeking parity of outcomes in classification problems, as it is possible to equalize outcomes across groups by depriving individuals from positive outcomes without redistributing them to disadvantaged individuals. In this thesis, we also demonstrate the undesirable consequences of enforcing fairness constraints on item exposure (Chapter 3). However, we argue that reducing inequalities in the distribution of outcomes is reasonable in the case of ranking, where the decision is *allocative*, because

[2] In classification, parity criteria are more often considered at the level of outcomes, i.e., of the decisions.

it can lead to positive changes in social welfare. In contrast, decisions in most fair classification problems are not allocative, because there is no budget on the number of positive classifications [Zafar et al., 2019, 2017a, Hardt et al., 2016b, Agarwal et al., 2018]. In other words, these works consider strict classification problems, rather than selection problems. In practice though, binary accept/reject decisions are often budgeted: there is typically a fixed budget to spend in lending problems, and a fixed number of slots in a college admissions. Budget considerations as in [Kleinberg et al., 2018b, Emelianov et al., 2022] bring classification problems closer to recommender systems where there is a fixed number of recommendation slots to allocate. In those budgeted settings, it is desirable to redistribute outcomes, since a positive outcome that is taken from someone is necessarily *transferred* to someone else. We present in the following section a main contribution of this thesis, which is a framework for guiding the allocative decision of ranking in recommender systems, rooted in distributive justice principles of social choice.

1.3 Social choice for fair recommendation

This section presents a core contribution of this thesis: a conceptual framework for fairness in recommender systems that is grounded in social choice theory.

1.3.1 Fair allocation of exposure in recommender systems

As we previously discussed, at the ranking stage, recommender systems make a collective *allocative decision* on which items receive exposure, and to which users they are exposed. Fairness in *allocation problems*, or *fair division*, has a long history in social choice theory, which is a branch of economics that studies collective decision-making processes based on the heterogeneous preferences of multiple agents [Arrow et al., 2010, Moulin, 2003]. In this thesis, we approach fairness in recommender systems as a new fair division problem, where the scarce resource to distribute is the amount of content that the system can display to its users, i.e., the total available exposure. Different item producers compete for a share of this limited resource. Our view is that the recommender is a social planner whose goal is to provide ranked recommendations to users by fairly allocating the exposure budget among item producers, while also taking into account the impact of the allocation mechanism on user satisfaction. We build on the extensive research on fair division that has been conducted in the past in social choice theory and cardinal welfare economics.

We use the term *utility* in its broad sense in cardinal welfare economics as a "*measurement of the higher-order characteristic that is relevant to the particular distributive justice problem at hand*" [Moulin, 2003]. In our allocation problem, there are two types of agents – users and item producers – who benefit differently from the rankings. Users value high quality rankings that best match their preferences, and items benefit from a high number of views. As a result, we define user utility as a ranking performance metric, and item utility as the expected number of views. We provide formal definitions of user utility and item utility in the next section, and a discussion of these modelling choices in Chapter 7. The allocation problem consists in choosing rankings by making trade-offs between user utilities and item utilities. We refer to this allocation problem as the *fair allocation of exposure* problem.

As discussed earlier, the traditional approach in recommender systems is to maximize average user utility only, by sorting items by decreasing relevance for each user. However, this approach can have undesirable effects, such as unfair winner-take-all effects and amplification of biases in estimated scores, as described in Section 1.2.1. Therefore, our motivation for considering the fairness of exposure allocation towards both users and item producers is to mitigate these negative

consequences.

1.3.2 Formal setting and utility models

We formalize the notions of user and item utilities in the following recommendation setting.

Formal setting. We consider a setting in which there is a set of n users and m items (e.g., videos) created by producers (e.g., video creators), and the recommender system must generate a top-K ranking of items for each user. We denote by $\mu_{ij} \in [0, 1]$ the ground-truth value of item j for user i. In practice, μ_{ij} represents a relevance score, or the probability that the user positively engages with the item (e.g., the probability of watching or liking a video). We denote by $P \in \mathbb{R}^{n \times m \times m}$ a *ranking policy* that defines a ranking for each user: P_{ijk} is equal to 1 if j is recommended to user i at position k, and 0 otherwise. The output of the recommender system is a ranking policy P.

Utility models. Following the academic literature on fairness of exposure [e.g. Singh and Joachims, 2018, Wang and Joachims, 2021, Biega et al., 2018], we assume that users examine ranked lists by following the *position-based model* [Craswell et al., 2008]. This model is based on the intuition that users examine items in order of their ranking, and that the probability of examining an item decreases as the item's rank increases. The position-based model is defined by a set of weights $\boldsymbol{b} \in \mathbb{R}_+^m$, where b_k represents the probability that a user examines an item at position k. We assume that the weights are non-increasing, i.e., $b_1 \geq \ldots \geq b_K$ and $b_k = 0$ for any $k > K$. In the position-based model, the user utility is measured with the following ranking performance metric:

$$\textit{User utility:}\ \ u_i(P) = \sum_{j=1}^{m} \sum_{k=1}^{m} \mu_{ij} P_{ijk} b_k \tag{1.1}$$

The user utility is higher when relevant items are ranked higher. When the weights are $b_k \propto \frac{1}{\log_2(1+k)}$, the user utility is the discounted cumulated gain (DCG) [Järvelin and Kekäläinen, 2002], a classical ranking performance metric. Note that the utility of user i only depends on their own ranking P_i and not the global ranking policy P.

Since the ranking policy also has an impact on the item producers, we also measure the exposure of an item, which is its expected number of views across all users' rankings in the position-based model. Formally, the exposure of an item $j \in [\![m]\!]$ is defined as:

$$\textit{Item exposure:}\ \ v_j(P) = \sum_{i=1}^{n} \sum_{k=1}^{m} P_{ijk} b_k.$$

Exposure is higher when the item appears in higher positions in more users' rankings. We define the *item utility* as the item exposure, and use the two terms interchangeably. To simplify the presentation, *we identify item producers with items*, but the framework would be conceptually equivalent by defining the exposure of a producer as the sum of the exposures of all the items produced by the producer. We defer the discussion of the implications and limitations of these utility models to Chapter 7, Section 7.2.3.1.

Note that unlike user utility, the exposure of an item j depends on the global ranking policy P, and not just the local ranking P_i of a given user i. Considering items' exposures' thus introduces a coupling between the rankings, and requires handling the global ranking policy P.

1.3.3 Maximizing welfare functions

The recommender system has to make a normative decision on how much utility should be redistributed a) between users and items, and b) among each population, between the better-off and the worse-off individuals. This involves a complex multidimensional trade-off. Boosting small item producers by reducing the exposure of popular items is costly for average user utility. At the same time, the least satisfied users should not bear that cost. It is therefore crucial to examine who benefits or bears the cost of reducing inequalities of exposure among items. Our goal is to provide a framework to assess the multi-dimensional trade-offs involved by a ranking policy, and to generate rankings that achieve a variety of these trade-offs. The choice of a specific trade-off is left to the designer of the recommender system.

We follow a general framework based on maximizing welfare functions in social choice [Moulin, 2003, Sen, 1970, Arrow, 1951]. *Welfare functions* specify an ordering of a set of alternatives $\mathcal{P}$ by mapping a utility profile to a real value that represents the aggregate utility of all individuals for a given alternative $P \in \mathcal{P}$. Socially preferred alternatives are those that maximize the welfare function. For the fair allocation of exposure problem, we propose to find a ranking policy P within a set of ranking policies $\mathcal{P}$ by maximizing a *global welfare function* $F(P)$, which is a weighted sum of welfare functions for users and items:

$$F(P) = (1-\lambda)g^{\text{user}}(\boldsymbol{u}(P)) + \lambda g^{\text{item}}(\boldsymbol{v}(P)), \tag{1.2}$$

where $g^{\text{user}} : \mathbb{R}^n \to \mathbb{R}$ and $g^{\text{item}} : \mathbb{R}^m \to \mathbb{R}$ are strictly concave, increasing functions that respectively aggregate the utilities of users and the utilities of items, and $\lambda \in [0, 1]$ controls the trade-off between the welfare of users and the welfare of items. The strength of the curvature of the concave welfare function g^{item} (resp. g^{user}) controls how much utility should be redistributed from better-off to worse-off items (resp. users).

1.3.4 Distributive justice principles in welfare economics

In this thesis, we study two classes of welfare functions from cardinal welfare economics for the choice of g^{user} and g^{item} in $F(P)$ in Equation (1.2). We focus on additive concave welfare functions in Chapter 3 and generalized Gini welfare functions in Chapter 4. We later introduce these classes of welfare functions in their respective chapters, as well as in the background section on social choice (Section 2.4). When using either of them in the global welfare function $F(P)$, we prove that the ranking policies obtained by maximizing $F(P)$ simultaneously satisfy two fundamental social choice properties. These properties are Pareto efficiency and the Pigou-Dalton transfer principle [Moulin, 2003], which we introduce for the problem of fair allocation of exposure in recommender systems. Considering the whole population of users and items, these properties are defined as follows:

1. **Pareto efficiency**: It is not possible to improve the utility of an individual (user or item), without decreasing the utility of another individual.
2. **Pigou-Dalton transfer principle**: At a given level of total utility, utility should be redistributed as much as possible from the better-off to the worse-off individuals.

Pareto efficiency is an efficiency criterion that avoids undesirable rankings where everyone is made worse-off. The Pigou-Dalton transfer principle is a distributive fairness criterion that allows to sort through Pareto-efficient solutions. It favours rankings that redistribute exposure from highly visible items to less visible items. Therefore, it promotes more equality among items, and it allows to mitigate the winner-take-all effects of the traditional ranking solution, which we described in Section 1.2.1. Furthermore, on the user side, since worst-off individuals are prioritized, the transfer

principle makes sure that the least satisfied users do not bear the cost of boosting the exposure of the least visible items.

The Pigou-Dalton transfer principle is equivalent to **Lorenz efficiency**, a criterion that we introduce in Chapter 3 that combines these efficiency and fairness guarantees [Hardy et al., 1952, Marshall et al., 1979]. The definition of the Lorenz efficiency criterion is based on the *generalized Lorenz curves* of utility profiles, which are a graphical representation of the cumulative utility detained by each fraction of a population, and are used in welfare economics to measure income inequality [Shorrocks, 1983, Kolm, 1976]. We introduce them in more detail in the background section (Section 2.4), and in Chapter 3 where we use generalized Lorenz curves to assess the fairness of rankings obtained by various methods for users and items.

1.3.5 Assessment of merit-based fairness constraints

In the previous section, we introduced fundamental social choice properties, particularly the Pigou-Dalton transfer principle (or equivalently, Lorenz efficiency), which had been overlooked by the fair ranking literature. These properties provide a principled basis to assess the fairness of rankings for users and items. In Chapter 3, we use their insights to examine the distributive fairness of existing approaches to fair ranking.

In the literature on fairness of exposure in rankings, fairness for items is often measured by a distance between the vector of items' exposures' (i.e., the item utility profile) and a target exposure vector, which represents the ideal distribution of exposure among items in a recommender system considered as fair [Diaz et al., 2020, Kletti et al., 2022a, Raj and Ekstrand, 2022]. A prominent fairness notion in this recent literature is *merit-based fairness* which states that the exposure of an item should be proportional to its merit – in these works the target exposure of an item is defined as an increasing function of its average value to users, which is used to measure the merit of an item [Biega et al., 2020, Diaz et al., 2020, Morik et al., 2020, Singh and Joachims, 2018, Biega et al., 2018]. Starting from this merit-based fairness measure, authors either proposed to minimize merit-based unfairness [Diaz et al., 2020, Biega et al., 2018], or optimize trade-offs between average user utility and merit-based fairness [Kletti et al., 2022a, Morik et al., 2020, Biega et al., 2020], or maximize user utility under merit-based fairness constraints [Singh and Joachims, 2018][3].

In Chapter 3, we assess these approaches in the light of distributive justice principles. Following these works, we define the merit of an item j as $q_j = \sum_{i=1}^{n} \mu_{ij}$. Let $E = n \|\boldsymbol{b}\|_1$ the total exposure and $Q = \sum_{j'=1}^{m} q_{j'}$ the total merit. Then the target of item j is $\frac{q_j E}{Q}$, so that if for all items j the exposure of j is equal to its target, then the ranking policy satisfies the merit-based fairness criterion stating that the exposure of an item should be proportional to its merit. We assess approaches which optimize the following trade-offs between total user utility and merit-based fairness, where $\beta > 0$ is a trade-off parameter:

$$F^{\text{merit}}(P) = \sum_{i=1}^{n} u_i(P) - \frac{\beta}{m} \sqrt{\sum_{j=1}^{m} \left(v_j(P) - \frac{q_j E}{Q} \right)^2} \tag{1.3}$$

We show in Chapter 3 (Proposition 2) that when increasing the strength of the penalty in favour of the merit-based fairness criterion, this can lead to increase inequalities among items while decreasing total user utility. In practice, in some recommendation problems, merit-based fairness can increase the exposure of popular items (items j with high merit q_j), leading to rich-gets-richer effects. Although this is compatible with Pareto efficiency, it is a clear violation of the Pigou-Dalton transfer

[3]Note that Singh and Joachims [2018] consider merit-based fairness for items at the level of a single ranking, while we (and the other works mentioned here) consider amortized fairness across the rankings of all users.

principle, which promotes transfers of exposure from "rich" to "poor" items. This fundamental fairness condition in social choice provides a more complete understanding of merit-based approaches, by demonstrating that they may unintentionally lead to distributive unfairness.

1.3.6 Reciprocal recommendation

Reciprocal recommender systems. The recommendation framework that we discussed thus far depicted *"one-sided" recommendation*, in the sense that only items are being recommended. Our conceptual framework for fair allocation of exposure also applies to *reciprocal recommendation* problems [Palomares et al., 2021]. Reciprocal recommender systems include the recommendation of friends or dating partners in social networks, or the recommendation of job seekers to recruiters and vice versa on job search platforms. The specificity of reciprocal recommender systems is that users are also items that can be recommended to other users (the item *per se* is the user's profile or CV). Since items are also users, they have meaningful preferences on which users they should be recommended to.

Ensuring fair recommendations in reciprocal recommender systems is a critical and complex issue. In professional matching platforms, uncarefully addressed popularity biases or learned stereotypes [Palomares et al., 2021, Geyik et al., 2019] can restrict the access of disadvantaged groups of eligible candidates to job opportunities, whenever the recommender system fails to give them enough exposure to the employers who are relevant to them, and to whom they would be relevant. Online dating platforms are also questioned about the fairness of their recommendation algorithms, which may exacerbate pre-existing biases in the dating market where users feel more entitled to express preferences based on race or economic status [Zheng et al., 2018, Hutson et al., 2018].

The key to extend our recommendation framework to reciprocal recommendation tasks is to redefine the utility of a user to account for the fact that (1) the user utility comes from both the recommendation they receive and who they are recommended to, and (2) users have preferences over who they are recommended to. In this setting, the set of users and the set of items are identical, so we have $n = m$. Let us denote by μ_{ij} the mutual preference value between two users i and j. We follow the common assumption in the reciprocal recommendation literature that $\mu_{ij} = \mu_{ji}$ [e.g. Palomares et al., 2021]. For instance, when recommending CVs to recruiters, μ_{ij} can be the probability of an interview, while in dating, it can be that of a "match". The *two-sided utility* of a user i is then the sum of the utility $\overline{u}_i(P)$ derived by i from the recommendations received, and the utility $\overline{v}_i(P)$ from being recommended to other users:

$$u_i(P) = \overline{u}_i(P) + \overline{v}_i(P) = \sum_{1 \leq i,j \leq n} (\mu_{ij} + \mu_{ji}) P_{ij}^{\intercal} \boldsymbol{b}$$

$$\text{where} \quad \overline{u}_i(P) = \sum_{j=1}^{n} \mu_{ij} P_{ij}^{\intercal} \boldsymbol{b} \quad \text{and} \quad \overline{v}_i(P) = \sum_{j=1}^{n} \mu_{ij} P_{ji}^{\intercal} \boldsymbol{b}\,.$$

In other words, the two-sided utility of i is the sum of its user-side utility and its item-side utility.

Fair allocation of exposure in reciprocal recommender systems. In the reciprocal recommendation setting, since there are only users, the welfare objective $F(P)$ of Equation (1.2) is simply a function that aggregates the two-sided utilities: $F(P) = g(\boldsymbol{u}(P))$, where $g : \mathbb{R}^n \to \mathbb{R}$ is a strictly concave, increasing function. In this setting, distributive fairness aims at improving the utility of the worse-off users, and the recommender system must make trade-offs between the utility of the worst-off users and total user utility. The curvature of g controls how much redistribution of two-sided utility is desired from better-off users to worse-off users. In practice, the utility of the

worse-off users can be improved by boosting their exposure, i.e. increasing their "item-side utility".

As for non-reciprocal recommendation problems, we address both additive concave welfare functions in Chapter 3 and generalized Gini welfare functions in Chapter 4 for the choice of g. We show that in both cases, ranking policies P obtained by maximizing $F(P)$ satisfy Lorenz efficiency (i.e., Pareto efficiency and the Pigou-Dalton transfer principle). We also show in Chapter 3 that striving for equal utilities in reciprocal recommendation is Pareto inefficient as it can destroy everyone's utility. Finally, we discuss in Chapter 7 the similarities and differences between reciprocal recommendation and matching problems.

Overall, our framework for fairness in recommender systems applies to both non-reciprocal and reciprocal recommendation tasks. As discussed in the related work chapter (Chapter 2, Section 2.1), the latter problem received considerably less attention in the fair recommendation literature. We believe that our framework for fair allocation of exposure in reciprocal recommendation tasks can be useful to address critical fairness issues in this overlooked setting.

1.4 Algorithms for maximizing concave functions of rankings

This section discusses the algorithmic challenges of maximizing concave functions of rankings and presents a high-level overview of our algorithmic contributions to overcome these challenges.

1.4.1 Batch and bandit settings

In this thesis, we consider either of two settings for recommender systems, which combine learning and ranking in different ways.

1. The **batch** setting: There is a fixed, large batch of n users and a set of m items. First, the recommender system operates the *learning step off-line*: it estimates the value of each item for all the users in the batch, i.e., it produces a full matrix of estimated scores $(\hat{\mu}_{ij})_{\substack{1\leq i\leq n\\1\leq j\leq m}}$ by learning from historical interactions and contextual information. Then, the system proceeds to the *ranking step*. It produces a global, static ranking policy P, i.e., one ranking of items per user, based on the estimated values $\hat{\mu}_{ij}$. Finally, the recommender system is able to produce *static* measures of users' satisfactions and items' exposures from the ranking policy P.
2. The **contextual bandit** setting: This is an *online* setting where the system observes users sequentially in sessions and learns from online interactions with users. We assume that the set of items is still fixed over time. At each timestep t, the system observes a user and their features $x_t \in \mathcal{X} \subset \mathbb{R}^d$, estimates the context-dependent values for the current user $\hat{\mu}(x_t)$, and produces a ranking based on the current value model. The system updates the value model based on the noisy feedback that the user gives on the ranking (e.g, which items of the ranking the user examined and engaged with). The system measures user and item utilities *dynamically* over timesteps. In summary, the contextual bandit setting consists in a *sequence of learning and ranking steps*, where the learning step is based on the observed user features x_t and past feedback from users, and the ranking step is based on the current model of user preferences $\hat{\mu}(x_t)$.

In Chapters 3 and 4, we address the batch setting and focus on the ranking problem, assuming that the ranking algorithm has access to a full matrix of user-item values $(\mu_{ij})_{\substack{1\leq i\leq n\\1\leq j\leq m}}$. We do not address the learning step of the batch setting in these chapters. Nonetheless, we do provide in Appendix A.3.3 an excess risk bound, which provides guarantees on the true value of the ranking objective when the algorithm uses estimated values $(\hat{\mu}_{ij})_{\substack{1\leq i\leq n\\1\leq j\leq m}}$, depending on the quality of the estimation. In Chapter 5, we address the contextual bandit setting, where the recommender system

balances exploration and exploitation to learn user preferences and rank items.

In the previous Section 1.3, we presented our conceptual framework and ranking objectives in the batch setting. The batch setting is useful for disentangling the learning step from the ranking step. This static setting allows to focus on the fairness of the allocative decision made in the ranking step. Specifically, it facilitates the modeling of trade-offs involved in the ranking policy and enables the definition of ideal objective functions for fair ranking by considering all users and items at once. This separation between learning and decision-making allows to think about fairness in allocative terms at the ranking stage, and to bring insights from the literature on fair division in social choice - an area that has been relatively unexplored in the context of fair machine learning.

The drawback of the batch setting is that it is impractical for real-world recommender systems, since it requires computing the rankings of all users at once, involving large computation and memory costs. In practice, rankings are generated online as users enter new sessions and request recommendations. It is much more efficient to compute the ranking of a single user at a time. The contextual bandit setting enables this, in combination with online learning of user preferences. However, the bandit setting is more challenging, as it requires to design algorithms that efficiently balance exploration and exploitation while optimizing for complex fairness objectives.

1.4.2 Maximizing concave ranking objectives

Algorithmic challenge: Maximization of concave functions of rankings. As described in Section 1.3.3, we propose in this thesis to find ranking policies P by maximizing global welfare functions of the form of $F(P)$ (Equation (1.2)), where F is a concave function of the ranking policy P. We choose to find P in the set $\mathcal{P}$ which is the convex hull of deterministic ranking policies, or equivalently, the convex set of stochastic ranking policies. We follow the line of work on fair ranking which also considers randomized rankings [e.g. Singh and Joachims, 2018], because they enable the use of convex optimization techniques to generate the recommendations, which would otherwise involve an intractable combinatorial optimization problem in the space of all users' rankings.

Despite the convex relaxation to stochastic rankings, optimizing welfare functions of the form of $F(P)$ is still computationally challenging, because ranking with fairness of exposure for items requires to solve a global optimization problem in the space of rankings of all users. Indeed, recall that the exposure of an item is the sum of its exposure to every users: $v_j(P) = \sum_{i=1}^{n} \sum_{k=1}^{m} P_{ijk} b_k$, which means that it is not possible *a priori* to decouple the global optimization problem into a set of local optimization problems where P_i is found independently for each user i. In contrast, the traditional solution for maximizing average user utility finds each P_i independently by sorting items by decreasing values μ_{ij} for each user i.

Generic algorithms for concave ranking objectives. In this thesis, we present computationally efficient algorithms that optimize ranking objectives of the form of $F(P)$.

1. In the batch setting, our algorithms output a randomized ranking policy that can be represented as a sparse convex combination of deterministic ranking policies.
2. In the contextual bandit setting, our algorithms produce one deterministic ranking at a time for each user observed in sequence, associated with a stochastic context.

The main algorithmic contributions of this thesis start from the result of Theorem 5, which we prove in Chapter 3. Our result states that iterations of the Frank-Wolfe algorithm [Frank and Wolfe, 1956] can be computed efficiently for concave functions of rankings in the position-based model, with *one decentralized sorting operation per user*. Based on this result, we leverage Frank-Wolfe

variants and their theoretical analyses to derive computationally efficient algorithms for ranking that provably optimize various fair ranking objectives, in the batch setting and in the bandit setting.

While our efficient ranking algorithms are primarily motivated by the optimization of global welfare functions of the form of $F(P)$ (Equation (1.2)), they apply to *all concave functions of users' and items' utilities*, and not just welfare functions. This includes objectives with convex item-side fairness penalties that have been proposed in the fair ranking literature. For example, our algorithms also apply to the merit-based fairness objective of Eq. (1.3) that we presented in Section 1.3.5.

1.5 Outline and contributions

We now present the outline of this thesis and summarize the contributions by chapter, which each corresponds to an article published during the PhD.

The first two chapters focus on the fairness of the ranking stage, in the batch setting.

Chapter 3: Fairness in rankings with additive concave welfare functions. We propose to assess the fairness of rankings for users and items in recommender systems based on fundamental distributive justice principles in welfare economics, based on Pareto efficiency and the Pigou-Dalton transfer principle. We show that some popular approaches to fair ranking fail to satisfy those principles. For example, merit-based fairness constraints can decrease user utility while increasing inequalities of exposure among item producers, which goes against the transfer principle that aims to reduce inequalities. To overcome the limitations of existing approaches, we propose a new approach to generating fair rankings that is grounded in cardinal welfare economics. It consists in maximizing additive concave welfare functions, which are a family of smooth welfare functions. These welfare functions can be interpreted as sums of utilities of agents who have *diminishing returns*. The property of *diminishing returns* for exposure means that "one additional view counts more for items who have 10 views than who have 10 million views", which is particularly relevant for recommender systems. Rankings produced by maximizing such welfare functions satisfy Pareto efficiency and the Pigou-Dalton transfer principle.

We also introduce the related tool of generalized Lorenz curves from welfare economics to assess the fairness of rankings. Generalized Lorenz curves are a graphical representation that allows to visualize the distribution of utilities among users and items, and in particular the utility of the worst-off individuals, which we aim to improve. Using this representation, we can observe how much utility is taken from the best-off to increase the utility of the worst-off individual users, when varying the parameters of the additive welfare function.

Our conceptual framework is also the first one to simultaneously address fairness in non-reciprocal and reciprocal recommendation problems. Reciprocal recommendation is a specific setting that has been relatively overlooked by the fairness literature, and where users are also items. Their utility is thus two-sided: they benefit from the recommendations they receive, and from being recommended to other users. We show that the welfare function approach for non-reciprocal recommendation can be extended to the reciprocal case by using our new notion of two-sided utility, in order to better serve the worst-off users.

On the algorithmic side, global welfare functions that account for items' exposures are challenging to optimize, because the exposure of an item depends on the rankings of all users. Prior to our work, existing methods addressed this challenge with heuristic approaches without any guarantees or control over the achievable trade-offs. We propose a computationally efficient algorithm for

fair ranking based on the Frank-Wolfe method [Frank and Wolfe, 1956]. The algorithm generates a stochastic ranking policy as a weighted sum of deterministic ranking policies. This eliminates the need for an additional Birkhoff-von-Neumann decomposition step [Birkhoff, 1940], which was required in prior work using stochastic rankings [Singh and Joachims, 2018, Wang and Joachims, 2021]. Our algorithm is capable of optimizing any concave function of the utilities of the rankings, which encompasses our additive welfare functions but also existing fair ranking criteria.

We simulate a music recommendation task based on data from Last.fm to evaluate the performance of our algorithm. Our experiments confirm that merit-based fairness approaches are unable to decrease item inequality and can exacerbate winner-take-all effects where popular items capture a large fraction of the total exposure. In contrast, our approach based on maximizing additive welfare functions obtains better trade-offs between total user utility and inequality of utilities among items (measured by the Gini index or the standard deviation). Moreover, by varying the parameters of the welfare function, we are able to drive item inequality close to zero. Finally, towards two-sided fairness, our approach is able to generate a wide range of trade-offs between fairness for items and fairness for users, measured by the utility of the 10% and 25% worst-off users.

Since our framework encompasses reciprocal recommendation problems, we also provide experimental evaluation on a social recommendation task based on Twitter data. By maximizing an additive concave welfare function of the two-sided utility of users, we are able to generate a wide range of trade-offs between total utility and utility of the 10% worst-off.

Chapter 4: Fairness in rankings with generalized Gini welfare functions. We propose an alternative fair ranking approach based on Generalized Gini welfare Functions (GGF), which are a more expressive class of welfare functions than the previous additive welfare functions. A drawback of GGFs compared to additive welfare functions is that they cannot be expressed as a sum of utilities of agents with diminishing returns. Although we lose this intuitive interpretation, we gain in *expressivity* since GGFs are able to directly express fairness criteria based on utility quantiles (e.g. "maximize the utility of the 10% worse-off"). GGFs also cover more classical inequality measures such as the Gini index, which is widely used in inequality measurement and more recently in the literature on fairness in recommender systems. Although GGFs do not have an intuitive interpretation as *"sum of utilities with diminishing returns"*, their main advantage is that they generalize various existing fairness criteria for ranking. Emphasizing the generality of GGFs, we also prove that all Lorenz-efficient rankings can be generated by maximizing GGFs.

The algorithmic challenge of optimizing GGFs is that they are nondifferentiable, and therefore not amenable to vanilla Frank-Wolfe optimization. We propose to adapt a Frank-Wolfe variant for nonsmooth problems [Lan, 2013] which uses the Moreau-Yosida envelope as smoothing technique [Moreau, 1962, Yosida et al., 1965], and present a computationally efficient procedure to compute the smooth approximation of GGFs.

We conduct experiments on movie and music recommendation tasks and compare our algorithm that optimizes GGFs to previous recommendation methods, including our own approach based on additive concave welfare functions from Chapter 3. As expected, we find that our GGF-based approach obtains better trade-offs between total user utility and item inequality measured by the Gini index. This is because GGFs can be instantiated to the Gini index and our Frank-Wolfe variant allows for direct optimization of this non-differentiable measure. For two-sided fairness, we also obtain superior trade-offs between the utility of the 25% worst-off users and the Gini index of items' utilities, when instantiating the user-side GGF and item-side GGF to these criteria. Experiments on a reciprocal recommendation task based on Twitter data demonstrate similar results when optimizing trade-offs between the utility of the 25% worst-off users and total user utility.

The two previous chapters focus on the ranking problem to analyze its properties from a fair allocation perspective, independently from potential biases arising at the learning stage. In practice though, there are real-world limitations to the previous batch setting in which learning and decision-making are decoupled, and one global decision is made for all users at once. Modern recommender systems interact with users in an online manner: they learn the personalized items' values from user feedback, and at the same time decide what content to show to the current user as they request recommendations. Contextual bandits are a popular paradigm to model this joint learning-and-decision-making setting in personalized recommender systems [Li et al., 2010].

Chapter 5: Fair ranking in the contextual bandit setting. We address the problem of fair ranking with contextual bandits, which is the paradigm of choice for online personalized recommender systems that learn to generate recommendations from user feedback. We present a generic algorithm that works for many fair ranking objectives, including the smooth welfare functions of Chapter 3 and the nonsmooth welfare functions of Chapter 4. This is the first algorithm with regret guarantees for fair ranking in the contextual bandit setting. Moreover, the algorithm is computationally fast and has an intuitive interpretation: At each timestep, the algorithm gives an adaptive boosts to items that received low exposure in past recommendations, and the boost depends on the gradient of the fair ranking objective.

In fact, we provide an extensive treatment of the more general problem of contextual bandits with concave rewards (CBCR) [Agrawal et al., 2016], which is a multi-objective bandit problem. In CBCR, there is a vector of multiple rewards that depends on a stochastic context, and the trade-off between the rewards is defined by a concave function. This CBCR setting covers a variety of problems beyond fair ranking, including optimizing multiple user engagement metrics (e.g., clicks, streaming time) in recommender systems. Prior theoretical works addressed simpler versions of CBCR with simple policy spaces: Agrawal and Devanur [2014], Busa-Fekete et al. [2017] focus on the non-contextual setting where policies are distributions over actions, and Agrawal et al. [2016] address a restriction of CBCR to a finite policy space. We remove these restrictions and present regret guarantees for the general CBCR problem by proving a reduction of CBCR to classical scalar-reward contextual bandits. Our proof is based on a geometric interpretation of CBCR as an optimization problem over the convex set of all achievable expected rewards, and leverages techniques from theoretical analyses of Frank-Wolfe algorithms in constrained convex optimization.

On the experimental side, we simulate an online ranking task based on music recommendation data. We observe that compared to heurisic contextual bandit algorithms for fair ranking, algorithms using our reduction reach the highest value of the fair ranking objective as the number of timesteps increases. This shows the advantage of a principled bandit algorithm compared to heuristics without theoretical guarantees. When the fair ranking objective is a trade-off between average user utility and item inequality, our reduction-based bandit algorithm obtains higher average utility than existing bandit algorithms, at all levels of inequality between items.

In Chapters 3 and 4, we addressed the problem of social planning in recommender systems, where we seek to trade-off users' and items' utilities for the *design* of *two-sided fair* rankings. In Chapter 6, we take a different perspective: we address the *audit* of recommender systems, and focus on *user-side fairness*. This work was mostly conducted at the beginning of the PhD program, motivated by the large resonance of audits for user fairness in ad systems. For instance, Datta et al. [2015] found that women received fewer online ads for high-paying jobs than equally qualified men, while Imana et al. [2021] observed gender-based disparities in ad delivery rates for different

companies proposing similar jobs. Our contribution to this research stream is a complement to existing audits for user fairness. We start from the observation that existing audits do not control for disparities that are in line with user preferences. To strengthen the conclusions of these audits, we propose to test for the preference-based criterion of *envy-freeness*, which stipulates that no user should prefer their recommendations to those of other users. Envy-freeness is a fairness criterion that was first studied in fair division [Foley, 1967], and it thus has similar roots to the main conceptual framework of this thesis. However, it leads to a different assessment in the context of recommender systems, and the choice of fairness criterion depends on the motivating application.

In the previous Chapters 3 and 4, our work was motivated by applications where item producers are not advertisers paying for users' attention, but rather content creators claiming a fair share of exposure on the platform. Typical examples are online video sharing platforms and music streaming services. Since item-side fairness is a key concern in these applications, we designed algorithms that improve the exposures of small items across the rankings of all users. We also aimed to ensure that the users whose rankings are impacted by boosting small items are not those for whom the boosts are the most costly. We addressed two-sided fairness in the sense of improving the exposure of the worst-off items, while also prioritizing the utilities of the worst-off users. The ranking algorithms that we developed in the previous chapters are not guaranteed to pass the audit for envy-freeness of Chapter 6, because optimal ranking policies for objectives that include a concave item fairness term are not envy-free for users in general. For example, if users Alice and Bob both want to receive job ads from a popular company, but the designer promotes less popular employers by boosting their ads in Bob's recommendations, then Bob will be envious of the recommendations of Alice. In practice though, we recommend that our audit for envy-freeness is used in applications where user-side fairness is the main concern and item-side fairness is not a priority, such as in ad systems.

Overall, our perspective in Chapter 6 is that of an auditor who is solely focused on assessing fairness for users, without considerations of whether user-side unfairness is a consequence of other objectives. We argue that the audit perspective is as important as that of the designer, given the significant role played by audits for user fairness in raising awareness about the need for fairness in recommender systems. Moreover, designers can use the evaluations produced by auditors as additional diagnoses to improve their systems. In fact, existing audits have led to settlements that drove online platforms to change their recommendation algorithms to comply with new requirements for user fairness [Bogen et al., 2023].

Chapter 6: User fairness as envy-freeness. In Chapter 6, we propose to assess the fairness of recommender systems for their users with the criterion of *envy-freeness* from fair division in social choice theory. Transposed to the recommendation setting, envy-freeness states that each user should prefer their recommendations to those of other users. For example, in a job recommender system where two users Alice and Bob seek taxi driver roles [Ali et al., 2019], if Bob is the only one to receive ads for driver jobs, then the system is deemed unfair by the envy-freeness criterion. Compared to our previous welfare function approach that relied on comparing utilities across users, envy-freeness avoids the difficult assumption of interpersonal comparisons of utilities. Indeed, in the envy-freeness criterion, different recommendations are compared from the perspective of the same user (e.g., Alice).

We present a formal analysis of the properties of envy-freeness as a user-side fairness criterion for recommender systems, and show its compatibility with optimal recommendations. We also show its incompatibility with item-side merit-based fairness constraints.[4] We also present a probabilistic

[4] In Chapter 6, the result is proved for merit-based fairnesss constraints applied at the level of each user, rather than across users.

relaxation of envy-freeness, in order to remove the quadratic dependence on the number of users and make the certification of envy-freeness tractable.

Compared to the previous chapters where we took the perspective of the designer of the recommender system as a social planner, in this chapter we take the perspective of an internal auditor of the recommender system. Auditing envy-freeness in recommender systems is technically challenging, because it requires probing users' preferences for the recommendations of others, in order to reliably answer the counterfactual questions: "would user Alice prefer the recommendations of Bob?". Our algorithmic contribution is that we cast the problem of certifying envy, or the absence thereof, as a new form of pure exploration bandit problem, with conservative exploration constraints. The conservative exploration constraints prevent the audit from significantly deteriorating recommendation performance for users, when switching their recommendations with those of other users. We present OCEF, an auditing algorithm with theoretical guarantees on its sample complexity and the satisfaction of the conservative exploration constraints. We experimentally confirm that the OCEF algorithm is able to certify envy-freeness on two recommendation tasks, while maintaining a performance close to the audited recommender system.

We conclude the thesis in Chapter 7, where we recapitulate the main contributions of this thesis and present additional contributions. This chapter also includes a critical examination of the limitations imposed by our modeling choices, as well as a discussion of the insights gained and questions that remain open.

Chapter 2

Related work

Contents

In this chapter, we present a comprehensive overview of the various literature related to this thesis. The main chapters of the thesis also include their own specific related work sections from the original publications.

2.1 Fairness in rankings and recommender systems

Fairness in machine learning Fairness in machine learning is an active research area that has gained increasing attention in recent years [Barocas et al., 2019, Corbett-Davies and Goel, 2018, Oneto and Chiappa, 2020, Kusner and Loftus, 2020, Mitchell et al., 2021, Chouldechova and Roth, 2020, Kamiran and Calders, 2009]. This field of study started from the realization that machine learning algorithms, if not designed and implemented carefully, can produce biased outcomes that disproportionately affect sensitive groups of people. This can perpetuate existing inequalities in the distribution of the benefits and harms of machine learning applications, and reinforce societal biases and stereotypes in learned representations. A large part of this literature has been first focused on classification and scoring tasks [Mitchell et al., 2021, Chouldechova, 2017, Kleinberg et al., 2016].

The domain of fairness in machine learning is organized along two main axes. The first axis is whether fairness is oriented towards individuals or groups defined by sensitive or protected attributes, such as race, gender, age, or socioeconomic status [Barocas and Selbst, 2016]. Group

fairness considers differences in average outcomes between salient social groups of people, usually to prevent discriminatory decisions [Barocas and Selbst, 2016, Feldman et al., 2015, Hardt et al., 2016a]. Individual fairness means that the algorithm should treat individuals fairly regardless of their group membership, often by considering a similarity measure between individuals [Dwork et al., 2012, Bower et al., 2021]. In Chapters 3,4,5 we measure the distributive fairness of outcomes at the level of individuals, without notion of similarity. We made this presentation choice to make the framework simpler, but as discussed in Appendix A.2, our framework can be expanded to groups using aggregate measures. On the other hand, the fairness criterion of envy-freeness, which we adapt to personalized recommender systems in Chapter 6, is primarily oriented towards individuals.

The second axis of fairness in machine learning is whether fairness is a question of parity or preference-based. Parity means that predictions, or prediction errors, should be the same between groups or individuals. Preference-based fairness means that predictions are allowed to be different as long as they faithfully reflect the preferences of all parties involved [Ustun et al., 2019, Kim et al., 2018, Zafar et al., 2017b]. In this thesis, we consider preference-based notions of fairness for users, because they are aligned with personalization: They allow for different recommendations to different users, as long as these recommendations are in line with the preferences of users. In Chapters 3 and 4, we consider preference-based fairness for users as improving the utility of the worst-off users (following the Pigou-Dalton transfer principle from social choice), using a utility measure that depends on users' preferences (defined in Eq. (1.1), Section 1.2.1). In Chapter 6, we consider envy-freeness, another preference-based fairness criterion for users, also derived from fair division in social choice. Envy-freeness ensures that no user prefers the recommendations of others.

Fairness of exposure in recommender systems A large part of the literature on fairness in machine learning, especially at the beginning of its rapid expansion, focused on classification and regression tasks. Fairness in ranking and recommendation systems is a growing subfield of fairness in machine learning, which directly involves multiple stakeholders [Burke et al., 2018, Abdollahpouri et al., 2020]. In this literature review, we put more emphasis on *exposure-based* fairness, which is a line of work on fairness in rankings for recommendation and retrieval systems that intervenes on the *exposure* or *attention* given to items, depending on their *position bias*. In the context of recommender systems, fairness has been considered from the perspective of both users and item producers, as these two stakeholders have different interests and goals which are mediated by the recommender system. The work of this thesis addresses two-sided, exposure-based fairness in rankings for recommender systems.

On the user side, the question of fairness in rankings originated from independent audits on recommender systems or search engines, which showed that results could exhibit bias against salient social groups by representing or exagerating stereotypes [Mattioli, 2012, Sweeney, 2013, Kay et al., 2015, Hannak et al., 2014, Mehrotra et al., 2017, Lambrecht and Tucker, 2019, Datta et al., 2015, Asplund et al., 2020, Ali et al., 2019, Vlasceanu and Amodio, 2022]. In Chapter 6, we propose to complement these audits with an alternative user-side fairness criterion, namely envy-freeness. By measuring disparities which are aligned with user preferences, audits for envy-freeness can strengthen the conclusions of audits for recommendation parity. In the literature, another common goal for user-side fairness is to prevent disparities in recommendation performance across sensitive groups of users [Mehrotra et al., 2017, Ekstrand et al., 2018]. For example, it is important to ensure that recommender systems do not systematically recommend lower-quality or less relevant items to disadvantaged users. In Chapters 3 and 4, we seek a similar goal without requiring strict equality of recommendation performance, but rather by improving recommendation performance for the worst-off users.

On the item side, there is an active stream of research on ranking algorithms that promote fairness for individual or sensitive groups of items [Celis et al., 2017b, Burke et al., 2018, Biega et al., 2018, Singh and Joachims, 2018, Morik et al., 2020, Zehlike and Castillo, 2020, Kletti et al., 2022a,b, Beutel et al., 2019a, Narasimhan et al., 2020, Heuss et al., 2022, Diaz et al., 2020, Oosterhuis, 2021, García-Soriano and Bonchi, 2021, Sarvi et al., 2022], for example when ranking resumes of job applicants (items) to recruiters (users) [Geyik et al., 2019] or ranking music tracks (items) to listeners (users) [Mehrotra et al., 2018]. The goal is often to prevent winner-take-all effects, combat popularity bias [Abdollahpouri et al., 2019b], promote smaller producers [Liu et al., 2019, Mehrotra et al., 2018] or diverse representation [Zehlike et al., 2022a]. A branch of this literature aims to ensure a minimal proportion of items from sensitive groups are shown in the top-K positions of a ranking [Asudeh et al., 2019, Celis et al., 2017b, Zehlike et al., 2017]. In the other stream of *fairness of exposure*, authors proposed methods that redistribute exposure across (groups of) producers, either towards equal exposure, or equal ratios of exposure to a measure of merit [Singh and Joachims, 2018, Biega et al., 2018, Diaz et al., 2020, Kletti et al., 2022a]. These approaches may be applied either *within* the recommendation list of each user [Singh and Joachims, 2018, Yang and Stoyanovich, 2017, Celis and Vishnoi, 2017, Zehlike et al., 2017], or on average over all users (in the literature, this is referred to as *amortized* fairness of exposure) [Biega et al., 2018, Beutel et al., 2019a, Kletti et al., 2022a, Usunier et al., 2022, Prost et al., 2022]. Section 3.3 of Chapter 3 is devoted to the assessment of these exposure-based approaches in our welfare-based framework, through the lens of distributive justice. Our framework focuses on the case of amortized exposure, which is more computationally challenging because it couples the rankings of all the users. Moreover, the works that focused on "within-list" fairness are often motivated by recruitment, college admissions and search engines, rather than *personalized* recommendation [Singh and Joachims, 2018, Celis et al., 2017b, Asudeh et al., 2019, Zehlike et al., 2017]. We assess the fairness of some of these approaches in Appendix A.8. In the fair ranking literature, inequalities among items are often measured by the classical Gini index [Morik et al., 2020, Mansoury et al., 2021b, Wang et al., 2023, Ge et al., 2021]. We study a generalized version of this inequality measure in Chapter 4, and address the challenge of directly optimizing this nondifferentiable measure over the space of ranking policies.

Some authors consider fairness for both users and items, often by applying existing user or item fairness criteria simultaneously to both sides, such as [Basu et al., 2020, Wu et al., 2021b, Wang and Joachims, 2021, Naghiaei et al., 2022, Wu et al., 2022b]. We address the problem of two-sided fairness in ranked recommendations in Chapters 3 and 4. Section 6.3.3 of Chapter 6 of this thesis discusses the compatibility of envy-freeness as user-side fairness criterion with usual item-side fairness criteria. Patro et al. [2020] also considers envy-freeness in a two-sided fairness framework, while [Deldjoo et al., 2021] propose to use generalized cross-entropy to measure unfairness among sensitive groups of users and items. [Wu et al., 2022a] recently considered two-sided fairness in recommendation as a multi-objective problem, where each objective corresponds to a different fairness notion, either for users or items. Other works consider additional stakeholders and interests, such as platform revenue [Burke et al., 2018, Abdollahpouri et al., 2020, Abdollahpouri and Burke, 2019, Gharahighehi et al., 2021].

Finally, we highlight that the majority of works that address fairness of exposure for items focus on the position-based model [Singh and Joachims, 2018, Morik et al., 2020, Zehlike and Castillo, 2020, Biega et al., 2018, Oosterhuis, 2021], where the exposure of an item only depends on its rank. The linear structure of the position-based model is algorithmically convenient, because it allows to express user utilities and item exposures as linear quantities. In this thesis, we follow these works and propose computationally efficient algorithms for fair ranking that leverage this linear structure.

Only a few works consider exposure in more general cascade models [Craswell et al., 2008] and dynamic bayesian network models [Chuklin et al., 2015, Chapelle and Zhang, 2009]. These works often propose heuristic algorithms focusing on empirical insights [Biega et al., 2020, Mansoury et al., 2022, Jeunen and Goethals, 2021], except for Kletti et al. [2022b] who propose a theoretical algorithm for Pareto optimal trade-offs between user utility and item-side fairness, in a single-user setting.

Fairness in reciprocal recommendation Most of the works mentioned above consider usual one-sided recommendation settings, such as music or movie recommendation, where items and users are separate entities, and only items are being recommended. In reciprocal recommender systems [Pizzato et al., 2013], such as dating applications or friends recommendation, users are recommended to other users. Reciprocal recommender systems received comparatively less attention in the fairness literature, to the exception of [Jia et al., 2018, Xia et al., 2015, Paraschakis and Nilsson, 2020]. In Chapters 3 and 4, we present the first generic framework to jointly address one-sided and reciprocal recommendation. Xia et al. [2019] aim at equalizing user utility between groups, which suffers from the problems discussed in Section 3.3 of Chapter 3: Striving for perfectly equal user utilities can lead to lower utility for everyone, and even zero utility. Jia et al. [2018] generate rankings using a welfare function approach, but optimizing only the utility of users *being recommended*, while we introduce a notion of two-sided utility which also accounts for a user's satisfaction of the recommendations they receive. Paraschakis and Nilsson [2020] postprocess rankings to correct for inconsistencies between estimated and declared preferences of users. We do not aim at correcting biases in preference estimates through post-processing. In contrast, we aim at fair trade-offs between utilities, under the assumption that biases in the preference estimates have been addressed earlier in the recommendation pipeline. Fairness is also studied in the context of ridesharing applications [Wolfson and Lin, 2017, Lesmana et al., 2019, Nanda et al., 2020], but they address matching rather than ranking problems.

For exhaustive surveys on fairness in ranked recommendations, we refer to [Zehlike et al., 2022a,b, Patro et al., 2022, Deldjoo et al., 2022, Wang et al., 2023, Ekstrand et al., 2022, Abdollahpouri et al., 2020, Li et al., 2022, Chen et al., 2023].

2.1.1 Normative analysis

Overall, the focus of the thesis work has been on the algorithmic aspects of fair recommendation as an allocation problem, aiming to develop solutions that are robust and applicable across various scenarios, independent of our specific normative reasons. Most technical papers on fairness in recommender systems, including the ones we published over the course of the PhD program, do not explicitly discuss the underlying normative framework. We aim to bridge this gap in this section by providing a normative analysis of our approach to fairness in recommender systems. This normative analysis draws on the classification frameworks of normative judgements proposed by Zehlike et al. [2022a].

Classification frameworks of Zehlike et al. [2022a]. In this survey article on fairness in ranking, existing fairness interventions are examined beyond mere technical considerations and analyzed into the underlying value framework and socio-technical context. The authors explore four normative dimensions, each contributing to a comprehensive understanding of fairness in ranking.

The first dimension pertains to the notion of *group structure*, which encompasses factors such as the number of sensitive groups involved, their cardinality and how multiple sensitive attributes

are handled. Under this category, the survey analyzes various approaches to ensure fairness across groups in recommender systems.

The second dimension is about *bias types*, including preexisting biases and technical biases (e.g., position bias), which can influence the outcomes of recommender systems.

The third dimension revolves around different *worldviews* adopted when defining fairness, as proposed in the taxonomy of Friedler et al. [2016]. One such perspective is the "What You See Is What You Get" (WYSIWYG) worldview, which advocates that observable scores accurately reflect the true merit of individuals. The survey discusses *"merit-based fairness"* (presented in Section 1.3.5) as an example of an approach aligned with this worldview. On the other hand, the "We are All Equal" (WAE) viewpoint suggests that any disparities observed are a result of biased observations rather than inherent differences.

The fourth dimension explores the concept of *Equal Opportunity* (EO), which is a broad philosophical doctrine aimed at rectifying morally irrelevant circumstances in accessing opportunities. Within this dimension, the survey distinguishes between different version of EO. Formal EO focuses on fair competitions where candidates are evaluated solely based on their qualifications, rejecting any irrelevant attributes but not addressing disparities due to prior disadvantages. Formal-plus EO extends this by considering how certain attributes can lead to disparities in qualifications. Substantive EO takes a broader view, considering lifetime opportunities and attempting to mitigate the impact of arbitrary factors on relevant qualifications. Luck-egalitarian EO and Rawls' Fair EO are examples of substantive EO approaches, aiming to make people's future prospects comparable and improve outcomes for the most disadvantaged, respectively.

Normative analysis of our approach We now apply the classification framework of Zehlike et al. [2022a] to our approach of fair allocation of exposure in recommender systems.

First, we do not consider fairness towards explicit groups in our framework, but rather distributive fairness at the level of individuals, following the paradigm of fair division. We discuss how to extend our approach to fairness across groups in Section 7.2.1.

Second, we focus on exposure-based fairness in recommender systems, which compares individual items based on position bias, which is a type of technical bias. If we consider exposure at the level of sensitive groups as in Appendix A.2 (following existing works on fairness of exposure [Singh and Joachims, 2018]), then our techniques can be used to mitigate preexisting biases in the observed data that affect the scoring model, and lead to unequal exposure across groups of items.

On the axis of worldview, the approach undertaken in our work can be categorized as "We Are All Equal" (WAE), to some extent. In contrast to merit-based approaches to fairness in recommender systems, we strive to equalize outcomes for items without defining a notion of merit. Still, we avoid giving too much exposure to items that are irrelevant to some users in their rankings by considering global recommendation objectives as trade-offs between user welfare and item welfare. While we do not define talent or merit explicitly, our approach ensures that item producers with similar relevance receive comparable exposure. On the user side, our welfare-based approach do not strictly equalize user utilities, but rather redistribute utility among users regardless of any measure of their "merit", and without destroying total utility.

Lastly, our work relates to the *priority view* or *prioritarianism* in political philosophy, which asserts that *"social welfare orderings should give explicit priority to the worse off"* [Temkin, 1993, Arneson, 2000, Parfit, 2018]. Prioritarianism is a concept distinct from Equal Opportunity and is often seen in contrast to strict egalitarianism. Prioritarianism is often associated to concave welfare functions in economics [Fleurbaey, 2015], which provide social welfare orderings that give explicit priority to the worst off individuals, and where the strength of the curvature controls the degree

of priority. The *absolute* view of prioritarianism considers that the importance of an individual should not depend on their relative position. This view aligns with the additively separable concave welfare functions from welfare economics that we consider in Chapter 3. In contrast, the *relative* view of prioritarianism considers the utility of individuals in relation to others: This includes non-additive welfare function such as the genaralized Gini welfare function that we study in Chapter 4 [Fleurbaey, 2015]. In our research, EO is not a primary consideration, as the main axiom in welfarism and social choice is anonymity, which conflicts with considering explicit attributes to identify individuals' circumstances. Since the main presentation of our fair allocation framework does not involve explicitly choosing groups, the language of Equal Opportunity is difficult to apply. The closest to our approach is Rawls' Fair EO, particularly in the aspect of trading-off between user welfare and exposure redistribution among items. This allows us to ensure that items of similar relevance receive comparable exposure, without defining a measure of "merit" or "effort". This compromise aligns with the Rawlsian Fair EO principle, which advocates for equal prospects of success among equally talented individuals, irrespective of arbitrary circumstances.

2.2 Frank-Wolfe algorithms

The backbone of the algorithmic contributions of this thesis is the family of Frank-Wolfe algorithms [Frank and Wolfe, 1956]. In this thesis, we show that they provide computationally efficient algorithms for fair ranking in the position-based model. In Chapter 3, we show how to use a vanilla Frank-Wolfe algorithm to generate rankings with smooth concave welfare functions in the batch setting, using only one top-K sorting operation per user of the batch at each iteration of the algorithm. In Chapter 4, we design an efficient smoothing method for the class of Generalized Gini welfare functions, which are non-differentiable, and we show how to apply a Frank-Wolfe variant for nonsmooth objectives from Lan [2013]. This section provides background on the Frank-Wolfe algorithms and smoothing techniques that we used *for the batch setting*, and situates our algorithmic contributions with respect to related techniques and existing algorithms for fair ranking.

We also use Frank-Wolfe in Chapter 5, to generate rankings with smooth and nonsmooth concave welfare functions in the contextual bandit setting, obtaining a fast algorithm that delivers fair rankings at the same cost as standard ranking-by-sorting algorithms. We discuss related approaches in the dedicated section on bandits with concave rewards (Section 2.3.2).

2.2.1 Background on Frank-Wolfe and algorithms for fair ranking

Background on Frank-Wolfe. The Frank-Wolfe algorithm [Frank and Wolfe, 1956], also known as the conditional gradient method, is an iterative optimization algorithm used for solving constrained convex optimization problems. Although it has been extensively used in machine learning applications, such as structured output prediction and low-rank matrix completion [Jaggi, 2013], to the best of our knowledge, it has not been used for ranking prior to the work of this thesis.

Consider a convex optimization problem of the form:

$$\max_{P \in \mathcal{P}} F(P), \tag{2.1}$$

where F is a smooth and concave function defined over a compact convex set $\mathcal{P}$. The Frank-Wolfe algorithm generates a sequence of solutions $P^{(t)}$ iteratively, where $P^{(t)}$ is the solution obtained after t iterations. The algorithm iteratively computes $\tilde{P}$ by solving the following linear optimization

problem:

$$\tilde{P} = \operatorname*{argmax}_{P \in \mathcal{P}} \langle P \,|\, \nabla F(P^{(t)}) \rangle, \tag{2.2}$$

where $\nabla F(P^{(t)})$ is the gradient of F evaluated at $P^{(t)}$.

The algorithm then updates the solution by performing a convex combination of the current solution and the newly obtained solution, i.e.,

$$P^{(t)} = (1 - \gamma^{(t)}) P^{(t-1)} + \gamma^{(t)} \tilde{P},$$

where $\gamma^{(t)} = \frac{2}{t+2}$ [Clarkson, 2010]. Notably, the algorithm always remains in the feasible region *without the need for any additional projection step.*

The Frank-Wolfe algorithm is particularly useful for simplex-type constraints [Clarkson, 2010]. In this case, each $\tilde{P}$ computed in the linear subproblem of (2.2) is a single element of the simplex. The Frank-Wolfe algorithm then constructs a solution that has a *sparse representation.*

An algorithmic contribution: Frank-Wolfe for fair ranking. The Frank-Wolfe algorithm is best used when $\operatorname{argmax}_{P \in \mathcal{P}} \langle P \,|\, \nabla F(P^{(t)}) \rangle$ (Eq. (2.2)) can be computed efficiently. In Chapter 3, we show that this is the case when $\mathcal{P}$ is the set of stochastic ranking policies. More precisely, we prove that the inner loop of Frank-Wolfe consists in computing one ranking for each user, which can be obtained with a straightforward *top-K sort operation per user*, when:

- $\mathcal{P}$ is the convex hull of tensors P where each slice P_i is a permutation matrix for one user i[1],
- and the concave objective F depends on user and item utilities defined in a position-based model with non-increasing weights.

Recalling that m is the number of items and K the number of ranking slots, each iteration of the Frank-Wolfe algorithm has a $O(m + K \ln K)$ time cost per user (as formally stated in Proposition 5 of Chapter 3 and Proposition 11 of Chapter 4). Importantly, this provides us with an algorithm that *decentralizes the computation of rankings across users*, while the main technical challenge brought by item-side fairness of exposure is the coupling of the users' rankings (since the exposure of an item is calculated across all rankings). This result also guides us towards the development of fast ranking algorithms in the *online setting* where users are served one at a time, which we present in Chapter 5.

Moreover, the algorithm outputs a *sparse* representation of a stochastic ranking policy, as a *convex combination of deterministic ranking policies.* Standard Frank-Wolfe convergence results guarantee that the algorithm finds an ϵ-optimal solution of the problem in (2.1) at a sublinear rate [Jaggi, 2013, Clarkson, 2010].

Comparison with existing algorithms for fairness of exposure in rankings. The usage of stochastic rankings was initiated by Singh and Joachims [2018] in the context of fairness in rankings to make inference a convex optimization problem. Singh and Joachims [2018] however considered a notion of item fairness applied within the ranking of each user separately, while we consider amortized fairness across users, similarly to [Morik et al., 2020, Biega et al., 2018, Kletti et al., 2022a]. Thus, they did not need to infer globally optimal ranking policies, and their optimization problem involved m^2 variables with m items, which was tractable in their case. In the *amortized* fairness setting, the optimization problem of Singh and Joachims [2018] would involve $n \times m^2$ variables where n is a typically large number of users. Our Frank-Wolfe approach is thus more efficient for amortized fairness in the batch setting.

[1] Or equivalently, the convex set of tensors P where each slice P_i is a bistochastic matrix for user i.

Moreover, the usage of stochastic rankings usually requires a post-processing stage in order to sample deterministic rankings. Singh and Joachims [2018] use the Birkhoff-von Neumann decomposition [Birkhoff, 1940], which decomposes a bistochastic matrix as a convex sum of permutations, with at most $(m-1)^2+1$ members in the decomposition, and Kletti et al. [2022a] propose an improved decomposition with m terms only. With the Frank-Wolfe algorithm, we remove the need for an additional decomposition step since Frank-Wolfe directly constructs a weighted sum of deterministic ranking policies.

In the line of works considering amortized fairness of exposure over multiple rankings, Wang and Joachims [2021], Kletti et al. [2022a] do not consider *personalized* ranking policies with one (stochastic) ranking for each user. Since the number of users is not an input variable in their problem settings, they do not seek algorithms that are scalable with respect to this variable. Biega et al. [2018] focus on amortized fairness across a fixed number n of rankings, and thus face the challenge of finding a global solution for the n rankings that are coupled by the items' exposures', similarly to us. They bypass this challenge by solving a linear program for each ranking separately, but this heuristic offers no global guarantee on items' exposures across rankings. In contrast, our Frank-Wolfe approach provably finds ϵ-optimal solutions to global optimization problems that consider the rankings of all users. In an online ranking setting, Morik et al. [2020] propose an algorithm for amortized fairness across rankings which ensures that a specific item-side disparity measure converges to zero. However, their algorithm cannot be used to optimize intermediate trade-offs between average user utility and item-side fairness, or two-sided fairness objectives. Moreover they do not provide guarantees on the users' utilities'. In contrast, the Frank-Wolfe algorithm provably maximizes concave functions that express a wide range of fairness-aware objectives as trade-offs between users' and items' utilities, including the merit-based fairness criterion of Biega et al. [2018], Morik et al. [2020].

Finally, Patro et al. [2020] consider a recommendation setting that is similar to our batch setting, where the recommender systems produces one personalized list of items for each users, with two-sided fairness considerations. Our Frank-Wolfe approach improves over their method in three ways. First, they considered unordered lists, while we consider the more challenging task of finding ranked lists, which increases the search space. Second, the complexity of their round robin algorithm is $O(nmK)$, and it is neither amenable to paralellization across the n users, nor adaptable to an online setting where users are observed in sequence. In contrast, our Frank-Wolfe algorithm decentralizes the top-K ranking operations for each user, and we adapt it to the online bandit setting in Chapter 5. Third, Patro et al. [2020]'s algorithm is limited to specific fairness criteria, while Frank-Wolfe allows for a broader variety of fairness-aware objectives.

2.2.2 Frank-Wolfe with smoothing

In Chapter 4, we propose to optimize welfare functions based on generalized Gini welfare functions (GGFs) [Weymark, 1981], which are ordered weighted averages of utilities, parameterized by a vector of non-increasing weights. Since these concave functions are non-differentiable, they cannot be optimized using the previous vanilla Frank-Wolfe algorithm. The technical contribution of Chapter 4 builds on nonsmooth convex optimization methods [Nesterov, 2005, Shamir and Zhang, 2013], and in particular variants of the Frank-Wolfe algorithm [Frank and Wolfe, 1956, Jaggi, 2013] for nonsmooth problems [Lan, 2013, Yurtsever et al., 2018, Ravi et al., 2019, Thekumparampil et al., 2020a]. The recent algorithm of [Thekumparampil et al., 2020a] is a Frank-Wolfe variant that uses the Moreau envelope like us. Its number of first-order calls is optimal, but this is at the cost of a more complex algorithm with inner loops that make it slow in practice. In our case, since

the calculation of the gradient is not a bottleneck, we use the simpler algorithm of Lan [2013], which applies Frank-Wolfe to the Moreau-Yosida envelope [Moreau, 1962, Yosida et al., 1965] of the nonsmooth objective.

The technical contribution of Chapter 4 is also related to the literature on differentiable ranking, which includes a large body of work on approximating learning-to-rank metrics [Chapelle and Wu, 2010, Taylor et al., 2008, Adams and Zemel, 2011], and recent growing interest in designing smooth ranking modules [Grover et al., 2019, Cuturi et al., 2019, Blondel et al., 2020] for end-to-end differentiation pipelines. The closest method to the algorithm that we present in Chapter 4 is the differentiable sorting operator of Blondel et al. [2020]. Blondel et al. [2020] use a regularization term to smooth the linear formulation of sorting. The regularized form can itself be written as a projection to a permutahedron, which can be efficiently computed using a well-known reduction to isotonic regression [Negrinho and Martins, 2014, Lim and Wright, 2016]. The problem they address is different since they differentiate the multi-dimensional sort operation, but eventually the techniques are similar to the ones we use because the smoothing is done in a similar way. In our case, the projection onto a permutahedron appears in the gradient of the GGF, rather than the GGF itself, which is important to unlock the result of Proposition 10. This is the key to fast Frank-Wolfe iterations in our optimization problem over stochastic ranking policies. Moreover, the weights of the GGF are also important in our case as they affect the Frank-Wolfe convergence guarantee, while Blondel et al. [2020] assign equal weights to utilities.

2.3 Bandit algorithms for fair and multi-objective recommender systems

In Chapter 5, we focus on the problem of online learning with bandit feedback and multiple rewards, where the desired trade-off between the rewards is defined by a known concave objective function. This problem is referred to as *Bandits with Concave Rewards* (BCR) [Agrawal and Devanur, 2014]. We provide regret guarantees for the Contextual setting of BCR (CBCR), where the vector of multiple rewards depends on a stochastic context. This setting is particularly relevant to fair machine learning in recommender systems and online allocation problems, where the overall welfare is naturally expressed as a (known) concave function of the (unknown) utilities of the agents [Moulin, 2003, Berthet and Perchet, 2017, Do et al., 2021c]. We review the literature on fairness in bandit-based recommendation in Section 2.3.1 and the literature on bandits with multiple rewards in Section 2.3.2.

Finally, we discuss the pure exploration bandit setting in Section 2.3.3, which is different from minimizing regret and is useful for the fairness certification problem of Chapter 6.

2.3.1 Fairness of exposure in bandits

In Chapter 5, we address the question of fairness of exposure in the contextual bandit setting, which is a popular paradigm for recommender systems that learn to generate personalized recommendations from online interations with users [Li et al., 2010, Lattimore and Szepesvári, 2020]. On the one hand, contextual bandit algorithms have been mostly developed to maximize a single scalar reward. In the case of recommendation, this reward usually corresponds to a proxy of user satisfaction based on engagement signals (e.g., clicks, shares, likes, etc.), and it thus ignores the impact of recommendations on item producers. On the other hand, most of the item-side fairness literature, which we reviewed in Section 2.1, focused on a *static* ranking setting, either without learning [Geyik et al., 2019, Beutel et al., 2019a, Yang and Stoyanovich, 2017, Singh and Joachims, 2018, Patro

et al., 2022, Kletti et al., 2022a, Diaz et al., 2020, Do and Usunier, 2022, Wu et al., 2022b] or with learning-to-rank [Bower et al., 2021, Singh and Joachims, 2019, Zehlike and Castillo, 2020].

Existing work on fairness of exposure in stochastic bandits focused on local exposure constraints on the probability of pulling an arm at each timestep, either in the form of lower/upper bounds [Celis et al., 2018b] or merit-based exposure targets [Wang et al., 2021a]. In contrast, we consider amortized exposure over time, in line with prior work on fair ranking [Biega et al., 2018, Morik et al., 2020, Usunier et al., 2022], along with fairness trade-offs defined by concave objective functions which are more flexible than fairness constraints [Zehlike and Castillo, 2020, Do et al., 2021c, Usunier et al., 2022, Wu et al., 2022a]. Moreover, these works [Celis et al., 2018b, Wang et al., 2021a] do not address combinatorial actions, while ours applies to ranking in the position-based model, which is more practical for recommender systems [Lagrée et al., 2016, Singh and Joachims, 2018]. The methods of [Patil et al., 2020, Chen et al., 2020] aim at guaranteeing a minimal cumulative exposure over time for each arm, but they also do not apply to ranking. In contrast, [Xu et al., 2021, Li et al., 2019] consider combinatorial bandits with fairness, but they do not address the contextual case, which limits their practical application to recommender systems. Mansoury et al. [2021a], Jeunen and Goethals [2021] propose heuristic algorithms for fairness in ranking in the contextual bandit setting, highlighting the problem's importance for real-world recommender systems, but these algorithms lack theoretical guarantees. In Chapter 5, we introduce the first principled bandit algorithms for this problem with provably vanishing regret.

Finally, several works on fairness in bandits focus on hiring rather than personalized recommendation [Joseph et al., 2016, Liu et al., 2017, Schumann et al., 2019b]. These works study criteria that are different from fairness of exposure since the goals and tasks involved are distinct.

2.3.2 Bandits with concave rewards

Several recent works on the societal impact of recommender systems and machine learning algorithms have advocated for the optimization of multiple rewards, instead of focusing on a single reward [Mehrotra et al., 2020, Stray et al., 2021, Vamplew et al., 2018]. The desired trade-off between the rewards is typically defined by a known concave function f, which is set by the practitioner depending on the application context [Mehrotra et al., 2020]. In the multi-objective bandit literature, the optimization of a known concave function of different rewards is known as Bandits with Concave Rewards (BCR) [Agrawal and Devanur, 2014]. Chapter 5 is dedicated to the Contextual setting of BCR (CBCR), where the vector of multiple rewards depends on a stochastic context.

The main challenge of CBCR is that the set of stationary policies are all mappings from a continuous context set to distributions over actions. In the non-contextual (BCR), which has been previously studied by Agrawal and Devanur [2014], and by Busa-Fekete et al. [2017] for the special case of Generalized Gini indices, policies are distributions over actions. These approaches perform a direct optimization in policy space, which is not possible in the contextual setup without restrictions or assumptions on optimal policies. Agrawal et al. [2016] study a setting of CBCR where the goal is to find the best policy in a finite set of policies. Because they rely on explicit search in the policy space, they do not resolve the main challenge of the general CBCR setting we address in Chapter 5. Cheung [2019], Siddique et al. [2020], Mandal and Gan [2022], Geist et al. [2021] address multi-objective reinforcement learning with concave aggregation functions, a problem more general than stochastic contextual bandits. In particular, Cheung [2019] use a Frank-Wolfe approach for this problem. However, these works rely on a tabular setting (i.e., finite state and action sets) and explicitly compute policies, which is not possible in our setting where policies are mappings from a continuous context set to distributions over actions. Our work is the only one

amenable to contextual bandits with concave rewards by removing the need for an explicit policy representation. Finally, compared to previous Frank-Wolfe approaches to bandits with concave rewards, e.g. [Agrawal and Devanur, 2014, Berthet and Perchet, 2017], our analysis is not limited to confidence-based exploration/exploitation algorithms.

CBCR is also related to the broad literature on bandit convex optimization (BCO) [Flaxman et al., 2004, Agarwal et al., 2011, Hazan et al., 2016, Shalev-Shwartz et al., 2012]. In BCO, the goal is to minimize a cumulative loss of the form $\sum_{t=1}^{T} \ell_t(\pi_t)$, where the convex loss function ℓ_t is *unknown* and the learner only observes the value $\ell_t(\pi_t)$ of the chosen parameter π_t at each timestep. Existing approaches to BCO perform gradient-free optimization in the parameter space. While BCR considers global objectives rather than cumulative ones, similar approaches have been used in non-contextual BCR [Berthet and Perchet, 2017] where the parameter space is the convex set of distributions over actions. As we previously highlighted, such parameterization does not apply to CBCR because direct optimization in policy space is infeasible.

CBCR is also related to multi-objective optimization [Miettinen, 2012, Drugan and Nowe, 2013], where the goal is to find all Pareto-efficient solutions. (C)BCR, focuses on one point of the Pareto front determined by the concave aggregation function f, which is more practical in our application settings where the decision-maker is interested in a specific (e.g., fairness) trade-off.

2.3.3 Pure exploration

We also leverage the multi-armed bandit paradigm for the online certification problem of Chapter 6, in which an auditor must collect user feedback to certify the preference-based fairness criterion of envy-freeness. However, unlike in Chapter 5, we focus on a pure exploration problem, rather than the regret minimization setting. The regret minimization problem deals with the exploration-exploitation trade-off, where bandit algorithms aim to achieve a cumulative performance at any time that is as close as possible to the optimal achievable performance [Robbins, 1952, Auer et al., 2002, Bubeck and Cesa-Bianchi, 2012]. In contrast, in the online certification setting, the goal is not to design a recommender system with cumulative performance guarantees, but rather to audit and evaluate an existing recommender system. We model this audit as a *pure exploration* problem, where the bandit algorithm must present a certificate regarding the arms after an exploration phase that should be as short as possible, and without taking cumulative performance into account. In our case, the certificate indicates whether an arm is better than the baseline, which is the audited recommender system.

The conservative exploration setting [Wu et al., 2016, Garcelon et al., 2020a], which was introduced for the regret minimization problem, adds the constraint that the anytime average performance should not be far worse than that of a special arm called the baseline. In Chapter 6, the baseline is the current recommender system, and the other "arms" are other users' personalized recommendations. The goal is to output a certificate indicating if an arm is better than the baseline, while not deteriorating performance compared to the baseline. We thus use a mix of pure exploration and conservative constraints.

In pure exploration, the most studied task is best-arm identification [Even-Dar et al., 2006, Gabillon et al., 2012, Audibert and Bubeck, 2010, Garivier and Kaufmann, 2016]. In Chapter 6, the problem is not to find the best arm but rather to decide whether an arm is better than the baseline, which is less demanding. The online certification problem is closer to threshold bandits [Locatelli et al., 2016], where the goal is to identify the set of arms with higher reward than a fixed threshold. Combinatorial pure exploration bandits [Chen et al., 2014] and multiple testing [Jamieson and Jain, 2018] address similar problems.

The differences with these settings, in addition to the conservative constraint, are twofold. First, they assume the threshold is known, i.e., the baseline performance is known, which we do not. Second, they aim at finding all arms that are better than the threshold, rather than deciding if there is such an arm. Although ideas from these works may be valuable in our context, we focused on ideas from best-arm identification methods [Audibert and Bubeck, 2010] to keep our proposed auditing algorithm and its analysis as simple as possible.

2.4 Background on fair division in social choice

Social choice theory is the study of collective decision-making based on the preferences of agents over alternatives [Arrow et al., 2010]. Fair division is one of the main branches of social choice, which addresses the allocation of resources among several agents in a manner that satisfies efficiency and fairness criteria [Moulin, 2003]. While the majority of this literature was first developed in microeconomics, it benefited from recent developments in computer science, with the rise of the field of computational social choice.

Fairness notions have been thoroughly examined from the perspective of distributive justice [Sen, 1970, Roemer, 1996]. The social choice literature proposed a variety of ways to translate these notions into mathematical definitions, and analyse their properties. In this section, we provide background on important fairness and efficiency criteria for allocation problems in social choice, in which the conceptual framework developed in this thesis is grounded. We discuss their axiomatic foundations and properties, as well as their interpretations in the context of recommender systems.

We refer to [Moulin, 2003] for comprehensive overviews of fair division, and to [Bouveret et al., 2016] for a recent survey of fair division in computational social choice.

2.4.1 Cardinal welfarism

Fair division is a social choice problem in which an *alternative* (or an *allocation*) is chosen from a set of feasible alternatives based on the individual preferences of a group of agents.

Problem definition. Classical fair division problems can be classified into two types, based on the nature of goods involved: either indivisible goods (e.g., books) or divisible goods (e.g., a cake). The inputs of a fair division problem are the set of agents, the set of goods, and the preferences of the agents over the goods, which are expressed as numerical values in cardinal welfare economics (rather than ordinal preferences). The output is an allocation of goods to the agents which specifies which (share of) goods are given to which agents (*"who gets what"*).

Let us now describe the inputs and outputs of our problem of fair allocation of exposure in recommender systems, focusing on the non-reciprocal setting[2]. We have the following inputs:

- The set of agents are the users *and* the items;
- The good with limited availability is the total exposure in the rankings (or equivalently, the slots in every users' ranking);
- Users have heterogeneous preferences over items, quantified by the relevance scores μ_{ij}, and they obtain higher utility when higher exposure is given to relevant items in their *own* ranking. Items all have the same preference for high exposure (or equivalently, slots in higher positions in *all* users' rankings).

[2]We also address the special case of reciprocal recommendation where items have preferences over users, in Chapters 3 and 4.

The output is a ranking policy that defines one ranking of items per user, and it is chosen from a set of alternatives which is the set of stochastic ranking policies. A ranking policy specifies which users' attention is given to which items. In fair division terms, it allocates to each item a share of total exposure, and to each user a ranking of items. Although the inputs and outputs are not standard for fair division, fair recommendation can still be framed as a fair division problem.[3]

Normative properties. The most important property is the criterion of Pareto efficiency. According to Moulin [2003], in distributive justice, *"its desirability is undisputed"*. In words, an allocation is Pareto-efficient if there is no other feasible allocation that would make at least one agent strictly better off while not making any of the others worse off.

We formally define Pareto-efficiency and Lorenz-efficiency by introducing the following notation. In this section, we denote by n the total number of agents, A a generic alternative and $\mathcal{A}$ the set of alternatives (which would be the set of stochastic ranking policies $\mathcal{P}$ in the fair allocation of exposure problem). We denote by $\mathcal{U} = \{(u_i(A))_{i=1}^n : A \in \mathcal{A}\}$ the set of achievable utility profiles. Let $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^n$ two utility profiles. We write $\boldsymbol{u} \succeq_{\mathrm{P}} \boldsymbol{v}$ if $\forall i \in [\![n]\!], u_i \geq v_i$. We say that $\boldsymbol{u}$ *Pareto-dominates* $\boldsymbol{v}$, noted $\boldsymbol{u} \succ_{\mathrm{P}} \boldsymbol{v}$, if $\boldsymbol{u} \succeq_{\mathrm{P}} \boldsymbol{v}$ and $\exists i, u_i > v_i$. Given $\mathcal{U} \subseteq \mathbb{R}^n$, $\boldsymbol{u} \in \mathcal{U}$ is said to be *Pareto-efficient* in $\mathcal{U}$ if no $\boldsymbol{v} \in \mathcal{U}$ Pareto-dominates $\boldsymbol{u}$, i.e., if $\forall \boldsymbol{v} \in \mathcal{U}, \neg(\boldsymbol{v} \succ_{\mathrm{P}} \boldsymbol{u})$. Similarly, $A \in \mathcal{A}$ is said to be *Pareto-efficient* if $\boldsymbol{u}(A)$ is Pareto-efficient.

We now describe the criterion of Lorenz efficiency which is at the core of the framework of Chapter 3. Let $(u_i^{\uparrow})_{i=1}^n$ (resp. $(v_i^{\uparrow})_{i=1}^n$) be the values in $\boldsymbol{u}$ (resp. $\boldsymbol{v}$) sorted in ascending order, i.e., $u_{(1)} \leq \ldots \leq u_{(n)}$. We write $\boldsymbol{u} \succeq_{\mathrm{L}} \boldsymbol{v}$ if $\forall k \in [\![n]\!], u_1^{\uparrow} + \ldots + u_k^{\uparrow} \geq v_i^{\uparrow} + \ldots + v_k^{\uparrow}$. We say that $\boldsymbol{u}$ *Lorenz-dominates* $\boldsymbol{v}$, denoted by $\boldsymbol{u} \succ_{\mathrm{L}} \boldsymbol{v}$, if $\boldsymbol{u} \succeq_{\mathrm{L}} \boldsymbol{v}$ and $\exists k, u_1^{\uparrow} + \ldots + u_k^{\uparrow} > v_i^{\uparrow} + \ldots + v_k^{\uparrow}$. $\boldsymbol{u} \in \mathcal{U}$ is said to be *Lorenz-efficient* in $\mathcal{U}$ if $\forall \boldsymbol{v} \in \mathcal{U}, \neg(\boldsymbol{v} \succ_{\mathrm{L}} \boldsymbol{u})$ [Shorrocks, 1983]. Similarly, $A \in \mathcal{A}$ is said to be *Lorenz-efficient* if $\boldsymbol{u}(A)$ is Pareto-efficient. Note that Lorenz efficiency implies Pareto efficiency.

The Lorenz dominance preorder is closely related to the mathematical notion of *majorization* [Hardy et al., 1952]. The definition of majorization is in fact the same as Lorenz dominance, except that the utilities are sorted in decreasing order instead of increasing order.

Cardinal social welfare functions. Given a set of possible alternatives $\mathcal{A}$, a *cardinal social welfare function* or more simply, a *welfare function*[4] F maps the utility profile $(u_i(A))_{i=1}^n$ for $A \in \mathcal{A}$ to a real value that represents the aggregate preference of all individuals for the alternative A. Socially preferred alternatives are those which maximize F over $\mathcal{U} = \{(u_i(A))_{i=1}^n : A \in \mathcal{A}\}$.

We now describe useful properties of social welfare functions F which relate to the efficiency of their maximizers. Let $F : \mathrm{dom}(F) \subseteq \mathbb{R}^n \to \mathbb{R}$ and $(\boldsymbol{u}, \boldsymbol{v}) \in \mathrm{dom}(F)^2$. F is *monotonic* if $\boldsymbol{u} \succeq_{\mathrm{P}} \boldsymbol{v} \Rightarrow F(\boldsymbol{u}) \geq F(\boldsymbol{v})$. F is *Schur-concave* if $\sum_{i=1}^n u_i = \sum_{i=1}^n v_i$ and $\boldsymbol{u} \succeq_{\mathrm{L}} \boldsymbol{v} \implies F(\boldsymbol{u}) \geq F(\boldsymbol{v})$. These extend to natural strict criteria of strict monotonicity ($\boldsymbol{u} \succ_{\mathrm{P}} \boldsymbol{v} \implies F(\boldsymbol{u}) > F(\boldsymbol{v})$) and strict Schur-concavity ($\sum_i u_i = \sum_i v_i$ and $\boldsymbol{u} \succ_{\mathrm{L}} \boldsymbol{v} \implies F(\boldsymbol{u}) > F(\boldsymbol{v})$). Given $\mathcal{U} \subseteq \mathrm{dom}(F)$, the definitions imply that for every $\boldsymbol{u} \in \mathrm{argmax}_{\boldsymbol{v} \in \mathcal{U}} F(\boldsymbol{v})$, if F is monotonic, then $\boldsymbol{u}$ is Pareto-efficient, and if F is both monotonic and strictly Schur-concave, then $\boldsymbol{u}$ is Lorenz-efficient [Shorrocks, 1983, Thistle, 1989].

[3] At an abstract level, recommendation is also reminiscent of many-to-one matchings, which are also largely studied in social choice. The output of a many-to-one matching would be an assignment of a set of items ("many") to a user ("one"). However, these matchings problems do not deal with exposure or users' attention as a scarce resource, while the primary goal of recommender systems is to support users with limited attention. We discuss the relationship between matching and recommendation in more detail in Chapter 7.

[4] An alternative naming is *collective utility function* as in [Moulin, 2003]. These functions induce a *social welfare ordering*, which is a binary relation over utility vectors that is reflexive, transitive and complete. In the social choice literature, the term *welfare function* is sometimes used for *social welfare ordering*, while we use it to refer to a *collective utility function*, following the common usage of "utilitarian welfare function" or "Nash welfare function".

The axiomatic approach. The axiomatic approach to cardinal welfare economics specifies desirable properties of social welfare functions based on the axioms of symmetry, continuity, independence of unconcerned agents and independence to scale, defined in [Moulin, 2003].

A fundamental result of axiomatic cardinal welfare economics is that welfare functions $F : \mathbb{R}_+^n \to \mathbb{R} \cup \{-\infty\}$ that satisfy monotonicity, symmetry, continuity, independence of unconcerned agents and independence to scale are additive and have the following form for $\alpha \in \mathbb{R}$:

$$W_\alpha(\boldsymbol{u}) = \sum_{i=1}^{n} \psi(u_i; \alpha) \quad \text{where} \quad \psi(x; \alpha) = \begin{cases} x^\alpha & \text{if } \alpha > 0 \\ log(x) & \text{if } \alpha = 0 \\ -x^\alpha & \text{if } \alpha < 0 \end{cases}. \tag{2.3}$$

In Chapter 3, we propose to find ranking policies by maximizing these additive welfare functions. More precisely, we maximize trade-offs between the welfare of users and the welfare of items, where the welfare on each side is defined by a function of the form of $W_\alpha(\boldsymbol{u})$.

Fairness: the Pigou-Dalton transfer principle and Lorenz efficiency. The fundamental axiom of fairness in cardinal welfare economics is the Pigou-Dalton transfer principle. It states that social welfare increases when we redistribute utility from a better-off individual to a worse-off, keeping others' utilities and the overall sum of utilities constant. This transfer principle is mathematically equivalent to Schur-concavity, which holds for welfare functions of the form (2.3) above when $\alpha \leq 1$ [Hardy et al., 1952, Marshall et al., 1979, Muirhead, 1902]. Since $\alpha = 1$ corresponds to the pure utilitarian welfare function, which is neutral with respect to mean-preserving redistributions of utilities, the Pigou-Dalton principle holds strictly for strictly Schur-concave functions, and thus for $\alpha < 1$ in (2.3). Then, socially preferred alternatives are Lorenz-efficient, which follows the fundamental concept underlying the welfarist measurement of inequalities [Shorrocks, 1983, Hardy et al., 1952]: a distribution of utility $\boldsymbol{u}$ which Lorenz-dominates a distribution $\boldsymbol{v}$ with the same mean is seen as more equitable since a larger share of utility is held by the worse-off individuals.

Assessing recommender systems in light of the Pigou-Dalton transfer principle (or with Lorenz efficiency) is useful to prohibit rich-gets-richer effects and promote less visible item producers. Since it favours utility transfers to the worst-off, it prevents ranking policies that unfairly put the burden of item-side redistrbution on the worst-off users.

Utilitarianism with diminishing returns. An alternative point of view that yields the same social preferences is to consider that $u_i(A)$ is not the underlying utility of i, but rather the amount of "value" received under A, and the true underlying utility of i exhibits *diminishing returns* with respect to $u_i(A)$. Assuming the same diminishing marginal utility curve for every individual, social welfare functions of the form (2.3) are then utilitarian social welfare functions, where α controls the diminishing marginal utility (see e.g., the discussion by Atkinson et al. [2015]). From either perspective (non-utilitarian/equity or utilitarian/diminishing marginal utility), Lorenz-efficiency of profiles $(u_i(A))_{i=1}^n$ is the refinement of Pareto-efficiency that leads to socially preferred outcomes. In other words, Lorenz-efficiency allows to choose between Pareto-efficient solutions.

The concept of diminishing returns is especially applicable to recommender systems. This is because item producers experience diminishing returns as they receive more exposure to users: "One extra view counts less for a producer with 10 million views than for one with only 10 views." This concept aligns with the goal of promoting smaller item producers and making them sustainable.

2.4.2 Inequality indices, welfare functions and Lorenz curves

An important class of fairness criteria aims to quantify the level of economic inequality caused by a given alternative. These criteria are based on inequality indices, which are often associated with welfare functions. In Chapter 4, we apply the class of generalized Gini welfare functions to the fair ranking problem, which are linked to the widely used Gini index for measuring inequality.

In this section, we present inequality indices and how they relate to welfare functions. Then we present the Gini index and other inequality indices, and discuss their properties. The Lorenz curve, a graphical representation of utility profiles, is also discussed, as is its connection to the Gini index. For a more detail survey on inequality measures, we refer the reader to Cowell [2000] for the theory of inequality measures in welfare economics and to Chakravarty et al. [2009] for an extensive survey of measures used in practice.

2.4.2.1 Inequality indices and welfare functions

Inequality indices are functions that measure the level of inequality in a population, in particular inequality of wealth or income. An important perspective on inequality is that more inequality in a society causes a loss of social welfare [Atkinson, 1970]. In this view, choosing an inequality index is similar to selecting a welfare function: it involves making a normative judgment.

Inequality indices like the well-known Gini index are often associated with welfare functions. In practice, inequality indices are typically used as *evaluation measures*, while welfare functions are used to *decide* on an allocation. In Chapter 4, we focus on the Gini index, which is also commonly used in recommendation papers to measure unfairness, and we maximize its associated welfare function, which we present in the following subsection.

The formal connection between inequality indices and welfare function is made in [Atkinson, 1970, Cowell, 2000]. In the following, we consider a generic population of n individuals and their utilities $\boldsymbol{u} \in \mathbb{R}^n$. An inequality index takes as input a utility profile $\boldsymbol{u}$ and outputs a measure $I(\boldsymbol{u}) \in \mathbb{R}$. The welfare function $W : \mathbb{R}_+^n \to \mathbb{R}$ associated to an inequality index I is an increasing function of the mean of utilities $\bar{\boldsymbol{u}} = \frac{1}{n}\sum_{i=1}^{n} u_i$, and a decreasing function of the inequality measure $I(\boldsymbol{u})$. It is often formulated as: $W(\boldsymbol{u}) = \bar{\boldsymbol{u}}(1 - I(\boldsymbol{u}))$. Conversely, it is possible to define an inequality index from a welfare function W as follows:

$$I(\boldsymbol{u}) = \begin{cases} 1 - \frac{W(\boldsymbol{u})}{\bar{\boldsymbol{u}}} & \text{if } \boldsymbol{u} \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The inequality index $I(\boldsymbol{u})$ thus represents the proportion of loss of welfare due to inequality. The concavity of the welfare function represents the degree of aversion to inequality of the evaluator. As for the choice of welfare function, choosing an index to measure inequality involves a normative judgement, since various choices can produce different conclusions.

A fundamental difference between inequality indices and their associated welfare function is the absence of normalization by the mean of utilities. As elaborated by Atkinson [1970] and discussed in the context of recommender systems in Chapter 3, inequality indices only focus on the shape of a utility profile, not on its mean. This implies that it is possible to decrease the value of an inequality index by making everyone worse-off, which is undesirable. The absence of normalization in the welfare function prevents this degenerate behavior. This is why inequality indices are used for evaluation rather than decision-making, while decisions are made by maximizing welfare functions. We follow this practice in Chapters 3 an 4.

2.4.2.2 Generalized Gini indices

The Gini index [Gini, 1921] is a well-known inequality index used in cardinal welfare economics. It is often calculated as [Yitzhaki and Schechtman, 2013]:

$$\text{Gini}(\boldsymbol{u}) = \frac{1}{n^2\bar{\boldsymbol{u}}} \sum_{i=1}^{n}\sum_{j=1}^{n} |u_i - u_j| \qquad \text{with } \bar{\boldsymbol{u}} = \frac{1}{n}\sum_{i=1}^{n} u_i.$$

Note that in addition to the Gini index being a well-known measure of inequality, the sum of absolute pairwise differences previously appeared routinely in papers on fairness of exposure [Morik et al., 2020] as measures of "unfairness", even though these papers do not explicitly mention the relationship with the Gini index.

The (Generalized) Lorenz curve The Gini index has many definitions. One of the most commonly used is based on the Lorenz curve, which plots cumulative fractions of utility owned by individuals ordered from those with less utility (the worse-off) to those with highest utility (the better-off). Formally, let $\boldsymbol{u}^{\uparrow}$ be the values of $\boldsymbol{u}$ sorted in increasing order, i.e., $u_1^{\uparrow} \leq ... \leq u_n^{\uparrow}$ and let $\boldsymbol{U} \in \mathbb{R}^n$ be the cumulative sum of $\boldsymbol{u}^{\uparrow}$, i.e., $U_i = u_1^{\uparrow} + ... + u_i^{\uparrow}$. The *Lorenz curve* of $\boldsymbol{u}$ is $i/n \mapsto \frac{U_i}{\|\boldsymbol{u}\|_1}$ (note that $\|\boldsymbol{u}\|_1 = U_n$, so the end point of the curve is 1). Then the Gini index is equal to 1-2*A where A is the area under the Lorenz curve:

$$\text{Gini}(\boldsymbol{u}) = 1 - \frac{2}{n\,\|\boldsymbol{u}\|_1} \sum_{i=1}^{n} U_i.$$

An example of Lorenz curve is given Fig. 2.1 (left). It provides a representation of how utility is distributed across the population. When there is perfect equality of utility, the Lorenz curve is a straight line from $(0, 0)$ to $(1, 1)$. On the other hand, the stronger the curvature, the more the utility is concentrated on the better-off individuals.

Generalized Lorenz curves are, which are at the core of Chapters 3 and 4, are Lorenz curves without the normalization by mean utility, i.e. the curve $i/n \mapsto U_i$ [Shorrocks, 1983] (Figure 2.1 right). Unlike Lorenz curves, generalized Lorenz curves uniquely characterize the distribution of utility in the population [Shorrocks, 1983], by taking into account the actual amount of utility possessed by each fraction of the population. Importantly, they make it possible to visualize which fractions of the population, ordered from worse-off to better-off, benefit the most from an allocation.

In Chapter 3, we propose to diagnose the fairness of rankings by looking at the generalized Lorenz curves of users and items to visualize *"who gets what"*. In particular, it allows to visualize how strongly our ranking methods redistribute utility from better-off to worst-off users or items, and to show that some existing ranking methods reduce the utility of the worst-off.

The welfare function of the Gini index The welfare function associated to the Gini index is the un-normalized value $\frac{1}{n}\sum_{i=1}^{n} U_i$. It can be written as the area under the generalized Lorenz curve, or equivalently as an ordered weighted average (OWA, [Yager, 1988]):

$$W_{\text{Gini}}(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} U_i = \sum_{i=1}^{n} \frac{n-i+1}{n} u_i^{\uparrow}. \tag{2.4}$$

This formula clarifies that the utility of the worse-off ($u_i^{\uparrow}$ for small i) accounts for more than the utility of the better-off. Note that the right-hand side of the formula above is called an ordered weighted average because the weight associated to a coordinate in $\boldsymbol{u}$ depends on its rank after

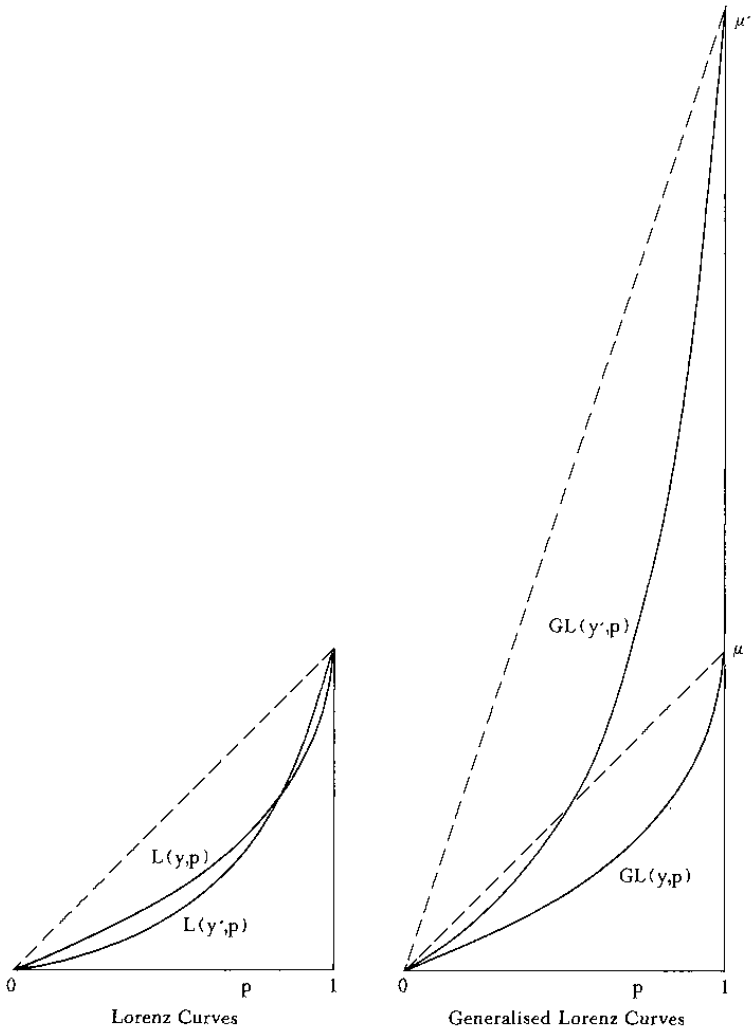


Figure 2.1: (Left) Lorenz curves and (Right) Generalized Lorenz curves of two income profiles $\mathbf{y}$ and $\mathbf{y}'$, from the original paper of [Shorrocks, 1983]. Because of Lorenz curves are normalized by mean income, they do not favour an income profile over the other since the two intersect. In contrast, Generalized Lorenz curves show that $\mathbf{y}'$ is preferrable to $\mathbf{y}$ because the cumulative income is higher for all fractions of the population, from the worst-off individuals to the whole population. In other words, the income profile $\mathbf{y}'$ Lorenz-dominates $\mathbf{y}$.

sorting in increasing order.

Generalized Gini welfare functions Welfare functions and inequality measures alike define normative judgements on how *less inequality* is desired. These values are often described in terms of redistribution or transfers from better-off to worse-off individuals in the population. The welfare function of the classical Gini index (2.4) gives equal weight to the cumulative share of each fraction of the population, in the sense that each point of the generalized Lorenz curve U_i has equal weight (2.4) (which corresponds to giving more weight to worse-off individuals of the population, where the weight is a linear function of the rank, as shown by the weight $\frac{n-i+1}{n}$ assigned to $u_i^{\uparrow}$).

Since generalized Lorenz curves uniquely characterizes the distribution of utility in the population [Shorrocks, 1983], a natural way to formulate normative judgements in terms of redistribution is by assigning weights to each point of the Lorenz curve. This family of welfare function is called Generalized Gini welfare Functions (GGFs, [Weymark, 1981]), which can be written as an OWA of utilities [Yager, 1988]. Given a vector $\boldsymbol{w}$ of non-increasing positive weights, s.t.

$w_1 = 1 \geq \ldots \geq w_n \geq 0$, and using the convention $w_{n+1} = 0$, the GGFs are defined as:

$$g_{\boldsymbol{w}}(\boldsymbol{u}) = \sum_{i=1}^{n}(w_i - w_{i+1})U_i = \sum_{i=1}^{n} w_i u_i^{\uparrow}.$$

The fact that the weights $\boldsymbol{w}$ are non-increasing imply that $w_i - w_{i+1} \geq 0$, and also that W_{Gini} is concave [Yager, 1988]. It is clear also that $g_{\boldsymbol{w}}(\boldsymbol{u})$ is increasing with respect to each coordinate in $\boldsymbol{u}$. This guarantees that maximizing W_{Gini} generates solutions that are Pareto-efficient and that W_{Gini} is monotonic with respect to the dominance of generalized Lorenz curves.

Inequality indices as special cases of Generalized Gini The Bonferroni and De Vergottini indices are two classical inequality indices, for which the associated welfare functions are special instantiations of Generalized Gini welfare Functions [Aristondo et al., 2013]. The Bonferroni index [Bonferroni, 1941] compares the overall income mean to the income means of the poorest individuals in the population to assess inequality. The De Vergottini index [De Vergottini, 1950] supplements the information given by the Bonferroni index by comparing the overall income mean to the income means of the population's wealthiest individuals. Unlike the classical Gini index, the Bonferroni and De Vergottini indices are sensitive to the precise location of utility transfers within the ordered utility profile. The Bonferroni welfare functions is a GGF with weights $w_i = \sum_{j=n-i+1}^{n} \frac{1}{jn}$, while the De Vergottini welfare functions is a GGF with weights $w_i = \sum_{j=i}^{n} \frac{1}{jn}$ [Aristondo et al., 2013].

Top wealth shares [Piketty and Saez, 2003] and quantile ratios [Burkhauser et al., 2009, Neves Costa et al., 2019] are other widely used examples of inequality measures based quantiles of utilities. A commonly used quantile ratio is the D9/D1 index which is the ratio of the 90th and 10th percentile values, and allows to compare the incomes of the wealthiest and poorest individuals of the population. GGFs also allow to express normative criteria based on utility quantiles, using e.g. $w_i = 1$ if $i \leq \lfloor qn \rfloor$ for the bottom q-th quantile.

2.4.2.3 Additively decomposable inequality indices

While the family of Generalized Gini indices provides an expressive framework for measuring inequality, there are several alternative inequality measures with interesting properties. A class of indices of interest are *generalized entropy indices* [Shorrocks, 1980], which are defined as:

$$E_\alpha(\boldsymbol{u}) = \begin{cases} \frac{1}{n\alpha(\alpha-1)} \sum_{i=1}^{n} \left[\left(\frac{u_i}{\bar{\boldsymbol{u}}}\right)^{\alpha} - 1\right], & \text{if } \alpha \neq 0, 1 \\ -\frac{1}{n} \sum_{i=1}^{n} \ln(\frac{u_i}{\bar{\boldsymbol{u}}}), & \text{if } \alpha = 0 \\ \frac{1}{n} \sum_{i=1}^{n} \frac{u_i}{\bar{\boldsymbol{u}}} \ln(\frac{u_i}{\bar{\boldsymbol{u}}}), & \text{if } \alpha = 1. \end{cases}$$

Several well-known inequality indices belong to the class of generalized entropy indices from [Shorrocks, 1980]. For example, $E_1(\boldsymbol{u})$ is the Theil index, and $E_2(\boldsymbol{u})$ is half the squared coefficient of variation. Note that the coefficient of variation is a normalized standard deviation, which is what we use to compute utility/inequality trade-offs in Chapter 3. Generalized Entropy indices are also related to the Atkinson index of [Atkinson, 1970], which is defined from a welfare function.

The appealing property of generalized entropy indices is that they are *additively decomposable*, meaning that they can be decomposed into a within-group and a between-group inequality term. In contrast, generalized Gini indices do not satisfy additive decomposability in general [Shorrocks, 1980]. Existing work on group fairness in ranking and recommendation define group utilities by averaging the utilities in the group, and measure unfairness by the Gini index [Morik et al., 2020]. The resulting measure does not account for within-group inequalities, while using an additively

decomposable inequality measure on the utility profile would provide a decomposition into within-group and between-group fairness. Speicher et al. [2018] advocate for this property in the context of fair machine learning. We propose a simple treatment of group fairness in Appendix A.2 that we discuss in more detail in Chapter 7, but we do not address the question of within-group vs. between-group fairness.

2.4.3 Envy-free allocations

The concepts from cardinal welfare economics that we presented in the previous section – i.e. welfare functions, inequality indices, Pareto and Lorenz dominance – are applicable to various social choice problems, including the allocation of private goods (e.g., a piece of cake) or public goods (e.g., a public road) [Le Breton and Weymark, 2011]. In this section, we discuss fairness criteria that are specific to allocation problems for private goods. As we discuss in Chapter 6, these criteria can be used to conceptualize user fairness in personalized recommender systems, where the personalized ranking assigned to a user can be seen as a private good.

Envy-freeness, which is usually credited to [Foley, 1967], is a desirable property in the fair division of private goods, in which a resource or a set of items must be divided among multiple agents. An allocation is said to be *envy-free* if no agent prefers the share of resource or bundle of items of another agent to their own, i.e., there is no envy. An extensive discussion of the axiomatic foundations of envy-freeness is found in [Thomson, 2011].

Another fairness criterion in fair division is proportionality, which is satisfied if each agent receives a share that they value at least as much as $1/n$ of the total resource's value to them. In the classical setting of additive utilities, envy-freeness implies proportionality [Thomson, 2011].

Unlike the welfare function approach described previously, envy-freeness can be defined in terms of ordinal preferences. Furthermore, it does not involve interpersonal comparison of utilities across agents, since different bundles are assessed by the preferences of the same agent. The latter property is interesting for recommender systems, since user utilities can be difficult to compare as they are based on patterns that differ across users (e.g., rating or browsing habits).

In Chapter 6, we propose envy-freeness as a fairness criterion for personalized recommendation, and analyse its properties and its relationship to other criteria for fair recommendation. The only agents that we consider in that chapter are the users.

The personalized recommendation setting is different from classical fair division in several ways. First, in recommender systems, the same item can be shown to an unrestricted number of users, whereas in fair division, a single good can be given to at most one agent. Second, the true user preferences in recommender systems are unknown and must be estimated from noisy feedback, while in fair division problems, the agents' preferences are known to the decision-maker. We address the technical challenge of exploring user preferences to certify envy-freeness in Chapter 6.

2.5 Social choice and welfare for fair machine learning

Recently, there has been growing interest in building connections between fairness in machine learning and social choice theory [Heidari et al., 2018, Ustun et al., 2019, Balcan et al., 2018, Gölz et al., 2019, Hossain et al., 2020, Chakraborty et al., 2019, Finocchiaro et al., 2021, Saito and Joachims, 2022], and welfare economics in particular [Speicher et al., 2018, Hu and Chen, 2020, Kleinberg et al., 2018b, Zimmer et al., 2021, Hossain et al., 2021]. In line with Hu and Chen [2020], who focused on classification tasks and parity constraints, we argue that the principle of Pareto efficiency should be part of fairness assessments. In Chapter 3, we are the first to propose

concave welfare functions and Lorenz efficiency to address two-sided fairness in recommendation. In particular, by introducing Lorenz efficiency, which combines Pareto efficiency and the Pigou-Dalton principle, we provide a refinement of Pareto efficiency which helps choosing ranking policies among Pareto-efficient solutions with a more complete assessment of "who gets what".

Among the recently proposed connections between fair machine learning and economic concepts, some authors proposed to use inequality indices to quantify and mitigate unfairness [Speicher et al., 2018, Heidari et al., 2018, Lazovich et al., 2022], take an axiomatic perspective [Gölz et al., 2019, Cousins, 2021, Williamson and Menon, 2019] or apply welfare economics principles [Hu and Chen, 2020, Rambachan et al., 2020]. In particular, the Generalized Gini welfare Functions (GGFs) that we study in Chapter 4, were recently applied to fair multi-agent reinforcement learning, with multiple reward functions [Busa-Fekete et al., 2017, Siddique et al., 2020, Zimmer et al., 2021]. These works consider sequential decision-making problems without ranking, and their GGFs aggregate the objectives of a few agents (typically $n < 20$), while in our ranking problem, there are as many objectives as there are users and items.

The integration of social choice concepts into fair machine learning is also useful for defining preference-based fairness criteria. Specifically, since social choice deals with fair decision-making based on the heterogeneous preferences of agents, its concepts are particularly suitable to personalized recommendation systems that aim to account for user preferences. In Chapter 6, we propose the social choice criterion of envy-freeness as a preference-based fairness criterion for personalized recommendation, which ensures that no user would prefer the recommendation received by another user. Envy-freeness was also studied as a user-side fairness criterion in Patro et al. [2020], but without addressing the challenge of measuring envy under noisy feedback. Saito and Joachims [2022] recently used it for item-side fairness in a model that considers the utilities of items beyond a mere preference for high exposure.

Preference-based fairness criteria have been discussed in several other aspects by the machine learning community. The framework of envy-free classification [Balcan et al., 2018] focuses on classification problems with a known auxiliary utility function of predictions. The recent work on preference-informed individual fairness [Kim et al., 2019] combines the notions of distance-based individual fairness of Dwork et al. [2012] and envy-freeness, but requires access to both user preferences and a measure of similarity between individuals. In Chapter 6, we address a personalized recommendation setting, where the preferences of users are unknown and must be estimated by the auditor to estimate envy from noisy feedback.

Notice that while the original frameworks of preference-based fairness for classification are defined at the level of groups [Zafar et al., 2017b, Ustun et al., 2019, Hossain et al., 2020, Suriyakumar et al., 2022], it is more challenging to define group envy-freeness when the recommendations are personalized. This is because in the classification setting, there is only one classifier per group, while in our case we have a recommendation policy per individual in the group. In our personalized recommendation setting, we would need a non-trivial definition to capture what it means for a group of users to be "envious of the recommendations of another group", since there is no single group-level recommendation.

Chapter 3

Fairness in rankings with additive concave welfare functions

Contents

This chapter is the article *Two-sided fairness in rankings via Lorenz dominance*, published at NeurIPS 2021 (see [Do et al., 2021c]). In this chapter, we approach fair recommendation as a fair division problem where the scarce resource is the total exposure and the agents are the users and items. We propose a conceptual framework for fairness in ranked recommendations grounded in cardinal welfarism in social choice. This chapter provides a basis for the design of fair ranking objectives, on which we rely in the next chapters that will focus on algorithmic challenges and online learning.

We introduce generalized Lorenz curves to the fair ranking problem, which are graphical representations of the distribution of utility among users and items. We formalize the criterion of *Lorenz efficiency* for fairness in rankings, which is satisfied by rankings with non-dominated generalized Lorenz curves. It ensures that rankings are Pareto-efficient and that they are maximally redistributive at a given level of overall utility (i.e., they follow the Pigou-Dalton transfer principle). This framework provides a better understanding of existing criteria for fairness in rankings, and shows that existing approaches amplify rich-gets-richer effects or destroy utility instead of redistributing it, in violation of Lorenz efficiency. We propose a principled approach to generate fair rankings by maximizing additive concave welfare functions of the utility profiles of users and items. The

curvature of the welfare function for users (resp. items) controls the degree of redistribution among users (resp. items).

In this chapter, we consider a batch setting, and focus on the ranking problem. We do not address the problem of learning the values μ_{ij} and assume they are given as input to the recommender system. Nonetheless, we provide in Appendix A.3.3 an excess risk bound on the true welfare of the ranking obtained when using estimates $\hat{\mu}_{ij}$.

In the batch setting, ranking with item-side fairness is challenging because items' utilities depend on the rankings of *all users*, requiring global inference. Previous methods that preceded the publication of this work addressed this issue with heuristic methods without guarantees or control on the achievable trade-offs. We show how the Frank-Wolfe algorithm can be leveraged for tractable fair ranking in the position-based model.

Our method can be applied to both one-sided and reciprocal recommendation tasks, such as music or movie recommendation, and dating or social recommendation, respectively. By proposing the first unified framework for these two settings, we provide a new opportunity to investigate the fairness of rankings in reciprocal recommender systems, an area that has received relatively little attention in prior research.

The modelling choices made in this chapter are further discussed in Chapter 7.

Note that this chapter uses the notation of the original publication, which is different than the notation of Chapter 1. This is because at the time of writing the article, we aimed at minimal changes between one-sided and reciprocal recommendation settings. We also use the term "quality-weighted exposure" instead of "merit-based exposure" to refer to a prior item-side fairness criterion.

Abstract

We consider the problem of generating rankings that are fair towards both users and item producers in recommender systems. We address both usual recommendation (e.g., of music or movies) and reciprocal recommendation (e.g., dating). Following concepts of distributive justice in welfare economics, our notion of fairness aims at increasing the utility of the worse-off individuals, which we formalize using the criterion of *Lorenz efficiency*. It guarantees that rankings are Pareto-efficient, and that they maximally redistribute utility from better-off to worse-off, at a given level of overall utility. We propose to generate rankings by maximizing concave welfare functions, and develop an efficient inference procedure based on the Frank-Wolfe algorithm. We prove that unlike existing approaches based on fairness constraints, our approach always produces fair rankings. Our experiments also show that it increases the utility of the worse-off at lower costs in terms of overall utility.

3.1 Introduction

Recommender systems have a growing impact on the information we see and on our life opportunities, as they help us browse news articles, find a new job, house, or people to connect with. While the objective of recommender systems is usually defined as maximizing the quality of recommendations from the user's perspective, the recommendations also have an impact on the recommended "items". News outlets rely on exposure to generate revenue, finding a job depends on which recruiter gets to see our resume, and the effectiveness of a dating application also depends on who we are recommended to—and if we are being recommended, then someone else is not. *Two-sided fairness in rankings* is the problem of generating personalized recommendations by fairly mediating between the interests of users and items. It involves a complex multidimensional trade-off. Fairness towards

item producers requires boosting the exposure of small producers (e.g., to avoid winner-take-all effects and popularity biases [Abdollahpouri et al., 2019b]) at the expense of average user utility. Fairness towards users aims at increasing the utility of the least served users (e.g., so that least served users do not support the cost of item-side fairness), once again at the expense of average user utility. The goal of this paper is to provide an algorithmic framework to generate rankings that achieve a variety of these trade-offs, leaving the choice of a specific trade-off to the practitioner.

The leading approach to fairness in rankings is to maximize user utility under constraints of equal item exposure (or equal quality-weighted exposure) [Singh and Joachims, 2018, Biega et al., 2018] or equal user satisfaction [Basu et al., 2020]. When these constraints imply an unacceptable decrease in average user utility, so-called "trade-offs between utility and fairness" [Zehlike and Castillo, 2020, Singh and Joachims, 2019] are obtained by relaxing the fairness constraints, leading to the optimization of a trade-off between average user utility and a measure of users' or items' inequality.

Thinking about fairness in terms of optimal utility/inequality trade-offs has, however, two fundamental limitations. First, the optimization of a utility/inequality trade-off is not necessarily Pareto-efficient from the point of view of users and items: it sometimes chooses solutions that decrease the utility of some individuals without making anybody else better off. We argue that reducing inequalities by decreasing the utility of the better-off is not desirable if it does not benefit anyone. The second limitation is that focusing on a single measure of inequality does not address the question of how inequality is reduced, and in particular, which fraction of the population benefits or bears the cost of reducing inequalities.

In this paper, we propose a new framework for two-sided fairness in rankings grounded in the analysis of generalized Lorenz curves of user and item utilities. Widely used to study efficiency and equity in cardinal welfare economics [Shorrocks, 1983], these curves plot the cumulative utility obtained by fractions of the population ordered from the worst-off to the best-off. A curve that is always above another means that all fractions of the populations are better off. We define fair rankings as those with non-dominated generalized Lorenz curves for users and items. First, this definition guarantees that fair rankings are Pareto-efficient. Second, examining the entirety of the generalized Lorenz curves provides a better understanding of which fractions of the population benefit from an intervention, and which ones have to pay for it. We present our general framework based on Lorenz dominance in usual recommendation settings (e.g., music or movie recommendation), and also show how extend it to *reciprocal recommendation* tasks such as dating applications or friends recommendation, where users are recommended to other users.

We present a new method for generating rankings based on the maximization of concave welfare functions of users' and items' utilities. The parameters of the welfare function control the relative weight of users and items, and how much focus is given to the worse-off fractions of users and items. We show that rankings generated by maximizing our welfare functions are fair for every value of the parameters. Our framework does not aim at defining what parameters are suitable in general — rather, the choice of a specific trade-off depends on the application.

From an algorithmic perspective, two-sided fairness is challenging because items' utilities depend on the rankings of all users, requiring global inference. Previous work on item-side fairness addressed this issue with heuristic methods without guarantees or control on the achievable trade-offs. We show how the Frank-Wolfe algorithm can be leveraged to make inference tractable, addressing both our welfare maximization approach and existing item-side fairness penalties.

We demonstrate that our welfare function approach enjoys stronger theoretical guarantees than existing methods. While it always generates rankings with non-dominated generalized Lorenz curves, many other approaches do not. We show that one of the main criteria of the literature,

called equity of attention by Biega et al. [2018], can lead to decrease user utility, while *increasing* inequalities of exposure between items. Moreover, equal user satisfaction criteria in reciprocal recommendation can lead to decrease the utility of *every user*, even the worse-off. Our notion of fairness prevents these undesirable behaviors. We report experimental results on music and friend recommendation tasks, where we analyze the trade-offs obtained by different methods by looking at different points of their Lorenz curves. Our welfare approach generates a wide variety of trade-offs, and is, in particular, more effective at improving the utility of worse-off users than the baselines.

We present our formal framework in Section 3.2. We discuss the theoretical properties of previous approaches in Section 3.3, and present our ranking algorithm in Section 3.4. Our experiments are described in Section 3.5, and the related work is discussed in Section 3.6.

3.2 Two-sided fairness via Lorenz dominance

3.2.1 Formal framework

Terminology and notation. We identify an item with its producer, so that "item utility" means "item producer's utility". The main paper focuses on fairness towards individual users and items. We describe in Appendix A.2 the extension of our approach to sensitive groups of users or items. $|\mathcal{X}|$ denotes the cardinal of the set $\mathcal{X}$. Given $n \in \mathbb{N}$, we denote by $[\![n]\!] = \{1, \dots, n\}$. The set of users $\mathcal{N}$ is identified with $\{1, ..., |\mathcal{N}|\}$ and the set of items $\mathcal{I}$ is identified with $\{|\mathcal{N}|+1, ..., n\}$ where $n = |\mathcal{N}| + |\mathcal{I}|$. For $(i, j) \in \mathcal{N} \times \mathcal{I}$, we denote by μ_{ij} the value of item j to user i.

A (deterministic) ranking $\sigma : \mathcal{I} \to [\![|\mathcal{I}|]\!]$ is a one-to-one mapping from items j to their rank $\sigma(j)$. Following [Singh and Joachims, 2018], we use *stochastic rankings* because they allow us to perform inference using convex optimization (see Section 3.4). The recommender system produces one stochastic ranking per user, represented by a 3-way *ranking tensor* P where P_{ijk} is the probability that j is recommended to i at rank k. We denote by $\mathcal{P}$ the set of ranking tensors.

Utilities of users and items are defined through a position-based model, as in previous work [Singh and Joachims, 2018, Biega et al., 2018, Wu et al., 2021b]. Let $v \in \mathbb{R}^{|\mathcal{I}|}$, where v_k is the exposure weight at rank k. We assume that lower ranks receive more exposure, so that $\forall k \in [\![|\mathcal{I}| - 1]\!], v_k \geq v_{k+1} \geq 0$.[1] Given a user i and a ranking σ_i, the *user-side utility* of i is the sum of the μ_{ij}s weighted by the exposure weight of their rank $\sigma_i(j)$: $u_i(\sigma_i) = \sum_{j \in \mathcal{I}} v_{\sigma_i(j)} \mu_{ij}$. Given an item j, the *item-side utility* of j is the sum over users i of the exposure of j to i. These definitions extend to stochastic rankings by taking the expectation over rankings, written in matrix form:[2]

$$\textit{user-side utility:} \quad u_i(P) = \sum_{j \in \mathcal{I}} \mu_{ij} P_{ij} v \qquad \textit{item-side utility (exposure):} \quad u_j(P) = \sum_{i \in \mathcal{N}} P_{ij} v$$

We denote by $\boldsymbol{u}(P) = (u_i(P))_{i=1}^n$ the utility profile for P, and by $\mathcal{U} = \{\boldsymbol{u}(P) : P \in \mathcal{P}\}$ the set of feasible profiles. For $\boldsymbol{u} \in \mathcal{U}$, $\boldsymbol{u}_\mathcal{N} = (u_i)_{i \in \mathcal{N}}$ and $\boldsymbol{u}_\mathcal{I} = (u_i)_{i \in \mathcal{I}}$ denote the utility profiles of users and items respectively.

Two-sided fairness in rankings. In practice, values of μ_{ij} are not known to the recommender system. Ranking algorithms use an estimate $\hat{\mu}$ of μ based on historical data. We address here the problem of *inference*: the task is to compute the ranking tensor given $\hat{\mu}$, with the goal of making fair trade-offs between (true) user and item utilities. Notice that the user-side utility depends only on the ranking of the user, but for every item, the exposure depends on the rankings of *all* users.

[1] We use a user-independent v for simplicity. Considering user-dependent weights is straightforward.

[2] We consider P_{ij} as a row vector in the formula, so that $P_{ij} v = \sum_{k=1}^{|\mathcal{I}|} P_{ijk} v_k$.

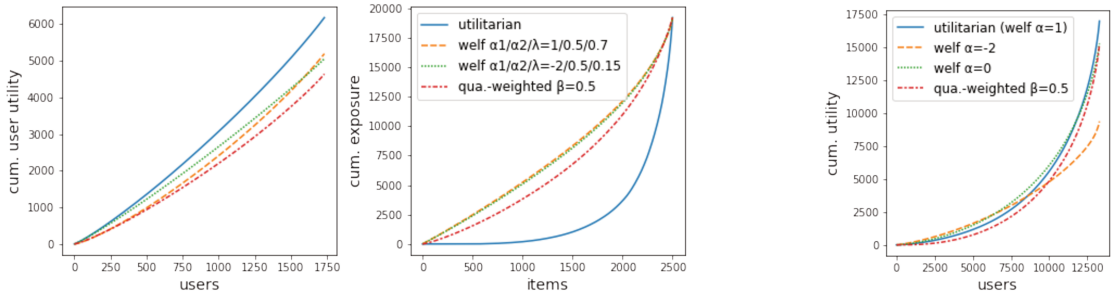


Figure 3.1: Generalized Lorenz curves for usual (left) and reciprocal (right) recommendation.

Thus, accounting for both users' and items' utilities in the recommendations is a global inference problem.

More general item utilities We consider exposure as the item-side utility to follow prior work and for simplicity. Our framework and algorithm readily applies in a more general case of *two-sided preferences*, where items also have preferences over users (for instance, in hiring, job seekers have preferences over which recruiters they are recommended to). Denoting μ_{ji} the value of user i to item j, the item side-utility is then $u_j(P) = \sum_{i \in \mathcal{N}} \mu_{ji} P_{ij} v$.

3.2.2 Lorenz efficiency and the welfare function approach

Our notion of fairness aims at improving the utility of the worse-off users and items. Since this does not prescribe exactly which fraction of the worse-off users/items should be prioritized, the assessment of trade-offs requires looking at all fractions of the population. This is captured by the generalized Lorenz curve used in cardinal welfare economics [Shorrocks, 1983]. Formally, given a utility profile $\boldsymbol{u}$, let $(u_{(i)})_{i=1}^n$ be the sorted values in $\boldsymbol{u}$ from smallest to largest, i.e., $u_{(1)} \leq \ldots \leq u_{(n)}$, then the generalized Lorenz curve plots $(U_i)_{i=1}^n$ where $U_i = u_{(1)} + \ldots + u_{(i)}$. To assess the fairness of trade-offs, we rely on the following dominance relations on utility profiles:

Pareto-dominance $\succ_{\mathrm{P}}$. $\boldsymbol{u} \succ_{\mathrm{P}} \boldsymbol{u'} \iff (\forall i \in [\![n]\!], u_i \geq u'_i$ and $\exists i \in [\![n]\!], u_i > u'_i)$.

Lorenz-dominance $\succ_{\mathrm{L}}$. Then $\boldsymbol{u} \succ_{\mathrm{L}} \boldsymbol{u'} \iff \boldsymbol{U} \succ_{\mathrm{P}} \boldsymbol{U'}$.

We write $\succeq_{\mathrm{L}}$ for non-strict Lorenz dominance (i.e., $\forall i, U_i \geq U'_i$). Notice that Pareto-dominance implies Lorenz-dominance. Our notion of fairness, which we call *Lorenz efficiency*, states that a ranking is fair if the utility profiles for users and for items are not jointly Lorenz-dominated:

Definition 1 (Lorenz efficiency)**.** *A utility profile $\boldsymbol{u} \in \mathcal{U}$ is* Lorenz-efficient *if there is no $\boldsymbol{u'} \in \mathcal{U}$ such that either ($\boldsymbol{u'}_{\mathcal{I}} \succeq_{\mathrm{L}} \boldsymbol{u}_{\mathcal{I}}$ and $\boldsymbol{u'}_{\mathcal{N}} \succ_{\mathrm{L}} \boldsymbol{u}_{\mathcal{N}}$) or ($\boldsymbol{u'}_{\mathcal{N}} \succeq_{\mathrm{L}} \boldsymbol{u}_{\mathcal{N}}$ and $\boldsymbol{u'}_{\mathcal{I}} \succ_{\mathrm{L}} \boldsymbol{u}_{\mathcal{I}}$).*

We consider that Lorenz-dominated profiles are undesirable (and unfair) because the utility of worse-off fractions of the population could have been increased at no cost for total utility. Examples of Lorenz-curves of users and items are given in Fig. 3.1. The blue solid, green dotted and orange dashed curves are all non-dominated (the blue solid ranking has higher user utility but high item inequality, the green dotted and orange dashed curves have similar item exposure profiles, but user curves that intersect). On the other hand, the red dot/dashed curve is an unfair ranking: compared to the green dotted and orange dashed curve, all fractions of the worse off users have lower utility, together with less exposure for worse-off items.

A fundamental result from cardinal welfare economics is that concave welfare functions of utility profiles order profiles according to Lorenz dominance [Atkinson, 1970, Shorrocks, 1983]. The choice of the welfare function specifies which (fair) trade-off is desirable in a specific context. This result

holds when all utilities are comparable. In our case where there are users and items, we propose the following welfare function parameterized by $\theta = (\lambda, \alpha_1, \alpha_2)$:[3]

$$\forall \boldsymbol{u} \in \mathbb{R}_+^n : \quad W_\theta(\boldsymbol{u}) = (1-\lambda) \sum_{i \in \mathcal{N}} \psi(u_i, \alpha_1) + \lambda \sum_{j \in \mathcal{I}} \psi(u_j, \alpha_2) \quad \text{with } \psi(x, \alpha) = \begin{cases} x^\alpha & \text{if } \alpha > 0 \\ \log(x) & \text{if } \alpha = 0 \\ -x^\alpha & \text{if } \alpha < 0 \end{cases}.$$

Inference is carried out by maximizing W_θ (an efficient algorithm is proposed in Section 3.4):

$$\textit{(ranking procedure)} \qquad P^* \in \operatorname*{argmax}_{P \in \mathcal{P}} W_\theta(\boldsymbol{u}(P)) \tag{3.1}$$

In W_θ, $\lambda \in [0, 1]$ controls the relative weight of users and items. The motivation for the specific choice of ψ is that it appears in scale invariant welfare functions [Moulin, 2003], but other families can be used as long as the functions are *increasing* and *concave*. Monotonicity implies that maxima of W_θ are Pareto-efficient. For $\alpha_1 < 1$ and $\alpha_2 < 1$, W_θ is strictly concave. Then, W_θ exhibits *diminishing returns*, which is the key to Lorenz efficiency: an increment in utility for a worse-off user/item increases welfare more than the same increment for a better-off user/item. The effect of the parameters is shown in Fig. 3.1 (left): For *item fairness* we obtain more item equality by using $\alpha_1 < 1$ (here, $\alpha_1 = 0.5$) and incrasing λ (see blue solid vs orange dashed curve). The parameter α_2 controls *user fairness*: smaller values yield more user utility for the worse-off users at the expense of total utility, with similar item exposure curve (green dotted vs orange dahsed curves). Let $\Theta = \{(\lambda, \alpha_1, \alpha_2) \in (0, 1) \times (-\infty, 1)^2\}$. For every $\theta \in \Theta, W_\theta$ is strictly concave, and users and items have non-zero weight. We then have (the result is a straightforward consequence of diminishing returns, see Appendix A.3):

Proposition 1. $\forall \theta \in \Theta, \forall P^* \in \operatorname*{argmax}_{P \in \mathcal{P}} W_\theta(\boldsymbol{u}(P))$, P^* *is Lorenz-efficient.*

Relationship to inequality measures A well-known measure of inequality is the Gini index, defined as $1 - 2 \times \text{AULC}$, where AULC is the area under the Lorenz curve. The difference between Lorenz and generalized Lorenz curves is that the former is normalized by the cumulative utility. This difference is fundamental: we can decrease inequalities while dragging everyone's utility to 0. However, this would lead to dominated *generalized* Lorenz curves. Interestingly, for *item-side* fairness, the cumulative exposure is a constant and thus trade-offs between user utility and item exposure inequality are not really problematic. However, for user-side fairness, the total utility is not constant and reducing inequalities might require dragging the utility of some users down for the benefit of no one.

Additional theoretical results In App. A.3.2, we show that as $\alpha_1, \alpha_2 \to -\infty$, utility profiles tend to leximin-optimal solutions [Moulin, 2003]. Leximin optimality corresponds to increasing the utility of the worst-off users/items one a a time, similarly to a lexical order. In App. A.3.3, we present an excess risk bound, which provides theoretical guarantees on the *true* welfare when computing rankings based on *estimated* preferences, depending on the quality of the estimates.

[3] $W_\theta(\boldsymbol{u}) = -\infty$ if $\alpha \leq 0$ and $\exists i, u_i = 0$. In practice, we use $\psi(x + \eta, \alpha)$ for $\eta > 0$ to avoid this case.

3.2.3 Extension to reciprocal recommendation

In reciprocal recommendation problems such as dating, the users are also items. The notion of fairness simplifies to increasing the utility of the worse-off users, which can in practice be done by boosting the exposure of worse-off users. Our framework above applies readily by taking $\mathcal{N} = \mathcal{I}$ and $n = |\mathcal{N}|$. The critical step however is to redefine the utility of a user to account for the fact that (1) the user utility comes from both the recommendation they receive and who they are recommended to, and (2) users have preferences over who they are recommended to.

To define this *two-sided utility*, let us denote by μ_{ij} the mutual preference value between i and j, and our examples follow the common assumption that $\mu_{ij} = \mu_{ji}$ (see e.g., Palomares et al. [2021]). For instance, when recommending CVs to recruiters, μ_{ij} can be the probability of interview, while in dating, it can be that of a "match". The two-sided utility is then the sum of the user-side utility and item-sided utility of the user:

$$\overbrace{\overline{u}_i(P) = \sum_{j \in \mathcal{I}} \mu_{ij} P_{ij} v}^{\substack{\text{user-side utility} \\ \textit{(j recommended to i)}}} \qquad \overbrace{\overline{v}_i(P) = \sum_{j \in \mathcal{N}} \mu_{ij} P_{ji} v}^{\substack{\text{item-side utility} \\ \textit{(i recommended to j)}}} \qquad \overbrace{u_i(P) = \overline{u}_i(P) + \overline{v}_i(P)}^{\text{(two-sided) utility}}$$

With this definition of two-sided utility, our previous framework can be readily applied using $\mathcal{N} = \mathcal{I}$. A (two-sided) utility profile $\boldsymbol{u} \in \mathcal{U}$ is *Lorenz-efficient* if there is no $\boldsymbol{u'} \in \mathcal{U}$ such that $\boldsymbol{u'} \succ_{\mathrm{L}} \boldsymbol{u}$. The welfare function simplifies to $W_\theta(\boldsymbol{u}) = \sum_{i=1}^{n} \psi(u_i, \alpha)$, and Proposition 1 also holds true in this setting: maximizing the welfare function always yields Lorenz-efficient rankings.

Fig. 3.1 (right) illustrates how decreasing α increases utilities for the worse-off users at the expense of total utility. It also shows a Lorenz-dominated (unfair) profile, in which all fractions from the worst-off to the better-off users have lower utility.

From now on, we refer to *one-sided* recommendation for non-reciprocal recommendation.

3.3 Comparison to utility/inequality trade-off approaches

As stated in the introduction, leading approaches to fairness in ranking are based on utility/inequality trade-offs. We describe here the representative approaches we consider as baselines in our experiments. We then present theoretical results illustrating the undesirable behavior of some of them.

3.3.1 Objective functions

One-sided recommendation In one-sided recommendation, the leading approach is to define exposure-based criteria for item fairness [Singh and Joachims, 2018, Biega et al., 2018]. The first criterion, *equality of exposure*, aims at equalizing exposure across items. The second one, *quality-weighted exposure*[4], which is advocated by many authors, defines the *quality* of an item as the sum of user values $q_j = \sum_{i \in \mathcal{N}} \mu_{ij}$ and aims for item exposure proportional to quality. The motivation of quality-weighted exposure is to take user utilities into account in the extreme case where the constraint is strictly enforced. Interestingly, as we show later, this approach has bad properties in terms of trading off user and item utilities.

[4]We use here the terminology of [Wu et al., 2021b]. This criterion has also been called "disparate treatment" [Singh and Joachims, 2018], "merit-based fairness" [Singh and Joachims, 2019] and "equity of attention" [Biega et al., 2018].

In our experiments, we use the standard deviation as a measure of inequality. Denoting by $E = |\mathcal{N}| \, \|v\|_1$ the total exposure and by $Q = \sum_{j \in \mathcal{I}} q_j$ the total quality:

$$\textit{quality-weighted exposure} \quad F_\beta^{\mathrm{qua}}(\boldsymbol{u}) = \sum_{i \in \mathcal{N}} u_i - \beta \sqrt{D^{qua}(\boldsymbol{u})} \quad \text{with} \quad D^{qua}(\boldsymbol{u}) = \frac{1}{|\mathcal{I}|} \sum_{j \in \mathcal{I}} \left(u_j - \frac{q_j E}{Q} \right)^2.$$

$$\textit{equality of exposure} \quad F_\beta(\boldsymbol{u}) = \sum_{i \in \mathcal{N}} u_i - \beta \sqrt{D(\boldsymbol{u})} \quad \text{with} \quad D(\boldsymbol{u}) = \frac{1}{|\mathcal{I}|} \sum_{j \in \mathcal{I}} \left(u_j - \frac{1}{|\mathcal{I}|} \sum_{j' \in \mathcal{I}} u_{j'} \right)^2.$$

Some authors use $D'(\boldsymbol{u}) = \sum_{(j,j') \in \mathcal{I}^2} |\frac{u_j}{q_j} - \frac{u_{j'}}{q_{j'}}|$ instead of $\sqrt{D^{qua}}$ [Singh and Joachims, 2019, Morik et al., 2020, Basu et al., 2020]. D^{qua} and D' have qualitatively the same behavior. We propose $D^{qua}(\boldsymbol{u})$ as a computationally efficient alternative to D', since it involves only a linear number of terms and $\sqrt{D^{qua}}$ is convex and differentiable except on 0.

Reciprocal recommendation For reciprocal recommendation, we consider as competing approach a trade-off between total (two-sided) utility and inequality of utilities, as measured by the standard deviation:

$$\textit{equality of utility} \quad F_\beta(\boldsymbol{u}) = \sum_{i \in \mathcal{N}} u_i - \beta \sqrt{D(\boldsymbol{u})} \quad \text{with} \quad D(\boldsymbol{u}) = \frac{1}{n} \sum_{j \in \mathcal{I}} \left(u_j - \frac{1}{n} \sum_{j' \in \mathcal{I}} u_{j'} \right)^2.$$

3.3.2 Inequity and inefficiency of some of the previous approaches

We point out here to two deficiencies of previous approaches.

First, for one-sided recommendation, we show that in some cases, compared to the welfare approach with any choice of the parameter $\theta \in \Theta$, quality-weighted exposure leads to the undesirable behavior of *decreasing user utility* while *increasing inequalities of exposure* between items. This is formalized by the proposition below, which uses the following notation: for $\theta \in \Theta$, let $\boldsymbol{u}^\theta = \operatorname{argmax}_{\boldsymbol{u} \in \mathcal{U}} W_\theta(\boldsymbol{u})$, and for $\beta > 0$, let $\mathcal{U}_\beta^{\mathrm{qua}} = \operatorname{argmax}_{\boldsymbol{u} \in \mathcal{U}} F_\beta^{\mathrm{qua}}(\boldsymbol{u})$.

Proposition 2. *The following claims hold irrespective of the choice of* $\boldsymbol{u}^{\mathrm{qua},\beta} \in \mathcal{U}_\beta^{\mathrm{qua}}$.

For every $d \in \mathbb{N}_*$ *and every* $N \in \mathbb{N}_*$, *there is a one-sided recommendation problem, with* $d+1$ *items and* $N(d+1)$ *users, such that* $\forall \theta \in \Theta$, *we have:*

$$\left(\exists \beta > 0, \boldsymbol{u}_{\mathcal{N}}^\theta \succ_{\mathrm{L}} \boldsymbol{u}_{\mathcal{N}}^{\mathrm{qua},\beta} \text{ and } \boldsymbol{u}_{\mathcal{I}}^\theta \succ_{\mathrm{L}} \boldsymbol{u}_{\mathcal{I}}^{\mathrm{qua},\beta} \right) \qquad \text{and} \qquad \lim_{\beta \to \infty} \frac{\sum_{i \in \mathcal{N}} u_i^{\mathrm{qua},\beta}}{\sum_{i \in \mathcal{N}} u_i^\theta} \xrightarrow[d \to \infty]{} \frac{5}{6}.$$

Second, in reciprocal recommendation, striving for pure equality can even lead to 0 utility for *every user*, even that of the worst-off user. More precisely, we show that in some cases, compared to the welfare approach with any choice of parameter $\theta \in \Theta$, there exists $\beta > 0$ such that equality of utility has lower utility for every user, eventually leading to 0 utility for everyone in the limit $\beta \to \infty$.

Proposition 3. *For* $\beta > 0$, *let* $\mathcal{U}_\beta^{\mathrm{eq}} = \operatorname{argmax}_{\boldsymbol{u} \in \mathcal{U}} F_\beta(\boldsymbol{u})$. *The claim below holds irrespective of the choice of* $\boldsymbol{u}^{\mathrm{eq},\beta} \in \mathcal{U}_\beta^{\mathrm{eq}}$. *Let* $n \geq 5$. *There is a reciprocal recommendation task with* n *users such that:*

$$\forall \theta \in \Theta, \boldsymbol{u}^\theta, \exists \beta > 0 : \quad \forall i \in [\![n]\!], u_i^\theta > u_i^{\mathrm{eq},\beta} \qquad \text{and} \qquad \lim_{\beta \to \infty} \sum_{i \in \mathcal{N}} u_i^{\mathrm{eq},\beta} = 0.$$

Proofs and additional results All proofs are deferred to App. A.4, where we provide several additional results regarding the use of quality-weighted exposure and equality of exposure in

reciprocal recommendation: We show in Prop. 25 that there are cases where both approaches lead to user utility profiles with Lorenz-dominated curves, and significantly lower total user utility than the welfare approach for any choice of the parameters.

3.4 Efficient inference of fair rankings with the Frank-Wolfe algorithm

We now present our inference algorithm for (3.1). Appendix A.5 contains the proofs of this section and describes a similar approach for the objective functions of the previous section. From an abstract perspective, the goal is to find a maximum P^* such that:

$$P^* \in \operatorname*{argmax}_{P \in \mathcal{P}} W(P) \quad \text{with} \quad W(P) = \sum_{i=1}^{n} \Phi_i\Big(\sum_{j=1}^{n} \mu_{ij}(P_{ij} + P_{ji})v \Big)$$

where for every i, $\Phi_i : \mathbb{R}_+ \to \mathbb{R}$ is concave increasing, $\mu_{ij} \geq 0$ and v is a vector of non-negative non-increasing values. Since W is concave and $\mathcal{P}$ is defined by equality constraints, the problem above is a convex optimization problem. However, this is a global optimization problem over the rankings of all users, so a naive approach would require $|\mathcal{N}||\mathcal{I}|^2$ parameters and $2|\mathcal{N}||\mathcal{I}|$ linear constraints. The same problem arises with the penalties of previous work. In the literature, authors either considered applying the item-fairness constraints to each ranking individually [Singh and Joachims, 2018, Basu et al., 2020], which leads to inefficiencies with our definition of utility (see Appendix A.8), or resort to heuristics to compute the rankings one by one without guarantees on the trade-offs that are achieved [Morik et al., 2020, Biega et al., 2018].

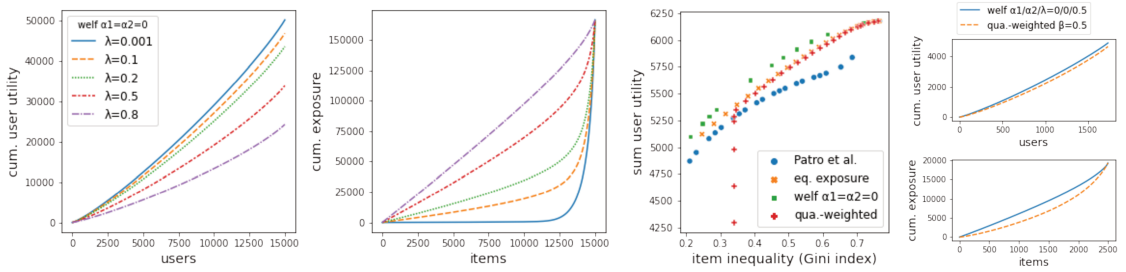
Our approach is based on the Frank-Wolfe algorithm [Frank and Wolfe, 1956], which was previously used in machine learning in e.g., structured output prediction or low-rank matrix completion [Jaggi, 2013], but to the best of our knowledge not for ranking. Denoting $\langle X \,|\, Y \rangle = \sum_{ijk} X_{ijk} Y_{ijk}$ the dot product between tensors, the algorithm creates iterates $P^{(t)}$ by first computing $\tilde{P} = \operatorname{argmax}_{P \in \mathcal{P}} \langle P \,|\, \nabla W(P^{(t)}) \rangle$ and then updating $P^{(t)} = (1 - \gamma^{(t)}) P^{(t-1)} + \gamma^{(t)} \tilde{P}$ with $\gamma^{(t)} = \frac{2}{t+2}$ [Clarkson, 2010]. Starting from an initial solution[5], the algorithm always stays in the feasible region without any additional projection step. Our main contribution of this section is to show that $\operatorname{argmax}_{P \in \mathcal{P}} \langle P \,|\, \nabla W(P^{(t)}) \rangle$ can be computed efficiently, requiring only one sort operation per user after computing the utilities. In the result below, for a ranking tensor P and a user i, we denote by $\mathfrak{S}(P_i)$ the support of P_i in ranking space.[6]

Theorem 4. *Let* $\tilde{\mu}_{ij} = \Phi_i'\big(u_i(P^{(t)})\big)\mu_{ij} + \Phi_j'\big(u_j(P^{(t)})\big)\mu_{ji}$. *Let* $\tilde{P}$ *such that:*
$\forall i \in \mathcal{N}, \forall \tilde{\sigma}_i \in \mathfrak{S}(\tilde{P}_i)$: $\tilde{\sigma}_i(j) < \tilde{\sigma}_i(j') \implies \tilde{\mu}_{ij} \geq \tilde{\mu}_{ij'}$. *Then* $\tilde{P} \in \operatorname*{argmax}_{P \in \mathcal{P}} \langle P \,|\, \nabla W(P^{(t)}) \rangle$.

Moreover, it produces a compact representation of the stochastic ranking as a weighted sum of permutation matrices. The number of iterations of the algorithm allows to control the trade-off between memory requirements and accuracy of the solution. Using previous convergence results for the Frank-Wolfe algorithm [Clarkson, 2010], assuming each Φ_i'' is bounded, we have:

Proposition 5. *Let* $B = \max_{i \in [\![n]\!]} \|\Phi_i''\|_\infty$ *and* $U = \max_{\boldsymbol{u} \in \mathcal{U}} \|\boldsymbol{u}\|_2^2$. *Let* K *be the maximum index of a nonzero value in* v *(or* $|\mathcal{I}|$*). Then* $\forall t \geq 1, W(P^{(t)}) \geq \max_{P \in \mathcal{P}} W(P) - O(\frac{BU}{t})$. *Moreover, for each user, an iteration costs* $O(|\mathcal{I}| \ln K)$ *operations and requires* $O(K)$ *additional bytes of storage.*

[5] In our experiments, we initialize with the utilitarian ranking (Proposition 21).
[6] Formally, $\mathfrak{S}(P_i) = \big\{\sigma : \mathcal{I} \to [\![|\mathcal{I}|]\!] \,\big|\, \sigma \text{ is one-to-one, and } \forall j \in \mathcal{I}, P_{ij\sigma(j)} > 0\big\}$.



(a) Examples of generalized Lorenz curves achieved by *welf*.(b) Summary of trade-offs(c) Dominated curve

Figure 3.2: Summary of results on Lastfm-2k, focusing on the user utility/item inequality trade-off.

3.5 Experiments

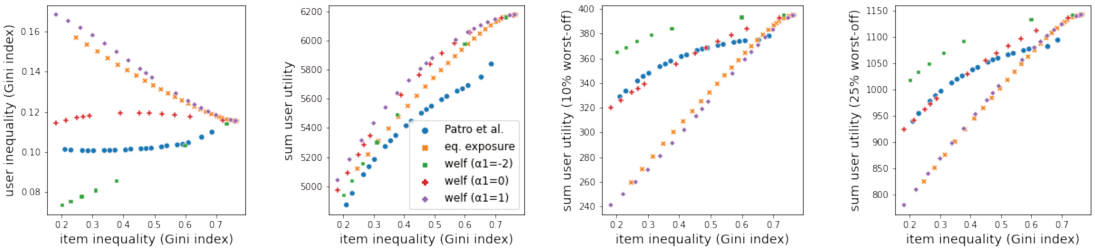
3.5.1 One-sided recommendation

We first present experiments on a music recommendation task. We report here our experiments with the Lastfm-2k dataset [Cantador et al., 2011, Patro et al., 2020], which contains the music listening histories of $1.9k$ users. We present in App. A.6.2 experiments on a larger portion of the Last.fm dataset, and in App. A.6.3 results using the MovieLens-20m dataset Harper and Konstan [2015]. Our results are qualitatively similar across the three datasets.

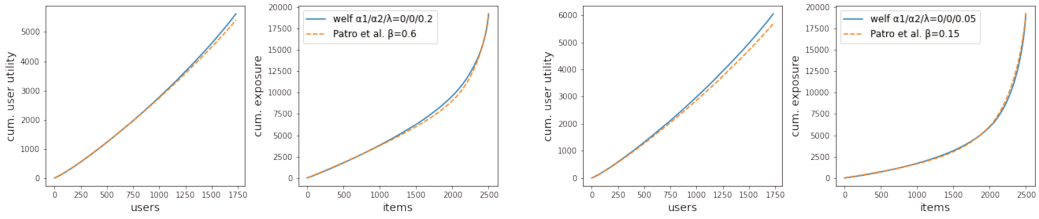
We select the top 2500 items most listened to, and estimate preferences with a matrix factorization algorithm using a random sample of 80% of the data. All experiments are carried out with three repetitions for this subsample. The details of the experimental protocol are in App. A.6.1. Since the goal is to analyze the behavior of the ranking algorithms rather than the quality of the preference estimates, we consider the estimated preferences as ground truth when computing user utilities and comparing methods, following previous work. We compare our welfare approach (*welf*) to three baselines. The first one is the algorithm of [Patro et al., 2020] (referred to as *Patro et al.* in the figures), who consider envy-freeness for user-side fairness and, for item-side fairness, a constraint that the minimum exposure of an item is $\beta \frac{E}{|\mathcal{I}|}$ where β is the trade-off parameter. The other baselines are quality-weighted exposure (*qua.-weighted*) and equality of exposure (*eq. exposure*) as described in Sec. 3.3.

Item-side fairness We first study in isolation item-side fairness, defined as improving the exposure of the worse-off item (producers). To summarize the trade-offs, we show the trade-offs by looking at exposure inequalities as measured by the Gini index (see Sec. 3.2.2). The results are given in Fig. 3.2:

- *Generating user utility/item inequality trade-offs* is performed with our approach by keeping $\alpha_1 = \alpha_2 = 0$ and varying the relative weight of items λ. Fig. 3.2a plots some trade-offs achieved by our approach. As expected, the user utility curve degrades as we increase the weight of items, while at the same time the curve of item exposure moves towards the straight line, which corresponds to strict equality of exposure. Fig. A.2 in the appendix provides analogous curves for all methods, obtained by varying the weight β of the inequality measure.
- *qua.-weighted yields unfair trade-offs* Fig. 3.2c shows a *welf* ranking that dominates a *qua.-weighted* ranking on both user and item curves. This is in line with the discussion of Section 3.3, *qua.-weighted* can lead to unfair rankings on utility/item inequality trade-offs.
- *welf dominates the user utility/item inequality (Gini) trade-offs* as seen on Fig. 3.2b: while all methods have the same total user utility when accepting high item inequality, *welf* dominates *Patro et al.*, *eq. exposure* and *qua.-weighted* as soon as Gini ≤ 0.5. Note, however, that the Gini



(a) User inequality. (b) Total user utility. (c) Cumulative utility at 10% worse-off users. (d) Cumulative utility at 25% worse-off users.



(e) Item inequality ≈ 0.4. (f) Item inequality ≈ 0.6.

Figure 3.3: Summary of results on Lastfm-2k for two-sided fairness: effect of varying α_1.

index is only one measure of inequality. When measuring item inequalities with the standard deviation, *eq. exposure* becomes optimal since our implementation optimizes a trade-off with this measure (see Fig. A.4 in App. A.6.1). Overall, *welf* and *eq. exposure* yield different fair trade-offs.

Two-sided fairness Fig. 3.3 shows the effect of the user curvature $\alpha_1 \in \{-2, 0, 1\}$, keeping $\alpha_2 = 0$. Fig. A.4 in App. A.6.1 shows similar plots when the item inequality is measured by the standard deviation rather than the Gini index.

- *Smaller α_1 reduce user inequalities at the expense of total user utility, at various levels of item inequality.* This is observed by comparing the results for $\alpha_1 \in \{-2, 0, 1\}$ in Fig. 3.3a and Fig. 3.3b.
- *welf $\alpha_1 = 0$ is better than Patro et al.*, which can be seen by jointly looking at Fig. 3.3c, 3.3d and Fig. 3.3b which give the cumulative utility at different points of the Lorenz curve (10%, 25% and 100% of the users respectively). We observe that *welf $\alpha_1 = 0$* is similar to *Patro et al.* at the 10% and 25% levels, but has higher total utility. Example curves are given in Fig. 3.3e and 3.3f which plot *welf $\alpha_1 = 0$* and *Patro et al.* at two levels of item inequality. *welf $\alpha_1 = 0$* obtains similar curves to *Patro et al.*, except that it performs better at the end of the curve. A similar comparison can be made with *welf $\alpha_1 = 1$* and *eq. exposure*.
- *More user inequalities is not necessarily unfair* as seen in Fig. 3.3a comparing *welf $\alpha_1 = 0$* and *Patro et al.*. We observe that *welf $\alpha_1 = 0$* has slightly higher Gini index, but this is not unfair: as seen in Fig. 3.3e and 3.3f, this is due to the higher utility at the end of the generalized Lorenz curve of *welf*, but the worse-off users have similar utilities with *welf* and *Patro et al.*.

3.5.2 Reciprocal recommendation

We now present results on a reciprocal recommendation task, where fairness refers to increasing the utility of the worse-off users (this can be done by boosting their exposure at the expense of total utility). Since there is no standard benchmark for reciprocal recommendation, we generate an artificial task based on the Higgs Twitter dataset [De Domenico et al., 2013], which contains follower links, and address the task of finding mutual followers (i.e., "matches"). We keep users having at least 20 mutual links, resulting in a subset of 13k users. We build estimated match probabilities

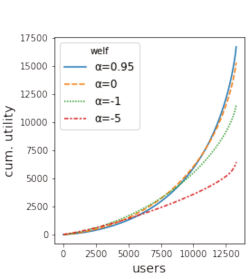
using matrix factorization. The experimental protocol is detailed in App. A.6.4. We also present in App. A.6.5 additional experiments using the Epinions dataset Richardson et al. [2003]. The results are qualitatively similar.

Our main baseline is equal utility (*eq. utility*) defined in Section 3.3. We also compare to quality-weighted exposure, and equality of exposure as baselines that ignore the reciprocal nature of the task. The results are summarized in Fig. 3.4:

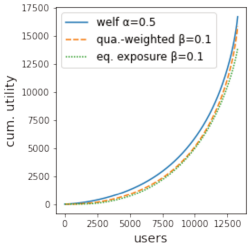
- *Example of trade-offs obtained by varying* α are plotted in Fig. 3.4a. As α decreases, the utility increases for the worse-off users at the expense of better-off users. We note that increasing the utility of worse-off users has a massive cost on total user utility: looking at the exact numbers we observe that $\alpha = -5$ has more than doubled the cumulative utility of the 10% worse off users compared to $\alpha = 1$ (120 vs 280), but at the cost of more than 60% of the total utility (17k vs 6.4k). Fig. A.2 in Appendix A.6.4 contains plots of the trade-offs achieved by the other methods.
- *qua.-weighted and eq. exposure are dominated by welf on a large range of hyperparameters.* An example is given in Fig. 3.4b, where *welf* $\alpha = 0.5$ already dominates some of their models, even though in this region of α there is little focus on worse-off users. More generally, all values of $\beta \geq 0.1$ for *qua.-weighted* and *eq. exposure* lead to rankings with dominated curves. This is expected since they ignore the reciprocal nature of the task.
- *eq. utility is dominated by welf near strict equality* as illustrated in Fig. 3.4c: for large values of β, it is not possible to increase the utility of the worse off users, and *eq. utility* only drags utility of better-off users down.
- *welf is more effective at increasing utility of the worse-off users* as can be seen in Fig. 3.4e-g, which plots the total utility as a function of the cumulative utility at different points of the Lorenz curve (10%, 20%, 50% worse-off users respectively). For total utilities larger than 50% of the maximum achievable, *welf* significantly dominates *eq. utility* in terms of utility of worse-off users (10% and 25%) at a given level of total utility. *welf* also dominates *eq. utility* on the 50% worse-off users (Fig. 3.4h) in the interesting region where the total utility is within 20% of the maximum.
- *More inequality is not necessarily unfair* As shown in Fig. 3.4d, we see that for the same utility for the 10% worse-off users, *welf* models have higher inequalities than *eq. utility*. As seen before, this higher inequality is due to a higher total utility (and higher total utilities for the 25% worse-off users. The analysis of these Lorenz curves allow us to conclude that these larger inequalities are not due to unfairness. They arise because *welf* optimizes the utility of the worse-off users at lower cost in terms of average utility than *eq. utility*.

3.6 Related work

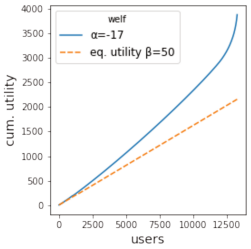
The question of fairness in rankings originated from independent audits on recommender systems or search engines, which showed that results could exhibit bias against relevant social groups [Sweeney, 2013, Kay et al., 2015, Hannak et al., 2014, Mehrotra et al., 2017, Lambrecht and Tucker, 2019] Our work follows the subsequent work on ranking algorithms that promote fairness of exposure for individual or sensitive groups of items [Celis et al., 2017b, Burke, 2017, Biega et al., 2018, Singh and Joachims, 2018, Morik et al., 2020, Zehlike and Castillo, 2020]. The goal is often to prevent winner-take-all effects, combat popularity bias [Abdollahpouri et al., 2019b] or promote smaller producers [Liu et al., 2019, Mehrotra et al., 2018]. Section 3.3 is devoted to the comparison with this type of approaches. Most of these works use a notion of fairness oriented towards items only. Towards two-sided fairness, Wang and Joachims [2020] promote user-side fairness using concave functions of user utilities, similarly to us. Other works use equality constraints to define user-side



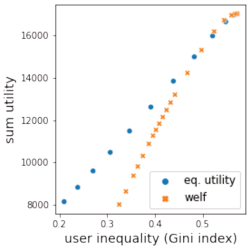
(a) Example trade-offs achieved by *welf*.



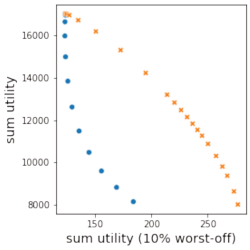
(b) *welf* dominates exposure-based methods.



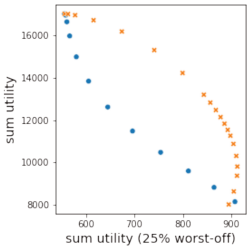
(c) *welf* dominates eq. utility near strict equality.



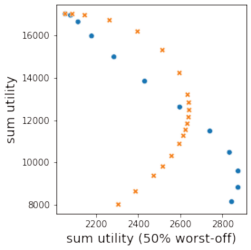
(d) Total utility vs inequality.



(e) Total utility vs utility of 10% worse-off users.



(f) Total utility vs utility of 25% worse-off users.



(g) Total utility vs utility of 50% worse-off users.

Figure 3.4: Results on the twitter dataset.

fairness [Basu et al., 2020, Wu et al., 2021b]. These three approaches rely on the definitions of item-side fairness discussed in Section 3.3. Patro et al. [2020] generate rankings that are envy-free on the user side, and guarantees the fair min-share for items. This approach is not amenable to controllable trade-offs between user and item utilities.

We are the first to address one-sided and reciprocal recommendation within the same framework. There is less existing work studying the fairness of rankings in the reciprocal setting. Xia et al. [2019] aim at equalizing user utility between groups, which suffers from the problems discussed in Section 3.3. Jia et al. [2018] generate rankings using a welfare function approach, but optimizing only the utility of users *being recommended*. Paraschakis and Nilsson [2020] postprocess rankings to correct for inconsistencies between estimated and declared preferences of users. In contrast, we aim at fair trade-offs between user and item utilities, under the assumption that biases in the preference estimates have been addressed earlier in the recommendation pipeline. Fairness is also studied in the context of ridesharing applications [Wolfson and Lin, 2017, Lesmana et al., 2019, Nanda et al., 2020], but they address matching rather than ranking problems.

There is growing interest in making the relationship between fairness in machine learning and social choice theory [Heidari et al., 2018, Ustun et al., 2019, Balcan et al., 2018, Gölz et al., 2019, Hossain et al., 2020, Chakraborty et al., 2019, Do et al., 2021b, Finocchiaro et al., 2021], and welfare economics in particular [Speicher et al., 2018, Hu and Chen, 2020, Kleinberg et al., 2018b, Zimmer et al., 2021]. In line with Hu and Chen [2020], who focused on classification and parity penalties, we argue that Pareto-efficiency should be part of fairness assessments. We are the first to propose concave welfare functions and Lorenz dominance to address two-sided fairness in recommendation.

3.7 Conclusion

We view fairness in rankings as optimizing the distribution of user and item utilities, giving priority to the worse-off. Following this view, we defined fair rankings as having non-dominated generalized

Lorenz curves of user and item utilities, and develop a new conceptual and algorithmic framework for fair ranking. The generality of the approach is showcased on several recommendation tasks, including reciprocal recommendation.

The expected positive societal impact of this work is to provide more principled approaches to mediating between several parties on a recommendation platform. Yet, we did not address several questions that are critical for the deployment of our approach. In particular, true user preferences are often not directly available, and we only observe proxies to them, such as clicks or likes. Second, interpersonal comparisons of utilities are critical in this work. It is thus necessary to make sure that the proxies we choose lead to meaningful comparisons of utilities between users. Third, estimating preferences or their proxies is itself not trivial in recommendation because of partial observability. The true fairness of our approach is bound to a careful analysis of (at least) these additional steps.

Chapter 4

Fairness in rankings with generalized Gini welfare functions

Contents

This chapter is the article *Optimizing generalized Gini indices for fairness in rankings*, published at SIGIR 2022 (see [Do and Usunier, 2022]). This chapter uses the notation of the original article, which is the same as the notation of Chapter 1.

In this chapter, we build on the previous conceptual framework and propose an alternative approach to additive welfare functions that also produces Lorenz-efficient rankings. We introduce the maximization of Generalized Gini welfare Functions (GGFs) for fair ranking, which allows to generate *all* Lorenz-efficient rankings. In contrast, maximizing additive concave welfare functions produces Lorenz-efficient rankings, but not all of them in general. While additive welfare functions have an intuitive interpretation as utilitarianism with diminishing returns, GGFs can express fairness criteria based on utility quantiles and classical inequality measures like the Gini index.

On the technical side, this chapter addresses the challenge of optimizing GGFs, in the batch setting. Since GGFs are nondifferentiable, we cannot use the Frank-Wolfe algorithm of the previous

chapter which was limited to smooth functions. To overcome this, we introduce a Frank-Wolfe variant that uses the Moreau-Yosida envelope as a smoothing technique, and present a computationally efficient procedure for computing the smooth approximation of GGFs.

The limitations of the modelling choices made in this chapter are further discussed in Chapter 7.

Abstract

There is growing interest in designing recommender systems that aim at being fair towards item producers or their least satisfied users. Inspired by the domain of inequality measurement in economics, this paper explores the use of generalized Gini welfare functions (GGFs) as a means to specify the normative criterion that recommender systems should optimize for. GGFs weight individuals depending on their ranks in the population, giving more weight to worse-off individuals to promote equality. Depending on these weights, GGFs minimize the Gini index of item exposure to promote equality between items, or focus on the performance on specific quantiles of least satisfied users. GGFs for ranking are challenging to optimize because they are non-differentiable. We resolve this challenge by leveraging tools from non-smooth optimization and projection operators used in differentiable sorting. We present experiments using real datasets with up to 15k users and items, which show that our approach obtains better trade-offs than the baselines on a variety of recommendation tasks and fairness criteria.

4.1 Introduction

Recommender systems play an important role in organizing the information available to us, by deciding which content should be exposed to users and how it should be prioritized. These decisions impact both the users and the item producers of the platform. While recommender systems are usually designed to maximize performance metrics of user satisfaction, several audits recently revealed potential performance disparities across users [Sweeney, 2013, Datta et al., 2015, Ekstrand et al., 2018, Mehrotra et al., 2017]. On the side of item producers, the growing literature on fairness of exposure aims to avoid popularity biases [Abdollahpouri et al., 2019b] by reducing inequalities in the exposure of different items [Singh and Joachims, 2018], or aiming for equal exposure weighted by relevance [Diaz et al., 2020, Biega et al., 2018, Morik et al., 2020]. In most cases, the approaches proposed for user- and item-side fairness aim to reduce inequalities.

In this paper, we propose a new approach to fair ranking based on Generalized Gini welfare Functions (GGFs, [Weymark, 1981]) from the economic literature on inequality measurement [Cowell, 2000]. GGFs are used to make decisions by maximizing a weighted sum of the utilities of individuals which gives more weight to those with lower utilities. By prioritizing the worse-off, GGFs promote more equality.

The normative appeal of GGFs lies in their ability to address a multiplicity of fairness criteria studied in the fair recommendation literature. Since GGFs include the well-known Gini inequality index as a special case [Gini, 1921], they can be used to optimize trade-offs between exposure inequality among items and user utility, a goal seeked by many authors [Morik et al., 2020, Zehlike and Castillo, 2020]. GGFs also conveniently specify normative criteria based on utility quantiles [Do et al., 2021c]: for instance, it is possible to improve the utility of the 10% worse-off users and/or items with GGFs, simply by assigning them more weight in the objective. Moreover, using techniques from convex multi-objective optimization, we show that GGFs cover *all* ranking policies that satisfy *Lorenz efficiency*, a distributive justice criterion which was recently introduced for

two-sided fairness in rankings [Do et al., 2021c].

The difficulty of using GGFs as objective functions for fairness in ranking stems from their non-differentiability, which leads to computational challenges. Indeed, ranking with fairness of exposure requires the solution of a global optimization problem in the space of (randomized) rankings of all users, because the exposure of an item is the sum of its exposure to every users. The Frank-Wolfe algorithm [Frank and Wolfe, 1956] was shown to be a computationally efficient method for maximizing globally fair ranking objectives, requiring only one top-K sort operation per user at each iteration [Do et al., 2021c]. However, vanilla Frank-Wolfe algorithms only apply to objective functions that are differentiable, which is not the case of GGFs.

We propose a new algorithm for the optimization of GGFs based on extensions of Frank-Wolfe algorithms for non-smooth optimization [Lan, 2013, Yurtsever et al., 2018, Thekumparampil et al., 2020a]. These methods usually optimize smoothed surrogate objective functions, while gradually decreasing a smoothing parameter, and a common smoothing technique uses the Moreau envelope [Moreau, 1962, Yosida et al., 1965]. Our main insight is that the gradient of the Moreau envelope of GGFs can be computed in $O(n \log n)$ operations, where n is the number of users or items. This result unlocks the use of Frank-Wolfe algorithms with GGFs, allowing us to efficiently find optimal ranking policies while optimizing GGFs.

We showcase the performances of the algorithm on two recommendations tasks of movies and music, and on a reciprocal recommendation problem (akin to dating platforms, where users are recommended to other users), with datasets involving up to $15k$ users and items. Compared to relevant baselines, we show that our algorithm successfully yields better trade-offs in terms of user utility and inequality in item exposure measured by the Gini index. Our approach also successfully finds better trade-offs in terms of two-sided fairness when maximizing the lower quantiles of user utility while minimizing the Gini index of item exposure.

In the remainder of the paper, we first describe our recommendation framework. We then present the family of generalized Gini welfare functions and its relationship to previously proposed fairness criteria in ranking. In Sec. 4.3 we provide the details of our algorithm and the convergence guarantees. Our experimental results are reported in Sec. 4.4, and an extension to reciprocal recommendation problems is discussed in Sec. 4.5. We position our approach with respect to the related work in Sec. 4.6, and Sec. 4.7 concludes the paper and discusses the limitations of our work.

4.2 Fair ranking with Generalized Gini functions

4.2.1 Recommendation framework

We consider a recommendation scenario with n users, and m items, and K recommendation slots. $\mu_{ij} \in [0, 1]$ denotes the value of item j for user i (e.g, a "liking" probability), and we assume the values μ are given as input to the system. The goal of the system is to produce a ranked list of items for each of the n users. Following previous work on fair rankings [e.g. Singh and Joachims, 2018], we consider randomized rankings because they enable the use of convex optimization techniques to generate the recommendations, which would otherwise involve an intractable combinatorial optimization problem in the space of all users' rankings. A *randomized ranking* for user i is represented by a bistochastic matrix $P_i \in \mathbb{R}^{m \times m}$, where P_{ijk} is the probability that item j is recommended to user i at position k. The recommender system is characterized by a *ranking policy* $P = (P_i)_{i=1}^n$. We denote the convex set of ranking policies by $\mathcal{P}$.

We use the term *utility* in its broad sense in cardinal welfare economics as a "*measurement of the higher-order characteristic that is relevant to the particular distributive justice problem at hand*"

[Moulin, 2003]. Similarly to Patro et al. [2020], Wang and Joachims [2021], Do et al. [2021c], we define the utility of a user as the ranking performance, and the utility of an item as its average exposure to users, which are formalized in (4.1) below. Utilities are defined according to the position-based model [Biega et al., 2018, Morik et al., 2020, Do et al., 2021c] with weights $\boldsymbol{b} \in \mathbb{R}^m_+$. The weight b_k is the probability that a user examines the item at position k, and we assume that the weights are non-increasing. Since there are K recommendation slots, we have $b_1 \geq \ldots \geq b_K$ and $b_k = 0$ for any $k > K$. The user and item utilities are then:

$$\textit{User utility: } u_i(P) = \sum_{j=1}^{m} \mu_{ij} P_{ij}^\top \boldsymbol{b} \qquad \textit{Item exposure: } v_j(P) = \sum_{i=1}^{n} P_{ij}^\top \boldsymbol{b}. \tag{4.1}$$

We follow a general framework where the ranking policy P is found by maximizing a global *welfare function* $F(P)$, and the welfare function is a weighted sum of welfare functions for users and items:

$$F(P) = (1-\lambda) g^{\text{user}}(\boldsymbol{u}(P)) + \lambda g^{\text{item}}(\boldsymbol{v}(P)), \tag{4.2}$$

where $g^{\text{user}} : \mathbb{R}^n \to \mathbb{R}$ and $g^{\text{item}} : \mathbb{R}^m \to \mathbb{R}$ respectively aggregate the utilities of users and item exposures and $\lambda \in [0, 1]$ specifies the relative weight of users and items.

4.2.2 Generalized Gini welfare functions

In this work, we focus on the case where g^{item} and g^{user} are based on Generalized Gini welfare Functions (GGFs) [Weymark, 1981]). A GGF $g_{\boldsymbol{w}} : \mathbb{R}^n \to \mathbb{R}$ is a function parameterized by a vector $\boldsymbol{w} \in \mathbb{R}^n$ of non-increasing positive weights such that $w_1 = 1 \geq \ldots \geq w_n \geq 0$, and defined by a weighted sum of its sorted inputs, which is also called an ordered weighted averaging operator (OWA) [Yager, 1988]. Formally, let $\boldsymbol{x} \in \mathbb{R}^n$ be a utility vector and denote by $\boldsymbol{x}^\uparrow$ the values of $\boldsymbol{x}$ sorted in increasing order, i.e., $x_1^\uparrow \leq ... \leq x_n^\uparrow$. Then:

$$g_{\boldsymbol{w}}(\boldsymbol{x}) = \sum_{i=1}^{n} w_i x_i^\uparrow.$$

Let $\mathcal{V}_n = \{\boldsymbol{w} \in \mathbb{R}^n : w_1 = 1 \geq \ldots \geq w_n \geq 0\}$ be the set of admissible weights of GGFs. Given $\boldsymbol{w^1} \in \mathcal{V}_n$, $\boldsymbol{w^2} \in \mathcal{V}_m$ and $\lambda \in (0, 1)$, we define the *two-sided GGF* as the welfare function (4.2) with $g^{\text{user}} = g_{\boldsymbol{w^1}}$ and $g^{\text{item}} = g_{\boldsymbol{w^2}}$:

$$F_{\lambda, \boldsymbol{w^1}, \boldsymbol{w^2}}(P) = (1-\lambda) g_{\boldsymbol{w^1}}\big(\boldsymbol{u}(P)\big) + \lambda g_{\boldsymbol{w^2}}\big((\boldsymbol{v}(P)\big). \tag{4.3}$$

With non-increasing, non-negative weights $\boldsymbol{w}$, OWA operators are concave [Yager, 1988]. The maximization of $F_{\lambda, \boldsymbol{w^1}, \boldsymbol{w^2}}(P)$ (4.3) is thus a convex optimization problem (maximization of a concave function over the convex set of ranking policies). GGFs address fairness from the point of view of distributive justice in welfare economics [Moulin, 2003], because they assign more weight to the portions of the population that have the least utility. Compared to a standard average, a GGF thus promotes more equality between individuals.

Relationship to the Gini index GGFs are welfare functions so they follow the convention that they should be maximized. Moreover, if $w_i > 0$ for all i, $g_{\boldsymbol{w}}$ is increasing with respect to every individual utilities, which ensures that maximizers of GGFs are Pareto-optimal [Moulin, 2003]. The Gini index of $\boldsymbol{x}$, denoted Gini($\boldsymbol{x}$) is associated to the GGF $g_{\boldsymbol{w}}(\boldsymbol{x})$ with $w_i = {}^{(n-i+1)}/_n$ [for formulas

of Gini index, see Yitzhaki and Schechtman, 2013]:

$$\begin{aligned}\text{Gini}(\boldsymbol{x}) &= 1 - \frac{2}{\|\boldsymbol{x}\|_1}\sum_{i=1}^{n}\frac{n-i+1}{n}x_i^{\uparrow} \\ &= \frac{1}{n^2\overline{\boldsymbol{x}}}\sum_{i=1}^{n}\sum_{j=1}^{n}|x_i - x_j| \qquad\qquad \text{with } \overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n}x_i.\end{aligned} \tag{4.4}$$

The second equality gives a more intuitive formula as a normalized average of absolute pairwise differences. The Gini index is an inequality measure, and therefore should be minimized, but, more importantly, it is normalized by the sum of utilities $\|\boldsymbol{x}\|_1$, which means that in general minimizing the Gini index does not yield Pareto-optimal solutions. The importance of this normalization is discussed by e.g., Atkinson [1970], and by [Do et al., 2021c] in the context of fairness in rankings. Yet, when $\boldsymbol{x}$ is a vector of item exposures $\boldsymbol{x} = \boldsymbol{v}(P)$, the normalization is not important because the total exposure is constant. It is then equivalent to minimize the Gini index of item exposures or to maximize its associated GGF.

Multi-objective optimization of Lorenz curves An alternative formula for $g_{\boldsymbol{w}}(\boldsymbol{x})$ is based on the generalized Lorenz curve[1] [Shorrocks, 1983] of $\boldsymbol{x}$, which is denoted $\boldsymbol{X}$ and is defined as the vector of cumulative sums of sorted utilities:

$$g_{\boldsymbol{w}}(\boldsymbol{x}) = \sum_{i=1}^{n} w_i' X_i \text{ where } w_i' = w_i - w_{i+1} \text{ and } X_i = x_1^{\uparrow} + \ldots + x_i^{\uparrow}. \tag{4.5}$$

We used the convention $w_{n+1} = 0$. Notice that since the weights $\boldsymbol{w}$ are non-increasing, we have that $w_i' \geq 0$. Thus, family of admissible OWA weights $\boldsymbol{w}$ yield weights $\boldsymbol{w}'$ that are non-negative and sum to 1. This formula offers the interpretation of GGFs as positively weighted averages of points of the generalized Lorenz curves. Every GGF thus corresponds to a scalarization of the multi-objective problem of maximizing every point of the generalized Lorenz curve [Geoffrion, 1968, Miettinen, 2012]. We get back to this interpretation in the next subsections.

4.2.3 GGFs for fairness in rankings

To give concrete examples of the relevance of GGFs for fairness in rankings, we provide here two fairness evaluation protocols that have been previously proposed and fall under the scope of maximizing of GGFs as in Eq. (4.3).

Trade-offs between user utility and inequality in item exposure The first task consists in mitigating inequalities of exposure between (groups of) items, and appears in many studies [Singh and Joachims, 2018, Zehlike and Castillo, 2020, Wu et al., 2021b]. This leads to a trade-off between the total utility of users and inequality among items, and such inequalities are usually measured by the Gini index (as in [Morik et al., 2020, Biega et al., 2018]). Removing the dependency on P to lighten the notation, a natural formulation of this trade-off uses the two-sided GGF (4.3) by setting $\boldsymbol{w^1} = (1, \ldots, 1)$ and $\boldsymbol{w^2} = \left(\frac{m-j+1}{m}\right)_{j=1}^{m}$, which yields:

$$g^{\text{user}}(\boldsymbol{u}) = \frac{1}{n}\sum_{i=1}^{n} u_i \qquad g^{\text{item}}(\boldsymbol{v}) = \sum_{j=1}^{m}\frac{m-j+1}{m}v_j^{\uparrow}. \tag{4.6}$$

[1] Lorenz curves are normalized so that the last value is 1, while generalized Lorenz curves are not normalized.

As stated in the previous section, for item exposure, maximizing g^{item} is equivalent to minimizing the Gini index. The Gini index for g^{item} has been routinely used for *evaluating* inequality in item exposure [Morik et al., 2020, Biega et al., 2018] but there is no algorithm to optimize general trade-offs between user utility and the Gini index of exposure. Morik et al. [2020] use the Gini index of exposures in the context of dynamic ranking (with the absolute pairwise differences formula (4.4)), where their algorithm is shown to asymptotically drive $g^{\text{item}}(\boldsymbol{v})$ to 0, equivalent to $\lambda \to 1$ in (4.2). However, their algorithm cannot be used to converge to the optimal rankings for other values of λ. Do et al. [2021c] use as baseline a variant using the standard deviation of exposures instead of absolute pairwise difference because it is easier to optimize (it is smooth except on 0). In contrast, our approach allows for the direct optimization of the welfare function (4.2) with this instantiation of g^{item} given by eq. (4.6).

Several authors [Morik et al., 2020, Biega et al., 2018] used *merit-weighted* exposure[2] $\boldsymbol{v}'_j(P) = \boldsymbol{v}(P)/\overline{\mu}_j$ where $\overline{\mu}_j = \frac{1}{n}\sum_{i=1}^{n}\mu_{ij}$ is the average value of item j across users, rather than the exposure itself. We keep the non-weighted exposure to simplify the exposition, but our method straightforwardly applies to merit-weighted exposure. Note however that the sum of weighted exposures is not constant, so using (4.6) with merit-weighted exposures is not strictly equivalent to minimizing the Gini index.

Two-sided fairness Do et al. [2021c] propose to add a user-side fairness criterion to the trade-off above, to ensure that worse-off users do not bear the cost of reducing exposure inequalities among items. Their evaluation involves multi-dimensional trade-offs between specific points of the generalized Lorenz curve. Using the formulation (4.5) of GGFs, trade-offs between maximizing the cumulative utility at a specific quantile q of users and total utility can be formulated using a parameter $\omega \in [0, 1]$ as follows:

$$g^{\text{user}}(\boldsymbol{u}) = \sum_{i=1}^{n} w'_i U_i \quad \text{with } w'_{\lfloor qn \rfloor} = \omega \text{ and } w'_n = 1 - \omega, \tag{4.7}$$

where all other values of $w'_i = 0$. In our experiments, we combine this g^{user} with the Gini index for g^{item} for two-sided fairness.

4.2.4 Generating all Lorenz-efficient solutions

In welfare economics, the fundamental property of concave welfare functions is that they are monotonic with respect to the dominance of generalized Lorenz curves [Atkinson, 1970, Shorrocks, 1983, Moulin, 2003], because this garantees that maximizing a welfare function performs an optimal redistribution from the better-off to the worse-off at every level of average utility. In the context of two-sided fairness in rankings, Do et al. [2021c] formalize their fairness criterion by stating that a ranking policy is fair as long as the generalized Lorenz curves of users and items are not jointly dominated. In this section, we show that the family of GGFs $F_{\lambda,\boldsymbol{w}^1,\boldsymbol{w}^2}(P)$ (4.3) allows to generate *every* ranking policy that are fair under this definition, and *only* those. The result follows from standard results of convex multi-objective optimization [Geoffrion, 1968, Miettinen, 2012]. We give here the formal statements for exhaustivity.

Let $\boldsymbol{x}$ and $\boldsymbol{x}'$ two vectors in $\mathbb{R}^n_+$. We say that $\boldsymbol{x}$ weakly-Lorenz-dominates $\boldsymbol{x}'$, denoted $\boldsymbol{x} \succeq_{\text{L}} \boldsymbol{x}'$, when the generalized Lorenz curve of $\boldsymbol{x}$ is always at least equal to that of $\boldsymbol{x}'$, i.e., $\boldsymbol{x} \succeq_{\text{L}} \boldsymbol{x}' \iff \forall i, X_i \geq X'_i$. We say that $\boldsymbol{x}$ Lorenz-dominates $\boldsymbol{x}'$, denoted $\boldsymbol{x} \succ_{\text{L}} \boldsymbol{x}'$ if $\boldsymbol{x} \succeq_{\text{L}} \boldsymbol{x}'$ and $\boldsymbol{x} \neq \boldsymbol{x}'$, i.e., if the generalized Lorenz curve of $\boldsymbol{x}$ is strictly larger than that of $\boldsymbol{x}'$ on at least one point. The

[2] also called *"equity of attention"* [Biega et al., 2018], *"disparate treatment"* [Singh and Joachims, 2018]

criterion that generalized Lorenz curves of users and items are not jointly-dominated is captured by the notion of *Lorenz-efficiency*:

Definition 2 (Do et al. [2021c]). *A ranking policy $P \in \mathcal{P}$ is* Lorenz-efficient *if there is no $P' \in \mathcal{P}$ such that either $[\boldsymbol{u}(P') \succeq_{\text{L}} \boldsymbol{u}(P)$ and $\boldsymbol{v}(P') \succ_{\text{L}} \boldsymbol{v}(P)]$ or $[\boldsymbol{v}(P') \succeq_{\text{L}} \boldsymbol{v}(P)$ and $\boldsymbol{u}(P') \succ_{\text{L}} \boldsymbol{u}(P)]$.*

We now present the main result of this section:

Proposition 6. *Let $\Theta = (0,1) \times \mathcal{V}_n \times \mathcal{V}_m$.*

1. *Let $(\lambda, \boldsymbol{w^1}, \boldsymbol{w^2}) \in \Theta$, where $\boldsymbol{w^1}$ and $\boldsymbol{w^2}$ have strictly decreasing weights, and $P^* \in \operatorname{argmax}_{P \in \mathcal{P}} F_{\lambda, \boldsymbol{w^1}, \boldsymbol{w^2}}(P)$. Then P^* is Lorenz-efficient.*
2. *If P is Lorenz-efficient, then there exists $(\lambda, \boldsymbol{w^1}, \boldsymbol{w^2}) \in \Theta$ such that $P \in \operatorname{argmax}_{P \in \mathcal{P}} F_{\lambda, \boldsymbol{w^1}, \boldsymbol{w^2}}(P)$.*

Proof. The proof uses standard results on convex multi-objective optimization from [Geoffrion, 1968, Miettinen, 2012]. Written in the form (4.5), the GGFs corresponds to the scalarization of the multi-objective problem of jointly maximizing the generalized Lorenz curves of users and items, which is a problem with $n + m$ objectives. Indeed, each objective function is a point of the generalized Lorenz curve $(\boldsymbol{U}(P), \boldsymbol{V}(P))$. Each objective $U_i(P)$ is concave because it corresponds to an OWA operator with non-increasing weights $\boldsymbol{\rho}$ with $\rho_{i'} = \mathbb{1}_{\{i' \leq i\}}$, applied to utilities, which are linear functions of the ranking policy. Each objective $V_i(P)$ is similarly concave. Moreover, we are optimizing over the convex set of stochastic ranking policies $\mathcal{P}$. The multi-objective problem is then concave, which means that the maximizers of all weighted sums of the objectives $(\boldsymbol{U}(P), \boldsymbol{V}(P))$ with strictly positive weights are Pareto-efficient. Reciprocally every Pareto-efficient solution is a solution of a non-negative weighted sum of the objectives $(\boldsymbol{U}(P), \boldsymbol{V}(P))$, where the weights sum to 1 [Miettinen, 2012].

The result follows from the observation that the Lorenz-efficiency of P, defined as the Lorenz-efficiency of $(\boldsymbol{u}(P), \boldsymbol{v}(P))$, is equivalent to the Pareto-efficiency of its joint user-item Lorenz curves $(\boldsymbol{U}(P), \boldsymbol{V}(P))$. This is because the Lorenz dominance relation between vectors $\boldsymbol{x}, \boldsymbol{x'}$ is defined as Pareto dominance in the space of their generalized Lorenz curves $\boldsymbol{X}, \boldsymbol{X'}$.

□

Additive welfare functions vs GGFs Do et al. [2021c] use additive concave welfare functions to generate Lorenz-efficient rankings. Let $\phi(x, \alpha) = x^\alpha$ if $\alpha > 0$, $\phi(x, \alpha) = \log(x)$ if $\alpha = 0$ and $\phi(x, \alpha) = -x^\alpha$ if $\alpha < 0$. Do et al. [2021c] use concave welfare functions of the form:

$$g^{\text{user}}(\boldsymbol{u}) = \sum_{i=1}^{n} \phi(u_i, \alpha_1) \qquad g^{\text{item}}(\boldsymbol{v}) = \sum_{j=1}^{m} \phi(v_j, \alpha_2) \tag{4.8}$$

Where α_1 (resp. α_2) specifies how much the rankings should redistribute utility to worse-off users (resp. least exposed items).

Additive separability plays an important role in the literature on inequality measures [Dalton, 1920, Atkinson, 1970, Cowell, 1988], as well as in the study of welfare functions because additive separability follows from a standard axiomatization [Moulin, 2003]. However, this leads to a restricted class of functions, so that varying α_1, α_2 and λ in (4.8) cannot generate all Lorenz-efficient solutions in general. The GGF approach provides a more general device to navigate the set of Lorenz-efficient solutions, with interpretable parameters since they are weights assigned to points of the generalized Lorenz curve.

4.3 Optimizing Generalized Gini Welfare

In this section, we provide a scalable method for optimizing two-sided GGFs welfare functions (4.3) $F_{\lambda,\boldsymbol{w}^1,\boldsymbol{w}^2}$. The challenge of optimizing GGFs is that they are nondifferentiable since they require sorting utilities. We first describe why existing approaches to optimize GGFs are not suited to ranking in Sec. 4.3.1. We then show how to efficiently compute the gradient of the Moreau envelope of GGFs in Sec. 4.3.2 and present the full algorithm in Sec. 4.3.3.

4.3.1 Challenges

In multi-objective optimization, a standard approach to optimizing OWAs is to solve the equivalent linear program derived by Ogryczak and Śliwiński [2003]. Because the utilities depend on 3d-tensors $P \in \mathcal{P}$ in our case, the linear program has $O(n \cdot m^2)$ variables and constraints, which is prohibitively large in practice. Another approach consists in using online subgradient descent to optimize GGFs, like [Busa-Fekete et al., 2017, Mehrotra et al., 2020]. This is not tractable in our case because it requires to project iterates onto the parameter space, which in our case involves costly projections onto the space of ranking policies $\mathcal{P}$. On the other hand, the Frank-Wolfe algorithm [Frank and Wolfe, 1956] was shown to provide a computationally efficient and provably convergent method to optimize over $\mathcal{P}$ [Do et al., 2021c]. However, it only applies to smooth functions, and Frank-Wolfe with subgradients may not converge to an optimal solution [Nesterov, 2018].

We turn to Frank-Wolfe variants for nonsmooth objectives, since Frank-Wolfe methods are well-suited to our structured ranking problem [Do et al., 2021c, Jaggi, 2013, Clarkson, 2010]. More precisely, following [Lan, 2013, Yurtsever et al., 2018, Thekumparampil et al., 2020a], our algorithm uses the Moreau envelope of GGFs for smoothing. The usefulness of this smooth approximation depends on its gradient, which computation is in some cases intractable [Chen et al., 2012]. Our main technical contribution is to show that the gradient of the Moreau envelope of GGFs can be computed in $O(n \log n)$ operations.

4.3.2 The Moreau envelope of GGFs

In the sequel, $\|\boldsymbol{z}\|$ denotes the ℓ_2 norm. Moreover, a function $L : \mathcal{X} \subseteq \mathbb{R}^n \to \mathbb{R}$ is C-smooth if it is differentiable with C-Lipschitz continuous gradients, i.e., if $\forall \boldsymbol{x}, \boldsymbol{x}' \in \mathcal{X}$, $\|\nabla L(\boldsymbol{x}) - \nabla L(\boldsymbol{x}')\| \leq C \|\boldsymbol{x} - \boldsymbol{x}'\|$.

4.3.2.1 Definition and properties

Let us fix weights $\boldsymbol{w} \in \mathcal{V}$ and focus on maximizing the GGF $g_{\boldsymbol{w}}$. Let $h(\boldsymbol{z}) := -g_{\boldsymbol{w}}(\boldsymbol{z})$ to obtain a convex function (this simplifies the overall discussion). The function h is $\|\boldsymbol{w}\|$-Lipschitz continuous, but non-smooth. We consider the smooth approximation of h given by its Moreau envelope [Parikh and Boyd, 2014] defined as:

$$h^\beta(\boldsymbol{z}) = \min_{\boldsymbol{z}' \in \mathbb{R}^n} h(\boldsymbol{z}') + \frac{1}{2\beta} \|\boldsymbol{z} - \boldsymbol{z}'\|^2 .$$

It is known that $h^\beta(\boldsymbol{z}) \leq h(\boldsymbol{z}) \leq h^\beta(\boldsymbol{z}) + \frac{\beta}{2} \|\boldsymbol{w}\|^2$ and that h^β is $\frac{1}{\beta}$-smooth [see e.g., Thekumparampil et al., 2020a]. The parameter β thus controls the trade-off between the smoothness and the quality of the approximation of h.

Algorithm 1: Computation of $\Pi_{\mathcal{C}(\tilde{\boldsymbol{w}})}$

input : GGF weights $\boldsymbol{w} \in \mathbb{R}^n$, $\boldsymbol{z} \in \mathbb{R}^n$
output : Projection of $\boldsymbol{z}$ onto the permutahedron $\mathcal{C}(\tilde{\boldsymbol{w}})$.

$\tilde{\boldsymbol{w}} \leftarrow -(w_n, ..., w_1)$ and $\sigma \leftarrow \text{argsort}(\boldsymbol{z})$
$\boldsymbol{x} \leftarrow \texttt{PAV}(z_\sigma - \tilde{\boldsymbol{w}})$
$\boldsymbol{y} \leftarrow \boldsymbol{z} + \boldsymbol{x}_{\boldsymbol{\sigma}^{-1}}$
Return $\boldsymbol{y}$.

4.3.2.2 Efficient computation of the gradient

We now present an efficient procedure to compute the gradient of $f^\beta(P) := h^\beta(\boldsymbol{u}(P))$.

Given an integer $n \in \mathbb{N}$, let $[\![n]\!] := \{1, \dots, n\}$ and let $\mathfrak{S}_n$ denotes the set of permutations of $[\![n]\!]$. For $\boldsymbol{x} \in \mathbb{R}^n$, and $\sigma \in \mathfrak{S}_n$, let us denote by $\boldsymbol{x}_{\boldsymbol{\sigma}} = (x_{\sigma(1)}, ..., x_{\sigma(n)})$. Furthermore, let $\mathcal{C}(\boldsymbol{x})$ denote the *permutahedron* induced by $\boldsymbol{x}$, defined as the convex hull of all permutations of the vector $\boldsymbol{x}$: $\mathcal{C}(\boldsymbol{x}) = \text{conv}\{\boldsymbol{x}_{\boldsymbol{\sigma}} : \sigma \in \mathfrak{S}_n\}$. Finally, let $\Pi_{\mathcal{X}}(\boldsymbol{z}) := \underset{\boldsymbol{z}' \in \mathcal{X}}{\text{argmin}} \|\boldsymbol{z} - \boldsymbol{z}'\|^2$. denote the projection onto a compact convex $\mathcal{X}$. The following proposition formulates ∇f^β as a projection onto a permutahedron:

Proposition 7. *Let $\tilde{\boldsymbol{w}} = -(w_n, ..., w_1)$. Let $P \in \mathcal{P}$. Then for all $(i, j, k) \in [\![n]\!] \times [\![m]\!]^2$, we have:*

$$\frac{\partial f^\beta}{\partial P_{ijk}}(P) = y_i \mu_{ij} b_k \qquad \text{where} \quad \boldsymbol{y} = \Pi_{\mathcal{C}(\tilde{\boldsymbol{w}})}\left(\frac{\boldsymbol{u}(P)}{\beta}\right). \tag{4.9}$$

Proof. Let $\text{prox}_{\beta h}(\boldsymbol{z}) = \underset{\boldsymbol{z}' \in \mathbb{R}^n}{\text{argmin}}\, \beta h(\boldsymbol{z}) + \frac{1}{2}\|\boldsymbol{z}' - \boldsymbol{z}\|^2$ denote the proximal operator of βh. Denoting by $\boldsymbol{u}*$ the adjoint of $\boldsymbol{u}$, it is known that $\nabla f^\beta(P) = \frac{1}{\beta}\boldsymbol{u}^*(\boldsymbol{u}(P) - \text{prox}_{\beta h}(\boldsymbol{u}(P))$ [Parikh and Boyd, 2014].

We first notice that since $\boldsymbol{w}$ are non-increasing, the rearrangement inequalities [Hardy et al., 1952] gives: $h(\boldsymbol{z}) = -\min_{\sigma \in \mathfrak{S}_n} \boldsymbol{w}_{\boldsymbol{\sigma}}^\intercal \boldsymbol{z} = \max_{\sigma \in \mathfrak{S}_n} -\boldsymbol{w}_{\boldsymbol{\sigma}}^\intercal \boldsymbol{z}$. Thus, h is the support function of the convex set $\mathcal{C}(\tilde{\boldsymbol{w}})$, since:

$$h(\boldsymbol{z}) = \max_{\sigma \in \mathfrak{S}_n} -\boldsymbol{w}_{\boldsymbol{\sigma}}^\intercal \boldsymbol{z} = \sup_{\boldsymbol{y} \in \mathcal{C}(\tilde{\boldsymbol{w}})} \boldsymbol{y}^\intercal \boldsymbol{z}.$$

Then the Fenchel conjugate of h is the indicator function of $\mathcal{C}(\tilde{\boldsymbol{w}})$, and its proximal is the projection $\Pi_{\mathcal{C}(\tilde{\boldsymbol{w}})}$ [Parikh and Boyd, 2014]. By Moreau decomposition, we get $\text{prox}(\boldsymbol{z}) = \boldsymbol{z} - \beta\, \Pi_{\mathcal{C}(\tilde{\boldsymbol{w}})}\left(\boldsymbol{z}/\beta\right)$, and thus:

$$\nabla f^\beta(P) = \boldsymbol{u}^*\left(\Pi_{\mathcal{C}(\tilde{\boldsymbol{w}})}\left(\boldsymbol{u}(P)/\beta\right)\right).$$

The result follows from the definition of $\boldsymbol{u}(P) = \left(\sum_{j,k} \mu_{ij} P_{ijk} b_k\right)_{i=1}^n$. □

Overall, computing the gradient of the Moreau envelope boils down to a projection onto the permutahedron $\mathcal{C}(\tilde{\boldsymbol{w}})$. This projection was shown by several authors to be reducible to isotonic regression:

Proposition 8 (Reduction to isotonic regression [Negrinho and Martins, 2014, Lim and Wright, 2016, Blondel et al., 2020])**.** *Let $\sigma \in \mathfrak{S}_n$ that sorts $\boldsymbol{z}$ decreasingly, i.e. $z_{\sigma(1)} \geq \ldots \geq z_{\sigma(n)}$. Let $\boldsymbol{x}$ be a solution to isotonic regression on $\boldsymbol{z}_{\boldsymbol{\sigma}} - \tilde{\boldsymbol{w}}$, i.e.*

$$\boldsymbol{x} = \underset{x'_1 \leq \ldots \leq x'_n}{\text{argmin}} \frac{1}{2}\|\boldsymbol{x}' - (\boldsymbol{z}_{\boldsymbol{\sigma}} - \tilde{\boldsymbol{w}})\|^2$$

Then we have: $\Pi_{\mathcal{C}(\tilde{\boldsymbol{w}})}(\boldsymbol{z}) = \boldsymbol{z} + \boldsymbol{x}_{\boldsymbol{\sigma}^{-1}}$.

Following these works, we use the Pool Adjacent Violators (PAV) algorithm for isotonic regression, which gives a solution in $O(n)$ iterations given a sorted input Best et al. [2000]. The algorithm for computing the projection is summarized in Alg. 1 where we use the notation $\text{argsort}(\boldsymbol{z}) = \{\sigma \in \mathfrak{S}_n : z_{\sigma(1)} \geq \ldots \geq z_{\sigma(n)}\}$ for permutations that sort $\boldsymbol{z} \in \mathbb{R}^n$ in decreasing order. Including the sorting of $\frac{\boldsymbol{u}(P)}{\beta}$, it costs $O(n \log n)$ time and $O(n)$ space.

Remark 1. *Our method is related to the differentiable sorting operator of Blondel et al. [2020], which uses a regularization term to smooth the linear formulation of sorting. The regularized form can itself be written as a projection to a permutahedron. The problem they address is different since they differentiate the multi-dimensional sort operation, but eventually the techniques are similar because the smoothing is done in a similar way.*

Remark 2. *We computed the gradient of $f^\beta(P) = h^\beta(\boldsymbol{u}(P))$ with user utilities. The gradient of $f^\beta(P) = h^\beta(\boldsymbol{v}(P))$ using item exposures is computed similarly: $\frac{\partial f^\beta}{\partial P_{ijk}}(P) = y_j b_k$ with $\boldsymbol{y} = \Pi_{\mathcal{C}(\tilde{\boldsymbol{w}})}\left(\frac{\boldsymbol{v}(P)}{\beta}\right)$.*

4.3.3 Frank-Wolfe with smoothing

We return to the optimization of the two-sided GGF objective (4.3). In this section, we fix the parameters $(\lambda, \boldsymbol{w^1}, \boldsymbol{w^2})$ and consider the minimization of $f := -F_{\lambda, \boldsymbol{w^1}, \boldsymbol{w^2}}$ over $\mathcal{P}$. For $\beta > 0$ we denote by h_1^β and h_2^β the Moreau envelopes of $-g_{\boldsymbol{w^1}}$ and $-g_{\boldsymbol{w^2}}$ respectively. The smooth approximation of f is then:

$$f^\beta(P) := (1-\lambda) h_1^\beta(\boldsymbol{u}(P)) + \lambda h_2^\beta(\boldsymbol{v}(P)).$$

Our algorithm FW-smoothing (Alg. 2) for minimizing f uses the Frank-Wolfe method for nonsmooth optimization from Lan [2013][3]. Given a sequence $(\beta_t)_{t\geq 1}$ of positive values decreasing to 0, the algorithm constructs iterates $P^{(t)}$ by applying Frank-Wolfe updates to f^{β_t} at each iteration t. More precisely, FW-smoothing finds an update direction with respect to ∇f^{β_t} by computing:

$$Q^{(t)} = \operatorname*{argmin}_{P \in \mathcal{P}} \langle P \,|\, \nabla f^{\beta_t}(P^{(t-1)}) \rangle. \tag{4.10}$$

The update rule is $P^{(t)} = P^{(t-1)} + \frac{2}{t+2}\left(Q^{(t)} - P^{(t-1)}\right)$.

Before giving the details of the computation of (4.10), we note that applying the convergence result of Lan [2013], and denoting $D_{\mathcal{P}} = \max_{P, P' \in \mathcal{P}} \|P - P'\|$ the diameter of $\mathcal{P}$, we obtain[4]:

Proposition 9 (Th. 4, [Lan, 2013]). *With $\beta_0 = \frac{2 D_{\mathcal{P}} b_1}{\|\boldsymbol{w}\|}$ and $\beta_t = \frac{\beta_0}{\sqrt{t}}$, FW-smoothing obtains the following convergence rate:*

$$f(P^{(T)}) - f(P^*) \leq \frac{2 D_{\mathcal{P}} b_1 \|\boldsymbol{w}\|}{\sqrt{T}}.$$

Efficient computation of the update direction For smooth welfare functions of user utilities and item exposures, the update direction (4.10) can be computed with only one top-K sorting operation per user [Do et al., 2021c]. In our case, the update is given by the following result, where $\text{top-}K(\boldsymbol{z}) = \{\sigma \in \mathfrak{S}_n : z_{\sigma(1)} \geq \ldots \geq z_{\sigma(K)} \text{ and } \forall k \geq K, z_{\sigma(K)} \geq z_{\sigma(k)}\}$ is the set of permutations that sort the k largest elements in $\boldsymbol{z}$.

[3] Lan [2013] uses the smoothing scheme of Nesterov [2005] which is in fact equal to the Moreau envelope (see [Beck and Teboulle, 2012, Sec. 4.3]).

[4] In more details, the convergence guarantee of Lan [2013] uses the operator norm of $\boldsymbol{u}$ and $\boldsymbol{v}$, which we bound as follows: $\|u(P)\|^2 \leq \sum_i \sum_{j,k} (\mu_{ij} P_{ijk} b_k)^2 \leq b_1^2 \|P\|^2$, because $\mu_{ij} \in [0,1]$ $b_k \in [0, b_1]$, and similarly $\|\boldsymbol{v}(P)\|^2 \leq b_1^2 \|P\|^2$.

Algorithm 2: FW-smoothing. Alg. 1 is used for $\boldsymbol{y}^1$ and $\boldsymbol{y}^2$.

input : values (μ_{ij}), # of iterations T, smoothing seq. $(\beta_t)_t$
output : ranking policy $P^{(T)}$

Initialize $P^{(0)}$ such that $P_i^{(0)}$ sorts μ_i in decreasing order
for *t=1, . . . , T* **do**
  Let $\boldsymbol{y}^1 = \Pi_{\mathcal{C}(\tilde{\boldsymbol{w}}^1)}\left(\frac{\boldsymbol{u}(P^{(t-1)})}{\beta_t}\right)$ and $\boldsymbol{y}^2 = \Pi_{\mathcal{C}(\tilde{\boldsymbol{w}}^2)}\left(\frac{\boldsymbol{v}(P^{(t-1)})}{\beta_t}\right)$
  for *i=1, . . . , n* **do**
    $\tilde{\mu}_{ij} = (1-\lambda)\, y_i^1 \mu_{ij} + \lambda y_j^2$
    $\tilde{\sigma}_i \leftarrow$ top-$K(-\tilde{\mu}_i)$ // Update direction (4.10)
  end
  Let $Q^{(t)} \in \mathcal{P}$ such that $Q_i^{(t)}$ represents $\tilde{\sigma}_i$
  $P^{(t)} \leftarrow (1 - \frac{2}{t+2})P^{(t-1)} + \frac{2}{t+2}Q^{(t)}$.
end
Return $P^{(T)}$.

Proposition 10. *Let $\tilde{\mu}$ defined by $\tilde{\mu}_{ij} = (1-\lambda)\, y_i^1 \mu_{ij} + \lambda y_j^2$ where $\boldsymbol{y}^1 = \Pi_{\mathcal{C}(\tilde{\boldsymbol{w}}^1)}\left(\boldsymbol{u}(P^{(t-1)})/\beta_t\right)$ and $\boldsymbol{y}^2 = \Pi_{\mathcal{C}(\tilde{\boldsymbol{w}}^2)}\left(\boldsymbol{v}(P^{(t-1)})/\beta_t\right)$.*

For all $i \in [\![n]\!]$, let $\tilde{\sigma}_i \in$ top-$K(-\tilde{\mu}_i)$ and $Q_i^{(t)}$ a permutation matrix representing $\tilde{\sigma}_i$. Then $Q^{(t)} \in \operatorname*{argmin}_{P\in\mathcal{P}} \langle P \,|\, \nabla f^{\beta_t}(P^{(t-1)})\rangle$.

Proof. Using the expression of the gradient of the Moreau envelope derived in Proposition 7, eq. (4.9), we have:

$$\frac{\partial f^{\beta_t}}{\partial P_{ijk}}(P^{(t-1)}) = (1-\lambda)\frac{\partial}{\partial P_{ijk}}(h_1^{\beta_t}(\boldsymbol{u}(P^{(t-1)})) + \lambda\frac{\partial}{\partial P_{ijk}}(h_2^{\beta_t}(\boldsymbol{v}(P^{(t-1)}))$$

And thus $\frac{\partial f^{\beta_t}}{\partial P_{ijk}}(P^{(t-1)}) = \tilde{\mu}_{ij} \times b_k$. The result then follows from [Do et al., 2021c, Lem. 3] and is a consequence of the rearrangement inequality Hardy et al. [1952]: $Q^{(t)}$ is obtained by sorting $\tilde{\mu}_{ij}$ in increasing order, or equivalently, by sorting $-\tilde{\mu}_{ij}$ in decreasing order. □

Since the computation of the gradient of Moreau envelopes costs $O(n \ln n + n \ln m)$ operations using Alg. 1, then by Prop. 10 at each iteration, the cost of the algorithm is dominated by the top-K sort per user, each of which has amortized complexity of $O(m + K \ln K)$:

Proposition 11. *Each iteration costs $O(nm + nK \ln K)$ operations. The total amount of storage required is $O(nKT)$.*

In conclusion, FW-smoothing has a cost per iteration similar to the standard Frank-Wolfe algorithm for ranking with smooth objective functions. The cost of the non-smoothness of the objective function is a convergence rate of $1/\sqrt{T}$, while the Frank-Wolfe algorithm converges in $O(1/T)$ when the objective is smooth [Clarkson, 2010].

Moreover, the algorithm produces a *sparse representation* of the stochastic ranking policy as a weighted sum of permutation matrices. In other words, this gives us a Birkhoff-von-Neumann decomposition [Birkhoff, 1940] of the bistochastic matrices *for free*, avoiding the overhead of an additional decomposition algorithm as in existing works on fair ranking [Singh and Joachims, 2018, Wang and Joachims, 2021, Su et al., 2021].

4.4 Experiments

We first present our experimental setting for recommendation of music and movies, together with the fairness criteria we explore and the baselines we consider. These fairness criteria have been chosen because they were used in the evaluation of prior work, and they exactly correspond to the optimization of a GGF. We thus expect our two-sided GGF $F_{\lambda,\boldsymbol{w}^1,\boldsymbol{w}^2}$ to fare better than the baselines, because they allow for the optimization of the exact evaluation criterion. We provide experimental results that demonstrate this claim in Sec. 4.4.2. Note that the GGFs are extremely flexible as we discussed in Sec. 4.2.1, so our experiments can only show a few illustrative examples of fairness criteria that can be defined with GGFs. In Sec. 4.4.3, we show the usefulness of FW-smoothing compared to the simpler baseline of Frank-Wolfe with subgradients.

4.4.1 Experimental setup

Our experiments are implemented in Python 3.9 using PyTorch[5]. For the `PAV` algorithm, we use the implementation of Scikit-Learn.[6]

4.4.1.1 Data and evaluation protocol

We present experiments on two recommendation tasks, following the protocols of [Do et al., 2021c, Patro et al., 2020]. First, we address music recommendation with **Lastfm-2k** from Cantador et al. [2011] which contains real listening counts of $2k$ users for $19k$ artists on the online music service Last.fm[7]. We filter the $2,500$ items having the most listeners. In order to show how the algorithm scales, we also consider the **MovieLens-20m** dataset Harper and Konstan [2015], which contains ratings in $[0.5, 5]$ of movies by users, and we select the top $15,000$ users and items with the most interactions.

We use an evaluation protocol similar to Patro et al. [2020], Do et al. [2021c], Wang and Joachims [2021]. For each dataset, a full user-item preference matrix $(\mu_{i,j})_{i,j}$ is obtained by standard matrix factorization algorithms[8] from the incomplete interaction matrix, following the protocol of [Do et al., 2021c]. Rankings are inferred from these estimated preferences. The exposure weights $\boldsymbol{b}$ are the standard weights of the *discounted cumulative gain* (DCG) (also used in e.g., Singh and Joachims [2018], Biega et al. [2018], Morik et al. [2020]): $\forall k \in [\![K]\!], b_k = \frac{1}{\log_2(1+k)}$.

The generated μ_{ij} are used as ground truth to evaluate rankings, in order to decouple the fairness evaluation of the ranking algorithms from the evaluation of biases in preference estimates (which are not addressed in the paper). The results are the average of three repetitions of the experiments over different random train/valid/test splits used to generate the μ_{ij}.

4.4.1.2 Fairness criteria

We remind two fairness tasks studied in the ranking literature and presented in Section 4.2.3, and describe existing approaches proposed to address them, which we consider as baselines for comparison with our two-sided GGF (4.3) $F_{\lambda,\boldsymbol{w}^1,\boldsymbol{w}^2}$.

Task 1: Trade-offs between user utility and inequality between items We use the two-sided GGF $F_{\lambda,\boldsymbol{w}^1,\boldsymbol{w}^2}$ instantiated as in Eq. (4.6), i.e., with $\boldsymbol{w}^1 = (1, \dots, 1)$ and $w^2_j = \frac{m-j+1}{m}$. This corresponds to a trade-off function between the sum of user utilities and a GGF for items with the

[5] http://pytorch.org
[6] https://github.com/scikit-learn/scikit-learn/blob/main/sklearn/isotonic.py
[7] https://www.last.fm/
[8] Using the Python library Implicit: https://github.com/benfred/implicit (MIT License).

Gini index weights, where the trade-off is controlled by varying $\lambda \in (0, 1)$. We remind though that unlike the standard Gini index, the GGF is un-normalized (see eq. (4.4), Sec 4.2.2).

We use three baselines for this task.

First, since the Gini index is non-differentiable, [Do et al., 2021c] proposed a differentiable surrogate using the standard deviation (std) instead, which we refer to as *eq. exposure*:

$$F^{\text{eq}}(P) = \sum_{i=1}^{n} u_i(P) - \frac{\lambda}{m}\sqrt{\sum_{j=1}^{m}\left(v_j(P) - \frac{1}{m}\sum_{j'=1}^{m} v_{j'}(P)\right)^2}$$

Second, Patro et al. [2020] address the trade-off of Task 1, since they compare various recommendation strategies based on the utility of users and the Lorenz curves of items (see [Patro et al., 2020, Fig. 1]), recalling that the standard Gini index is often defined as $1 - 2A$ where A is the area under the Lorenz curve [Yitzhaki and Schechtman, 2013]. Their fairness constraints are slightly different though, as their algorithm *FairRec*[9] guarantees envy-freeness for users, and a minimum exposure of $\frac{\lambda n \|\boldsymbol{b}\|}{m}$ for every item, where λ is the user-item tradeoff parameter.

Finally, we use the additive welfare function (4.8) (refered to as *welf*) with the recommended values $\alpha_1 \in \{-2, 0, 1\}$ and $\alpha_2 = 0$ [Do et al., 2021c], and varying $\lambda \in (0, 1)$ as third baseline. We only report the result of $\alpha_1 = 1$ since it obtained overall better performances on this task.

Task 2: Two-sided fairness We consider trade-offs between the cumulative utility of the q fraction of worst-off users, where $q \in \{0.25, 0.5\}$, and inequality between items measured by the Gini index, as in [Do et al., 2021c]. For this task, we instantiate the two-sided GGF $F_{\lambda,\boldsymbol{w^1},\boldsymbol{w^2}}$ as follows: the GGF for users is given by Eq. (4.7) with parameters (q, ω) in $\{0.25, 0.5\} \times \{0.25, 0.5, 1\}$, and the GGF for items uses the Gini index weights $w_j = \frac{m-j+1}{m}$. We generate trade-offs between user fairness and item fairness by varying $\lambda \in (0, 1)$.

The baseline approach for this task is *welf*, the additive welfare function (4.8), still with the recommended values $\alpha_1 \in \{-2, 0, 1\}$ and $\alpha_2 = 0$ and varying $\lambda \in (0, 1)$. We only report the results of $\alpha_1 = -2$ as they obtained the best performances on this task.

4.4.2 Results

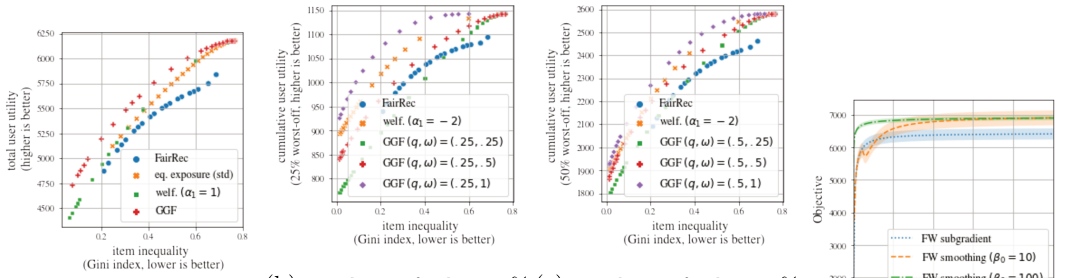
We now present experiments that illustrate the effectiveness of the two-sided GGF approach on Task 1 and 2.

For each fairness method, Pareto frontiers are generated by varying λ. Since Patro et al. [2020]'s algorithm *FairRec* does not scale, we compare to *FairRec* only on Lastfm-2k.

We optimize $F_{\lambda,\boldsymbol{w^1},\boldsymbol{w^2}}$ using FW-smoothing with $\beta_0 = 100$ and $T = 5k$ for Lastfm-2k, and $\beta_0 = 1000$ and $T = 50k$ for MovieLens. F^{welf} and F^{eq} are optimized with the Frank-Wolfe method of [Do et al., 2021c] for $T = 1k$ and $T = 5k$ iterations respectively for Lastfm-2k and MovieLens. This is the number of iterations recommended by [Do et al., 2021c], while we need more interactions for FW-smoothing because its convergence is $O(\frac{1}{\sqrt{T}})$ rather than $O(\frac{1}{T})$ because of non-smoothness.

We first focus on Lastfm-2k. On Task 1, *Fig. 4.1a*, the GGF (red + curve) obtains the best trade-off between total utility of users and Gini inequality between items, compared to *FairRec* and *eq. exposure*. It fares better than *eq. exposure* (orange ×) on this task because *eq. exposure* reduces inequality between items by minimizing the std of exposures, while GGF with weights $w_j^2 = \frac{m-j+1}{m}$

[9] [Patro et al., 2020] consider unordered recommendation lists with a uniform attention model. We transform them into ordered lists using the order output by *FairRec*, and adapt the item-side criterion of minimal exposure to the position-based model.

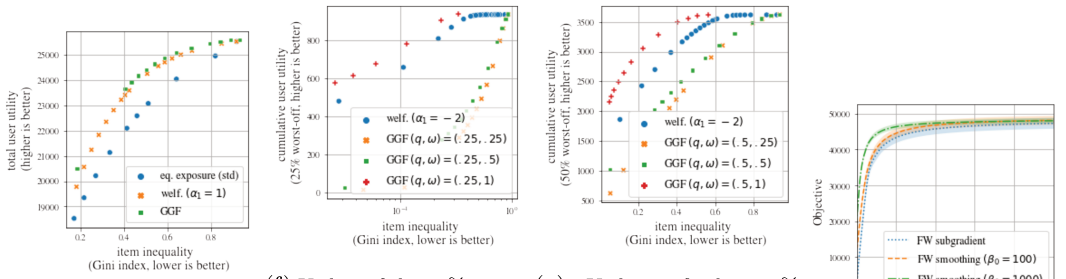


(a) Total user utility vs. inequality between items. GGF is instantiated as (4.6).

(b) Utility of the 25% worse-off users vs. inequality between items. GGF is instantiated as (4.7) with parameters (q, ω).

(c) Utility of the 50% worse-off users vs. inequality between items. GGF is instantiated as (4.7) with parameters (q, ω).

(d) Optimization of GGF (4.6) with $\lambda = 0.5$



(e) Total user utility vs. inequality between items. GGF is instantiated as (4.6).

(f) Utility of the 25% worse-off users vs. inequality between items. GGF is instantiated as (4.7) with parameters (q, ω).

(g) Utility of the 50% worse-off users vs. inequality between items. GGF is instantiated as (4.7) with parameters (q, ω).

(h) Optimization of GGF (4.6) with $\lambda = 0.5$

Figure 4.1: Summary of the results on Lastfm-2k (top row) and MovieLens (bottom row). (Left 3 columns): Trade-offs achieved by competing methods on various fairness criteria, when varying $\lambda \in [0, 1]$. (Right column): Convergence of FW-subgradient compared to FW-smoothing, for various values of β_0. FW-subgradient is not guaranteed to converge to an optimum.

minimizes the Gini index. This shows that Task 1 can be addressed by directly optimizing the Gini index, thanks to GGFs with FW-smoothing.

For Task 2, *Fig. 4.1b and 4.1c* depicts the trade-offs achieved between the utility of the 25% / 50% worst-off users and inequality among items. First, we observe that the more weight ω is put on the q-% worst-off in GGF, the higher the curve, which is why we observe the ordering green $\square$ $\prec$ red $+$ $\prec$ purple $\diamond$ for GGF, on both 25% and 50% trade-offs plots. Second, as expected, *welf* is outperformed by our two-sided GGF with instantiation (4.7) and $(q, \omega) = (q, 1)$ (purple $\diamond$), since it corresponds to the optimal settings for this task.

Figures 4.1e,4.1f,4.1g illustrates the same trade-offs on MovieLens. Results are qualitatively similar: by adequately parameterizing GGFs, we obtain the best guarantees on each fairness task.

Overall, these results show that even though the baseline approaches obtain non-trivial performances on the two fairness tasks above, the direct optimization of the trade-offs involving the Gini index or points of the Lorenz curves, which is possible thanks to our algorithm, yields significant performance gains. Moreover, we reiterate that these two tasks are only examples of fairness criteria that GGFs can formalize, since by varying the weights we can obtain all Lorenz-efficient rankings (Prop. 6).

4.4.3 Convergence diagnostics

We now demonstrate the usefulness of FW-smoothing for optimizing GGF objectives, compared to simply using the Frank-Wolfe method of [Do et al., 2021c] with a subgradient of the GGF (FW-subgradient). We note that a subgradient of $g_{\boldsymbol{w}}(\boldsymbol{x})$ is given by $\boldsymbol{w}_{\boldsymbol{\sigma}^{-1}}$, where $\sigma \in \text{argsort}(-\boldsymbol{x})$.

More precisely, FW-subgradient is also equivalent to using subgradients of $-g_{\boldsymbol{w}^1}$ and $-g_{\boldsymbol{w}^2}$ in Line 3 of Alg. 2, instead of $\nabla f^{\beta_t}(P^{(t)})$, ignoring the smoothing parameters β_t. FW-subgradient is simpler than FW-smoothing, but it is not guaranteed to converge [Nesterov, 2018]. The goal of this section is to assess whether the smoothing is necessary in practice.

We focus on the two-sided GGF (4.6) of Task 1 on Lastfm-2k and MovieLens, using FW-subgradient and FW-smoothing with different values of β_0. Figure 4.1d depicts the objective value as a function of the number of iterations, averaged over three seeds (the colored bands represent the std), on Lastfm-2k. We observe that FW-subgradient (blue dotted curve) plateaus at a suboptimum. In contrast, FW-smoothing converges (orange dotted and green dash-dot curves), and the convergence is faster for larger β_0. On MovieLens (Fig 4.1h), FW-subgradient converges to the optimal solution, but it is still slower than FW-smoothing with $\beta_0 = 1000$.

In conclusion, even though FW-subgradient reaches the optimal performance on Movielens for this set of parameters, it is still possible that FW-subgradient plateaus at significantly suboptimal solutions. The use of smoothing is thus not only necessary for theoretical convergence guarantees, but also in practice. In addition, FW-smoothing has comparable computational complexity to FW-subgradient since the computation cost is dominated by the sort operations in Alg. 2.

4.5 Reciprocal recommendation

4.5.1 Extension of the framework and algorithm

We show that our whole method for fair ranking readily applies to reciprocal recommendation tasks, such as the recommendation of friends or dating partners, or in job search platforms.

Reciprocal recommendation framework The recommendation framework we discussed thus far depicted *"one-sided" recommendation*, in the sense that only items are being recommended. In *reciprocal recommendation* problems [Palomares et al., 2021], users are also items who can be recommended to other users (the item *per se* is the user's profile or CV), and they have preferences over other users.

In this setting, $n = m$ and μ_{ij} denotes the mutual preference value between i and j (e.g., the probability of a "match" between i and j). Following [Do et al., 2021c], we extend our previous framework to reciprocal recommendation by introducing the *two-sided utility* of a user i, which sums the utility $\overline{u}_i(P)$ derived by i from the recommendations it gets, and the utility $\overline{v}_i(P)$ from being recommended to other users:

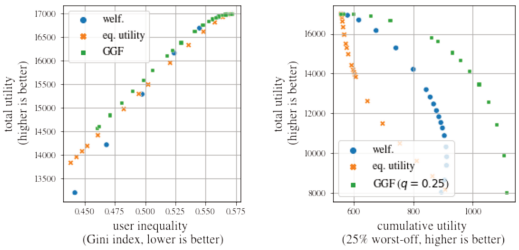
$$u_i(P) = \overline{u}_i(P) + \overline{v}_i(P) = \sum_{i,j} (\mu_{ij} + \mu_{ji}) P_{ij}^{\intercal} \boldsymbol{b}$$

$$\text{where} \quad \overline{u}_i(P) = \sum_{j=1}^{n} \mu_{ij} P_{ij}^{\intercal} \boldsymbol{b} \quad \text{and} \quad \overline{v}_i(P) = \sum_{j=1}^{n} \mu_{ij} P_{ji}^{\intercal} \boldsymbol{b}\,.$$

Objective and optimization The two-sided GGF objective (4.3) in reciprocal recommendation simply becomes one GGF of two-sided utilities, and it is specified by a single weighting vector $\boldsymbol{w}$:

$$\max_{P \in \mathcal{P}} \{F_{\boldsymbol{w}}(P) := g_{\boldsymbol{w}}(\boldsymbol{u}(P))\}. \tag{4.11}$$

The choice of $\boldsymbol{w}$ controls the degree of priority to the worse-off in the user population. We show in our experiments in Section 4.5.2 that in reciprocal recommendation too, the GGF objective can be adequately parameterized to address existing fairness criteria.



(a) Total utility vs. inequality. GGF is instantiated as (4.12) and we vary λ.

(b) Total utility vs. utility of the 25% worse-off. GGF is instantiated as (4.7) with $q = 0.25$ and varying ω.

Figure 4.2: Fairness trade-offs achieved by competing methods on reciprocal recommendation on Twitter. They are generated by varying λ for *eq. utility* and GGF, and α for *welf.*

$F_{\boldsymbol{w}}$ can be optimized using our algorithm FW-smoothing. Since there is only one GGF, the subroutine Alg. 1 is simply used once per iteration to project onto $\mathcal{C}(\tilde{\boldsymbol{w}})$, and obtain $\boldsymbol{y} = \Pi_{\mathcal{C}(\tilde{\boldsymbol{w}})}\left(\frac{\boldsymbol{u}(P^{(t-1)})}{\beta_t}\right)$ in Line 3 of Algorithm 2. Line 5 becomes $\tilde{\mu}_{ij} = (1-\lambda)y_i\mu_{ij} + \lambda y_j\mu_{ji}$.

Our method for fair ranking is thus general enough to address both one-sided and reciprocal recommendation, using the notion of two-sided utility from Do et al. [2021c].

4.5.2 Experiments

Similarly to our experiments in Sec. 4.4.2, the goal of these experiments is to demonstrate that in reciprocal recommendation, GGF can be parameterized to exactly optimize for existing fairness criteria, outperforming previous approaches designed to address them.

Data We reproduce the experimental setting of [Do et al., 2021c] who also study fairness in reciprocal recommendation. We simulate a friend recommendation task from the Higgs **Twitter** dataset [De Domenico et al., 2013], which contains directed follower links on the social network Twitter. We consider a mutual follow as a "match", and we keep users having at least 20 matches, resulting in a subset of $13k$ users. We estimate mutual scores μ_{ij} (i.e., match probabilities) by matrix factorization.

4.5.2.1 Fairness criteria

Similarly to Section 4.4.1.2, we state two fairness tasks that previously appeared in the literature, we instantiate the GGF objective $F_{\boldsymbol{w}}$ for each task and describe existing baselines.

Task 1: Trade-offs between total utility and inequality of utility among users Although reciprocal recommendation received less attention in the fairness literature, an existing requirement is to mitigate inequalities in utility between users [Jia et al., 2018, Basu et al., 2020], similarly to the *eq. exposure* criterion in one-sided recommendation. This leads to a trade-off between the sum of utilities and inequality typically measured by the Gini index. For Task 1, we use the GGF (4.11) with $w_i = (1-\lambda) + \lambda \cdot \frac{n-i+1}{n}$, which yields a trade-off function between the sum of utilities and inequality of utilities, and we vary λ in $(0, 1)$ to generate such trade-offs:

$$F_{\boldsymbol{w}}(P) = (1-\lambda)\sum_{i=1}^{n} u_i(P) + \lambda \sum_{i=1}^{n} \frac{n-i+1}{n} u_i^{\uparrow}(P). \tag{4.12}$$

We use two baselines for this task.

First, similarly to *eq. exposure*, to bypass the nonsmoothness of the Gini index, [Do et al., 2021c] optimize a surrogate with std, named *eq. utility*:

$$F^{\text{eq}}(P) = \sum_{i=1}^{n} u_i(P) - \frac{\lambda}{n}\sqrt{\sum_{i=1}^{n}\left(u_i(P) - \frac{1}{n}\sum_{i'=1}^{n} u_{i'}(P)\right)^2}.$$

Second, the welfare function *welf* (4.8) of [Do et al., 2021c] is used in reciprocal recommendation as a single sum: $F^{welf}(P) = \sum_{i=1}^{n} \phi(u_i(P), \alpha)$ where ϕ is defined in Sec. 4.2.3. We study *welf* as baseline by varying α, which controls the redistribution of utility in the user population.

Task 2: Trade-offs between total utility and utility of the worse-off The main task studied by [Do et al., 2021c] with *welf* is to trade-off between the total utility and the cumulative utility of the q fraction of worse-off users. For this task, we instantiate the GGF with (4.7), with fixed quantile $q = 0.25$ and we vary ω to generate trade-offs between total utility and cumulative utility of the 25% worst-off.

We compare it to the *welf* baseline where α is varied as in [Do et al., 2021c].

4.5.2.2 Fairness trade-offs results

Results We now demonstrate that in reciprocal recommendation too, GGF is the most effective approach in addressing existing fairness criteria. We optimize the GGF $F_{\boldsymbol{w}}(P)$ using FW-smoothing with $\beta_0 = 10$ for $T = 50k$ iterations, and optimize F^{welf} and F^{eq} using Frank-Wolfe for $T = 5k$ iterations.

Figure 4.2 depicts the trade-offs obtained by the competing approaches on the fairness tasks 1 and 2, on the Twitter dataset. Fig. 4.2a illustrates the superiority of GGF (green □) on Task 1, despite good performance of the baselines *eq. utility* (orange ×) and *welf* (blue ○). As in one-sided recommendation with *eq. exposure*, the reason why *eq. utility* achieves slightly worse trade-offs on this fairness task is because it minimizes the std as a surrogate to the Gini index, instead of the Gini index itself as GGF does. For Task 2, on Fig.4.2b, we observe that GGF with parameterization (4.7) (green □) is the most effective. This is because unlike the *welf* approach (blue ○) of Do et al. [2021c] who address this fairness task, this form of GGF is exactly designed to optimize for utility quantiles.

4.6 Related work

Algorithmic fairness Fairness in ranking and recommendation systems is an active area of research. Since recommender systems involve multiple stakeholders [Burke, 2017, Abdollahpouri et al., 2020], fairness has been considered from the perspective of both users and item producers. On the user side, a common goal is to prevent disparities in recommendation performance across sensitive groups of users [Mehrotra et al., 2017, Ekstrand et al., 2018]. On the item side, authors aim to prevent winner-take-all effects [Abdollahpouri et al., 2019b] by redistributing exposure across groups of producers, either towards equal exposure, or equal ratios of exposure to relevance [Singh and Joachims, 2018, Biega et al., 2018, Diaz et al., 2020, Kletti et al., 2022a], sometimes measured by the classical Gini index [Morik et al., 2020, Wilkie and Azzopardi, 2014].

Some authors consider fairness for both users and items, often by applying existing user or item criteria simultaneously to both sides, such as [Basu et al., 2020, Wu et al., 2021b, Wang and Joachims, 2021]. [Patro et al., 2020, Do et al., 2022a] instead discuss two-sided fairness with

envy-freeness as user-side criterion, while [Deldjoo et al., 2021] propose to use generalized cross entropy to measure unfairness among sensitive groups of users and items. [Wu et al., 2021a] recently considered two-sided fairness in recommendation as a multi-objective problem, where each objective corresponds to a different fairness notion, either for users or items. Similarly, Mehrotra et al. [2020] aggregate multiple recommendation objectives using a GGF, in a contextual bandit setting. In their case, the aggregated objectives represent various metrics (e.g., clicks, dwell time) for various stakeholders. Unlike these two works [Wu et al., 2021a, Mehrotra et al., 2020], in our case the multiple objectives are the individual utilities of each user and item, and our goal is to be fair towards each entity by redistributing utility. To our knowledge, we are the first to use GGFs as *welfare functions* of users' and items' utilities for two-sided fairness in rankings.

Reciprocal recommender systems received comparatively less attention in the fairness literature, to the exception of [Jia et al., 2018, Xia et al., 2015, Paraschakis and Nilsson, 2020]. The closest to our work is the additive welfare approach of [Do et al., 2021c], which addresses fairness in both one-sided and reciprocal recommendation, and is extensively discussed in the paper, see Sec. 4.2.1.

In the broader fair machine learning community, several authors advocated for economic concepts [Finocchiaro et al., 2020], using inequality indices to quantify and mitigate unfairness [**?**Speicher et al., 2018, Heidari et al., 2018, Lazovich et al., 2022], taking an axiomatic perspective [Gölz et al., 2019, Cousins, 2021, Williamson and Menon, 2019] or applying welfare economics principles [Hu and Chen, 2020, Rambachan et al., 2020]. GGFs, in particular, were recently applied to fair multi-agent reinforcement learning, with multiple reward functions [Busa-Fekete et al., 2017, Siddique et al., 2020, Zimmer et al., 2021]. These works consider sequential decision-making problems without ranking, and their GGFs aggregate the objectives of a few agents (typically $n < 20$), while in our ranking problem, there are as many objectives as there are users and items.

Nonsmooth convex optimization and differentiable ranking Our work builds on nonsmooth convex optimization methods [Nesterov, 2005, Shamir and Zhang, 2013], and in particular variants of the Frank-Wolfe algorithm [Frank and Wolfe, 1956, Jaggi, 2013] for nonsmooth problems [Lan, 2013, Yurtsever et al., 2018, Ravi et al., 2019, Thekumparampil et al., 2020a]. The recent algorithm of [Thekumparampil et al., 2020a] is a Frank-Wolfe variant which uses the Moreau envelope like us. Its number of first-order calls is optimal, but this is at the cost of a more complex algorithm with inner loops that make it slow in practice. In our case, since the calculation of the gradient is not a bottleneck, we use the simpler algorithm of Lan [2013], which applies Frank-Wolfe to the Moreau envelope of the nonsmooth objective.

Our technical contribution is also related to the literature on differentiable ranking, which includes a large body of work on approximating learning-to-rank metrics [Chapelle and Wu, 2010, Taylor et al., 2008, Adams and Zemel, 2011], and recent growing interest in designing smooth ranking modules [Grover et al., 2019, Cuturi et al., 2019, Blondel et al., 2020] for end-to-end differentiation pipelines. The closest method to ours is the differentiable sorting operator of Blondel et al. [2020], which also relies on isotonic regression. The differences between our approaches are explained in Remark 1.

4.7 Conclusion

We proposed generalized Gini welfare functions as a flexible method to produce fair rankings. We addressed the challenges of optimizing these welfare functions by leveraging Frank-Wolfe methods for nonsmooth objectives, and demonstrated their efficiency in ranking applications.

Our framework and algorithm applies to both usual recommendation of movies or music, and to reciprocal recommendation scenarios, such as dating or hiring.

Generalized Gini welfare functions successfully address a large variety of fairness requirements for ranking algorithms. On the one hand, GGFs are effective in reducing inequalities, since they generalize the Gini index in economics. Optimizing them allows to meet the requirements of equal utility criteria, largely advocated by existing work on fair recommendation [Singh and Joachims, 2018, Basu et al., 2020, Patro et al., 2020, Wu et al., 2021b]. On the other hand, GGFs effectively increase the utility of the worse-off, which is usually measured by quantile ratios in economics, and has been recently considered as a fairness criterion in ranking [Do et al., 2021c].

Our approach is limited to fairness considerations at the stage of inference. It does not address potential biases arising at other parts of the recommendation pipeline, such as in the estimation of preferences. Moreover, we considered a static model, which does not accounts for real-world dynamics, such as responsiveness in two-sided markets [Su et al., 2021], feedback loops in the learning process [Bottou et al., 2013], and the changing nature of the users' and items' populations [Morik et al., 2020] and preferences [Kalimeris et al., 2021]. Addressing these limitations, in combination with our method, are interesting directions for future research.

Chapter 5

Fair ranking in the contextual bandit setting

Contents

This chapter is the article *Contextual bandits with concave rewards, and an application to fair ranking*, published at ICLR 2023 (see [Do et al., 2023]). In this chapter, we address fair ranking in the contextual bandit setting, which we first described in Section 1.4.1 of Chapter 1. In the contextual bandit setting, rankings are computed one at a time as the users request recommendations, and user preferences are learned online through sequential interactions. This setting is more practical than the batch setting, since it is more efficient to compute the ranking of the current user, instead of all the users in a large batch, as in the previous chapters.

In this paper, we address a more generic contextual bandit problem with multiple rewards, where the trade-off between the rewards is defined by a known concave function f. This bandit problem, called Contextual Bandits with Concave Rewards (CBCR), encompasses our fair ranking problem, but also the optimization of multiple metrics on online platforms. Our work provides the first general solution to CBCR that does not impose restrictions on the policy space. This was made possible through a novel use of theoretical analyses of Frank-Wolfe algorithms, which allowed us to prove a reduction to scalar-reward contextual bandits. Motivated by fairness in rankings, we show how CBCR applies to fairness-aware objectives for ranking in Section 5.4, and derive the first algorithm with regret guarantees for fair ranking in the contextual bandit setting.

In the application of CBCR to fair ranking (Section 5.4), we solely focus on ranking objectives with item-side fairness, while the previous chapters addressed two-sided fairness. More precisely, we only consider trade-offs between item-side fairness and average user utility, which is the average

of user rewards over contexts. Addressing user-side fairness would require explicitly encoding user identifiers in the context vector x_t and keeping track of user activity. This would require a lot of additional formalism just for the fair ranking application, at the cost of clarity for the rest of the paper which core matter is our solution to the general CBCR problem. For insights on how to integrate user fairness in an online setting, we refer to [Usunier et al., 2022] which addresses two-sided fairness in online ranking, without learning.

Abstract

We consider Contextual Bandits with Concave Rewards (CBCR), a multi-objective bandit problem where the desired trade-off between the rewards is defined by a known concave objective function, and the reward vector depends on an observed stochastic context. We present the first algorithm with provably vanishing regret for CBCR without restrictions on the policy space, whereas prior works were restricted to finite policy spaces or tabular representations. Our solution is based on a geometric interpretation of CBCR algorithms as optimization algorithms over the convex set of expected rewards spanned by all stochastic policies. Building on Frank-Wolfe analyses in constrained convex optimization, we derive a novel reduction from the CBCR regret to the regret of a *scalar-reward* bandit problem. We illustrate how to apply the reduction off-the-shelf to obtain algorithms for CBCR with both linear and general reward functions, in the case of non-combinatorial actions. Motivated by fairness in recommendation, we describe a special case of CBCR with rankings and fairness-aware objectives, leading to the first algorithm with regret guarantees for contextual combinatorial bandits with fairness of exposure.

5.1 Introduction

Contextual bandits are a popular paradigm for online recommender systems that learn to generate personalized recommendations from user feedback. These algorithms have been mostly developed to maximize a single scalar reward which measures recommendation performance for users. Recent fairness concerns have shifted the focus towards item producers whom are also impacted by the exposure they receive [Biega et al., 2018, Geyik et al., 2019], leading to optimize trade-offs between recommendation performance for users and fairness of exposure for items [Singh and Joachims, 2019, Zehlike and Castillo, 2020]. More generally, there is an increasing pressure to insist on the multi-objective nature of recommender systems [Vamplew et al., 2018, Stray et al., 2021], which need to optimize for several engagement metrics and account for multiple stakeholders' interests [Mehrotra et al., 2020, Abdollahpouri et al., 2019a]. In this paper, we focus on the problem of contextual bandits with multiple rewards, where the desired trade-off between the rewards is defined by a known concave objective function, which we refer to as *Contextual Bandits with Concave Rewards* (CBCR). Concave rewards are particularly relevant to fair recommendation, where several objectives can be expressed as (known) concave functions of the (unknown) utilities of users and items [Do et al., 2021c].

Our CBCR problem is an extension of Bandits with Concave Rewards (BCR) [Agrawal and Devanur, 2014] where the vector of multiple rewards depends on an observed stochastic context. We address this extension because contexts are necessary to model the user/item features required for personalized recommendation. Compared to BCR, the main challenge of CBCR is that optimal policies depend on the entire distribution of contexts and rewards. In BCR, optimal policies are distributions over actions, and are found by direct optimization in policy space [Agrawal and Devanur, 2014,

Berthet and Perchet, 2017]. In CBCR, stationary policies are mappings from a continuous context space to distributions over actions. This makes existing BCR approaches inapplicable to CBCR because the policy space is not amenable to tractable optimization without further assumptions or restrictions. As a matter of fact, the only prior theoretical work on CBCR is restricted to a finite policy set [Agrawal et al., 2016].

We present *the first algorithms with provably vanishing regret* for CBCR without restriction on the policy space. Our main theoretical result is a reduction where the CBCR regret of an algorithm is bounded by its regret on a proxy bandit task with *single* (scalar) reward. This reduction shows that it is straightforward to turn *any* contextual (scalar reward) bandits into algorithms for CBCR. We prove this reduction by first re-parameterizing CBCR as an optimization problem in the space of feasible rewards, and then revealing connections between Frank-Wolfe (FW) optimization in reward space and a decision problem in action space. This bypasses the challenges of optimization in policy space.

To illustrate how to apply the reduction, we provide two example algorithms for CBCR with non-combinatorial actions, one for linear rewards based on LinUCB [Abbasi-Yadkori et al., 2011], and one for *general reward functions* based on the SquareCB algorithm [Foster and Rakhlin, 2020] which uses online regression oracles. In particular, we highlight that our reduction can be used together with any exploration/exploitation principle, while previous FW approaches to BCR relied exclusively on upper confidence bounds [Agrawal and Devanur, 2014, Berthet and Perchet, 2017, Cheung, 2019].

Since fairness of exposure is our main motivation for CBCR, we show how our reduction also applies to the *combinatorial* task of fair ranking with contextual bandits, leading to the *first algorithm with regret guarantees* for this problem, and we show it is *computationally efficient*. We compare the empirical performance of our algorithm to relevant baselines on a music recommendation task.

Related work. Agrawal et al. [2016] address a restriction of CBCR to a finite set of policies, where explicit search is possible. Cheung [2019] use FW for reinforcement learning with concave rewards, a similar problem to CBCR. However, they rely on a tabular setting where there are few enough policies to compute them explicitly. Our approach is the only one to apply to CBCR without restriction on the policy space, by removing the need for explicit representation and search of optimal policies.

Our work is also related to fairness of exposure in bandits. Most previous works on this topic either do not consider rankings [Celis et al., 2018b, Wang et al., 2021a, Patil et al., 2020, Chen et al., 2020], or apply to combinatorial bandits without contexts [Xu et al., 2021]. Both these restrictions are impractical for recommender systems. Mansoury et al. [2021a], Jeunen and Goethals [2021] propose heuristics with experimental support that apply to both ranking and contexts in this space, but they lack theoretical guarantees. We present the first algorithm with regret guarantees for fair ranking with contextual bandits. We provide a more detailed discussion of the related work in Appendix B.1.

5.2 Maximization of concave rewards in contextual bandits

Notation. For any $n \in \mathbb{N}$, we denote by $[\![n]\!] = \{1, \dots, n\}$. The dot product of two vectors x and y in $\mathbb{R}^n$ is either denoted $x^\intercal y$ or using braket notation $\langle x \,|\, y \rangle$, depending on which one is more readable.

Setting. We define a stochastic contextual bandit [Langford and Zhang, 2007] problem with D rewards. At each time step t, the environment draws a context $x_t \sim P$, where $x \in \mathcal{X} \subseteq \mathbb{R}^q$

and P is a probability measure over $\mathcal{X}$. The learner chooses an action $a_t \in \mathcal{A}$ where $\mathcal{A} \subseteq \mathbb{R}^K$ is the action space, and receives a noisy multi-dimensional reward $r_t \in \mathbb{R}^D$, with expectation $\mathbb{E}[r_t|x_t, a_t] = \mu(x_t)a_t$, where $\mu : \mathcal{X} \to \mathbb{R}^{D\times K}$ is the matrix-value contextual expected reward function.[1] The trade-off between the D cumulative rewards is specified by a known concave function $f : \mathbb{R}^D \to \mathbb{R} \cup \{\pm\infty\}$. Let $\overline{\mathcal{A}}$ denote the convex hull of $\mathcal{A}$ and $\overline{\pi} : \mathcal{X} \to \overline{\mathcal{A}}$ be a stationary policy,[2] then the optimal value for the problem is defined as $f^* = \sup_{\overline{\pi}:\mathcal{X}\to\overline{\mathcal{A}}} f\Big(\mathbb{E}_{x\sim P}\big[\mu(x)\overline{\pi}(x)\big]\Big)$.

We rely on either of the following assumptions on f:

Assumption A. *f is closed proper concave[3] on $\mathbb{R}^D$ and $\mathcal{A}$ is a compact subset of $\mathbb{R}^K$. Moreover, there is a compact convex set $\mathcal{K} \subseteq \mathbb{R}^D$ such that*

- (Bounded rewards) $\forall (x, a) \in \mathcal{X} \times \mathcal{A}, \mu(x)a \in \mathcal{K}$ *and for all* $t \in \mathbb{N}_*$, $r_t \in \mathcal{K}$ *with probability* 1.
- (Local Lipschitzness) *f is L-Lipschitz continuous with respect to $\|.\|_2$ on an open set containing $\mathcal{K}$.*

Assumption B. *Assumption A holds and f has C-Lipschitz-continuous gradients w.r.t. $\|.\|_2$ on $\mathcal{K}$.*

The most general version of our algorithm, described in Appendix B.4, removes the need for the smoothness assumption using smoothing techniques. We describe an example in Section 5.3.3. In the rest of the paper, we denote by $D_{\mathcal{K}} = \sup_{z,z'\in\mathcal{K}} \|z - z'\|_2$ the diameter of $\mathcal{K}$, and use $\tilde{C} = \frac{C}{2}D_{\mathcal{K}}^2$.

We now give two examples of this problem setting, motivated by real-world applications in recommender systems, and which satisfy Assumption A.

Example 3 (Optimizing multiple metrics in recommender systems.)**.** *Mehrotra et al. [2020] formalized the problem of optimizing D engagement metrics (e.g. clicks, streaming time) in a bandit-based recommender system. At each t, x_t represents the current user's features. The system chooses one arm among K, represented by a vector a_t in the canonical basis of $\mathbb{R}^K$ which is the action space $\mathcal{A}$. Each entry of the observed reward vector $(r_{t,i})_{i=1}^D$ corresponds to a metric's value. The trade-off between the metrics is defined by the Generalized Gini Function: $f(z) = \sum_{i=1}^D w_i z_i^{\uparrow}$, where $(z_i^{\uparrow})_{i=1}^D$ denotes the values of z sorted increasingly and $w \in \mathbb{R}^D$ is a vector of non-increasing weights.*

Example 4 (Fairness of exposure in rankings.)**.** *The goal is to balance the traditional objective of maximizing user satisfaction in recommender systems and the inequality of exposure between item producers [Singh and Joachims, 2018, Zehlike and Castillo, 2020]. For a recommendation task with m items to rank, this leads to a problem with $D = m + 1$ objectives, which correspond to the m items' exposures, plus the user satisfaction metric. The context $x_t \in \mathcal{X} \subset \mathbb{R}^{md}$ is a matrix where each $x_{t,i} \in \mathbb{R}^d$ represents a feature vector of item i for the current user. The action space $\mathcal{A}$ is combinatorial, i.e. it is the space of rankings represented by permutation matrices:*

$$\mathcal{A} = \Big\{a \in \{0,1\}^{m\times m} : \forall i \in [\![m]\!], \sum_{k=1}^{m} a_{i,k} = 1 \text{ and } \forall k \in [\![m]\!], \sum_{i=1}^{m} a_{i,k} = 1\Big\} \tag{5.1}$$

For $a \in \mathcal{A}$, $a_{i,k} = 1$ if item i is at rank k. Even though we use a double-index notation and call a a permutation matrix*, we flatten a as a vector of dimension $K = m^2$ for consistency of notation.*

[1] Notice that linear structure between $\mu(x_t)$ and a_t is standard in combinatorial bandits [Cesa-Bianchi and Lugosi, 2012] and it reduces to the usual multi-armed bandit setting when $\mathcal{A}$ is the canonical basis of $\mathbb{R}^K$.

[2] In the multi-armed setting, stationary policies return a distribution over arms given a context vector. In the combinatorial setup, $\overline{\pi}(x) \in \overline{\mathcal{A}}$ is the average feature vector of a stochastic policy over $\mathcal{A}$. For the benchmark, we are only interested in expected rewards so there is to need to specify the full distribution over $\mathcal{A}$.

[3] This means that f is concave and upper semi-continuous, is never equal to $+\infty$ and is finite somewhere.

We now give a concrete example for f, which is concave as usual for objective functions in fairness of exposure [Do et al., 2021c]. It is inspired by Morik et al. [2020], who study trade-offs between average user utility and inequality[4] of item exposure:

$$f(z) = \underbrace{z_{m+1}}_{\text{user utility}} - \beta \underbrace{\frac{1}{2m}\sum_{i=1}^{m}\sum_{j=1}^{m}|z_i - z_j|}_{\text{inequality of item exposure}} \qquad \text{where } \beta > 0 \text{ is a trade-off parameter.} \quad (5.2)$$

The learning problem. In the bandit setting, P and μ are unknown and the learner can only interact online with the environment.Let $h_T = \left(x_t, a_t, r_t\right)_{t\in[\![T-1]\!]}$ be the history of contexts, actions, and reward observed up to time $T-1$ and $\delta' > 0$ be a confidence level, then at step t a bandit algorithm $\mathfrak{A}$ receives in input the history h_t, the current context x_t, and it returns a distribution over actions $\mathcal{A}$ and selects an action $a_t \sim \mathfrak{A}(h_t, x_t, \delta')$. The objective of the algorithm is to minimize the regret

$$R_T = f^* - f(\hat{s}_T) \qquad \text{where } \hat{s}_T = \frac{1}{T}\sum_{t=1}^{T} r_t.$$

Note that our setting subsumes classical stochastic contextual bandits: when $D = 1$ and $f(z) = z$, maximizing $f(\hat{s}_T)$ amounts to maximizing a cumulative scalar reward $\sum_{t=1}^{T} r_t$. In Lem. 32 (App. B.3.3), we show that alternative definitions of regret, with different choices of comparator or performance measure, would yield a difference of order $O(1/\sqrt{T})$, and hence not substantially change our results.

5.3 A general reduction-based approach for CBCR

In this section we describe our general approach for CBCR. We first derive our key reduction from CBCR to a specific scalar-reward bandit problem. We then instantiate our algorithm to the case of linear and general reward functions for smooth objectives f. Finally, we extend to the case of non-smooth objective functions using Moreau-Yosida regularization [Rockafellar and Wets, 2009].

5.3.1 Reduction from CBCR to scalar-reward contextual bandits

There are two challenges in the CBCR problem: 1) the computation of the optimal policy $\sup_{\overline{\pi}:\mathcal{X}\to\overline{\mathcal{A}}} f\Big(\mathbb{E}_{x\sim P}\big[\mu(x)\overline{\pi}(x)\big]\Big)$ even with known μ; 2) the learning problem when μ is unknown.

1: Reparameterization of the optimization problem. The first challenge is that optimizing directly in policy space for the benchmark problem $\sup_{\overline{\pi}:\mathcal{X}\to\overline{\mathcal{A}}} f\Big(\mathbb{E}_{x\sim P}\big[\mu(x)\overline{\pi}(x)\big]\Big)$ is intractable without any restriction, because the policy space includes all mappings from the continuous context space $\mathcal{X}$ to distributions over actions. Our solution is to rewrite the optimization problem as a standard convex constrained problem by introducing the convex set $\mathcal{S}$ of feasible rewards:

$$\mathcal{S} = \left\{\mathbb{E}_{x\sim P}\big[\mu(x)\overline{\pi}(x)\big] \,\middle|\, \overline{\pi} : \mathcal{X}\to\overline{\mathcal{A}}\right\} \text{ so that } f^* = \sup_{\overline{\pi}:\mathcal{X}\to\overline{\mathcal{A}}} f\Big(\mathbb{E}_{x\sim P}\big[\mu(x)\overline{\pi}(x)\big]\Big) = \max_{s\in\mathcal{S}} f(s).$$

Under Assumption A, $\mathcal{S}$ is a compact subset of $\mathcal{K}$ (see Lemma 30 in App. B.3) so f attains its maximum over $\mathcal{S}$. We have thus reduced the complex initial optimization problem to a concave optimization problem over a compact convex set.

[4] $\mathrm{Gini}(z_1,\dots,z_m) = \frac{1}{2m}\sum_{i=1}^{m}\sum_{j=1}^{m}|z_i - z_j|$ is an unnormalized Gini coefficient.

2: Reducing the learning problem to scalar-reward bandits. Unfortunately, since P and μ are unknown, the set $\mathcal{S}$ is unknown. This precludes the possibility of directly using standard constrained optimization techniques, including gradient descent with projections onto $\mathcal{S}$. We consider Frank-Wolfe, a projection-free optimization method robust to approximate gradients [Lacoste-Julien et al., 2013, Kerdreux et al., 2018]. At each iteration t of FW, the update direction is given by the linear subproblem: $\operatorname{argmax}_{s\in\mathcal{S}}\langle \nabla f(z_{t-1}) \,|\, s\rangle$, where z_{t-1} is the current iterate. Our main technical tool, Lemma 12, allows to connect the FW subproblem in the unknown reward space $\mathcal{S}$ to a workable decision problem in the action space (see Lemma 36 in Appendix B.5 for a proof):

Lemma 12. *Let $\mathbb{E}_t[.]$ be the expectation conditional on h_t. Let $z_t \in \mathcal{K}$ be a function of contexts, actions and rewards up to time t. Under Assumption A, we have:*

$$\forall t \in \mathbb{N}_*, \mathbb{E}_t\Big[\max_{a\in\mathcal{A}} \langle \nabla f(z_{t-1}) \,|\, \mu(x_t)a\rangle\Big] = \max_{s\in\mathcal{S}} \langle \nabla f(z_{t-1}) \,|\, s\rangle.$$

For all $\delta \in (0,1]$, with probability at least $1-\delta$, we have:

$$\sum_{t=1}^{T} \Big(\max_{s\in\mathcal{S}} \langle \nabla f(z_{t-1}) \,|\, s\rangle - \max_{a\in\mathcal{A}} \langle \nabla f(z_{t-1}) \,|\, \mu(x_t)a\rangle \Big) \leq LD_{\mathcal{K}} \sqrt{2T\ln(\delta^{-1})}.$$

Lemma 12 shows that FW for CBCR operates closely to a sequence of decision problems of the form $(\max_{a\in\mathcal{A}} \langle \nabla f(z_{t-1}) \,|\, \mu(x_t)a\rangle)_{t=1}^T$. However, we have yet to address the problem that P and μ are unknown. To solve this issue, we introduce a **reduction to scalar-reward contextual bandits**. We can notice that solving for the sequence of actions maximizing $\sum_{t=1}^T \langle \nabla f(z_{t-1}) \,|\, \mu(x_t)a\rangle$ corresponds to solving a contextual bandit problem with adversarial contexts and stochastic rewards. Formally, using $z_t = \hat{s}_t$[5], we define the extended context $\tilde{x}_t = (\nabla f(\hat{s}_{t-1}), x_t)$, the average scalar reward $\tilde{\mu}(\tilde{x}_t) = \nabla f(\hat{s}_{t-1})^\intercal \mu(x_t)$ and the observed scalar reward $\tilde{r}_t = \langle \nabla f(\hat{s}_{t-1}) \,|\, r_t\rangle$. This fully defines a contextual bandit problem with *scalar reward*. Then, the objective of the algorithm is to minimize the following *scalar regret*:

$$R_T^{\text{scal}} = \sum_{t=1}^{T} \max_{a\in\mathcal{A}} \tilde{\mu}(\tilde{x}_t)^\intercal a - \sum_{t=1}^{T} \tilde{r}_t = \sum_{t=1}^{T} \max_{a\in\mathcal{A}} \langle \nabla f(\hat{s}_{t-1}) \,|\, \mu(x_t)a\rangle - \sum_{t=1}^{T} \langle \nabla f(\hat{s}_{t-1}) \,|\, r_t\rangle. \quad (5.3)$$

In this framework, the only information observed by the learning algorithm is $\tilde{h}_t := \big(\tilde{x}_{t'}, a_{t'}, \tilde{r}_{t'}\big)_{t'\in[\![t-1]\!]}$. This regret minimization problem has been extensively studied [see e.g., Slivkins, 2019, Chap. 8 for an overview]. The following key reduction result[6] relates R_T^{scal} to R_T, the regret of the original CBCR problem:

Theorem 13. *Under Assmpt. B, for every $T \in \mathbb{N}_*$ and $\delta > 0$, algorithm $\mathfrak{A}$ satisfies, with prob. $\geq 1-\delta$:*

$$R_T = f^* - f(\hat{s}_T) \leq \frac{R_T^{\text{scal}} + LD_{\mathcal{K}}\sqrt{2T\ln(1/\delta)} + \tilde{C}\ln(eT)}{T}.$$

The reduction shown in Thm. 13 hints us at how to use or adapt scalar bandit algorithms for CBCR. In particular, any algorithm with sublinear regret will lead to a vanishing regret for CBCR. Since the worst-case regret of contextual bandits is $\Omega(\sqrt{T})$ [Dani et al., 2008], we obtain

[5] For simplicity, we presented our reduction with $z_t = \hat{s}_t$ but other choices of z_t are possible (see Appendix B.4). The important point is that the reduction works without restricting z_t to $\mathcal{S}$.

[6] In practice, this result is used in conjunction with an upper bound $\overline{R}^{\text{scal}}(T,\delta')$ on R_T^{scal} that holds with probability $\geq 1-\delta'$, which gives $R_T \leq \overline{R}^{\text{scal}}(T,\delta')/T + O(\sqrt{\ln(1/\delta)/T})$ with probability at least $1-\delta-\delta'$ using the union bound.

near minimax optimal algorithms for CBCR. We illustrate this with two algorithms derived from our reduction in Sec. 5.3.2.

Proof sketch of Theorem 13: CBCR and Frank-Wolfe algorithms (full proof in Appendix B.5). Although the set $\mathcal{S}$ is not known, the standard telescoping sum argument for the analysis of Frank-Wolfe algorithms (see Lemma 37 in Appendix B.5, and e.g., [Berthet and Perchet, 2017, Lemma 12] for similar derivations) gives that under Assumption B, denoting $g_t = \nabla f(\hat{s}_{t-1})$:

$$TR_T \leq \sum_{t=1}^{T} \max_{s\in\mathcal{S}} \langle g_t \,|\, s - r_t\rangle + \tilde{C}\ln(eT).$$

The result is true for every sequence $(r_t)_{t\in[\![T]\!]} \in \mathcal{K}^T$, and only tracks the trajectory of $\hat{s}_t$ in reward space. We introduce now the reference of the scalar regret:

$$TR_T = \sum_{t=1}^{T} \big(\max_{s\in\mathcal{S}} \langle g_t \,|\, s\rangle - \max_{a\in\mathcal{A}} \langle g_t \,|\, \mu(x_t)a\rangle\big) + \underbrace{\sum_{t=1}^{T} \max_{a\in\mathcal{A}} \langle g_t \,|\, \mu(x_t)a - r_t\rangle}_{=R_T^{\text{scal}}} + \tilde{C}\ln(eT) \tag{5.4}$$

Lemma 12 bounds the leftmost term, from which Theorem 13 immediately follows using (5.4). □

5.3.2 Practical application: Two algorithms for multi-armed CBCR

To illustrate the effectiveness of the reduction from CBCR to scalar-reward bandits, we focus on the case where the action space $\mathcal{A}$ is the canonical basis of $\mathbb{R}^K$ (as in Example 3). We first study the case of linear rewards. Then, for general reward functions, we introduce the FW-SquareCB algorithm, the first example of a FW-based approach combined with an exploration principle other than optimism. This shows our approach has a much broader applicability to solve (C)BCR than previous strategies.

From LinUCB to FW-LinUCB (details in Appendix B.7). We consider a CBCR with linear reward function, i.e., $\mu(x) = \theta x$ where $\theta \in \mathbb{R}^{D\times d}$ (recall we have D rewards) and $x \in \mathbb{R}^{d\times K}$, where d is the number of features. Let $\tilde{\theta} := \text{flatten}(\theta)$ and $g_t = \nabla f(\hat{s}_{t-1})$. Using $[.;.]$ to denote the vertical concatenation of matrices, the expected reward for action a in context x at time t can be written $\langle g_t \,|\, \mu(x)a\rangle = g_t^\intercal \theta x a = \langle \tilde{\theta} \,|\, \tilde{x}_t a\rangle$ where $\tilde{x}_t \in \mathbb{R}^{Dd\times K}$ is the *extended* context with entries $\tilde{x}_t = [g_{t,0}x_t; \ldots; g_{t,D}x_t] \in \mathbb{R}^{Dd\times K}$. This is an instance of a linear bandit problem, where at each time t, action a is associated to the vector $\tilde{x}_t a$ and its expected reward is $\langle \tilde{\theta} \,|\, \tilde{x}_t a\rangle$. As a result, we can immediately derive a LinUCB-based algorithm for linear CBCR by leveraging the equivalence $\text{FW-LinUCB}(h_t, x_t, \delta') = \text{LinUCB}(\tilde{h}_t, \tilde{x}_t, \delta')$. LinUCB's regret guarantees imply $R_T^{\text{scal}} = O(d\sqrt{T})$ with high probability, which, in turn give a $O(1/\sqrt{T})$ for R_T.

From SquareCB to FW-SquareCB (details in Appendix B.8). We now consider a CBCR with general reward function $\mu(x)$. The SquareCB algorithm [Foster and Rakhlin, 2020] is a randomized exploration strategy that delegates the learning of rewards to an arbitrary online regression algorithm. The scalar regret of SquareCB is bounded depending on the regret of the base regression algorithm.

For FW-SquareCB, we have access to an online regression oracle $\hat{\mu}_t$, an estimate of μ which is a function of h_t, which has regression regret bounded by $R_{\text{oracle}}(T)$. The exploration strategy of FW-SquareCB follows the same principles as SquareCB: let $g_t = \nabla f(\hat{s}_{t-1})$ and denote $\underline{\hat{\mu}}_t = g_t^\intercal \hat{\mu}_t(x_t)$,

Table 5.1: Regret bounds depending on assumptions and base algorithm $\mathfrak{A}$, for multi-armed bandits with K arms (in dimension d for LinUCB). See Appendix B.7 and B.8 for the full details.

Algorithm (FW-<bandit>)	Assumptions (informal)	Bound on R_T (simplified, using $\delta' = \delta$)
FW-LinUCB	$\mu(x)a = \theta x a$ for $\theta \in \mathbb{R}^{D\times d}, x \in \mathbb{R}^{d\times K}$	$\frac{LD_{\mathcal{K}} dD \ln\left((1+\frac{TLD_{\mathcal{K}}}{dD})/\delta\right)}{\sqrt{T}}$
FW-SquareCB	$\sum_{t=1}^{T} \mathbb{I}[\![\,]\!]\, \hat{\mu}_t(x_t)a_t - \mu(x_t)a_t\|_2^2 \le R_{\text{oracle}}(T)$	$\frac{L\sqrt{K\left(R_{\text{oracle}}(T) + D_{\mathcal{K}}^2 \ln(T/\delta)\right)}}{\sqrt{T}}$

so that $\underline{\hat{\mu}}_t^\intercal a = \langle g_t \,|\, \hat{\mu}_t(x_t)a\rangle$. Let $\mathfrak{A}_t = \text{FW-SquareCB}(h_t, x_t, \delta')$ defined as

$$\forall a \in \mathcal{A}, \mathfrak{A}_t(a) = \begin{cases} \frac{1}{K+\gamma_t\left(\underline{\hat{\mu}}_t^* - \underline{\hat{\mu}}_t^\intercal a\right)} & \text{if } a \neq \underline{a}_t \\ 1 - \sum_{a\in\mathcal{A}, a\neq\underline{a}_t} \mathfrak{A}_t(a) & \text{if } a = \underline{a}_t \end{cases} \quad \text{where } \underline{a}_t \in \operatorname*{argmax}_{a\in\mathcal{A}} \underline{\hat{\mu}}_t^\intercal a \text{ and } \underline{\hat{\mu}}_t^* = \underline{\hat{\mu}}_t \underline{a}_t$$

Then FW-SquareCB has R_T in $O(\sqrt{R_{\text{oracle}}(T)}/\sqrt{T})$ with high probability.

5.3.3 The case of nonsmooth f

When f is nonsmooth, we use a smoothing technique where the scalar regret is not measured using $\nabla f(\hat{s}_{t-1})$, but rather using gradients of a sequence $(f_t)_{t\in\mathbb{N}}$ of smooth approximations of f, whose smoothness decrease over time [see e.g., Lan, 2013, for applications of smoothing to FW]. We provide a comprehensive treatment of smoothing in our general approach described in Appendix B.4, while specific smoothing techniques are discussed in Appendix B.6.

We now describe the use of Moreau-Yosida regularization [Rockafellar and Wets, 2009, Def. 1.22]: $f_t(z) = \max_{y\in\mathbb{R}^D}\left(f(y) - \frac{\sqrt{t+1}}{2\beta_0}\|y-z\|_2^2\right)$. It is well-known that f_t is concave and L-Lipschitz whenever f is, and f_t is $\frac{\sqrt{t+1}}{\beta_0}$-smooth (see Lemma 38 in Appendix B.6). A related smoothing method was used by Agrawal and Devanur [2014] for (non-contextual) BCR. Our treatment of smoothing is more systematic than theirs, since we use a smoothing factor $\beta_0/\sqrt{t+1}$ that decreases over time rather than a fixed smoothing factor that depends on a pre-specified horizon. Our regret bound for CBCR is based on a scalar regret $R_T^{\text{scal,sm}}$ where $\nabla f_{t-1}(\hat{s}_{t-1})$ is used instead of $\nabla f(\hat{s}_{t-1})$:

$$R_T^{\text{scal,sm}} = \sum_{t=1}^{T} \max_{a\in\mathcal{A}} \langle \nabla f_{t-1}(\hat{s}_{t-1}) \,|\, \mu(x_t)a\rangle - \sum_{t=1}^{T} \langle \nabla f_{t-1}(\hat{s}_{t-1}) \,|\, r_t\rangle.$$

Theorem 14. *Under Assumptions A, for every $z_0 \in \mathcal{K}$, every $T \ge 1$ and every $\delta > 0, \delta' > 0$, Algorithm $\mathfrak{A}$ satisfies, with probability at least $1-\delta-\delta'$:*

$$R_T \le \frac{R_T^{\text{scal,sm}}}{T} + \frac{LD_{\mathcal{K}}}{\sqrt{T}}\left(\frac{D_{\mathcal{K}}}{L\beta_0} + 3\frac{L\beta_0}{D_{\mathcal{K}}} + \sqrt{2\ln\frac{1}{\delta}}\right).$$

The proof is given in Appendix B.6. Taking $\beta_0 = \frac{D_{\mathcal{K}}}{L}$ leads to a simpler bound where $\frac{D_{\mathcal{K}}}{L\beta_0} + 3\frac{L\beta_0}{D_{\mathcal{K}}} = 4$.

Algorithm 3: FW-LinUCBRank: linear contextual bandits for fair ranking.

input : $\delta' > 0, \lambda > 0, \hat{s}_0 \in \mathcal{K}$ $V_0 = \lambda \boldsymbol{I}_d, y_0 = \mathbf{0}_d, \hat{\theta}_0 = \mathbf{0}_d$

for $t = 1, \dots$ **do**

  Observe context $x_t \sim P$

  $\forall i, \hat{v}_{t,i} \leftarrow \hat{\theta}_{t-1}^\intercal x_{t,i} + \alpha_t\left(\frac{\delta'}{3}\right) \|x_{t,i}\|_{V_{t-1}^{-1}}$ // UCB on $v_i(x_t)$ (def. of α_t in Lem. 49, App. B.9)

  $a_t \leftarrow \text{top-}\overline{k}\{\frac{\partial f}{\partial z_{m+1}}(\hat{s}_{t-1})\hat{v}_{t,i} + \frac{\partial f}{\partial z_i}(\hat{s}_{t-1})\}_{i=1}^m$ // FW linear optimization step

  Observe exposed items $e_t \in \{0,1\}^m$ and user feedback $c_t \in \{0,1\}^m$

  Update $\hat{s}_t \leftarrow \hat{s}_{t-1} + \frac{1}{t}(r_t - \hat{s}_{t-1})$

  $V_t \leftarrow V_{t-1} + \sum_{i=1}^m e_{t,i} x_{t,i} x_{t,i}^\intercal, \quad y_t \leftarrow y_{t-1} + \sum_{i=1}^m c_{t,i} x_{t,i}$ and $\hat{\theta}_t \leftarrow V_t^{-1} y_t$ // regression

end

5.4 Contextual ranking bandits with fairness of exposure

In this section, we apply our reduction to the combinatorial bandit task of fair ranking, and obtain the first algorithm with regret guarantees in the contextual setting. This task is described in Example 4 (Sec. 5.2). We remind that there is a fixed set of m items to rank at each timestep t, and that actions are flattened permutation matrices ($\mathcal{A}$ is defined in Ex. 4, Eq. (5.1)). The context $x_t \sim P$ is a matrix $x_t = (x_{t,i})_{i \in [\![m]\!]}$ where each $x_{t,i} \in \mathbb{R}^d$ represents a feature vector of item i for the current user.

Observation model. The *user utility* $u(x_t)$ is given by a position-based model with position weights $b(x_t) \in [0,1]^m$ and expected value for each item $v(x_t) \in [0,1]^m$. Denoting $u(x_t)$ the flattened version of $v(x_t)b(x_t)^\intercal \in \mathbb{R}^{m \times m}$, the user utility is [Lagrée et al., 2016, Singh and Joachims, 2018]:

$$\langle u(x_t) \,|\, a \rangle = \sum_{i=1}^m v_i(x_t) \sum_{k=1}^m a_{i,k} b_k(x_t).$$

In this model, $b_k(x_t) \in [0,1]$ is the probability that the user observes the item at rank k. The quantity $\sum_{k=1}^m a_{i,k} b_k(x_t)$ is thus the probability that the user observes item i given ranking a. We denote $\overline{k} = \max_{x \in \mathcal{X}} \|b(x)\|_0 \leq m$ the maximum rank that can be exposed to any user. In most practical applications, $\overline{k} \ll m$. As formalized in Assumption D below, the position weights $b_k(x)$ are always non-increasing with k since the user browses the recommended items in order of their rank. We use a linear assumption for item values, where $D_\mathcal{X}$ and D_θ are known constants:

Assumption C. $\sup_{x \in \mathcal{X}} \|x\|_2 \leq D_\mathcal{X}$ *and* $\exists \theta \in \mathbb{R}^d, \|\theta\|_2 \leq D_\theta$ *s.t.* $\forall x \in \mathcal{X}, \forall i \in [\![m]\!], v_i(x) = \theta^\intercal x_i$.

We propose an observation model where values $v_i(x)$ *and* position weights $b(x)$ are *unknown*. However, we assume that at each time step t, after computing the ranking a_t, we have two types of feedback: first, $e_{t,i} \in \{0,1\}$ is 1 if item i has been exposed to the user, and 0 otherwise. Second $c_{t,i} \in \{0,1\}$ which represents a binary `like`/`dislike` feedback from the user. We have

$$\mathbb{E}[e_{t,i} | x_t, a_t] = \sum_{k=1}^m a_{t,i,k} b_k(x_t) \qquad \mathbb{E}[c_{t,i} | x_t, e_{t,i}] = \begin{cases} v_i(x_t) & \text{if } e_{t,i} = 1 \\ 0 & \text{if } e_{t,i} = 0 \end{cases}$$

This observation model captures well applications such as newsfeed ranking on mobile devices or dating applications where only one post/profile is shown at a time. What we gain with this model is that $b(x)$ can *depend arbitrarily on the context* x, while previous work on bandits in the position-based model assumes b known and context-independent [Lagrée et al., 2016].[7]

[7] When b is unknown, depends on the context x, and we do not observe e_t, several approaches have been proposed to estimate the position weights [see e.g., Fang et al., 2019]. Incorporating these approaches in contextual bandits for ranking is likely feasible but out of the scope of this work.

Fairness of exposure. There are $D = m + 1$ rewards, i.e., $\mu(x) \in \mathbb{R}^{(m+1)\times m^2}$. Denoting $\mu_i(x)$ the ith-row of $\mu(x)$, seen as a column vector, each of the m first rewards is the exposure of a specific item, while the $m + 1$-th reward is the user utility:

$$\forall i \in [\![m]\!], \langle \mu_i(x) \,|\, a \rangle = \sum_{k=1}^{m} a_{i,k} b_k(x) \qquad \text{and} \qquad \mu_{m+1}(x) = u(x)$$

The observed reward vector $r_t \in \mathbb{R}^D$ is defined by $\forall i \in [\![m]\!], r_{t,i} = e_{t,i}$ and $r_{t,m+1} = \sum_{i=1}^m c_{t,i}$. Notice that $\mathbb{E}\big[r_{t,m+1}\big|x_t\big] = u(x_t)$. Let $\mathcal{K}$ be the convex hull of $\{z \in \{0,1\}^{m+1} : \sum_{i=1}^m z_i \leq \overline{k} \text{ and } z_{m+1} \leq \sum_{i=1}^m z_i\}$, we have $D_{\mathcal{K}} \leq \sqrt{\overline{k}}\sqrt{\overline{k}+2} \leq \overline{k}+1$ and $r_t \in \mathcal{K}$ with probability 1. The objective function $f : \mathbb{R}^D \to \mathbb{R}$ makes a trade-off between average user utility and inequalities in item exposure (we gave an example in Eq. (5.2)). The remaining assumptions of our framework are that the objective function is non-decreasing with respect to average user utility. This is not required but it is natural (see Example 4) and slightly simplifies the algorithm.

Assumption D. *The assumptions of the framework described above hold, as well as Assumption B. Moreover,* $\forall z \in \mathcal{K}\ \frac{\partial f}{\partial z_{m+1}}(z) > 0$, *and* $\forall x \in \mathcal{X}, 1 \geq b_1(x) \geq \ldots \geq b_{\overline{k}}(x) = \ldots = b_m(x) = 0$.

Algorithm and results. We present the algorithm in the setting of linear contextual bandits, using LinUCB [Abbasi-Yadkori et al., 2011, Li et al., 2010] as scalar exploration/exploitation algorithm in Algorithm 3. It builds reward estimates based on Ridge regression with regularization parameter λ. As in the previous section, we focus on the case where f is smooth but the extension to nonsmooth f is straightforward, as described in Section 5.3. Appendix B.9 provides the analysis for the general case.

As noted by Do et al. [2021c], Frank-Wolfe algorithms are particularly suited for fair ranking in the position-based model. This is illustrated by line 4 of Alg. 3, where for $\tilde{u} \in \mathbb{R}^m$, top-$\overline{k}(\tilde{u})$ outputs a permutation (matrix) of $[\![m]\!]$ that sorts the top-$\overline{k}$ elements of $\tilde{u}$. Alg. 3 is thus *computationally fast*, with a cost dominated by the top-$\overline{k}$ sort. It also has an intuitive interpretation as giving items an adaptive bonus depending on ∇f (e.g., boosting the scores of items which received low exposure in previous steps). The following result is a consequence of [Do et al., 2021c, Theorem 1]:

Proposition 15. *Let* $t \in \mathbb{N}_*$ *and* $\hat{\mu}_t$ *such that* $\forall i \in [\![m]\!], \hat{\mu}_{t,i} = \mu_i(x_t)$ *and* $\hat{\mu}_{t,m+1} = \hat{v}_t b(x_t)^\intercal$ *viewed as a column vector, with* $\hat{v}$ *defined in line 3 of Algorithm 3. Then, under Assumption D,* a_t *defined on line 4 of Algorithm 3 satisfies:* $\langle \nabla f(\hat{s}_{t-1}) \,|\, \hat{\mu}_t a_t \rangle = \operatorname*{argmax}_{a \in \mathcal{A}} \langle \nabla f(\hat{s}_{t-1}) \,|\, \hat{\mu}_t a \rangle$.

The proposition says that even though computing a_t as in line 4 of Alg. 3 does not require the knowledge of $b(x_t)$, we still obtain the optimal update direction according to $\hat{\mu}_t$. Together with the usage of the observed reward r_t in FW iterates (instead of e.g., $\hat{\mu}_t a_t$ as would be done by Agrawal and Devanur [2014]), this removes the need for explicit estimates of $\mu(x_t)$. This is how our algorithm works *without knowing the position weights* $b(x_t)$, which are then allowed to depend on the context.

The usage of $\hat{v}_t$ to compute a_t follows the usual confidence-based approach to explore/exploitation principles for linear bandits, which leads to the following result (proven in Appendix B.9):

Theorem 16. *Under Assumptions B, C and D, for every* $\delta' > 0$, *every* $T \in \mathbb{N}_*$, *every* $\lambda \geq D_{\mathcal{X}}^2 \overline{k}$, *with probability at least* $1 - \delta'$, *Algorithm 3 has scalar regret bounded by*

$$R_T^{\text{scal}} = O\Big(L\sqrt{T\overline{k}}\sqrt{d\ln(T/\delta')}\Big(\sqrt{d\ln(T/\delta')} + D_\theta\sqrt{\lambda} + \sqrt{\overline{k}/d}\Big)\Big).$$

Thus, considering only $d, T, \overline{k}$ *and* $\delta = \delta'$ *Alg. 3 has regret* $R_T \leq O\big(\frac{d\overline{k}\ln(T/\delta)}{\sqrt{T}}\big)$ *w.p. at least* $1 - \delta$.

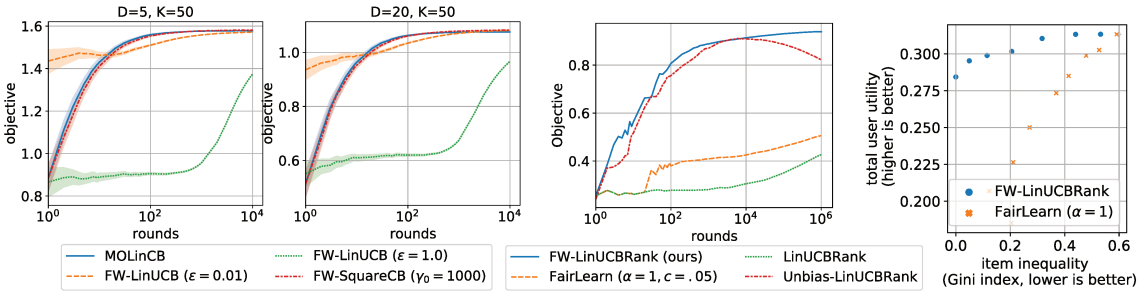


Figure 5.1: **(left)** Multi-armed CBCR: Objective values on environments from [Mehrotra et al., 2020]. **(middle)** Ranking CBCR: Fairness objective value over timesteps on Last.fm data. **(right)** Ranking CBCR: Trade-off between user utility and item inequality after 5×10^6 iterations on Last.fm data.

5.5 Experiments

We present two experimental evaluations of our approach, which are fully detailed in App. B.2.

5.5.1 Multi-armed CBCR: Application to multi-objective bandits

We first focus on the multi-objective recommendation task of Example 3 where $f(z) = \sum_{i=1}^{D} w_i z_i^{\uparrow}$.

Algorithms. We evaluate our two instantiations presented in Sec. 5.3.2 with the Moreau-Yosida smoothing technique of Sec. 5.3.3: (i) FW-SquareCB with Ridge regression and (ii) FW-LinUCB, where exploration is controlled by a scaling variable ϵ on the exploration bonus of each arm. We compare them to MOLinCB from [Mehrotra et al., 2020].

Environments. We reproduce the synthetic environments of Mehrotra et al. [2020], where the context and reward parameters are generated randomly, and $w_i = \frac{1}{2^{i-1}}$. We set $K = 50$ and $D \in \{5, 20\}$ (we also vary K in App. B.2). Each simulation is repeated with 100 random seeds.

Results. Following [Mehrotra et al., 2020], we evaluate the algorithms' performance by measuring the value of $f(\frac{1}{T}\sum_{t=1}^{T} \mu(x_t)a_t)$ over time. Our results are shown in Figure 5.1 (left). We observe that our algorithm FW-SquareCB obtains comparable performance with the baseline MOLinCB. These algorithms converge after ≈ 100 rounds. In this environment from [Mehrotra et al., 2020], only little exploration is needed, hence FW-LinUCB obtains better performance when ϵ is smaller ($\epsilon = 0.01$). The advantage of using an FW instantiation for the multi-objective bandit optimization task is that unlike MOLinCB, its convergence is also supported by our theoretical regret guarantees.

5.5.2 Ranking CBCR: Application to fairness of exposure in rankings

We now tackle the ranking problem of Section 5.4. We show how FW-LinUCBRank allows to fairly distribute exposure among items on a music recommendation task with bandit user feedback.

Environment. Following [Patro et al., 2020], we use the Last.fm music dataset from [Cantador et al., 2011], from which we extract the top 50 users and items with the most listening counts. We use a protocol similar to Li et al. [2016] to generate context and rewards from those. We use $\bar{k} = 10$ ranking slots, and exposure weights $b_k(x) = \frac{\log(2)}{1+\log(k)}$. Simulations are repeated with 10 seeds.

Algorithms. Our algorithm is FW-LinUCBRank with the nonsmooth objective f of Eq. (5.2), which trades off between user utility and item inequality. We study other fairness objectives in App. B.2. Our first baseline is *LinUCBRank* [Ermis et al., 2020], designed for ranking without fairness. Then, we study two baselines with amortized fairness of exposure criteria. Mansoury et al. [2021a] proposed a fairness module for UCB-based ranking algorithms, which we plug into

LinUCBRank. We refer to this baseline as *Unbiased-LinUCBRank*. Finally, the *FairLearn*(c, α) algorithm [Patil et al., 2020] enforces as fairness constraint that the pulling frequency of each arm be $\geq c$, up to a tolerance α. We implement as third baseline a simple adaptation of *FairLearn* to contextual bandits and ranking.

Dynamics. Figure 5.1 (middle) represents the values of f over time achieved by the competing algorithms, for fixed $\beta = 1$. As expected, compared to the fairness-aware and -unaware baselines, our algorithm FW-LinUCBRank reaches the best values of f. Interestingly, *Unbiased-LinUCBRank* also obtains high values of f on the first 10^4 rounds, but its performance starts decreasing after more iterations. This is because *Unbiased-LinUCBRank* is not guaranteed to converge to an optimal trade-off between user fairness and item inequality.

At convergence. We analyse the trade-offs achieved after $5 \cdot 10^6$ rounds between user utility and item inequality measured by the Gini index. We vary β in the objective f of Eq. (5.2) for FW-LinUCBRank and the strength c in *FairLearn*(c, α), with tolerance $\alpha = 1$. In Fig. 5.1 (right), we observe that compared to *FairLearn*, FW-LinUCBRank converges to much higher user utility at all levels of inequality among items. In particular, it achieves zero-unfairness at little cost for user utility.

5.6 Conclusion

We presented the first general approach to contextual bandits with concave rewards. To illustrate the usefulness of the approach, we show that our results extend randomized exploration with generic online regression oracles to the concave rewards setting, and extend existing ranking bandit algorithms to fairness-aware objective functions. The strength of our reduction is that it can produce algorithms for CBCR from any contextual bandit algorithm, including recent extensions of SquareCB to infinite compact action spaces [Zhu and Mineiro, 2022, Zhu et al., 2022] and future ones.

In our main application to fair ranking, the designer sets a fairness trade-off f to optimize. In practice, they may choose f among a small class by varying hyperparameters (e.g. β in Eq. (5.2)). An interesting open problem is the integration of recent elicitation methods for f [e.g., Lin et al., 2022] in the bandit setting. Another interesting issue is the generalization of our framework to include constraints [Agrawal and Devanur, 2016]. Finally, we note that the deployment of our algorithms requires to carefully design the whole machine learning setup, including the specification of reward functions [Stray et al., 2021], the design of online experiments [Bird et al., 2016], while taking feedback loops into account [Bottou et al., 2013, Jiang et al., 2019, Dean and Morgenstern, 2022].

Chapter 6

User fairness as envy-freeness

Contents

This chapter is the article *Online certification of preference-based fairness for personalized recommender systems*, published at AAAI 2022 (see [Do et al., 2022a]).

In Chapters 3 and 4, we considered the problem of the *designer* of a recommender systems who is concerned with *two-sided fairness* for users and items. In this chapter, we shift our focus to *auditing* recommender systems and prioritizing *user-side fairness*. This work was inspired by the growing concerns raised by audits for user fairness in advertising systems, such as the gender-based disparities observed in ad delivery rates for different companies proposing similar jobs [Imana et al., 2021, Lambrecht and Tucker, 2019, Datta et al., 2015]. Our contribution to this research is a complement to existing audits, most of which do not control for disparities that align with user preferences. To address this limitation, we proposed to test for the preference-based criterion of envy-freeness, which stipulates that no user should prefer their recommendations to those of other users. For example, in a job ad system where two users Alice and Bob are interested in taxi driver roles [Ali et al., 2019], if Bob is the only one to receive ads for driver jobs, then the system is deemed unfair by the envy-freeness criterion.

Envy-freeness is a fairness criterion that has roots in fair division. In the context of recommender systems, it leads to different assessments than our previous framework which was also rooted in fair

division, but approached fairness as redistribution of utility, following the Pigou-Dalton transfer principle. One advantage of envy-freeness is that it avoids the challenge of interpersonal comparisons of utilities across users, which are difficult due to the different scaling of performance metrics used to measure user utilities (since users have different browsing or rating patterns). Comparing and aggregating user utilities is necessary in the design of recommender systems, where practitioners traditionally maximize measures of performance on average over users, and in the design of *two-sided fair* recommender systems, where designers need to make trade-offs between users' and items' utilities. In contrast, interpersonal comparisons can be avoided for the auditor, who only seeks reliable evaluations of how the system serves some users compared to others.

In the previous chapters 3 and 4, we designed ranking algorithms to improve the exposure of small items while prioritizing the utilities of the worst-off users. These algorithms produce rankings which are suboptimal for average user utility, since one of the main motivations of two-sided fairness in rankings is to mitigate the winner-take-all effects of the user-side optimal rankings. In contrast, we show that envy-freeness is compatible with providing optimal recommendations for users. Further, our previous two-sided fair ranking algorithms may not pass the audit for envy-freeness, as optimal ranking policies for objectives that include a concave item fairness term are not envy-free for users in general. For example, promoting less popular employers by boosting their ads in one user's recommendations may lead this user to envy another user who receives recommendations of popular employers. In contrast, our perspective in this chapter is that of an auditor solely focused on assessing fairness for users, regardless of whether user-side unfairness is a consequence of other objectives, such as item-side fairness.

We argue that the audit perspective is just as important as that of the designer, given the significant role played by audits for user fairness in raising awareness about the need for fairness in recommender systems. In fact, existing audits have led to settlements that drove online platforms to change their ad recommendation algorithms to comply with new requirements for user fairness [Bogen et al., 2023]. Moreover, designers can use the evaluations produced by auditors as additional diagnoses to improve their systems. If an internal auditor detects envy in a recommender system, then the designer can examine whether removing user envy would lead to increased inequalities on the item side, and assess whether this trade-off is acceptable with respect to the objective set for item fairness. In practice though, we recommend that our audit for envy-freeness be used in applications where user-side fairness, rather than item-side fairness, is the main concern, such as in the line of work on auditing ad delivery systems.

On the algorithmic side, the auditing problem is completely different from the designer's problem. We cast the audit for envy-freeness as a *pure exploration* bandit problem, since the goal is to provide high-confidence envy-freeness certificates from as few samples as possibles. This is different from Chapter 5 where the designer addresses a regret minimization problem, in order to deliver recommendations while balancing exploration and exploitation. The auditing algorithm OCEF that we introduce in this chapter is meant as an auditing tool, not a recommendation strategy.

We also note the following presentation differences with the previous chapters:

1. Recommendation policies are distributions over single items. We do not consider rankings.
2. We use the original notation of the article, which is different from the rest of the thesis since we address a different problem.
3. In the formal analysis of the compatibility between envy-freeness and item-side fairness criteria in Section 6.3.3, we consider fairness criteria within the recommendations of a single user ("within list"), rather than across users. Moreover, compared to Chapter 3, we use the terminology "equity of exposure" instead of "quality-weighted exposure", and "parity of exposure" instead of "equal

exposure".

Abstract

Recommender systems are facing scrutiny because of their growing impact on the opportunities we have access to. Current audits for fairness are limited to coarse-grained parity assessments at the level of sensitive groups. We propose to audit for *envy-freeness*, a more granular criterion aligned with individual preferences: every user should prefer their recommendations to those of other users. Since auditing for envy requires to estimate the preferences of users beyond their existing recommendations, we cast the audit as a new pure exploration problem in multi-armed bandits. We propose a sample-efficient algorithm with theoretical guarantees that it does not deteriorate user experience. We also study the trade-offs achieved on real-world recommendation datasets.

6.1 Introduction

Recommender systems shape the information and opportunities available to us, as they help us prioritize content from news outlets and social networks, sort job postings, or find new people to connect with. To prevent the risk of unfair delivery of opportunities across users, substantial work has been done to audit recommender systems Sweeney [2013], Asplund et al. [2020], Imana et al. [2021]. For instance, Datta et al. [2015] found that women received fewer online ads for high-paying jobs than equally qualified men, while Imana et al. [2021] observed different delivery rates of ads depending on gender for different companies proposing similar jobs.

The audits above aim at controlling for the possible acceptable justifications of the disparities, such as education level in job recommendation audits. Yet, the observed disparities in recommendation do not necessarily imply that a group has a less favorable treatment: they might as well reflect that individuals of different groups tend to prefer different items. To strengthen the conclusions of the audits, it is necessary to develop methods that account for user preferences. Audits for equal satisfaction between user groups follow this direction [Mehrotra et al., 2017], but they also have limitations. For example, they require interpersonal comparisons of measures of satisfaction, a notoriously difficult task [Sen, 1999].

We propose an alternative approach to incorporating user preferences in audits which focuses on *envy-free recommendations*: the recommender system is deemed fair if each user prefers their recommendation to those of all other users. Envy-freeness allows a system to be fair even in the presence of disparities between groups as long as these are justified by user preferences. On the other hand, if user B systematically receives better opportunities than user A *from A's perspective*, the system is unfair. The criterion does not require interpersonal comparisons of satisfaction, since it relies on comparisons of different recommendations from the perspective of the same user. Similar fairness concepts have been studied in classification tasks under the umbrella of preference-based fairness [Zafar et al., 2017b, Kim et al., 2019, Ustun et al., 2019]. Envy-free recommendation is the extension of these approaches to personalized recommender systems.

Compared to auditing for recommendation parity or equal satisfaction, auditing for envy-freeness poses new challenges. First, envy-freeness requires answering counterfactual questions such as "would user A get higher utility from the recommendations of user B than their own?", while searching for the users who most likely have the best recommendations from A's perspective. This type of question can be answered reliably only through active exploration, hence we cast it in the framework of pure exploration bandits Bubeck et al. [2009]. To make such an exploration

possible, we consider a scenario where the auditor is allowed to replace a user's recommendations with those that another user would have received in the same context. Envy, or the absence thereof, is estimated by suitably choosing whose recommendations should be shown to whom. While this scenario is more intrusive than some black-box audits of parity, auditing for envy-freeness provides a more compelling guarantee on the wellbeing of users subject to the recommendations.

The second challenge is that active exploration requires randomizing the recommendations, which in turn might alter the user experience. In order to control this cost of the audit (in terms of user utility), we follow the framework of conservative exploration Wu et al. [2016], Garcelon et al. [2020a], which guarantees a performance close to the audited system. We provide a theoretical analysis of the trade-offs that arise, in terms of the cost and duration of the audit (measured in the number of timesteps required to output a certificate).

Our technical contributions are twofold. **(1)** We provide a novel formal analysis of envy-free recommender systems, including a comparison with existing item-side fairness criteria and a probabilistic relaxation of the criterion. **(2)** We cast the problem of auditing for envy-freeness as a new pure exploration problem in bandits with conservative exploration constraints, and propose a sample-efficient auditing algorithm which provably maintains, throughout the course of the audit, a performance close to the audited system.

We discuss the related work in Sec. 6.2. Envy-free recommender systems are studied in Sec. 6.3. In Sec. 6.4, we present the bandit-based auditing algorithm. In Sec. 6.5, we investigate the trade-offs achieved on real-world datasets.

6.2 Related work

Fair recommendation The domain of fair machine learning is organized along two orthogonal axes. The first axis is whether fairness is oriented towards groups defined by protected attributes Barocas and Selbst [2016], or rather oriented towards individuals Dwork et al. [2012]. The second axis is whether fairness is a question of *parity* (predictions [or prediction errors] should be invariant by group or individual) Corbett-Davies and Goel [2018], Kusner et al. [2017], or *preference-based* (predictions are allowed to be different if they faithfully reflect the preferences of all parties) Zafar et al. [2017b], Kim et al. [2019], Ustun et al. [2019]. Our work takes the perspective of envy-freeness, which follows the preference-based approach and is aimed towards individuals.

The literature on fair recommender systems covers two problems: *auditing* existing systems, and *designing* fair recommendation algorithms. Most of the *auditing* literature focused on group parity in recommendations Hannak et al. [2014], Lambrecht and Tucker [2019], and equal user utility Mehrotra et al. [2017], Ekstrand et al. [2018], while our audit for envy-freeness focuses on whether personalized results are aligned with (unknown) user preferences. On the *designing* side, Patro et al. [2020], Ilvento et al. [2020] cast fair recommendation as an allocation problem, with criteria akin to envy-freeness. They do not address the partial observability of preferences, so they cannot guarantee user-side fairness without an additional certificate that the estimated preferences effectively represent the true user preferences. Our work is thus complementary to theirs.

While we study fairness for users, recommender systems are multi-sided Burke [2017], Patro et al. [2020], thus fairness can also be oriented towards recommended items Celis et al. [2017b], Biega et al. [2018], Geyik et al. [2019].

Multi-armed bandits In *pure exploration* bandits Bubeck et al. [2009], Audibert and Bubeck [2010], an agent has to identify a specific set of arms after exploring as quickly as possible, without performance constraints. Our setting is close to threshold bandits Locatelli et al. [2016], **?** where

the goal is to find arms with better performance than a given baseline. Outside pure exploration, in the *regret minimization* setting, conservative exploration Wu et al. [2016] enforces the anytime average performance to be not too far worse than that of a baseline arm.

In our work, the baseline is *unknown* – it is the current recommender system – and the other "arms" are other users' policies. The goal is to make the decision as to whether an arm is better than the baseline, while not deteriorating performance compared to the baseline. We thus combine pure exploration and conservative constraints.

Existing work on fairness in exploration/exploitation Joseph et al. [2016], Jabbari et al. [2017], Liu et al. [2017] is different from ours because unrelated to personalization.

Fair allocation Envy-freeness was first studied in fair allocation Foley [1967] in social choice. Our setting is different because: a) the same item can be given to an unrestricted number of users, and b) true user preferences are unknown.

6.3 Envy-free recommendations

6.3.1 Framework

There are M users, and we identify the set of users with $[\![M]\!] = \{1, \dots, M\}$. A personalized recommender system has one stochastic recommendation policy π^m per user m. We denote by $\pi^m(a|x)$ the probability of recommending item $a \in \mathcal{A}$ for user $m \in [\![M]\!]$ in context $x \in \mathcal{X}$. We assume that $\mathcal{X}$ and $\mathcal{A}$ are finite to simplify notation, but this has no impact on the results. We consider a synchronous setting where at each time step t, the recommender system observes a context $x_t^m \sim q^m$ for each user, selects an item $a_t^m \sim \pi^m(.|x_t^m)$ and observes reward $r_t^m \sim \nu^m(a_t^m|x_t^m) \in [0, 1]$. We denote by $\rho^m(a|x)$ the expected reward for user m and item a in context x, and, for any recommendation policy π, $u^m(\pi)$ is the utility of m for π:

$$\begin{aligned} u^m(\pi) &= \mathbb{E}_{x\sim q^m}\mathbb{E}_{a\sim\pi(.|x)}\mathbb{E}_{r\sim\nu^m(a|x)}\left[r\right] \\ &= \sum_{x\in\mathcal{X}}\sum_{a\in\mathcal{A}} q^m(x)\pi(a|x)\rho^m(a|x) \end{aligned} \tag{6.1}$$

We assume that the environment is *stationary*: the context and reward distributions q^m and ν^m, as well as the policies π^m are fixed. Even though in practice policies evolve as they learn from user interactions and user needs change over time, we leave the study of non-stationarities for future work. The stationary assumption approximately holds when these changes are slow compared to the time horizon of the audit, which is reasonable when significant changes in user needs or recommendation policies take e.g., weeks. Our approach applies when items a are single products as well as when items are structured objects such as rankings. Examples of (context x, item a) pairs include: x is a query to a search engine and a is a document or a ranking of documents, or x is a song chosen by the user and a a song to play next or an entire playlist. Remember, our goal is *not* to learn the user policies π^m, but rather to audit existing π^ms for fairness.

6.3.2 ϵ-envy-free recommendations

Existing audits for user-side fairness in recommender systems are based on two main criteria:

1. *recommendation parity*: the distribution of recommended items should be equal across (groups of) users,

2. *equal user utility*: all (groups of) users should receive the same utility, i.e. $\forall m, n,\ u^m(\pi^m) = u^n(\pi^n)$.

There are two ways in which these criteria conflict with the goal of personalized recommender systems to best accomodate user preferences. First, recommendation parity does not control for disparities that are aligned with user preferences. Second, equal user utility drives utility down as soon as users have different best achievable utilities. To address these shortfalls, we propose envy-freeness as a complementary diagnosis for the fairness assessment of personalized recommender systems. In this context, envy-freeness requires that users prefer their recommendations to those of any other user:

Definition 5. *Let* $\epsilon \geq 0$. *A recommender system is* ϵ-envy-free *if:* $\forall m, n \in [M]:\quad u^m(\pi^n) \leq \epsilon + u^m(\pi^m)$.

Envy-freeness, originally studied in fair allocation Foley [1967] and more recently fair classification Balcan et al. [2018], Ustun et al. [2019], Kim et al. [2019], stipulates that it is fair to apply different policies to different individuals or groups as long as it benefits everyone. Following this principle, we consider the personalization of recommendations as fair only if it better accommodates individuals' preferences. In contrast, we consider unfair the failure to give users a better recommendation when one such is available to others.

Unlike parity or equal utility, envy-freeness is in line with giving users their most preferred recommendations (see Sec. 6.3.3). Another improvement from equal user utility is that it does not involve interpersonal utility comparisons.

Envy can arise from a variety of sources, for which we provide concrete examples in our experiments (Sec. 6.5.1).

Remark 3. *We discuss an immediate extension of envy-freeness from individuals to groups of users in App. C.2, in the special case where groups have homogeneous preferences and policies. Defining group envy-free recommendations in the general case is nontrivial and left for future work.*

6.3.3 Compatibility of envy-freeness

Optimal recommendations are envy-free. [1] Let $\pi^{m,*} \in \operatorname{argmax}_\pi u^m(\pi)$ denote an optimal recommendation policy for m. Then *the optimal recommender system* $(\pi^{m,*})_{m \in M}$ *is envy-free* since: $u^m(\pi^{m,*}) = \max_\pi u^m(\pi) \geq u^m(\pi^{n,*})$. In contrast, achieving equal user utility in general can only be achieved by decreasing the utility of best-served users for the benefit of no one. It is also well-known that achieving parity in general requires to deviate from optimal predictions [Barocas et al., 2018].

Envy-freeness vs. item-side fairness Envy-freeness is a user-centric notion. Towards multi-sided fairness [Burke, 2017], we analyze the compatibility of envy-freeness with item-side fairness criteria for rankings from Singh and Joachims [2018], based on sensitive categories of items (denoted $\mathcal{A}_1, ..., \mathcal{A}_S$). *Parity of exposure* prescribes that for each user, the exposure of an item category should be proportional to the number of items in that category. In *Equity of exposure*[2], the exposure of item categories should be proportional to their average relevance to the user.

The optimal policies under parity and equity of exposure constraints, denoted respectively by $(\pi^{m,\text{par}})_{m=1}^M$ and $(\pi^{m,\text{eq}})_{m=1}^M$, are defined given user m and context x as:

[1] App.C.1 shows the difference between envy-freeness and optimality certificates.

[2] Singh and Joachims [2018] use the terminology of demographic parity (resp. disparate treatment) for what we call parity (resp. equity) of exposure. Our use of "equity" follows Biega et al. [2018].

(parity)
$$\pi^{m,\text{par}}(.|x) = \underset{\substack{p:\mathcal{A}\to[0,1]\\ \sum_a p(a)=1}}{\text{argmax}} \sum_{a\in\mathcal{A}} p(a)\rho^m(a|x)$$
$$\text{u.c. } \forall s \in [\![S]\!], \sum_{a\in\mathcal{A}_s} p(a) = \frac{|\mathcal{A}_s|}{|\mathcal{A}|}. \tag{6.2}$$

Optimal policies under equity of exposure are defined similarly[3], but the constraints are $\forall s, \sum_{a\in\mathcal{A}_s} p(a) = \frac{\sum_{a\in\mathcal{A}_s}\rho^m(a|x)}{\sum_{a\in\mathcal{A}}\rho^m(a|x)}$.

We show their relation to envy-freeness:

Proposition 17. *With the above notation:*

- *the policies* $(\pi^{m,\text{par}})_{m=1}^M$ *are envy-free, while*
- *the policies* $(\pi^{m,\text{eq}})_{m=1}^M$ *are not envy-free in general.*

Optimal recommendations under parity of exposure are envy-free because the parity constraint (6.2) is the same for all users. Given two users m and n, $\pi^{m,\text{par}}$ is optimal for m under (6.2) and $\pi^{n,\text{par}}$ satisfies the same constraint, so we have $u^m(\pi^{m,\text{par}}) \geq u^m(\pi^{n,\text{par}})$.

In contrast, the optimal recommendations under equity of exposure are, in general, not envy-free. A first reason is that less relevant item categories reduce the exposure of more relevant categories: a user who prefers item a but who also likes item b from another category envies a user who only liked item is a. Note that *amortized* versions of the criterion and other variants considering constraint averages over user/contexts [Biega et al., 2018, Patro et al., 2020] have similar pitfalls unless envy-freeness is explictly enforced, as in Patro et al. [2020] who developed an envy-free algorithm assuming the true preferences are known. For completeness, we describe in App.C.1 a second reason why equity of exposure constraints create envy, and an edge case where they do not.

6.3.4 Probabilistic relaxation of envy-freeness

Envy-freeness, as defined in Sec. 6.3.2, (a) compares the recommendations of a target user to those of *all* other users, and (b) these comparisons must be made for *all* users. In practice, as we show, this means that the sample complexity of the audit increases with the number of users, and that all users must be part of the audit.

In practice, it is likely sufficient to relax both conditions on all users to give a guarantee for most recommendation policies and most users. Given two small probabilities λ and γ, the relaxed criterion we propose requires that for at least $1-\lambda$ fraction of users, the utility of users for their own policy is in the top-γ% of their utilities for anyone else's policy. The formal definition is given below. The fundamental observation, which we prove in Th. 19 in Sec. 6.4.5, is that the sample complexity of the audit and the number of users impacted by the audit are now *independent on the total number of users*. We believe that these relaxed criteria are thus likely to encourage the deployment of envy-free audits in practice.

Definition 6. *Let* $\epsilon, \gamma, \lambda \geq 0$. *Let* U_M *denote the discrete uniform distribution over* $[\![M]\!]$. *A user* m *is* (ϵ,γ)-envious *if:*

$$\mathbb{P}_{n\sim U_M}\left[u^m(\pi^m) + \epsilon < u^m(\pi^n)\right] > \gamma.$$

[3]The original criterion [Singh and Joachims, 2018, Eq. 4] would be written in our case as $\forall s, s' \in [\![S]\!]$, $\frac{1}{|\mathcal{A}_s|}\sum_{a\in\mathcal{A}_s} p(a) = \frac{1}{|\mathcal{A}_{s'}|}\sum_{a\in\mathcal{A}_{s'}} p(a)$, which is equivalent to (6.2). A similar remark holds for the equity constraint.

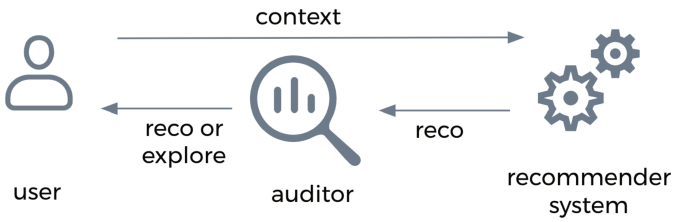


Figure 6.1: Auditing scenario: the auditor either shows the user their recommendation in the current rec. system, or explores by showing the recommendation given to another user.

A recommender system is $(\epsilon, \gamma, \lambda)$-envy-free *if at least a* $(1 - \lambda)$ *fraction of its users are not* (ϵ, γ)-envious.

6.4 Certifying envy-freeness

6.4.1 Auditing scenario

The envy-freeness auditor must answer the counterfactual question: "had user m been given the recommendations of user n, would m get higher utility?". The main challenge is that the answer requires to access to user preferences, which are only partially observed since users only interact with recommended items. There is thus a need for an active exploration process that recommends items which would not have been recommended otherwise.

To make such an exploration possible, we consider the following auditing scenario: at each time step t, the auditor chooses to either (a) give the user a "normal" recommendation, or (b) explore user preferences by giving the user a recommendation from another user (see Fig. 6.1) . This scenario has the advantage of lightweight infrastructure requirements, since the auditor only needs to query another user's policy, rather than implementing a full recommender system within the operational constraints of the platform. Moreover, this interface is sufficient to estimate envy because envy is defined based on the performance of other user's policies. This type of internal audit Raji et al. [2020] requires more access than usual external audits that focus on recommendation parity, but this is necessary to explore user preferences.

We note that the auditor must make sure that this approach follows the relevant ethical standard for randomized experiments in the context of the audited system. The auditor must also check that using other users' recommendation policies does not pose privacy problems. From now on, we assume these issues have been resolved.

6.4.2 The equivalent bandit problem

We now cast the audit for envy-freeness as a new variant of pure exploration bandit problems. We first focus on auditing envy for a single target user and define the corresponding objectives, then we present our auditing algorithm. Finally we specify how to use it for the certification of either the exact or probabilistic envy-freeness criteria.

For a target user m, the auditor must estimate whether $u^m(\pi^m) + \epsilon \geq u^m(\pi^n)$, for n in a subset $\{n_1, ..., n_K\}$ of K users from $[\![M]\!]$ (where K is specified later, depending on the criterion). As we first focus on auditing envy for one target user m, we drop all superscripts m to simplify notation. We identify $\{n_1, ..., n_K\}$ with $[\![K]\!]$ and rename $\left(u^m(\pi^{n_1}), ..., u^m(\pi^{n_K})\right)$ as $(\mu_1, ..., \mu_K)$. To estimate μ_k, we obtain samples by making recommendations using the policy π^k and observing the reward. The remaining challenge is to choose which user k to sample at each time step while not deteriorating the experience of the target user too much. Index 0 represents the target user: we

Algorithm 4: OCEF algorithm. ξ_t (line 4) evaluates the conservative exploration constraint and is defined in (6.4). Values for $\beta_k(t)$ and confidence bounds $\underline{\mu}_k$ and $\overline{\mu}_k$ are given in Lemma 53.

input : Confidence parameter δ, conservative exploration parameter α, envy parameter ϵ
output : `envy` or ϵ−`no-envy`

1 $S_0 \leftarrow [\![K]\!]$ // all arms except 0
2 **for** *t=1, ...* **do**
3 Choose ℓ_t from S_{t-1} // e.g., unifsample
4 **if** $\beta_0(t-1) > \min_{k \in S_{t-1}} \beta_k(t-1)$ **or** $\xi_t < 0$ **then** $k_t \leftarrow 0$
5 **else** $k_t \leftarrow \ell_t$
6 Observe context $x_t \sim q$, show $a_t \sim \pi^{k_t}(.|x_t)$ and observe $r_t \sim \nu(a_t|x_t)$ // i.e., pull arm k_t and update confintervals with Lem53
7 $S_t \leftarrow \left\{k \in S_{t-1} : \overline{\mu}_k(t) > \underline{\mu}_0(t) + \epsilon\right\}$
8 **if** $\exists k \in S_t, \underline{\mu}_k(t) > \overline{\mu}_0(t)$ **then return** `envy`
9 **if** $S_t = \emptyset$ **then return** ϵ-`no-envy`
10 **end**

use μ_0 for the utility of the user for their policy (i.e., $u^m(\pi^m)$). Because the audit is a special form of bandit problem, following the bandit literature, an index of a user is called an *arm*, and arm 0 is the *baseline*.

Objectives and evaluation metrics We present our algorithm OCEF (Online Certification of Envy-Freeness) in the next subsection. Given $\epsilon > 0$ and $\alpha \geq 0$, OCEF returns either `envy` or ϵ-`no-envy` and has two objectives:

1. Correctness: if OCEF returns `envy`, then $\exists k, \mu_k > \mu_0$. If OCEF returns ϵ-`no-envy` then $\max_{k \in [\![K]\!]} \mu_k \leq \mu_0 + \epsilon$.
2. Recommendation performance: during the audit, OCEF must maintain a fraction $1-\alpha$ of the baseline performance. Denoting by $k_s \in \{0, \ldots, K\}$ the arm (group index) chosen at round s, this requirement is formalized as a conservative exploration constraint Wu et al. [2016]:

$$\forall t, \frac{1}{t} \sum_{s=1}^{t} \mu_{k_s} \geq (1-\alpha)\mu_0 \,. \tag{6.3}$$

We focus on the *fixed confidence* setting, where given a confidence parameter $\delta \in (0,1)$ the algorithm provably satisfies both objectives with probability $1-\delta$. In addition, there are two criteria to assess an online auditing algorithm:

1. Duration of the audit: the number of time-steps before the algorithm stops.
2. Cost of the audit: the cumulative loss of rewards incurred. Denoting the duration by τ, the cost is $\tau\mu_0 - \sum_{s=1}^{\tau} \mu_{k_s}$.
   It is possible that the cost is negative when there is envy. In that case, the audit increased recommendation performance by finding better recommendations for the group.

We note the asymmetry in the return statements of the algorithm: `envy` does not depend on ϵ. This asymmetry is necessary to obtain finite worst-case bounds on the duration and the cost of audit, as we see in Theorem 18.

Our setting had not yet been addressed by the pure exploration bandit literature, which mainly studies the identification of (ϵ-)optimal arms [Audibert and Bubeck, 2010]. Auditing for envy-freeness requires proper strategies in order to efficiently estimate the arm performances compared to the unknown baseline. Additionally, by making the cost of the audit a primary evaluation

criterion, we also bring the principle of conservative exploration to the pure exploration setting, while it had only been studied in regret minimization [Wu et al., 2016]. In our setting, conservative constraints involve nontrivial trade-offs between the duration and cost of the audit. We now present the algorithm, and then the theoretical guarantees for the objectives and evaluation measures.

6.4.3 The OCEF algorithm

OCEF is described in Alg. 4. It maintains confidence intervals on arm performances $(\mu_k)_{k=0}^K$. Given the confidence parameter δ, the lower and upper bounds on μ_k at time step t, denoted by $\underline{\mu}_k(t)$ and $\overline{\mu}_k(t)$, are chosen so that with probability at least $1-\delta$, we have $\forall k, t, \mu_k \in [\underline{\mu}_k(t), \overline{\mu}_k(t)]$. In the algorithm, $\beta_k(t) = (\overline{\mu}_k(t) - \underline{\mu}_k(t))/2$. As Jamieson et al. [2014], we use anytime bounds inspired by the law of the iterated logarithm. These are given in Lem. 53 in App. C.5.

OCEF maintains an active set S_t of all arms in $[\![K]\!]$ (i.e., excluding the baseline) whose performance are not confidently less than $\mu_0 + \epsilon$. It is initialized to $S_0 = [\![K]\!]$ (line 1). At each round t, the algorithm selects an arm $\ell_t \in S_t$ (line 3). Then, depending on the state of the conservative exploration constraint (described later), the algorithm pulls k_t, which is either ℓ_t or the baseline (lines 4-6). After observing the reward r_t, the confidence interval of μ_{ℓ_t} is updated, and all active arms that are confidently worse than the baseline plus ϵ are de-activated (line 7). The algorithm returns `envy` if an arm k is confidently better than the baseline (line 8), returns ϵ-`no-envy` if there are no more active arms, (line 9) or continues if neither of these conditions are met.

Conservative exploration To deal with the conservative exploration constraint (6.3), we follow Garcelon et al. [2020a]. Denoting $A_t = \{s \leq t : k_s \neq 0\}$ the time steps at which the baseline was not pulled, we maintain a confidence interval such that with probability $\geq 1-\delta$, we have $\forall t > 0, \left|\sum_{s\in A_t}(\mu_{k_s} - r_s)\right| \leq \Phi(t)$. The formula for Φ is given in Lem. 55 in App. C.5. This confidence interval is used to estimate whether the conservative constraint (6.3) is met at round t as follows. First, let us denote by $N_k(t)$ the number of times arm k has been pulled until t, and notice that (6.3) is equivalent to $\sum_{s\in A_t}\mu_{k_s} - ((1-\alpha)t - N_0(t))\mu_0 \geq 0$. After choosing ℓ_t (line 3), we use the lower bound on $\sum_{s\in A_t}\mu_{k_s}$ and the upper bound for μ_0 to obtain a conservative estimate of (6.3). Using $\tau = t-1$, this leads to:

$$\xi_t = \sum_{s\in A_\tau} r_s - \Phi(t) + \underline{\mu}_{\ell_t}(\tau) + (N_0(\tau) - (1-\alpha)t)\overline{\mu}_0(\tau)\,. \tag{6.4}$$

Then, as long as the confidence intervals hold, pulling ℓ_t does not break the constraint (6.3) if $\xi_t \geq 0$. The algorithm thus pulls the baseline arm when $\xi_t < 0$. To simplify the theoretical analysis, OCEF also pulls the baseline if it does not have the tightest confidence interval (lines 4-6).

6.4.4 Analysis

The main theoretical result of the paper is the following:

Theorem 18. *Let $\epsilon \in (0,1]$, $\alpha \in (0,1], \delta \in (0, \frac{1}{2})$ and $\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k)$ and $h_k = \max(1, \frac{1}{\eta_k})$. Using $\underline{\mu}, \overline{\mu}$ and Φ given in Lemmas 53 and 55 (App. C.5), OCEF achieves the following guarantees with probability $\geq 1-\delta$:*

- *OCEF is correct and satisfies the conservative constraint on the recommendation performance (6.3).*
- *The duration is in $O\left(\sum_{k=1}^K \frac{h_k \log\left(\frac{K\log(Kh_k/\delta\eta_k)}{\delta}\right)}{\min(\alpha\mu_0, \eta_k)}\right)$.*

- *The cost is in* $O\Big(\sum_{k:\mu_k<\mu_0} \frac{(\mu_0-\mu_k)h_k}{\eta_k} \log\big(\frac{K\log(Kh_k/\delta\eta_k))}{\delta}\big)\Big)$.

The important problem-dependent quantity η_k is the gap between the baseline and other arms k. It is asymmetric depending on whether the arm is better than the baseline $(\mu_k - \mu_0)$ or the converse $(\mu_0 - \mu_k + \epsilon)$ because the stopping condition for `envy` does not depend on ϵ. This leads to a worst case that only depends on ϵ, since $\eta_k = \max(\mu_k - \mu_0, \mu_0 - \mu_k + \epsilon) \geq \frac{\epsilon}{2}$, while if the condition was symmetric, we would have possibly unbounded duration when $\mu_k = \mu_0 + \epsilon$ for some $k \neq 0$. Overall, ignoring log terms, we conclude that when $\alpha\mu_0$ is large, the duration is of order $\sum_k \frac{1}{\eta_k^2}$ and the cost is of order $\sum_k \frac{1}{\eta_k}$. This becomes $\sum_k \frac{1}{\alpha\mu_0\eta_k}$ and $\sum_k \frac{1}{\eta_k}$ when $\alpha\mu_0$ is small compared to η_k. This means that the conservative constraint has an impact mostly when it is strict. It also means that when either $\alpha\mu_0 \ll \eta_k$ or $\eta_k^2 \ll \eta_k$ the cost can be small even when the duration is fairly high.

6.4.5 Full audit

Exact criterion To audit for envy-freeness on the full system, we apply OCEF to all M users simultaneously and with $K = M$, meaning that the set of arms corresponds to all the users' policies. By the union bound, using $\delta' = \frac{\delta}{M}$ instead of δ in OCEF's confidence intervals, the guarantees of Theorem 18 hold simultaneously for all users.

For recommender systems with large user databases, the duration of OCEF thus becomes less manageable as M increases. We show how to use OCEF to certify the probabilistic criterion with guarantees that do not depend on M.

Probabilistic criterion The AUDIT algorithm for auditing the full recommender system is described in Alg. 5. AUDIT samples a subset of users and a subset of arms for each sampled user. Then it applies OCEF to each user simultaneously with their sampled arms. It stops either upon finding an envious user, or when all sampled users are certified with ϵ-no envy. Again there is a necessary asymmetry in the return statements of AUDIT to obtain finite worst-case bounds whether or not the system is envy-free.

The number of target users $\tilde{M}$ and arms K in Alg. 5 are chosen so that ϵ-envy-freeness *w.r.t.* the sampled users and arms translates into $(\epsilon, \gamma, \lambda)$-envy-freeness. Combining these random approximation guarantees with Th. 18, we get:

Theorem 19. *Let* $\tilde{M} = \left\lceil \frac{\log(3/\delta)}{\lambda} \right\rceil$ *and* $K = \left\lceil \frac{\log(3\tilde{M}/\delta)}{\log(1/(1-\gamma))} \right\rceil$. *With probability* $1-\delta$, *AUDIT is correct, it satisfies the conservative constraint* (6.3) *for all* $\tilde{M}$ *target users, and the bounds on duration and cost from Th. 18 (using* $\frac{\delta}{3\tilde{M}}$ *instead of* δ*) are simultaneously valid.*

Importantly, in contrast to naively using OCEF to compare all users against all, the audit for the probabilistic relaxation of envy-freeness only requires to query a constant number of users and policies that *does not depend on the total number of users* M. Therefore, the bounds on duration and cost are also independent of M, which is a drastic improvement.

6.5 Experiments

We present experiments describing sources of envy (Sec. 6.5.1) and evaluating the auditing algorithm OCEF on two recommendation tasks (Sec. 6.5.2).

We create a music recommendation task based on the Last.fm dataset from Cantador et al. [2011], which contains the music listening histories of 1.9k users. We select the 2500 items most listened to, and simulate ground truth user preferences by filling in missing entries with a popular

Algorithm 5: AUDIT algorithm. The algorithm either outputs a probabilistic certificate of $(\epsilon, \gamma, \lambda)$-envy-freeness, or evidence of envy.

input : Confidence parameter δ, conservative exploration parameter α, envy parameters $(\epsilon, \gamma, \lambda)$
output : $(\epsilon, \gamma, \lambda)$-`envy-free` or `not-envy-free`

Draw a sample $\tilde{S}$ of $\tilde{M} = \left\lceil \frac{\log(3/\delta)}{\lambda} \right\rceil$ users from $[\![M]\!]$
for *each user* $m \in \tilde{S}$ *in parallel* **do**
  Sample $K = \left\lceil \frac{\log(3\tilde{M}/\delta)}{\log(1/(1-\gamma))} \right\rceil$ arms from $[\![M]\!] \setminus \{m\}$
  Run OCEF$\left(\frac{\delta}{3\tilde{M}}, \alpha, \epsilon\right)$ for user m with the K arms
  if *OCEF outputs* `envy` **then return** `not-envy-free`
end
return $(\epsilon, \gamma, \lambda)$-`envy-free`

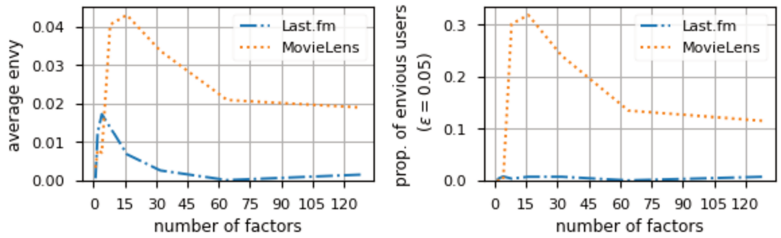


Figure 6.2: Envy from model mispecification on MovieLens and Lastfm: envy is high when the latent factor model is mispecified, but it decreases as the number of factors increases.

matrix completion algorithm for implicit feedback data[4]. We also address movie recommendation with the MovieLens-1M dataset Harper and Konstan [2015], which contains ratings of movies by real users, and from which we extract the top 2000 users and 2500 items with the most ratings. We binarize ratings by setting those < 3 to zero, and as for Last.fm we complete the matrix to generate ground truth preferences.

For both recommendation tasks, the simulated recommender system estimates relevance scores using low-rank matrix completion Bell and Sejnowski [1995] on a training sample of 70% of the ground truth preferences, where the rated / played items are sampled uniformly at random. Recommendations are given by a fixed-temperature *softmax* policy over the predicted scores. We generate binary rewards using a Bernoulli distribution with expectation given by our ground truth preferences.

6.5.1 Sources of envy

We consider two measures of the degree of envy. Denoting $\Delta^m = \max(\max_{n \in [\![M]\!]} u^m(\pi^n) - u^m(\pi^m), 0)$, these are:

- the average envy experienced by users: $\frac{1}{M} \sum_{m \in [\![M]\!]} \Delta^m$,
- the proportion of ϵ-envious users: $\frac{1}{M} \sum_{m \in [\![M]\!]} \mathbb{1}_{\{\Delta^m > \epsilon\}}$.

6.5.1.1 Envy from model mispecification

We demonstrate that envy arises from a standard recommendation model when the modeling assumptions are too strong. We vary the number of latent factors of the matrix completion model

[4]Using the Python library Implicit: `https://github.com/benfred/implicit` (MIT License).

	Last.fm		MovieLens	
	EUU	OPT	EUU	OPT
Total utility	1552	**1726**	1671	**1761**
Average envy	0.10	**0**	0.04	**0**
Prop. 0.05-envious	0.61	**0**	0.13	**0**

Table 6.1: Optimal policies with equal user utility penalty (EUU) vs. Unconstrained optimal policies (OPT), computed on ground truth preferences: EUU deteriorates total utility and creates envy between users.

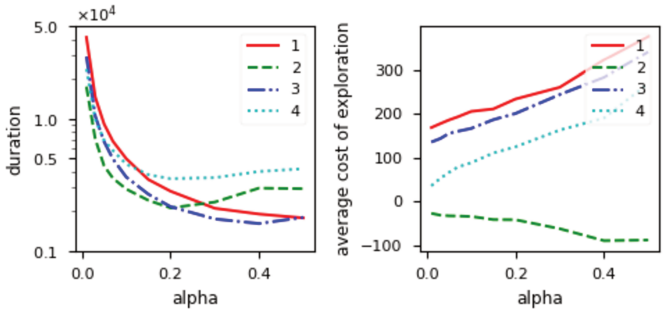


Figure 6.3: Effect of the conservative exploration parameter α on the duration and cost of auditing on Bandit experiments.

and evaluate a softmax policy with inverse temperature set to 5. In Fig. 6.2, with one latent factor we observe no envy. This is because all users receive the same recommendations since matrix completion is then equivalent to a popularity-based recommender system. With enough latent factors, preferences are properly captured by the model and the degree of envy decreases. For intermediate number of latent factors, envy is visible.

6.5.1.2 Envy from equal user utility

We show that in contrast to envy-freeness, enforcing equal user utility (EUU) degrades user satisfaction and creates envy between users. We compute optimal EUU policies and unconstrained optimal policies (OPT) on the ground truth preferences of Last.fm and MovieLens. Our results in Table 6.1 confirm the pitfalls of EUU, while illustrating that OPT policies are always envy-free.

We discuss more sources of envy and provide the details of these computations in App. C.3.

6.5.2 Evaluation of the auditing algorithm

Our goal is now to answer for OCEF: in practice, what is the interplay between the required sample size per user, the cost of exploration and the conservative exploration parameter?

6.5.2.1 Bandit experiments

We first study the trade-off between duration and cost of the audit on 4 bandit problems with Bernoulli rewards and 10 arms. In Problem 1, the baseline is the best arm and all other arms are equally bad. In Prob. 2, arm 1 is best and all other arms are as bad as the baseline. In Prob.3 the baseline is best and the means of arms from best to worst decrease rapidly. Prob. 4 uses the same means as Prob. 3, but the means of the baseline and arm 1 are swapped, making the baseline second-to-best. We set $\delta = \epsilon = 0.05$ and report results averaged over 100 trials. The details of the bandit configurations are given in Appendix C.4.1.

Figure 6.3 plots the duration and the cost of exploration (C.1) as a function of the conservative

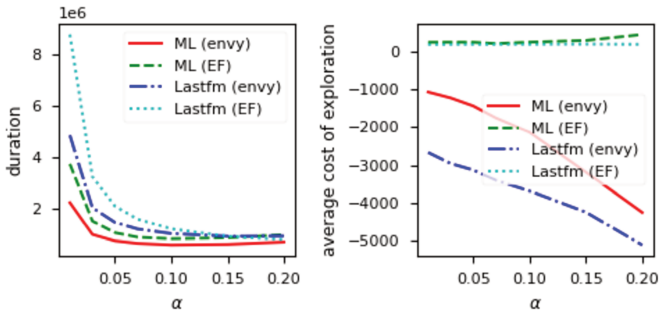


Figure 6.4: Scaling w.r.t. α on MovieLens (ML) and Last.fm, for recommender systems that are either envy-free (EF) or with envy. There are 41 target users and 75 arms.

constraint parameter α (smaller α means more conservative). The curves show that for Problems 2, 3, and 4, duration is minimal for a non-trivial α. This is because when α is large, all arms are pulled as much as the baseline, so their confidence intervals are similar. When α decreases, the baseline is pulled more, which reduces the length of the relevant confidence intervals $\beta_0(t) + \beta_k(t)$ for *all* arms k. This, in turn, shortens the audit because non-baseline arms are more rapidly discarded or declared better. When α becomes too small, however, the additional pulls of the baseline have no effect on $\beta_0(t) + \beta_k(t)$ because it is dominated by $\beta_k(t)$, so the duration only increases. This subtle phenomenon is not captured by our analysis (Th. 18), because the ratios $\beta_0(t)/\beta_k(t)$ are difficult to track formally.

The sign of the cost of exploration depends on whether there is envy. In Prob. 2 where the baseline has the worst performance, exploration is beneficial to the user and so the cost is negative. On all other instances however, the cost is positive. The cost of exploration is closest to 0 when α becomes small because then $\beta_0(t) + \beta_k(t)$ is the smallest possible for a given number of pulls of k. For instance, in Prob. 4, the cost is close to 0 when α is very small and increases with α. It is the case where the baseline is not the best arm but is close to it, and there are many bad arms. When the algorithm is very conservative, bad arms are discarded rapidly thanks to the good estimation of the baseline performance. In this "low-cost" regime however, the audit is significantly longer.

Appendix C.4.1 contains additional results when varying the number of arms and the confidence parameter δ.

6.5.2.2 MovieLens and Last.fm experiments

We now evaluate the certification of the (absence of) envy of recommendation policies on MovieLens (ML) and Last.fm. We consider two recommendation policies which are softmax functions over predicted relevance scores with inverse temperature set to either 5 or 10. These scores were obtained by matrix completion with 48 latent factors. On both datasets, with inverse temperature equal to 5, the softmax recommender system is envy-free, whereas there is envy when it is set to 10. We use AUDIT with OCEF to certify the probabilistic criterion. The envy parameters are set to $\epsilon = \delta = 0.05$ and $\lambda = \gamma = 0.1$, therefore we have $\tilde{M} = 41$ target users and $K = 75$ arms, independently on the number of users in each dataset.

The results of applying OCEF on each dataset (ML or Last.fm) with each policy (envy-free or with envy) are shown in Fig. 6.4. For the $(\epsilon, \gamma, \lambda)$-envy-free policies, results are averaged over 20 trials and over all the non-(ϵ, γ)-envious users, whereas when there is envy, results are averaged over the target users who are ϵ-envious. We observe clear tendencies similar to those of the previous section, although the exact sweet spots in terms of α depends on the specific configuration. In

particular, on envy-free configurations, the cost of the audit is positive and grows when relaxing the conservative constraint, while it is negative and decreasing with α when there is envy. More details are provided in App. C.4.2.

6.6 Conclusion

We proposed the audit of recommender systems for user-side fairness with the criterion of envy-freeness. The auditing problem requires an explicit exploration of user preferences, which leads to a formulation as a bandit problem with conservative constraints. We presented an algorithm for this problem and analyzed its performance experimentally. In future work, we plan to extend envy-freeness to heterogeneous groups of users, in order to generalize existing definitions of preference-based fairness to personalized predictions.

Chapter 7

Conclusion

Contents

7.1 Summary of contributions

This thesis makes both conceptual and algorithmic contributions.

In this thesis, we developed a conceptual framework based on distributive justice principles from social choice theory to assess the fairness of ranked recommendations. We approach recommendation as a fair allocation problem where the designer makes trade-offs between the utilities of the users and items. Within this framework, we proposed a principled approach to generate fair rankings by maximizing concave welfare functions of users' and items' utilities. In Chapter 3, we started with additive concave welfare functions, which encode the intuition of diminishing marginal utilities, and then treated in Chapter 4 the case of generalized Gini welfare functions, which have a more complex form but are more expressive. The perspective of social choice also gives a better understanding of existing ranking approaches, in which we show that popular merit-based approaches can lead to undesirable distributive unfairness (Chapter 3).

Along with the conceptual framework of this thesis, we made several algorithmic contributions, built around Frank-Wolfe methods. We addressed the challenge of optimizing concave functions of stochastic ranking policies, which can be used to express many objectives for fair and multi-objective recommendation. We first showed how to efficiently leverage Frank-Wolfe methods in the batch setting, for ranking in the position-based model in Chapter 3. Then we showed how to extend this approach to the case of the non-differentiable GGFs in Chapter 4. In Chapter 5, we addressed the problem of fair ranking in the contextual bandit setting, and presented the first bandit algorithm with regret guarantees for the problem. All the algorithms developed in this thesis are supported by theoretical guarantees on their convergence and complexity. We also evaluated our algorithms against relevant benchmarks on simulated environments based on public datasets such as MovieLens, Last.fm and Twitter data, which include up to $15k$ users and items.

In addition to proposing new methods for *designing* recommender systems that are fair towards

both users and items, we also addressed a different *auditing* problem, which is focused on *user-side fairness* in Chapter 6. Motivated by prominent audits for parity in the delivery of job ads, we propose an audit for envy-freeness, which provides more refined conclusions but is more technically challenging. We address this technical challenge by developing a sample-efficient pure exploration bandit algorithm for the task, that does not significantly degrade recommendation performance for the users sampled for the audit.

As we will discuss in the last section, our research leaves several open questions. These include a more detailed treatment of two-sided fairness at the group level, more general modeling of user and item utilities, and the incorporation of real-world dynamics that affect user and item preferences and behaviors. Additionally, while our work focuses on the perspective of fair division, the field of social choice offers valuable insights for the recommendation community that warrant further exploration. Addressing these challenging questions in conjunction with our contributions can lead to exciting research avenues. Despite the remaining open questions, our research has made significant strides in improving the current state-of-the-art in fairness for recommender systems. We have gained a better understanding of the limitations of equality and merit-based constraints on exposure, as well as how to design principled ranking objectives. Our results have led to the development of efficient algorithms that can be practically implemented, serving as a stepping stone towards the development of principled approaches to fairness in recommender systems in more complex settings. We hope that our work will inspire further progress in this field.

7.2 Discussion

In this final section, we discuss additional relevant topics that we did not include in the main body of this thesis, but to which we contributed. These topics are group fairness (Section 7.2.1) and other perspectives from social choice (Section 7.2.2). Then we discuss the limitations of our framework and open questions (Section 7.2.3), and the challenges of implementing fair recommender systems in practice (Section 7.2.4).

7.2.1 Towards group fairness

We described our framework for fair allocation of exposure at the level of individuals. Our framework can be extended to groups, following prior work on fair ranking which considered the utility of a group as the sum or the average of utilities of its members [Singh and Joachims, 2018, Morik et al., 2020, Singh and Joachims, 2019]. We provide the technical details of this extension in Appendix A.2, using the sum to aggregate utilities. In Appendix A.2, we define Lorenz efficiency at the level of group utilities, and show that maximizing additive concave welfare functions of group-level utilities yields Lorenz-efficient ranking policies. Note that this extension also allows to consider item-side fairness at the level of item *producers* instead of single items, by defining the utility of a producer as the sum or mean of their items' utilities.

However, this treatment of groups by adding up individual utilities is not the only method to assess fairness at the level of groups and has some limitations. In particular, it does not account for individual differences inside groups. Considerations for both individual differences within groups and redistribution between groups have been extensively studied in the economic literature on *equality of opportunity* [Roemer and Trannoy, 2016, Roemer, 1996], which has inspired several works on algorithmic fairness [Hardt et al., 2016b, Heidari et al., 2019, Arif Khan et al., 2022]. Another perspective is the economic literature on inequality measurement, where the decomposition of inequality into a within-group term and between-group term was studied through the property

of *additive decomposability* [Cowell, 2011], and was recently discussed in the context of fair machine learning [Speicher et al., 2018, Williamson and Menon, 2019]. The future implementation of these principles to two-sided fairness in recommendation are a promising extension of our efforts to integrate distributive justice principles into the assessment and design of recommender systems.

In all cases, fairness at the level of groups is predicated on access to a discrete sensitive attribute of users and/or items. Consequently, the practical application of group fairness notions is restricted by real-world constraints on the direct usage of sensitive attributes. Such restrictions exist when the sensitive attribute is not available, when collecting or inferring information about group membership is illegal, or when the delineation of groups into discrete categories is impractical or unethical [Tomasev et al., 2021, Andrus and Villeneuve, 2022]. Addressing group fairness without access to sensitive attributes is considered a key open problem for practical applications of fairness-aware measures and algorithms [Holstein et al., 2019, Veale and Binns, 2017, Andrus and Villeneuve, 2022, Kallus et al., 2022]. In a recent work [Liu et al., 2023], we leverage homophily in social networks to derive group fairness measures for recommender systems that do not rely on discrete group labels, while satisfying a notion of additive decomposability of inequality measures.

7.2.2 Opportunities of social choice for modern selection problems

Election problems Social choice problems fall into two broad categories: *public outcomes* (e.g., elections) and *private outcomes* (e.g., fair division) [Arrow et al., 2010, Donaldson and Weymark, 1988]. In this thesis, we focused on personalized recommender systems, in which the outcomes are private. Indeed, for users, the rankings are personal, and for items, the amount of exposure received by an item is not shared with other items. This motivated us to leverage fair division as a conceptual framework for personalized recommender systems. In non-personalized search engines and group recommender systems (e.g., lists of *"Trending topics"* or *"Top restaurants in Paris"*), the recommendations are the same for all users, and hence the outcome is public, from the perspective of users. In this *non-personalized* setting, concepts from fair *public decision-making* in social choice present interesting opportunities.

During the PhD program, we also made contributions to this branch of social choice where outcomes are public. In [Do et al., 2021a, 2022b], we addressed *committee elections*, a popular class of social choice problems where the public outcome is a subset of individuals elected from a larger pool of candidates [Lackner and Skowron, 2020]. In committee elections, fairness is often understood as a form of *proportional representation*, meaning that the elected committee is representative of the population of voters [Lackner and Skowron, 2020]. While proportionality is mainly considered with respect to the preferences of voters, a few recent works have considered representation based on demographic attributes [Lang and Skowron, 2018, Celis et al., 2017a, Bredereck et al., 2018].

The connection between voting problems and group recommender systems has already received significant attention in the computational social choice literature [Skowron et al., 2016a,b] and in the literature on diversity in information retrieval [Dang and Croft, 2012], and has been studied in fair recommendation more recently [Chakraborty et al., 2019, Allouah et al., 2022]. This connection is often made by casting users as voters and items as candidates. In Appendix D, we included a contribution made to the social choice literature, in which we address a specific committee election problem (see [Do et al., 2021a]). In that piece of work, we focused on designing algorithms for selecting committees that satisfy a proportional representation criterion with respect to multiple demographic attributes, in online settings.

Although we did not develop a formal connection between recommender systems and the committee selection problem addressed in Appendix D, several prominent concepts and tools

independently developed in the two fields are closely related. The proportionality criteria developed in the committee election literature are in fact similar to some diversity and fairness criteria developed in the recommender systems and information retrieval literature. For instance, *intent-based diversification* of search results consists in finding a set of items that covers the various intents behind a specific query (e.g., the query *"jaguar"* can have the animal or the car brand as intent) [Chapelle et al., 2011]. This problem can be seen as a voting problem where each item is a candidate and each intent is a voter. As a matter of fact, a few works on proportionality in committee elections also use query ambiguity in search engines as a motivating example [Skowron et al., 2016b].

Another example is the problem of proportional representation of political parties based on representation targets in *party-list elections* [Lang and Skowron, 2018]. Existing rules for electing a committee (i.e., an assembly) are closely related to some metrics proposed for fair ranking with respect to sensitive groups of items. The D'Hondt rule is mathematically similar to the KL-metric for fair ranking proposed in [Yang and Stoyanovich, 2017, Geyik et al., 2019], and the Hamilton rule is the ℓ_1 metric of [Yang and Stoyanovich, 2017]. Leveraging deeper connections between proportionality in committee elections and fairness and diversity in information retrieval is a promising avenue for future research.

Matching Matching problems are also widely studied in game theory and social choice, and fall in the category of private outcomes [Gale and Shapley, 1962]. In the fair machine learning literature, a few recent works explore fairness in matchings when a centralized matching algorithm uses noisy estimates of agents' merit as input [Castera et al., 2022, Devic et al., 2023].

While the conventional examples of matching problems in social choice are college admissions and hospital-resident matching, two-sided matching markets are widespread in online platforms for job search, dating and friend recommendation. In this thesis, we modelled these applications as ranking tasks in reciprocal recommender systems, because their main purpose is to filter profiles among an overloaded candidate space, in order to support users with limited attention. The rankings produced by the recommender system assist users in finding other users. Users then act autonomously to match with each other. For example, on a job search platform, a recruiter can decide to connect with a candidate that was recommended to them, and the candidate can accept or decline the invitation to connect. Unlike in more traditional matching problems such as college admissions, the matching itself is not computed by the algorithm.

Nonetheless, the matching literature is still relevant in the context of reciprocal recommender systems, since it focuses more explicitly on the actual capacity constraints of agents (e.g., the actual number of slots in a university program or the headcount of a recruiter), while recommender systems focus on their limited attention on the online platform.

7.2.3 Limitations of recommendation as fair allocation

7.2.3.1 Challenges of defining utilities

Challenges of measuring user preferences and utility. Our framework for fair allocation of exposure in recommender systems is based on careful definition and measurement of the users' preference values μ_{ij}. It does not address potential biases arising at other parts of the recommendation pipeline, such as in the learning stage. These include selection bias (e.g., users only give feedback on items that were recommended to them) [Marlin and Zemel, 2009], position bias (e.g., users tend to click on items that are shown first) [Craswell et al., 2008], and estimation bias (e.g., in the learning model used to produce estimates $\hat{\mu}_{ij}$) [Chen et al., 2023]. These can create feedback

loops that reinforce suboptimalities in the learning of μ_{ij} if exploration is not sufficient [Bottou et al., 2013].

Beyond the challenges of producing unbiased engagement predictions, measuring the true values of items to users is a fundamentally difficult task because of the lack of observable ground truth. In our experimental analysis, we followed the common practice of online platforms which rely on engagement signals such as clicks, likes, and play counts to measure the values μ_{ij}, which are the main signals available at large scale. However, there may be a mismatch between these engagement signals and the true unobservable user preferences. Furthermore, those signals differ in strength, e.g. a like is probably more informative of a user's preference than a click. Several methods have been proposed by researchers in academia and industry to overcome these challenges. These include the use of surrogates of long-term user value [Wang et al., 2022], and of measurement theory in social sciences to provide a more principled approach to measuring value from existing signals [Milli et al., 2021, Jacobs and Wallach, 2021]. Other works proposed psychologically-grounded models of user preferences and behaviour [Curmei et al., 2022, Kleinberg et al., 2022].

Moreover, we assumed a stationary model of user preferences μ_{ij}. This stationarity assumption ignores the feedback loops involved in recommender sytems. These include feedback loops caused by the impact of the recommender system on users' preferences themselves [Adomavicius et al., 2013, Kalimeris et al., 2021, Carroll et al., 2022, Jiang et al., 2019, Warlop et al., 2018], and the patterns of consumption used to estimate them [Anderson et al., 2020]. An interesting direction would be the incorporation in our work of some recently proposed dynamic models of user preferences in recommender systems [Dean and Morgenstern, 2022, Curmei et al., 2022, Jiang et al., 2019].

Note that fair classification problems also suffer from measurement issues [Suresh and Guttag, 2019, Corbett-Davies and Goel, 2018, Kilbertus et al., 2020, Kleinberg et al., 2018a]. In the bank loan example of Section 1.2.2, repaid/default outcomes are only observed for individuals whose loan application was accepted. Kilbertus et al. [2020] emphasize the importance of focusing on fair decisions rather than on predictions, especially in settings where data availability depends on past decisions. We take a similar stance for recommendation problems by focusing on the fairness of rankings, even in the presence of imperfect measures of μ_{ij}.

Challenges of defining item producer utility. Following the academic literature on fairness of exposure [Kletti et al., 2022b, Singh and Joachims, 2018, Biega et al., 2018, Diaz et al., 2020], we identified item producers with their items and defined the item utility as the expected number of views received by an item, i.e., its exposure. In order to define an item producer's utility, we suggest to follow the TREC fair ranking track [Biega et al., 2020], wherein the producer's utility is the cumulative utility of their items. These modeling choices aim at simplifying the formal framework and the presentation of our approach and results.

The definition of an item's utility as its expected number of views comes with certain limitations that need to be considered carefully in real-world contexts. There are various settings where what item producers seek is not mere exposure but active engagement with their content. These include streaming platforms, where artists value the number of playcounts, or social media platforms, where people seek user interaction in the form of likes and shares. In such situations, it could be more apt to define an item's utility as the expected positive engagement received (i.e., number of likes), as opposed to just views. This aligns with the concept of utility as "impact" proposed by Saito and Joachims [2022], as well as the conditions for long-term sustainability of item producers considered in [Mladenov et al., 2020, Zhan et al., 2021].

However, there are also circumstances where the expected exposure is a reasonable proxy for item-side utility. Consider a business such as a shop or restaurant listed on a mapping application

like Google Maps. In such a scenario, the number of views could be a satisfactory measure of item utility, as users cannot engage more beyond views at the recommendation stage to contribute to the establishment's success. These variations highlight the difficulty of crafting a universally applicable measure of item utility, and the importance of taking the application context into account.

Transitioning the definition of item utility from views to engagement does not poses significant changes on the algorithmic side. Our Frank-Wolfe algorithms can still be used to efficiently optimize concave welfare functions of users' and items' utilities, since the item utility still has a similar linear form as in the case of utility-as-exposure. In fact, this definition of item utility bears resemblance to our notion of utility in the reciprocal recommendation setting, where our Frank-Wolfe variants can also be used in an efficient way.

The shift from views to engagement more importantly alters the implications of fairness for items. The redistribution of engagement among items, as opposed to redistributing exposure, necessitates a significant boost for items that have no relevance to the majority of users. Moreover, the pursuit of engagement equality among items, as opposed to exposure, could impose a substantial burden on the user side. This has been highlighted by LinkedIn's research [Basu et al., 2020], demonstrating that it may inadvertently lead to the intensified recommendation of less relevant items. In our approach, where we promote trade-offs over strict equality constraints, such a change would necessitate careful adjustment of the trade-off parameters, in order to decrease inequality in item utilities at a reasonable cost for user welfare. Consequently, when item utilities are defined in terms of engagement rather than exposure, it may be appropriate to consider alternative notions of fairness beyond mere redistribution. For instance, Saito and Joachims [2022] advocates for envy-freeness as a criterion for item-side fairness, when engagement metrics take precedence over views. This choice avoids the degenerate behaviour of redistributing engagement across items.

The definition of utilities requires a nuanced understanding of what users and item producers' actually value, while also keeping in mind the consequences of different definitions on the recommender systems' stakeholders.

7.2.3.2 Fixed exposure in the position-based model

We followed the literature on fairness of exposure which defines exposure in the position-based user model (PBM) [Singh and Joachims, 2018, Sapiezynski et al., 2019, Biega et al., 2018, Zehlike and Castillo, 2020], where the probability that a user observes an item only depends on its rank. Fairness of exposure in cascade models [Craswell et al., 2008] and in more general dynamic bayesian network models [Chapelle and Zhang, 2009] is more challenging from an algorithmic perspective. Indeed, these general exposure models do not have the linear structure of the PBM which enabled linear programming formulations [Singh and Joachims, 2018] and the computational efficiency of the Frank-Wolfe-based algorithms developed in this thesis. While the PBM is still widely used for its manageability and the (normalized) DCG metric [Järvelin and Kekäläinen, 2002], there has been interest in evaluating fairness of exposure in rankings in cascade models, as in the TREC 2019 fair ranking track [Biega et al., 2020]. It is only recently that an algorithm with optimality guarantees was proposed for fair trade-offs in general dynamic bayesian network models [Kletti et al., 2022b].

From a fair division perspective, cascade models challenge the notion of exposure as a fixed quantity to allocate. In the PBM, the total exposure is fixed and equal to $E = n \|\boldsymbol{b}\|_1$ where n is the number of users b_k is the weight associated to the rank k. In practice though, the system has an impact on the actual budget of exposure to allocate, since the number of viewed items varies dynamically depending on whether the user keeps browsing the ranked list. In cascade models, the rank at which a user stops browsing depends on whether items ranked at higher positions are

relevant to the user. Therefore, the number of exposed items depends on the ranking (*via* the interdependence of the values of ranked items).[1]

It is not clear how to assess the fairness of the allocation of exposure when exposure is a dynamic quantity. In theory, it is possible to increase the total exposure available by showing irrelevant items in the highest positions of the ranking, in order to keep a patient user captive and have more exposure to give to small items. In practice though, the user's patience is likely to decrease in the long run. Jeunen and Goethals [2021], who propose a heuristic for item-side fairness in a cascade model, make a similar observation in their experiments. They find that shuffling the items in the first positions yields different user utility/item inequality trade-offs, depending on the user's openness to randomization. We also suspect that the effectiveness of fairness-aware ranking policies in cascade models depends on how much patience users have, and this likely varies depending on their satisfaction from past recommendations.

7.2.3.3 Fairness beyond fair division

Fairness is a complex, multi-faceted, contextual and much debated upon concept, and fair division is only *one* way to frame it. We focused on the perspective of fair division because of its historical importance, its strong foundations in decades of research in social choice, and its relevance to the problem of making trade-offs between the interests of the stakeholders of recommender systems.

However, there are many other ways to frame fairness which could help improve recommender systems. In Section 7.2.1, we suggested that other economic models of distributive justice, such as the theory of equality of opportunity, could provide a better treatment of groups of items and users in recommender systems. Fair division is historically not concerned with groups, and does not explicitly address the historical and societal disadvantage of social groups [Moulin, 2003]. In contrast, in the theory of equality of opportunity of Roemer [1996], outcomes are redistributed after correcting for arbitrary circumstances that are not in the control of individuals (e.g., race or socio-economic background). Several works proposed to connect doctrines of equality of opportunity with group-level fairness notions in machine learning [Heidari et al., 2019, Zehlike et al., 2022a, Arif Khan et al., 2022], but the adaptation of these frameworks to user-side and item-side fairness in recommender systems remains an open issue.

Finally, distributive justice is one fundamental axis of theories of justice that has been considered as separate and complementary to recognition justice [Fraser and Honneth, 2003]. In the context of fair machine learning, this distinction has been discussed in terms of distributive harms and representational harms [Binns, 2017, Barocas et al., 2017]. Distributive harms are about the fair distribution of outcomes of machine learning applications, while representational harms are concerned with biases and stereotypes in learned representations, such as gendered associations in word embeddings [Bolukbasi et al., 2016]. In the context of recommender systems, this thesis deals with the former axis, through explicit anchoring in distributive justice principles. However, fairness in rankings as in [Zehlike et al., 2017, Yang and Stoyanovich, 2017, Geyik et al., 2019, Singh and Joachims, 2018] is often motivated by the mitigation of representational harms [Binns, 2017]. The difference between our work and these works on fair ranking is that we consider the distribution of exposure across the lists of multiple users, while they consider exposure within a single ranked list. Fair ranking *within lists* can be seen as a way to promote diverse representation within a

[1] Using the notation of Chapter 1: In the cascade model with weights $\boldsymbol{b}$, given a *deterministic* ranking policy P, the exposure of an item j is: $v_j(P) = \sum_{i=1}^{n}\sum_{k=1}^{m} P_{ijk} b_k \prod_{r<k}(1-\mu_i^\intercal P_{i:r})$. The total exposure is: $E(P) = \sum_{i=1}^{n}\sum_{k=1}^{m} b_k \prod_{r<k}(1-\mu_i^\intercal P_{i:r})$, where $\mu_i^\intercal P_{i:r}$ is the value of the item at rank r for user i in the ranking policy P. The total exposure to allocate thus depends on the ranking policy P.

ranking, by matching group-based representation targets [Zehlike et al., 2022a]. This brings the purpose of fair ranking within lists closer to topic-based diversification [Zhai et al., 2015, Ziegler et al., 2005], although they have been considered as separate problems in the literature [Burke et al., 2018, Zehlike et al., 2017, Yang et al., 2019]. In our case (as well as in [Biega et al., 2018, Kletti et al., 2022a, Diaz et al., 2020, Morik et al., 2020]), we provide exposure guarantees for (groups of) items across lists, because the utility that an item derives from a recommender system is its overall exposure. These guarantees of fair distribution of utility in the overall recommender system do not *a priori* translate into guarantees of fair representation or diversity within each of the lists. This makes existing work on within-list fairness and diversity complementary to ours, from a conceptual perspective. From an algorithmic perspective, our approach deals with the within-list setting simply by considering a separate objective function for each user (as explained in Appendix A.8).

7.2.4 Practical challenges of real-world recommender systems

Several industry practitioners highlighted the challenges of integrating academic research on fairness into production systems [Bakalar et al., 2021, Holstein et al., 2019, Beutel et al., 2019b]. This section discusses some of the remaining gaps between the normative question addressed in this thesis ("how should the system trade-off between the interests of users and item producers?") and the practical challenges of implementing and evaluating algorithms that respond to it.

Dynamics of recommender systems In the framework of this thesis, we assumed that several aspects of the environment were static, such as the set of items (or item producers) and the preferences and engagement patterns of users. These assumptions can be challenging in practice because they ignore the impact of the recommender system on the environment. We previously mentioned the effect of recommendations on user preferences in Section 7.2.3. Recommendations also impact users' perception of the platform, affecting user retention in the long run through complex mechanisms. It is thus important to consider this long-term impact when designing fairness-aware recommendation strategies, through models of leaving/returning behaviour [Wu et al., 2017, Jing and Smola, 2017, Ben-Porat et al., 2022, Chandar et al., 2022], surrogate measures of long-term user experience [Wang et al., 2022], and models of user trust [Cen et al., 2022]. Some works also discussed how recommendations affect the long-term dynamics of content production [Mladenov et al., 2020, Zhan et al., 2021]. Recent work also proposed game-theoretic frameworks for recommender systems, which model the strategic behaviour of item producers [Hron et al., 2023, Ben-Porat and Tennenholtz, 2018].

Because of feedback loops, static fairness interventions can fail to improve global welfare in the long run [Akpinar et al., 2022, Peysakhovich et al., 2023]. A few works use reinforcement learning to account for the recommendations' impact on the environment with long-term fairness constraints [Ge et al., 2021, Yu et al., 2022]. An interesting direction would be to build upon simulation studies and open-sourced environments that have been proposed to model feedback loops and long-term effects in recommender systems [Ie et al., 2019, Yao et al., 2021, Krauth et al., 2020, Huang et al., 2020, Rohde et al., 2018, Bountouridis et al., 2019, Zhan et al., 2021]. Another possible direction is the use of causal inference methods to tackle feedback loops in recommender systems [Bottou et al., 2013, Schnabel et al., 2016, Sinha et al., 2016, Wang et al., 2020, Krauth et al., 2022].

Multi-stage recommendation pipelines Real-world recommender systems are part of pipelines that are more complex than the one described in Section 1.2.1. As documented by existing platforms [Twitter, 2023, YouTube, 2021, Instagram, 2022], those pipelines include more components, such as

a candidate sourcing stage, where a few thousand recent items are extracted from a pool of hundreds of million items, before the learning stage. Several recent works have studied the interaction of multiple components in a multi-stage pipeline [Hron et al., 2021, Wang et al., 2021b], emphasizing the implications of unfair candidate sourcing on the ranking stage [Wang and Joachims, 2023, Bower et al., 2022].

Choice of trade-off in practice In the fair ranking problem, the designer decides on a specific welfare function F to optimize. In practice, this can be accomplished by varying hyperparameters within a predefined class of welfare objectives (i.e., by varying λ and the hyperparameters of g^{user}, g^{item} in Eq. (1.2)). The task of choosing a trade-off between different metrics in recommender systems is a general problem that practitioners face [Kohavi et al., 2009, Gunawardana et al., 2012]. The gold standard for evaluating and choosing a recommendation algorithm based on overall evaluation criteria (OEC) is the use of online controlled experiments [Kohavi et al., 2009], which must be carefully designed with awareness of their social and ethical implications [Bird et al., 2016].

Beyond the fairness trade-offs involving users and items that we specifically address in this thesis, other trade-offs, objectives, and stakeholders are also relevant in the design of recommender systems. Practitioners must consider the interests of the platform itself: For example, when revenue is drawn from advertising, the trade-off between revenue and user experience is a common concern [l'Ecuyer et al., 2017]. Platform policies and regulations also require compliance with additional ethical and regulatory principles, such as privacy [McSherry and Mironov, 2009] and integrity [Kalimeris et al., 2021, Facebook, 2020, YouTube, 2021]. Moreover, the overall performance of recommender systems that drive the choice of an algorithm is often measured by OECs which are more focused on long-term goals, such as daily or monthly active users. These metrics are typically prioritized over offline metrics like DCG, that we use to measure user utility in this thesis [Kohavi et al., 2012].

The task of balancing multiple OECs in recommender systems is akin to a macroeconomic problem. The framework developed in this thesis focuses on the microeconomic problem of deciding which users get to see which items. It is one piece of the bigger picture: The overall performance of the system is the result of the interaction between the microeconomic decisions and the macroeconomic dynamics of the system.

Bibliography

Yasin Abbasi-Yadkori, Dávid Pál, and Csaba Szepesvári. Improved algorithms for linear stochastic bandits. *Advances in neural information processing systems*, 24, 2011. 75, 82, 181, 182, 188, 189

Himan Abdollahpouri and Robin Burke. Multi-stakeholder recommendation and its connection to multi-sided fairness. *arXiv preprint arXiv:1907.13158*, 2019. 21

Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. Beyond personalization: Research directions in multistakeholder recommendation. *arXiv preprint arXiv:1905.01986*, 2019a. 74

Himan Abdollahpouri, Masoud Mansoury, Robin Burke, and Bamshad Mobasher. The unfairness of popularity bias in recommendation. *arXiv preprint arXiv:1907.13286*, 2019b. 4, 21, 41, 50, 54, 69, 242

Himan Abdollahpouri, Gediminas Adomavicius, Robin Burke, Ido Guy, Dietmar Jannach, Toshihiro Kamishima, Jan Krasnodebski, and Luiz Pizzato. Multistakeholder recommendation: Survey and research directions. *User Modeling and User-Adapted Interaction*, 30(1):127–158, 2020. 3, 20, 21, 22, 69, 240

Ryan Prescott Adams and Richard S Zemel. Ranking via sinkhorn propagation. *arXiv preprint arXiv:1106.1925*, 2011. 27, 70

Gediminas Adomavicius, Jesse C Bockstedt, Shawn P Curley, and Jingjing Zhang. Do recommender systems manipulate consumer preferences? a study of anchoring effects. *Information Systems Research*, 24(4):956–975, 2013. 105

Alekh Agarwal, Dean P Foster, Daniel J Hsu, Sham M Kakade, and Alexander Rakhlin. Stochastic convex optimization with bandit feedback. *Advances in Neural Information Processing Systems*, 24, 2011. 29, 163

Alekh Agarwal, Alina Beygelzimer, Miroslav Dudík, John Langford, and Hanna Wallach. A reductions approach to fair classification. *arXiv preprint arXiv:1803.02453*, 2018. 7, 245

Shivani Agarwal. Surrogate regret bounds for bipartite ranking via strongly proper losses. *The Journal of Machine Learning Research*, 15(1):1653–1674, 2014. 145

Shipra Agrawal and Nikhil Devanur. Linear contextual bandits with knapsacks. *Advances in Neural Information Processing Systems*, 29, 2016. 84, 217

Shipra Agrawal and Nikhil R Devanur. Bandits with concave rewards and convex knapsacks. In *Proceedings of the fifteenth ACM conference on Economics and computation*, pages 989–1006, 2014. 16, 27, 28, 29, 74, 75, 80, 82, 163, 173, 175, 250

Shipra Agrawal, Nikhil R Devanur, and Lihong Li. An efficient algorithm for contextual bandits with knapsacks, and an extension to concave objectives. In *Conference on Learning Theory*, pages 4–18. PMLR, 2016. 16, 28, 75, 163, 250, 251

Nil-Jana Akpinar, Cyrus DiCiccio, Preetam Nandy, and Kinjal Basu. Long-term dynamics of fairness intervention in connection recommender systems. In *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, AIES '22, page 22–35, New York, NY, USA, 2022. Association for Computing Machinery. ISBN 9781450392471. doi: 10.1145/3514094.3534173. URL https://doi.org/10.1145/3514094.3534173. 108

Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. Discrimination through optimization: How facebook's ad delivery can lead to biased outcomes. *Proceedings of the ACM on human-computer interaction*, 3(CSCW):1–30, 2019. 5, 17, 20, 85, 243, 253

Youssef Allouah, Rachid Guerraoui, Lê-Nguyên Hoang, and Oscar Villemaud. Robust sparse voting. *arXiv preprint arXiv:2202.08656*, 2022. 103

Eitan Altman. *Constrained Markov decision processes*, volume 7. CRC Press, 1999. 217, 218

Ashton Anderson, Lucas Maystre, Ian Anderson, Rishabh Mehrotra, and Mounia Lalmas. Algorithmic effects on the diversity of consumption on spotify. In *Proceedings of The Web Conference 2020*, pages 2155–2165, 2020. 105

McKane Andrus and Sarah Villeneuve. Demographic-reliant algorithmic fairness: characterizing the risks of demographic data collection in the pursuit of fairness. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 1709–1721, 2022. 103

Falaah Arif Khan, Eleni Manis, and Julia Stoyanovich. Towards substantive conceptions of algorithmic fairness: Normative guidance from equal opportunity doctrines. In *Equity and Access in Algorithms, Mechanisms, and Optimization*, pages 1–10. 2022. 102, 107

Oihana Aristondo, José García-Lapresta, Casilda Lasso de la Vega, and Ricardo Pereira. Classical inequality indices, welfare and illfare functions, and the dual decomposition. *Fuzzy Sets and Systems*, 228, 10 2013. doi: 10.1016/j.fss.2013.02.001. 36

Richard J Arneson. Luck egalitarianism and prioritarianism. *Ethics*, 110(2):339–349, 2000. 23

Kenneth J Arrow, Amartya Sen, and Kotaro Suzumura. *Handbook of social choice and welfare*, volume 2. Elsevier, 2010. 7, 30, 103, 246

Kenneth Joseph Arrow. *Social Choice and Individual Values*. New York, NY, USA: Wiley: New York, 1951. 9

Joshua Asplund, Motahhare Eslami, Hari Sundaram, Christian Sandvig, and Karrie Karahalios. Auditing race and gender discrimination in online housing markets. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 14, pages 24–35, 2020. 20, 87

Abolfazl Asudeh, HV Jagadish, Julia Stoyanovich, and Gautam Das. Designing fair ranking schemes. In *Proceedings of the 2019 International Conference on Management of Data*, pages 1259–1276, 2019. 21

Anthony B Atkinson. On the measurement of inequality. *Journal of economic theory*, 2(3):244–263, 1970. 33, 36, 43, 57, 58, 59

Anthony B Atkinson, Andrea Brandolini, and Hugh Dalton. Unveiling the ethics behind inequality measurement: Dalton's contribution to economics. *The Economic Journal*, pages 209–234, 2015. 32

Jean-Yves Audibert and Sébastien Bubeck. Best arm identification in multi-armed bandits. 2010. 29, 30, 88, 93

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multiarmed bandit problem. *Machine learning*, 47(2-3):235–256, 2002. 29

Haris Aziz. A rule for committee selection with soft diversity constraints. *Group Decision and Negotiation*, 28:1193–1200, 2019. 214

Moshe Babaioff, Nicole Immorlica, David Kempe, and Robert Kleinberg. Online auctions and generalized secretary problems. *ACM SIGecom Exchanges*, 7(2):1–11, 2008. 215

Ashwinkumar Badanidiyuru, Baharan Mirzasoleiman, Amin Karbasi, and Andreas Krause. Streaming submodular maximization: Massive data summarization on the fly. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 671–680, 2014. 215

Chloé Bakalar, Renata Barreto, Stevie Bergman, Miranda Bogen, Bobbie Chern, Sam Corbett-Davies, Melissa Hall, Isabel Kloumann, Michelle Lam, Joaquin Quiñonero Candela, et al. Fairness on the ground: Applying algorithmic fairness approaches to production systems. *arXiv preprint arXiv:2103.06172*, 2021. 108

Marina-Florica Balcan, Travis Dick, Ritesh Noothigattu, and Ariel D. Procaccia. Envy-free classificatoion. *arXiv preprint arXiv:1809.08700*, 2018. 37, 38, 51, 90

Solon Barocas and Andrew D Selbst. Big data's disparate impact. *Calif. L. Rev.*, 104:671–769, 2016. 3, 19, 20, 88, 240

Solon Barocas, Kate Crawford, Aaron Shapiro, and Hanna Wallach. The problem with bias: From allocative to representational harms in machine learning'. In *SIGCIS conference paper*, 2017. 107

Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and Machine Learning*. fairmlbook.org, 2018. http://www.fairmlbook.org. 90

Solon Barocas, Moritz Hardt, and Arvind Narayanan. *Fairness and Machine Learning: Limitations and Opportunities*. fairmlbook.org, 2019. http://www.fairmlbook.org. 3, 5, 19, 240, 243

Peter L Bartlett, Michael I Jordan, and Jon D McAuliffe. Convexity, classification, and risk bounds. *Journal of the American Statistical Association*, 101(473):138–156, 2006. 145

Kinjal Basu, Cyrus DiCiccio, Heloise Logan, and Noureddine El Karoui. A framework for fairness in two-sided marketplaces. *arXiv preprint arXiv:2006.12756*, 2020. 21, 41, 46, 47, 51, 68, 69, 71, 106, 160, 161

Mohammad Hossein Bateni, Mohammadtaghi Hajiaghayi, and Morteza Zadimoghaddam. Submodular secretary problem and extensions. *ACM Transactions on Algorithms (TALG)*, 9(4):1–23, 2013. 215

Amir Beck and Marc Teboulle. Smoothing and first order methods: A unified framework. *SIAM Journal on Optimization*, 22(2):557–580, 2012. 62

Xiaohui Bei, Shengxin Liu, Chung Keung Poon, and Hongao Wang. Candidate selections with proportional fairness constraints. In *Proceedings of the 19th International Conference on Autonomous Agents and Multiagent Systems, AAMAS '20, Auckland, New Zealand, May 9-13, 2020*, pages 150–158, 2020. 214

Anthony J Bell and Terrence J Sejnowski. An information-maximization approach to blind separation and blind deconvolution. *Neural computation*, 7(6):1129–1159, 1995. 96, 198

Omer Ben-Porat and Moshe Tennenholtz. A game-theoretic approach to recommendation systems with strategic content providers. In *Advances in Neural Information Processing Systems*, pages 1110–1120, 2018. 108

Omer Ben-Porat, Lee Cohen, Liu Leqi, Zachary C Lipton, and Yishay Mansour. Modeling attrition in recommender systems with departing bandits. *arXiv preprint arXiv:2203.13423*, 2022. 108

Gerdus Benadè, Paul Gölz, and Ariel D Procaccia. No stratification without representation. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 281–314, 2019. 214

Quentin Berthet and Vianney Perchet. Fast rates for bandit optimization with upper-confidence frank-wolfe. *Advances in Neural Information Processing Systems*, 30, 2017. 27, 29, 75, 79, 163, 177

Michael J Best, Nilotpal Chakravarti, and Vasant A Ubhaya. Minimizing separable convex functions subject to simple chain constraints. *SIAM Journal on Optimization*, 10(3):658–672, 2000. 62

Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Li Wei, Yi Wu, Lukasz Heldt, Zhe Zhao, Lichan Hong, Ed H Chi, et al. Fairness in recommendation ranking through pairwise comparisons. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2212–2220, 2019a. 6, 21, 27, 163, 245

Alex Beutel, Jilin Chen, Tulsee Doshi, Hai Qian, Allison Woodruff, Christine Luu, Pierre Kreitmann, Jonathan Bischof, and Ed H Chi. Putting fairness principles into practice: Challenges, metrics, and improvements. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 453–459, 2019b. 108

Asia J Biega, Krishna P Gummadi, and Gerhard Weikum. Equity of attention: Amortizing individual fairness in rankings. In *The 41st international acm sigir conference on research & development in information retrieval*, pages 405–414, 2018. 8, 10, 21, 25, 26, 28, 41, 42, 45, 47, 50, 54, 56, 57, 58, 64, 69, 74, 88, 90, 91, 105, 106, 108, 154, 164, 166

Asia J Biega, Fernando Diaz, Michael D Ekstrand, and Sebastian Kohlmeier. Overview of the trec 2019 fair ranking track. *arXiv preprint arXiv:2003.11650*, 2020. 10, 22, 105, 106

Reuben Binns. Fairness in machine learning: Lessons from political philosophy. *arXiv preprint arXiv:1712.03586*, 2017. 107

Sarah Bird, Solon Barocas, Kate Crawford, Fernando Diaz, and Hanna Wallach. Exploring or exploiting? social and ethical implications of autonomous experimentation in ai. In *Workshop on Fairness, Accountability, and Transparency in Machine Learning*, 2016. 84, 109

Garrett Birkhoff. *Lattice theory*, volume 25. American Mathematical Soc., 1940. 15, 26, 63, 248

Mathieu Blondel, Olivier Teboul, Quentin Berthet, and Josip Djolonga. Fast differentiable sorting and ranking. In *International Conference on Machine Learning*, pages 950–959. PMLR, 2020. 27, 61, 62, 70

Miranda Bogen, Pushkar Tripathi, Aditya Srinivas Timmaraju, Mashayekhi Mehdi, Zeng Qi, Roudani Rabyd, Gahagan Sean, Howard Andrew, and Leone Isabella. Toward fairness in personalized ads. Technical report, Meta, 2023. 17, 86, 252

Tolga Bolukbasi, Kai-Wei Chang, James Y Zou, Venkatesh Saligrama, and Adam T Kalai. Man is to computer programmer as woman is to homemaker? debiasing word embeddings. In *Advances in Neural Information Processing Systems*, pages 4349–4357, 2016. 107

Carlo E Bonferroni. *Elementi di statistica generale*. Universitacommerciale Bocconi, 1941. 36

Léon Bottou, Jonas Peters, Joaquin Quiñonero-Candela, Denis X Charles, D Max Chickering, Elon Portugaly, Dipankar Ray, Patrice Simard, and Ed Snelson. Counterfactual reasoning and learning systems: The example of computational advertising. *Journal of Machine Learning Research*, 14 (11), 2013. 71, 84, 105, 108

Léon Bottou, Frank E Curtis, and Jorge Nocedal. Optimization methods for large-scale machine learning. *Siam Review*, 60(2):223–311, 2018. 170, 177

Dimitrios Bountouridis, Jaron Harambam, Mykola Makhortykh, Mónica Marrero, Nava Tintarev, and Claudia Hauff. Siren: A simulation framework for understanding the effects of recommender systems in online news environments. In *Proceedings of the conference on fairness, accountability, and transparency*, pages 150–159, 2019. 108

S. Bouveret, Y. Chevaleyre, and N. Maudet. Fair allocation of indivisible goods. In *Handbook of Computational Social Choice*, 2016. 30

Amanda Bower, Hamid Eftekhari, Mikhail Yurochkin, and Yuekai Sun. Individually fair rankings. In *International Conference on Learning Representations*, 2021. 20, 28

Amanda Bower, Kristian Lum, Tomo Lazovich, Kyra Yee, and Luca Belli. Random isn't always fair: Candidate set imbalance and exposure inequality in recommender systems. *arXiv preprint arXiv:2209.05000*, 2022. 109

Robert Bredereck, Piotr Faliszewski, Ayumi Igarashi, Martin Lackner, and Piotr Skowron. Multiwinner elections with diversity constraints. In *Thirty-Second AAAI Conference on Artificial Intelligence*, 2018. 103, 214

Sébastien Bubeck and Nicolo Cesa-Bianchi. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *arXiv preprint arXiv:1204.5721*, 2012. 29, 172, 175

Sébastien Bubeck, Rémi Munos, Gilles Stoltz, and . Pure exploration in multi-armed bandits problems. In *International conference on Algorithmic learning theory*, pages 23–37. Springer, 2009. 87, 88

Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In *Conference on Fairness, Accountability and Transparency*, pages 77–91, 2018. 3, 240

Robin Burke. Hybrid recommender systems: Survey and experiments. *User modeling and user-adapted interaction*, 12(4):331–370, 2002. 197

Robin Burke. Multisided fairness for recommendation. *arXiv preprint arXiv:1707.00093*, 2017. 50, 69, 88, 90

Robin Burke, Nasim Sonboli, and Aldo Ordonez-Gauger. Balanced neighborhoods for multi-sided fairness in recommendation. In *Conference on fairness, accountability and transparency*, pages 202–214. PMLR, 2018. 20, 21, 108

Richard V Burkhauser, Shuaizhang Feng, and Stephen P Jenkins. Using the p90/p10 index to measure us inequality trends with current population survey data: A view from inside the census bureau vaults. *Review of Income and Wealth*, 55(1):166–185, 2009. 36

Róbert Busa-Fekete, Balázs Szörényi, Paul Weng, and Shie Mannor. Multi-objective bandits: Optimizing the generalized gini index. In *International Conference on Machine Learning*, pages 625–634. PMLR, 2017. 16, 28, 38, 60, 70, 163, 168, 250

Iván Cantador, Peter Brusilovsky, and Tsvi Kuflik. 2nd workshop on information heterogeneity and fusion in recommender systems (hetrec 2011). In *Proceedings of the 5th ACM conference on Recommender systems*, RecSys 2011, New York, NY, USA, 2011. ACM. 48, 64, 83, 95, 154, 164, 198

Micah D Carroll, Anca Dragan, Stuart Russell, and Dylan Hadfield-Menell. Estimating and penalizing induced preference shifts in recommender systems. In *International Conference on Machine Learning*, pages 2686–2708. PMLR, 2022. 105

Rémi Castera, Patrick Loiseau, and Bary SR Pradelski. Statistical discrimination in stable matchings. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 373–374, 2022. 104

L Elisa Celis and Nisheeth K Vishnoi. Fair personalization. *arXiv preprint arXiv:1707.02260*, 2017. 21

L Elisa Celis, Lingxiao Huang, and Nisheeth K Vishnoi. Multiwinner voting with fairness constraints. *arXiv preprint arXiv:1710.10057*, 2017a. 103

L Elisa Celis, Damian Straszak, and Nisheeth K Vishnoi. Ranking with fairness constraints. *arXiv preprint arXiv:1704.06840*, 2017b. 21, 50, 88

L. Elisa Celis, Lingxiao Huang, and Nisheeth K. Vishnoi. Multiwinner voting with fairness constraints. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI-18*, pages 144–151. International Joint Conferences on Artificial Intelligence Organization, 7 2018a. doi: 10.24963/ijcai.2018/20. URL https://doi.org/10.24963/ijcai.2018/20. 214

L Elisa Celis, Sayash Kapoor, Farnood Salehi, and Nisheeth K Vishnoi. An algorithmic framework to control bias in bandit-based personalization. *arXiv preprint arXiv:1802.08674*, 2018b. 28, 75, 164

O. Celma. *Music Recommendation and Discovery in the Long Tail*. Springer, 2010. 155

Sarah H Cen, Andrew Ilyas, and Aleksander Mądry. A game-theoretic perspective on trust in recommendation. *ICML 2022 Workshop on Responsible Decision-Making in Dynamic Environments*, 2022. URL https://responsibledecisionmaking.github.io/assets/pdf/papers/36.pdf. 108

Nicolo Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012. 76

Abhijnan Chakraborty, Gourab K Patro, Niloy Ganguly, Krishna P Gummadi, and Patrick Loiseau. Equality of voice: Towards fair representation in crowdsourced top-k recommendations. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 129–138. ACM, 2019. 37, 51, 103

Satya R Chakravarty et al. Inequality, polarization and poverty. *Advances in distributional analysis. New York*, 2009. 33

Praveen Chandar, Brian St. Thomas, Lucas Maystre, Vijay Pappu, Roberto Sanchis-Ojeda, Tiffany Wu, Ben Carterette, Mounia Lalmas, and Tony Jebara. Using survival models to estimate user engagement in online experiments. In *Proceedings of the ACM Web Conference 2022*, pages 3186–3195, 2022. 108

Olivier Chapelle and Mingrui Wu. Gradient descent optimization of smoothed information retrieval metrics. *Information retrieval*, 13(3):216–235, 2010. 27, 70

Olivier Chapelle and Ya Zhang. A dynamic bayesian network click model for web search ranking. In *Proceedings of the 18th international conference on World wide web*, pages 1–10, 2009. 22, 106

Olivier Chapelle, Shihao Ji, Ciya Liao, Emre Velipasaoglu, Larry Lai, and Su-Lin Wu. Intent-based diversification of web search results: metrics and algorithms. *Information Retrieval*, 14(6):572–592, 2011. 104

Jiawei Chen, Hande Dong, Xiang Wang, Fuli Feng, Meng Wang, and Xiangnan He. Bias and debias in recommender system: A survey and future directions. *ACM Transactions on Information Systems*, 41(3):1–39, 2023. 22, 104

Shouyuan Chen, Tian Lin, Irwin King, Michael R Lyu, and Wei Chen. Combinatorial pure exploration of multi-armed bandits. In *Advances in Neural Information Processing Systems*, pages 379–387, 2014. 29

Xi Chen, Qihang Lin, Seyoung Kim, Jaime G Carbonell, and Eric P Xing. Smoothing proximal gradient method for general structured sparse learning. *arXiv preprint arXiv:1202.3708*, 2012. 60

Yifang Chen, Alex Cuellar, Haipeng Luo, Jignesh Modi, Heramb Nemlekar, and Stefanos Nikolaidis. Fair contextual multi-armed bandits: Theory and experiments. In *Conference on Uncertainty in Artificial Intelligence*, pages 181–190. PMLR, 2020. 28, 75, 164

Wang Chi Cheung. Regret minimization for reinforcement learning with vectorial feedback and complex objectives. *Advances in Neural Information Processing Systems*, 32, 2019. 28, 75, 163

Yann Chevaleyre, Ulle Endriss, and Nicolas Maudet. Distributed fair allocation of indivisible goods. *Artificial Intelligence*, 242:1–22, 2017. 198

Alexandra Chouldechova. Fair prediction with disparate impact: A study of bias in recidivism prediction instruments. *arXiv preprint arXiv:1703.00056*, 2017. 19

Alexandra Chouldechova and Aaron Roth. A snapshot of the frontiers of fairness in machine learning. *Communications of the ACM*, 63(5):82–89, 2020. 19

Aleksandr Chuklin, Ilya Markov, and Maarten de Rijke. Click models for web search. *Synthesis lectures on information concepts, retrieval, and services*, 7(3):1–115, 2015. 22

Kenneth L Clarkson. Coresets, sparse greedy approximation, and the frank-wolfe algorithm. *ACM Transactions on Algorithms (TALG)*, 6(4):1–30, 2010. 25, 47, 60, 63, 151, 152, 175

Sam Corbett-Davies and Sharad Goel. The measure and mismeasure of fairness: A critical review of fair machine learning. *arXiv preprint arXiv:1808.00023*, 2018. 5, 19, 88, 105, 243

Sam Corbett-Davies, Emma Pierson, Avi Feller, Sharad Goel, and Aziz Huq. Algorithmic decision making and the cost of fairness. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 797–806. ACM, 2017. 5, 6, 243, 244, 245

David Cossock and Tong Zhang. Statistical analysis of bayes optimal subset ranking. *IEEE Transactions on Information Theory*, 54(11):5140–5154, 2008. 143, 145, 146, 152

Cyrus Cousins. An axiomatic theory of provably-fair welfare-centric machine learning. *Advances in Neural Information Processing Systems*, 34, 2021. 38, 70

Frank Cowell. *Measuring inequality*. Oxford University Press, 2011. 103

Frank A Cowell. Inequality decomposition: three bad measures. *Bulletin of Economic Research*, 40 (4):309–312, 1988. 59

Frank A Cowell. Measurement of inequality. *Handbook of income distribution*, 1:87–166, 2000. 33, 54

Nick Craswell, Onno Zoeter, Michael Taylor, and Bill Ramsey. An experimental comparison of click position-bias models. In *International conference on web search and data mining*, 2008. 8, 22, 104, 106

Mihaela Curmei, Andreas A Haupt, Benjamin Recht, and Dylan Hadfield-Menell. Towards psychologically-grounded dynamic preference models. In *Proceedings of the 16th ACM Conference on Recommender Systems*, pages 35–48, 2022. 105

Marco Cuturi, Olivier Teboul, and Jean-Philippe Vert. Differentiable ranking and sorting using optimal transport. In *Advances in Neural Information Processing Systems*, pages 6858–6868, 2019. 27, 70

Hugh Dalton. The measurement of the inequality of incomes. *The Economic Journal*, 30(119): 348–361, 1920. 59

Van Dang and W Bruce Croft. Diversity by proportionality: an election-based approach to search result diversification. In *Proceedings of the 35th international ACM SIGIR conference on Research and development in information retrieval*, pages 65–74, 2012. 103

Varsha Dani, Thomas P Hayes, and Sham M Kakade. Stochastic linear optimization under bandit feedback. 2008. 78, 175

Amit Datta, Michael Carl Tschantz, and Anupam Datta. Automated experiments on ad privacy settings. *Proceedings on Privacy Enhancing Technologies*, 2015(1):92–112, 2015. 3, 16, 20, 54, 85, 87, 240, 251

Manlio De Domenico, Antonio Lima, Paul Mougel, and Mirco Musolesi. The anatomy of a scientific rumor. *Scientific reports*, 3(1):1–9, 2013. 49, 68

Mario De Vergottini. Sugli indici di concentrazione. *Statistica*, 10(4):445–454, 1950. 36

Sarah Dean and Jamie Morgenstern. Preference dynamics under personalized recommendations. *arXiv preprint arXiv:2205.13026*, 2022. 84, 105

Yashar Deldjoo, Vito Walter Anelli, Hamed Zamani, Alejandro Bellogin, and Tommaso Di Noia. A flexible framework for evaluating user and item fairness in recommender systems. *User Modeling and User-Adapted Interaction*, pages 1–55, 2021. 21, 70

Yashar Deldjoo, Dietmar Jannach, Alejandro Bellogin, Alessandro Difonzo, and Dario Zanzonelli. A survey of research on fair recommender systems. *arXiv preprint arXiv:2205.11127*, 2022. 22

Siddartha Devic, David Kempe, Vatsal Sharan, and Aleksandra Korolova. Fairness in matching under uncertainty. *arXiv preprint arXiv:2302.03810*, 2023. 104

Fernando Diaz, Bhaskar Mitra, Michael D Ekstrand, Asia J Biega, and Ben Carterette. Evaluating stochastic rankings with expected exposure. In *Proceedings of the 29th ACM international conference on information & knowledge management*, pages 275–284, 2020. 10, 21, 28, 54, 69, 105, 108, 164

Virginie Do and Nicolas Usunier. Optimizing generalized gini indices for fairness in rankings. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '22, page 737–747, 2022. v, 28, 53, 164, 166, 167, 168

Virginie Do, Jamal Atif, Jérôme Lang, and Nicolas Usunier. Online selection of diverse committees. In Zhi-Hua Zhou, editor, *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI-21*, pages 154–160. International Joint Conferences on Artificial Intelligence Organization, 8 2021a. doi: 10.24963/ijcai.2021/22. URL https://doi.org/10.24963/ijcai.2021/22. Main Track. v, 103

Virginie Do, Sam Corbett-Davies, Jamal Atif, and Nicolas Usunier. Online certification of preference-based fairness for personalized recommender systems. *arXiv preprint arXiv:2104.14527*, 2021b. 51

Virginie Do, Sam Corbett-Davies, Jamal Atif, and Nicolas Usunier. Two-sided fairness in rankings via lorenz dominance. *Advances in Neural Information Processing Systems*, 34, 2021c. v, 27, 28, 39, 54, 55, 56, 57, 58, 59, 60, 62, 63, 64, 65, 66, 67, 68, 69, 70, 71, 74, 77, 82, 164, 167

Virginie Do, Sam Corbett-Davies, Jamal Atif, and Nicolas Usunier. Online certification of preference-based fairness for personalized recommender systems. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 36, pages 6532–6540, 2022a. v, 69, 85

Virginie Do, Matthieu Hervouin, Jérôme Lang, and Piotr Skowron. Online approval committee elections. In Lud De Raedt, editor, *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22*, pages 251–257. International Joint Conferences on Artificial Intelligence Organization, 7 2022b. doi: 10.24963/ijcai.2022/36. URL https://doi.org/10.24963/ijcai.2022/36. Main Track. v, 103

Virginie Do, Elvis Dohmatob, Matteo Pirotta, Alessandro Lazaric, and Nicolas Usunier. Contextual bandits with concave rewards, and an application to fair ranking. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=UT-_SVOyD1H. v, 73

David Donaldson and John A Weymark. Social choice in economic environments. *Journal of Economic Theory*, 46(2):291–308, 1988. 103

Madalina M Drugan and Ann Nowe. Designing multi-objective multi-armed bandits algorithms: A study. In *The 2013 International Joint Conference on Neural Networks (IJCNN)*, pages 1–8. IEEE, 2013. 29, 163

Dheeru Dua and Casey Graff. UCI machine learning repository, 2017. URL `http://archive.ics.uci.edu/ml`. 234

John C Duchi, Peter L Bartlett, and Martin J Wainwright. Randomized smoothing for stochastic optimization. *SIAM Journal on Optimization*, 22(2):674–701, 2012. 180

Cynthia Dwork, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Richard Zemel. Fairness through awareness. In *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference*, pages 214–226. ACM, 2012. 20, 38, 88

Yonathan Efroni, Shie Mannor, and Matteo Pirotta. Exploration-exploitation in constrained mdps. *arXiv preprint arXiv:2003.02189*, 2020. 220, 231

Michael D Ekstrand, Mucun Tian, Ion Madrazo Azpiazu, Jennifer D Ekstrand, Oghenemaro Anuyah, David McNeill, and Maria Soledad Pera. All the cool kids, how do they fit in?: Popularity and demographic biases in recommender evaluation and effectiveness. In *Conference on Fairness, Accountability and Transparency*, pages 172–186. PMLR, 2018. 5, 20, 54, 69, 88, 242

Michael D Ekstrand, Anubrata Das, Robin Burke, and Fernando Diaz. Fairness in recommender systems. In *Recommender systems handbook*, pages 679–707. Springer, 2022. 22

Vitalii Emelianov, Nicolas Gast, and Patrick Loiseau. Fairness in selection problems with strategic candidates. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 375–403, 2022. 7, 245

Beyza Ermis, Patrick Ernst, Yannik Stein, and Giovanni Zappella. Learning to rank in the position based model with bandit feedback. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, pages 2405–2412, 2020. 83, 165

Eyal Even-Dar, Shie Mannor, and Yishay Mansour. Action elimination and stopping conditions for the multi-armed bandit and reinforcement learning problems. *Journal of machine learning research*, 7(Jun):1079–1105, 2006. 29

Facebook. Community standards enforcement report. `https://transparency.facebook.com/community-standards-enforcement`, 2020. 109

Piotr Faliszewski, Piotr Skowron, Arkadii Slinko, and Nimrod Talmon. Multiwinner voting: A new challenge for social choice theory. *Trends in computational social choice*, 74:27–47, 2017. 214

Zhichong Fang, Aman Agarwal, and Thorsten Joachims. Intervention harvesting for context-dependent examination-bias estimation. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 825–834, 2019. 81

Michael Feldman, Sorelle A Friedler, John Moeller, Carlos Scheidegger, and Suresh Venkatasubramanian. Certifying and removing disparate impact. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 259–268. ACM, 2015. 6, 20, 245

Jessie Finocchiaro, Roland Maio, Faidra Monachou, Gourab K Patro, Manish Raghavan, Ana-Andreea Stoica, and Stratis Tsirtsis. Bridging machine learning and mechanism design towards algorithmic fairness. *arXiv preprint arXiv:2010.05434*, 2020. 70

Jessie Finocchiaro, Roland Maio, Faidra Monachou, Gourab K Patro, Manish Raghavan, Ana-Andreea Stoica, and Stratis Tsirtsis. Bridging machine learning and mechanism design towards algorithmic fairness. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 489–503, 2021. 37, 51

Bailey Flanigan, Paul Gölz, Anupam Gupta, and Ariel D. Procaccia. Neutralizing self-selection bias in sampling for sortition. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. 215

Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. *arXiv preprint cs/0408007*, 2004. 29, 163

Marc Fleurbaey. Equality versus priority: how relevant is the distinction? *Economics & Philosophy*, 31(2):203–217, 2015. 23, 24

Duncan K Foley. Resource allocation and the public sector. 1967. 17, 37, 89, 90, 251

Dylan Foster and Alexander Rakhlin. Beyond ucb: Optimal and efficient contextual bandits with regression oracles. In *International Conference on Machine Learning*, pages 3199–3210. PMLR, 2020. 75, 79, 183, 184, 185, 186

Marguerite Frank and Philip Wolfe. An algorithm for quadratic programming. *Naval research logistics quarterly*, 3(1-2):95–110, 1956. 13, 15, 24, 26, 47, 55, 60, 70, 199, 248

Nancy Fraser and Axel Honneth. *Redistribution or recognition?: a political-philosophical exchange*. Verso, 2003. 107

David A Freedman. On tail probabilities for martingales. *the Annals of Probability*, pages 100–118, 1975. 202

Sorelle A Friedler, Carlos Scheidegger, and Suresh Venkatasubramanian. On the (im) possibility of fairness. *arXiv preprint arXiv:1609.07236*, 2016. 23

Victor Gabillon, Mohammad Ghavamzadeh, and Alessandro Lazaric. Best arm identification: A unified approach to fixed budget and fixed confidence. In *Advances in Neural Information Processing Systems*, pages 3212–3220, 2012. 29

David Gale and Lloyd S Shapley. College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15, 1962. 104

Evrard Garcelon, Mohammad Ghavamzadeh, Alessandro Lazaric, and Matteo Pirotta. Improved algorithms for conservative exploration in bandits. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 3962–3969, 2020a. 29, 88, 94, 202

Evrard Garcelon, Baptiste Roziere, Laurent Meunier, Jean Tarbouriech, Olivier Teytaud, Alessandro Lazaric, and Matteo Pirotta. Adversarial attacks on linear contextual bandits. *Advances in Neural Information Processing Systems*, 33:14362–14373, 2020b. 164

David García-Soriano and Francesco Bonchi. Maxmin-fair ranking: individual fairness under group-fairness constraints. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 436–446, 2021. 21

Aurélien Garivier and Emilie Kaufmann. Optimal best arm identification with fixed confidence. In *Conference on Learning Theory*, pages 998–1027, 2016. 29

Yingqiang Ge, Shuchang Liu, Ruoyuan Gao, Yikun Xian, Yunqi Li, Xiangyu Zhao, Changhua Pei, Fei Sun, Junfeng Ge, Wenwu Ou, et al. Towards long-term fairness in recommendation. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 445–453, 2021. 21, 108

Matthieu Geist, Julien Pérolat, Mathieu Laurière, Romuald Elie, Sarah Perrin, Olivier Bachem, Rémi Munos, and Olivier Pietquin. Concave utility reinforcement learning: the mean-field game viewpoint. *arXiv preprint arXiv:2106.03787*, 2021. 28, 163

Arthur M Geoffrion. Proper efficiency and the theory of vector maximization. *Journal of mathematical analysis and applications*, 22(3):618–630, 1968. 57, 58, 59

Sahin Cem Geyik, Stuart Ambler, and Krishnaram Kenthapadi. Fairness-aware ranking in search & recommendation systems with application to linkedin talent search. In *Proceedings of the 25th acm sigkdd international conference on knowledge discovery & data mining*, pages 2221–2231, 2019. 2, 6, 11, 21, 27, 74, 88, 104, 107, 163, 239, 245

Alireza Gharahighehi, Celine Vens, and Konstantinos Pliakos. Fair multi-stakeholder news recommender system with hypergraph ranking. *Information Processing & Management*, 58(5):102663, 2021. 21

Corrado Gini. Measurement of inequality of incomes. *The economic journal*, 31(121):124–126, 1921. 34, 54, 160

Paul Gölz, Anson Kahng, and Ariel D Procaccia. Paradoxes in fair machine learning. *Advances in Neural Information Processing Systems (NeurIPS).*, 2019. 37, 38, 51, 70

Aditya Grover, Eric Wang, Aaron Zweig, and Stefano Ermon. Stochastic optimization of sorting networks via continuous relaxations. *arXiv preprint arXiv:1903.08850*, 2019. 27, 70

Asela Gunawardana, Guy Shani, and Sivan Yogev. Evaluating recommender systems. In *Recommender systems handbook*, pages 547–601. Springer, 2012. 109

Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. Measuring price discrimination and steering on e-commerce web sites. In *Proceedings of the 2014 conference on internet measurement conference*, pages 305–318, 2014. 20, 50, 88

Moritz Hardt, Eric Price, Eric Price, and Nati Srebro. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems*, volume 29. Curran Associates, Inc., 2016a. URL https://proceedings.neurips.cc/paper/2016/file/9d2682367c3935defcb1f9e247a97c0d-Paper.pdf. 5, 20

Moritz Hardt, Eric Price, Nati Srebro, et al. Equality of opportunity in supervised learning. In *Advances in Neural Information Processing Systems*, pages 3315–3323, 2016b. 6, 7, 102, 142, 243, 245

G. H. Hardy, J. E. Littlewood, and George Pólya. Inequalities. 2nd ed. Cambridge, Engl.: At the University Press. XII, 324 p. (1952)., 1952. 10, 31, 32, 61, 63, 151

F Maxwell Harper and Joseph A Konstan. The movielens datasets: History and context. *Acm transactions on interactive intelligent systems (tiis)*, 5(4):1–19, 2015. 4, 48, 64, 96, 157, 198

Thomas P Hayes. A large-deviation inequality for vector-valued martingales. *Combinatorics, Probability and Computing*, 2005. 172, 173

Elad Hazan et al. Introduction to online convex optimization. *Foundations and Trends® in Optimization*, 2(3-4):157–325, 2016. 29, 163

Hoda Heidari, Claudio Ferrari, Krishna Gummadi, and Andreas Krause. Fairness behind a veil of ignorance: A welfare analysis for automated decision making. In *Advances in Neural Information Processing Systems*, pages 1265–1276, 2018. 37, 38, 51, 70

Hoda Heidari, Michele Loi, Krishna P Gummadi, and Andreas Krause. A moral framework for understanding fair ml through economic models of equality of opportunity. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, pages 181–190, 2019. 102, 107, 142

Maria Heuss, Fatemeh Sarvi, and Maarten de Rijke. Fairness of exposure in light of incomplete exposure estimation. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 759–769, 2022. 21

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. 207

Wassily Hoeffding. Probability inequalities for sums of bounded random variables. In *The Collected Works of Wassily Hoeffding*, pages 409–426. Springer, 1994. 227

Kenneth Holstein, Jennifer Wortman Vaughan, Hal Daumé III, Miro Dudik, and Hanna Wallach. Improving fairness in machine learning systems: What do industry practitioners need? In *Proceedings of the 2019 CHI conference on human factors in computing systems*, pages 1–16, 2019. 103, 108

Safwan Hossain, Andjela Mladenovic, and Nisarg Shah. Designing fairly fair classifiers via economic fairness notions. In *Proceedings of The Web Conference 2020*, pages 1559–1569, 2020. 37, 38, 51

Safwan Hossain, Evi Micha, and Nisarg Shah. Fair algorithms for multi-agent multi-armed bandits. *Advances in Neural Information Processing Systems*, 34:24005–24017, 2021. 37

Jiri Hron, Karl Krauth, Michael Jordan, and Niki Kilbertus. On component interactions in two-stage recommender systems. *Advances in neural information processing systems*, 34:2744–2757, 2021. 109

Jiri Hron, Karl Krauth, Michael Jordan, Niki Kilbertus, and Sarah Dean. Modeling content creator incentives on algorithm-curated platforms. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=l6CpxixmUg. 108

Lily Hu and Yiling Chen. Fair classification and social welfare. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 535–545, 2020. 6, 37, 38, 51, 70, 245

Yifan Hu, Yehuda Koren, and Chris Volinsky. Collaborative filtering for implicit feedback datasets. In *2008 Eighth IEEE International Conference on Data Mining*, pages 263–272. Ieee, 2008. 154, 198

Jin Huang, Harrie Oosterhuis, Maarten De Rijke, and Herke Van Hoof. Keeping dataset biases out of the simulation: A debiased simulator for reinforcement learning based recommender systems. In *Fourteenth ACM conference on recommender systems*, pages 190–199, 2020. 108

Ferenc Huszár, Sofia Ira Ktena, Conor O'Brien, Luca Belli, Andrew Schlaikjer, and Moritz Hardt. Algorithmic amplification of politics on twitter. *Proceedings of the National Academy of Sciences*, 119(1):e2025334119, 2022. 2, 239

Jevan A Hutson, Jessie G Taft, Solon Barocas, and Karen Levy. Debiasing desire: Addressing bias & discrimination on intimate platforms. *Proceedings of the ACM on Human-Computer Interaction*, 2(CSCW):1–18, 2018. 2, 11, 239

Eugene Ie, Chih-wei Hsu, Martin Mladenov, Vihan Jain, Sanmit Narvekar, Jing Wang, Rui Wu, and Craig Boutilier. Recsim: A configurable simulation platform for recommender systems. *arXiv preprint arXiv:1909.04847*, 2019. 108

Christina Ilvento, Meena Jagadeesan, and Shuchi Chawla. Multi-category fairness in sponsored search auctions. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 348–358, 2020. 88

Basileal Imana, Aleksandra Korolova, John Heidemann, and . Auditing for discrimination in algorithms delivering job ads. In *Proceedings of the Web Conference 2021*, pages 3767–3778, 2021. 5, 16, 85, 87, 242, 251

Instagram. What is the instagram feed? `https://ai.facebook.com/tools/system-cards/instagram-feed-ranking/`, 2022. 108

Rashidul Islam, Kamrun Naher Keya, Ziqian Zeng, Shimei Pan, and James Foulds. Debiasing career recommendations with neural fair collaborative filtering. In *Proceedings of the Web Conference 2021*, pages 3779–3790, 2021. 6, 245

Shahin Jabbari, Matthew Joseph, Michael Kearns, Jamie Morgenstern, and Aaron Roth. Fairness in reinforcement learning. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*, pages 1617–1626. JMLR. org, 2017. 89

Abigail Z Jacobs and Hanna Wallach. Measurement and fairness. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, pages 375–385, 2021. 105

Martin Jaggi. Revisiting frank-wolfe: Projection-free sparse convex optimization. In *International Conference on Machine Learning*, pages 427–435. PMLR, 2013. 24, 25, 26, 47, 60, 70, 175

Thomas Jaksch, Ronald Ortner, and Peter Auer. Near-optimal regret bounds for reinforcement learning. *Journal of Machine Learning Research*, 11(4), 2010. 220, 227

Kevin Jamieson and Robert Nowak. Best-arm identification algorithms for multi-armed bandits in the fixed confidence setting. In *2014 48th Annual Conference on Information Sciences and Systems (CISS)*, pages 1–6. IEEE, 2014. 199

Kevin Jamieson, Matthew Malloy, Robert Nowak, and Sébastien Bubeck. lil'ucb: An optimal exploration algorithm for multi-armed bandits. In *Conference on Learning Theory*, pages 423–439, 2014. 94, 200, 201, 205

Kevin G Jamieson and Lalit Jain. A bandit approach to sequential experimental design with false discovery control. In *Advances in Neural Information Processing Systems*, pages 3660–3670, 2018. 29

Kalervo Järvelin and Jaana Kekäläinen. Cumulated gain-based evaluation of ir techniques. *ACM Transactions on Information Systems (TOIS)*, 20(4):422–446, 2002. 4, 8, 106, 241

Olivier Jeunen and Bart Goethals. Top-k contextual bandits with equity of exposure. In *Fifteenth ACM Conference on Recommender Systems*, pages 310–320, 2021. 22, 28, 75, 107, 164

Yongzheng Jia, Xue Liu, and Wei Xu. When online dating meets nash social welfare: Achieving efficiency and fairness. In *Proceedings of the 2018 World Wide Web Conference*, pages 429–438, 2018. 22, 51, 68, 70

Ray Jiang, Silvia Chiappa, Tor Lattimore, András György, and Pushmeet Kohli. Degenerate feedback loops in recommender systems. In *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, pages 383–390, 2019. 84, 105

How Jing and Alexander J Smola. Neural survival recommender. In *Proceedings of the Tenth ACM International Conference on Web Search and Data Mining*, pages 515–524, 2017. 108

Christopher C Johnson. Logistic matrix factorization for implicit feedback data. *Advances in Neural Information Processing Systems*, 27(78):1–9, 2014. 154, 158, 198

Matthew Joseph, Michael Kearns, Jamie H Morgenstern, and Aaron Roth. Fairness in learning: Classic and contextual bandits. In *Advances in Neural Information Processing Systems*, pages 325–333, 2016. 28, 89

Dimitris Kalimeris, Smriti Bhagat, Shankar Kalyanaraman, and Udi Weinsberg. Preference amplification in recommender systems. In *Proceedings of the 27th ACM SIGKDD Conference on Knowledge Discovery & Data Mining*, pages 805–815, 2021. 71, 105, 109

Nathan Kallus, Xiaojie Mao, and Angela Zhou. Assessing algorithmic fairness with unobserved protected class using data combination. *Management Science*, 68(3):1959–1981, 2022. 103

Faisal Kamiran and Toon Calders. Classifying without discriminating. In *Computer, Control and Communication, 2009. IC4 2009. 2nd International Conference on*, pages 1–6. IEEE, 2009. 19

Matthew Kay, Cynthia Matuszek, and Sean A Munson. Unequal representation and gender stereotypes in image search results for occupations. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 3819–3828. ACM, 2015. 2, 20, 50, 142, 239

Thomas Kerdreux, Fabian Pedregosa, and Alexandre d'Aspremont. Frank-wolfe with subsampling oracle. In *International Conference on Machine Learning*, pages 2591–2600. PMLR, 2018. 78

Niki Kilbertus, Manuel Gomez Rodriguez, Bernhard Schölkopf, Krikamol Muandet, and Isabel Valera. Fair decisions despite imperfect predictions. In *International Conference on Artificial Intelligence and Statistics*, pages 277–287. PMLR, 2020. 5, 105, 243

Michael Kim, Omer Reingold, and Guy Rothblum. Fairness through computationally-bounded awareness. In *Advances in Neural Information Processing Systems*, pages 4842–4852, 2018. 20

Michael P Kim, Aleksandra Korolova, Guy N Rothblum, and Gal Yona. Preference-informed fairness. *arXiv preprint arXiv:1904.01793*, 2019. 38, 87, 88, 90

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. Inherent trade-offs in the fair determination of risk scores. *arXiv preprint arXiv:1609.05807*, 2016. 6, 19, 244

Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan. Human decisions and machine predictions. *The quarterly journal of economics*, 133(1):237–293, 2018a. 105

Jon Kleinberg, Jens Ludwig, Sendhil Mullainathan, and Ashesh Rambachan. Algorithmic fairness. In *Aea papers and proceedings*, volume 108, pages 22–27, 2018b. 5, 6, 7, 37, 51, 243, 244, 245

Jon Kleinberg, Sendhil Mullainathan, and Manish Raghavan. The challenge of understanding what users want: Inconsistent preferences and engagement optimization. *arXiv preprint arXiv:2202.11776*, 2022. 105

Till Kletti, Jean-Michel Renders, and Patrick Loiseau. Introducing the expohedron for efficient pareto-optimal fairness-utility amortizations in repeated rankings. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 498–507, 2022a. 10, 21, 25, 26, 28, 69, 108, 164

Till Kletti, Jean-Michel Renders, and Patrick Loiseau. Pareto-optimal fairness-utility amortizations in rankings with a dbn exposure model. *arXiv preprint arXiv:2205.07647*, 2022b. 21, 22, 105, 106

Ron Kohavi, Roger Longbotham, Dan Sommerfield, and Randal M Henne. Controlled experiments on the web: survey and practical guide. *Data mining and knowledge discovery*, 18:140–181, 2009. 109

Ron Kohavi, Alex Deng, Brian Frasca, Roger Longbotham, Toby Walker, and Ya Xu. Trustworthy online controlled experiments: Five puzzling outcomes explained. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 786–794, 2012. 109

Serge-Christophe Kolm. Unequal inequalities. ii. *Journal of economic theory*, 13(1):82–111, 1976. 10

Yehuda Koren, Robert Bell, and Chris Volinsky. Matrix factorization techniques for recommender systems. *Computer*, 42(8):30–37, 2009. 4, 197, 242

Karl Krauth, Sarah Dean, Alex Zhao, Wenshuo Guo, Mihaela Curmei, Benjamin Recht, and Michael I Jordan. Do offline metrics predict online performance in recommender systems? *arXiv preprint arXiv:2011.07931*, 2020. 108

Karl Krauth, Yixin Wang, and Michael I Jordan. Breaking feedback loops in recommender systems with causal inference. *arXiv preprint arXiv:2207.01616*, 2022. 108

Matt J Kusner and Joshua R Loftus. The long road to fairer algorithms. *Nature*, 578(7793):34–36, 2020. 19

Matt J Kusner, Joshua Loftus, Chris Russell, and Ricardo Silva. Counterfactual fairness. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems 30*, pages 4069–4079. Curran Associates, Inc., 2017. URL `http://papers.nips.cc/paper/6995-counterfactual-fairness.pdf`. 88

Martin Lackner and Piotr Skowron. Multi-winner voting with approval preferences. Technical Report arXiv:2007.01795 [cs.GT], arXiv.org, 2020. 103

Simon Lacoste-Julien, Martin Jaggi, Mark Schmidt, and Patrick Pletscher. Block-coordinate frank-wolfe optimization for structural svms. In *International Conference on Machine Learning*, 2013. 78

Paul Lagrée, Claire Vernade, and Olivier Cappe. Multiple-play bandits in the position-based model. *Advances in Neural Information Processing Systems*, 29, 2016. 28, 81, 164, 188

Anja Lambrecht and Catherine Tucker. Algorithmic bias? an empirical study of apparent gender-based discrimination in the display of stem career ads. *Management Science*, 65(7):2966–2981, 2019. 20, 50, 85, 88

Guanghui Lan. The complexity of large-scale convex programming under a linear optimization oracle. *arXiv preprint arXiv:1309.5550*, 2013. 15, 24, 26, 27, 55, 60, 62, 70, 80, 175, 179, 180, 249

Jérôme Lang and Piotr Skowron. Multi-attribute proportional representation. *Artificial Intelligence*, 263:74–106, 2018. 103, 104, 214

John Langford. Tutorial on practical prediction theory for classification. *Journal of machine learning research*, 6(3), 2005. 211

John Langford and Tong Zhang. The epoch-greedy algorithm for contextual multi-armed bandits. *Advances in neural information processing systems*, 20(1):96–1, 2007. 75

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020. 27, 175

Tomo Lazovich, Luca Belli, Aaron Gonzales, Amanda Bower, Uthaipon Tantipongpipat, Kristian Lum, Ferenc Huszar, and Rumman Chowdhury. Measuring disparate outcomes of content recommendation algorithms with distributional inequality metrics. *arXiv preprint arXiv:2202.01615*, 2022. 38, 70

Michel Le Breton and John A Weymark. Arrovian social choice theory on economic domains. *Handbook of social choice and welfare*, 2:191–299, 2011. 37

Nixie S Lesmana, Xuan Zhang, and Xiaohui Bei. Balancing efficiency and fairness in on-demand ridesourcing. In H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 32. Curran Associates, Inc., 2019. 22, 51

Fengjiao Li, Jia Liu, and Bo Ji. Combinatorial sleeping bandits with fairness constraints. *IEEE Transactions on Network Science and Engineering*, 7(3):1799–1813, 2019. 28, 164

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010. 16, 27, 82, 250

Shuai Li, Baoxiang Wang, Shengyu Zhang, and Wei Chen. Contextual combinatorial cascading bandits. In *International conference on machine learning*, pages 1245–1253. PMLR, 2016. 83, 164, 188, 189

Yunqi Li, Hanxiong Chen, Shuyuan Xu, Yingqiang Ge, Juntao Tan, Shuchang Liu, and Yongfeng Zhang. Fairness in recommendation: A survey. *arXiv preprint arXiv:2205.13619*, 2022. 22

Cong Han Lim and Stephen J Wright. Efficient bregman projections onto the permutahedron and related polytopes. In *Artificial Intelligence and Statistics*, pages 1205–1213. PMLR, 2016. 27, 61

Daryl Lim, Julian McAuley, and Gert Lanckriet. Top-n recommendation with missing implicit feedback. In *Proceedings of the 9th ACM Conference on Recommender Systems*, pages 309–312, 2015. 157

Zhiyuan Jerry Lin, Raul Astudillo, Peter Frazier, and Eytan Bakshy. Preference exploration for efficient bayesian optimization with multiple outcomes. In *International Conference on Artificial Intelligence and Statistics*, pages 4235–4258. PMLR, 2022. 84

David Liu, Virginie Do, Nicolas Usunier, and Maximilian Nickel. Group fairness without demographics using social networks. 2023. 2023 ACM Conference on Fairness, Accountability, and Transparency. v, 103

Lydia T Liu, Sarah Dean, Esther Rolf, Max Simchowitz, and Moritz Hardt. Delayed impact of fair machine learning. *arXiv preprint arXiv:1803.04383*, 2018. 5, 243

Weiwen Liu, Jun Guo, Nasim Sonboli, Robin Burke, and Shengyu Zhang. Personalized fairness-aware re-ranking for microlending. In *Proceedings of the 13th ACM Conference on Recommender Systems*, pages 467–471, 2019. 21, 50

Yang Liu, Goran Radanovic, Christos Dimitrakakis, Debmalya Mandal, and David C Parkes. Calibrated fairness in bandits. *arXiv preprint arXiv:1707.01875*, 2017. 28, 89

Andrea Locatelli, Maurilio Gutzeit, and Alexandra Carpentier. An optimal algorithm for the thresholding bandit problem. *arXiv preprint arXiv:1605.08671*, 2016. 29, 88

Pierre l'Ecuyer, Patrick Maillé, Nicolás E Stier-Moses, and Bruno Tuffin. Revenue-maximizing rankings for online platforms with quality-sensitive consumers. *Operations Research*, 65(2): 408–423, 2017. 109

Debmalya Mandal and Jiarui Gan. Socially fair reinforcement learning. *arXiv preprint arXiv:2208.12584*, 2022. 28, 163

Masoud Mansoury, Himan Abdollahpouri, Bamshad Mobasher, Mykola Pechenizkiy, Robin Burke, and Milad Sabouri. Unbiased cascade bandits: Mitigating exposure bias in online learning to rank recommendation. *arXiv preprint arXiv:2108.03440*, 2021a. 28, 75, 83, 164, 165, 166

Masoud Mansoury, Himan Abdollahpouri, Mykola Pechenizkiy, Bamshad Mobasher, and Robin Burke. A graph-based approach for mitigating multi-sided exposure bias in recommender systems. *ACM Transactions on Information Systems (TOIS)*, 40(2):1–31, 2021b. 21

Masoud Mansoury, Bamshad Mobasher, and Herke van Hoof. Exposure-aware recommendation using contextual bandits. *arXiv preprint arXiv:2209.01665*, 2022. 22

Benjamin M Marlin and Richard S Zemel. Collaborative prediction and ranking with non-random missing data. In *Proceedings of the third ACM conference on Recommender systems*, pages 5–12, 2009. 104

Albert W Marshall, Ingram Olkin, and Barry C Arnold. Inequalities: theory of majorization and its applications. 1979. 10, 32

Dana Mattioli. On orbitz, mac users steered to pricier hotels, 2012. 20

Andreas Maurer and Massimiliano Pontil. Empirical bernstein bounds and sample variance penalization. *arXiv preprint arXiv:0907.3740*, 2009. 221, 231

Frank McSherry and Ilya Mironov. Differentially private recommender systems: Building privacy into the netflix prize contenders. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 627–636, 2009. 109

Rishabh Mehrotra, Ashton Anderson, Fernando Diaz, Amit Sharma, Hanna Wallach, and Emine Yilmaz. Auditing search engines for differential satisfaction across demographics. In *Proceedings of the 26th international conference on World Wide Web companion*, pages 626–633, 2017. 20, 50, 54, 69, 87, 88

Rishabh Mehrotra, James McInerney, Hugues Bouchard, Mounia Lalmas, and Fernando Diaz. Towards a fair marketplace: Counterfactual evaluation of the trade-off between relevance, fairness & satisfaction in recommendation systems. In *Proceedings of the 27th acm international conference on information and knowledge management*, pages 2243–2251, 2018. 2, 4, 21, 50, 142, 169, 239, 242

Rishabh Mehrotra, Niannan Xue, and Mounia Lalmas. Bandit based optimization of multiple objectives on a music streaming platform. In *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 3224–3233, 2020. 28, 60, 70, 74, 76, 83, 168, 169

Kaisa Miettinen. *Nonlinear multiobjective optimization*, volume 12. Springer Science & Business Media, 2012. 29, 57, 58, 59, 163

Smitha Milli, Luca Belli, and Moritz Hardt. From optimizing engagement to measuring value. In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pages 714–722, 2021. 105

Shira Mitchell, Eric Potash, Solon Barocas, Alexander D'Amour, and Kristian Lum. Algorithmic fairness: Choices, assumptions, and definitions. *Annual Review of Statistics and Its Application*, 8:141–163, 2021. 19

Martin Mladenov, Elliot Creager, Omer Ben-Porat, Kevin Swersky, Richard Zemel, and Craig Boutilier. Optimizing long-term social welfare in recommender systems: A constrained matching approach. In *International Conference on Machine Learning*, pages 6987–6998. PMLR, 2020. 105, 108

Jean Jacques Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 255:2897–2899, 1962. 15, 27, 55, 249

Marco Morik, Ashudeep Singh, Jessica Hong, and Thorsten Joachims. Controlling fairness and bias in dynamic learning-to-rank. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 429–438, 2020. 10, 21, 25, 26, 28, 34, 36, 46, 47, 50, 54, 56, 57, 58, 64, 69, 71, 77, 102, 108, 141, 154, 160, 164, 166

Hervé Moulin. *Fair division and collective welfare*. MIT press, 2003. 7, 9, 27, 30, 31, 32, 44, 56, 58, 59, 107, 167, 246

Robert Franklin Muirhead. Some methods applicable to identities and inequalities of symmetric algebraic functions of n letters. *Proceedings of the Edinburgh Mathematical Society*, 21:144–162, 1902. 32

Mohammadmehdi Naghiaei, Hossein A Rahmani, and Yashar Deldjoo. Cpfair: Personalized consumer and producer fairness re-ranking for recommender systems. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 770–779, 2022. 21

Vedant Nanda, Pan Xu, Karthik Abhinav Sankararaman, John Dickerson, and Aravind Srinivasan. Balancing the tradeoff between profit and fairness in rideshare platforms during high-demand hours. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 2210–2217, 2020. 22, 51

Harikrishna Narasimhan, Andrew Cotter, Maya Gupta, and Serena Wang. Pairwise fairness for ranking and regression. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 5248–5255, 2020. 21

Renato Negrinho and Andre Martins. Orbit regularization. *Advances in neural information processing systems*, 27:3221–3229, 2014. 27, 61

Yu Nesterov. Smooth minimization of non-smooth functions. *Mathematical programming*, 103(1): 127–152, 2005. 26, 62, 70

Yu Nesterov. Complexity bounds for primal-dual methods minimizing the model of objective function. *Mathematical Programming*, 171(1):311–330, 2018. 60, 67

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017. 180

Rita Neves Costa, Sebastien Perez-Duarte, et al. Not all inequality measures were created equal-the measurement of wealth inequality, its decompositions, and an application to european household wealth. Technical report, European Central Bank, 2019. 36

Włodzimierz Ogryczak and Tomasz Śliwiński. On solving linear programs with the ordered weighted averaging objective. *European Journal of Operational Research*, 148(1):80–91, 2003. 60

Luca Oneto and Silvia Chiappa. Fairness in machine learning. In *Recent trends in learning from data: Tutorials from the inns big data and deep learning conference (innsbddl2019)*, pages 155–196. Springer, 2020. 19

Harrie Oosterhuis. Computationally efficient optimization of plackett-luce ranking models for relevance and fairness. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 1023–1032, 2021. 21

Iván Palomares, Carlos Porcel, Luiz Pizzato, Ido Guy, and Enrique Herrera-Viedma. Reciprocal recommender systems: Analysis of state-of-art literature, challenges and opportunities towards social recommendation. *Information Fusion*, 69:103–127, 2021. 11, 45, 67, 144, 158

Debmalya Panigrahi, Atish Das Sarma, Gagan Aggarwal, and Andrew Tomkins. Online selection of diverse results. In *Proceedings of the fifth ACM international conference on Web search and data mining*, pages 263–272, 2012. 215

Dimitris Paraschakis and Bengt J Nilsson. Matchmaking under fairness constraints: a speed dating case study. In *International Workshop on Algorithmic Bias in Search and Recommendation*, pages 43–57. Springer, 2020. 22, 51, 70

Derek Parfit. Equality and priority 1. In *The Notion of Equality*, pages 427–446. Routledge, 2018. 23

Neal Parikh and Stephen Boyd. Proximal algorithms. *Foundations and Trends in optimization*, 1 (3):127–239, 2014. 60, 61

Vishakha Patil, Ganesh Ghalme, Vineet Nair, and Yadati Narahari. Achieving fairness in the stochastic multi-armed bandit problem. In *AAAI*, pages 5379–5386, 2020. 28, 75, 84, 164, 166

Gourab K Patro, Arpita Biswas, Niloy Ganguly, Krishna P Gummadi, and Abhijnan Chakraborty. Fairrec: Two-sided fairness for personalized recommendations in two-sided platforms. In *Proceedings of The Web Conference 2020*, pages 1194–1204, 2020. 3, 21, 26, 38, 48, 51, 56, 64, 65, 69, 71, 83, 88, 91, 154, 157, 164, 198, 240

Gourab K Patro, Lorenzo Porcaro, Laura Mitchell, Qiuyue Zhang, Meike Zehlike, and Nikhil Garg. Fair ranking: a critical review, challenges, and future directions. *arXiv preprint arXiv:2201.12662*, 2022. 22, 27, 164

Alexander Peysakhovich, Christian Kroer, and Nicolas Usunier. Implementing fairness constraints in markets using taxes and subsidies. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pages 916–930, 2023. 108

Thomas Piketty and Emmanuel Saez. Income inequality in the united states, 1913–1998. *The Quarterly journal of economics*, 118(1):1–41, 2003. 36

Luiz Pizzato, Tomasz Rej, Joshua Akehurst, Irena Koprinska, Kalina Yacef, and Judy Kay. Recommending people to people: the nature of reciprocal recommenders with a case study in online dating. *User Modeling and User-Adapted Interaction*, 23(5):447–488, 2013. 22

Geoff Pleiss, Manish Raghavan, Felix Wu, Jon Kleinberg, and Kilian Q Weinberger. On fairness and calibration. In *Advances in Neural Information Processing Systems*, pages 5684–5693, 2017. 6, 244

Flavien Prost, Ben Packer, Jilin Chen, Li Wei, Pierre Kremp, Nicholas Blumm, Susan Wang, Tulsee Doshi, Tonia Osadebe, Lukasz Heldt, et al. Simpson's paradox in recommender fairness: Reconciling differences between per-user and aggregated evaluations. *arXiv preprint arXiv:2210.07755*, 2022. 21

Lijing Qin, Shouyuan Chen, and Xiaoyan Zhu. Contextual combinatorial bandit and its application on diversified online recommendation. In *Proceedings of the 2014 SIAM International Conference on Data Mining*, pages 461–469. SIAM, 2014. 164

Amifa Raj and Michael D Ekstrand. Measuring fairness in ranked results: An analytical and empirical comparison. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval*, pages 726–736, 2022. 10

Inioluwa Deborah Raji, Andrew Smart, Rebecca N White, Margaret Mitchell, Timnit Gebru, Ben Hutchinson, Jamila Smith-Loud, Daniel Theron, and Parker Barnes. Closing the ai accountability gap: defining an end-to-end framework for internal algorithmic auditing. In *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, pages 33–44, 2020. 92

Ashesh Rambachan, Jon Kleinberg, Sendhil Mullainathan, and Jens Ludwig. An economic approach to regulating algorithms. Technical report, National Bureau of Economic Research, 2020. 38, 70

Sathya N Ravi, Maxwell D Collins, and Vikas Singh. A deterministic nonsmooth frank wolfe algorithm with coreset guarantees. *Informs Journal on Optimization*, 1(2):120–142, 2019. 26, 70

Pradeep Ravikumar, Ambuj Tewari, and Eunho Yang. On ndcg consistency of listwise ranking methods. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 618–626. JMLR Workshop and Conference Proceedings, 2011. 145

A Renwick, S Allan, W Jennings, R McKee, M Russell, and G Smith. A considered public voice on brexit: The report of the citizens' assembly on brexit. 2017. 222, 233

Matthew Richardson, Rakesh Agrawal, and Pedro Domingos. Trust management for the semantic web. In *International semantic Web conference*, pages 351–368. Springer, 2003. 50, 159

Herbert Robbins. Some aspects of the sequential design of experiments. *Bulletin of the American Mathematical Society*, 58(5):527–535, 1952. 29

R Tyrrell Rockafellar and Roger J-B Wets. *Variational analysis*, volume 317. Springer Science & Business Media, 2009. 77, 80, 170, 179

John E Roemer. *Theories of distributive justice*. Harvard University Press, 1996. 30, 102, 107

John E Roemer and Alain Trannoy. Equality of opportunity: Theory and measurement. *Journal of Economic Literature*, 54(4):1288–1332, 2016. 102, 142

David Rohde, Stephen Bonner, Travis Dunlop, Flavian Vasile, and Alexandros Karatzoglou. Recogym: A reinforcement learning environment for the problem of product recommendation in online advertising. *arXiv preprint arXiv:1808.00720*, 2018. 108

Aviv Rosenberg and Yishay Mansour. Online convex optimization in adversarial markov decision processes. *arXiv preprint arXiv:1905.07773*, 2019. 220

Yuta Saito and Thorsten Joachims. Fair ranking as fair division: Impact-based individual fairness in ranking. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 1514–1524, 2022. 37, 38, 105, 106

Piotr Sapiezynski, Wesley Zeng, Ronald E Robertson, Alan Mislove, and Christo Wilson. Quantifying the impact of user attentionon fair group representation in ranked lists. In *Companion proceedings of the 2019 world wide web conference*, pages 553–562, 2019. 106

Fatemeh Sarvi, Maria Heuss, Mohammad Aliannejadi, Sebastian Schelter, and Maarten de Rijke. Understanding and mitigating the effect of outliers in fair ranking. In *Proceedings of the Fifteenth ACM International Conference on Web Search and Data Mining*, pages 861–869, 2022. 21

Tobias Schnabel, Adith Swaminathan, Ashudeep Singh, Navin Chandak, and Thorsten Joachims. Recommendations as treatments: Debiasing learning and evaluation. In *international conference on machine learning*, pages 1670–1679. PMLR, 2016. 108

Bernhard Schölkopf and Alexander J Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002. 210

Candice Schumann, Samsara N Counts, Jeffrey S Foster, and John P Dickerson. The diverse cohort selection problem. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, pages 601–609. International Foundation for Autonomous Agents and Multiagent Systems, 2019a. 215

Candice Schumann, Zhi Lang, Nicholas Mattei, and John P Dickerson. Group fairness in bandit arm selection. *arXiv preprint arXiv:1912.03802*, 2019b. 28

Amartya Sen. *Collective Choice and Social Welfare*. Holden Day, San Francisco, 1970. URL http://www.amazon.com/Collective-Choice-Social-Welfare-K/dp/0444851275. Edinburgh: Oliver and Boyd, 197l; Amsterdam: North-Holland, 1979. Swedish translation: Bokforlaget Thales, 1988. 9, 30

Amartya Sen. Equality of what? *The Tanner lecture on human values*, 1, 1979. 144

Amartya Sen. The possibility of social choice. *American economic review*, 89(3):349–378, 1999. 87

Shai Shalev-Shwartz et al. Online learning and online convex optimization. *Foundations and Trends® in Machine Learning*, 4(2):107–194, 2012. 29, 163

Ohad Shamir and Tong Zhang. Stochastic gradient descent for non-smooth optimization: Convergence results and optimal averaging schemes. In *International conference on machine learning*, pages 71–79. PMLR, 2013. 26, 70

Anthony F Shorrocks. The class of additively decomposable inequality measures. *Econometrica: Journal of the Econometric Society*, pages 613–625, 1980. 36, 160

Anthony F Shorrocks. Ranking income distributions. *Economica*, 50(197):3–17, 1983. 10, 31, 32, 34, 35, 41, 43, 57, 58, 143

Umer Siddique, Paul Weng, and Matthieu Zimmer. Learning fair policies in multi-objective (deep) reinforcement learning with average and discounted rewards. In *International Conference on Machine Learning*, pages 8905–8915. PMLR, 2020. 28, 38, 70, 163

Ashudeep Singh and Thorsten Joachims. Fairness of exposure in rankings. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2219–2228. ACM, 2018. 6, 8, 10, 13, 15, 21, 23, 25, 26, 27, 28, 41, 42, 45, 47, 50, 54, 55, 57, 58, 63, 64, 69, 71, 76, 81, 90, 91, 102, 105, 106, 107, 141, 154, 161, 164, 245, 248

Ashudeep Singh and Thorsten Joachims. Policy learning for fairness in ranking. *Advances in Neural Information Processing Systems*, 32, 2019. 28, 41, 45, 46, 74, 102, 141, 160

Rahul Singh, Abhishek Gupta, and Ness B Shroff. Learning in markov decision processes under constraints. *arXiv preprint arXiv:2002.12435*, 2020. 220

Ayan Sinha, David F Gleich, and Karthik Ramani. Deconvolving feedback loops in recommender systems. *Advances in neural information processing systems*, 29, 2016. 108

Piotr Skowron, Piotr Faliszewski, and Jérôme Lang. Finding a collective set of items: From proportional multirepresentation to group recommendation. *Artificial Intelligence*, 241:191–216, 2016a. 103

Piotr Skowron, Martin Lackner, Markus Brill, Dominik Peters, and Edith Elkind. Proportional rankings. *arXiv preprint arXiv:1612.01434*, 2016b. 103, 104

Aleksandrs Slivkins. Introduction to multi-armed bandits. *arXiv preprint arXiv:1904.07272*, 2019. 78

Till Speicher, Hoda Heidari, Nina Grgic-Hlaca, Krishna P Gummadi, Adish Singla, Adrian Weller, and Muhammad Bilal Zafar. A unified approach to quantifying algorithmic unfairness: Measuring individual &group unfairness via inequality indices. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 2239–2248. ACM, 2018. 37, 38, 51, 70, 103

Harald Steck. Calibrated recommendations. In *Proceedings of the 12th ACM conference on recommender systems*, pages 154–162, 2018. 6, 245

Julia Stoyanovich, Ke Yang, and HV Jagadish. Online set selection with fairness and diversity constraints. In *Proceedings of the EDBT Conference*, 2018. 215

Jonathan Stray, Ivan Vendrov, Jeremy Nixon, Steven Adler, and Dylan Hadfield-Menell. What are you optimizing for? aligning recommender systems with human values. *arXiv preprint arXiv:2107.10939*, 2021. 28, 74, 84

Yi Su, Magd Bayoumi, and Thorsten Joachims. Optimizing rankings for recommendation in matching markets. *arXiv preprint arXiv:2106.01941*, 2021. 63, 71

Harini Suresh and John V Guttag. A framework for understanding unintended consequences of machine learning. *arXiv preprint arXiv:1901.10002*, 2019. 105, 197

Vinith M Suriyakumar, Marzyeh Ghassemi, and Berk Ustun. When personalization harms: Reconsidering the use of group attributes in prediction. *arXiv preprint arXiv:2206.02058*, 2022. 38

Latanya Sweeney. Discrimination in online ad delivery. *Queue*, 11(3):10, 2013. 5, 20, 50, 54, 87, 142, 242

Michael Taylor, John Guiver, Stephen Robertson, and Tom Minka. Softrank: optimizing non-smooth rank metrics. In *Proceedings of the 2008 International Conference on Web Search and Data Mining*, pages 77–86, 2008. 27, 70

Larry S Temkin. *Inequality*. Oxford University Press, 1993. 23

Kiran K Thekumparampil, Prateek Jain, Praneeth Netrapalli, and Sewoong Oh. Projection efficient subgradient method and optimal nonsmooth frank-wolfe method. In H. Larochelle, M. Ranzato, R. Hadsell, M. F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 12211–12224. Curran Associates, Inc., 2020a. URL https://proceedings.neurips.cc/paper/2020/file/8f468c873a32bb0619eaeb2050ba45d1-Paper.pdf. 26, 55, 60, 70

Kiran K Thekumparampil, Prateek Jain, Praneeth Netrapalli, and Sewoong Oh. Projection efficient subgradient method and optimal nonsmooth frank-wolfe method. *Advances in Neural Information Processing Systems*, 33:12211–12224, 2020b. 179

Paul D Thistle. Ranking distributions with generalized lorenz curves. *Southern Economic Journal*, pages 1–12, 1989. 31, 143

William Thomson. Fair allocation rules. In *Handbook of social choice and welfare*, volume 2, pages 393–506. Elsevier, 2011. 37

Nenad Tomasev, Kevin R McKee, Jackie Kay, and Shakir Mohamed. Fairness for unobserved characteristics: Insights from technological impacts on queer communities. In *Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society*, pages 254–265, 2021. 103

Twitter. Twitter's recommendation algorithm. https://blog.twitter.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm, 2023. 108

Berk Ustun, Yang Liu, and David Parkes. Fairness without harm: Decoupled classifiers with preference guarantees. In *International Conference on Machine Learning*, pages 6373–6382, 2019. 20, 37, 38, 51, 87, 88, 90

Nicolas Usunier, Virginie Do, and Elvis Dohmatob. Fast online ranking with fairness of exposure. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 2157–2167, 2022. v, 21, 28, 74, 164

Peter Vamplew, Richard Dazeley, Cameron Foale, Sally Firmin, and Jane Mummery. Human-aligned artificial intelligence is a multiobjective problem. *Ethics and Information Technology*, 20(1): 27–40, 2018. 28, 74

Michael Veale and Reuben Binns. Fairer machine learning in the real world: Mitigating discrimination without collecting sensitive data. *Big Data & Society*, 4(2):2053951717743530, 2017. 103

Madalina Vlasceanu and David M Amodio. Propagation of societal gender inequality by internet search algorithms. *Proceedings of the National Academy of Sciences*, 119(29):e2204529119, 2022. 20

Lequn Wang and Thorsten Joachims. Fairness and diversity for rankings in two-sided markets. *arXiv preprint arXiv:2010.01470*, 2020. 50, 154

Lequn Wang and Thorsten Joachims. User fairness, item fairness, and diversity for rankings in two-sided markets. In *Proceedings of the 2021 ACM SIGIR International Conference on Theory of Information Retrieval*, pages 23–41, 2021. 8, 15, 21, 26, 56, 63, 64, 69, 248

Lequn Wang and Thorsten Joachims. Uncertainty quantification for fairness in two-stage recommender systems. In *Proceedings of the Sixteenth ACM International Conference on Web Search and Data Mining*, pages 940–948, 2023. 109

Lequn Wang, Yiwei Bai, Wen Sun, and Thorsten Joachims. Fairness of exposure in stochastic bandits. In *International Conference on Machine Learning*, pages 10686–10696. PMLR, 2021a. 28, 75, 164

Menghan Wang, Mingming Gong, Xiaolin Zheng, and Kun Zhang. Modeling dynamic missingness of implicit feedback for recommendation. *Advances in neural information processing systems*, 31: 6669, 2018. 157, 198

Xuezhi Wang, Nithum Thain, Anu Sinha, Flavien Prost, Ed H Chi, Jilin Chen, and Alex Beutel. Practical compositional fairness: Understanding fairness in multi-component recommender systems. In *Proceedings of the 14th ACM International Conference on Web Search and Data Mining*, pages 436–444, 2021b. 109

Yifan Wang, Weizhi Ma, Min Zhang, Yiqun Liu, and Shaoping Ma. A survey on the fairness of recommender systems. *ACM Transactions on Information Systems*, 41(3):1–43, 2023. 21, 22

Yixin Wang, Dawen Liang, Laurent Charlin, and David M Blei. Causal inference for recommender systems. In *Proceedings of the 14th ACM Conference on Recommender Systems*, pages 426–431, 2020. 108

Yuyan Wang, Mohit Sharma, Can Xu, Sriraj Badam, Qian Sun, Lee Richardson, Lisa Chung, Ed H Chi, and Minmin Chen. Surrogate for long-term user experience in recommender systems. In *Proceedings of the 28th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 4100–4109, 2022. 105, 108

Romain Warlop, Alessandro Lazaric, and Jérémie Mary. Fighting boredom in recommender systems with linear reinforcement learning. *Advances in Neural Information Processing Systems*, 31, 2018. 105

Tsachy Weissman, Erik Ordentlich, Gadiel Seroussi, Sergio Verdu, and Marcelo J Weinberger. Inequalities for the l1 deviation of the empirical distribution. *Hewlett-Packard Labs, Tech. Rep*, 2003. 220

John A Weymark. Generalized gini inequality indices. *Mathematical Social Sciences*, 1(4):409–430, 1981. 26, 35, 54, 56

Colin Wilkie and Leif Azzopardi. Best and fairest: An empirical analysis of retrieval system bias. In *European Conference on Information Retrieval*, pages 13–25. Springer, 2014. 69

Robert Williamson and Aditya Menon. Fairness risk measures. In *International Conference on Machine Learning*, pages 6786–6797. PMLR, 2019. 38, 70, 103

Ouri Wolfson and Jane Lin. Fairness versus optimality in ridesharing. In *2017 18th IEEE International Conference on Mobile Data Management (MDM)*, pages 118–123. IEEE, 2017. 22, 51

Haolun Wu, Chen Ma, Bhaskar Mitra, Fernando Diaz, and Xue Liu. Multi-fr: A multi-objective optimization method for achieving two-sided fairness in e-commerce recommendation. *arXiv preprint arXiv:2105.02951*, 2021a. 70

Haolun Wu, Chen Ma, Bhaskar Mitra, Fernando Diaz, and Xue Liu. A multi-objective optimization framework for multi-stakeholder fairness-aware recommendation. *ACM Transactions on Information Systems (TOIS)*, 2022a. 21, 28

Haolun Wu, Bhaskar Mitra, Chen Ma, Fernando Diaz, and Xue Liu. Joint multisided exposure fairness for recommendation. *arXiv preprint arXiv:2205.00048*, 2022b. 21, 28, 164

Qingyun Wu, Hongning Wang, Liangjie Hong, and Yue Shi. Returning is believing: Optimizing long-term user engagement in recommender systems. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1927–1936, 2017. 108

Yao Wu, Jian Cao, Guandong Xu, and Yudong Tan. Tfrom: A two-sided fairness-aware recommendation model for both customers and providers. *arXiv preprint arXiv:2104.09024*, 2021b. 21, 42, 45, 51, 57, 69, 71, 154

Yifan Wu, Roshan Shariff, Tor Lattimore, and Csaba Szepesvári. Conservative bandits. In *International Conference on Machine Learning*, pages 1254–1262, 2016. 29, 88, 89, 93, 94, 206

Bin Xia, Junjie Yin, Jian Xu, and Yun Li. We-rec: A fairness-aware reciprocal recommendation based on walrasian equilibrium. *Knowledge-Based Systems*, 182:104857, 2019. 22, 51

Peng Xia, Benyuan Liu, Yizhou Sun, and Cindy Chen. Reciprocal recommendation system for online dating. In *2015 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 234–241. IEEE, 2015. 22, 70

Huanle Xu, Yang Liu, Wing Cheong Lau, and Rui Li. Combinatorial multi-armed bandits with concave rewards and fairness constraints. In *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, pages 2554–2560, 2021. 28, 75, 164

Ronald R Yager. On ordered weighted averaging aggregation operators in multicriteria decision-making. *IEEE Transactions on systems, Man, and Cybernetics*, 18(1):183–190, 1988. 34, 35, 36, 56

Ke Yang and Julia Stoyanovich. Measuring fairness in ranked outputs. In *Proceedings of the 29th international conference on scientific and statistical database management*, pages 1–6, 2017. 21, 27, 104, 107, 163

Ke Yang, Vasilis Gkatzelis, and Julia Stoyanovich. Balanced ranking with diversity constraints. *arXiv preprint arXiv:1906.01747*, 2019. 108

Nicholas C. Yannelis. *Integration of Banach-Valued Correspondence*, pages 2–35. Springer Berlin Heidelberg, Berlin, Heidelberg, 1991. ISBN 978-3-662-07071-0. 192

Sirui Yao and Bert Huang. Beyond parity: Fairness objectives for collaborative filtering. In *Advances in Neural Information Processing Systems*, pages 2921–2930, 2017. 6, 245

Sirui Yao, Yoni Halpern, Nithum Thain, Xuezhi Wang, Kang Lee, Flavien Prost, Ed H Chi, Jilin Chen, and Alex Beutel. Measuring recommender system effects with simulated users. *arXiv preprint arXiv:2101.04526*, 2021. 108

Shlomo Yitzhaki and Edna Schechtman. More than a dozen alternative ways of spelling gini. In *The Gini Methodology*, pages 11–31. Springer, 2013. 34, 57, 65

Kôsaku Yosida et al. Functional analysis. 1965. 15, 27, 55, 249

Farzad Yousefian, Angelia Nedić, and Uday V Shanbhag. On stochastic gradient and subgradient methods with adaptive steplength sequences. *Automatica*, 48(1):56–67, 2012. 180

YouTube. On youtube's recommendation system. https://blog.youtube/inside-youtube/on-youtubes-recommendation-system/, 2021. 108, 109

Eric Yang Yu, Zhizhen Qin, Min Kyung Lee, and Sicun Gao. Policy optimization with advantage regularization for long-term fairness in decision systems. *arXiv preprint arXiv:2210.12546*, 2022. 108

Alp Yurtsever, Olivier Fercoq, Francesco Locatello, and Volkan Cevher. A conditional gradient framework for composite convex minimization with applications to semidefinite programming. In *International Conference on Machine Learning*, pages 5727–5736. PMLR, 2018. 26, 55, 60, 70, 179

Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez Rodriguez, and Krishna P Gummadi. Fairness beyond disparate treatment & disparate impact: Learning classification without disparate mistreatment. In *Proceedings of the 26th International Conference on World Wide Web*, pages 1171–1180. International World Wide Web Conferences Steering Committee, 2017a. 6, 7, 245

Muhammad Bilal Zafar, Isabel Valera, Manuel Rodriguez, Krishna Gummadi, and Adrian Weller. From parity to preference-based notions of fairness in classification. In *Advances in Neural Information Processing Systems*, pages 228–238, 2017b. 20, 38, 87, 88

Muhammad Bilal Zafar, Isabel Valera, Manuel Gomez-Rodriguez, and Krishna P Gummadi. Fairness constraints: A flexible approach for fair classification. *The Journal of Machine Learning Research*, 20(1):2737–2778, 2019. 6, 7, 245

Andrea Zanette and Emma Brunskill. Tighter problem-dependent regret bounds in reinforcement learning without domain knowledge using value function bounds. *arXiv preprint arXiv:1901.00210*, 2019. 227, 232

Meike Zehlike and Carlos Castillo. Reducing disparate exposure in ranking: A learning to rank approach. In *Proceedings of The Web Conference 2020*, pages 2849–2855, 2020. 21, 28, 41, 50, 54, 57, 74, 76, 106, 164

Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, and Ricardo Baeza-Yates. Fa* ir: A fair top-k ranking algorithm. In *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*, pages 1569–1578, 2017. 21, 107, 108

Meike Zehlike, Ke Yang, and Julia Stoyanovich. Fairness in ranking: A survey. *arXiv preprint arXiv:2103.14000*, 2021. 164

Meike Zehlike, Ke Yang, and Julia Stoyanovich. Fairness in ranking, part i: Score-based ranking. *ACM Computing Surveys*, 55(6):1–36, 2022a. 21, 22, 23, 107, 108

Meike Zehlike, Ke Yang, and Julia Stoyanovich. Fairness in ranking, part ii: Learning-to-rank and recommender systems. *ACM Computing Surveys*, 55(6):1–41, 2022b. 22

ChengXiang Zhai, William W Cohen, and John Lafferty. Beyond independent relevance: methods and evaluation metrics for subtopic retrieval. In *Acm sigir forum*, volume 49, pages 2–9. ACM New York, NY, USA, 2015. 108

Ruohan Zhan, Konstantina Christakopoulou, Ya Le, Jayden Ooi, Martin Mladenov, Alex Beutel, Craig Boutilier, Ed Chi, and Minmin Chen. Towards content provider aware recommender systems: A simulation study on the interplay between user and provider utilities. In *Proceedings of the Web Conference 2021*, pages 3872–3883, 2021. 105, 108

Tong Zhang et al. Statistical behavior and consistency of classification methods based on convex risk minimization. *The Annals of Statistics*, 32(1):56–85, 2004. 145

Liyuan Zheng and Lillian J Ratliff. Constrained upper confidence reinforcement learning. *arXiv preprint arXiv:2001.09377*, 2020. 220

Yong Zheng, Tanaya Dave, Neha Mishra, and Harshit Kumar. Fairness in reciprocal recommendations: A speed-dating study. In *Adjunct publication of the 26th conference on user modeling, adaptation and personalization*, pages 29–34, 2018. 11

Yinglun Zhu and Paul Mineiro. Contextual bandits with smooth regret: Efficient learning in continuous action spaces. In *International Conference on Machine Learning*, pages 27574–27590. PMLR, 2022. 84

Yinglun Zhu, Dylan J Foster, John Langford, and Paul Mineiro. Contextual bandits with large action spaces: Made practical. In *International Conference on Machine Learning*, pages 27428–27453. PMLR, 2022. 84

Cai-Nicolas Ziegler, Sean M McNee, Joseph A Konstan, and Georg Lausen. Improving recommendation lists through topic diversification. In *Proceedings of the 14th international conference on World Wide Web*, pages 22–32, 2005. 108

Matthieu Zimmer, Claire Glanois, Umer Siddique, and Paul Weng. Learning fair policies in decentralized cooperative multi-agent reinforcement learning. In *International Conference on Machine Learning*, pages 12967–12978. PMLR, 2021. 37, 38, 51, 70

Indre Zliobaite. On the relation between accuracy and fairness in binary classification. *arXiv preprint arXiv:1505.05723*, 2015. 6, 245

Appendix A

Appendix of Chapter 3

A.1 Outline of the appendix

These appendices are structured as follows:

- In Appendix A.2, we present how our fairness framework can be applied to sensitive groups of users or categories of items.
- In Appendix A.3, we present a deeper analysis of the trade-offs achieved by the welfare approach. We also provide a theoretical guarantee relating the true welfare obtained by maximizing the welfare using estimated preferences, depending on the quality of the estimates.
- In Appendix A.4, we present the proofs for the theoretical results comparing our results and previous criteria of fairness in rankings. In addition, in Appendix A.4.3, we describe how to extend the criteria of equality of exposure and quality-weighted exposure in a reciprocal recommendation setting. This is the extension used in our experiments on reciprocal recommendation. In Proposition 25, we present an additional result regarding the inefficiency of these criteria in reciprocal recommendation.
- In Appendix A.5, we present the more general version of the Frank-Wolfe algorithm, which we use both to optimize the welfare function over stochastic rankings, as well as the penalty-based baselines. This appendix also contains the proofs of the results in Section 3.4. In addition, this appendix contains fundamental lemmas that are used in other appendices.
- Appendix A.6 gives the details of the experiments presented in Section 3.5, as well we many additional experiments (two additional, larger scale datasets on one-sided recommendation, and an additional dataset for reciprocal recommendation)
- Appendix A.7 briefly discusses the difference between the penalty we use in our implementation of the baseline approaches and an alternative penalty used by some authors.
- Finally, Appendix A.8 discusses the difference between applying item-side fairness criteria for every ranking, compared to what we do in the paper, which defines item-side utility as an aggregate over the rankings of all users.

A.2 Fairness towards sensitive groups rather than individuals

In all the paper we focus on fairness towards individual users and items rather than groups of users or items. Prior work [Singh and Joachims, 2018, Morik et al., 2020, Singh and Joachims, 2019] considered the utlity of a group as the sum or the average utility of its members. Using this definition of group utility, our framework dirrectly extends to groups rather than individuals.

In this section we describe the case of one-sided recommendation with groups of users and item categories. The case of reciprocal recommendation (with user groups only) is similar but simpler.

Let $\mathcal{S} = (s_p)_{p=1}^{|\mathcal{S}|}$ be (possibly overlapping) user groups, i.e., $\forall p \in [\![|\mathcal{S}|]\!], s_p \subseteq \mathcal{N}$ and $\cup_{p\in[\![|\mathcal{S}|]\!]} s_p = \mathcal{N}$. Similarly, let $\mathcal{C} = (c_q)_{q=1}^{|\mathcal{C}|}$ be (possibly overlapping) item categories, i.e., $\forall q \in [\![|\mathcal{C}|]\!], c_q \subseteq \mathcal{I}$ and $\cup_{q\in[\![|\mathcal{C}|]\!]} c_q = \mathcal{I}$. On the user side, such groups would typically correspond to demographic groups considered sensitive for the application at hand [Sweeney, 2013]. On the item side, groups can represent a single producer for the case where we want to be fair to producers based on the aggregate utility they obtain from their products [Mehrotra et al., 2018], or demographic groups as well [Kay et al., 2015].

In all cases, we redefine the user-side utility for groups and the item-side utility for categories:

$$u_{s_p}^{\mathrm{gr}}(P) = \sum_{i\in s_p} u_i(P) \qquad\qquad u_{c_q}^{\mathrm{cat}}(P) = \sum_{j\in c_q} u_j(P)$$

Let $\boldsymbol{u}^{\mathrm{gr}}(P) = (u_{s_p}^{\mathrm{gr}}(P))_{p=1}^{|\mathcal{S}|}$ and $\boldsymbol{u}^{\mathrm{cat}}(P) = (u_{c_q}^{\mathrm{cat}}(P))_{q=1}^{|\mathcal{C}|}$ be the utility profiles of user groups and item categories associated to P respectively. The two-sided Lorenz efficiency for groups and categories is defined as:

Definition 7. *Let $\mathcal{S}$ be a set of user groups and $\mathcal{C}$ a set of item categories. Let $P \in \mathcal{P}$. P is $(\mathcal{S},\mathcal{C})$-Lorenz-efficient if there is no $P' \in \mathcal{P}$ such that either condition holds:*

1. $\boldsymbol{u}^{\mathrm{gr}}(P') \succeq_{\mathrm{L}} \boldsymbol{u}^{\mathrm{gr}}(P)$ *and* $\boldsymbol{u}^{\mathrm{cat}}(P') \succ_{\mathrm{L}} \boldsymbol{u}^{\mathrm{cat}}(P)$, *or*
2. $\boldsymbol{u}^{\mathrm{cat}}(P') \succeq_{\mathrm{L}} \boldsymbol{u}^{\mathrm{cat}}(P)$ *and* $\boldsymbol{u}^{\mathrm{gr}}(P') \succ_{\mathrm{L}} \boldsymbol{u}^{\mathrm{gr}}(P)$.

The welfare function associated to $(\mathcal{S},\mathcal{C})$, still parametrized by $\theta = (\lambda, \alpha_1, \alpha_2) \in \Theta$, is defined as

$$W_\theta^{\mathrm{gr}}(P) = (1-\lambda)\sum_{s\in\mathcal{S}} \psi(u_s^{\mathrm{gr}}(P), \alpha_1) + \lambda \sum_{c\in\mathcal{C}} \psi(u_c^{\mathrm{cat}}(P), \alpha_2)$$

The welfare function follows the general form of objective function used for the algorithm in Appendix A.5, so the optimization of W_θ^{gr} requires similar computational complexity as W_θ.

Finally, the extension of Proposition 1 is straightforward. Its proof is similar to the proof presented in Appendix A.3.

Proposition 20. $\forall \theta \in \Theta, \forall P^* \in \operatorname*{argmax}_{P\in\mathcal{P}} W_\theta^{\mathrm{gr}}(P)$, *$P^*$ is $(\mathcal{S},\mathcal{C})$-Lorenz-efficient.*

Note that this way of treating groups is not necessarily optimal. In particular, in does not account for within-group fairness. The separate consideration of within-group and between-group fairness has been studied extensively in the literature on equality of opportunity [Roemer and Trannoy, 2016], which has inspired several works on algorithmic fairness [Hardt et al., 2016b, Heidari et al., 2019]. Yet, how to apply these principles to two-sided fairness in recommendation is still open, and is left as future work.

A.3 More on welfare functions

This appendix provides an in-depth analysis of the trade-offs that are achievable by the welfare approach. We first pove the proposition of Section 3.2.2, and analyze the utilitarian rankings (obtained with $\alpha_1 = \alpha_2 = 1$). We then analyze how to obtain leximin optimal solutions on the side of the items in Appendix A.3.2, as mentioned in Section 3.2.2. Finally, we prove Theorem 24 in Appendix A.3.3, which provides a regret bound relating the true welfare achieved when maximizing

welfare on estimated preferences. Some results in this section use Lemma 27 of Appendix 3.4, which is proved in Appendix 3.4.

Throughout the appendices, we use the more general version of item utilities (two-sided preferences), described at the end of Section 3.2.1. Moreover, to clarify the notation, we remind that a *ranking tensor* is a three-way tensor P where P_{ijk} is the probability that item j is recommended to user i at rank k. We consider P as an $n \times n \times |\mathcal{I}|$ tensor, where irrelevant entries are set to 0. With this notation, the utility for both users and items can be written with the same formula:

$$\forall i \in [\![n]\!], u_i(P) = \sum_{j=1}^{n} \mu_{ij}(P_{ij} + P_{ji})v.$$

Note that this formula also corresponds to the two-sided utility in reciprocal recommendation. In general, the results in this appendix can be extended to reciprocal recommendation with minimal changes to their proofs, using $\mathcal{N} = \mathcal{I} = [\![n]\!]$ and the formula above for the utility.

A.3.1 Lorenz efficiency and utilitarian ranking

We first prove Proposition 1:

Proposition 1. $\forall \theta \in \Theta, \forall P^* \in \underset{P \in \mathcal{P}}{\operatorname{argmax}} W_\theta(\boldsymbol{u}(P))$, P^* *is Lorenz-efficient.*

Proof. It is well known that if Φ is increasing and strictly concave, then $F(\boldsymbol{u}) = \sum_{i=1}^{n} \Phi(u_i)$ is monotonic with respect to Lorenz dominance [Shorrocks, 1983, Thistle, 1989]: $\boldsymbol{u} \succ_{\mathrm{L}} \boldsymbol{u'} \implies F(\boldsymbol{u}) > F(\boldsymbol{u'})$.

In the case of W_θ, for every $\theta = (\lambda, \alpha_1, \alpha_2) \in \Theta$, both $\psi(., \alpha_1)$ and $\psi(., \alpha_2)$ are strictly concave by the definition of Θ (recall that in Θ, we have $\alpha_1, \alpha_2 < 1$).

The partial function[1] $\boldsymbol{u}'_{\mathcal{N}} \mapsto W_\theta((\boldsymbol{u}_{\mathcal{I}}, \boldsymbol{u}'_{\mathcal{N}}))$ is, up to a constant, of the form of F and likewise for the partial function $\boldsymbol{u}'_{\mathcal{I}} \mapsto W_\theta((\boldsymbol{u}'_{\mathcal{I}}, \boldsymbol{u}_{\mathcal{N}}))$. We now prove the result by contradiction. Assume that $\boldsymbol{u} \in \operatorname{argmax}_{\boldsymbol{u} \in \mathcal{U}} W_\theta(\boldsymbol{u})$ is not Lorenz-efficient. Then there is $\boldsymbol{u'} \in \mathcal{U}$ such that ($\boldsymbol{u}'_{\mathcal{N}} \succeq_{\mathrm{L}} \boldsymbol{u}_{\mathcal{N}}$ and $\boldsymbol{u}'_{\mathcal{I}} \succ_{\mathrm{L}} \boldsymbol{u}_{\mathcal{I}}$) or ($\boldsymbol{u}'_{\mathcal{N}} \succ_{\mathrm{L}} \boldsymbol{u}_{\mathcal{N}}$ and $\boldsymbol{u}'_{\mathcal{I}} \succeq_{\mathrm{L}} \boldsymbol{u}_{\mathcal{I}}$). Let us assume ($\boldsymbol{u}'_{\mathcal{N}} \succeq_{\mathrm{L}} \boldsymbol{u}_{\mathcal{N}}$ and $\boldsymbol{u}'_{\mathcal{I}} \succ_{\mathrm{L}} \boldsymbol{u}_{\mathcal{I}}$), the other case is dealt with similarly. We then have:

$$\begin{aligned} W_\theta(\boldsymbol{u'}) &\geq W_\theta((\boldsymbol{u}'_{\mathcal{I}}, \boldsymbol{u}_{\mathcal{N}})) && \text{(because } \boldsymbol{u}'_{\mathcal{N}} \succeq_{\mathrm{L}} \boldsymbol{u}_{\mathcal{N}}\text{)} \\ &> W_\theta((\boldsymbol{u}_{\mathcal{I}}, \boldsymbol{u}_{\mathcal{N}})) && \text{(because } \boldsymbol{u}'_{\mathcal{I}} \succ_{\mathrm{L}} \boldsymbol{u}_{\mathcal{I}}\text{)} \end{aligned}$$

which contradicts the maximality of $\boldsymbol{u}$. □

The analogous for Proposition 1 for reciprocal recommendation is a direct consequence of standard results that concave welfare functions are monotonic with respect to Lorenz dominance [Shorrocks, 1983, Thistle, 1989].

Utilitarian ranking Proposition 21 below generalizes to two-sided utilities the well-known result that maximizing user-side utility is achieved by sorting $j \in \mathcal{I}$ by decreasing μ_{ij} (see e.g., [Cossock and Zhang, 2008]). For a ranking tensor P and a user i, we denote by $\mathfrak{S}(P_i)$ the support of P_i in ranking space.[2] We remind that $\sigma(j)$ is the rank of item j, and that lower ranks are better. For a user i and item j, we use $\mu_{ji} = 1$.

[1] We denote by $(\boldsymbol{u}_{\mathcal{I}}, \boldsymbol{u}'_{\mathcal{N}})$ the vector $\mathbb{R}^d$ such that $(\boldsymbol{u}_{\mathcal{I}}, \boldsymbol{u}'_{\mathcal{N}})_i = u_i$ if $i \in \mathcal{I}$ and $(\boldsymbol{u}_{\mathcal{I}}, \boldsymbol{u}'_{\mathcal{N}})_i = u'_i$ if $i \in \mathcal{N}$.

[2] Formally, $\mathfrak{S}(P_i) = \left\{\sigma : \mathcal{I} \to [\![|\mathcal{I}|]\!] \,\middle|\, \sigma \text{ is one-to-one, and } \forall j \in \mathcal{I}, P_{ij\sigma(j)} > 0\right\}$.

Proposition 21 (Utilitarian ranking). *Assume* $\forall k \in [\![n-1]\!], v_k > v_{k+1} \geq 0$ *and let*

$$P^* \in \underset{P\in\mathcal{P}}{\operatorname{argmax}} W_{\frac{1}{2},1,1}(P) = \underset{P\in\mathcal{P}}{\operatorname{argmax}} \sum_{i\in[\![n]\!]} u_i(P).$$

1. $\forall i \in \mathcal{N}, \forall \sigma \in \mathfrak{S}(P_i^*) : \sigma(j) < \sigma(j') \implies \tilde{\mu}_{ij} \geq \tilde{\mu}_{ij'}$ *with* $\tilde{\mu}_{ij} = \mu_{ij} + \mu_{ji}$.
2. *If* $\forall (i,j) \in [\![n]\!]^2, \mu_{ij} = \mu_{ji}$, *then* $\tilde{\mu}_{ij} \geq \tilde{\mu}_{ij'} \iff \mu_{ij} \geq \mu_{ij'}$.

When mutual preferences are symmetric (i.e., $\mu_{ij} = \mu_{ji}$), the utilitarian ranking is the same as the usual sort by decreasing μ_{ij}. This also obviously holds when we consider exposue as item utility ($\mu_{ji} = 1$). This means that without considerations of two-sided fairness ($\alpha_1, \alpha_2 < 1$), the optimal ranking for two-sided utilities is the same as the usual ranking. This might explain why the two-sided utility has never been studied before, even in reciprocal recommendation [Palomares et al., 2021].

For the proof of Proposition 21, the main part is the following lemma:

Lemma 22. *Let* $F(\boldsymbol{u}(P)) = \sum_{i=1}^{n} u_i(P)$ *and* $\tilde{\mu}_{ij} = \mu_{ij} + \mu_{ji}$. *Assume* $\forall k \in [\![n-1]\!], v_k \geq v_{k+1} \geq 0$.
If $P^* \in \mathcal{P}$ *is such that* $\forall \sigma \in \mathfrak{S}(P_i^*), \forall j, j', \sigma(j) < \sigma(j') \implies \tilde{\mu}_{ij} \geq \tilde{\mu}_{ij'}$ *then* $P^* \in \underset{P\in\mathcal{P}}{\operatorname{argmax}} \boldsymbol{u}(P)$.
Moreover, if $\forall k \in [\![n-1]\!], v_k > v_{k+1} \geq 0$, *then the reciprocal is true.*

Proof. Notice that, thanks to the completion of P with zeros on irrelevant entries and formula A.3, $F(\boldsymbol{u}(P))$ can be rewritten as:

$$F(\boldsymbol{u}(P)) = \sum_{i=1}^{n} u_i(P) = \sum_{i=1}^{n}\sum_{j=1}^{n} \mu_{ij}(P_{ij} + P_{ji})v = \sum_{i=1}^{n}\sum_{j=1}^{n} (\mu_{ij} + \mu_{ji})P_{ij}v$$

where the last equality is obtained by swapping i and j in the second sum, which is possible since i and j span the same range.

The result is then a direct consequence of Lemma 27 in Appendix A.5, using $A_{ij} = \mu_{ij} + \mu_{ji}$. □

The first of statement of Proposition 21 assumes that the exposure weights v are non-negative and strictly decreasing as per the second point of Lemma 22. Lemma 22 above gives the statement for the more general case of non-increasing v.

Proof of Proposition 21. The first statement is the consequence of Lemma 22 above, noticing that $F(\boldsymbol{u}(P))$ in Lemma 22 always has the same argmax. The second statement is obvious from the assumptions.

□

A.3.2 Item-side leximin optimality

The most egalitarian trade-off achievable by our method is described by the leximin order [Sen, 1979]. Given two utility profiles $\boldsymbol{u}$ and $\boldsymbol{u'}$, $\boldsymbol{u} \geq_{\text{lex}} \boldsymbol{u'}$ if $\boldsymbol{U}$ is greater than $\boldsymbol{U'}$ according to the lexicographic order.[3] The leximin optimal profile is egalitarian in the sense that it maximizes the utility of individuals in sequence, from the worse-off to the better-off. Depending on the set of feasible profiles, this may not lead to equal utility for everyone, but any further reduction of inequality can only be achieved by making people worse off for the benefit of no other, in violation of Pareto-dominance.

The proposition below formalizes how leximin optimal solutions on the side of items are found. It shows that item-side leximin solutions are obtained by having $\alpha_2 \to -\infty$ and $\lambda \to 1$ at the same

[3] Formally, $\boldsymbol{u} >_{\text{lex}} \boldsymbol{u'}$ if $(\exists k \in [\![d]\!]$ s.t. $\forall i < k$, $U_i = U'_i$ and $U_k > U'_k)$. $\boldsymbol{u} \geq_{\text{lex}} \boldsymbol{u'} \iff \neg(\boldsymbol{u'} \geq_{\text{lex}} \boldsymbol{u})$.

time. The proposition gives a formal statement of the rate at which λ should converge to 1 relative to α.

In the statement of the proposition, given two functions F and G, we use $F(\alpha) \underset{\alpha\to-\infty}{\geq} G(\alpha)$ as a shorthand for $F(\alpha) \geq G(\alpha)$ for α sufficiently small.[4].

Proposition 23. *Let* $\mathcal{U}_{\text{lex}}^{\text{item}} = \{\boldsymbol{u} \in \mathcal{U} : \forall \boldsymbol{u'} \in \mathcal{U}, \boldsymbol{u}_{\mathcal{I}} \geq_{\text{lex}} \boldsymbol{u'}_{\mathcal{I}}\}$ *and let* $\boldsymbol{u}^* = \underset{\boldsymbol{u}\in\mathcal{U}_{\text{lex}}^{\text{item}}}{\operatorname{argmax}} \sum_{i\in\mathcal{N}} \psi(u_i, \alpha_1)$.

$$\forall \eta > \max(1, \|\boldsymbol{u^*}_{\mathcal{I}}\|_\infty), \forall \boldsymbol{u} \in \mathcal{U} : W_{1-\eta^\alpha,\alpha_1,\alpha}(\boldsymbol{u}^*) \underset{\alpha\to-\infty}{\geq} W_{1-\eta^\alpha,\alpha_1,\alpha}(\boldsymbol{u}).$$

This means that among the leximin-optimal item-side utility profiles, α_1 still controls the redistribution profile on the user side, since it is possible that $|\mathcal{U}_{\text{lex}}^{\text{item}}| > 1$ in one-sided recommendation. A similar result holds for user-side item leximin.

Proof. Let $\boldsymbol{u}^* = \underset{\boldsymbol{u}\in\mathcal{U}_{\text{lex}}^{\text{item}}}{\operatorname{argmax}} \sum_{i\in\mathcal{N}} \psi(u_i, \alpha_1)$ and $\boldsymbol{u} \in \mathcal{U}$. Let $\theta = (\lambda, \alpha_1, \alpha)$ and take $\alpha < \min(0, \alpha_1)$.

Let $(j_1, j_2, \dots, j_{|\mathcal{I}|})$ be the ranking of $\boldsymbol{u}^*_{\mathcal{I}}$ in increasing order: $u^*_{j_1} \leq \dots u^*_{j_{|\mathcal{I}|}}$. Likewise, let $(j'_1, j'_2, \dots, j'_{|\mathcal{I}|})$ be the ranking of $\boldsymbol{u}_{\mathcal{I}}$ in increasing order: $u_{j'_1} \leq \dots \leq u_{j'_{|\mathcal{I}|}}$.

Let $m = \max\{k \in [\![|\mathcal{I}|]\!] \cup \{0\} : \forall \ell \leq k, u^*_{j_\ell} = u_{j'_\ell}\} + 1$, be the last index (+1) such that the smallest values of $\boldsymbol{u}^*$ and $\boldsymbol{u}$ are equal ($m = 1$ if the smallest values are different).

Let $C(\alpha) = W_{1-\eta^\alpha,\alpha_1,\alpha}(\boldsymbol{u}^*) - W_{1-\eta^\alpha,\alpha_1,\alpha}(\boldsymbol{u})$.

Let $K = \sum_{i\in\mathcal{N}} \big(\psi(u^*_i, \alpha_1) - \psi(u'_i, \alpha_1)\big)$.

case 1: $m = |\mathcal{I}| + 1$. Then $C(\alpha) = (1-\eta^\alpha)K \geq 0$ since $\boldsymbol{u}^*_{\mathcal{I}} = \boldsymbol{u}_{\mathcal{I}}$ and $\boldsymbol{u}^*$ maximizes the user-side welfare.

case 2: $m < |\mathcal{I}|$. Then, we have $u_{j'_m} < u^*_{j_m}$ by the leximin optimality of $\boldsymbol{u}^*_{\mathcal{I}}$. We then have:

$$\begin{aligned} C(\alpha) &= (1-\eta^\alpha)K + \eta^\alpha \sum_{j\in\mathcal{I}} -(u^*_j)^\alpha + (u_j)^\alpha \\ &= -(1-\eta^\alpha)(u^*_{j_m})^\alpha \Big(\frac{K}{1-\eta^\alpha} \underbrace{\big(\frac{\eta}{u^*_{j_m}}\big)^\alpha}_{\xrightarrow[\alpha\to-\infty]{} 0} + 1 + \sum_{k>m} \underbrace{\big(\frac{u^*_{j_k}}{u^*_{j_m}}\big)^\alpha}_{\xrightarrow[\alpha\to-\infty]{} 0} - \underbrace{\big(\frac{u_{j'_m}}{u^*_{j_m}}\big)^\alpha}_{\xrightarrow[\alpha\to-\infty]{} +\infty} - \sum_{k>m} \underbrace{\big(\frac{u_{j'_m}}{u^*_{j_m}}\big)^\alpha}_{\geq 0} \Big) \end{aligned}$$

which implies $\lim_{\alpha\to-\infty} C(\alpha) = +\infty$ and thus the desired result. □

A.3.3 Guarantees when performing inference with estimated preferences

In practice, inference is carried out on an estimate $\hat{\mu}$ of μ, meaning that, denoting $\hat{\boldsymbol{u}}$ the resulting estimated utility[5] the system output $\hat{P} = \operatorname{argmax}_{P\in\mathcal{P}} W_\theta(\hat{\boldsymbol{u}}(P))$. The following result extends surrogate regret bounds that exist in classification [Bartlett et al., 2006, Zhang et al., 2004] and learning to rank [Cossock and Zhang, 2008, Ravikumar et al., 2011, Agarwal, 2014]to the case of welfare functions and global stochastic rankings. It makes the link between the quality of the estimate $\hat{\mu}$ and an optimality guarantee for $\boldsymbol{u}(\hat{P})$ (i.e., the true welfare of the ranking inferred on the estimated values). We prove the result for $\theta = (\frac{1}{2}, \alpha, \alpha)$ for $\alpha \leq 1$ to simplify notation.[6]

Theorem 24. *Let* $\alpha \leq 1$ *and* $\theta = (\frac{1}{2}, \alpha, \alpha) \in \Theta$. *Let* $\hat{\mu} \in \mathbb{R}_+^{|\mathcal{N}|\times|\mathcal{I}|}$, $\hat{P} = \operatorname{argmax}_{P\in\mathcal{P}} W_\theta(\hat{\boldsymbol{u}}(P))$, *and* $P^* = \operatorname{argmax}_{P\in\mathcal{P}} W_\theta(\boldsymbol{u}(P))$.

[4] Formally, $F(\alpha) \underset{\alpha\to-\infty}{\geq} G(\alpha) \iff \exists \alpha_0 \in \mathbb{R}, \forall \alpha \leq \alpha_0, F(\alpha) \geq G(\alpha)$.

[5] We have $\hat{\boldsymbol{u}}_i(P) = \sum_{j\in\mathcal{I}} \hat{\mu}_{ij} P_{ij} v$ for $i \in \mathcal{N}$.

[6] The dependency on $\hat{\mu}$ in $B(\hat{\mu})$ is because $\psi'(.,\alpha)$ is not bounded in general. In practice, we use $\psi(x+\eta,\alpha)$ for a small $\eta > 0$ to avoid the singular point at 0, in which case $B < \psi'(\eta,\alpha)$.

Let furthermore $B(\hat{\mu}) = \max\left(\max_{i\in[\![n]\!]} \psi'(u_i(\hat{P}),\alpha), \max_{i\in[\![n]\!]} \psi'(\hat{u}_i(P^*),\alpha)\right)$. *We have:*

$$W_\theta(\boldsymbol{u}(P^*)) - W_\theta(\boldsymbol{u}(\hat{P})) \leq 4B(\hat{\mu})\sqrt{n\left\|v\right\|_2^2}\sqrt{\sum_{(i,j)\in\mathcal{N}\times\mathcal{I}}(\hat{\mu}_{ij}-\mu_{ij})^2}.$$

The existing results closest to our Theorem 24 are Theorem 2 of [Cossock and Zhang, 2008]. Here the result is substantially more difficult to prove because of the concave function and the fact that utilities are two-sided, calling for considering the rankings of multiple users at once.

Proof. We have:

$$\begin{aligned} W_\theta(\boldsymbol{u}(P^*)) - W_\theta(\boldsymbol{u}(\hat{P})) &= W_\theta(\boldsymbol{u}(P^*)) - \underbrace{W_\theta(\hat{\boldsymbol{u}}(\hat{P}))}_{\geq W_\theta(\hat{\boldsymbol{u}}(P^*)} + W_\theta(\hat{\boldsymbol{u}}(\hat{P})) - W_\theta(\boldsymbol{u}(\hat{P})) \\ &\leq \underbrace{W_\theta(\boldsymbol{u}(P^*)) - W_\theta(\hat{\boldsymbol{u}}(P^*))}_{=C_1} + \underbrace{W_\theta(\hat{\boldsymbol{u}}(\hat{P})) - W_\theta(\boldsymbol{u}(\hat{P}))}_{=C_2} \end{aligned}$$

Let $B_1(\hat{\mu}) = \max_{i\in[\![n]\!]} \psi'(\hat{\boldsymbol{u}}_i(P^*),\alpha)$.

We first prove:

$$C_1 \leq 2B_1(\hat{\mu})\sqrt{n\left\|v\right\|_2^2}\sqrt{\sum_{(i,j)\in[\![n]\!]^2}(\hat{\mu}_{ij}-\mu_{ij})^2}. \tag{A.1}$$

To prove (A.1), we start by using the concavity of $\psi(.,\alpha)$ for $\alpha \leq 1$. Let $\Phi(.) = \frac{1}{2}\psi(.,\alpha)$. We have:

$$\begin{aligned} C_1 &= \sum_{i=1}^n \left(\Phi(\boldsymbol{u}(P^*)) - \Phi(\hat{\boldsymbol{u}}(P^*))\right) \leq \sum_{i=1}^n \Phi'(\hat{\boldsymbol{u}}_i(P^*))\big(\boldsymbol{u}(P^*)) - \hat{\boldsymbol{u}}(P^*)\big) \\ \text{thus}\quad C_1 &\leq \sum_{i=1}^n\sum_{j=1}^n \Phi'(\hat{\boldsymbol{u}}_i(P^*))(\mu_{ij}-\hat{\mu}_{ij})(P^*_{ij}+P^*_{ji})v \\ &= \sum_{i=1}^n\sum_{j=1}^n \big(\underbrace{\Phi'(\hat{\boldsymbol{u}}_i(P^*))(\mu_{ij}-\hat{\mu}_{ij}) + \Phi'(\hat{\boldsymbol{u}}_j(P^*))(\mu_{ji}-\hat{\mu}_{ji})}_{=A_{ij}}\big)P^*_{ij}v \end{aligned}$$

where, similarly to the proof of Lemma 22, we swapped the indexed (i,j) in the $\Phi'(\hat{\boldsymbol{u}}_i(P^*))\mu_{ij}P^*_{ji})v$, which is possible because i and j span the same range in the sum.

Notice that the terms $A_{ij}P^*_{ij}v$ are all zero except if $i\in\mathcal{N}$ and $j\in\mathcal{I}$ (because $P^*_{ijk}=0$ otherwise). For $i\in\mathcal{N}$, let σ_i be a ranking which ranks $(A_{ij})_{j\in\mathcal{I}}$ in decreasing order, i.e., $\sigma_i(j) < \sigma_i(j') \implies A_{ij} \geq A_{ij'}$. Using Lemma 27 in Appendix A.5, we have:

$$C_1 \leq \max_{P\in\mathcal{P}} \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{I}} A_{ij}P_{ij}v = \sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{I}} A_{ij}v_{\sigma_i(j)}$$

Now let $V = [v_{\sigma_i(j)}]_{\substack{i\in\mathcal{N}\\ j\in\mathcal{I}}}$. By Cauchy-Shwarz inequality and denoting $\left\|X\right\|_F = \sqrt{\sum_{ij}X_{ij}^2}$ the Frobenius norm of matrix X, we have $\left\|V\right\|_F = \sqrt{n\left\|v\right\|_2^2}$ and $\left\|A\right\|_F \leq B_1(\hat{\mu})(\left\|\mu-\hat{\mu}\right\|_F + \left\|\mu^\top-\hat{\mu}^\top\right\|_F)$, leading to:

$$C_1 \leq \sqrt{n\left\|v\right\|_2^2}\left\|A\right\|_F \leq 2B_1(\hat{\mu})\sqrt{n\left\|v\right\|_2^2}\left\|\mu-\hat{\mu}\right\|_F$$

which proves (A.1).

Similarly, using $B_2(\hat{\mu}) = \max_{i \in [\![n]\!]} \psi'(u_i(\hat{P}), \alpha)$ and the same arguments as above, we obtain:

$$C_2 \leq \sqrt{n \left\|v\right\|_2^2} \left\|A\right\|_F \leq 2B_2(\hat{\mu}) \sqrt{n \left\|v\right\|_2^2} \left\|\mu - \hat{\mu}\right\|_F$$

which yields the desired result. □

A.4 Comparison to utility/inequality trade-offs

In this appendix, we provide the proofs of Section 3.3, and describe more precisely how we applied quality-weighted exposure and equality of exposure in reciprocal recommendation.

A.4.1 One-sided recommendation: quality-weighted exposure

We prove here Proposition 2 of Section 3.3. The result shows that in some cases, compared to any choice of the parameter $\theta \in \Theta$ of the welfare approach, quality-weighted exposure leads to the undesirable behavior of *decreasing user utility* while *increasing inequalities of exposure* between items. Figure A.1 gives an example.

Proposition 2. *The following claims hold irrespective of the choice of* $\boldsymbol{u}^{\mathrm{qua},\beta} \in \mathcal{U}_\beta^{\mathrm{qua}}$.

For every $d \in \mathbb{N}_*$ *and every* $N \in \mathbb{N}_*$, *there is a one-sided recommendation problem, with* $d+1$ *items and* $N(d+1)$ *users, such that* $\forall \theta \in \Theta$, *we have:*

$$\left(\exists \beta > 0, \boldsymbol{u}_{\mathcal{N}}^{\theta} \succ_{\mathrm{L}} \boldsymbol{u}_{\mathcal{N}}^{\mathrm{qua},\beta} \text{ and } \boldsymbol{u}_{\mathcal{I}}^{\theta} \succ_{\mathrm{L}} \boldsymbol{u}_{\mathcal{I}}^{\mathrm{qua},\beta}\right) \qquad \text{and} \qquad \lim_{\beta \to \infty} \frac{\sum_{i \in \mathcal{N}} u_i^{\mathrm{qua},\beta}}{\sum_{i \in \mathcal{N}} u_i^{\theta}} \xrightarrow[d \to \infty]{} \frac{5}{6}.$$

Proof. We prove it for $N = 1$, the more general case is just obtained by repeating the pattern with $d+1$ items and $d+1$ users.

Let $i_1, ..., i_{d+1}$ be the indexes of the users and $j_1, ..., j_{d+1}$ the indexes of the items. The preferences have the following pattern:

$$\forall k \in [\![d+1]\!], \mu_{i_k j_k} = 1 \qquad\qquad \forall k \in [\![d]\!], \mu_{i_k j_{d+1}} = \frac{1}{2}$$

all other μ_{ij} (for user i and item j) are set to 0 (note that we are in a problem with one-sided preferences, which means $\mu_{ji} = 1$ for every item j and user i.

We consider a task with a single recommendation slot ($v_1 = 1, v_2 = \ldots = v_{|\mathcal{I}|} = 0$). On that problem, the optimal ranking for every $\theta \in \Theta$ is to show item j_k to user i_k, which leads to perfect equality in terms of item exposure, and maximizes every user utility. It is thus leximin optimal for both users and items for every $\theta \in \Theta$.

Then, the qualities are equal to:

$$\forall k \in [\![d]\!], q_{j_k} = 1 \qquad\qquad q_{j_{d+1}} = \frac{1}{2}d + 1$$

the target exposure is thus $t_{j_k} = \frac{d+1}{\frac{3}{2}d+1}$ for $k \in [\![d]\!]$ and $t_{j_{d+1}} = (d+1)\frac{\frac{1}{2}d+1}{\frac{3}{2}d+1}$.

Since the problem is symmetric in the users $i_1, ..., i_d$, by the concavity of $F_\beta^{\mathrm{qua}}(\boldsymbol{u}(P))$ with respect to P, there is an optimal ranking described by a single probability p as:

$$\forall k \in [\![d]\!], P_{i_k j_k} = 1 - p \qquad\qquad P_{i_k j_{d+1}} = p \qquad\qquad P_{i_{d+1} j_{d+1}} = 1$$

Note that for such a P, $\forall k \in [\![d]\!]$, $u_{i_k}^{\text{qua},\beta}(P) = 1 - \frac{1}{2}p$, and it is clearr that there is $\beta > 0$ such that $p > 0$, which then implies $\boldsymbol{u}^\theta \succ_{\text{L}} \boldsymbol{u}_{\mathcal{N}}^{\text{qua},\beta}$ and $\boldsymbol{u}^\theta \succ_{\text{L}} \boldsymbol{u}_{\mathcal{I}}^{\text{qua},\beta}$.

Now, as $\beta \to \infty$, p is such that exposure equals its target, which leads to the following equation:

$$dp + 1 = (d+1)\frac{\frac{1}{2}d + 1}{\frac{3}{2}d + 1}.$$

We thus get $p = \frac{d+1}{d}\frac{d+2}{3d+2} - \frac{1}{d} \xrightarrow[d\to\infty]{} \frac{1}{3}$, which gives the result $u_{i_k}^{\text{qua},\beta}(P) = 1 - \frac{1}{2}p \xrightarrow[p\to\frac{1}{3}]{} \frac{5}{6}$.

Notice that similarly to Proposition 3, the result does not depend on the choice of $\boldsymbol{u}^{\text{qua},\beta}$ because the sum of user utilities converges. □

A.4.2 Reciprocal recommendation: equality of exposure

We now prove Proposition 3.

Proposition 3. *For $\beta > 0$, let $\mathcal{U}_\beta^{\text{eq}} = \text{argmax}_{\boldsymbol{u}\in\mathcal{U}} F_\beta(\boldsymbol{u})$. The claim below holds irrespective of the choice of $\boldsymbol{u}^{\text{eq},\beta} \in \mathcal{U}_\beta^{\text{eq}}$. Let $n \geq 5$. There is a reciprocal recommendation task with n users such that:*

$$\forall \theta \in \Theta, \boldsymbol{u}^\theta, \exists \beta > 0 : \quad \forall i \in [\![n]\!], u_i^\theta > u_i^{\text{eq},\beta} \qquad \textit{and} \qquad \lim_{\beta\to\infty} \sum_{i\in\mathcal{N}} u_i^{\text{eq},\beta} = 0.$$

Proof. The example is given in Figure A.1. We still consider a recommendation task with a single recommendation slot.

Let us rename the users by $i_1, i_2, ..., i_5$. The preference patterns are $\mu_{i_1 i_2} = \mu_{i_1 i_3} = 1$ and $\mu_{i_4 i_5} = 1$. Apart from $\mu_{ij} = \mu_{ji}$, other μ_{ij}s are 0. In this proof, we show that $u_{i_1}^{\text{eq},\beta} = 2u_{i_2}^{\text{eq},\beta}$ for every β, which implies that $u_{i_1}^{\text{eq},\beta} \xrightarrow[\beta\to\infty]{} 0$ because 0 utility for every user is feasible. On this task, the leximin ranking also maximizes the sum of users utilities (as shown in Figure A.1), so the optimal ranking is the same for every $\theta \in \Theta$, and every user has a two-sided utility of at least 1.5.

Since $F_\beta(\boldsymbol{u})$ is stricly Schur-concave for $\beta > 0$, i_2 and i_3 always have the same utility in an optimal utility profile (because they play a symmetric role). i_4 and i_5 also have the same utility. Note that the interest of i_4 and i_5 in that problem is to make it possible to recommend them to i_1, which has 0 value.

Similarly to the problem in one-sided recommendation, the only way to decrease the penalty is to reduce the utility of i_1, i_4, i_5. However, reducing the utility of i_1 can only be done by either recommending i_4 or i_5 to i_1, or recommending i_4/i_5 to i_2/i_3. In all cases, decreasing i_1's utility decreases i_2/i_3's utilities.

More precisely, because of the symmetries and the concavity of $F_\beta(\boldsymbol{u}(P))$ with respect to P, for every $\beta > 0$, there is an optimal ranking tensor described by three probabilities p, q, q' such that:[7]

$$\begin{array}{lll}
P_{i_1 i_2} = P_{i_1 i_3} = \frac{1}{2}p & P_{i_2 i_1} = P_{i_3 i_1} = q & P_{i_4 i_5} = P_{i_5 i_4} = q' \\
P_{i_1 i_4} = P_{i_1 i_5} = \frac{1}{2}(1-p) & P_{i_2 i_3} = P_{i_2 i_4} = P_{i_2 i_5} = \frac{1}{3}(1-q) & P_{i_4 i_1} = P_{i_4 i_2} = P_{i_4 i_3} = \frac{1}{3}(1-q') \\
 & P_{i_3 i_2} = P_{i_3 i_4} = P_{i_3 i_5} = \frac{1}{3}(1-q) & P_{i_5 i_1} = P_{i_5 i_2} = P_{i_5 i_3} = \frac{1}{3}(1-q')
\end{array}$$

[7] Since there is a single recommendation slot, we identify P_{ij1} with P_{ij}

In all cases, the two-sided utility are

$$u_{i_1}(P) = \underbrace{p}_{\substack{P_{i_1 i_2}\mu_{i_1 i_2}+P_{i_1 i_3}\mu_{i_1 i_3} \\ \text{user-side utility}}} + \underbrace{2q}_{\substack{P_{i_2 i_1}\mu_{i_2 i_1}+P_{i_3 i_1}\mu_{i_3 i_1} \\ \text{item-side utility}}} \qquad \text{and} \qquad u_{i_2}(P) = q + \frac{1}{2}p$$

Thus, in an optimal ranking for $F_\beta(\boldsymbol{u})$, we must have $u_{i_1}(P) = 2u_{i_2}(P)$. Equality, which is achieved at $\beta \to \infty$ can then only be at 0 utility for every user (since 0 is feasible).

The task used in the proof contains only 5 users. Any number of users can be added to the group $\{i_4, i_5\}$, with a "complete" preference profile ($\mu_{ij} = 1$ for all pair i, j in that group). □

The Lorenz efficiency of our welfare approach guarantees that it cannot exhibit the undesirable behaviors of equality or quality-weighted exposure penalties described in Propositions 2 and 25.

A.4.3 Equality of exposure and quality-weighted exposure in reciprocal recommendation

In one-sided recommendation with one-sided preferences, equality of exposure is the same as equality of utility. More generally, let $e_j(P) = \sum_{i \in \mathcal{N}} P_{ij} v$ the total exposure of item j. Equality of exposure is defined by:

$$F_\beta^{\text{expo}}(P) = \sum_{i \in \mathcal{N}} \overline{u}_i(P) - \beta \sqrt{\sum_{j \in \mathcal{I}} \left(e_j(P) - \frac{|\mathcal{N}|}{|\mathcal{I}|} \|v\|_1 \right)^2}$$

In one-sided recommendation, parity of exposure is relatively well behaved because the exposure target $\frac{|\mathcal{N}|}{|\mathcal{I}|}\|v\|_1$ is constant. Driving towards equality can thus not lead to a decrease of the total exposure budget, which was the problem with equality of utility in settings with two-sided preferences (driving towards equality of utility leads to a decrease of total utility), as we described in Section 3.3.

The formula allows us to extend parity of exposure in the next section and in our experiments, since it is also valid in reciprocal recommendation. Likewise, the formula of quality-weighted exposure that is also valid in reciprocal recommendation is given by:

$$F_\beta^{\text{qua}}(P) = \sum_{i \in \mathcal{N}} \overline{u}_i(P) - \beta \sqrt{\sum_{j \in \mathcal{I}} \left(e_j(P) - \frac{q_j E}{Q} \right)^2}.$$

The result below shows that equality of exposure and quality-weighted exposure lead to inefficiencies in reciprocal recommendation settings:

Proposition 25. *For every $n \in \mathbb{N}_*$, there is a reciprocal recommendation task with n users such that:*

$$\forall \theta \in \Theta, \exists \beta > 0 : \qquad \boldsymbol{u}^\theta \succ_{\text{L}} \boldsymbol{u}^{\text{expo},\beta} \qquad \textit{and} \qquad \boldsymbol{u}^\theta \succ_{\text{L}} \boldsymbol{u}^{\text{qua},\beta}.$$

Moreover, $\lim_{\beta \to \infty} \sum_{i \in \mathcal{N}} u_i^{\text{expo},\beta} = \frac{2}{n} \sum_{i \in \mathcal{N}} u_i^\theta$ *and* $\lim_{\beta \to \infty} \sum_{i \in \mathcal{N}} u_i^{\text{qua},\beta} = \frac{2+n}{2n} \sum_{i \in \mathcal{N}} u_i^{\text{sum}}$.

Proof. An example of extreme case is with n users when there is a "leader" who is the only possible match with other users. We consider a single recommendation slot. The preferences are:

$$\forall j \in \{2, \dots, n\}, \mu_{1j} = \mu_{j1} = 1 \qquad \forall (i,j) \in \{2, \dots, n\}^2, \mu_{ij} = 0.$$

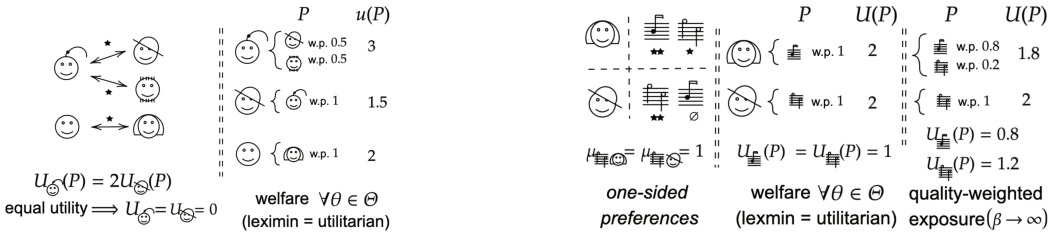


Figure A.1: **Left:** Example of a reciprocal recommendation task where equality of utility leads to 0 utility (see the proof of Prop. 3 in App. A.4). There is one recommendation slot per user. We give the recommendation probabilities and utilities for the utilitarian ranking and three users, the other ones are obtained by the symmetry of the problem. The utilitarian ranking is also leximin optimal, so our approach yields the same recommendations for all θ. **Right:** Example where quality-weighted exposure reduces user utility while increasing inequalities between items.

On this task, the for every $\theta \in \Theta$, the optimal ranking is given by:

$$\forall j \in \{2, \dots, n\}, P_{1j} = \frac{1}{n-1} \qquad \forall i \in \{2, \dots, n\} P_{i1} = 1.$$

The reason it is the only possible optimal ranking is because it is leximin optimal and has the maximum achievable sum of utilities. The utilities are then $u_1(P) = n$ and $u_i(P) = 1 + \frac{1}{n-1}$, which leads to $\sum_{i=1}^{n} u_i = 2n$.

Equality of exposure Driving towards equality of exposure requires to reduce the exposure of user 1, which in turn reduces the utility of user 1 and the utilities of those who user 1 is less exposed to. Thus, there is $\beta > 0$ such that $\boldsymbol{u}^{\theta} \succ_{\mathrm{L}} \boldsymbol{u}^{\mathrm{expo},\beta}$ because of the loss of efficiency. Finally, by the concavity of the objective with respect to P, and by the symmetry of the problem with respect to $i_2, \dots, i_n$, we can conclude that an optimal way to achieve perfect equality of exposure is to recommend, to every user i, every user $j \neq i$ with probability $\frac{1}{n-1}$. The utility is then $u_1(P) = 1 + (n-1)\frac{1}{n-1}$ and $u_i(P) = \frac{2}{n-1}$ for $i \geq 2$, leading to $\sum_{i=1}^{n} u_i = 4$, which gives the result.

Quality-weighted exposure On the same example, the qualities are $q_1 = n - 1$ and $q_i = 1$ for $1 \geq 2$. The total exposure targets are then $t_1 = \frac{1}{2}n$ and $t_i = \frac{n}{2(n-1)}$. These exposure targets mean less exposure for 1 than in the leximin ranking. Thus β sufficiently large has the effect of reducing 1's exposure[8], which reduces the utility of 1 and the users to whom 1 is less recommended. Thus $\boldsymbol{u}^{\theta} \succ_{\mathrm{L}} \boldsymbol{u}^{\mathrm{qua},\beta}$. By the symmetry of the problem, as $\beta \to \infty$, quality weighted exposure is achieved by setting:

$$\forall j \in \{2, \dots, n\} : P_{1j} = \frac{1}{n-1} \qquad P_{j1} = \frac{n}{2(n-1)}$$
$$\forall j' \in \{2, \dots, n\}, j' \neq j, P_{jj'} = \frac{1 - \frac{n}{2(n-1)}}{n-2} = \frac{1}{2(n-1)}$$

The utilities are then $u_1(P) = 1 + (n-1)\frac{n}{2(n-1)} = 1 + \frac{n}{2}$ and $u_i(P) = \frac{n}{2(n-1)} + \frac{1}{n-1} = \frac{n+2}{2(n-1)}$. The total utility is thus $2 + n$, which gives the result. □

The Lorenz efficiency of our welfare approach guarantees that it cannot exhibit the undesirable behaviors of parity or quality-weighted exposure penalties described in Propositions 2 and 25.

[8] Direct calculations of the derivatives show that when $\beta > 0$ is too small the penalty has no effect.

A.5 A generic Frank-Wolfe algorithm for ranking

In this section, we present a general form of our algorithm presented in Section 3.4, as well as the proofs of the claims.

Let $F : \mathbb{R}^n \to \mathbb{R}$, concave, and we want to find

$$P^* \in \operatorname*{argmax}_{P \in \mathcal{P}} F(\boldsymbol{u}(P)). \tag{A.2}$$

Let $\langle X \,|\, Y \rangle = \sum_{ijk} X_{ijk} Y_{ijk}$ be the dot product between three-way tensors, and let $\nabla(F \circ \boldsymbol{u})(P)$ be the gradient of $P \mapsto F(\boldsymbol{u}(P))$ taken at P, i.e., $(\nabla(F \circ \boldsymbol{u}))_{ijk} = \frac{\partial F \circ \boldsymbol{u}}{\partial P_{ijk}}$

Starting from $P^{(0)} \in \mathcal{P}$ (in our experiments we always use a utilitarian ranking $P^{(0)} \in \operatorname{argmax}_{P \in \mathcal{P}} \sum_{i=1}^n u_i(P)$), the Frank-Wolfe algorithm alternates two steps for $t \geq 1$:

1. let $\tilde{P} \in \operatorname{argmax}_{P \in \mathcal{P}} \langle P \,|\, \nabla(F \circ \boldsymbol{u})(P^{t-1}) \rangle$
2. $P^{(t)} = (1 - \gamma^{(t)}) P^{(t-1)} + \gamma^{(t)} \tilde{P}$ with $\gamma^{(t)} = \frac{2}{t+2}$

The stepsize $\frac{2}{t+2}$ is from Clarkson [2010, Section 3], which avoids a line search and in our experiments seemed to yield acceptable results. Irrespective of the step size, the fundamental results which allows to use Frank-Wolfe in the setting of (A.2) are the two following lemmas:

Lemma 26. *Recall that $u_i(P) = \sum_{i=1}^n \mu_{ij}(P_{ij} + P_{ji})v$. Let $\frac{\partial F}{\partial u_i}$ denote the derivative of F with respect to its i-th argument and $\frac{\partial F}{\partial u_i}(\boldsymbol{u}(P))$ the value of this derivative at $\boldsymbol{u}(P)$.*

Then, $\forall i \in \mathcal{N}, \forall j \in \mathcal{I}, \forall k \in [\![|\mathcal{I}|]\!]$, we have:

$$\frac{\partial F \circ \boldsymbol{u}}{\partial P_{ijk}}(P) = \Big(\mu_{ij} \frac{\partial F}{\partial u_i}\big(\boldsymbol{u}(P)\big) + \mu_{ji} \frac{\partial F}{\partial u_j}\big(\boldsymbol{u}(P)\big)\Big) v_k.$$

Proof. The result is a consequence of the chain rule:

$$\frac{\partial F \circ \boldsymbol{u}}{\partial P_{ijk}}(P) = \sum_{p=1}^n \frac{\partial F}{\partial u_p}(\boldsymbol{u}(P)) \frac{\partial u_p(P)}{\partial P_{ijk}}$$

With

$$u_p(P) = \sum_{q=1}^n \mu_{pq} \sum_{r=1}^{|\mathcal{I}|} (P_{pqr} + P_{qpr}) v_k.$$

Thus $\frac{\partial u_p(P)}{\partial P_{ijk}} = (\mu_{ij} \mathbb{1}_{\{p=i\}} + \mu_{ji} \mathbb{1}_{\{p=j\}}) v_k$, which gives the desired result. □

Lemma 27. *Let A be an $n \times n$ matrix with $A_{ij} \in \mathbb{R}$ (not necessarily non-negative). Let $v \in \mathbb{R}^{|\mathcal{I}|}$ with non-negative and non-increasing entries, i.e., $\forall k \in [\![|\mathcal{I}| - 1]\!]$, $v_k \geq v_{k+1} \geq 0$. Let K be the last index such that $v_K > 0$ (or $K = |\mathcal{I}|$ if there is no such index).*

Let $P \in \mathcal{P}$ such that:

$$\forall i, \forall \sigma_i \in \mathfrak{S}(P_i), \forall (j, j') \in \mathcal{I}^2 : \quad \Big(\sigma_i(j) \leq K \text{ and } \sigma_i(j) < \sigma_i(j') \implies A_{ij} \geq A_{ij'}\Big).$$

And let X be the $n \times n \times |\mathcal{I}|$ tensor defined as $X_{ijk} = A_{ij} v_k$.

Then $P \in \operatorname{argmax}_{P \in \mathcal{P}} \langle P \,|\, X \rangle$.

Moreover, if $\forall k \in [\![|\mathcal{I}| - 1]\!]$, $v_k > v_{k+1} \geq 0$, then for every $P \in \operatorname{argmax}_{P \in \mathcal{P}} \langle P \,|\, X \rangle$, we have:

$$\forall i, \forall \sigma_i \in \mathfrak{S}(P_i), \forall (j, j') \in \mathcal{I}^2 : \quad \Big(\sigma_i(j) < \sigma_i(j') \implies A_{ij} \geq A_{ij'}\Big).$$

Proof. The result stems from the rearrangement inequality (also known as the Hardy-Littlewood inequality [Hardy et al., 1952]), which states that for two vectors $a \in \mathbb{R}^n_+$, and $b \in \mathbb{R}^n$, $\operatorname{argmax}_\nu \sum_{j=1}^n a_{\nu(j)} b_j$,

where ν spans the permutations of $[\![n]\!]$, is the set of permutations such that b is ordered similarly to $(a_{\nu(i)})_{i=1}^n$. If the a_ks are non-increasing, then every permutation that sorts b in decreasing order is in the argmax. We need the reciprocal statement for the second part of our Lemma: if the a_is are strictly decreasing, then only the permutations that sort b in decreasing order are in $\operatorname{argmax}_\nu \sum_{j=1}^n a_{\nu(j)} b_j$. Note that these arguments are well-known in learning to rank [see, e.g., Cossock and Zhang, 2008].

In our case, notice that

$$\langle P \,|\, X \rangle = \sum_{i \in \mathcal{N}} \Big(\sum_{j \in \mathcal{I}} A_{ij} P_{ijk} v_k \Big)$$

The maximization over P can then be performed over each user i (and thus each bistochastic matrix P_i separately). Now, if P_i is such that every $\sigma_i \in \mathfrak{S}(P_i)$ orders A_{ij} in decreasing order, then by the rearrangement inequality $\sigma_i \in \operatorname{argmax}_\nu \sum_{j \in \mathcal{I}} A_{ij} v_{\nu(j)}$. Notice that if only the K first elements of v are non-zero, we only need a top-K ranking. This gives us the first part of the thoerem.

The second part of the theorem follows from the reciprocal of the rearrangement inequality, since for P_i to be an optimal stochastic ranking for $\sum_{j \in \mathcal{I}} A_{ij} P_{ijk} v_k$, every permutation σ_i in its support must be in $\operatorname{argmax}_\nu \sum_{j \in \mathcal{I}} A_{ij} v_{\nu(j)}$. □

A.5.1 Proof of Theorem 4

Lemma 26 and 27 together are sufficient to give algorithms for the inference of stochastic rankings using our welfare function (3.1) and using the penalties of Section 3.3, by computing the partial derivatives $\frac{\partial F}{\partial u_i}$. The main result of Section 3.4, which we prove now, instantiates this principle for the welfare function approach:

Theorem 4. *Let* $\tilde{\mu}_{ij} = \Phi_i'\big(u_i(P^{(t)})\big)\mu_{ij} + \Phi_j'\big(u_j(P^{(t)})\big)\mu_{ji}$. *Let* $\tilde{P}$ *such that:*
$\forall i \in \mathcal{N}, \forall \tilde{\sigma}_i \in \mathfrak{S}(\tilde{P}_i)$: $\tilde{\sigma}_i(j) < \tilde{\sigma}_i(j') \implies \tilde{\mu}_{ij} \geq \tilde{\mu}_{ij'}$. *Then* $\tilde{P} \in \operatorname*{argmax}_{P \in \mathcal{P}} \langle P \,|\, \nabla W(P^{(t)}) \rangle$.

Proof. Notice that with $W(P) = F(\boldsymbol{u}(P)) = \sum_{i=1}^n \Phi_i(u_i(P))$, then $\frac{\partial F}{\partial u_i}(\boldsymbol{u}(P)) = \Phi_i'(u_i(P))$. By Lemma 26, we have that $\langle P \,|\, \nabla F(P^{(t)}) \rangle$ is of the form $\langle P \,|\, X \rangle$ with $X_{ijk} = A_{ij} v_k$ with $A_{ij} = \tilde{\mu}_{ij}$, so the result is implied by Lemma 27. □

A.5.2 Proof of Proposition 5

Proposition 5. *Let* $B = \max_{i \in [\![n]\!]} \|\Phi_i''\|_\infty$ *and* $U = \max_{\boldsymbol{u} \in \mathcal{U}} \|\boldsymbol{u}\|_2^2$. *Let* K *be the maximum index of a nonzero value in* v *(or* $|\mathcal{I}|$*). Then* $\forall t \geq 1, W(P^{(t)}) \geq \max_{P \in \mathcal{P}} W(P) - O(\frac{BU}{t})$. *Moreover, for each user, an iteration costs* $O(|\mathcal{I}| \ln K)$ *operations and requires* $O(K)$ *additional bytes of storage.*

Proof. Note that $\mathcal{P}$ is a simplex over ranking tensors containing one deterministic ranking for each user. Using [Clarkson, 2010, Section 3], the Frank-Wolfe algorithm with our step-size converges in $O\big(\frac{C_W}{t}\big)$, where, using [Clarkson, 2010, Equation 11] and denoting by $\nabla^2 W$ the Hessian of W, we have

$$C_W \leq \sup_{\substack{\boldsymbol{u}, \boldsymbol{u}' \in \mathcal{U} \\ \tilde{\mathbf{u}} \in \mathcal{U}}} -\frac{1}{2} (\boldsymbol{u} - \boldsymbol{u}')^\top \nabla^2 W(\tilde{\mathbf{u}}) (\boldsymbol{u} - \boldsymbol{u}') \leq \frac{B}{2} \sup_{\boldsymbol{u}, \boldsymbol{u}' \in \mathcal{U}} \|\boldsymbol{u} - \boldsymbol{u}'\|_2^2 \leq 2BU.$$

where we used $\|\boldsymbol{u} - \boldsymbol{u}'\|_2^2 \leq 2\|\boldsymbol{u}\|_2^2 + 2\|\boldsymbol{u}'\|_2^2$.

For the computation cost, we use Lemma 27, which is more precise than Theorem 4, to see that finding the argmax only requires a top-K ranking. While technically any $P \in \mathcal{P}$ should contain a whole bistochastic matrix, it is not necessary to store a completion of the top-K rankings because

they have no impact on the utility. As such, storing each $\tilde{P}$ only costs $O(K)$ bytes per user, which contain the indices of the top-K items in the ranking found by Theorem 2.

Computing the two-sided utilities costs $O(|\mathcal{N}||\mathcal{I}|)$, and thus $O(|\mathcal{I}|)$ per user. Moreover, computing the top-K ranking costs $O(|\mathcal{I}| \ln K)$ in the worst case, with a streaming method that maintains a min-heap of the top-K elements seen so far, and finish with sorting the top-K elements. □

Notice that for faster average performance, the top-K sort can be performed using a fast selection algorithm (such as quickselect), to obtain the top-K elements with $O(|\mathcal{I}|)$ expected time complexity, and then sorting, yielding $O(|\mathcal{I}| + K \ln K)$ expected time complexity per user at each iteration.

A.6 Additional experimental results

Our experiments are fully implemented in Python 3.9 using PyTorch[9]. We provide the code as supplementary material. We compare our welfare maximization approach with the fairness penalties presented in Section 3.3.

We also compare ourselves to the algorithm FairRec from Patro et al. [2020] (referred to as *Patro et al.* in the figures and description), who consider envy-freeness as user-side fairness criterion, and max-min share of exposure as item-side fairness criterion. Envy-freeness states that every user should prefer their recommendation list to that of any other user. The max-min exposure criterion on the item side means that each user should receive an exposure of at least $\beta\frac{E}{|\mathcal{I}|}$, where β is a parameter allowing to control how much exposure is guaranteed to items. We vary this parameter in our experiments to show the trade-offs achieved by *Patro et al.*. Since *Patro et al.* does not produce rankings, we took the recommendation list with the given order as a ranked list.

A.6.1 One-sided recommendation: Lastfm-2k dataset

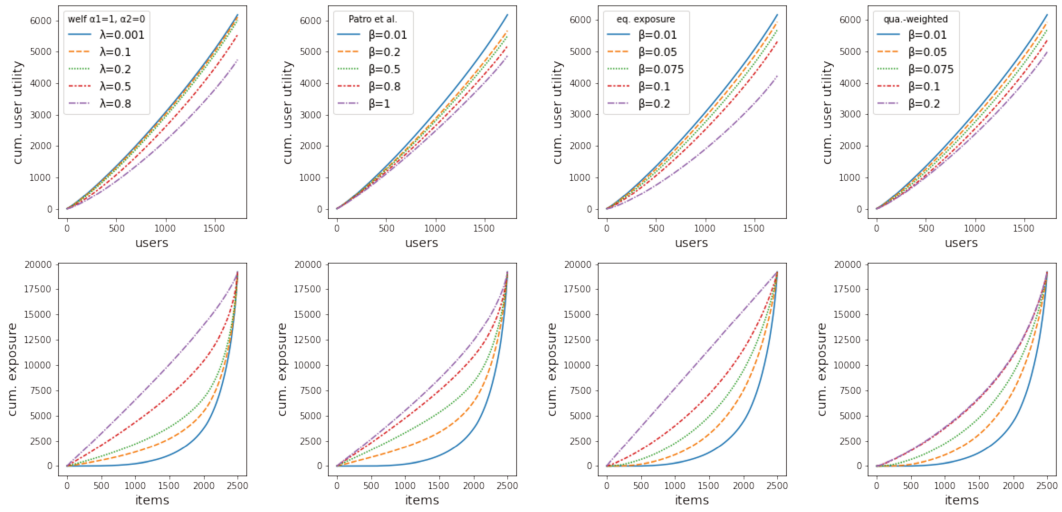
We describe in this section the details of the experiments presented in Section 3.5.1. We use a dataset from the online music service Last.fm[10]. In the main paper, we presented results on **Lastfm-2k** from Cantador et al. [2011] which contains real play counts of $2k$ users for $19k$ artists, and was used by Patro et al. [2020] who also study two-sided fairness in recommendation. We filter the top $2,500$ items most listened to. Following Johnson [2014], we pre-process the raw counts with log-transformation. We split the dataset into train/validation/test sets, each including 70%/10%/20% of the user-item play counts. We create three different splits using three random seeds. One-sided preferences are estimated using the standard matrix factorization algorithm[11] of Hu et al. [2008] trained on the train set, with hyperparameters selected on the validation set by grid search. The number of latent factors is chosen in $[16, 32, 64, 128]$, the regularization in $[0.1, 1., 10., 20., 50.]$, and the confidence weighting parameter in $[0.1, 1., 10., 100.]$. The estimated preferences we use are the positive part of the resulting estimates.

Rankings are inferred from these estimated preferences. The exposure weights we use in the computation of utilities are the standard weights of the *discounted cumulative gain* (DCG) (also used in e.g., Singh and Joachims [2018], Biega et al. [2018], Morik et al. [2020]): $\forall k \in [\![|\mathcal{I}|]\!], v_k = \frac{1}{\log_2(1+k)}$. For each ranking approach, the Frank-Wolfe algorithm is run with 5000 iterations to make sure we are close to convergence, and the number of recommendation slots is set to 40.

We evaluate rankings on estimated preferences, considered as ground truth, following many works on fair recommendation [Singh and Joachims, 2018, Patro et al., 2020, Wang and Joachims, 2020, Wu et al., 2021b]. This is because the goal is to evaluate the fairness of ranking algorithms themselves, rather than biases in preference estimates. All results are averaged over three random seeds. To obtain various trade-offs, for *welf* we vary λ in $[0.001, 0.01, 0.05, 0.075, 0.1, 0.125, 0.15, 0.2, 0.3, 0.325, 0.35]$ and $[0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.9, 0.95, 0.99, 0.999]$. For *Patro et al.* we vary β in $[0.01, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4]$ and $[0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75, 0.8, 0.85, 0.9, 0.95, 1]$, and for other methods we vary β in $[0.001, 0.005, 0.01, 0.015, 0.0175, 0.02, 0.025, 0.03, 0.035, 0.04, 0.045, 0.05, 0.055, 0.06]$ and $[0.065, 0.07, 0.075, 0.08, 0.085, 0.09, 0.095, 0.1, 0.105, 0.11, 0.2, 0.5, 1, 2, 5, 10, 20, 30, 40, 50, 70, 100]$.

Item-side fairness Figure A.2 presents the various trade-offs achieved by each method in one-sided recommendation, as discussed in Section 3.5.1. We observe that only *qua.-weighted* is unable

[9] http://pytorch.org
[10] https://www.last.fm/
[11] Using the Python library Implicit: https://github.com/benfred/implicit (MIT License).



(a) Welfare (ours) (b) *Patro et al.* (c) *eq. exposure* (d) *qua.-weighted*

Figure A.2: representative trade-offs achieved by the various compared methods on Lastfm-2k. The trade-offs achieved by the different methods look alike, except that *qua.-weighted* does not aim at reaching equality of exposure for exteme values of β. See Section 3.5.1 for the discussions on the differences between the trade-offs achieved by the different approaches.

to reach equal exposure because of its quality-weighted exposure target: perfectly equal exposure is only permitted when all items have the same quality.

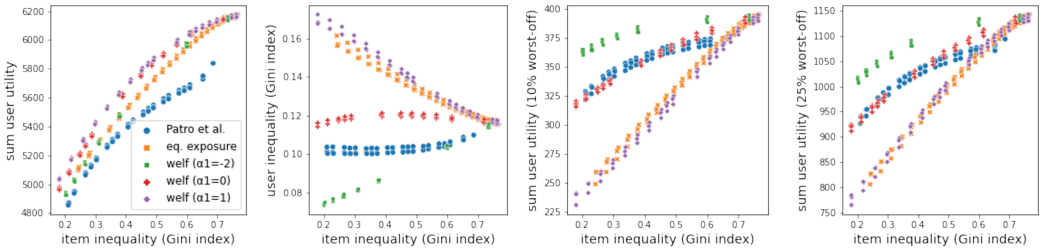
Two-sided fairness Figure A.3 shows the effect of varying α_1 and λ on user fairness as in Figure 3.3 of the main paper, but with results repeated over three random seeds. We observe the same trade-offs and conclude again that *welf* is better than *Patro et al.* and *eq. exposure*, in terms of its impact on worse-off users.

The importance of considering the whole Lorenz curve In Fig. A.4 we show the results of the same models as before, but changing the way we measure the item inequality: using the standard deviation of exposure rather than the Gini index. Now, *eq. exposure* dominates the total utility/item inequality plot, since the plot corresponds exactly to the objective function of the algorithm. Comparing *eq. exposure* with *welf* $\alpha_1 = 1$, we now see that the trade-offs are different, with *eq. exposure* performing better on the worse-off users. Comparing *welf* $\alpha_1 = 0$ and *Patro et al.*, we see that they still exhibit similar behaviors, with *welf* $\alpha_1 = 0$ being better for better off users. Finally, *welf* $\alpha_1 = -2$ still dominates the othe methods in terms of performance on the worse-off users.

A.6.2 One-sided recommendation: Lastfm-15k dataset

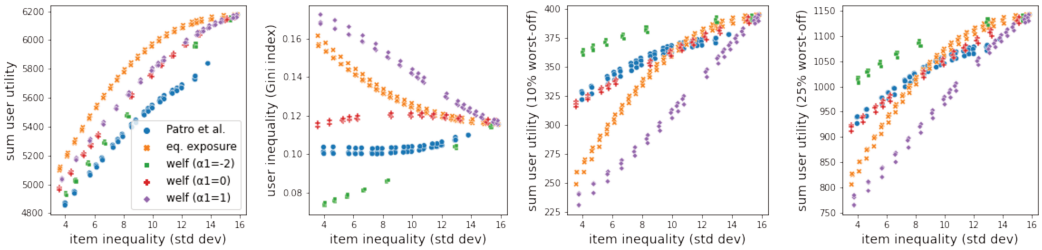
We replicate the experiments on a larger dataset to verify our conclusions at a larger scale. We consider another Lastfm dataset from Celma [2010], which includes $360k$ users and $180k$ items (artists). We select the top $15,000$ users and items having the most interactions, so we refer to this dataset as Lastfm-15k. We apply exactly the same experimental protocol as for Lastfm-2k, with the same range of hyperparameters for the different methods.

Results Fig. A.7 and A.6 show the results obtained by *welf*, *Patro et al.* and *eq. exposure*. The conclusions are similar to those on Lastfm-2k, with the results of *welf* $\alpha = 0$ being more uniformly



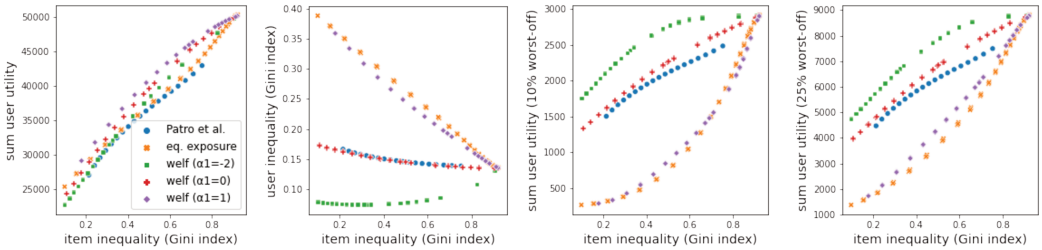
(a) Total user utility. (b) User inequality. (c) Cumulative utility for 10% worse-off users (d) Cumulative utility for 25% worse-off users

Figure A.3: Focus on user fairness on Lastfm-2k: effect of varying α_1 (user-side curvature of the welfare function) keeping $\alpha_2 = 0$. The figure shows all the results obtained with a repetition of three seeds. Overlapping points correspond to the same model parameter across different seeds. We can see that the variance is negligible compared to the observed differences.



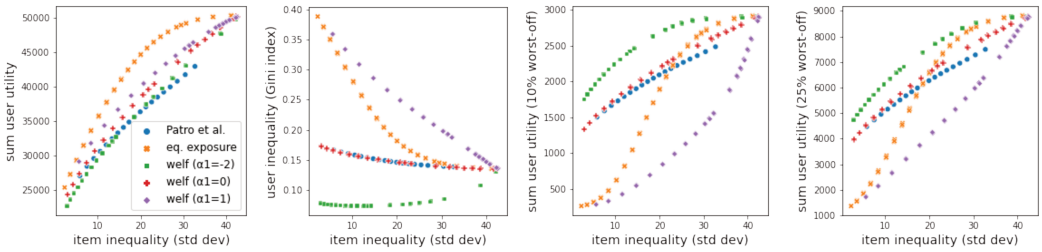
(a) Total user utility. (b) User inequality. (c) Cumulative utility for 10% worse-off users (d) Cumulative utility for 25% worse-off users

Figure A.4: Focus on user fairness on Lastfm-2k, measuring item inequality with standard deviation rather than Gini index. We observe a similar relative behavior between *welf* and *Patro et al.*, but now equality of exposure is optimal on the total utility/item inequality trade-off since it corresponds exactly to the objective of the algorithm. Nonetheless, *welf* $\alpha_1 = -2$ still obtains higher performance on 10%-25% worse-off users, showing that *welf* offers a larger range of trade-offs than *eq. exposure*.



(a) Total user utility. (b) User inequality. (c) Cumulative utility for 10% worse-off users (d) Cumulative utility for 25% worse-off users

Figure A.5: Results on Lastfm-15k when measuring the inequality between items with the Gini coefficient.



(a) Total user utility. (b) User inequality. (c) Cumulative utility for 10% worse-off users (d) Cumulative utility for 25% worse-off users

Figure A.6: Results on Lastfm-15k when measuring inequalities between items with the standard deviation.

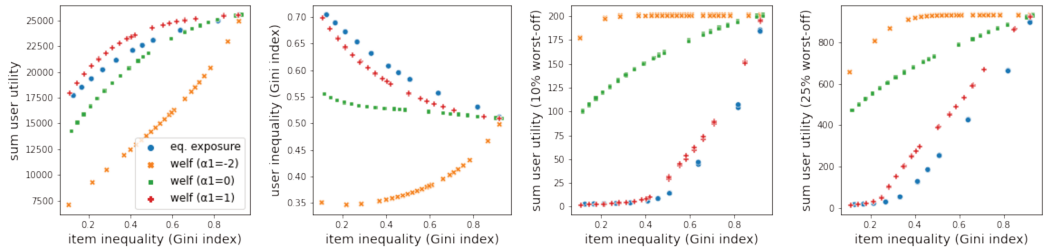
better than those of *Patro et al.*, even though overall similar. *welf* $\alpha_1 = -2$ dominates in terms of user utility on worse-off users. *welf* and *eq. exposure* still find different trade-offs, with *welf* dominating *eq. exposure* when inequality between items is measured by the Gini index, and *eq. exposure* dominating *welf* when inequality is measured by the standard deviation.

A.6.3 One-sided recommendation: Movielens dataset

We provide additional results on the **MovieLens**-20m dataset Harper and Konstan [2015], which contains ratings on a 5-star scale of movies by real users. To simulate a collaborative filtering task with implicit feedback similar to Last.fm, we consider missing ratings as negative feedback and the task is to predict positive values. Since ratings < 3 are usually considered as negative Lim et al. [2015], Wang et al. [2018], we set ratings < 3 to zero, resulting in a dataset with preference values among $\{0, 3, 3.5, 4, 4.5, 5\}$. As for Lastfm-15k, we select the top $15,000$ users and items with the most interactions. For the inference and evaluation of rankings, we follow the same protocols as for Last.fm.

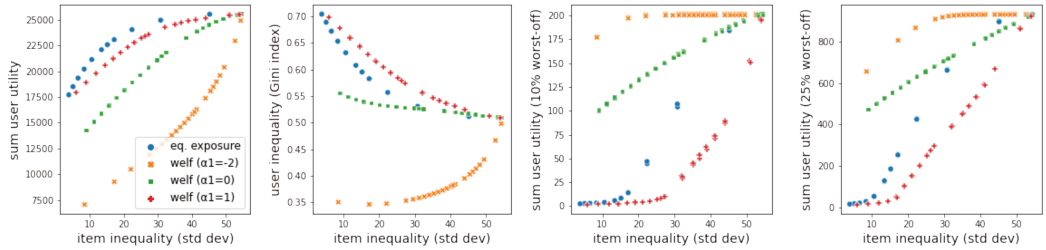
The experimental protocol is the same as for Lastfm-2k and Lastfm-15k except that we do not run the algorithm by [Patro et al., 2020] because its runtime was prohibitive.

results The results are qualitatviely similar to those on Lastfm-2k and Lastfm-15k except that the trends are magnified. *welf* $\alpha = 1$ and *eq. exposure* seem more similar, with *welf* $\alpha = 0$ dominating the trade-off total utility/item iniequality when item inequality is measured with the Gini index, and *eq. exposure* dominating this trade-off when item inequality is measured with standard deviation. *welf* $\alpha = -2$ has great performance on worse-off users compared to *eq. exposure* or *welf* with larger α, but also comes at a significant cost in terms of total user utility, which is very rapidly driven



(a) Total user utility. (b) User inequality. (c) Cumulative utility for 10% worse-off users (d) Cumulative utility for 25% worse-off users

Figure A.7: Results on Movielens when measuring the inequality between items with the Gini coefficient.



(a) Total user utility. (b) User inequality. (c) Cumulative utility for 10% worse-off users (d) Cumulative utility for 25% worse-off users

Figure A.8: Results on Movielens when measuring inequalities between items with the standard deviation.

down.

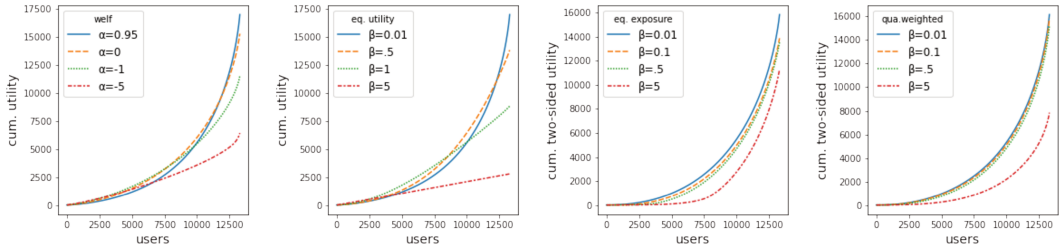
A.6.4 Reciprocal recommendation: Twitter-13k dataset

We now provide the full details of the experiments on Twitter presented in Section 3.5.2 of the main body. Given the lack of common benchmark for reciprocal recommendation [Palomares et al., 2021], we generate a reciprocal recommendation task for people-to-people recommendation problems based on the social network Twitter. We use the Higgs Twitter-13k dataset which includes (directed) follower relationships between users.[12] We keep users having at least 20 mutual follows, resulting in a subset of 13k users. We use the directed links to estimate the probability ϕ_{ij} that i follows j, and the (symmetric) probability of a mutual follow, which is $\mu_{ij} = \phi_{ij} \times \phi_{ji}$. As in the experiments for one-sided recommendation, we split the dataset into train/validation/test sets, each including 70%/10%/20% of the *directed* follower links. We create three random uniform splits, corresponding to three different seeds.

Estimates $\hat{\phi}_{ij}$ are built with logistic matrix factorization[13] [Johnson, 2014] trained on the train set with hyperparameter selection on the validation set. The number of latent factors is chosen in $[16, 32, 64, 128]$, the regularization in $[0.1, 1., 10., 20., 50.]$, and the confidence weighting parameter in $[0.1, 1., 10., 100.]$. Rankings are inferred from all estimated mutual preferences $\hat{\mu}_{ij} = \max(\hat{\phi}_{ij}\hat{\phi}_{ji}, 0)$. For each ranking method, the Frank-Wolfe algorithm is run with 5000 iterations, and the number of recommendation slots is set to 40. As for one-sided recommendation, rankings are estimated on estimated mutual preferences taken as ground truth.

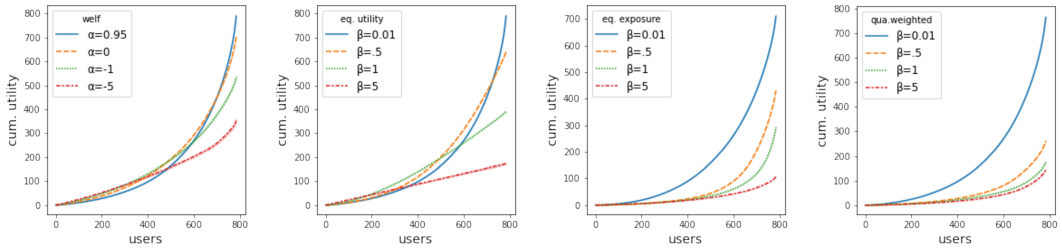
We generate different trade-offs with *welf* by varying α in $[0.99, 0.9, 0.75, 0.5, 0.25, 0, -0.25, -0.5, -0.6, -0.7, -0.8, -0.9$

[12] It was collected following the discovery of the Higgs boson in July, 2012.
[13] Using the Python library Implicit (MIT License).



(a) Welfare (ours) (b) *eq. utility* (c) *eq. exposure* (d) *qua.-weighted*

Figure A.9: representative trade-offs achieved by the various compared methods on Twitter-13k. Exposure-based approaches (*qua.-weighted* and *eq. exposure*) do not yield interesting trade-offs as they are unable to increase the utility of worse-off users. The trade-offs achieved by the *welf* and *eq. utility* are different. Equal utility rapidly generates near-flat curves without really focusing on the very first users, while *welf* increases the utility of the worst-off users while keeping the total utility relatively high.



(a) Welfare (ours) (b) *eq. utility* (c) *eq. exposure* (d) *qua.-weighted*

Figure A.10: representative trade-offs achieved by the various compared methods on Epinions. The results are qualitatively similar to those on Twitter-13k.

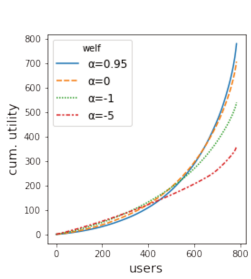
$[-1.1, -1.25, -1.5, -1.75, -2.0, -2.5, -3, -5, -10, -15, -16, -17, -18]$. For all other methods, we vary β in $[0.01, 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9, 1, 1.1, 1.25, 1.5, 2, 5, 10, 50, 100]$.

All presented results are obtained by averaging performance over the three seeds.

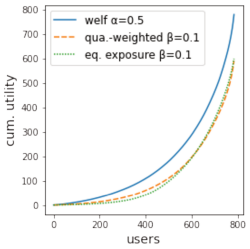
Results Figure A.9 presents the trade-offs achieved by the different methods on Twitter-13k. As expected, *qua.-weighted* and *eq. exposure* do not exhibit a good behavior: stronger penalties lead to more dominated curves where the utility of every user is decreased. This is because constraining item exposure is not meaningful in reciprocal recommendation, where the relevant utility is the two-sided utility. The trade-offs achieved by the *welf* and *eq. utility* are different. Equal utility rapidly generates near-flat curves without really focusing on the very first users, while *welf* increases the utility of the worst-off users while keeping the total utility relatively high.

A.6.5 Reciprocal recommendation: Epinions dataset

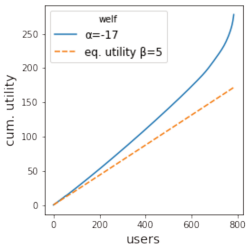
We present additional experiments on reciprocal recommendation with the Epinions dataset Richardson et al. [2003]. Epinions.com is a consumer review site with a who-trust-whom network, and the dataset gathers (directed) trust relationships between members of the platform. Here, we consider the task of finding mutual trust links. We keep users having at least 20 mutual trust links, resulting in a subset of 800 entities. For the inference and evaluation of rankings, we use the same protocols as for the Twitter experiments described in the previous subsection. The experimental parameters are the same as for the Twitter-13k dataset.



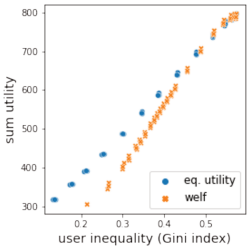
(a) Example trade-offs achieved by *welf*.



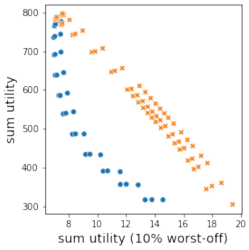
(b) *welf* dominates exposure-based methods.



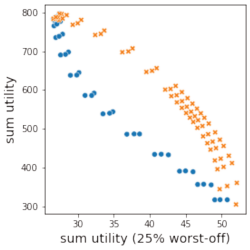
(c) *welf* dominates eq. utility near strict equality.



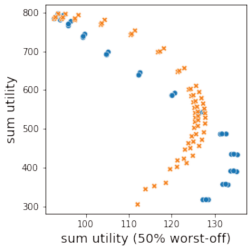
(d) Total utility vs inequality.



(e) Total utility vs utility of 10% worse-off users.



(f) Total utility vs utility of 25% worse-off users.



(g) Total utility vs utility of 50% worse-off users.

Figure A.11: Results on the epinions dataset.

Results Figure A.10 presents the trade-offs achieved by the different methods on Epinions. As expected, *qua.-weighted* and *eq. exposure* do not exhibit a good behavior: stronger penalties lead to more dominated curves where the utility of every user is decreased. In Figure A.11 plots the equivalent of Fig. 3.4. The results are similar: all of *qua.-weighted*, *eq. exposure* and *eq. utility* have dominated curves. We also observe that in the more interesting region where we are closer to the maximum achievable utility, *welf* optimizes better the utility of worse-off users. Yet, in that region, there is no strict dominance of *welf* over *eq. utility*.

A.7 Pairwise vs pointwise penalties

Our penalty-based approach uses the penalty $\sqrt{D(\boldsymbol{u})}$ with:

$$D(\boldsymbol{u}) = \sum_{j \in \mathcal{I}} \left(u_j - \frac{1}{|\mathcal{I}|} \sum_{j' \in \mathcal{I}} u_{j'} \right)^2.$$

Some authors use $D'(\boldsymbol{u}) = \sum_{(j,j') \in \mathcal{I}^2} |u_j - u'_j|$ instead of $\sqrt{D(\boldsymbol{u})}$ [Singh and Joachims, 2019, Morik et al., 2020, Basu et al., 2020], but it is less computationally efficient than our penalty because it involves a quadratic number of terms.

The penalties are similar in that they are related to well-known measures of inequalities:

- $\frac{D'(\boldsymbol{u})}{2|\mathcal{I}| \sum_{j \in \mathcal{I}} u_j}$ is the Gini index of $\boldsymbol{u}_\mathcal{I}$ [Gini, 1921], which, up to an affine transform is the area under the Lorenz curve.
- $D(\boldsymbol{u})$, which is (up to a constant) the variance of $\boldsymbol{u}_\mathcal{I}$ is part of the family of additively decomposable inequality measures [Shorrocks, 1980]. We use $\sqrt{D(\boldsymbol{u})}$ to scale the penalty with the sum of users' utilities.

Note that $\sqrt{D(\boldsymbol{u})}$ and $D'(\boldsymbol{u})$ have the same dependency to the overall scale of the utilities (i.e., multiplying all utilities by a constant factor has the effect of multiplying both penalties by the

same factor). Since both penalties drive towards equality, it is straightforward to show that the results of Section 3.3 as $\beta \to \infty$ also apply to $D'(\boldsymbol{u})$.

A.8 Exposure constraints at the level of every ranking

The notions of fairness of exposure are sometimes defined with item-side constraints defined at the level of *every ranking* [Singh and Joachims, 2018, Basu et al., 2020]. We give here the examples of constraints for equality of exposure and quality-weighted exposure:

$$\text{equality of exposure} \qquad P^{\text{expo}} \in \operatorname*{argmax}_{P \in \mathcal{P}} \sum_{i \in \mathcal{N}} u_i(P) \qquad \text{u.c.} \quad \forall (i,j) \in \mathcal{N} \times \mathcal{I}, P_{ij} v = \frac{\|v\|_1}{|\mathcal{I}|}$$

$$\text{quality-weighted exposure} \qquad P^{\text{qua}} \in \operatorname*{argmax}_{P \in \mathcal{P}} \sum_{i \in \mathcal{N}} u_i(P) \qquad \text{u.c.} \quad \forall (i,j) \in \mathcal{N} \times \mathcal{I}, P_{ij} v = \frac{\mu_{ij} \|v\|_1}{\sum_{j' \in \mathcal{I}} \mu_{ij'}}$$

The advantage of this formulation is that it leads to optimization problems that can be solved locally for every user, since there is no dependency between user rankings through item utility anymore.

However, applying the exposure criterion at the level of every ranking effectively applies a different notion of fairness. In our setting, this corresponds to defining a separate recommendation task for every user, i.e., taking $|\mathcal{N}| = 1$. The welfare function then mediates, within a single ranking, between the user utility and the utility of the different items.

When evaluated on exposures aggregated over all users, as we do in the paper, applying the fairness constraints at the level of individual rankings can lead to drastic reductions of user utility for no benefit in terms of total item exposure. This is summarized in the following result, which shows that there exists problems for which the optimal rankings for every $\theta \in \Theta$ satisfy the constraints of equality of exposure and quality-weighted exposure as we define them in Section 3.3, but when applying the constraints at the level of every ranking, it has the effect of reducing user utility. In the proposition, we use the notation of the objective function for parity of exposure F_β and F_β^{qua} of Section 3.3.

Proposition 28. *For every $d \in \mathbb{N}_*$ and every $N \in \mathbb{N}_*$, there is a one-sided recommendation task with $d+1$ items and $N(d+1)$ users such that, $\forall \theta \in \Theta$:*

$\forall \boldsymbol{u}^\theta \in \operatorname{argmax}_{\boldsymbol{u} \in \mathcal{U}} W_\theta(\boldsymbol{u})$, $\forall \beta > 0$ we have: $\boldsymbol{u}^\theta \in \operatorname*{argmax}_{\boldsymbol{u} \in \mathcal{U}} F_\beta(\boldsymbol{u})$ and $\boldsymbol{u}^\theta \in \operatorname*{argmax}_{\boldsymbol{u} \in \mathcal{U}} F_\beta^{\text{qua}}(\boldsymbol{u})$, and

$$\sum_{i \in \mathcal{N}} u_i(P^{\text{expo}}) = \frac{2}{d+1} \sum_{i \in \mathcal{N}} u_i^\theta \qquad \text{and} \qquad \sum_{i \in \mathcal{N}} u_i(P^{\text{qua}}) = (\frac{1}{2} + \frac{1}{d}) \sum_{i \in \mathcal{N}} u_i^\theta.$$

In other words, applying the constraints at the level of every ranking might lead to a drastic decrease of user utilities, even in tasks where satisfying the constraints on average over users (as we do in this paper) does not conflict with the optimal ranking.

Proof. We describe the problem with $N = 1$, the general case is obtained by repeating the preference pattern. Let us consider a task with $d+1$ users, $d+1$ items and a single recommendation slot. Let $i_1, \dots, i_{d+1}$ be the user indexes, and $j_1, \dots, j_{d+1}$ the item indexes. The preferences are defined as:

$$\forall k \in [\![d+1]\!], \mu_{i_k j_k} = 1 \qquad \forall j \neq j_k, \mu_{i_k j} = \frac{1}{d}.$$

All items have the same quality. For every $\theta \in \Theta$, $\boldsymbol{u}^\theta$ is given by the utilitarian ranking, which gives probability 1 to item j_k for user i_k, which leads to optimal user utility $u_i^\theta = 1$ and equal exposure to every item $u_j^\theta = 1$. Since the quality is the same for all items (equal to $1 + d\frac{1}{d}$), the ranking for $\boldsymbol{u}^\theta$ satisfies both equality of exposure and quality-weighted exposure constraints. Thus, for every $\beta > 0$, $\boldsymbol{u}^\theta \in \operatorname{argmax}_{\boldsymbol{u}\in\mathcal{U}} F_\beta(\boldsymbol{u})$ and $\boldsymbol{u}^\theta \in \operatorname{argmax}_{\boldsymbol{u}\in\mathcal{U}} F_\beta^{\text{qua}}(\boldsymbol{u})$.

On the other hand, satisfying equality of exposure at the level of every ranking requires $P_{ij}^{\text{expo}} = \frac{1}{d+1}$ for every user i and item j, which leads to $u_i(P^{\text{expo}}) = \frac{1}{d+1} + d \times \frac{1}{d} \times \frac{1}{d+1} = \frac{2}{d+1}$ for every user.

For quality-weighted exposure for every ranking, it leads to:

$$\forall k \in [\![d+1]\!], P_{i_k j_k}^{\text{qua}} = \frac{1}{2} \qquad\qquad \forall j \neq j_k, P_{i_k j}^{\text{qua}} = \frac{1}{d}$$

and thus a user utility $u_i(P^{\text{qua}}) = \frac{1}{2} + d \times \frac{1}{d} \times \frac{1}{d} = \frac{1}{2} + \frac{1}{d}$. □

Notice that in the examples of the proof, the global exposure of items is constant in P^{expo} and P^{qua}, as well as in the ranking given by optimal welfare. So from the point of view of our definitions of utility, applying the constraints at the level of every ranking only decreased user utility for the benefit of no items. Yet, we re-iterate that applying item-side fairness at the level of every ranking might be meaningful in some contexts. The goal of this section is to highlight the difference between using global and local definitions of item utilities.

Appendix B

Appendix of Chapter 5

B.1 Related work

The non-contextual setting of bandits with concave rewards (BCR) has been previously studied by Agrawal and Devanur [2014], and by Busa-Fekete et al. [2017] for the special case of Generalized Gini indices. In BCR, policies are distributions over actions. These approaches perform a direct optimization in policy space, which is not possible in the contextual setup without restrictions or assumptions on optimal policies. Agrawal et al. [2016] study a setting of CBCR where the goal is to find the best policy in a finite set of policies. Because they rely on explicit search in the policy space, they do not resolve the main challenge of the general CBCR setting we address here. Cheung [2019], Siddique et al. [2020], Mandal and Gan [2022], Geist et al. [2021] address multi-objective reinforcement learning with concave aggregation functions, a problem more general than stochastic contextual bandits. In particular, Cheung [2019] use a FW approach for this problem. However, these works rely on a tabular setting (i.e., finite state and action sets) and explicitly compute policies, which is not possible in our setting where policies are mappings from a continuous context set to distributions over actions. Our work is the only one amenable to contextual bandits with concave rewards by removing the need for an explicit policy representation. Finally, compared to previous FW approaches to bandits with concave rewards, e.g. [Agrawal and Devanur, 2014, Berthet and Perchet, 2017], our analysis is not limited to confidence-based exploration/exploitation algorithms.

CBCR is also related to the broad literature on bandit convex optimization (BCO) [Flaxman et al., 2004, Agarwal et al., 2011, Hazan et al., 2016, Shalev-Shwartz et al., 2012]. In BCO, the goal is to minimize a cumulative loss of the form $\sum_{t=1}^{T} \ell_t(\pi_t)$, where the convex loss function ℓ_t is *unknown* and the learner only observes the value $\ell_t(\pi_t)$ of the chosen parameter π_t at each timestep. Existing approaches to BCO perform gradient-free optimization in the parameter space. While BCR considers global objectives rather than cumulative ones, similar approaches have been used in non-contextual BCR [Berthet and Perchet, 2017] where the parameter space is the convex set of distributions over actions. As we previously highlighted, such parameterization does not apply to CBCR because direct optimization in policy space is infeasible.

CBCR is also related to multi-objective optimization [Miettinen, 2012, Drugan and Nowe, 2013], where the goal is to find all Pareto-efficient solutions. (C)BCR, focuses on one point of the Pareto front determined by the concave aggregation function f, which is more practical in our application settings where the decision-maker is interested in a specific (e.g., fairness) trade-off.

In recent years, the question of fairness of exposure attracted a lot of attention, and has been mostly studied in a static ranking setting [Geyik et al., 2019, Beutel et al., 2019a, Yang and

Stoyanovich, 2017, Singh and Joachims, 2018, Patro et al., 2022, Zehlike et al., 2021, Kletti et al., 2022a, Diaz et al., 2020, Do and Usunier, 2022, Wu et al., 2022b]. Existing work on fairness of exposure in bandits focused on local exposure constraints on the probability of pulling an arm at each timestep, either in the form of lower/upper bounds [Celis et al., 2018b] or merit-based exposure targets [Wang et al., 2021a]. In contrast, we consider amortized exposure over time, in line with prior work on fair ranking [Biega et al., 2018, Morik et al., 2020, Usunier et al., 2022], along with fairness trade-offs defined by concave objective functions which are more flexible than fairness constraints [Zehlike and Castillo, 2020, Do et al., 2021c, Usunier et al., 2022]. Moreover, these works [Celis et al., 2018b, Wang et al., 2021a] do not address combinatorial actions, while ours applies to ranking in the position-based model, which is more practical for recommender systems [Lagrée et al., 2016, Singh and Joachims, 2018]. The methods of [Patil et al., 2020, Chen et al., 2020] aim at guaranteeing a minimal cumulative exposure over time for each arm, but they also do not apply to ranking. In contrast, [Xu et al., 2021, Li et al., 2019] consider combinatorial bandits with fairness, but they do not address the contextual case, which limits their practical application to recommender systems. [Mansoury et al., 2021a, Jeunen and Goethals, 2021] propose heuristic algorithms for fairness in ranking in the contextual bandit setting, highlighting the problem's importance for real-world recommender systems, but they lack theoretical guarantees. Using our FW reduction with techniques from contextual combinatorial bandits [Lagrée et al., 2016, Li et al., 2016, Qin et al., 2014], we obtain the first principled bandit algorithms for this problem with provably vanishing regret.

B.2 More on experiments

Our experiments are fully implemented in Python 3.9.

B.2.1 Ranking CBCR: Application to fairness of exposure in rankings with bandit feedback

B.2.1.1 Details of the environment and algorithms

Environment Following [Patro et al., 2020] who also address fairness in recommender systems, we use the Last.fm music dataset[1] from [Cantador et al., 2011], which includes the listening counts of $1,892$ users for the tracks of $17,632$ artists, which we identify as the items. For the first environment, which we presented in Section 5.5 and which we call *Lastfm-50* here, we extract the top $n = 50$ users and $m = 50$ items having the most interactions. In order to examine algorithms at larger scale, we also design another environment, *Lastfm-2k*, where we keep all $n = 1.9k$ users and the top $m = 2.5k$ items having the most interactions. In both cases, to generate contexts and rewards, we follow a protocol similar to other works on linear contextual bandits [Garcelon et al., 2020b, Li et al., 2016]. Using low-rank matrix factorization with d' latent factors[2], we obtain user factors $u_j \in \mathbb{R}^{d'}$ and item factors $v_i \in \mathbb{R}^{d'}$ for all $j, i \in [\![n]\!] \times [\![m]\!]$. We design the context set as $\mathcal{X} = \{\text{flatten}(u_j v_i^\intercal) : j, i \in [\![n]\!] \times [\![m]\!]\} \subset \mathbb{R}^d$, where $d = d'^2$. At each time step t, the environment draws a user j_t uniformly at random from $[\![n]\!]$ and sends context $x_t = \text{flatten}(u_{j_t} v_i^\intercal)$. Given a context x_t and item i, clicks are drawn from a Bernoulli distribution: $c_{t,i} \sim \mathcal{B}(u_{j_t}^\intercal v_i)$.

We set $\bar{k} = 10$, and for the position weights, we use the standard weights of the discounted cumulative gain (DCG): $\forall k \in [\![\bar{k}]\!], b_k = \frac{1}{\log_2(1+k)}$ and $b_{\bar{k}+1}, \dots, b_m = 0$.

[1] https://www.last.fm, the dataset is publicly available for non-commercial use.
[2] Using the Python library Implicit, MIT License: https://implicit.readthedocs.io/

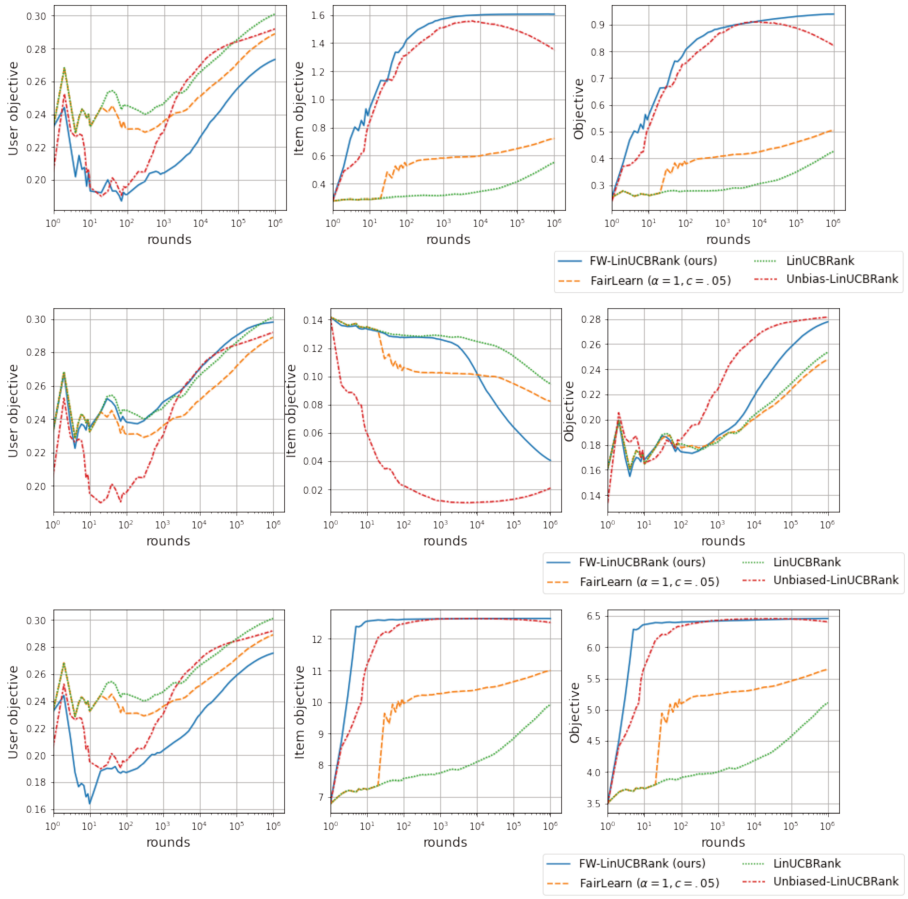


Figure B.1: *Lastfm-50*: Objective values over time for (top) *Gini*, (middle) *eq. exposure*, (bottom) *welf.*

Details of the algorithms For all algorithms, the regularization parameter of the Ridge regression is set to $\lambda = 0.1$.

The first baseline we consider is the algorithm *LinUCBRank*[3] of [Ermis et al., 2020], which is a top-$\overline{k}$ ranking bandit algorithm without fairness. It is equivalent to using FW-LinUCBRank with $f(s) = s_{m+1}$, which corresponds to the usual top-$\overline{k}$ ranking objective without item fairness. More precisely, at each timestep, the algorithm produces a top-$\overline{k}$ ranking of $\left(\hat{\theta}_{t-1}^{\intercal} x_{t,i} + \alpha_t(\frac{\delta'}{3}) \left\|x_{t,i}\right\|_{V_{t-1}^{-1}}\right)_{i=1}^{m}$.

We also consider as baselines two bandit algorithms with amortized fairness of exposure criteria. First, Mansoury et al. [2021a] proposed a fairness module for cascade ranking bandits, which can be easily adapted to the position-based model (PBM). Their goals include reducing inequality in exposure between items, measured by the Gini index of exposures in their experiments. While they measure the exposure of an item as their recommendation frequency over time, we adapt their module to the PBM by using the observation frequency, i.e. $\sum_{t'=1}^{t} e_{t',i}$ for item i at time t. Transposed to our setting, their module consists in a simple modification of *LinUCBRank* by multiplying the exploration bonus of each item i by a factor:

$$\eta_{t,i} = 1 - \frac{\sum_{t'=1}^{t-1} e_{t',i}}{\sum_{t'=1}^{t-1} \frac{1}{k} \sum_{i'=1}^{m} e_{t',i'}}.$$

More precisely, at each timestep, the algorithm produces a top-$\overline{k}$ ranking of $\left(\hat{\theta}_{t-1}^{\intercal} x_{t,i} + \eta_{t,i} \times \alpha_t(\frac{\delta'}{3}) \left\|x_{t,i}\right\|_{V_{t-1}^{-1}}\right)_{i=1}^{m}$.

[3] *LinUCBRank* appears under various names in the literature, including PBMLinUCBRank [Ermis et al., 2020] and CascadeLinUCB [?].

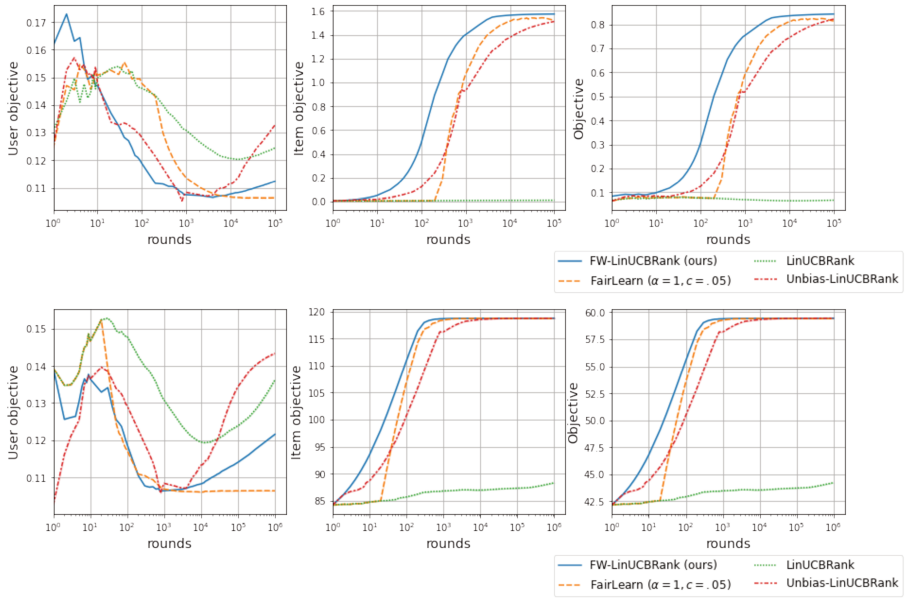


Figure B.2: *Lastfm-2k*: Objective values over time for (top) *Gini*, (bottom) *welf*.

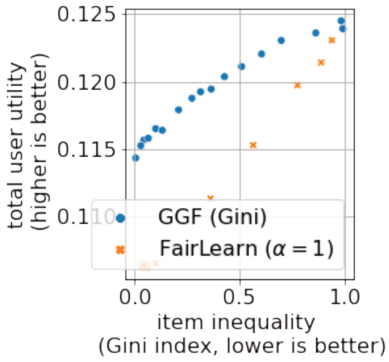


Figure B.3: Trade-offs between user utility and inequality on *Lastfm-2k*, after $T = 10^6$ rounds.

Following [Mansoury et al., 2021a], we call this baseline *Unbiased-LinUCBRank*.

Our second baseline with fairness is the *FairLearn*(c, α) algorithm of Patil et al. [2020] for stochastic bandits with a fairness constraint on the pulling frequency $N_{t,i}$ of each arm i at each timestep t. The constraint is parameterized by a variable c and a tolerance parameter α: $\lfloor ct \rfloor - N_{t,i} \leq \alpha$. We adapt *FairLearn*$(c, \alpha)$ to ranking by applying the algorithm sequentially for each recommendation slot, while constraining the algorithm not to choose the same item twice for a given ranked list. We also adapt *FairLearn* to contextual bandits by using LinUCB as underlying learning algorithm. More precisely, for the current timestep and slot, if the constraint is not violated, then the algorithm plays the item with the highest LinUCB upper confidence bound.

Objectives To illustrate the flexibility of our approach, we use algorithm FW-LinUCBRank to optimize three existing objectives which trade off between user utility and item fairness, in the form: $f(s) = s_{m+1} + \beta f^{\text{item}}(s_{1:m})$. *Gini* measures item inequality by the Gini index, as in [Biega et al., 2018, Morik et al., 2020, Do and Usunier, 2022], and *eq. exposure* uses the standard deviation

[Do et al., 2021c]:

$$\textit{(Gini)}\quad f^{\text{item}}(s) = \sum_{j=1}^{m} \frac{m-j+1}{m} s_j^{\uparrow} \qquad \textit{(eq. expo)}\quad f^{\text{item}}(s) = -\frac{1}{m}\sqrt{\sum_{j=1}^{m}\left(s_j - \frac{1}{m}\sum_{j'=1}^{m} s_{j'}\right)^2}$$

Since *Gini* is nonsmooth, we apply the FW-LinUCBRank algorithm for nonsmooth f with Moreau-Yosida regularization, presented in Section 5.3.3 and detailed in Appendix B.6.1 (we use $\beta_0 = 1$ in our experiments). To compute the gradient of the Moreau envelope f_t, we use the algorithm of Do and Usunier [2022] which specifically applies to generalized Gini functions and top-$\overline{k}$ ranking.

We also study additive concave welfare functions [Do et al., 2021c, Moulin, 2003] where α is a parameter controlling the degree of redistribution of exposure to the worse-off items:

$$\textit{(Welf)}\quad f^{\text{item}}(s) = \sum_{j=1}^{m} s_j^{\alpha}\,, \quad \alpha > 0$$

B.2.1.2 Additional results

We now present additional results, which are obtained by repeating each simulation with 10 different random seeds.

Dynamics For the three objectives described, Figure B.1 represents the values of the user and item objectives (left and middle), and the value of the objective f (right) over time, achieved by the competing algorithms on *Lastfm-50*. We set $\beta = 0.5$ for all objectives and for *welf*, we set $\alpha = 0.5$. We observe that with this value of β, the item objective f^{item} is given more importance in f than the user utility.

We observe that for *Gini* and *welf*, *FW-LinUCBRank* achieves the highest value of f across timesteps. This is because unlike *LinUCBRank*, it accounts for the item objective f^{item}. In both cases, *Unbiased-LinUCBRank* achieves a high value of f over time but starts decreasing, after 10^4 iterations for *Gini* and 5.10^5 iterations for *welf*. This is because *Unbiased-LinUCBRank* is not designed to converge towards an optimum of f. For *eq. exposure*, when $\beta = 0.5$, *Unbiased-LinUCBRank* obtains surprisingly better values of f than *FW-LinUCBRank*. Therefore, depending on the objective to optimize and the timeframe, *Unbiased-LinUCBRank* can be chosen as an alternative to *FW-LinUCBRank*. However, due to its lack of theoretical guarantees, it is more difficult to understand in which cases it may work, and for how many iterations. Furthermore, unlike *Unbiased-LinUCBRank*, *FW-LinUCBRank* can be chosen to optimise a wide variety of functions by varying the tradeoff parameter β in all objectives, and α in *welf* to control the degree of redistribution. *Unbiased-LinUCBRank* does not have such controllability and flexibility.

Figure B.2 shows the objective values for *Gini* and *welf* on *Lastfm-2k*. We observe similar results where *FW-LinUCBRank* converges more quickly than its competitors ($\approx 5,000$ iterations for *Gini* and ≈ 500 iterations for *welf*) and obtains the highest values of f. For the first 10^5 iterations of optimizing *Gini*, *Unbiased-LinUCBRank* obtains significantly lower values than *FW-LinUCBRank* on *welf*.

Fairness trade-off for fixed T On the larger *Lastfm-2k* dataset, we study the tradeoffs between user utility and item inequality obtained by *FW-LinUCBRank* and *FairLearn* on Figure B.3 after $T = 10^6$ rounds. The Pareto frontiers are obtained as follows: *FW-LinUCBRank* optimises for *Gini*, in which we vary β, and for *FairLearn* we vary the constraint value c at fixed $\alpha = 1$. Figure 5.1 in Section 5.5 of the main paper illustrated the same Pareto frontier but for $5\times$ more iterations and

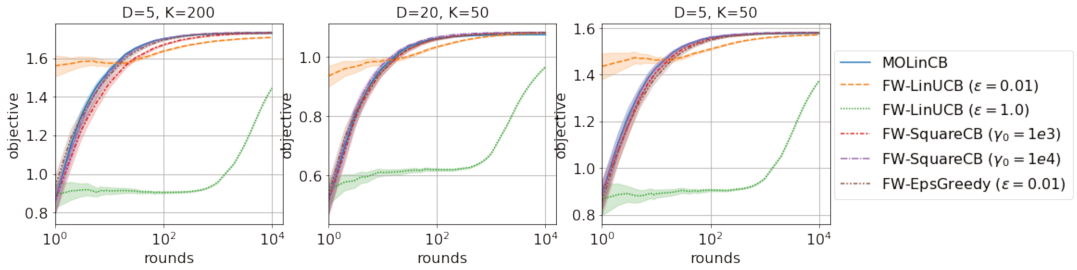


Figure B.4: Multi-objective bandits: GGF value achieved on various synthetic environments.

on the smaller *Lastfm-50* dataset. Although the algorithms might not have converged for this larger dataset, we observe that *FW-LinUCBRank* obtains better trade-offs than *FairLearn*, achieving higher user utility at all levels of inequality. We conclude that even in a setting with more items and shorter learning time, *FW-LinUCBRank* effectively reduces item inequality, at lower cost for user utility than the baseline.

B.2.2 Multi-armed CBCR: Application to multi-objective bandits with generalized Gini function

We provide the details and additional simulations on the task of optimizing the Generalized Gini aggregation Function (GGF) in multi-objective bandits [Busa-Fekete et al., 2017, Mehrotra et al., 2020]. We remind that the goal is to maximize a GGF of the D-dimensional rewards, which is a nonsmooth concave aggregation function parameterized by nonincreasing weights $w_1 = 1 \geq \ldots \geq w_D \geq 0$: $f(s) = \sum_{i=1}^{D} w_i s_i^{\uparrow}$, where $(s_i^{\uparrow})_{i=1}^{D}$ denotes the values of s sorted in increasing order.

Mehrotra et al. [2020] study the contextual bandit setting, motivated by music recommendation on Spotify with multiple metrics. They consider atomic actions $a_t \in \mathcal{A}$ (i.e., $\mathcal{A}$ is the canonical basis of $\mathbb{R}^K$) and a linear reward model: $\forall i \in [\![D]\!]$, $\exists \theta_i \in \mathbb{R}^d$, $\mathbb{E}_t[r_{t,i}] = \theta_i^\intercal x_t^\intercal a_t$. These are the same assumptions as described in Table 5.1 of Section 5.3.2 and in Appendix B.7.

GGFs are concave functions, but they are nondifferentiable. Therefore, we use the variant of our FW approach for nonsmooth f (see Section 5.3.3), where we smooth the objective via Moreau-Yosida regularization with parameter $\beta_0 = 0.01$, using the algorithm of [Do and Usunier, 2022] to compute the gradients of the smooth approximations f_t.

Algorithms In the main body, we evaluated two instantiations of our FW meta-algorithm, namely FW-LinUCB and FW-SquareCB. The level of exploration in FW-LinUCB is controlled by a variable ϵ. More precisely, the exploration bonus is multiplied by $\sqrt{\epsilon}$, i.e. the UCBs are calculated as: $\hat{\theta}_{t-1,i}^\intercal x_{t,k} + \sqrt{\epsilon}\,\alpha_t(\delta)\,\|x_{t,k}\|_{V_{t-1}^{-1}}$. In FW-Square-CB, as detailed in Appendix B.8, the exploration is controlled by a sequence $(\gamma_t)_{t\geq 1}$, growing as $\sqrt{t}$ (higher γ_t means less exploration). We set it to $\gamma_t = \gamma_0\sqrt{t}$ with $\gamma_0 \in \{10^3, 10^4\}$.

In addition to the two algorithms presented in Section 5.5, to show the flexibility of our FW approach, we also implement FW-ϵ-greedy, another instantiation of our FW algorithm which uses $\epsilon-$greedy as scalar bandit algorithm.

We compare our algorithms with MOLinCB of Mehrotra et al. [2020], an online gradient descent-style algorithm which was designed for this task, but was introduced without theoretical guarantees, as an extension of the MO-OGDE algorithm of Busa-Fekete et al. [2017] who study the non-contextual problem. We use the default parameters of MOLinCB recommended by Mehrotra et al. [2020].

Environments Since the Spotify dataset of Mehrotra et al. [2020] is not publicly available, we only focus on their simulated, controlled environments. We reproduced these environments exactly as described in Appendix A of their paper. For completeness, we restate the protocol here: we draw a hidden parameter $\theta \in \mathbb{R}^{D\times d}$ uniformly at random in $[0,1]$, and each element of a context-arm vector $x_{t,k}$ is drawn from $\mathcal{N}(\frac{1}{d}, \frac{1}{d^2})$. Given a context x_t and arm k_t, the D-dimensional reward is generated as a draw from $\mathcal{N}(\theta x_{t,k_t}, 0.01(\theta x_{t,k_t})^2)$. We choose $d = 10$ in the data generation and $\lambda = 0.1$ in the Ridge regression, as recommended by Mehrotra et al. [2018].

In Section 5.5 of the main body, we varied the number of objectives $D \in \{5, 20\}$ and set $K = 50$. Here we also experiment with $K = 200$ to see the effect of varying the number of arms. The GGF weights are set to $w_j = \frac{1}{2^{j-1}}$. Each simulation is repeated with 100 different random seeds.

Results The extended results, with more arms and algorithms, are depicted in Figure B.4. We observe that FW-ϵ-greedy achieves similar performance to the baseline MOLinCB, with small exploration $\epsilon = 0.01$. FW-SquareCB also achieves comparable performance to MOLinCB when there is little exploration, i.e. with $\gamma_0 = 10^4$ rather than 10^3. This is coherent with our observation in Section 5.5 that FW-LinUCB obtains better performance when there is very little exploration on this environment from Mehrotra et al. [2018]. Note that there is no forced exploration in their algorithm MOLinCB. Overall, we obtain qualitatively similar results when $K = 200$ compared to $K = 50$.

B.3 Proofs of Section 5.2

In this section we give the missing details of Section 5.2. For completeness, we remind the definitions of Lipschitz-continuity and super-gradients in the next subsection. Then, we start in Section B.3.2 the analysis of the structure of the set $\mathcal{S}$ defined in Section 5.3 of the main paper, and more precisely its support function $g \mapsto \max_{s\in\mathcal{S}} g^\intercal s$. This contains new lemmas that are fundamental for the analysis throughout the paper, in particular in the proof of Lemma 32, which is given in Section B.3.3.

B.3.1 Brief reminder on Lipschitz functions and super-gradients

We remind the following definitions. Let D and D' be two integers, and f a function $f : \mathbb{R}^D \to \mathbb{R}^{D'}$. We have:

- *(Lipschitz continuity)* f is L-Lipschitz continuous with respect to $\|.\|_2$ on a set $\mathcal{Z} \subseteq \mathbb{R}^D$ if

$$\forall z, z' \in \mathcal{Z}, \|f(z) - f(z')\|_2 \leq L \|z - z'\|_2 .$$

- *(super-gradients)* If $f : \mathbb{R}^D \to \mathbb{R}\cup\{\pm\infty\}$, a super-gradient of f at a point $z \in \mathbb{R}^D$ where $f(z) \in \mathbb{R}$ is a vector g such that for all $z' \in \mathbb{R}^D$, $f(z') \leq f(z) + \langle g \,|\, z' - z\rangle$.

We remind the following results when $f : \mathbb{R}^D \to \mathbb{R}\cup\{\pm\infty\}$ is a proper closed concave function:

- f has non-empty set of super-gradients at every point z where $f(z) \in \mathbb{R}$,
- if f is L-Lipschitz on $\mathcal{Z} \subseteq \mathbb{R}^D$ and $\mathcal{Z}$ is open, then for every $z \in \mathcal{Z}$ and every super-gradient g of f at z, we have $\|g\|_2 \leq L$.

The assumption of Lipschitz-continuity of f on a set $\mathcal{Z}$ implicitly implies the assumption that $\mathcal{Z}$ is in the domain of f.

Remark 4 (About our Lipschitzness assumptions). *We use Lipschitzness over an open set containing $\mathcal{K}$ in Assumption A because we use boundedness of the super-gradients of f. In fact, a more precise alternative would be to require that super-gradients are bounded uniformly on $\mathcal{K}$ by L. We choose the Lipschitz formulation because we believe it is more natural.*

As a side note, in assumption B, we use Lipschitzness of the gradients on $\mathcal{K}$, not on an open set containing $\mathcal{K}$. This is because smoothness in used in the ascent lemma (see Eq. B.5), which uses Inequality 4.3 of Bottou et al. [2018], the proof of which directly uses Lipschitz-continuity of the gradients on $\mathcal{K}$ [Bottou et al., 2018, Appendix B], without relying on an argument of boundedness of gradients.

B.3.2 Preliminaries: the structure of the set $\mathcal{S}$

We denote by $x_{1:T} = (x_1, \dots, x_T)$ a sequence of contexts of length T. Let

$$\mathcal{S} = \left\{ \mathbb{E}_{x\sim P}\left[\mu(x)\overline{\pi}(x)\right] \middle| \overline{\pi} : \mathcal{X} \to \overline{\mathcal{A}} \right\}$$
$$\forall x_{1:T} \in \mathcal{X}^T, \mathcal{S}(x_{1:T}) = \left\{ \frac{1}{T}\sum_{t=1}^{T} \mu(x_t)\overline{\pi}(x_t) \middle| \overline{\pi} : \mathcal{X} \to \overline{\mathcal{A}} \right\}$$

It is straightforward to show that $\mathcal{S}(x_{1:T}) = \left\{ \frac{1}{T}\sum_{t=1}^{T} \mu(x_t)\pi_t \middle| (\pi_1, \dots, \pi_T) \in \overline{\mathcal{A}}^T \right\}$. These sets are particularly relevant because of the following equality, for every $f : \mathbb{R}^D \to \mathbb{R} \cup \{\pm\infty\}$:

$$f^* = \sup_{\overline{\pi}:\mathcal{X}\to\overline{\mathcal{A}}} f\Big(\mathbb{E}_{x\sim P}\left[\mu(x)\overline{\pi}(x)\right]\Big) = \sup_{s\in\mathcal{S}} f(s) \tag{B.1}$$

and

$$f_T^+ = \sup_{(\pi_t)_{t\in[\![T]\!]}\in\overline{\mathcal{A}}^T} f\Big(\frac{1}{T}\sum_{t=1}^{T}\mu(x_t)\pi_t\Big) = \sup_{s\in\mathcal{S}(x_{1:T})} f(s).$$

We study in this section the structure of these sets. We provide here the part of Assumption A that is relevant to this section:

Assumption $\tilde{\text{A}}$. *$\mathcal{A}$ is a compact subset of $\mathbb{R}^K$ and there is a compact convex set $\mathcal{K} \subseteq \mathbb{R}^D$ such that $\forall (x,a) \in \mathcal{X}\times\mathcal{A}, \mu(x)a \in \mathcal{K}$.*

We remind the following basic results from convex sets in Euclidian spaces that we use throughout the paper without reference:

Lemma 29. *Let $\mathcal{A}$ be a compact subset of $\mathbb{R}^K$. We have:*

- *[Rockafellar and Wets, 2009, Corollary 2.30] The convex hull $\overline{\mathcal{A}}$ of a, denoted by $\overline{\mathcal{A}}$, is compact.*
- *For every $w \in \mathbb{R}^K$, $\max_{a\in\mathcal{A}} w^\intercal a = \max_{a\in\overline{\mathcal{A}}} w^\intercal a$.*

The following lemma allows us to use maxima instead of suprema over $\mathcal{S}$ and $\mathcal{S}(x_{1:T})$. The proof of this lemma is deferred to Appendix B.10.1.

Lemma 30. *Under Assumption $\tilde{A}$, $\mathcal{S}$ is compact and $\forall T \in \mathbb{N}_*$, $\forall x_{1:T} \in \mathcal{X}^T, \mathcal{S}(x_{1:T})$ is compact.*

The next result regarding the support functions of $\mathcal{S}$ and $\mathcal{S}(x_{1:T})$ is the key to our approach:

Lemma 31. *Let $w \in \mathbb{R}^D$ and $T \in \mathbb{N}_*$. Under Assumption $\tilde{A}$, we have*

$$\mathbb{E}_{x_{1:T}\sim P^T}\Big[\max_{s\in\mathcal{S}(x_{1:T})} w^\intercal s\Big] = \max_{s\in\mathcal{S}} w^\intercal s.$$

Moreover, for every $\delta \in (0,1]$, we have with probability at least $1-\delta$:

$$\max_{s\in\mathcal{S}(x_{1:T})} w^\intercal s \le \max_{s\in\mathcal{S}} w^\intercal s + \|w\|_2 D_{\mathcal{K}} \sqrt{\frac{2\ln \delta^{-1}}{T}}.$$

The inequality $\max_{s\in\mathcal{S}} w^\intercal s \le \max_{s\in\mathcal{S}(x_{1:T})} w^\intercal s + \|w\|_2 D_{\mathcal{K}} \sqrt{\frac{2\ln \delta^{-1}}{T}}$ *also holds with probability* $1-\delta$.

Proof. The first result is a direct consequence of the maximization of linear functions over the simplex. Using (B.1) with $f(s) = w^\intercal s$ and the linearity of expectations, we have

$$\max_{s\in\mathcal{S}} w^\intercal s = \max_{\overline{\pi}:\mathcal{X}\to\overline{\mathcal{A}}} \mathbb{E}_{x\sim P}\big[w^\intercal \mu(x)\overline{\pi}(x)\big].$$

The optimal policy given w, denoted by $\overline{\pi}^w$ is thus obtained by optimizing for every x the dot product between $w^\intercal\mu(x) \in \mathbb{R}^K$ and $\overline{\pi}(x) \in \overline{\mathcal{A}} \subseteq \mathbb{R}^K$. Since, for each x, it is a linear optimization, we can find an optimizer in $\mathcal{A}$ (see Lemma 29), which gives:

$$\max_{s\in\mathcal{S}} w^\intercal s = \mathbb{E}_{x\sim P}\big[\underbrace{w^\intercal\mu(x)\overline{\pi}^w(x)}_{\eta^w(x)}\big] \qquad \text{where } \overline{\pi}^w(x) \in \operatorname*{argmax}_{a\in\mathcal{A}} w^\intercal\mu(x)a,$$

where in the equation above we mean that $\overline{\pi}^w$ is a measurable selection of $x \mapsto \operatorname{argmax}_{a\in\mathcal{A}} w^\intercal\mu(x)a$. For the same reason, we have $\max_{s\in\mathcal{S}(x_{1:T})} w^\intercal s = \frac{1}{T}\sum_{t=1}^T \eta^w(x_t)$. We obtain

$$\mathbb{E}_{x_{1:T}\sim P^T}\Big[\max_{s\in\mathcal{S}(x_{1:T})} w^\intercal s\Big] = \mathbb{E}_{x_{1:T}\sim P^T}\Big[\frac{1}{T}\sum_{t=1}^T \eta^w(x_t)\Big] = \mathbb{E}_{x\sim P}\big[\eta^w(x)\big] = \max_{s\in\mathcal{S}} w^\intercal s.$$

which is the first equality.

For the high-probability inequality, let $X_t = \eta^w(x_t) - \mathbb{E}_{x\sim P}\big[\eta^w(x)\big]$. Since the $(x_t)_{t\in[\![T]\!]}$ are independent and identically distributed (i.i.d.), the variables $(X_t)_{t\in[\![T]\!]}$ are also i.i.d., and we have

$$|X_t| \le w^\intercal\Big(\underbrace{\mu(x_t)\overline{\pi}^w(x_t)}_{\in\mathcal{K}} - \underbrace{\mathbb{E}_{x\sim P}\big[\mu(x)\overline{\pi}^w(x)\big]}_{\in\mathcal{K}}\Big) \le \|w\|_2 D_{\mathcal{K}} \qquad \text{and } \mathbb{E}\big[X_t\big] = 0.$$

Given $\delta \in (0,1]$, Hoeffding's inequality applied to $\frac{1}{T}\sum_{t=1}^T X_t$ gives, with probability at least $1-\delta$:

$$\max_{s\in\mathcal{S}(x_{1:T})} w^\intercal s - \max_{s\in\mathcal{S}} w^\intercal s = \frac{1}{T}\sum_{t=1}^T X_t \le \|w\|_2 D_{\mathcal{K}} \sqrt{\frac{2\ln\delta^{-1}}{T}}.$$

The reverse equation is obtained by applying Hoeffding's inequality to $-\frac{1}{T}\sum_{t=1}^T X_t$. □

B.3.3 Proof of Lemma 32

Lemma 32. *Under Assumption A, $\forall T \in \mathbb{N}_*, \forall \delta \in (0,1]$, we have, with probability at least $1-\delta$:*

$$\big|f_T^+ - f^*\big| \le L D_{\mathcal{K}} \sqrt{\frac{2\ln\frac{4e^2}{\delta}}{T}} \qquad \textit{where } f_T^+ = \max_{(\pi_1,\dots,\pi_T)\in\overline{\mathcal{A}}^T} f\Big(\frac{1}{T}\sum_{t=1}^T \mu(x_t)\pi_t\Big)$$

We also have, with probability $1-\delta$ *over contexts, actions, and rewards:*

$$\left|f(s_T) - f(\hat{s}_T)\right| \leq LD_{\mathcal{K}}\sqrt{\frac{2\ln(2e^2\delta^{-1})}{T}} \quad \textit{where } s_T = \frac{1}{T}\sum_{t=1}^{T}\mu(x_t)a_t.$$

The first statement shows that the performance of the optimal non-stationary policy over T steps converges to f^* at a rate $O(1/\sqrt{T})$. Furthermore, measuring the algorithm's performance by expected rewards instead of observed rewards would also amount to a difference of order $O(1/\sqrt{T})$. This choice would lead to what is commonly referred to as a *pseudo*-regret. Since the worst-case regret of BCR is $\Omega(1/\sqrt{T})$ [Bubeck and Cesa-Bianchi, 2012], the previous lemma shows that the alternative definitions of regret would not substantially change our results.

Proof. We start with the first inequality.

We first prove that w.p. greater than $1-\delta/2$, we have $f_T^+ \leq f^* + LD_{\mathcal{K}}\sqrt{\frac{2\ln\frac{2}{\delta}}{T}}$.

Since f is continuous on $\mathcal{K}$ and since $\mathcal{S} \subseteq \mathcal{K}$ and $\mathcal{S}$ is compact by Lemma 30, there is $s^* \in \mathcal{S}$ such that $f^* = f(s^*)$. Similarly, since $\mathcal{S}(x_{1:T})$ is compact, there is s_T^* such that $f(s_T^*) = \max_{s\in\mathcal{S}(x_{1:T})} f(s)$. Using (B.1), we need to prove that with probability at least $1-\delta/2$, we have $f(s_T^*) \leq f(s^*) + LD_{\mathcal{K}}\sqrt{\frac{2\ln\frac{2}{\delta}}{T}}$.

Using the concavity of f, let g^* be a supergradient of f at s^*. We have

$$\begin{aligned}
f(s_T^*) &\leq f(s^*) + \langle g^* \,|\, s_T^* - s^*\rangle \\
&\leq f(s^*) + \max_{s\in\mathcal{S}(x_{1:T})} \langle g^* \,|\, s - s^*\rangle \\
\implies \text{w.p. } \geq 1-\delta/2: \quad f(s_T^*) &\leq f(s^*) + \underbrace{\max_{s\in\mathcal{S}} \langle g^* \,|\, s - s^*\rangle}_{\leq 0 \text{ by def. of } s^*} + \|g^*\|_2 D_{\mathcal{K}}\sqrt{\frac{2\ln\frac{2}{\delta}}{T}} \quad \text{(by Lemma 31)} \\
&\leq f(s^*) + LD_{\mathcal{K}}\sqrt{\frac{2\ln\frac{2}{\delta}}{T}}. \qquad \text{(by the Lipschitz assumption)}
\end{aligned}$$

We now prove $f^* \leq f_T^+ + LD_{\mathcal{K}}\sqrt{\frac{2\ln\frac{4e^2}{\delta}}{T}}$ with probability at least $1-\delta/2$.

Let $\overline{\pi}^* \in \operatorname{argmax}_{\overline{\pi}:\mathcal{X}\to\overline{\mathcal{A}}} f\Big(\mathbb{E}_{x\sim P}\big[\mu(x)\overline{\pi}(x)\big]\Big)$ (an optimal policy exists by Lemma 30). Denote by $(X_t = \mu(x_t)\overline{\pi}^*(x_t))_{t\in[\![T]\!]}$ a sequence of independent and identically distributed random variables obtained by sampling $x_t \sim P$.

We have $|X_t - \mathbb{E}X_t| \leq D_{\mathcal{K}}$ and $\mathbb{E}X_t = s^*$. By the Lipschitz property of f, we obtain

$$f(s^*) \leq f\Big(\frac{1}{T}\sum_{t=1}^{T} X_t\Big) + L\,\|[\|\Big]\frac{1}{T}\sum_{t=1}^{T} X_t - s^*\Big\|_2.$$

Using the version of Azuma's inequality for vector-valued martingale with bounded increments of Hayes [2005, Theorem 1.8] to obtain, for every $\epsilon > 0$:

$$\mathbb{P}\Big(\frac{1}{D_{\mathcal{K}}}\,\|[\|\Big]\frac{1}{T}\sum_{t=1}^{T} X_t - s^*\Big\|_2 \geq \epsilon\Big) \leq 2e^2 e^{-T\epsilon^2/2}.$$

Setting $\frac{\delta}{2} = 2e^2 e^{-T\epsilon^2/2}$ and solving for ϵ gives, with probability at least $1-\delta/2$:

$$f^* \leq f\Big(\frac{1}{T}\sum_{t=1}^{T} X_t\Big) + LD_{\mathcal{K}}\sqrt{\frac{2\ln\frac{4e^2}{\delta}}{T}} \leq f_T^+ + LD_{\mathcal{K}}\sqrt{\frac{2\ln\frac{4e^2}{\delta}}{T}}.$$

For the second inequality: using L-Lipschitzness of f, the inequality is a direct consequence of the lemma below, which is itself a direct consequence of [Hayes, 2005, Theorem 1.8]. □

In the following lemma and its proof, we use the two following filtrations:

- $\mathfrak{F} = (\mathfrak{F}_t)_{t\in\mathbb{N}_*}$ where $\mathfrak{F}_t$ is the σ-algebra generated by $(x_1, a_1, r_1, \dots, x_{t-1}, a_{t-1}, r_{t-1}, x_t)$,
- $\overline{\mathfrak{F}} = (\overline{\mathfrak{F}}_T)_{T\in\mathbb{N}_*}$ where $\overline{\mathfrak{F}}_T$ is the σ-algebra generated by $(x_1, a_1, r_1, \dots, x_{t-1}, a_{t-1}, r_{t-1}, x_t, a_t)$.

Our setup implies that the process $(a_t)_{t\in\mathbb{N}_*}$ is adapted to $\mathfrak{F}$ while $(r_t)_{t\in\mathbb{N}_*}$ is adapted to $\overline{\mathfrak{F}}$.

Lemma 33. *Under Assumption A, if the actions $(a_1, \dots, a_T)$ define a process adapted to $(\mathfrak{F}_T)_{T\in\mathbb{N}}$, then, for every $T \in \mathbb{N}$, for every δ, with probability $1-\delta$, we have:*

$$\|s_T - \hat{s}_T\|_2 \leq D_{\mathcal{K}} \sqrt{\frac{2 \ln \frac{2e^2}{\delta}}{T}}$$

Proof. Let $X_T = \sum_{t=1}^{T} r_t - \mu(x_t)a_t$. We have $\|X_T - X_{T-1}\|_2 \leq D_{\mathcal{K}}$, and $(X_T)_{T\in\mathbb{N}}$ is a martingale adapted to $(\overline{\mathfrak{F}}_T)_{T\in\mathbb{N}}$ satisfying $X_0 = 0$. We can then use the version of Azuma's inequality for vector-valued martingale with bounded increments of Hayes [2005, Theorem 1.8] to obtain, for every $\epsilon > 0$:

$$\mathbb{P}\Big(\|[\|\Big] \frac{X_T}{D_{\mathcal{K}}}_2 \geq \epsilon \Big) \leq 2e^2 e^{-\epsilon^2/(2T)}.$$

Solving for ϵ gives the desired result. □

B.4 The general template Frank-Wolfe algorithm

Algorithm 6: Generic Frank-Wolfe algorithm for CBCR.

input: initial point $z_0 \in \mathcal{K}$, Approx. RLOO confidence parameter δ'

1 **for** $t = 1 \dots T$ **do**
2 Observe $x_t \sim P$
3 Pull $a_t \sim \mathfrak{A}(h_t, x_t, \delta')$ // Explore/exploit step
4 Observe reward $r_t \in \mathcal{K}$, update temporal average of observed rewards $\hat{s}_t$
5 Let $\rho_t = \mathfrak{U}(h_{t+1}, \delta')$ // Generic Frank-Wolfe update
6 Update $z_t = z_{t-1} + \frac{1}{t}(\rho_t - z_{t-1})$
7 **end**

A more general framework The analysis of the next sections is done within a more general famework than that of the main paper, which is described in Algorithm 6. Similarly to the main paper, the action is drawn according to $a_t \sim \mathfrak{A}(h_t, x_t, \delta')$ (Line 3 of Alg. 6). However, we allow for a generic choice of Frank-Wolfe iterate with respect to which we compute (an extension of) the scalar regret (presented in (B.2) below). The *update direction* is denoted by ρ_t and is chosen according to a function $\mathfrak{U}(h_{t+1}, \delta')$, a companion function from $\mathfrak{A}(h_t, x_t, \delta')$. Note that the update direction is chosen given $h_{t+1} = (h_t, (x_t, a_t, r_t))$, the history after the actions and rewards have been taken.

The proofs of the main paper apply to the special case of Alg. 6 where $\forall t \geq 1, \rho_t = r_t$. We then have the FW iterate z_t in Line 6 of the algorithm satisfy $\forall t \geq 1, z_t = \hat{s}_t$.

The reason we study this generalization is to show how our analysis applies in cases where the FW iterate is not the observed reward. In prior work on (non-contextual) BCR, Agrawal and

Devanur [2014, Algorithm 4] use an upper-confidence approach and use the upper confidence on the expected reward as update direction. The generalization made by introducing $\mathfrak{U}(h_{t+1}, \delta')$ compared to the main paper allows for our analysis to encompass their approach.

We need to update Assumptions A and B to account for the fact that ρ_t is used in place of r_t.

Assumption A′. *f is closed proper concave on $\mathbb{R}^D$ and $\mathcal{A}$ is a compact subset of $\mathbb{R}^K$. Moreover, there is a compact convex set $\mathcal{K} \subseteq \mathbb{R}^D$ such that*

- (Bounded rewards and iterates) *For all $t \in \mathbb{N}_*$, $r_t \in \mathcal{K}$ and $\rho_t \in \mathcal{K}$ with probability 1.*
- (Local Lipschitzness) *f is L-Lipschitz continuous with respect to $\|.\|_2$ on an open set containing $\mathcal{K}$.*

Assumption B′. *Assumption A′ holds and f has C-Lipschitz-continuous gradients w.r.t. $\|.\|_2$ on $\mathcal{K}$.*

In Assumption A we added $\mu(x_t)a_t \in \mathcal{K}$ for clarity, but it is not necessary since $\mu(x_t)a_t \in \mathcal{K}$ with probability 1 is implied by $r_t \in \mathcal{K}$ with probability 1. The difference between Assumption A′ and Assumption A is to make sure that the updates ρ_t, and thus the iterates z_t belong to $\mathcal{K}$ and are in the domain of definition of f. Notice that in the special case of $\rho_t = r_t$, Assumption A′ reduces to Assumption A and, similarly, Assumption B reduces to Assumption B′. We use the term *smooth* as a synonym of Lipschitz-continuous gradients.

Analysis for (possibly) non-smooth objective functions We are going to present a single analysis that encompasses both the case where f is smooth (Assumption B of the main paper), and the case where f may not be smooth, which we briefly discussed in Section 5.3.3. In order for our analysis to be agnostic to the type of smoothing used and to also encompass the case where f is smooth, we propose the following assumption, where $(f_t)_{t\in\mathbb{N}}$ is a sequence of smooth approximations of f:

Assumption E. *Assumption A′ holds and $\exists(\beta_0, L, M_1, M_2) \in \mathbb{R}_+^4$ such that $(f_t)_{t\in\mathbb{N}}$ satisfy:*

1. *$\forall t \in \mathbb{N}, f_t : \mathbb{R}^D \to \mathbb{R} \cup \{\pm\infty\}$ is proper closed concave on $\mathbb{R}^D$,*
2. *$\forall t \in \mathbb{N}, f_t$ is differentiable on $\mathcal{K}$ with $\sup_{z\in\mathcal{K}} \|\nabla f_t(z)\|_2 \leq L$, and f_t is $\frac{\sqrt{t+1}}{\beta_0}$-smooth on $\mathcal{K}$,*
3. *$\forall t \in \mathbb{N}_*, \forall z \in \mathcal{K}, |f_t(z) - f_{t-1}(z)| \leq \frac{M_1}{t\sqrt{t}}$ and $|f_t(z) - f(z)| \leq \frac{M_2}{\sqrt{t}}$.*

Notice that any function f satisfying Assumption B with coefficient of smoothness C satisfies Assumption E with $\beta_0 = 1/C$, $M_1 = M_2 = 0$. Regarding non-smooth f, we discuss in more details in Appendix B.6 specific methods to perform this smoothing, including the Moreau envelope used in Section 5.3.3.

The generalization of the scalar regret takes into account both the approximation functions $(f_t)_{t\in\mathbb{N}}$ and the general update z_t:

$$R_T^{\text{gen}} = \sum_{t=1}^{T} \max_{a\in\mathcal{A}} \langle \nabla f_{t-1}(z_{t-1}) \,|\, \mu(x_t)a \rangle - \sum_{t=1}^{T} \langle \nabla f_{t-1}(z_{t-1}) \,|\, \rho_t \rangle + LT \|z_T - \hat{s}_T\|_2 \,. \tag{B.2}$$

The general regret bound then takes the following form, where we distinguish between smooth and non-smooth f. Recall that $\tilde{C} = CD_{\mathcal{K}}^2/2$.

Theorem 34. *Under Assumptions B′, using $\forall T \in \mathbb{N}, f_T = f$.*

For every $T \in \mathbb{N}$, every $z_0 \in \mathcal{K}$, every $\delta > 0$, Algorithm 6 satisfies, with probability at least $1 - \delta$:

$$R_T \leq \frac{R_T^{\text{gen}} + LD_{\mathcal{K}}\sqrt{2T \ln \frac{1}{\delta}} + \tilde{C} \ln(eT)}{T}$$

Theorem 35. *Under Assumptions E, for every $z_0 \in \mathcal{K}$, every $T \geq 1$ and every $\delta > 0$, Algorithm 6 satisfies, with probability at least $1-\delta$:*

$$R_T \leq \frac{R_T^{\text{gen}}}{T} + \frac{\frac{D_{\mathcal{K}}^2}{\beta_0} + 4M_1 + 2M_2 + LD_{\mathcal{K}}\sqrt{2\ln\frac{1}{\delta}}}{\sqrt{T}}$$

The proofs are given in Appendix B.5.

The worst-case regret of contextual bandits is $\Omega(\sqrt{T})$ [Bubeck and Cesa-Bianchi, 2012, Dani et al., 2008, Lattimore and Szepesvári, 2020], which gives a lower bound for the worst-case regret of CBCR in $\Omega(\frac{1}{\sqrt{T}})$. The dependencies on the problem parameters are all directly derived from the regret bounds R_T^{gen} of the underlying scalar bandit algorithm (LinUCB, SquareCB, etc.). Therefore we obtain CBCR algorithms that are near minimax optimal as soon as $R_T^{\text{gen}} \leq O(\sqrt{T})$. The residual terms $O(\frac{1}{\sqrt{T}})$ terms are tied to the use of Azuma's inequality (Lemma 36) and FW analysis (using Lipschitz and smoothness parameters), and the dependencies to these parameters match usual convergence guarantees in optimization [Jaggi, 2013, Clarkson, 2010, Lan, 2013]. As we rely on a worst-case analysis in deriving our reduction guarantees, it remains an open question whether problem-dependent optimal bounds could be recovered as well.

We make three remarks in order:

Remark 5 (Why we need a specific result for smooth f)**.** *The result for C-smooth f has a better dependency than the general result using $\beta_0 = 1/C$ ($\ln(eT)$ instead of $\sqrt{T}$), which makes a fundamental difference in practice if the smoothness coefficient is close to $\sqrt{T}$. This is why we keep the two results separate.*

Remark 6 (Comparison to the smoothing as used by Agrawal and Devanur [2014])**.** *Agrawal and Devanur [2014, Thm 5.4] present an analysis for non-smooth f where, at a high-level, they run the smooth algorithm using f_T instead of a sequence $(f_t)_{t\in\mathbb{N}}$, and then apply the convergence bound for smooth f. Our analysis has two advantages:*

1. *Anytime bounds: our approach does not require the horizon to be known in advance.*
2. *Better bound: they obtain a bound on $\sqrt{\ln T/T}$ by suitably choosing the smoothing parameter, whereas we obtain a bound of $1/\sqrt{T}$. In practice, it may not make a difference if $\frac{R_T^{\text{gen}}}{T}$ is itself in $\sqrt{\ln T}/T$, but the advantage of our approach is clear as far as the analysis of FW for (C)BCR is concerned.*

Remark 7 (About the confidence parameter δ' in $\mathfrak{A}(h_t, x_t, \delta')$ and $\mathfrak{U}(h_{t+1}, \delta')$)**.** *In practice, exploration/exploitation algorithms need a confidence parameter that defines the probability of their regret guarantee. For instance, in confidence-based approaches, it is the probability with which the confidence intervals are valid at every time step. In our case, it means that explicit upper bounds on R_T^{gen} are of the form $\overline{R}^{\text{gen}}(T, \delta')$ which hold with probability $1-\delta'$, where $\delta'$is the confidence parameter in $\mathfrak{A}(h_t, x_t, \delta')$. Using the union bound, we obtain bounds of the form $R_T \leq \overline{R}^{\text{gen}}(T, \delta')/T + O\big(\sqrt{\frac{\ln(1/\delta)}{T}}\big)$ that are valid with probability $1-\delta-\delta'$.*

Note the difference in the roles of δ and δ': δ is not a parameter of the algorithm, it is only here to account for the randomization over contexts.

B.5 Proofs for Section 5.3 and Appendix B.4

This section contains the proofs for the results of Section 5.3. All the proofs are made for the more general framework described in Appendix B.4. The framework of the paper can be recovered as the special case $\forall t \in \mathbb{N}, \rho_t = r_t$ and $z_t = \hat{s}_t$.

Proof of Lemma 12. Lemma 12 is the special case of Lemma 36 when f is smooth. Note that every f satisfying Assumption A satisfies the assumptions of Lemma 36. □

Proof of Theorem 13. Thm. 13 is a special case of Theorem 34 of Appendix B.4, using $\forall t \in \mathbb{N}, \rho_t = r_t$ and $z_t = \hat{s}_t$. The proof of Theorem 34 is given in Section B.5.1. □

Lemma 36. *Assume that $\forall T, f_T$ is differentiable on $\mathcal{K}$ with $\forall z \in \mathcal{K}, \|\nabla f_T(z)\|_2 \leq L$. Then, for every $z \in \mathcal{K}$, we have:*

$$\mathbb{E}_{x\sim P}\big[\max_{a\in\mathcal{A}}\langle \nabla f_{t-1}(z) \,|\, \mu(x)a\rangle\big] = \max_{\overline{\pi}:\mathcal{X}\to\overline{\mathcal{A}}} \mathbb{E}_{x\sim P}\big[\langle \nabla f_{t-1}(z) \,|\, \mu(x)\overline{\pi}(x)\rangle\big] = \max_{s\in\mathcal{S}}\langle \nabla f_{t-1}(z_{t-1}) \,|\, s\rangle \tag{B.3}$$

Assume furthermore that z_t is a function of contexts, actions and rewards up to time t. Let $a_t^* \in \operatorname{argmax}_{a\in\mathcal{A}}\langle \nabla f_{t-1}(z_{t-1}) \,|\, \mu(x_t)a\rangle$. For all $\delta \in (0,1]$, with probability at least $1-\delta$, we have:

$$\sum_{t=1}^{T} \max_{s\in\mathcal{S}}\langle \nabla f_{t-1}(z_{t-1}) \,|\, s - \mu(x_t)a_t^*\rangle \leq LD_{\mathcal{K}}\sqrt{2T\ln\frac{1}{\delta}} \tag{B.4}$$

Proof. Let $z \in \mathcal{K}$. We first prove (B.3). The first equality in (B.3) comes from the maximization over functions over the simplex with a linear objective: define

$$\overline{\pi}_t^* : \mathcal{X} \mapsto \overline{\mathcal{A}} \qquad \text{such that } \overline{\pi}_t^*(x) \in \operatorname*{argmax}_{a\in\mathcal{A}}\langle \nabla f_{t-1}(z_{t-1}) \,|\, \mu(x)a\rangle,$$

using some arbitrary tie-breaking rule when the argmax is not unique. We have, for every policy $\overline{\pi}$:

$$\begin{aligned} & \mathbb{E}_{x\sim P}\big[\langle \nabla f_{t-1}(z) \,|\, \mu(x)\overline{\pi}(x)\rangle\big] \leq \mathbb{E}_{x\sim P}\big[\max_{a\in\mathcal{A}}\langle \nabla f_{t-1}(z) \,|\, \mu(x)a\rangle\big] \\ \implies \quad & \max_{\overline{\pi}:\mathcal{X}\to\overline{\mathcal{A}}} \mathbb{E}_{x\sim P}\big[\langle \nabla f_{t-1}(z) \,|\, \mu(x)\overline{\pi}(x)\rangle\big] \leq \mathbb{E}_{x\sim P}\big[\langle \nabla f_{t-1}(z) \,|\, \mu(x)\overline{\pi}_t^*(x)\rangle\big]. \end{aligned}$$

On the other hand, it is clear that

$$\mathbb{E}_{x\sim P}\big[\langle \nabla f_{t-1}(z) \,|\, \mu(x)\overline{\pi}_t^*(x)\rangle\big] \leq \max_{\overline{\pi}:\mathcal{X}\to\overline{\mathcal{A}}} \mathbb{E}_{x\sim P}\big[\langle \nabla f_{t-1}(z) \,|\, \mu(x)\overline{\pi}(x)\rangle\big],$$

and we get the first equality of (B.3).

The second equality in (B.3) holds by the definition of $\mathcal{S}$ since for every policy $\overline{\pi}$, we have

$$\mathbb{E}_{x\sim P}\big[\langle \nabla f_{t-1}(z) \,|\, \mu(x)\overline{\pi}(x)\rangle\big] = \langle \nabla f_{t-1}(z) \,|\, \mathbb{E}_{x\sim P}\big[\mu(x)\overline{\pi}(x)\big]\rangle.$$

We now prove (B.4). Let $\big(\mathbb{E}_t[.]\big)_{t\geq 1}$ be the conditional expectations with respect to the filtration $\widetilde{\mathfrak{F}} = (\widetilde{\mathfrak{F}}_t)_{t\geq 1}$ where $\mathfrak{F}_t$ is the σ-algebra generated by $(x'_t, a'_t, r'_t)_{t'\in[\![t-1]\!]}$, i.e., contexts, actions and rewards up to time $t-1$, so that we have:

$$\mathbb{E}_t\big[\langle \nabla f_{t-1}(z_{t-1}) \,|\, \mu(x_t)\overline{\pi}_t^*(x_t)\rangle\big] = \mathbb{E}_{x\sim P}\big[\langle \nabla f_{t-1}(z_{t-1}) \,|\, \mu(x)\overline{\pi}_t^*(x)\rangle\big].$$

Using (B.3) gives $\mathbb{E}_t\big[\langle \nabla f_{t-1}(z_{t-1}) \,|\, \mu(x_t)\overline{\pi}_t^*(x_t)\rangle\big] = \max_{s\in\mathcal{S}}\langle \nabla f_{t-1}(z_{t-1}) \,|\, s\rangle$, from which we obtain

$$\begin{aligned} & \max_{s\in\mathcal{S}}\langle \nabla f_{t-1}(z_{t-1}) \,|\, s - \mu(x_t)a_t^*\rangle \\ & \qquad = \mathbb{E}_t\big[\langle \nabla f_{t-1}(z_{t-1}) \,|\, \mu(x_t)\overline{\pi}_t^*(x_t)\rangle\big] - \langle \nabla f_{t-1}(z_{t-1}) \,|\, \mu(x_t)\overline{\pi}_t^*(x_t)\rangle \end{aligned}$$

$X_T = \sum_{t=1}^{T} \max_{s\in\mathcal{S}}\langle \nabla f_{t-1}(z_{t-1}) \,|\, s - \mu(x_t)a_t^*\rangle$ thus defines a martingale adapted to $\mathfrak{F}$, and,

using $X_0 = 0$, we have, for all $_t$:

$$|X_t - X_{t-1}| \leq L \sup_{\substack{s\in\mathcal{S}\\ x\in\mathcal{X}\\ a\in\mathcal{A}}} \|s - \mu(x)a\|_2 \leq L \sup_{z,z'\in\mathcal{K}} \|z - z'\|_2 \leq LD_{\mathcal{K}}.$$

The results then follows from Azuma's inequality. □

The next lemma is the main technical tool of the paper. The proof is not technically difficult given the previous result, using the telescoping sum approach of the proof of Lemma 12 of Berthet and Perchet [2017] and organizing the residual terms.

Lemma 37. *Under Assumption E, denote* $\forall t \in \mathbb{N}, f_t^* = \max_{s\in\mathcal{S}} f_t(s)$, *and* $\tilde{R}_t(z) = f_t^* - f_t(z)$.
Let $\overline{C}(T)$, $\overline{F}^*(T)$ *in* $\mathbb{R}\cup\{+\infty\}$ *such that,* $\forall T \in \mathbb{N}_*$, *we have:*

$$\sum_{t=1}^{T} \frac{D_{\mathcal{K}}^2}{2}\frac{C_{t-1}}{t} \leq \overline{C}(T), \qquad \sum_{t=1}^{T} t\big(\tilde{R}_t(z_t) - \tilde{R}_{t-1}(z_t)\big) \leq \overline{F}^*(T)$$

And let $\overline{B}(T) = \overline{C}(T) + \overline{F}^*(T)$. *Then, for all* $z_0 \in \mathcal{K}$, $\forall T, \forall\delta > 0, \forall\delta' > 0$, *Algorithm 6 satisfies, with probability at least* $1-\delta$*:*

$$f_T^* - f_T(\hat{s}_T) \leq \frac{\overline{B}(T) + R_T^{\text{gen}} + LD_{\mathcal{K}}\sqrt{2T\ln\frac{1}{\delta}}}{T}$$

Proof. We start with the standard ascent lemma using bounded curvature on $\mathcal{K}$ [Bottou et al., 2018, Inequality 4.3], denoting $\tilde{C}_{_T} = \frac{D_{\mathcal{K}}^2}{2}C_T$:

$$f_{t-1}(z_t) \geq f_{t-1}(z_{t-1}) + \frac{1}{t}\langle\nabla f_{t-1}(z_{t-1}) \,|\, \rho_t - z_{t-1}\rangle - \frac{\tilde{C}_{_{t-1}}}{t^2}$$
$$f_{t-1}^* - f_{t-1}(z_t) \leq f_{t-1}^* - f_{t-1}(z_{t-1}) - \frac{1}{t}\langle\nabla f_{t-1}(z_{t-1}) \,|\, \rho_t - z_{t-1}\rangle + \frac{\tilde{C}_{_{t-1}}}{t^2}$$

Let us denote by $g_t = \nabla f_{t-1}(z_{t-1})$ and let $a_t^* \in \operatorname{argmax}_{a\in\mathcal{A}}\langle g_t \,|\, \mu(x_t)a\rangle$. We first decompose the middle term:

$$\begin{aligned}\langle g_t \,|\, \rho_t - z_{t-1}\rangle &= \max_{s\in\mathcal{S}}\langle g_t \,|\, s - z_{t-1}\rangle - \max_{s\in\mathcal{S}}\langle g_t \,|\, s - \mu(x_t)a_t^*\rangle - \langle g_t \,|\, \mu(x_t)a_t^* - \rho_t\rangle \\ &\geq f_{t-1}^* - f_{t-1}(z_{t-1}) - \underbrace{\max_{s\in\mathcal{S}}\langle g_t \,|\, s - \mu(x_t)a_t^*\rangle}_{\alpha_t} - \underbrace{\langle g_t \,|\, \mu(x_t)a_t^* - \rho_t\rangle}_{\rho_t} \quad \text{(by (B.5) below)}\end{aligned}$$

Where the last inequality uses the concavity of f_t: for all $s_{t-1}^* \in \operatorname{argmax}_{s\in\mathcal{S}} f_{t-1}(s)$, we have:

$$f_{t-1}^* - f_{t-1}(z_{t-1}) \leq \langle\nabla f_{t-1}(z_{t-1}) \,|\, s_{t-1}^* - z_{t-1}\rangle \leq \max_{s\in\mathcal{S}}\langle\nabla f_{t-1}(z_{t-1}) \,|\, s - z_{t-1}\rangle \tag{B.5}$$

and thus we get

$$\begin{aligned} & f_{t-1}^* - f_{t-1}(z_t) \leq \big(f_{t-1}^* - f_{t-1}(z_{t-1})\big)(1-\frac{1}{t}) + \frac{1}{t}(\alpha_t + \rho_t) + \frac{\tilde{C}_{_{t-1}}}{t^2} \\ \implies \quad & t\tilde{R}_t(z_t) \leq (t-1)\tilde{R}_{t-1}(z_{t-1}) + \alpha_t + \rho_t + \frac{\tilde{C}_{_{t-1}}}{t} + t\big(\tilde{R}_t(z_t) - \tilde{R}_{t-1}(z_t)\big) \\ \implies \quad & T\tilde{R}_T(z_T) \leq \sum_{t=1}^{T}\alpha_t + \sum_{t=1}^{T}\rho_t + \sum_{t=1}^{T} t\big(\tilde{R}_t(z_t) - \tilde{R}_{t-1}(z_t)\big) + \sum_{t=1}^{T}\frac{\tilde{C}_{_{t-1}}}{t}\end{aligned}$$

Using the Lipschitz property for f_T, we finally obtain

$$T\tilde{R}_T(\hat{s}_T) \leq \underbrace{\sum_{t=1}^{T}\alpha_t + \sum_{t=1}^{T}\rho_t + TL\left\|z_T - \hat{s}_T\right\|_2}_{\substack{\leq LD_{\mathcal{K}}\sqrt{2T\ln(1/\delta)}+R_T^{\text{gen}} \\ \text{w.p. } \geq 1-\delta \text{ by (B.2) and Lemma 36.}}} + \underbrace{\sum_{t=1}^{T} t\big(\tilde{R}_t(z_t) - \tilde{R}_{t-1}(z_t)\big)}_{\leq \overline{F}^*(T)} + \underbrace{\sum_{t=1}^{T}\frac{\tilde{C}_{t-1}}{t}}_{\leq \overline{C}(T)}$$

Which is the desired result. □

B.5.1 proofs of the main results

We now prove the results of Appendix B.4.

Proof of Theorem 34. First, notice that since f differentiable on $\mathcal{K}$ (since it is smooth) and since both z_T and $\frac{1}{T}\sum_{t=1}^{T}\mu(x_t)a_t$ are in $\mathcal{K}$, using $\forall t, f_t = f$, we have $R_T = f^* - f(\hat{s}_T) = f_T^* - f_T(\hat{s}_T)$. Using the notation of Lemma 37, we then have $\overline{C}(T) = 0$ and $D(T) = 0$. Also:

$$\sum_{t=1}^{T}\frac{D_{\mathcal{K}}^2}{2}\frac{C_t}{t} = \sum_{t=1}^{T}\frac{\tilde{C}}{t} \leq \tilde{C}(\ln(t)+1)$$

The result then follows from Lemma 37. □

Proof of Theorem 35. Using the notation of Lemma 37, we specify $\overline{C}(T)$, $\overline{F}^*(T)$ in turn.

$$\sum_{t=1}^{T}\frac{D_{\mathcal{K}}^2}{2}\frac{C_{t-1}}{t} = \sum_{t=1}^{T}\frac{D_{\mathcal{K}}^2}{2\beta_0\sqrt{t}} \leq \frac{D_{\mathcal{K}}^2}{\beta_0}\sqrt{T}.$$

For $\overline{F}^*(T)$, we decompose $\tilde{R}_t(z_t) - \tilde{R}_{t-1}(z_t)$ into two terms:

$$\tilde{R}_t(z_t) - \tilde{R}_{t-1}(z_t) = f_t^* - f_{t-1}^* + f_{t-1}(z_t) - f_t(z_t) \leq \frac{2M_1}{t\sqrt{t}}$$

Using $\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \leq 2\sqrt{T}$, we obtain $\overline{F}^*(T) \leq 2M_1\sum_{t=1}^{T}\frac{t}{t\sqrt{t}} \leq 4M_1\sqrt{T}$. Lemma 37 gives

$$f_T^* - f_T(\hat{s}_T) \leq R_T^{\text{gen}} + \frac{\frac{D_{\mathcal{K}}^2}{\beta_0} + 4M_1 + LD_{\mathcal{K}}\sqrt{2\ln(\delta^{-1}/2)}}{\sqrt{T}} \tag{B.6}$$

To finish the proof, notice that:

$$\left|f^* - f(\hat{s}_T) - \big(f_T^* - f_T(\hat{s}_T)\big)\right| \leq 2\sup_{z'\in\mathcal{K}}|f_T(z') - f(z')| \leq \frac{2M_2}{\sqrt{T}}. \tag{B.7}$$

The result follows from (B.6) and (B.7) using:

$$R_T = f^* - f(\hat{s}_T) \leq f_T^* - f_T(\hat{s}_T) + \frac{2M_2}{\sqrt{T}}.$$

□

B.6 Smooth approximations of non-smooth functions

We discuss here in more details two specific smoothing techniques: the Moreau envelope, also called Moreau-Yosida regularization in Section B.6.1, then randomized smoothing in Section B.6.2. As in

Appendices B.4 and B.5, we focus on the general framework described in Algorithm 6.

Proof of Theorem 14. Usinh Theorem 35 above and Lemma 39 below gives the result since

$$\frac{D_{\mathcal{K}}^2}{\beta_0} + 4M_1 + 2M_2 = \frac{D_{\mathcal{K}}^2}{\beta_0} + 3L^2\beta_0 = LD_{\mathcal{K}}\big(\frac{D_{\mathcal{K}}}{L\beta_0} + 3\frac{L\beta_0}{D_{\mathcal{K}}}\big).$$

□

B.6.1 Smoothing with the Moreau envelope

For functions that are non-smooth, we propose first a smoothing technique based on the Moreau envelope, following the approach described by Lan [2013]. Let $f : \mathbb{R}^D \to \mathbb{R} \cup \{\pm\infty\}$ be a closed proper concave function. The Moreau envelope (or Moreau-Yosida regularization) of f with parameter β_T [Rockafellar and Wets, 2009, Def. 1.22] is defined as

$$\tilde{f}_\beta(z) = \max_{y\in\mathbb{R}^D} \Big(f(y) - \frac{1}{2\beta}\|y - z\|_2^2\Big).$$

For $\beta > 0$, let the proximal operator $\text{prox}_\beta = \text{argmax}_{y\in\mathbb{R}^D} \tilde{f}_\beta(y)$. The basic properties of the Moreau envelope [Rockafellar and Wets, 2009, Th. 2.26] are that if $f : \mathbb{R}^D \to \mathbb{R} \cup \{\pm\infty\}$ is an upper semicontinuous, proper concave function then $\tilde{f}_\beta$ is concave, finite everywhere, continuously differentiable with $\frac{1}{\beta}$-Lipschitz gradients. We also have that the proximal operator prox_β is well-defined (the argmax is attained in a single point) and we have

$$\nabla\tilde{f}_\beta(z) = \frac{1}{\beta}\big(z - \text{prox}_\beta(z)\big).$$

It is immediate to prove the following inequalities for every $z \in \mathbb{R}^n$ and every $\beta > 0$:

$$f(z) \le \tilde{f}_\beta(z) \le f(\text{prox}_\beta(z)).$$

The following properties of the Moreau envelope (See [Yurtsever et al., 2018, Appendix A.1] and [Thekumparampil et al., 2020b, Lemma 1]) are key to the main results:

Lemma 38. *Let $\beta > 0$, $f : \mathbb{R}^D \to \mathbb{R} \cup \{\pm\infty\}$ be a proper closed concave function, and $\mathcal{Z} \subseteq \mathbb{R}^D$ be a convex set such that f is locally L-Lipschitz-continuous on $\mathcal{Z}$. Then:*

- *$\forall z \in \mathcal{Z}$ such that $\text{prox}_\beta(z) \in \mathcal{Z}$, we have $\big\|z - \text{prox}_\beta(z)\big\| \le L\beta$ and:*

$$\tilde{f}_\beta(z) - \frac{L^2\beta}{2} \le f(z) \le \tilde{f}_\beta(z).$$

- *$\forall z \in \mathcal{Z}$ such that $\text{prox}_\beta(z) \in \mathcal{Z}$, $\forall\beta > 0$ and $\beta' > 0$, we have:*

$$\tilde{f}_\beta \le \tilde{f}_{\beta'} + \frac{1}{2}\Big(\frac{1}{\beta'} - \frac{1}{\beta}\Big)\|[\|\,]z - \text{prox}_\beta(z)\|_2^2 \le \frac{L^2\beta}{2}\Big(\frac{\beta}{\beta'} - 1\Big)$$

We reformulate the lemma above in the language of Appendix B.4:

Lemma 39. *Under Assumption A, assuming furthermore that f is L-Lipschitz on $\mathbb{R}^D$.*

Let $f_t = \tilde{f}_{\beta_t}$ with $\beta_t = \frac{\beta_0}{\sqrt{t+1}}$. Then f and $(f_t)_{t\in\mathbb{N}}$ satisfy Assumption E with the corresponding values of β_0 and L, $M_2 = \frac{L^2\beta_0}{2}$ and $M_1 = \frac{L^2\beta_0}{2}$.

Proof. By Lemma 38, f_t is L-Lipschitz on $\mathbb{R}^D$ for every t, and we have $M_2 = \frac{L^2\beta_0}{2}$. Moreover, Lemma 38 also gives $0 \le f_{t-1}(z) - f_t(z) \le \frac{L^2\beta_0}{2t}(\sqrt{t+1} - \sqrt{t}) \le \frac{L^2\beta_0}{2t\sqrt{t}}$. and thus $M_1 = \frac{L^2\beta_0}{2}$. □

B.6.2 Randomized smoothing

We now describe the randomized smoothing technique [Lan, 2013, Nesterov and Spokoiny, 2017, Duchi et al., 2012, Yousefian et al., 2012], which consists in convolving f with a probability density function Λ. Following Lan [2013] who combines Frank-Wolfe with randomized smoothing for nonsmooth optimization, we present our results with Λ as the random uniform distribution in the ℓ_2-ball $\{z \in \mathbb{R}^D : \|z\|_2 \leq 1\}$. Let $\beta > 0$ and ξ a random variable with density Λ. Then the randomized smoothing approximation of f is defined as:

$$f_\beta(z) := \mathbb{E}_\Lambda[f(x + \beta\xi)] = \int_{\mathbb{R}^D} f(x + \beta y)\Lambda(y)dy. \tag{B.8}$$

Following [Lan, 2013, Duchi et al., 2012], we abuse notation and take the "gradient" of f inside integrals and expectation below, because f is almost-everywhere differentiable since it is concave. We restate the following well-known properties of randomized smoothing (see e.g., [Yousefian et al., 2012, Lemma 8]):

Lemma 40. *Let $\beta > 0$ and f_β be defined as in Eq. (B.8).*

- $\forall z \in \mathcal{K}$, $f(z) \leq f_\beta(z) \leq f(z) + L\beta$.
- *f_β is L-Lipschitz continuous over $\mathcal{K}$.*
- *f_β is continuously differentiable and its gradient is $\frac{L\sqrt{D}}{\beta}$-Lipschitz continuous.*
- $\forall z \in \mathcal{K}, \nabla f_\beta(z) = \mathbb{E}[\nabla f(z + \beta\xi)]$.

We obtain the following results, stated in the language of Theorem 35 of Appendix B.4.

Lemma 41. *Under Assumption A, assuming furthermore that f is L-Lipschitz on $\mathbb{R}^D$.*

For $t \geq 1$, let $f_t = f_{\beta_t}$ with $\beta_t = \frac{D^{\frac{1}{4}} D_\mathcal{K}}{\sqrt{t+1}}$, and let $\beta_0 = \frac{\sqrt{D} D_\mathcal{K}}{L}$.

Then f and $(f_t)_{t\in\mathbb{N}}$ satisfy Assumption E with the corresponding values of β_0 and L, $M_2 = LD^{\frac{1}{4}} D_\mathcal{K}$ and $M_1 = 2LD^{\frac{1}{4}} D_\mathcal{K}$.

Proof. By Lemma 40, f_t is L-Lipschitz on $\mathbb{R}^D$ for every t, so that f_t has L-bounded gradient. Moreover, with this definition of β_0, f_t is $\frac{\sqrt{t+1}}{\beta_0}$-smooth.

We have $M_2 = LD^{\frac{1}{4}} D_\mathcal{K}$ because:

$$|f_t(z) - f(z)| = |\mathbb{E}[f(z + \beta_t\xi)] - \mathbb{E}[f(z)]| \leq \mathbb{E}[|f(z + \beta_t\xi) - f(z)|] \leq \mathbb{E}[\|L\beta_t\xi\|_2] \leq \frac{LD^{\frac{1}{4}} D_\mathcal{K}}{\sqrt{t}}$$

We also have $M_1 = 2LD^{\frac{1}{4}} D_\mathcal{K}$ because:

$$\begin{aligned} |f_{t-1} - f_t| \leq \mathbb{E}[|f(x + \beta_{t-1}\xi) - f(x + \beta_t\xi)|] &\leq L\,|\beta_{t-1} - \beta_t|\,\mathbb{E}[\|\xi\|_2] \\ &= LD^{\frac{1}{4}} D_\mathcal{K}(\frac{1}{\sqrt{t}} - \frac{1}{\sqrt{t+1}}) \leq \frac{2LD^{\frac{1}{4}} D_\mathcal{K}}{t^{\frac{3}{2}}}. \end{aligned}$$

□

B.7 FW-LinUCB: upper-confidence bounds for linear bandits with K arms

In this section, we have:

Algorithm 7: FW-linUCB: linear CBCR with K arms.

input : $\delta' > 0, \lambda > 0, \hat{s}_0 \in \mathcal{K}$ $V_0 = \lambda \boldsymbol{I}_{dD}, y_0 = \mathbf{0}_{dD}, \hat{\theta}_0 = \mathbf{0}_{dD}$

for $t = 1, \dots$ **do**

  Observe context $x_t \sim P$, $x_t \in \mathbb{R}^{d \times K}$

  $g_t \leftarrow \nabla f_{t-1}(\hat{s}_{t-1}), \tilde{x}_t \leftarrow [g_{t,0} x_t; \dots; g_{t,D} x_t]$

  $\forall i \in [\![K]\!], \hat{u}_{t,i} \leftarrow \hat{\theta}_{t-1}^\intercal \tilde{x}_{t,i} + \alpha_t\left(\frac{\delta'}{2}\right) \|\tilde{x}_{t,i}\|_{V_{t-1}^{-1}}$ // see (B.10) and (B.11) for def. of $\|.\|_{V_{t-1}^{-1}}$ and α_t.

  $a_t \leftarrow \operatorname{argmax}_{a \in \mathcal{A}} \hat{u}_t a$

  Observe reward r_t, let $\tilde{r}_t = g_t^\intercal r_t$

  Update $\hat{s}_t \leftarrow \hat{s}_{t-1} + \frac{1}{t}(r_t - \hat{s}_{t-1})$

  $V_t \leftarrow V_{t-1} + (\tilde{x}_t a_t)(\tilde{x}_t a_t)^\intercal$, $y_t \leftarrow y_{t-1} + \tilde{r}_t \tilde{x}_t a_t$ and $\hat{\theta}_t \leftarrow V_t^{-1} y_t$ // regression

end

- a finite action space $\mathcal{A}$ which is the canonical basis of $\mathbb{R}^K$, i.e., we focus on the multi-armed bandit setting
- $\mathcal{X} \subseteq \mathbb{R}^{d \times K}$, where d is the dimension of the feature space. Given $x \in \mathcal{X}$, the feature representation of arm $a \in \mathcal{A}$ is given by the matrix-vector product xa,
- Given a matrix $\theta \in \mathbb{R}^{D \times d}$, we denote by $\|\theta\|_{\mathrm{F}}$ the frobenius norm of θ, i.e., $\|\theta\|_{\mathrm{F}} = \|\mathrm{flatten}(\theta)\|_2$.

In addition, we make here the following linear assumption on the rewards:

Assumption F. *There is $\theta \in \mathbb{R}^{D \times d}$ such that $\|\theta\|_{\mathrm{F}} \leq D_\theta$ such that $\forall x \in \mathcal{X}, \mu(x)a = \theta x a$. Moreover, there is $D_{\mathcal{X}} > 0$ such that $\sup_{\substack{x \in \mathcal{X} \\ a \in \mathcal{A}}} \|xa\|_2 \leq D_{\mathcal{X}}$.*

We perform the analysis under Assumption E, which is the more general we have. In particular, we assume that we have access to a sequence $(f_t)_{t \in [\![T]\!]}$ of smooth approximations of f. We focus on the special case of Algorithm 6 that is described in the main paper, i.e., where $\rho_T = r_t$.

The algorithm. As hinted in Section 5.3.2, FW-LinUCB applies the LinUCB algorithm [Abbasi-Yadkori et al., 2011], designed for scalar-reward contextual bandits with adversarial contexts and stochastic rewards, to the following extended rewards and contexts, where we use $[.;.]$ to denote the vertical concatenation of matrices and $g_t = \nabla f_{t-1}(\hat{s}_{t-1})$:

- $\tilde{x}_t \in \mathbb{R}^{Dd \times K}$ is the extended context with entries $\tilde{x}_t = [g_{t,0} x_t; \dots; g_{t,D} x_t] \in \mathbb{R}^{Dd \times K}$, so that the feature vector of action a at time t is $\tilde{x}_t a$;
- $\tilde{r}_t = g_t^\intercal r_t$ is the scalar observed reward,
- $\tilde{\theta} = \mathrm{flatten}(\theta) \in \mathbb{R}^{dD}$ is the ground-truth parameter vector and $\tilde{\mu}(x) = \tilde{\theta}^\intercal \tilde{x}_t$ is the average reward function.

Notice that under assumption A and F, denoting

$$\widetilde{\mathcal{X}} = \left\{[g_{t,0} x_t; \dots; g_{t,D} x_t] : \|g\|_2 \leq L, x \in \mathcal{X}\right\} \qquad \text{and } D_{\widetilde{\mathcal{X}}} = \max_{\substack{\tilde{x} \in \widetilde{\mathcal{X}} \\ a \in \mathcal{A}}} \|\tilde{x} a\|_2 , \tag{B.9}$$

we have $\forall t, \tilde{x}_t \in \widetilde{\mathcal{X}}$ with probability 1 and $D_{\widetilde{\mathcal{X}}} \leq L D_{\mathcal{X}}$. Moreover, $|\tilde{r}_t - \tilde{\mu}(x_t) a_t| \leq L D_{\mathcal{K}}$, which implies in particular that for every $t \in [\![T]\!]$, $\tilde{r}_t$ is $LD_{\mathcal{K}}/2$-subgaussian.

Given this notation, the FW-LinUCB algorithm is LinUCB applied to the scalar-reward bandit problem above. The algorithm is summarized in Algorithm 7 for completeness, where λ is the regularization parameter of the ridge regression, $\hat{\theta}_t$ is the current regression parameters, the matrix V_t and the vector y_t are incremental computations of the relevant matrices to compute $\hat{\theta}_t$. The

crucial part of the algorithm is Line 3 which defines an upper confidence bound on $\tilde{\mu}(x_t)a$, denoted by $\hat{u}_t \in \mathbb{R}^K$ and defined by:

$$\forall i \in [\![K]\!], \hat{u}_{t,i} = \hat{\theta}_{t-1}^\intercal \tilde{x}_{t,i} + \alpha_t(\delta'/2) \left\|\tilde{x}_{t,i}\right\|_{V_{t-1}^{-1}} \qquad \text{where } \left\|\tilde{x}_{t,i}\right\|_{V_{t-1}^{-1}} = \sqrt{\tilde{x}_{t,i}^\intercal V_{t-1}^{-1} \tilde{x}_{t,i}}, \tag{B.10}$$

and α_t is defined according to Theorem 2 of Abbasi-Yadkori et al. [2011]:

$$\alpha_t(\delta') = \frac{LD_{\mathcal{K}}}{2}\sqrt{dD \ln\left(\frac{1 + TD_{\tilde{\mathcal{X}}}^2/\lambda}{\delta'}\right)} + \sqrt{\lambda} D_\theta. \tag{B.11}$$

Under Assumption E, we have with probability $\geq 1 - \delta'/2$: $\forall t \in \mathbb{N}_*, \hat{u}_t a \geq \tilde{\mu}(x_t)a$ [Abbasi-Yadkori et al., 2011, Theorem 2].

The result. Let $\tilde{d} = dD$. The regret bound of LinUCB [Abbasi-Yadkori et al., 2011, Theorem 3] and Azuma inequality give:

Theorem 42. *Under Assumption E, for every $T \in \mathbb{N}_*$, for every $\delta' > 0$, Algorithm 7 satisfies, with probability at least $1 - \delta'$:*

$$\begin{aligned} R_T^{\text{scal}} \leq & 4\sqrt{T\tilde{d}\log(1 + TD_{\tilde{\mathcal{X}}}/\tilde{d})}\left(\sqrt{\lambda} D_\theta + \frac{LD_{\mathcal{K}}}{2}\sqrt{2\ln(2/\delta') + \tilde{d}\ln\left(1 + TD_{\tilde{\mathcal{X}}}/(\lambda\tilde{d})\right)}\right) \\ & + LD_{\mathcal{K}}\sqrt{2\ln(2/\delta')}. \end{aligned}$$

Proof. Recall that as noted in (5.3) We decompose the scalar regret R_T^{scal} into a pseudo regret and a residual term:

$$R_T^{\text{scal}} = \underbrace{\sum_{t=1}^T \max_{a \in \mathcal{A}} \tilde{\mu}(\tilde{x}_t)^\intercal a - \sum_{t=1}^T \tilde{\mu}(\tilde{x}_t)^\intercal a_t}_{\text{pseudo-regret}} + \sum_{t=1}^T \underbrace{\left(\tilde{\mu}(\tilde{x}_t)^\intercal a_t - \tilde{r}_t\right)}_{X_t}$$

The pseudo-regret term is bounded using Theorem 3 by Abbasi-Yadkori et al. [2011]. The result applies as-is, except that they assume rewards $|\theta^\intercal \tilde{x}_t| \leq 1$, which is not the case here. The bound is still valid without changes, as in our case we have $|\max_{a \in \mathcal{A}} \tilde{\mu}(\tilde{x}_t)^\intercal a - \tilde{\mu}(\tilde{x}_t)^\intercal a_t| \leq LD_{\mathcal{K}}$. The steps in the proof where they use the assumption $|\theta^\intercal \tilde{x}_t| \leq 1$ is below Equation 7 [Abbasi-Yadkori et al., 2011, Appendix C], which in our notation and our assumption can be written as:

$$\begin{aligned} \max_{a \in \mathcal{A}} \tilde{\mu}(\tilde{x}_t)^\intercal a - \tilde{\mu}(\tilde{x}_t)^\intercal a_t &\leq \min\left(2\alpha_t(\delta'/2)\left\|\tilde{x}_t a_t\right\|_{V_{t-1}^{-1}}, LD_{\mathcal{K}}\right) \\ &\leq 2\alpha_t(\delta'/2)\min(\left\|\tilde{x}_t a_t\right\|_{V_{t-1}^{-1}}, 1) \end{aligned}$$

where the first inequality comes from Abbasi-Yadkori et al. [2011] and the second one is true in our case because $2\alpha_t(\delta') \geq LD_{\mathcal{K}}$. From here on, the proof of Abbasi-Yadkori et al. [2011]'s regret bound follows the same as the original result.[4] Theorem 3 from Abbasi-Yadkori et al. [2011] gives us the first term of the regret bound of the theorem, which is true with probability at least $1 - \delta'/2$ in our case because we use $\alpha_t(\delta'/2)$.

For the rightmost term, let $\overline{\mathfrak{F}} = \left(\overline{\mathfrak{F}}_t\right)_{t \in \mathbb{N}_*}$ be the filtration where $\overline{\mathfrak{F}}_t$ is the σ-algebra generated by $(x_1, a_1, r_1, \ldots, x_{t-1}, a_{t-1}, r_{t-1}, x_t, a_t)$. Then $(X_t)_{t \in \mathbb{N}_*}$ is a martingale difference sequence adapted to $\overline{\mathfrak{F}}$ with $|X_t| \leq LD_{\mathcal{K}}$. By Azuma's inequality, we have $\sum_{t=1}^T X_t \leq LD_{\mathcal{K}}\sqrt{2T\overline{k}\ln\frac{2}{\delta'}}$ with probability

[4] In short, they have different bounds, one involving the varance of r_t and the other one involving average rewards $\tilde{\mu}(\tilde{x})$. We assume rewards r_T are uniformly bounded in $\mathcal{K}$, so we do not have to deal with two different quantities in our bounds and have $LD_{\mathcal{K}}$ everywhere.

Algorithm 8: FW-SquareCB: contextual bandits with concave rewards and regression oracles

input : initial point $\hat{s}_0 \in \mathcal{K}$, exploration parameters $(\gamma_t)_{t\in\mathbb{N}}$. $\mathcal{A}$ is the canonical basis of $\mathbb{R}^K$.

for $t = 1 \dots$ **do**
  Observe $x_t \sim P$
  Compute $\hat{\mu}_t(x_t)$ using RegSq // see (B.12)
  Let $g_t = \nabla f_{t-1}(\hat{s}_{t-1})$ and $\underline{\hat{\mu}}_t = g_t^\intercal \hat{\mu}_t(x_t) \in \mathbb{R}^K$
  Let $\underline{a}_t \in \operatorname*{argmax}_{a\in\mathcal{A}} \underline{\hat{\mu}}_t^\intercal a$ and $\underline{\hat{\mu}}_t^* = \underline{\hat{\mu}}_t \underline{a}_t$ // use arbitrary tie breaking rule
  Let $\forall a \in \mathcal{A}, \mathfrak{A}_t(a) = \begin{cases} \frac{1}{K+\gamma_t\left(\underline{\hat{\mu}}_t^* - \underline{\hat{\mu}}_t^\intercal a\right)} & \text{if } a \neq \underline{a}_t \\ 1 - \sum_{\substack{a\in\mathcal{A} \\ a\neq\underline{a}_t}} \mathfrak{A}_t(a) & \text{if } a = \underline{a}_t \end{cases}$ // Exploration/exploitation step
  Draw $a_t \sim \mathfrak{A}_t$ // Action taken at time step t
  Observe reward r_t and update $\hat{s}_t = \hat{s}_{t-1} + \frac{1}{t}(r_t - \hat{s}_{t-1})$
end

$1 - \delta'/2$. The final result holds using a union bound. □

Bound of Table 5.1. The bound is obtained by keeping the main dependencies in $T, \tilde{d}, L$ and $D_{\mathcal{K}}$, ignoring the dependencies in λ and D_θ, and using the fact that $D_{\widetilde{\mathcal{X}}} \leq L D_{\mathcal{K}}$ (as described below (B.9)). □

B.8 FW-SquareCB: CBCR with general reward fuctions

The SquareCB algorithm was recently proposed by Foster and Rakhlin [2020] for zero-regret contextual multi-armed bandit *with general reward functions*, based on the notion of online regression oracles. They propose, for single-reward contextual bandits with adversarial contexts and stochastic rewards, a generic randomized exploration scheme that delegates learning to an online regression algorithm. Their exploration/exploitation strategy then has (bandit) regret bounded as a function of the online regret of the regression algorithm. In this section, we extend the SquareCB approach to the case of CBCR. The main interest of this section is that by building on the work of Foster and Rakhlin [2020], we obtain at nearly no cost an algorithm for general reward functions for multi-armed CBCR problems.

This section shows how to extend this algorithm to our setting of concave rewards. To simplify the notation, we consider the case of finite K with atomic actions, i.e., $|\mathcal{A}| = K$. Our algorithm is based on an oracle for multi-dimensional regression RegSq, which provides approximate values for μ:

$$\forall T, \forall x \in \mathcal{X}, \quad \hat{\mu}_T(x) = \mathrm{RegSq}\big(x, (x_1, a_1, r_1, \dots, a_{T-1}, r_{T-1})\big). \tag{B.12}$$

The key assumption is that the problem is realizable and that RegSq has bounded regret:

Assumption G. *There is a function $T \mapsto R_{\mathrm{oracle}}(T) \in \mathbb{R}$, non-decreasing in T,[5] and Φ, a class of functions from $\mathcal{X}$ to $\mathbb{R}^{D\times K}$ such that, for every $T \in \mathbb{N}$:*

1. (Realizability) $\mu \in \Phi$,

[5] Monotonicity of R_{oracle} is not required in [Foster and Rakhlin, 2020]. We use it in (B.14) below to deal with time-dependent γ_t. Meaningful $R_{\mathrm{oracle}}(T)$ are non-decreasing with T since they bound a cumulative regret.

2. (Regret bound) *For every* $(x_t, a_t, r_t)_{t\in[\![T]\!]}, \in (\mathcal{X}\times\mathcal{A}\times\mathcal{K})^T$*, we have:*

$$\sum_{t=1}^{T} \|[\|\,]\hat{\mu}_t(x_t)a_t - r_{t2}^2 - \inf_{\phi\in\Phi}\sum_{t=1}^{T} \|[\|\,]\phi(x_t)a_t - r_{t2}^2 \leq R_{\text{oracle}}(T).$$

3. For every $(x_t, a_t, r_t)_{t\in[\![T]\!]} \in (\mathcal{X}\times\mathcal{A}\times\mathcal{K})^T, \hat{\mu}_T(x_T)a_T \in \mathcal{K}$.

Assumption G is the counterpart for multidimensional regression of Assumptions 1 and 2a of Foster and Rakhlin [2020], which are the basis of the original SquareCB algorithm.

Remark 8 (The "informal" assumption used in Table 5.1)**.** *Notice that in Table 5.1, we describe an "informal" version of this assumption, which reads* $\sum_{t=1}^{T} \|[\|\,]\hat{\mu}_t(x_t)a_t - \mu(x_t)a_{t2}^2 \leq R_{\text{oracle}}(T)$*, which is the counterpart for multi-dimensional regression of Assumption 2b by Foster and Rakhlin [2020]. Our choice in the table was to simplify the presentation, as this assumption is shorter. Our analysis is also valid under this alternative assumption. Our proofs are made under Assumption G because it is more widely applicable (more discussion of these assumptions can be found in [Foster and Rakhlin, 2020]).*

Algorithm 8 describes how SquareCB principles apply to our framework. We use the framework of the main paper, or, equivalently, the special case of Algorithm 6 where $\forall t \in \mathbb{N}, \rho_t = r_t$ and $z_t = \hat{s}_t$. Note that the algorithm is parameterized by $(\gamma_t)_{t\in\mathbb{N}_*}$ instead of the desired confidence level δ' to make the analysis more general. Theorem 43 gives a formula for γ_t as a function of the desired confidence δ'. As for the previous sections, we describe the algorithm for the general case of smooth approximations of f, using ∇f_{t-1} rather than ∇f in Line 4 of the algorithm.

At time step t, the regression oracle provides an estimate of $\mu(x_t)$, then the algorithm computes $\mathfrak{A}_t$, with a larger probability for the action which maximizes $a \mapsto \langle \nabla f(\hat{s}_{t-1}) \,|\, \hat{\mu}_t(x_t)a\rangle$. The exact formula for these probabilities $\mathfrak{A}_t$ follow the original SquareCB algorithm, with the exception that we use an iteration-dependent γ_t instead of a constant γ.[6]

The main result of this section is the following (see Section B.8.2 and the next section for intermediate lemmas):

Theorem 43. *Let* $\delta' > 0$*. For every* $t \in \mathbb{N}_*$*, let* $\gamma_t = \frac{2}{L}\sqrt{\frac{tK}{R_{\text{oracle}}(t)+8D_{\mathcal{K}}^2 \ln \frac{4t^2}{\delta'}}}$*. Then, under Assumptions E and G, Algorithm 8 satisfies, with probability at least* $1-\delta'$ *:*

$$R_T^{\text{gen}} \leq 4L\sqrt{KT\big(R_{\text{oracle}}(T) + 8D_{\mathcal{K}}^2 \ln \frac{4T^2}{\delta'}\big)} + LD_{\mathcal{K}}\sqrt{2T \ln \frac{2}{\delta'}}$$

Recall that Assumption B is a special case of E when $\rho_t = r_t$, as we are here. Thus, the bound on R_T^{gen} is the same irrespective of whether we use the algorithm for smooth f (in which case $R_T^{\text{scal}} = R_T^{\text{gen}}$) or with smooth approximations (in which case $R_T^{\text{scal,sm}} = R_T^{\text{gen}}$). This is because only the Lipschitzness of $(f_t)_{t\in\mathbb{N}}$ is used in the analysis of R_T^{gen} for FW-SquareCB.

The following result is a direct corollary of Theorem 43, and gives the order of magnitude we obtain for smooth f. Obtaining a similar for smooth approximations of f, using Theorem 35 instead of Theorem 34 is straightforward.

Proof of the FW-SquareCB regret bound of Table 5.1. We apply the bound obtained by Theorem

[6] Throughout the paper, we chose to provide anytime bounds rather than bounds that depend on horizon-dependent parameters. The analysis with fixed γ is easier.

43 within the bound of Theorem 34, using $\delta' := 2\delta/3$ and $\delta := \delta/3$. We obtain:

$$R_T \leq \frac{4L\sqrt{KT\big(R_{\text{oracle}}(T) + 8D_{\mathcal{K}}^2 \ln \frac{12t^2}{\delta}\big)} + 2LD_{\mathcal{K}}\sqrt{2T \ln \frac{3}{\delta}} + \tilde{C}\ln(eT)}{T}.$$

The bound given in the theorem uses the sub-additivity of $\sqrt{\cdot}$ to group the terms in $\sqrt{\ln \delta^{-1}}$ for better readability. □

The proof of Theorem 43 is decomposed into two subsections: in the next subsection, we make the necessary adaptations to the SquareCB analysis to account for multi-dimensional regression. This proof follows essentially the same steps as the original analysis of SquareCB. There are only two changes:

- We use multi-dimensional regression instead of scalar regression, while we need to bound a scalar regret. There is an additional step to go from the scalar regret to the multi-dimensional regression, but it turns out there is no added difficulty (see first line of the proof of Lemma 46).
- For coherence with the overall bounds of the paper, we use an anytime analysis using an increasing sequence of $(\gamma_t)_{t\in[\![T]\!]}$, instead of a fixed exploration parameter γ that needs be tuned for a specific horizon determined *a priori*. This introduces a bit more difficulty, where the main tool is Lemma 47. Our choice of anytime bound is more for coherence in the presentation of the paper than an intended contribution.

  Nonetheless, what we gain with our anytime bound is that the exploration parameter γ does not depend on a fixed horizon. What we lose, however, is that we need a high-probability bound on cumulative errors based on $R_{\text{oracle}}(t)$ that is valid for every t (see Lemma 46), while the "fixed gamma" case only requires this bound to hold for the horizon T. This is the reason for the $\ln T$ factor in our bound, which is not present in the original paper.

In the next sections, we use the following notation:

$$g_t = \nabla f_{t-1}(\hat{s}_{t-1}), \qquad \underline{\mu}_t = g_t^\intercal \mu_t(x_t), \qquad \underline{\mu}_t^* = \max_{a\in\mathcal{A}} \underline{\mu}_t a,$$
$$\underline{\hat{\mu}}_t = g_t^\intercal \hat{\mu}_t(x_t), \qquad \underline{\hat{\mu}}_t^* = \max_{a\in\mathcal{A}} \underline{\hat{\mu}}_t a.$$

B.8.1 Adaptation of SquareCB proof to CBCR

In the SquareCB paper, Foster and Rakhlin [2020] study high probability bounds on a different type of regret, based on average rewards associated to the actions $\mu(x_t)a_t$ rather than observed rewards r_t. However, this difference has little influence since we can start with the following inequality, which is similar to [Foster and Rakhlin, 2020, Lemma 2].

Lemma 44. *Under Assumption E, for every $T \in \mathbb{N}_*$, every $\delta' > 0$, Algorithm 8 satisfies*

$$\sum_{t=1}^{T} \big(\underline{\mu}_t^* - g_t^\intercal r_t\big) \leq \sum_{t=1}^{T} \mathbb{E}_{a\sim\mathfrak{A}_t}\big[\underline{\mu}_t^* - \underline{\mu}_t^\intercal a\big] + LD_{\mathcal{K}}\sqrt{2T\ln(1/\delta')}.$$

Proof. The proof is by Azuma's inequality. Let $\mathfrak{F} = (\mathfrak{F}_t)_{t\in\mathbb{N}_*}$ be the filtration where $\mathfrak{F}_t$ is the σ-algebra generated by $(x_1, a_1, r_1, \ldots, x_{t-1}, a_{t-1}, r_{t-1}, x_t)$, and let us denote $X_T = \sum_{t=1}^{T} \mathbb{E}_{a\sim\mathfrak{A}_t}\big[\underline{\mu}_t^\intercal a\big] - g_t^\intercal r_t$. Then, $(X_T)_{T\in\mathbb{N}}$ is a martingale adapted to filtration $\mathfrak{F}$ and satisfies $|X_t - X_{t-1}| \leq LD_{\mathcal{K}}$. We obtain the result by noticing that $X_T = \sum_{t=1}^{T}\big(\underline{\mu}_t^* - g_t^\intercal r_t\big) - \sum_{t=1}^{T}\mathbb{E}_{a\sim\mathfrak{A}_t}\big[\underline{\mu}_t^* - \underline{\mu}_t^\intercal a\big]$ and applying Azuma's inequality to X_T. □

Notice that the difference between [Foster and Rakhlin, 2020, Lemma 2] and our Lemma 44 is that

we consider the randomization over actions and rewards, while they only consider the randomization over actions because they study average rewards. However, since it does not change the upper bound on the variations of the martingale, this additional randomness does not change the bound.

The next step is the fundamental step in the proof of the original SquareCB algorithm. Even though the notation differ slightly from the original paper, the proof is the same as in [Foster and Rakhlin, 2020, Appendix B]:

Lemma 45. *[Foster and Rakhlin, 2020, Lemma 3] For every $t \in \mathbb{N}_*$, the choice of γ_t and $\mathfrak{A}(h_t, x_t, \delta')$ of Algorithm 8 guarantees:*

$$\mathbb{E}_{a\sim\mathfrak{A}_t}\left[\underline{\mu}_t^* - \underline{\mu}_t^\intercal a\right] \leq \frac{2K}{\gamma_t} + \frac{\gamma_t}{4}\mathbb{E}_{a\sim\mathfrak{A}_t}\left[\left(\underline{\hat{\mu}}_t^\intercal a - \underline{\mu}_t^\intercal a\right)^2\right].$$

The last step of these preliminary lemmas is to relate the cumulative expected error to the oracle regret bound. We use here the same proof as [Foster and Rakhlin, 2020, Lemma 2]. We then have:

Lemma 46. *Under Assumption E, for every $\delta' > 0$, Algorithm 8 satisfies, w.p. at least $1-\delta'$:*

$$\forall T \in \mathbb{N}_*, \sum_{t=1}^{T} \mathbb{E}_{a\sim\mathfrak{A}_t}\left[\left(\underline{\hat{\mu}}_t^\intercal a - \underline{\mu}_t^\intercal a_t\right)^2\right] \leq 2L^2 R_{\text{oracle}}(T) + 16L^2 D_{\mathcal{K}}^2 \ln\frac{2T^2}{\delta'}$$

Proof. We first notice that $\sum_{t=1}^{T}\mathbb{E}_{a\sim\mathfrak{A}_t}\left[\left(\underline{\hat{\mu}}_t^\intercal a - \underline{\mu}_t^\intercal a_t\right)^2\right] \leq L^2\sum_{t=1}^{T}\mathbb{E}_{a\sim\mathfrak{A}_t}\left[\|[\|\,]\hat{\mu}(x_t)a - \mu(x_t)a\|_2^2\right]$. We then apply the same steps as in the proof of [Foster and Rakhlin, 2020, Lemma 2] to $\sum_{t=1}^{T}\mathbb{E}_{a\sim\mathfrak{A}_t}\left[\|[\|\,]\hat{\mu}(x_t)a - \mu(x_t)a\|_2^2\right]$ (which we do not reproduce here) to obtain: for every every $T \in \mathbb{N}$, every $\delta'_T > 0$, with probability at least $1-\delta'_T$:

$$\sum_{t=1}^{T} \mathbb{E}_{a\sim\mathfrak{A}_t}\left[\left(\underline{\hat{\mu}}_t^\intercal a - \underline{\mu}_t^\intercal a_t\right)^2\right] \leq 2L^2 R_{\text{oracle}}(T) + 16L^2 D_{\mathcal{K}}^2 \ln\frac{1}{\delta'_T}$$

Let $\delta' > 0$. Applying a union bound and taking $\delta'_t = \frac{\delta'}{2t^2}$ so that $\sum_{t=1}^{T}\delta'_t \leq \frac{\pi^2}{12}\delta' \leq \delta'$, we obtain the desired result. □

Notice the $\log T$ factor in the bound, which appears because the bound is valid for all time steps. This is because we propose anytime convergence bounds, with the exploration parameter that decreases with time, whereas [Foster and Rakhlin, 2020] only prove their result in the case where the exploration parameter is chosen for a specific horizon.

As the main first step for the final result, we need these two lemmas which are the main technical steps to our anytime bound. The proof is deferred to Appendix B.10.2

Lemma 47. *Let $(\lambda_t)_{t\in\mathbb{N}} \in \mathbb{R}_+^T$ be a sequence of non-negative numbers, denote $\Lambda_T = \sum_{t=1}^{T}\lambda_t$ and let $(\overline{\Lambda}_T)_{T\in\mathbb{N}}$ such that $\forall T \in \mathbb{N}, \overline{\Lambda}_T > 0$ and $\overline{\Lambda}_T \geq \Lambda_T$.*

$$\sum_{t=1}^{T}\frac{\lambda_t}{\sqrt{\overline{\Lambda}_t}} \leq 2\sqrt{\overline{\Lambda}_T}.$$

We get the following corollary

Lemma 48. *Let $R'_{\text{oracle}}(T,\delta') = 2L^2 R_{\text{oracle}}(T) + 16L^2 D_{\mathcal{K}}^2 \ln\frac{2T^2}{\delta'}$. Under the conditions of Lemma 46, assume that there is $\gamma_0 > 0$ such that $\forall t \in [\![T]\!], \gamma_t = \gamma_0\sqrt{\frac{t}{R'_{\text{oracle}}(t,\delta')}}$. Then, for every $\delta' > 0$,*

Algorithm 8 satisfies, w.p. at least $1-\delta'$*:*

$$\sum_{t=1}^{T} \gamma_t \mathbb{E}_{a\sim\mathfrak{A}_t}\left[\left(\hat{\underline{\mu}}_t^\intercal a - \underline{\mu}_t^\intercal a\right)^2\right] \leq 2\gamma_0\sqrt{TR'_{\text{oracle}}(T,\delta')}. \tag{B.13}$$

Proof. Using $\gamma_t \leq \gamma_0\sqrt{\frac{T}{R'_{\text{oracle}}(t,\delta')}}$, the sum on the left hand side of (B.13) has the form of Lemma 47 multiplied by $\gamma_0\sqrt{T}$, with probability $1-\delta'$ by Lemma 46. The result thus follows from applying both Lemmas. □

B.8.2 Final result

Proof of Theorem 43. Notice that the value of γ_t given in the theorem is equal to

$$\gamma_t = 2\sqrt{\frac{2tK}{R'_{\text{oracle}}(t,\delta'/2)}}.$$

Using this formula, we have

$$\begin{aligned}\sum_{t=1}^{T}\frac{2K}{\gamma_t} &= \sqrt{\frac{K}{2}}\sum_{t=1}^{T}\sqrt{\frac{R'_{\text{oracle}}(t,\delta'/2)}{t}} \leq \sqrt{\frac{R'_{\text{oracle}}(T,\delta'/2)K}{2}}\sum_{t=1}^{T}\frac{1}{\sqrt{t}} \\ &\leq \sqrt{2KTR'_{\text{oracle}}(T,\delta'/2)}.\end{aligned} \tag{B.14}$$

Where the first line comes from the monotonicity of $R_{\text{oracle}}(T)$ of Assumption G.

Using Lemmas 45 and 48, we thus have, with probability $1-\delta'/2$:

$$\sum_{t=1}^{T}\mathbb{E}_{a\sim\mathfrak{A}_t}\left[\underline{\mu}_t^* - \underline{\mu}_t^\intercal a\right] \leq 2\sqrt{2KTR'_{\text{oracle}}(T,\delta'/2)}.$$

Using a union bound and Lemma 44, we obtain, with probability at least $1-\delta'$:

$$\sum_{t=1}^{T}\left(\underline{\mu}_t^* - g_t^\intercal r_t\right) \leq 2\sqrt{2KTR'_{\text{oracle}}(T,\delta'/2)} + LD_{\mathcal{K}}\sqrt{2T\ln\frac{2}{\delta'}}.$$

□

B.9 FW-LinUCBRank: CBCR for fair ranking with linear contextual bandits

In this section and following the previous sections, we analyze Algorithm 9 under Assumption E, which is more general than the bound proposed in the main paper, which used Algorithm 3 under Assumption B. The only difference in the algorithms is the use of f_{t-1} instead of f in Line 4 of Algorithm 9. This allows us to provide the algorithm for both smooth and non-smooth objective functions f.

The bound is decomposed into two parts: we describe the results for online regression within our observation model for ranking in the next subsection. Then we dive into the final result.

Algorithm 9: FW-linUCBRank: linear contextual bandits for fair ranking.

input : $\delta' > 0, \lambda > 0, \hat{s}_0 \in \mathcal{K}$ $V_0 = \lambda \boldsymbol{I}_d, y_0 = \mathbf{0}_d, \hat{\theta}_0 = \mathbf{0}_d$

1 **for** $t = 1, \dots$ **do**

2 Observe context $x_t \sim P$

3 $\forall i, \hat{v}_{t,i} \leftarrow \hat{\theta}_{t-1}^\intercal x_{t,i} + \alpha_t\left(\frac{\delta'}{3}\right) \|x_{t,i}\|_{V_{t-1}^{-1}}$ // UCB on $v_i(x_t)$, see Lem. 49 for def. of α_t

4 $a_t \leftarrow \text{top-}\overline{k}\{\frac{\partial f_{t-1}}{\partial z_{m+1}}(\hat{s}_{t-1})\hat{v}_{t,i} + \frac{\partial f_{t-1}}{\partial z_i}(\hat{s}_{t-1})\}_{i=1}^m$ // FW linear optimization step

5 Observe exposed items $e_t \in \{0,1\}^m$ and user feedback $c_t \in \{0,1\}^m$

6 Update $\hat{s}_t \leftarrow \hat{s}_{t-1} + \frac{1}{t}(r_t - \hat{s}_{t-1})$

7 $V_t \leftarrow V_{t-1} + \sum_{i=1}^m e_{t,i} x_{t,i} x_{t,i}^\intercal, \quad y_t \leftarrow y_{t-1} + \sum_{i=1}^m c_{t,i} x_{t,i}$ and $\hat{\theta}_t \leftarrow V_t^{-1} y_t$ // regression

8 **end**

B.9.1 Results for online linear regression (from [Li et al., 2016])

Even though our linear contextual bandit setup is different from e.g., [Lagrée et al., 2016] for ranking, the availability of the feedback $e_{t,i}$, which tells us whether item i has been exposed, makes the analysis of the online linear regression similar to the general setup of linear bandits. Our approach builds on the confidence intervals developed by Li et al. [2016], which expands the analysis of confidence ellipsoids for linear regression of Abbasi-Yadkori et al. [2011] to cascade user models in rankings.

Each $c_{t,i}$ is $\frac{1}{2}$-subgaussian (because Bernoulli), and is conditionally independent of the other random variables conditioned and on $e_{t,i}$ and $x_{t,i}$. The incremental linear regression of line 7 of Algorithm 9 is the same as [Abbasi-Yadkori et al., 2011]. Our observation model satisfies the conditions of the analysis of confidence ellipsoids of Li et al. [2016], from which we obtain:

Lemma 49. *Under the probabilistic model described in Section 5.4, and under Assumption C. Let $\delta' > 0$ and $\lambda \geq D_{\mathcal{X}}^2 \overline{k}$, and let*

$$\alpha_T(\delta') = \frac{1}{2}\sqrt{\ln\left(\frac{\det(V_T)}{V_0 \delta'^2}\right)} + \sqrt{\lambda} D_\theta.$$

Then, under Assumption C and with the notation of Algorithm 9, we have:

- *(*[Li et al., 2016, Lemma 4.2]*) with probability $\geq 1 - \delta'$, for all $T \geq 0$, θ lies in the confidence ellipsoid:*

$$C_T = \{\tilde{\theta} \in \mathbb{R}^d : \left\|\hat{\theta}_T - \tilde{\theta}\right\|_{V_T} \leq \alpha_T(\delta')\}$$

- *(*[Li et al., 2016, Lemma 4.4]*):*

$$\alpha_T(\delta') \leq \frac{1}{2}\sqrt{2\ln\left(\frac{1}{\delta'}\right) + d\ln\left(1 + \frac{T D_{\mathcal{X}}^2 \overline{k}}{\lambda d}\right)} + \sqrt{\lambda} D_\theta.$$

These results stem from [Li et al., 2016, Lemma A.4 and A.5] that claim that, with the assumptions of Lemma 49, the following inequality holds with probability 1:

$$\sum_{t=1}^T \sum_{i=1}^m \|x_{t,i}\|_{V_{t-1}^{-1}}^2 e_{t,i} \leq 2\ln\frac{\det V_T}{\det(V_0)} \leq 2d\ln\left(1 + \frac{T D_{\mathcal{X}}^2 \overline{k}}{\lambda d}\right).$$

Notice that terms equivalent to $D_{\mathcal{X}}$ and D_θ do not appear in [Li et al., 2016] because they assume they are ≤ 1. The $D_{\mathcal{X}}^2$ term comes from a modification necessary in [Li et al., 2016, Lemma

A.4] while D_θ is required by the initial confidence bound proved by Abbasi-Yadkori et al. [2011]. The term $\overline{k}$ plays the constant C_γ of [Li et al., 2016].

B.9.2 Guarantees for FW-LinUCB

We start by writing an alternative to Assumption D for the case where f is not smooth to carry out our analysis with as little assumptions on f as possible:

Assumption D′. *The assumptions of the framework of Sec. 5.4 hold, as well as Ass. E. Moreover,* $\forall t \in \mathbb{N}, \forall z \in \mathcal{K}\ \frac{\partial f_t}{\partial z_{m+1}}(z) > 0$, *and* $\forall x \in \mathcal{X}, 1 \geq b_1(x) \geq \ldots \geq b_{\overline{k}}(x) = \ldots = b_m(x) = 0$.

Lemma 50. *Under Assumptions D′ and C Let* $T > 0, \delta' > 0$ *and* $\lambda \geq D_{\mathcal{X}}^2 \overline{k}$. *Then for every* $\delta' > 0$, *Algorithm 9 satisfies, with probability at least* $1 - \delta'$*:*

$$R_T^{\text{gen}} \leq 2L\alpha_T(\delta'/3)\sqrt{T\overline{k}}\left(\sqrt{2\ln(\frac{3}{\delta'})} + \sqrt{2d\ln\left(1 + \frac{TD_{\mathcal{X}}^2\overline{k}}{\lambda d}\right)}\right) + LD_{\mathcal{K}}\sqrt{2T\ln\frac{3}{\delta'}}.$$

where α_T *is defined in Lemma 49.*

Proof. Let $g_t = \nabla f_{t-1}(\hat{s}_{t-1})$, and $a_t^* \in \operatorname{argmax}_{a\in\mathcal{A}}\langle g_t \,|\, \mu(x_t)a - r_t\rangle$. Let furthermore $\delta' > 0$. Assume the algorithm uses $\alpha_t(\delta'/3)$, so that $C_t = \{\tilde{\theta} \in \mathbb{R}^d : \left\|\hat{\theta}_t - \tilde{\theta}\right\|_{V_t} \leq \alpha_t(\delta'/3)\}$.

Let us define $\hat{\mu}_t$ similarly to Proposition 15, i.e., $\forall t \in \mathbb{N}_*$, $\hat{\mu}_t$ such that $\forall i \in [\![m]\!], \hat{\mu}_{t,i} = \mu_i(x_t)$ and $\hat{\mu}_{t,m+1} = \hat{v}_t b(x_t)^\intercal$ viewed as a column vector, with $\hat{v}$ defined in line 3 of Algorithm 9. We have:

$$\sum_{t=1}^T \max_{a\in\mathcal{A}}\langle g_t \,|\, \mu(x_t)a - r_t\rangle = \sum_{t=1}^T \underbrace{\langle g_t \,|\, \mu(x_t)a_t^* - \hat{\mu}_t a_t\rangle}_{:=A_t} + \sum_{t=1}^T \underbrace{\langle g_t \,|\, \hat{\mu}_t a_t - \mu(x_t)a_t\rangle}_{:=B_t} + \sum_{t=1}^T \underbrace{\langle g_t \,|\, \mu(x_t)a_t - r_t\rangle}_{:=X_t}$$

Step 1: Upper bound on $\sum_{t=1}^T A_t$ via optimism Let $t \geq 0$. For $\tilde{\theta} \in \mathbb{R}^d$, denote $\mu_{\tilde{\theta}}(x) \in \mathbb{R}^{D\times K}$ (recall $D = m + 1$), the average reward function where parameters $\tilde{\theta}$ replace θ. We first show that for every $a \in \mathcal{A}$, we have $\max_{\tilde{\theta}\in C_t}\langle g_t \,|\, \mu_{\tilde{\theta}}(x_t)a\rangle \leq \langle g_t \,|\, \hat{v}_t a\rangle$, where $\hat{v}_t$ is given in Line 3 of Algorithm 9.

Given $a \in \mathcal{A}$, let us denote by $\text{mat}(a)$ the view of a as an $m \times m$ permutation matrix (instead of an m^2-dimensional column vector). Recalling that x_t is a $m \times d$ matrix and $g_t \in \mathbb{R}^{m+1}$, let us denote by $g_{t,1:m}$ the vector containing the first m dimensions of g_t. We have:

$$\langle g_t \,|\, \mu_{\tilde{\theta}}(x_t)a\rangle = g_{t,1:m}^\intercal \text{mat}(a)b(x_t) + g_{t,m+1}\tilde{\theta}^\intercal x_t^\intercal \text{mat}(a)b(x_t), \tag{B.15}$$

therefore:

$$\begin{aligned}\max_{\tilde{\theta}\in C_t}\langle g_t \,|\, \mu_{\tilde{\theta}}(x_t)a\rangle &= g_{t,1:m}^\intercal \text{mat}(a)b(x_t) + g_{t,m+1}\max_{\tilde{\theta}\in C_T}\left(g_{t,m+1}\tilde{\theta}^\intercal x_t^\intercal \text{mat}(a)b(x_t)\right)\\ &\leq g_{t,1:m}^\intercal \text{mat}(a)b(x_t) + g_{t,m+1}\hat{v}_t^\intercal \text{mat}(a)b(x_t) = \langle g_t \,|\, \hat{\mu}_t a\rangle.\end{aligned}$$

The first equality is because $g_{t,m+1} \geq 0$. The second equality is deduced by direct calculation from the definition of C_t in Lemma 49, which gives $\hat{v}_{t,i} = \max_{\tilde{\theta}\in C_t}\tilde{\theta}^\intercal x_{t,i}$.

By Proposition 15 we have that a_t defined at Line 4 of Algorithm 9 maximizes $\langle g_t \,|\, \hat{\mu}_t a\rangle$ over a. We thus have $\max_{a\in\mathcal{A}}\max_{\tilde{\theta}\in C_t}\langle g_t \,|\, \mu_{\tilde{\theta}}(x_t)a\rangle \leq \langle g_t \,|\, \hat{\mu}_t a_t\rangle$.

By Lemma 49, we have $\theta \in C_t$ for all $t \geq 0$ with probability $1-\delta'/3$. Therefore, with probability $1-\delta'/3$, we have for all $t \geq 0$: $\langle g_t \,|\, \mu_\theta(x_t) a_t^* \rangle \leq \langle g_t \,|\, \hat{\mu}_t a_t \rangle$. Noting that $\mu_\theta(x_t) = \mu(x_t)$ by definition of θ, we obtain that $\forall t, A_t \leq 0$ and thus $\sum_{t=1}^T A_t \leq 0$ with probability $1-\delta'/3$.

Step 2: Upper bound on $\sum_{t=1}^T B_t$ using linear bandit techniques Let $a_{t,i} \in \mathbb{R}^m$ denote the i-th row of $\mathrm{mat}(a_t)$, which contains only 0s except a 1 at the rank of item i in a. Since $\hat{\mu}_t$ and $\mu(x_t)$ only differ in the last dimension, which is the user utility, we have, using (B.15):

$$B_t = g_{t,m+1}\left((\hat{v}_t - v(x_t))^\intercal \mathrm{mat}(a_t) b(x_t)\right) = g_{t,m+1} \sum_{i=1}^m \left(\hat{v}_{t,i} - v_i(x_t)\right) a_{t,i}^\intercal b(x_t)$$

Denoting $\overline{e}_{t,i} = a_{t,i}^\intercal b(x_t) \in \mathbb{R}$ the expected exposure of item i in ranking a_t given context x_t, we have:

$$\begin{aligned} B_t &= \underbrace{g_{t,m+1}}_{\in[0,L]} \sum_{i=1}^m \left(\hat{v}_{t,i} - v_i(x_t)\right)\overline{e}_{t,i} \leq L \sum_{i=1}^m \Big((\hat{\theta}_{t-1} - \theta)^\intercal x_{t,i} + \alpha_t(\delta'/3) \left\|x_{t,i}\right\|_{V_{t-1}^{-1}} \Big)\overline{e}_{t,i} \\ &\leq L \sum_{i=1}^m \Big(\left\|\hat{\theta}_{t-1} - \theta\right\|_{V_{t-1}} \left\|x_{t,i}\right\|_{V_{t-1}^{-1}} + \alpha_t(\delta'/3) \left\|x_{t,i}\right\|_{V_{t-1}^{-1}} \Big)\overline{e}_{t,i} \qquad \text{(by Cauchy-Schwarz)} \end{aligned}$$

By Lemma 49, we have, with probability $1-\delta'/3$: $\left\|\hat{\theta}_{t-1} - \theta\right\|_{V_{t-1}} \leq \alpha_t(\delta'/3)$, and thus:

$$\begin{aligned} B_t &\leq 2L\alpha_t\Big(\frac{\delta'}{3}\Big) \sum_{i=1}^m \left\|x_{t,i}\right\|_{V_{t-1}^{-1}} \overline{e}_{t,i} \\ &= 2L\alpha_t\Big(\frac{\delta'}{3}\Big) \Bigg(\underbrace{\Big(\sum_{i=1}^m \left\|x_{t,i}\right\|_{V_{t-1}^{-1}} (\overline{e}_{t,i} - e_{t,i}) \Big)}_{X_t'} + \Big(\sum_{i=1}^m \left\|x_{t,i}\right\|_{V_{t-1}^{-1}} e_{t,i} \Big) \Bigg) \end{aligned}$$

We first deal with the sum over t of the right-hand side, using $e_{t,i} \in \{0,1\}$:

$$\begin{aligned} \sum_{t=1}^T \sum_{i=1}^m \left\|x_{t,i}\right\|_{V_{t-1}^{-1}} e_{t,i} &= \sum_{t=1}^T \sum_{i=1}^m (\left\|x_{t,i}\right\|_{V_{t-1}^{-1}} e_{t,i}) \times e_{t,i} \\ &\leq \sqrt{\sum_{t=1}^T \sum_{i=1}^m e_{t,i}^2} \sqrt{\sum_{t=1}^T \sum_{i=1}^m (\left\|x_{t,i}\right\|_{V_{t-1}^{-1}}^2 e_{t,i}^2)} \qquad \text{(by Cauchy-Schwarz)} \\ &\leq \sqrt{T\overline{k}} \sqrt{d \ln\Big(1 + \frac{T D_{\mathcal{X}}^2 \overline{k}}{\lambda d}\Big)}. \qquad \text{(by B.9.1)} \end{aligned}$$

For the left-hand term, we have that $\left(\sum_{t=1}^T X_t'\right)_{T \in \mathbb{N}_*}$ is a martingale adapted to the filtration $\overline{\mathfrak{F}} = (\overline{\mathfrak{F}}_T)_{T \in \mathbb{N}_*}$ where $\overline{\mathfrak{F}}_T$ is the σ-algebra generated by $(x_1, a_1, r_1, \ldots, x_{T-1}, a_{T-1}, r_{T-1}, x_T, a_T)$, with $\left|X_t'\right| \leq \frac{D_{\mathcal{X}} \overline{k}}{\sqrt{\lambda}}$. Thus, with probability at least $1-\delta'/3$, we have

$$\sum_{t=1}^T \sum_{i=1}^m \sum_{i=1}^m \left\|x_{t,i}\right\|_{V_{t-1}^{-1}} (\overline{e}_{t,i} - e_{t,i}) \leq \frac{D_{\mathcal{X}} \overline{k}}{\sqrt{\lambda}} \sqrt{2T \ln \frac{3}{\delta'}} \leq \sqrt{2T\overline{k} \ln \frac{3}{\delta'}}.$$

Where the last inequality comes from the assumption $\lambda \geq D_{\mathcal{X}}^2 \overline{k}$. We conclude this step by saying that with probability $1 - 2\delta'/3$, we have:

$$\sum_{t=1}^{T} B_t \leq 2L\alpha_t\Big(\frac{\delta'}{3}\Big)\sqrt{T\overline{k}}\Bigg(\sqrt{2\ln\frac{3}{\delta'}} + \sqrt{d\ln\Big(1 + \frac{TD_{\mathcal{X}}^2\overline{k}}{\lambda d}\Big)}\Bigg).$$

Step 3: Upper bound on $\sum_{t=1}^{T} X_t$ using Azuma's inequality Following the same arguments as in the proof of Thm. 42, let $\overline{\mathfrak{F}} = \big(\overline{\mathfrak{F}}_t\big)_{t\in\mathbb{N}_*}$ be the filtration where $\overline{\mathfrak{F}}_t$ is the σ-algebra generated by $(x_1, a_1, r_1, \ldots, x_{t-1}, a_{t-1}, r_{t-1}, x_t, a_t)$. Then $(X_t)_{t\in\mathbb{N}}$ is a martingale difference sequence adapted to $\overline{\mathfrak{F}}$ with $|X_t| \leq LD_{\mathcal{K}}$, so that $\sum_{t=1}^{T} X_t \leq L\sqrt{2T\overline{k}\ln\frac{3}{\delta'}}$ with probability $1 - \delta'/3$.

The final result is obtained using a union bound, considering that Step 1 and Step 2 use the same confidence interval given by Lemma 49 which is valid w.p. $\geq 1 - \delta'/3$, Step 2 uses an addition Azuma inequality valid w.p. $1 - \delta'/3$, and step 3 uses an additional Azuma inequality which valid with probability $\geq 1 - \delta'/3$. □

Theorem 16. *Under Assumptions B, C and D, for every $\delta' > 0$, every $T \in \mathbb{N}_*$, every $\lambda \geq D_{\mathcal{X}}^2\overline{k}$, with probability at least $1 - \delta'$, Algorithm 3 has scalar regret bounded by*

$$R_T^{\text{scal}} = O\Bigg(L\sqrt{T\overline{k}}\sqrt{d\ln(T/\delta')}\Big(\sqrt{d\ln(T/\delta')} + D_\theta\sqrt{\lambda} + \sqrt{\overline{k}/d}\Big)\Bigg).$$

Thus, considering only $d, T, \overline{k}$ and $\delta = \delta'$ Alg. 3 has regret $R_T \leq O\big(\frac{d\overline{k}\ln(T/\delta)}{\sqrt{T}}\big)$ w.p. at least $1 - \delta$.

Proof. Let $\delta > 0$ and use $\delta' := 3\delta/4$ and $\delta := \delta/4$ in the bound on R_T obtained by applying Lemma 50 and Theorem 34. Notice that Using $\lambda \geq D_{\mathcal{X}}^2\overline{k}$ and $D_{\mathcal{K}} = O(\overline{k})$, we have:

$$\overline{R}^{\text{scal}}(T, 3\delta/4) = O\Bigg(L\alpha_T(\delta)\sqrt{T\overline{k}d\ln(T/\delta)} + L\overline{k}\sqrt{T\ln(1/\delta)}\Bigg)$$

and $\alpha_T(\delta) = O\Big(\sqrt{d\ln(T/\delta)} + D_\theta\sqrt{\lambda}\Big)$.

We thus get

$$\begin{aligned}\overline{R}^{\text{scal}}(T, \delta) &= O\Bigg(L\sqrt{T\overline{k}}\sqrt{d\ln(T/\delta)}\Big(\sqrt{d\ln(T/\delta)} + D_\theta\sqrt{\lambda}\Big) + L\overline{k}\sqrt{T\ln(1/\delta)}\Bigg)\\ &= O\Bigg(L\sqrt{T\overline{k}}\sqrt{d\ln(T/\delta)}\Big(\sqrt{d\ln(T/\delta)} + D_\theta\sqrt{\lambda} + \sqrt{\overline{k}/d}\Big)\Bigg)\end{aligned}$$

For the smooth case, the total bound adds $O(L\overline{k}\sqrt{T\ln(1/\delta)} + \tilde{C}\frac{\ln T}{T})$. A bound on the complete regret is thus

$$R_T = O\Bigg(L\sqrt{T\overline{k}}\sqrt{d\ln(T/\delta)}\Big(\sqrt{d\ln(T/\delta)} + D_\theta\sqrt{\lambda} + \sqrt{\overline{k}/d} + \tilde{C}\frac{\ln T}{T}\Big)\Bigg)$$

□

B.10 Additional technical lemmas

B.10.1 Proof of Lemma 30 ($\mathcal{S}$ is compact)

Lemma 30. *Under Assumption $\tilde{A}$, $\mathcal{S}$ is compact and $\forall T \in \mathbb{N}_*$, $\forall x_{1:T} \in \mathcal{X}^T, \mathcal{S}(x_{1:T})$ is compact.*

Proof. We start with $\mathcal{S}(x_{1:T})$. Let $x_{1:T} \in \mathcal{X}^T$. We notice that $\mathcal{S}(x_{1:T})$ is the image of $\overline{\mathcal{A}}^T$ by the continuous mapping $\phi : (\mathbb{R}^K)^T \to \mathbb{R}^D$ defined by $\phi(a_1, ..., a_T) = \frac{1}{T}\sum_{t=1}^T \mu(x_t)a_t$. Since $\overline{\mathcal{A}}$ is compact, $\overline{\mathcal{A}}^T$ is compact as well. $\mathcal{S}(x_{1:T})$ is thus the image of a compact set by a continuous function, and is therefore compact.

For the set $\mathcal{S}$, we provide a proof here using Diestel's theorem (see [Yannelis, 1991]). Consider the set-valued map defined by $G : \mathcal{X} \to \{B \mid B \subseteq \mathbb{R}^D\}$

$$G(x) := \mu(x)\overline{\mathcal{A}} := \{\mu(x)\overline{a} \mid \overline{a} \in \overline{\mathcal{A}}\}.$$

Then, $\mathcal{S}$ can be written as the *Aumann integral* of G over $\mathcal{X}$ w.r.t P, i.e.

$$\mathcal{S} = \int_{\mathcal{X}} G \,\mathrm{d}P := \left\{ \int_{\mathcal{X}} g \,\mathrm{d}P \;\middle|\; g \in \mathcal{G} \right\}, \tag{B.16}$$

where $\mathcal{G} \subseteq L^1(\mathcal{X}, P)$ is the collection of all P-integrable selections of G, i.e. the collection of all P-integrable functions $g : \mathcal{X} \to \mathbb{R}$ such that $g(x) \in G(x)$ for P-a.e $x \in \mathcal{X}$.

Now, since $\overline{\mathcal{A}}$ is compact, convex and nonempty, the values of the set-valued function G are nonempty, convex, and compact. Moreover, since $\sup_{x\in\mathcal{X}, a\in\overline{\mathcal{A}}} \|\mu(x)a\|_2 < +\infty$ because $\forall x, a, \mu(x)a \in \mathcal{K}$, the set-valued function G is P-integrably bounded in the sense of [Yannelis, 1991, Section 2.2]. It then follows from *Diestel's Theorem* [Yannelis, 1991, Theorem 3.1] that the collection $\mathcal{G}$ of P-integrable selections of G is weakly compact in $L^1(\mathcal{X}, P)$. Finally, since $g \mapsto \int_{\mathcal{X}} g \,\mathrm{d}P$ is a weakly continuous mapping from $L^1(\mathcal{X}, P)$ to $\mathbb{R}^D$, and $\mathcal{S} \subseteq \mathbb{R}^D$ is the image of $\mathcal{G}$ under this mapping (refer to the correspondence (B.16)), we deduce that $\mathcal{S}$ is weakly compact as a subset of $\mathbb{R}^D$, and therefore compact since $\mathbb{R}^D$ is finite-dimensional. □

B.10.2 Proof of Lemma 47

Lemma 47. *Let $(\lambda_t)_{t\in\mathbb{N}} \in \mathbb{R}_+^T$ be a sequence of non-negative numbers, denote $\Lambda_T = \sum_{t=1}^T \lambda_t$ and let $(\overline{\Lambda}_T)_{T\in\mathbb{N}}$ such that $\forall T \in \mathbb{N}, \overline{\Lambda}_T > 0$ and $\overline{\Lambda}_T \geq \Lambda_T$.*

$$\sum_{t=1}^T \frac{\lambda_t}{\sqrt{\overline{\Lambda}_t}} \leq 2\sqrt{\overline{\Lambda}_T}.$$

Proof. First, we treat the case where $\lambda_0 > 0$. Then $\forall_t \in [\![T]\!], \Lambda_t > 0$. We thus have

$$\sum_{t=1}^T \frac{\lambda_t}{\sqrt{\overline{\Lambda}_t}} \leq \sum_{t=1}^T \frac{\lambda_t}{\sqrt{\Lambda_t}}$$

We now prove that the right-hand term is $\leq \sqrt{\Lambda_T}$. Let us observe that, for every $\alpha \geq 0, \beta > \alpha$:

$$\frac{1}{2}\frac{\alpha}{\sqrt{\beta}} \leq \sqrt{\beta} - \sqrt{\beta - \alpha},$$

which is proved using $\sqrt{\beta} - \sqrt{\beta - \alpha} = \int_{\beta-\alpha}^{\beta} \frac{1}{2\sqrt{s}} ds \geq \alpha \frac{1}{2\sqrt{\beta}}$. Using the telescoping sum (with $\Lambda_0 = 0$):

$$\sum_{t=1}^T \frac{\lambda_t}{\sqrt{\Lambda_t}} \leq 2\sum_{t=1}^T \left(\sqrt{\Lambda_t} - \sqrt{\underbrace{\Lambda_t - \lambda_t}_{=\Lambda_{t-1}}}\right) = 2\sqrt{\Lambda_T} \leq 2\sqrt{\overline{\Lambda}_T},$$

we obtain the desired result.

More generally, if $\lambda_0 = 0$, there are two cases:

1. if $\forall_T \in [\![T]\!], \lambda_t = 0$ then the result is true;
2. otherwise, let $T_0 = \min\{t \in [\![T]\!] : \lambda_t > 0\}$. Using the result above, we have:

$$\sum_{t=1}^{T} \frac{\lambda_t}{\sqrt{\overline{\Lambda}_t}} = \sum_{t=T_0}^{T} \frac{\lambda_t}{\sqrt{\overline{\Lambda}_t}} \leq 2\sqrt{\overline{\Lambda}_T}.$$

□

Appendix C

Appendix of Chapter 6

C.1 (In-)Compatibility of envy-freeness

C.1.1 Envy-freeness vs. optimality certificates

We showed in Section 6.3.3 that envy-freeness is compatible with optimal predictions. To understand the differences between a certificate of envy-freeness and a certificate of optimality, let us denote by $\Pi^* = \{\pi : \exists u \text{ satisfying } (6.1) \ , \pi \in \text{argmax}_{\pi'} u(\pi')\}$ the set of potentially optimal policies. If the set of users policies approximately covers the set of potentially optimal policies Π^*, then an envy-free system is also optimal. Formally, let $D(\pi, \pi')$ such that $|u(\pi) - u(\pi')| \leq D(\pi, \pi')$. It is easy to see that if $\max_{\pi \in \Pi^*} \min_{m \in M} D(\pi, \pi^m) \leq \tilde{\epsilon}$, then ϵ-envy-freeness implies $\epsilon + \tilde{\epsilon}$-optimality.

In practice, the space of optimal policies is much larger than the number of users (for instance, there are $|\mathcal{A}|^{|\mathcal{X}|}$ optimal policies in our setting), so that auditing for envy is tractable in cases where auditing for optimality is not.

C.1.2 Envy-freeness vs. equity of exposure

We remind the definition of optimal policies with equity of exposure constraints from Section 6.3.3:

$$\textit{(equity)} \qquad \pi^{m,\text{eq}}(.|x) = \underset{\substack{p:\mathcal{A}\to[0,1] \\ \sum_a p(a)=1}}{\text{argmax}} \sum_{a \in \mathcal{A}} p(a)\rho^m(a|x)$$

$$\text{u.c. } \forall s \in [\![S]\!], \sum_{a \in \mathcal{A}_s} p(a) = \frac{\sum_{a \in \mathcal{A}_s} \rho^m(a|x)}{\sum_{a \in \mathcal{A}} \rho^m(a|x)}$$

The constraints should be ignored when $\sum_{a \in \mathcal{A}} \rho^m(a|x) = 0$.

Following Proposition 17 from Section 6.3.3, we describe here a second source of envy when using optimal policies with equity of exposure constraints. By the linearity of the optimization problem for $\pi^{m,\text{eq}}$, the policy assigns to the best item in a category the exposure of the entire category. It implies that categories with high average relevance have more exposure than categories with few but highly relevant items. Table C.1 gives an example with two users and two categories of items where both users envy each other with the optimal recommendations under equity of exposure constraints.

In some degenerate cases though, equity of exposure policies are envy-free.

		item cat. 1		item cat. 2		utilities	
(item idx)		1	2	3	4	u^1	u^2
(rewards)	ρ^1	1	0	0.8	0.7		
	ρ^2	0.8	0.7	1	0		
(policies)	$\pi^{1,\mathrm{eq}}$	0.4	0	0.6	0	**0.88**	**0.92**
	$\pi^{2,\mathrm{eq}}$	0.6	0	0.4	0	**0.92**	**0.88**

Table C.1: Example where the optimal recommendations under item-side equity of exposure constraints are not user-side fair because both users envy each other. There are 4 items, 2 item categories and 2 users. User 1 envies user 2 since $u^1(\pi^{2,\mathrm{eq}}) > u^1(\pi^{1,\mathrm{eq}})$. Also, $u^2(\pi^{1,\mathrm{eq}}) > u^2(\pi^{2,\mathrm{eq}})$.

Lemma 51. *If for all contexts $x \in \mathcal{X}$, each user $m \in [\![M]\!]$ only likes a single item category $\mathcal{A}_{s_m}$, i.e. $\forall a \in \mathcal{A} \setminus \mathcal{A}_{s_m}, \rho^m(a|x) = 0$, then the policies $(\pi^{m,\mathrm{eq}})_{m=1}^M$ are envy-free.*

Proof. We set contexts x aside to simplify notation, but the generalization is straightforward.

We actually prove a stronger result than the lemma: if each user m only likes a single item, then $(\pi^{m,\mathrm{eq}})_{m=1}^M = (\pi^{m,*})_{m=1}^M$, where $\pi^{m,*}$ is the optimal unconstrained policy for m.

Let $a_s^m = \mathrm{argmax}_{a \in \mathcal{A}_s} \rho^m(a)$ be the favorite item in category $\mathcal{A}_s$ for user m, then the optimal equity of exposure constrained policies has the following analytical expression:

$$\forall s \in S, \forall a \in \mathcal{A}_s, \quad \pi^{m,\mathrm{eq}}(a) = \mathbb{1}_{\{a=a_s^m\}} \frac{\sum\limits_{a \in \mathcal{A}_s} \rho^m(a')}{\sum\limits_{a' \in \mathcal{A}} \rho^m(a')},$$

and we thus have:

$$u^m(\pi^{m,\mathrm{eq}}) = \sum_{s \in [\![S]\!]} \rho^m(a_s^m) \frac{\sum\limits_{a \in \mathcal{A}_s} \rho^m(a)}{\sum\limits_{a \in \mathcal{A}} \rho^m(a)}.$$

If each user $m \in [\![M]\!]$ only likes a single item category $s_m \in [\![S]\!]$, i.e. $\forall a \in \mathcal{A} \setminus \mathcal{A}_{s_m}, \rho^m(a) = 0$, then $\frac{\sum\limits_{a \in \mathcal{A}_s} \rho^m(a)}{\sum\limits_{a \in \mathcal{A}} \rho^m(a)} = \mathbb{1}_{\{s=s_m\}}$.

Then $u^m(\pi^{m,\mathrm{eq}}) = \rho^m(a_{s_m}^m) = \max_{a \in \mathcal{A}} \rho^m(a)$.

Then $\pi^{m,\mathrm{eq}}$ is the optimal unconstrained policy for user m, meaning the whole system is envy-free (cf. Sec 6.3.2).

From Eq. C.1.2, we actually note that $(\pi^{m,\mathrm{eq}})_{m=1}^M = (\pi^{m,*})_{m=1}^M$ if and only if each user m equally values their favorite items in each category they like, i.e. $\forall m, \exists \kappa > 0, \forall s \in S, \rho^m(a_s^m) > 0 \Rightarrow \rho^m(a_s^m) = \kappa$.

□

C.2 Extension to group envy-freeness

We briefly discuss an extension of envy-free recommendation to groups, since most of the literature on fair machine learning focuses on systematic differences between groups. Certifying envy-freeness at the level of groups rather than individuals also relaxes the criterion because it requires less exploration. Let us assume we are given a partition G of the users into disjoint groups. For $g, g' \in G$, we define the group utility of g with respect to g' as:

$$U(g, g') = \frac{1}{|g|} \sum_{m \in g} u^m \left(\frac{1}{|g'|} \sum_{n \in g'} \pi^n \right).$$

Definition 8. *Given* $\epsilon \geq 0$*, the recommender system is* ϵ-group-envy-free *if:* $\forall g, g' \in G, \quad U(g, g') \leq U(g, g) + \epsilon$.

Group envy-freeness is equivalent to envy-freeness when each group is a singleton. When we have prior knowledge that user preferences and policies are homogeneous within each group, ϵ-envy-freeness translates to ϵ'-group envy-freeness, with $\epsilon' \approx \epsilon$, and the reciprocal is also true:

Proposition 52. *Let* $\epsilon, \tilde{\epsilon} > 0$*, and assume that for all groups and all pairs of users* m, n *in the same group* g*, we have* $\sup_{x \in \mathcal{X}} \|\pi^m(.|x) - \pi^n(.|x)\|_1 \leq \tilde{\epsilon}$ *and* $\sup_{x \in \mathcal{X}} \|\rho^m(.|x) - \rho^n(.|x)\|_1 \leq \tilde{\epsilon}$*. Then,* ϵ*-group envy-freeness implies* $(\epsilon + 4\tilde{\epsilon})$*-envy-freeness.*

The result is natural since when all groups have users with homogeneous preferences and policies, groups and users are a similar entity as regards the assessment of envy-freeness. The proof is straightforward and omitted. When groups have heterogeneous policies, the "average policy" $\frac{1}{|g|} \sum_{n \in g} \pi^n$ is uninformative because it does not represent any user's policy. Defining a notion of group utility in the general case is thus nontrivial and left for future work.

C.3 Sources of envy

In this section, we first list a few possible sources of envy in recommender systems. Then we provide the details of experiments[1] which showcase one of these sources, namely model mispecification (App. C.3.2).

C.3.1 Examples of sources of envy

Model mispecification Recommender systems often rely on strong modeling assumptions and multi-task learning, with methods such as low-rank matrix factorization Koren et al. [2009]. The limited capacity of the models (e.g., a rank that is too low) or incorrect assumptions might leave aside users with less common preference patterns. Appendix C.3.2 gives a more detailed example on two simulated recommendation tasks.

Misaligned incentives A recommender system might have incentives to recommend some items to specific users, e.g., sponsored content. Envy appears when there is a mismatch between users who like these items and users to whom they are recommended.

Measurement bias Many hybrid recommender systems rely on user interactions together with user-side data Burke [2002]. This includes side-information such as browsing history on third-party, partner websites. Envy arises in these settings if there is measurement bias Suresh and Guttag [2019], e.g., if the side information is unevenly collected for all users (e.g., browsing patterns are different across users and partners are aligned with the patterns of a user groups only).

Operational constraints Regardless of incentives, recommendations might need to obey additional constraints. As described in Proposition 17, the item-side fairness constraint of equity of exposure is an example of possible source of (user-side) envy. The user-side fairness constraint of equal utility also creates envy, as we showed in Sec. 6.5.1.

[1] For all our experiments, we used Python and a machine with Intel Xeon Gold 6230 CPUs, 2.10 GHz, 1.3 MiB of cache.

In the following, we provide the details of our experiments from Sec. 6.5.1 where we showcase examples of environments with envy based on movie and music recommendation tasks.

In these experiments, we measure envy based on the quantity:

$$\Delta^m = \max\Big(\max_{n\in[\![M]\!]} u^m(\pi^n) - u^m(\pi^m), 0\Big)$$

In line with Chevaleyre et al. [2017], we consider two ways of measuring the degree of envy:

- the average envy experienced by users: $\frac{1}{M}\sum_{m\in[\![M]\!]}\Delta^m$,
- the proportion of ϵ-envious users: $\frac{1}{M}\sum_{m\in[\![M]\!]}\mathbb{1}_{\{\Delta^m>\epsilon\}}$.

C.3.2 Setup of the experiments on envy from model mispecification

We describe in this section the details of the experiments on envy from mispecification presented in Section 6.5.1. We used Lastfm-2k [Cantador et al., 2011], a dataset from the online music service Last.fm[2] which contains real play counts of $2k$ users for $19k$ artists, and was used by Patro et al. [2020] who also study envy-freeness as a user-side fairness criterion. We filter the top $2,500$ items most listened to. Following Johnson [2014], we pre-process the raw counts with log-transformation. We split the dataset into train/validation/test sets, each including 70%/10%/20% of the user-item listening counts. We create three different splits using three random seeds. We estimate relevance scores for the whole user-item matrix using the standard matrix factorization algorithm[3] of Hu et al. [2008] trained on the train set, with hyperparameters selected on the validation set by grid search with DCG@40 as metric. The number of latent factors is chosen in $[16, 32, 64, 128]$, the regularization in $[0.01, 0.1, 1., 10.]$, and the confidence weighting parameter in $[0.1, 1., 10., 100.]$. The resulted matrix of estimated relevance scores serves as the ground truth preferences.

We also address movie recommendation using the MovieLens-1M dataset Harper and Konstan [2015], which contains 1 million ratings on a 5-star scale from approximately 6000 users and 4000 movies. We extract a 2000×2500 user $\times$ items matrix, keeping users and items with the most rating. We transform MovieLens ratings into an implicit feedback dataset similar to Last.fm. Since setting ratings < 3 are usually considered as negative Wang et al. [2018], we set ratings < 3 to zero, resulting in a dataset with preference values among $\{0, 3, 3.5, 4, 4.5, 5\}$. We then use the same algorithm as for Last.fm to obtain relevance scores that we use to simulate ground truth preferences.

We then simulate a recommender system's estimation of preferences using low-rank matrix completion[4] Bell and Sejnowski [1995] on a training sample of 70% of the whole "ground truth" preferences, with hyperparameter selection on a 10% validation sample. Here, the regularization is chosen in $[0.001, 0.01, 0.1, 1.]$, and the confidence weighting parameter in $[0.1, 1., 10., 100.]$. The estimated preference scores are given as input to the recommendation policies.

The recommendation policies we consider are softmax distributions over the predicted scores with fixed inverse temperature. These policies recommend a single item, drawn from the softmax distribution.

We generate binary rewards using a Bernoulli distribution with expectation given by our ground truth. We consider no context in these experiments, so that the policies and rewards only depend on the user and the item.

Figure 6.2 in Sec. 6.5.1 was generated by varying the number of latent factors in the recommender system's preference estimation model. For each number of latent factors in the range

[2] http://www.lastfm.com
[3] Using the Python library Implicit: https://github.com/benfred/implicit (MIT License).
[4] Using the implementation of https://github.com/gbolmier/funk-svd (MIT License).

$[1, 2, 4, 8, 16, 32, 64, 128, 256]$, a new model was trained on the train set with hyperparameter selection on the validation set. The degrees of envy are measured on the whole ground truth preference matrix.

C.3.3 Envy from equal user utility constraints

We provide the full details of the experiments on envy from equal user utility presented in Sec. 6.5.1 from the main paper. The goal of these experiments is to show that in contrast to envy-freeness, enforcing equal user utility (EUU) degrades user satisfaction and creates envy between users. We remind from Sec. 6.3.2 that the fairness constraint of EUU is defined as:

$$\forall m, n \in [\![M]\!], u^m(\pi^m) = u^n(\pi^n),$$

or equivalently:

$$\forall m \in [\![M]\!], u^m(\pi^m) = \frac{1}{M} \sum_{n \in [\![M]\!]} u^n(\pi^n).$$

Equal user utility is enforced by adding a penalty to the maximization of user utilities. Optimal EUU policies are found by maximizing the following concave objective function, where the parameter $b > 0$ controls the strength of the penalty:

$$\textit{(EUU)} \quad \pi_b^{\text{euu}} = \underset{\substack{p:\mathcal{A}\to[0,1]^M \\ \forall m, \sum_a p^m(a)=1}}{\operatorname{argmax}} \sum_{m \in [\![M]\!]} u^m(p^m) - b\sqrt{D(p)}$$

$$\text{with} \quad D(p) = \sum_{m \in [\![M]\!]} \left(u^m(p^m) - \frac{1}{M} \sum_{n \in [\![M]\!]} u^n(p^n) \right)^2.$$

We infer EUU policies using the Frank-Wolfe algorithm Frank and Wolfe [1956] with the ground truth preferences given as input. The parameter of the penalty is set to $b = 50$. We also generate the unconstrained optimal policies (OPT) based on the ground truth (recall that these are $u^m(\pi^{m,*}) = \max_\pi u^m(\pi) \geq u^m(\pi^{n,*})$).

A comparison of EUU and OPT is provided in Table 6.1 in Sec. 6.5.1, with the following evaluation measures : total utility (higher is better), average envy and proportion of 0.05-envious users (lower is better). The results on both dataset confirm the claim that enforcing EUU penalties deteriorates total utility and creates envy between users, while illustrating the known property that OPT policies are compatible with envy-freeness.

C.4 OCEF experiments

C.4.1 Bandit experiments

We performed experiments on toy bandit environments to assess the performance of our algorithm OCEF on various configurations, which were also considered in Jamieson and Nowak [2014]. The four bandits instances have 10 arms. They are Bernoulli variables with means equal to

1) $\mu_0 = 0.6$ and $\mu_k = 0.3$ for $k \in [\![9]\!]$,
2) $\mu_0 = 0.3$, $\mu_1 = 0.6$ and $\mu_k = 0.3$ for $k = 2..9$,
3) $\mu_k = 0.7 - 0.7 * \left(\frac{k}{10}\right)^{0.6}, k = 0, ..., 9$, and the baseline is μ_0,

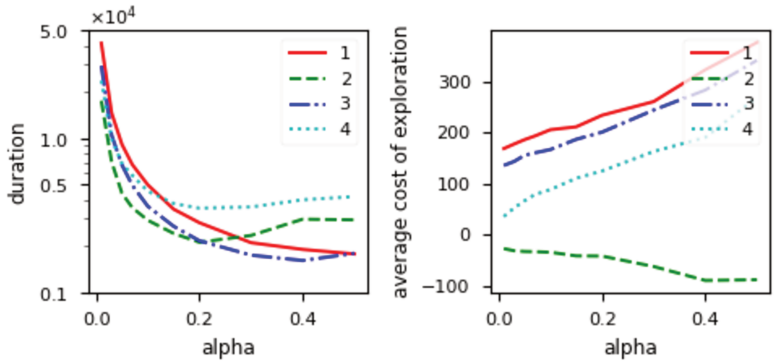


Figure C.1: Effect of the conservative exploration parameter α on the duration and cost of auditing on Bandit experiments.

4) same as 3), but permuting μ_0 and μ_1.

Fig. C.1 shows the result of applying OCEF on the various configurations, where we set $\delta = \epsilon = 0.05$, $\omega = 0.99$ [5] and report results averaged over 100 trials. We observe clear tendencies similar to those presented in Section 6.5.2, although the exact sweet spots in terms of α depends on the specific configuration.

The cost of exploration follows similar patterns as in in Section 6.5.2. In Prob. 2, the baseline has the worst performance, so exploration is beneficial to the user and the cost is negative. On the other hand, for instance in Prob. 4, the cost is close to 0 when α is very small and increases with α. It is the case where the baseline is not the best arm but is close to it, and there are many bad arms. When the algorithm is very conservative, bad arms are discarded rapidly thanks to the good estimation of the baseline performance. In this "low-cost" regime however, the audit is significantly longer.

We show additional results when varying δ in Figure C.2. Results are averaged over 100 simulations and the conservative exploration parameter is set to $\alpha = 0.05$. The duration decreases as δ increases, i.e. a lower confidence certificate requires fewer samples per user. The duration for Problem 1 is longer than for the other instances. This is because with α set to 0.05 and the baseline mean being much higher than non-baseline arms, the conservative constraint 6.3 enforces many pulls of the baseline, since each exploration round is very costly. As a consequence, too little data is collected on the non-baseline arms to conclude that they are below $\mu_0 + \epsilon$. Since all non-baseline arms have equal means, the size of the active set remains the same for a long time, while in Problem 3, where the baseline is also the best arm, arms are eliminated one at a time.

We show how OCEF scales with the number of arms in Figure C.3, for fixed values $\alpha = \delta = \epsilon = 0.05$. We set $K_{\max} = 100$ and define 4 instances as in the list above, except that $K = K_{\max}$ instead of $K = 9$. We run OCEF on the instances $\mu_{0:K'}$ and vary the value of $K' \leq K_{\max}$. The duration increases for all problems, and the slope depends on the gaps between μ_0 and the μ_k.

C.4.2 Setup of the MovieLens and Last.fm experiments

We now provide additional details on the experimental evaluation of OCEF on MovieLens and Last.fm presented in Sec. 6.5.2. The protocole to generate the recommendation task is the same as the one described in App. C.3 for the experiments on sources of envy. The policies are softmax distributions over scores predicted by the matrix factorization model with a number of factors equal to 48.

[5] Following [Jamieson et al., 2014] who recommend ω close to 1.

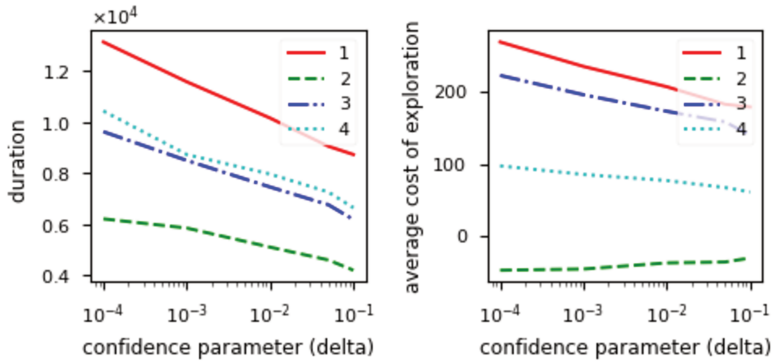


Figure C.2: Effect of the confidence parameter δ on the duration and cost on 4 different bandit instances.

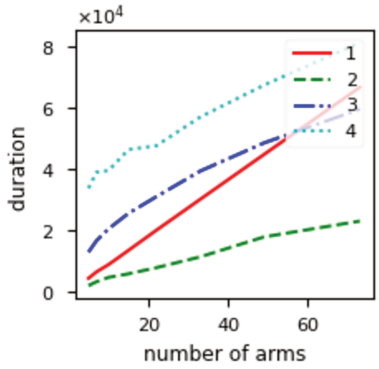


Figure C.3: Effect of the number of arms on the duration on 4 different bandit instances.

In these experiments, the auditor interacts with the audited users. Rewards are drawn from Bernoulli distributions with expectation equal to the ground truth preferences.

Two recommendation policies are audited. The first one is a softmax with inverse temperature equal to 5. Since the inverse temperature is small, the softmax distribution is closer to random, which means users get more similar recommendations: the recommender system is thus envy-free. The second one is a softmax with inverse temperature equal to 15. With higher inverse temperature, the distribution is more peaked, which exacerbates differences between policies. Since the model with 48 factors is mispecified (see Sec.6.5.1), envy is visible.

C.5 Proofs

C.5.1 Theoretical results

C.5.1.1 Useful lemmas

Recall that OCEF considers a single audited group m, therefore we do not use superscripts m in the following (e.g., $\mu_k, r_t ...$).

The algorithm relies on valid confidence intervals. As in Jamieson et al. [2014], we use anytime bounds inspired by the law of the iterated algorithm (LIL), and a union bound.

We say that a random variable is σ-subgaussian if it is subgaussian with variance proxy σ^2. Since we assume the rewards for each user are bounded, more precisely $r_t \in [0, 1]$, they are $\frac{1}{2}$-subgaussian.

Throughout the paper, we assume that rewards for each user are independent conditionally to

the arm played.

Lemma 53. *Let $\delta \in (0,1)$. Assume the rewards are σ-subgaussian.*

Let $\omega \in (0,1)$, $\quad \theta = \log(1+\omega)\left(\frac{\omega\delta}{2(2+\omega)}\right)^{\frac{1}{1+\omega}}$.

$$\text{Let} \quad N_k(t) = \sum_{s=1}^{t} \mathbb{1}_{\{k_s=k\}} \qquad\qquad \widehat{\mu}_k(t) = \frac{\sum_{s=1}^{t} r_s \mathbb{1}_{\{k_s=k\}}}{N_k(t)}$$

$$\beta_k(t) = \sqrt{\frac{2\sigma^2(1+\sqrt{\omega})^2(1+\omega)}{N_k(t)}} \times \sqrt{\log\left(\frac{2(K+1)}{\theta}\log((1+\omega)N_k(t))\right)}$$

$$\underline{\mu}_k(t) = \widehat{\mu}_k(t) - \beta_k(t) \qquad\qquad \overline{\mu}_k(t) = \widehat{\mu}_k(t) + \beta_k(t)$$

Then,

$$\mathbb{P}\left[\forall t > 0, \forall k \in [\![K]\!], \quad \mu_k \in [\underline{\mu}_k(t); \overline{\mu}_k(t)]\right] \geq 1 - \frac{\delta}{2}.$$

Notice that the choice of θ makes sure that β_k is well defined as long as $N_k(t) > 0$. We use the convention that when $N_k(t) = 0$, $\beta_k(t)$ is strictly larger than when $N_k(t) = 1$ to ensure β_k is strictly decreasing with N_k. Also, when $N_k(t) = 0$, we set $\widehat{\mu}_k(t) = 0$.

Following Garcelon et al. [2020a], our lower bound on the conservative constraint relies on Freedman's martingale inequality Freedman [1975].

Lemma 54. *Assume all rewards are σ-subgaussian. Let $A_t = \{s \leq t : k_s \neq 0\}$ be the number of times a non-baseline arm $k \neq 0$ has been pulled up to time t. Let $\phi(t) = \sigma\sqrt{2|A_{t-1}|\log\left(\frac{6|A_{t-1}|^2}{\delta}\right)} + \frac{2}{3}\log\left(\frac{6|A_{t-1}|^2}{\delta}\right)$.*

Then, $\quad \forall \delta > 0,$

$$\mathbb{P}\left[\forall t > 0, \left|\sum_{s \in A_{t-1}} (\mu_{k_s} - r_s)\right| \leq \phi(t)\right] \geq 1 - \frac{\delta}{2}.$$

As in Lemma 53, we use the convention $\phi(t) = 0$ when $|A_{t-1}| = 0$.

Lemma 55. *Let $\delta \in (0,1)$.*

Let $\Phi(t) = \min\left(\sum_{k=1}^{K} \beta_k(t-1)N_k(t-1), \phi(t)\right)$, with $\phi(t)$ defined in Lemma 54. Let $\mathcal{E}$ be the event under which all confidence intervals are valid, i.e.:

$$\begin{aligned}
&\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 \qquad \text{with} \\
&\mathcal{E}_1 = \left\{\forall k \in \{0, \dots, K\}, \forall t > 0, \mu_k(t) \in [\underline{\mu}_k(t); \overline{\mu}_k(t)\right\} \\
&\mathcal{E}_2 = \left\{\forall t > 0, \left|\sum_{s \in A_{t-1}} (\mu_{k_s} - r_s)\right| \leq \Phi(t)\right\}.
\end{aligned}$$

Then $\mathbb{P}[\mathcal{E}] \geq 1 - \delta$.

Proof. By Lemma 53, $\mathbb{P}[\mathcal{E}_1] \geq 1 - \frac{\delta}{2}$. By the lemma above, with probability $1 - \frac{\delta}{2}$, we have for all $t > 0$, $\left|\sum_{s \in A_{t-1}} (\mu_{k_s} - r_s)\right| \leq \phi(t)$.

Then, notice that

$$\Big|\sum_{s\in A_{t-1}}(\mu_{k_s}-r_s)\Big| = \Big|\sum_{k=1}^{K} N_k(t-1)(\mu_k-\widehat{\mu}_k(t-1))\Big|.$$

Hence under $\mathcal{E}_1$ we also have:

$$\Big|\sum_{s\in A_{t-1}}(\mu_{k_s}-r_s)\Big| \leq \sum_{k=1}^{K} N_k(t-1)\beta_k(t-1).$$

Therefore,

$$\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 = \mathcal{E}_1 \cap \Big\{\big|\sum_{s\in A_{t-1}}(\mu_{k_s}-r_s)\big| \leq \phi(t)\Big\},$$

and thus, by a union bound, we have: $\mathbb{P}[\mathcal{E}] \geq 1-\delta$. □

C.5.1.2 Theorems

We now provide our complete theoretical guarantees for correctness (Theorem 56), duration (Theorem 57) and cost (Theorem 58), which we then prove in App. C.5.2 and C.5.3. From these results, we derive Theorem 18 in the main paper, which we prove in App. C.5.4.

Theorem 56 (Correctness)**.** *With probability at least* $1-\delta$*:*

1. *OCEF satisfies the safety constraint* (6.3) *at every time step,*
2. *if OCEF outputs* ϵ-`no-envy` *then the user* m *is not* ϵ*-envious, and if it outputs* `envy`*, then* m *is envious.*

We denote $\log^+(.) = \max(1, \log(.))$.

Theorem 57 (Duration)**.** *Let* $\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k)$*,* $\delta \in (0,1)$*,* $\theta = \log(2)\sqrt{\frac{\delta}{6}}$*, and*

$$\forall k \neq 0, \quad H_k = 1 + \frac{64}{\eta_k^2}\log\left(\frac{2(K+1)\log^+\left(\frac{128(K+1)}{\theta\eta_k^2}\right)}{\theta}\right),$$

$$H_0 = \max\left(\max_{k\in[\![K]\!]} H_k, \frac{6K+2}{\alpha\mu_0} + \sum_{k=1}^{K}\frac{256\log\left(\frac{2(K+1)\log(2H_k)}{\theta}\right)}{\alpha\mu_0\eta_k}\right).$$

With probability at least $1-\delta$*, OCEF stops in at most* τ *steps, with*

$$\tau \leq \sum_{k=0}^{K} H_k\,.$$

Finally, we define the *cost of exploration* as the potential reward lost because of exploration actions, in our case the cumulative reward lost, on average over users in the group:

$$C_t = t\mu_0 - \sum_{s=1}^{t}\mu_{k_s}\,. \tag{C.1}$$

In the worst case, the following bound holds:

Theorem 58 (Cost of exploration). *Under the assumptions and notation of Theorem 57, let τ be the time step where OCEF stops. With probability $1-\delta$, we have:*

$$C_\tau \leq \sum_{k:\mu_k<\mu_0} (\mu_0 - \mu_k) H_k$$

Certification of the exact criterion for all users The audit of the full system for the exact envy-freeness criterion consists in running OCEF for every user. Since we are making multiple tests, we need to use a tighter confidence parameter for each user so that the confidence intervals simultaneously hold for all users.

Corollary 59 (Online certification). *With probability at least $1-\delta$, running OCEF simultaneously for all M users, each with confidence parameter $\delta' = \frac{\delta}{M}$, we have:*

1. *for all $m \in [M]$ OCEF satisfies the constraints (6.3),*
2. *all users for which OCEF returns ϵ-NO ENVY are not ϵ-envious of any other users, and all users for which OCEF returns ENVY are envious of another user.*
3. *For every user, the bounds on the duration of the experiment and the cost of exploration given by Theorems 57 and 58 (using δ/M instead of δ) are simultaneously valid.*

For the certification of the probabilistic envy-freeness criterion, we refer to Theorem 19 in the main paper, which we prove in App. C.5.5.

C.5.2 Proof of Theorem 56

Proof. We assume that event $\mathcal{E}$ holds true. Then all confidence intervals are valid, i.e., for all $k = 0, ..., K$, $\underline{\mu}_k(t) \leq \mu_k \leq \overline{\mu}_k(t)$, and $\sum_{s\in A_{t-1}} \mu_{k_s} \geq \sum_{s\in A_{t-1}} r_s - \Phi(t)$.

Let Z_t be the safety budget, defined as $Z_t = \sum_{s=1}^{t} \mu_{k_s} - (1-\alpha)\mu_0 t$, so that the conservative constraint (6.3) is equivalent to $\forall t, Z_t \geq 0$. We have $Z_t = \sum_{s\in A_{t-1}} \mu_{k_s} + \mu_{k_t} + (N_0(t-1) - (1-\alpha)t)\mu_0$. Therefore, ξ_t (eq. (6.4)) is a lower bound on the safety budget Z_t if ℓ_t is played. By construction of the algorithm, the safety constraint (6.3) is immediately satisfied since a pull that could violate it is not permitted.

By the validity of confidence intervals under $\mathcal{E}$, if OCEF stops because of the first condition, then $\exists k, \mu_k > \mu_0$. Therefore 0 is not ϵ-envious of k and OCEF is correct.

If OCEF stops because of the second condition, i.e., $\forall k, \overline{\mu}_k(t) \leq \underline{\mu}_0(t) + \epsilon$, then $\forall k, \mu_k \leq \mu_0 + \epsilon$. Therefore 0 is not envious and OCEF is correct.

Since $\mathbb{P}[\mathcal{E}] \geq 1-\delta$, OCEF satisfies the safety constraint and is correct with probability $\geq 1-\delta$. □

C.5.3 Proofs of Theorem 57 and Theorem 58

Notation For conciseness, we use $\tilde{K} = K+1$, and

$$\psi_k(t) = 2\sigma^2(1+\sqrt{\omega})^2(1+\omega)\log\left(\frac{2\tilde{K}}{\theta}\log((1+\omega)N_k(t))\right),$$

$$\text{so that } \beta_k(t) = \sqrt{\frac{\psi_k(t)}{N_k(t)}}.$$

We shall also use $\Gamma_\omega = 2\sigma^2(1+\sqrt{\omega})^2(1+\omega)$. We use the convention $\psi_k(t) = 0$ when $N_k(t) = 0$, and set $\beta_k(t)$ to some value strictly larger than when $N_k(t) = 1$.

We remind that $\omega \in (0,1), \quad \theta = \log(1+\omega)\left(\frac{\omega\delta}{2(2+\omega)}\right)^{\frac{1}{1+\omega}}$ and $\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k)$. We denote by $\eta_{\min} = \min_{k\in[\![K]\!]} \eta_k$.

Finally, we notice that under event $\mathcal{E}$ (as defined in Sec. C.5.1.1), we have for all $k \in \{0,\dots,K\}$ and all t:

$$\mu_k + 2\beta_k(t) \geq \overline{\mu}_k(t) \geq \mu_k \geq \underline{\mu}_k(t) \geq \mu_k - 2\beta_k(t). \tag{C.2}$$

Lemma 60. *Under event $\mathcal{E}$, for every $k \in [\![K]\!]$, if k is pulled at round t, then $4\beta_k(t) \geq \eta_k$.*

Proof of Lemma 60. Since k is pulled at t, the two following inequalities hold:

$$\overline{\mu}_k(t-1) > \underline{\mu}_0(t-1) + \epsilon \tag{C.3}$$

$$\underline{\mu}_k(t-1) \leq \overline{\mu}_0(t-1) \tag{C.4}$$

We prove them by contradiction. If (C.3) does not hold, then k should be discarded from the active set at time $t-1$, and therefore cannot be pulled at t. Likewise, if (C.4) does not hold, then the algorithm stops at $t-1$, so k cannot be pulled at t.

Using (C.3) and (C.2), we have:

$$\mu_k + 2\beta_k(t-1) \geq \overline{\mu}_k(t-1) > \underline{\mu}_0(t-1) + \epsilon \geq \mu_0 - 2\beta_0(t-1) + \epsilon.$$

Since 0 was not pulled at time t, we also have $\beta_0(t-1) \leq \beta_k(t-1)$, hence $4\beta_k(t-1) \geq \mu_0 + \epsilon - \mu_k$.

Using (C.4) and (C.2) we have $\mu_k - 2\beta_k(t) \leq \mu_0 + 2\beta_0(t)$ and since $\beta_0(t) \leq \beta_k(t)$, we obtain $4\beta_k(t-1) \geq \mu_k - \mu_0$.

□

In the following lemma, we recall that we denote $\log^+(.) = \max(1, \log(.))$.

Lemma 61. *Under event $\mathcal{E}$, $\forall \tau > 0, \forall k \in [\![K]\!]$, we have*

$$N_k(\tau) \leq H_k \qquad \textit{with}$$
$$H_k = 1 + \frac{32\sigma^2(1+\sqrt{\omega})^2(1+\omega)}{\eta_k^2} \times \log\left(\frac{2(K+1)\log^+\left(\frac{64(K+1)\sigma^2(1+\sqrt{\omega})^2(1+\omega)^2}{\theta\eta_k^2}\right)}{\theta}\right)$$

Proof. Let $\tau > 0$, $k \in [\![K]\!]$, and let $t \leq \tau$ be last time step before τ at which k was pulled. If such a t does not exist, then $N_k(\tau) = 0$ and the result holds. In all cases, we have $N_k(t) = N_k(\tau)$.

We consider $t > 0$ from now on.

By Lemma 60, we have $4\beta_k(t-1) \geq \eta_k$, and thus $N_k(t-1) \leq \frac{16\psi_k(t-1)}{\eta_k^2}$, which writes, if $N_k(t) > 0$:

$$\begin{aligned} N_k(t-1) &\leq \frac{16\psi_k(t-1)}{\eta_k^2} \\ &\leq \frac{16\Gamma_\omega}{\eta_k^2}\log\left(\frac{2\tilde{K}}{\theta}\log\left((1+\omega)N_k(t-1)\right)\right). \end{aligned} \tag{C.5}$$

Using $\frac{1}{t}\log\left(\frac{\log((1+\omega)t)}{\Omega}\right) \geq c \Rightarrow t \leq \frac{1}{c}\log\left(\frac{\log((1+\omega)/c\Omega)}{\Omega}\right)$ (see Equation (1) in Jamieson et al. [2014]) with $\Omega = \frac{\theta}{2\tilde{K}}$ and $c = \frac{\eta_k^2}{16\Gamma_\omega}$, we obtain

$$N_k(t-1) \leq \frac{16\Gamma_\omega}{\eta_k^2}\log\left(\frac{2\tilde{K}}{\theta}\log\left(\frac{(1+\omega)32\tilde{K}\Gamma_\omega}{\theta\eta_k^2}\right)\right)$$

Since $N_k(t) = N_k(t-1) + 1$, using $\log^+$ instead of log inside to deal with the case $N_k(t-1) = 0$ gives the desired result. □

Lemma 62. *Under event $\mathcal{E}$, at every time step τ, we have*

$$N_0(\tau) \leq \max\left(\max_{k\in[\![K]\!]} H_k, \frac{6K+2}{\alpha\mu_0} + \sum_{k=1}^{K} \frac{64\sigma^2(1+\sqrt{\omega})^2(1+\omega)\log\left(\frac{2(K+1)\log((1+\omega)H_k)}{\theta}\right)}{\alpha\mu_0\eta_k}\right)$$

Proof. Let $\tau > 0$ and $t \leq \tau$ the last time 0 was pulled before τ. We assume $t > 0$.

Case 1: 0 was pulled because $\beta_0(t-1) > \min_{k\in[\![K]\!]} \beta_k(t-1)$.
Then $N_0(\tau) = N_0(t-1) + 1 \leq 1 + \max_{k\neq 0} N_k(t-1)$.
By lemma 60, we thus have $N_0(\tau) \leq \max_{k\in[\![K]\!]} H_k$.

Case 2: 0 was pulled because $\xi_t < 0$. Here the proof follows similar steps as that of Theorem 5 in Wu et al. [2016].

$$\sum_{s\in A_{t-1}} r_s - \Phi(t) + \underline{\mu}_{\ell_t}(t-1) + (N_0(t-1) - (1-\alpha)t)\overline{\mu}_0(t-1) < 0$$

We drop $\underline{\mu}_{\ell_t}(t-1)$, replace t by $\sum_{k=0}^{K} N_k(t-1) + 1$ and rearrange terms to obtain:

$$\alpha N_0(t-1)\overline{\mu}_0(t-1) \leq (1-\alpha)\overline{\mu}_0(t-1) + (1-\alpha)\sum_{k=1}^{K} N_k(t-1)\overline{\mu}_0(t-1) - \sum_{s\in A_{t-1}} r_s + \Phi(t) \tag{C.6}$$

Since we have $\beta_0(t-1) \leq \beta_k(t-1)$ (otherwise we would be in case 1), and $A_{t-1} = \sum_{k=1}^{K} N_k(t-1)$, we bound the the sum over arms in (C.6):

$$\begin{aligned}\sum_{k=1}^{K} N_k(t-1)\overline{\mu}_0(t-1) &\leq \sum_{k=1}^{K} N_k(t-1)(\mu_0 + 2\beta_0(t-1)) \\ &\leq \sum_{k=1}^{K} N_k(t-1)(\mu_0 + 2\beta_k(t-1)) \\ &= \sum_{s\in A_{t-1}} \mu_0 + \sum_{k=1}^{K} 2\beta_k(t-1)N_k(t-1).\end{aligned}$$

Using Lemma 55, we also bound $-\sum_{s\in A_{t-1}} r_s \geq \sum_{s\in A_{t-1}} \mu_s + \Phi(t)$ (under $\mathcal{E}$).
Plugging this into (C.6) gives:

$$\alpha N_0(t-1)\overline{\mu}_0(t-1) \leq (1-\alpha)\overline{\mu}_0(t-1) + 2(1-\alpha)\sum_{k=1}^{K} N_k(t-1)\beta_k(t-1) + \sum_{s\in A_{t-1}} ((1-\alpha)\mu_0 - \mu_{k_s}) + 2\Phi(t).$$

Recall that $\Phi(t) = \min(\sum_{k=1}^{K} N_k(t-1)\beta_k(t-1), \phi(t))$, and therefore $\Phi(t) \leq \sum_{k=1}^{K} N_k(t-1)\beta_k(t-1)$.

Using $\mu_0 - \mu_{k_s} \leq \eta_{k_s}$ and $\sum_{s \in A_{t-1}} \eta_{k_s} = \sum_{k=1}^{K} N_k(t-1)\eta_k$, we obtain:

$$\begin{aligned}\alpha N_0(t-1)\overline{\mu}_0(t-1) \leq (1-\alpha)\overline{\mu}_0(t-1) \\ + \sum_{k=1}^{K} \Big((\eta_k - \alpha\mu_0) N_k(t-1) \\ + 4\sqrt{\Psi_k(t-1) N_k(t-1)} \Big).\end{aligned}$$

We bound $f_k := (\eta_k - \alpha\mu_0)N_k(t-1) + 4\sqrt{\Psi_k(t-1)N_k(t-1)}$.

Since (C.5) $N_k(t-1) \leq \frac{16\psi_k(t-1)}{\eta_k^2} + 1$, and $\eta_k - \alpha\mu_0 \leq \eta_k$, we have

$$f_k \leq \frac{16\psi_k(t-1)}{\eta_k} + \eta_k + 4\sqrt{\frac{16\psi_k(t-1)^2}{\eta_k^2} + \psi_k(t-1)}$$

Using $\sqrt{(\frac{x}{z})^2 + x} \leq \frac{x}{z} + \frac{z}{2}$ for $x \geq 0, z > 0$, with $x = 4\psi_k(t-1)$ and $z = \eta_k$, we obtain:

$$\begin{aligned} f_k &\leq \frac{16\psi_k(t-1)}{\eta_k} + \frac{16\psi_k(t-1)}{\eta_k} + 3\eta_k \\ &\leq \frac{32\psi_k(t-1)}{\eta_k} + 3\eta_k. \end{aligned}$$

Using $\psi_k(t-1) = \Gamma_\omega \log\left(\frac{2\tilde{K}}{\theta}\log((1+\omega)N_k(t-1))\right)$ if $N_k(t) > 0$ and $N_k(t-1) \leq H_k$ by Lemma 61, we obtain

$$f_k \leq \frac{32\Gamma_\omega}{\eta_k} \log\left(\frac{2\tilde{K}}{\theta}\log\left((1+\omega)H_k\right)\right) + 3\eta_k\,.$$

This bound is also valid when $N_k(t) > 0$.

Going back to (C.6), and since $\mu_0 \leq \overline{\mu}_0(t-1)$ under $\mathcal{E}$, we have (notice $\eta_k \leq 2$ since $\mu_k \in [0,1]$ and $\epsilon \in [0,1]$):

$$\begin{aligned}\alpha N_0(t-1)\mu_0 \leq &(1-\alpha)\overline{\mu}_0(t-1) + 6K \\ &+ \sum_{k=1}^{K} \frac{32\Gamma_\omega}{\eta_k} \log\left(\frac{2\tilde{K}}{\theta}\log\left((1+\omega)H_k\right)\right).\end{aligned} \tag{C.7}$$

To bound the first term of the right-hand side, let us first notice that the final result holds if $N_0(t-1) \leq \max_{k \in [\![K]\!]} H_k$. So we can assume $N_0(t-1) > \max_{k \in [\![K]\!]} H_k$ from now on. By the definition of the H_ks (see above (C.5)), this implies $N_0(t-1) > \frac{16\psi_0(t-1)}{\eta_{\min}^2}$, which in turn implies $4\beta_0(t-1) \leq \eta_{\min}$.

We thus use $\overline{\mu}_0(t-1) \leq \mu_0 + 2\beta_0(t-1) \leq \mu_0 + \frac{\eta_{\min}}{2} \leq 2$, which gives the final result.

The result directly follows from (C.7). □

The proof of Theorem 57 follows from $\tau = \sum_{k=1}^{K} N_k(\tau) + N_0(\tau)$, by setting $\omega = 1$ for ease of reading, and $\sigma = \frac{1}{2}$ since Bernoulli variables are $\frac{1}{2}$-subgaussian (using Hoeffding's inequality Hoeffding [1963]).

We prove Corollary 59 from Theorem 56 and Theorem 57.

We now prove Theorem 58:

Proof. Since playing the baseline is neutral in the cost of exploration, it can be re-written as:

$$C_\tau = \sum_{k=1}^{K} (\mu_0 - \mu_k) N_k(\tau) \leq \sum_{k:\mu_k<\mu_0} (\mu_0 - \mu_k) N_k(\tau),$$

where τ is the time the algorithm stops. Using Lemma 61 to upper bound $N_k(\tau)$, we obtain the result. □

Corollary 59 simply follows from the fact that by applying each algorithm with confidence δ/M, the confidence intervals are then simultaneously valid for all users with probability $1-\delta$, so all the correctness/duration/cost proofs holds for all groups simultaneously with probability $1-\delta$. For the statistical guarantees on certifying the probabilistic envy-freeness criterion, we provide the proof of Theorem 19 in App. C.5.5.

C.5.4 Proof of Theorem 18

Theorems 56, 57, and 58 are summarized in Theorem 18 in the main paper. We restate Theorem 18 and prove it below:

Theorem. *Let* $\epsilon \in (0,1]$, $\alpha \in (0,1], \delta \in (0, \frac{1}{2})$ *and*

$$\eta_k = \max(\mu_k - \mu_0, \mu_0 + \epsilon - \mu_k) \text{ and } h_k = \max(1, \frac{1}{\eta_k}).$$

Using $\underline{\mu}, \overline{\mu}$ *and* Φ *given in Lemmas 53 and 55, OCEF achieves the following guarantees with probability at least* $1-\delta$*:*

- *OCEF is correct and satisfies the conservative constraint on the recommendation performance (6.3).*
- *The duration is in* $O\Big(\sum_{k=1}^{K} \frac{h_k \log\big(\frac{K \log(\frac{K h_k}{\delta \eta_k})}{\delta}\big)}{\min(\alpha\mu_0, \eta_k)}\Big)$.
- *The cost is in* $O\Big(\sum_{k:\mu_k<\mu_0} \frac{(\mu_0-\mu_k)h_k}{\eta_k} \log\big(\frac{K \log(\frac{K h_k}{\delta \eta_k})}{\delta}\big)\Big)$.

Proof. With $\delta \in (0, \frac{1}{2})$, let $\theta = \log(2)\sqrt{\frac{\delta}{6}}$. Then Theorems 57 and 58 hold for (δ, θ).

Duration We first show that:

$$H_k = O\Big(\frac{h_k}{\eta_k} \log\big(\frac{K h_k}{\delta \eta_k}\big)\Big), \tag{C.8}$$

$$\log(H_k) = O\Big(\log\big(\frac{K h_k}{\delta \eta_k}\big)\Big). \tag{C.9}$$

Recall from Th. 57 that H_k is defined as:

$$H_k = 1 + \frac{64}{\eta_k^2} \log\left(\frac{2(K+1)\log^+\big(\frac{256(K+1)}{\theta \eta_k^2}\big)}{\theta}\right)$$

We replace the $\log^+$ term from Th. 57 by $\log\big(\frac{K h_k}{\delta \eta_k}\big) > 0$, because $\frac{K h_k}{\delta} \geq 3$ as soon as $K \geq 2$.

We thus have

$$H_k = 1 + O\Big(\frac{1}{\eta_k^2} \underbrace{\log\Big(\frac{K}{\delta}\log\big(\frac{Kh_k}{\delta\eta_k}\big)\Big)}_{=B}\Big), \tag{C.10}$$

Using $\log(x) \le x \Rightarrow x\log(x) \le x^2$ for $x \ge 0$, and the fact that $\log\big(\frac{Kh_k}{\delta\eta_k}\big) \ge 0$, we have:

$$B \le \log\Big(\frac{Kh_k}{\delta\eta_k}\log\big(\frac{Kh_k}{\delta\eta_k}\big)\Big) \le 2\log\big(\frac{Kh_k}{\delta\eta_k}\big).$$

Since $1 + \frac{1}{\eta_k^2} \le 2\frac{h_k}{\eta_k}$, eq. (C.8) holds.

We now bound $\log(H_k)$:

$$\begin{aligned}\log(H_k) &= O\Big(\log\Big(\frac{h_k}{\eta_k}\log\big(\frac{Kh_k}{\delta\eta_k}\big)\Big)\Big)\\ &= O\Big(\log\Big(\frac{Kh_k}{\delta\eta_k}\log\big(\frac{Kh_k}{\delta\eta_k}\big)\Big)\Big)\\ &= O\Big(\log\Big(\frac{Kh_k}{\delta\eta_k}\Big)\Big)\end{aligned}$$

where the last line comes from $\frac{Kh_k}{\delta\eta_k}\log\big(\frac{Kh_k}{\delta\eta_k}\big) \le \big(\frac{Kh_k}{\delta\eta_k}\big)^2$.

Therefore, eq. (C.9) holds.

Now, let

$$\Gamma = \frac{6K+2}{\alpha\mu_0} + \sum_{k=1}^{K} \frac{128\log\Big(\frac{2(K+1)\log(2H_k)}{\theta}\Big)}{\alpha\mu_0\eta_k},$$

so that $H_0 = \max(\max_{k\in[\![K]\!]} H_k, \Gamma)$.

We have:

$$\begin{aligned}\Gamma &= O\Big(\frac{K}{\alpha\mu_0} + \sum_{k=1}^{K}\frac{h_k}{\alpha\mu_0}\log\big(\frac{K\log(H_k)}{\delta}\big)\Big)\\ &= O\Big(\sum_{k=1}^{K}\frac{h_k}{\alpha\mu_0}\log\big(\frac{K\log(H_k)}{\delta}\big)\Big)\\ &= O\Big(\sum_{k=1}^{K}\frac{h_k}{\alpha\mu_0}\log\big(\frac{K\log(\frac{Kh_k}{\delta\eta_k})}{\delta}\big)\Big),\end{aligned}$$

where the second equality is because $K = \sum_{k=1}^{K} 1 \le \sum_{k=1}^{K} h_k$, and the last equality uses eq. (C.9). Combining this with eq. (C.8) we have:

$$H_0 = O\Big(\sum_{k=1}^{K}\frac{h_k}{\min(\alpha\mu_0,\eta_k)}\log\big(\frac{K\log(\frac{Kh_k}{\delta\eta_k})}{\delta}\big)\Big).$$

Using eq. (C.8) again to bound $\tau = H_0 + \sum_{k=1}^{K} H_k$, , we get the desired bound for duration.

Cost For the cost, we remind the bound given in Th. 58:

$$\begin{aligned} C_\tau &\leq \sum_{k:\mu_k<\mu_0} (\mu_0 - \mu_k) H_k \\ &= O\Big(\sum_{k:\mu_k<\mu_0} \frac{(\mu_0-\mu_k)h_k}{\eta_k} \log\Big(\frac{K}{\delta}\log\Big(\frac{Kh_k}{\delta\eta_k}\Big)\Big)\Big) \end{aligned}$$

using (C.10) and $1 + \frac{1}{\eta_k^2} = O(\frac{h_k}{\eta_k})$.

□

C.5.5 Proof of Theorem 19

We restate Theorem 19 which summarizes the guarantees for the audit of the probabilistic envy-freeness criterion with AUDIT, and we prove it below:

Theorem. *Let* $\epsilon, \gamma, \lambda \in (0,1], \delta \in (0, \frac{1}{2})$. *Let* $\tilde{M} = \left\lceil \frac{\log(3/\delta)}{\lambda} \right\rceil$ *and* $K = \left\lceil \frac{\log(3\tilde{M}/\delta)}{\log(1/(1-\gamma))} \right\rceil$. *With probability at least* $1-\delta$,

- *AUDIT satisfies the conservative constraint (6.3) for all* $\tilde{M}$ *audited users,*
- *the bounds on duration and cost from Th. 18 (using* $\frac{\delta}{3\tilde{M}}$ *instead of* δ*) are simultaneously valid,*
- *if AUDIT outputs* $(\epsilon,\gamma,\lambda)$-`envy-free`*, then the recommender system is* $(\epsilon,\gamma,\lambda)$*-envy-free, and if it outputs* `not-envy-free`*, then* $\exists (m,n)$, $u^m(\pi^m) < u^m(\pi^n)$.

Proof. The first point is a consequence of Theorem 56 and the second point is a consequence of Theorems 57 and 58. Since we apply OCEF to each target user with confidence $\frac{\delta}{3\tilde{M}}$, by the union bound the confidence intervals are simultaneously valid for all $\tilde{M}$ target users with probability $1-\frac{\delta}{3}$. Therefore, with probability at least $1-\frac{\delta}{3}$, the conservative constraint is satisfied for all $\tilde{M}$ users *and* the bounds on cost and duration hold simultaneously for all $\tilde{M}$ users.

We now prove the third bullet point in two steps.

Step 1 We show that the value of $K = \frac{\log(3\tilde{M}/\delta)}{\log(1/(1-\gamma))}$ is chosen to guarantee the following result: with probability $1-\frac{\delta}{3\tilde{M}}$, if for a user we have $\mu_0 + \epsilon \geq \max_{k\in[\![K]\!]} \mu_k$, then the user is not (ϵ,γ)-envious.

First, we apply the theorem on random subset selection from (Schölkopf and Smola [2002], Theorem 6.33), which guarantees that with probability $1-(1-\gamma)^K$, the arm with maximal reward among the K arms is in the $(1-\gamma)$-quantile range of all possible M arms. Solving for $(1-\gamma)^K = \frac{\delta}{3\tilde{M}}$, we get that when $K = \left\lceil \frac{\log(3\tilde{M}/\delta)}{\log(1/(1-\gamma))} \right\rceil$, the arm with maximal reward among the K is in the $(1-\gamma)$ quantile range with probability $1-\frac{\delta}{3\tilde{M}}$. This means that if for a target user m, we have $u^m(\pi^m) + \epsilon = \mu_0 + \epsilon \geq \max_{k\in[\![K]\!]} \mu_k$, then with probability $1-\frac{\delta}{3\tilde{M}}$, we also have:

$$\mathbb{P}_{n\sim U_M}[u^m(\pi^m) + \epsilon \geq u^m(\pi^n)] \geq 1-\gamma,$$

meaning the user is not (ϵ,γ)-envious. By a union bound over the $\tilde{M}$ target users, the property holds simultaneously for all $\tilde{M}$ target users with probability $1-\frac{\delta}{3}$.

Step 2 We now show that the number of users to audit $\tilde{M} = \left\lceil \frac{\log(3/\delta)}{\lambda} \right\rceil$ is chosen to guarantee that if none of the $\tilde{M}$ sampled users are (ϵ,γ)-envious, then this holds true for an $(1-\lambda)$ fraction of the whole population with probability $1-\frac{\delta}{3}$.

Let $\delta' = \frac{\delta}{3}$. Denoting q the probability that a user is not (ϵ,γ)-envious, we want to guarantee that $q \geq 1-\lambda$ with probability at least $1-\delta'$, using $\tilde{M}$ Bernoulli trials where $p := 1-q$ is the probability of success.

Let $\bar{B}(\tilde{M}, k, \delta')$ denote the largest p' such that the probability of observing k or more successes is at least $1-\delta'$ (i.e., $\bar{B}(\tilde{M}, k, \delta')$ is the binomial tail inversion). By definition, we have $p \leq \bar{B}(\tilde{M}, 0, \delta')$. Using the property that $\bar{B}(\tilde{M}, 0, \delta') \leq \frac{\log(1/\delta')}{\tilde{M}}$ (see e.g., Langford [2005]), we can guarantee that $p \leq \lambda$ as soon as $\frac{\log(1/\delta')}{\tilde{M}} \leq \lambda$. Solving for $\tilde{M}$, we obtain that $\tilde{M} = \left\lceil \frac{\log(1/\delta')}{\lambda} \right\rceil = \left\lceil \frac{\log(3/\delta)}{\lambda} \right\rceil$ is sufficient to guarantee $p \leq \lambda$, or equivalenly $q \geq 1-\lambda$ with probability $1 - \frac{\delta}{3}$.

We combining Step 1 and 2 by a union bound: if for $\tilde{M}$ users and K arms, we have $\mu_0 + \epsilon \geq \max_{k \in [\![K]\!]} \mu_k$, then with probability $1 - \frac{2\delta}{3}$, an $(1-\lambda)$ fraction of the whole population is not (ϵ, γ)-envious – or equivalently, the recommender system is $(\epsilon, \gamma, \lambda)$-envy-free. Since OCEF is correct with probability $1 - \frac{\delta}{3}$ when outputting that $\mu_0 + \epsilon \geq \max_{k \in [\![K]\!]} \mu_k$ (i.e., ϵ-`no-envy`), the union bound guarantees with probability $1-\delta$ that AUDIT is correct when outputting $(\epsilon, \gamma, \lambda)$-`envy-free`. Since OCEF is correct with probability $\geq 1-\delta$ when outputting `envy`, then so is AUDIT when outputting `not-envy-free`, which achieves the proof of the third bullet point.

□

Appendix D

Online selection of diverse committees

Abstract

Citizens' assemblies need to represent subpopulations according to their proportions in the general population. These large committees are often constructed in an online fashion by contacting people, asking for the demographic features of the volunteers, and deciding to include them or not. This raises a trade-off between the number of people contacted (and the incurring cost) and the representativeness of the committee. We study three methods, theoretically and experimentally: a greedy algorithm that includes volunteers as long as proportionality is not violated; a non-adaptive method that includes a volunteer with a probability depending only on their features, assuming that the joint feature distribution in the volunteer pool is known; and a reinforcement learning based approach when this distribution is not known a priori but learnt online.

D.1 Introduction

Forming a representative committee consists in selecting a set of individuals, who agree to serve, in such a way that every part of the population, defined by specific features, is represented proportionally to its size. As a paradigmatic example, the Climate Assembly in the UK and the Citizens' Convention for Climate in France brought together 108 and 150 participants respectively, representing sociodemographic categories such as gender, age, education level, professional activity, residency, and location, in proportion to their importance in the wider society. Beyond citizens' deliberative assemblies, proportional representation often has to be respected when forming an evaluation committee, selecting a diverse pool of students or employees, and so on.

Two key criteria for evaluating the committee formation process are the representativeness of the final selection and the number of persons contacted (each of these incurring a cost). The trade-off is that the higher the number of people contacted, the more proportional the resulting committee.

A first possibility is to use an offline strategy (as for the UK assembly): invitations are sent to a large number of people (30,000), and the final group is selected among the pool of volunteers. An alternative setting which is common in hiring is to consider an online process: the decision-maker is given a stream of candidates and has to decide at each timestep whether or not to admit the candidate to the final committee. This work focuses on the latter setting.

A further difficulty is that the distribution of *volunteers* is not necessarily known in advance. For

example, although the target is to represent distinct age groups proportionally to their distribution in the wider population, it may be the case that older people are predominant among volunteers.

Multi-attribute proportional representation in committee selection in an off-line setting usually assumes full access to a finite (typically large) database of candidates. This assumption is impractical in a variety of real-world settings: first, the database does not exist beforehand and constructing it would require contacting many more people than necessary; second, in some domains, the decision to hire someone should be made immediately so that people don't change their mind in the meantime (which is typical in professional contexts).

An online strategy must achieve a good trade-off between sample complexity, i.e. the number of timesteps needed to construct a full committee, and the quality of the final committee, as measured by its distance to the target distribution.

We focus on the online setting. We introduce a new model and offer three different strategies, which rely on different assumptions on the input (and the process). The *greedy* strategy selects volunteers as long as their inclusion does not jeopardize the size and representation constraints; it does not assume any prior distribution on the volunteer pool. The *nonadaptive* strategy, based on constrained Markov decision processes, repeatedly chooses a random person, and decides whether to include or not a volonteer with a probability that depends only on their features; it assumes the joint distribution in the volunteer pool is known; it can be parallelised. Finally, the *reinforcement learning* strategy assumes this distribution is not known a priori but can be learnt online.

Which of these strategies are interesting depends on domain specificities. For each, we study bounds for expected quality and sample complexity, and perform experiments using real data from the UK Citizens' Assembly on Brexit.

The outline of the paper is as follows. We discuss related work in Section D.2, define the problem in Section D.3, define and study our three strategies in Sections D.3.2, D.4 and D.5, analyse our experiments in Section D.6 and conclude in Section D.7.

D.2 Related work

Diversity and representation in committee (s)election The problem of selecting a diverse set of candidates from a candidate database, where each candidate is described by a vector of attribute values, has been considered in several places. In Lang and Skowron [2018], the goal is to find a committee of a fixed size whose distribution of attribute values is as close as possible to a given target distribution. In Celis et al. [2018a], Bredereck et al. [2018], each candidate has a score, obtained from a set of votes, and some constraints on the proportion of selected candidates with a given attribute value are specified; the goal is to find a fixed-size committee of maximal score satisfying the constraints. In the same vein, Aziz [2019] considers soft constraints, and Bei et al. [2020] do not require the size of the committee to be fixed.[1]

Our online setting shifts the difficulty of the multi-attribute representation problem from computational complexity analyses, to the need for probabilistic guarantees on the tradeoffs between sample complexity and achieved proportionality.

Representative and fair sortition Finding a representative committee (typically, a panel of citizens) with respect to a set of attributes, using *sortition*, is the topic of at least two recent papers. Benadè et al. [2019] show that stratification (random selection from small subgroups defined by

[1]Note that *diversity* and *proportional representation* are often used with a different meaning in multiwinner elections, namely, in the sense that each voter should feel represented in an elected committee, regardless of attributes. A good entry to this literature is the survey Faliszewski et al. [2017].

attribute values, rather than from the larger group) only helps marginally. Flanigan et al. [2020] go further and consider this three-stage selection process: (1) letters are sent to a large number of random individuals (the *recipients*); (2) these recipients answer whether they agree to participate, and if so, give their features; those individuals constitute the *pool*; (3) a sampling algorithm is used to select the final *panel* from the pool. As the probability of willingness to participate is different across demographic groups, each person is selected with a probability that depends on their features, so as to correct this self-selection bias. This guarantees that the whole process be fair to all individuals of the population, with respect of going from the initial population to the panel.[2]

The main differences between this work and ours are: (1) (once again) our process is online; (2) we do not consider individual fairness, only group representativeness; (3) we care about minimizing the number of people contacted. Moreover, unlike off-line processes, our process can be applied in contexts where hiring a person just interviewed cannot be delayed; this may not be crucial for citizens' assemblies (although someone who volunteers at first contact may change their mind if the delay until the final selection is long), but this is definitely so when hiring a diverse team of employees.

Online selection problems Generalized secretary problems Babaioff et al. [2008] are optimal stopping problems where the goal is to hire the best possible subset of persons, assuming that persons arrive one at a time, their value is observed at that time, and the decision to hire or not them must be taken immediately. The problem has been generalized to finding a set of items maximizing a submodular value function Bateni et al. [2013], Badanidiyuru et al. [2014] While the latter models do not deal with diversity constraints, Stoyanovich et al. [2018] aims at selecting a group of people arriving in a streaming fashion from a finite pool, with the goal of optimizing their overall quality subject to diversity constraints. The common point with our approach is the online nature of the selection process. The main differences are that they consider only one attribute, the size of the pool is known, and yet more importantly, what is optimized is the intrinsic quality values of the candidates and not the number of persons interviewed. Closer to our setting is Panigrahi et al. [2012] who consider diversity along multiple features in online selection of search results, regardless of item quality. They only seek to maximise diversity, and do not consider trade-offs with the number of items observed.

The diverse hiring setting of Schumann et al. [2019a] is very different. At each time step, the decision-maker chooses which candidate to interview and only decides on which subset to hire after multiple rounds, whereas in our setting, candidates arrive one by one and decisions are made immediately.

D.3 Formal setting

D.3.1 Problem definition

Let $\mathcal{X} = \mathcal{X}_1 \times ... \times \mathcal{X}_d$ be the product space of d finite domains, each of size $D_i = |\mathcal{X}_i|$, and where we identify $\mathcal{X}_i$ with $[\![D_i]\!] = \{1, ..., D_i\}$. Each candidate is represented by a *characteristic vector* $x \in \mathcal{X}$ with d *features*. Let $x^i \in \mathcal{X}_i$ denote the value of the i-th feature. For each $i \in [\![d]\!]$, we consider a *target vector* $\rho^i \in (0,1)^{D_i}$ with $\sum_{j=1}^{D_i} \rho^i_j = 1$.

The candidate database is infinite and the horizon as well. At each timestep $t \geq 1$, the agent observes a candidate x_t drawn i.i.d. from a stationary distribution p over $\mathcal{X}$, i.e. $x_t \sim p$. The

[2] Fairness guarantees are pushed further in following (yet unpublished) work by the authors: see https://youtu.be/x_1Ce1kT7vc.

gender \ age	S	J
M	$1/2 - \epsilon'$	$1/4$
F	$1/4$	ϵ'

Table D.1: Example candidate distribution p with 2 binary features.

decision-maker must immediately decide between two actions: *accept* or *reject* the candidate, which we respectively denote as $a_t = 1$ and $a_t = 0$.

The goal is to select a *committee* C of K candidates that matches the target vectors as closely as possible, while minimizing the number of candidates screened.

For some set C, let $\lambda(C) \in \prod_{i=1}^{d} [0,1]^{D_i}$ be the *representation profile* of C, where $\lambda_j^i(C) = \frac{|\{x \in C : x^i = j\}|}{|C|}$. We define the *representation loss* as $\|\lambda(C) - \rho\|_\infty = \max_{i \in [\![d]\!], j \in [\![D_i]\!]} |\lambda_j^i(C) - \rho_j^i|$. We evaluate how much C matches the target ρ by the ℓ_∞ metric, because it is harsher than ℓ_1, ℓ_2 on committees that are unacceptable in our applications (e.g. committees with no women that achieve perfect representation on all other categories than gender).

Let $C_t = \{x_{t'} : t' \leq t, a_{t'} = 1\}$ denote the set of all accepted candidates at the end of step t. The agent stops at τ, where τ is the first time when K candidates have been accepted, i.e. the total number of candidates screened. The agent following a (possibly randomized) algorithm ALG must minimize the *sample complexity* $\mathbb{E}^{p, ALG}[\tau]$.

Importantly, we consider two settings: whether the candidate distribution p is *known* or *unknown*.

Remark 9. *In this model, we simply ignore non-volunteers, since the agent only needs to make decisions for volunteers, which from now on we call* candidates. *The joint distribution of characteristic vectors in the population of candidates is p.*

D.3.2 Greedy strategy

We describe a first simple strategy. In `Greedy`, the agent greedily accepts any candidate as long as the number of people in the committee with $x^i = j$ does not exceed the quota $\lceil \rho_j^i K \rceil + \frac{\epsilon K}{(D_i - 1)}$ for any i, j, where $\epsilon > 0$ is some tolerance parameter for the representation quality.

Proposition 63. *The representation loss incurred by* `Greedy` *is bounded as follows:*

$$\|\lambda(C_\tau) - \rho\|_\infty \underset{a.s.}{\leq} \left(\frac{\max_{i \in [d]} D_i - 1}{K} + \epsilon\right).$$

The proof and pseudocode are included in App. E.1.

This method is simple to interpret and implement, and can even be used when the candidate distribution p is unknown. However, in the following example, we see that `Greedy` may be inefficient because it requires interacting with an arbitrarily large number of candidates to recruit a full committee.

Example 9. *Let $\epsilon' > 0, \ll 1$. There are 2 binary features,* gender *and* age*, with domains $\mathcal{X}_{gender} = \{M, F\}$ and $\mathcal{X}_{age} = \{S, J\}$. The candidates are distributed as p given in Table D.1. We want a committee of size $K = 4$ (e.g., a thesis committee) and the target is $\rho^{gender} = (1/2, 1/2)$ and $\rho^{age} = (3/4, 1/4)$.*

Let A be the event that in the first 3 timesteps, the agent observes candidates with characteristic vectors $\{FS, MS, MS\}$ in any order. Then `Greedy` *accepts all of them, i.e. $A = \{C_3 = \{FS, MS, MS\}\}$. We have: $\mathbb{P}[A] = 1/4(1/2 - \epsilon')^2 \times 3! = 3/2(1/2 - \epsilon')^2 \geq 3/2\big(1/3\big)^2 = 1/6$.*

Under event A, `Greedy` *can only stop upon finding FJ in order to satisfy the representation constraints. Therefore, $\tau | A$ follows a geometric distribution with success probability ϵ', hence its*

expectation is $1/\epsilon'$*, and* $\mathbb{E}^{p,\texttt{Greedy}}[\tau] \geq \mathbb{E}[\tau|A] \times \mathbb{P}[A] = 1/6\epsilon'$*. Therefore, the sample complexity of* `Greedy` *in this example is arbitrarily large.*

This example shows the limits of directly applying a naive strategy to our online selection problem, where the difficulty arises from considering multiple features simultaneously, even when there are only 2 binary features. We further discuss the strengths and weaknesses of `Greedy`, and its sensitivity to the tolerance ϵ in our experiments in Section D.6.

The greedy strategy is adaptive, in the sense that decisions are made based on the current candidate and candidates accepted in the past. In the following section, we present, with theoretical guarantees, an efficient yet non-adaptive algorithm based on constrained MDPs for the setting in which the candidate distribution is known. We then adapt this approach to the case when this distribution is unknown, using techniques for efficient exploration / exploitation in constrained MDPs relying on the principle of optimism in the face of uncertainty.

D.4 p is known: constrained MDP strategy

In this section, we assume the distribution p is known, and we place ourselves in the limit where we would select a committee of infinite size, and aim to maximize the rate at which candidates are selected, under the constraint that the proportion of accepted candidates per feature value is controlled by ρ. One advantage of this approximation is that the optimal policy is stationary, thus simple to represent. Moreover, as stationary policies can be very well parallelized, in the case where multiple candidates can be interviewed simultaneously. To apply this approach to the finite-size committee selection problem, one needs to interrupt the agent when K candidates have been selected. We showcase a high probability bound of $O(\sqrt{1/K})$ on the representation loss, which guarantees that for large enough values of K, the resulting committee is representative.

From now on, we assume that any feature vector can be observed, i.e., $p(x) > 0$ for all x, so that proportional representation constraints can be satisfied.

D.4.1 Our model

Fundamentally, our problem could be seen as a contextual bandit with stochastic contexts $x_t \sim p$ and two actions $a_t = 0$ or 1. However, the type of constraints incurred by proportional representation are well studied in constrained MDPs (CMDPs) Altman [1999], whereas the contextual bandits literature focused on other constraints (e.g., knapsack constraints Agrawal and Devanur [2016]). We show how we can efficiently leverage the CMDP framework for our online committee selection problem.

Formally, we introduce an MDP $M = (\mathcal{X}, \mathcal{A}, P, r)$, where the set of states is the d-dimensional candidate space $\mathcal{X}$, the set of actions is $\mathcal{A} = \{0, 1\}$, and the (deterministic) reward is $r(x, a) = \mathbb{1}_{\{a=1\}}$. The transition kernel P, which defines the probability to be in state x' given that the previous state was x and the agent took action a, is very simple in our case: we simply have $P(x'|x, a) = p(x')$ since candidates are drawn i.i.d regardless of the previous actions and candidates.

We consider the *average reward* setting in which the performance of a policy $\pi : \mathcal{X} \times \mathcal{A} \to [0, 1]$ is measured by its *gain* $g^{p,\pi}$, defined as:

$$g^{p,\pi}(x) = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}^{p,\pi}\left[\sum_{t=1}^{T} r(x_t, a_t)\middle| x_1 = x\right].$$

We simply write $g^{p,\pi} := g^{\pi}$ when the underlying transition is p without ambiguity.

We include proportional representation constraints following the framework of CMDPs, where the set of allowed policies is restricted by a set of additional constraints specified by reward functions. In our case, for $i \in [\![d]\!], j \in [\![D_i]\!]$, we introduce $r^i_j(x,a) = \mathbb{1}_{\{x^i=j, a=1\}}$, and let $\xi^i_j = r^i_j - \rho^i_j r$ be the reward function for the constraint indexed by i,j. Similarly to the gain, we define $h^{i\,\pi}_j = \lim_{T\to\infty} \frac{1}{T}\mathbb{E}^\pi\left[\sum_{t=1}^T \xi^i_j(x_t,a_t)\right]$. The CMDP is defined by:

$$\max_\pi \{g^\pi \mid \forall i \in [\![d]\!], \forall j \in [\![D_i]\!], h^{i\,\pi}_j = 0\}. \tag{D.1}$$

Given the simplicity of the transition kernel, and since the MDP is ergodic by the assumption $p > 0$, the gain is constant, i.e. $\forall x \in \mathcal{X}, g^\pi(x) = g^\pi$, and problem (D.1) is well defined. From now on, we only write g^π and $\xi^{i\,\pi}_j$. Moreover, the optimal policy for the CMDP (D.1) is denoted π^* and is *stationary* Altman [1999].

Lemma 64. *g^π is the* selection rate *under policy π:*

$$g^\pi = \sum_x p(x)\pi(x,1) = \mathbb{P}^{p,\pi}[a=1]$$

Moreover, if π is feasible for CMDP (D.1), then:

$$\forall i \in [d], \forall j \in [\![D_i]\!], \mathbb{P}^{p,\pi}[x^i = j | a = 1] = \rho^i_j.$$

Lemma 64 implies that (a) π^* maximises the selection rate of candidates, and (b) the constraints of (D.1) force candidates x with $x^i = j$ to be accepted in proportions given by ρ^i_j.

The CMDP can be expressed as the linear program:

$$\begin{aligned} \max_{\pi \in \mathbb{R}_+^{\mathcal{X}\times\mathcal{A}}} \quad & \sum_{x,a} \pi(x,a)p(x)r(x,a) \\ \text{u.c.} \quad & \forall x \in \mathcal{X}, \sum_a \pi(x,a) = 1 \\ & \forall i,j, \sum_{x,a} \pi(x,a)p(x)\xi^i_j(x,a) = 0. \end{aligned} \tag{D.2}$$

Notice that problem (D.2) is feasible by the assumption that $\forall x \in \mathcal{X}, p(x) > 0$. Next we study how well the proportional selection along features is respected when we shift from infinite to finite-sized committee selection.

D.4.2 Theoretical guarantees

We analyze the `CMDP-based` strategy where at each timestep, the agent observes candidates $x_t \sim p$, decides to accept x_t by playing $a_t \sim \pi^*(.|x_t)$ and stops when K candidates have been accepted. We later refer to it as `CMDP` for brevity.

First, we formally relate the gain g^π that we optimize for in (D.1) to the quantity of interest $\mathbb{E}^{p,\pi}[\tau]$.

Lemma 65. *For any stationary policy π, $\mathbb{E}^{p,\pi}[\tau] = \frac{K}{g^\pi}$.*

Lemma 65 is a direct consequence of the fact that $\tau + K$ follows a negative binomial distribution with parameters K and $1 - g^\pi$, which are respectively the number of successes and the probability of failure, i.e. of rejecting a candidate under π. Note that this is only true because in our case the transition structure of the MDP ensures constant gain. A quick sanity check shows that if the

Algorithm 10: `RL-CMDP` algorithm.

input : confidence δ, committee size K, targets ρ
output : committee C_τ

$t \leftarrow 0$, $C_0 \leftarrow \emptyset$;
while $|C_t| < K$ **do**
  for *episode* $l = 1, 2, \ldots$ **do**
    $\tau_l = t + 1$;
    $\pi_l \leftarrow$ sol. of (D.3) via the extended LP (D.4);
    while $n_t(x_t) < 2n_{\tau_l - 1}(x_t)$ **do**
      $t \leftarrow t + 1$, Execute π_l;
    end
  end
end
return C_t

agent systematically accepts all candidates, i.e. $g^\pi = 1$, then $\mathbb{E}^{p,\pi}[\tau] = K$, and that maximizing g^π is equivalent to minimizing $\mathbb{E}^{p,\pi}[\tau]$.

We exhibit a bound on the representation loss of `CMDP` which follows the optimal stationary policy π^* of CMDP (D.1). Let $\tilde{d} = \sum_{i=1}^{d}(D_i - 1)$. ($\tilde{d} = d$ when all features are binary.)

Proposition 66. *Let π^* be an optimal stationary policy for CMDP (D.1). Let $\delta > 0$. Then,*

$$\mathbb{P}^{p,\pi^*}\left[\|\lambda(C_\tau) - \rho\|_\infty \leq \sqrt{\frac{\log(\frac{2\tilde{d}}{\delta})}{2K}}\right] \geq 1 - \delta.$$

All proofs of this section are available in Appendix E.2.1.

The upper bound on the representation loss of `CMDP` decreases with the committee size in $\sqrt{1/K}$. This shows that the stationary policy π^* works well for larger committees, although it acts independently from previously accepted candidates. The intuition is that for larger committees, adding a candidate has less impact on the current representation vector.

Example 10. *We take the same attributes and same distribution as in Table D.1, with $\epsilon' = 1/6$. Here, the target vectors are $\rho^{gender} = (1/2, 1/2)$ and $\rho^{age} = (1/2, 1/2)$: an ideal committee contains as many women as men, as many senior as junior.*

With the optimal policy for LP (D.2), each time the current volunteer is a senior male, we select him with probability 1/2; all other volunteers are selected with probability 1. The expected final composition of the pool is 30% of junior male, 30% of senior female, 20% of junior female and 20% of senior male. As the policy selects in average 5/6 of the volunteers, the expected time until we select K candidates is $\mathbb{E}^{p,\pi^*}[\tau] = (6/5)K$. More details can be found in App. E.5.

D.5 p is unknown: optimistic CMDP strategy

We now tackle the committee selection problem when the candidate distribution p is unknown and must be learned online. Let $g^* = g^{\pi^*}$ be the value of (D.1), which is the optimal gain of the CMDP when the distribution p is known. We evaluate a learning algorithm by:

1. the performance regret: $R(T) = \sum_{t=1}^{T}(g^* - r(x_t, a_t))$,
2. the cost of constraint violations:
   $R^c(T) = \max_{i,j} \left| \sum_{t=1}^{T} \xi_j^i(x_t, a_t) \right|$.

We propose an algorithm that we call `RL-CMDP` (Reinforcement Learning in CMDP, Alg. 10). It is an adaptation of the *optimistic* algorithm UCRL2 Jaksch et al. [2010], and it also builds on the algorithm OptCMDP proposed by Efroni et al. [2020] for finite-horizon CMDPs. Learning in average-reward CMDPs involves different challenges, because there is no guarantee that the policy at each episode has constant gain. It does not matter in our case, since as we noted in Sec. D.4, the simple structure of the transition kernel ensures constant gain, and does not require to use the Bellman equation. The few works on learning in average-reward CMDPs make unsuitable assumptions for our setting Zheng and Ratliff [2020], Singh et al. [2020].

`RL-CMDP` proceeds in episodes, which end each time the number of observations for some candidate x doubles. During each episode l, observed candidates x_t are accepted on the basis of a single stationary policy π_l.

Let τ_l denote the start time of episode l and $E_l = [\tau_l, \tau_{l+1}]$. Let $n_t(x) = \sum_{t'=1}^{t} \mathbb{1}_{\{x_{t'}=x\}}$ and $N(t) = |C_{t-1}| = \sum_{t'=1}^{t-1} \mathbb{1}_{\{a_{t'}=1\}}$. Let $N_j^i(t) = \sum_{t'=1}^{t-1} \mathbb{1}_{\{x_{t'}^i=j, a_{t'}=1\}}$ be the number of accepted candidates x such that $x^i = j$ before t.

At each episode l, the algorithm estimates the true candidate distribution by the empirical distribution $\hat{p}_l(x) = \frac{n_{\tau_l-1}(x)}{\tau_l - 1}$ and maintains confidence sets B_l on p. As in UCRL2, these are built using the inequality on the ℓ_1-deviation of p and $\hat{p}_l$ from Weissman et al. [2003]:

Lemma 67. *With probability* $\geq 1 - \frac{\delta}{3}$,

$$\|\hat{p}_l - p\|_1 \leq \sqrt{\frac{2|\mathcal{X}| \log\left(6|\mathcal{X}|\tau_l(\tau_l - 1)/\delta\right)}{\tau_l - 1}} := \beta_l$$

Let $B_l = \{\tilde{p} \in \Delta(\mathcal{X}) : \|\hat{p}_l - \tilde{p}\|_1 \leq \beta_l\}$ be the confidence set for p at episode l. The associated set of compatible CMDPs is then $\{\tilde{M} = (\mathcal{X}, \mathcal{A}, \tilde{p}, r, \xi) : \tilde{p} \in B_l\}$. At the beginning of each episode, `RL-CMDP` finds the optimum of:

$$\max_{\pi \in \Pi, \tilde{p} \in B_l} \{g^{\tilde{p},\pi} \,|\, \forall i, j,\ h_j^{i\,\tilde{p},\pi} = 0\}. \tag{D.3}$$

Extended LP In order to optimize this problem, we re-write (D.3) as an extended LP. Following Rosenberg and Mansour [2019] and the CMDP literature, we introduce the state-action occupation measure $\mu(x, a) = \pi(x, a)p(x)$ and variables $\beta(x)$ to linearize the ℓ_1 constraint induced by the confidence set:

$$\begin{aligned}
\max_{\substack{\mu \in \mathbb{R}^{\mathcal{X}\times\mathcal{A}} \\ \beta \in \mathbb{R}^{\mathcal{X}}}} \quad & \sum_{x,a} \mu(x,a) r(x,a) \\
\text{u.c.} \quad & \mu \geq 0, \sum_{x,a} \mu(x,a) = 1 \\
& \forall x, \sum_a \mu(x,a) \leq \hat{p}_l(x) + \beta(x) \\
& \forall x, \sum_a \mu(x,a) \geq \hat{p}_l(x) - \beta(x) \\
& \forall x, a, \sum_y \beta(y) \leq \mu(x,a)\beta_l \\
& \forall i, j, \sum_{x,a} \mu(x,a)\xi_j^i(x,a) = 0.
\end{aligned} \tag{D.4}$$

The last constraint is the proportional representation constraint. The second to fourth constraints enforce the compatibility of μ with the ℓ_1 confidence set. We retrieve the distribution as $\tilde{p}_l(x) =$

$\sum_a \mu(x,a)$, and the policy as:

$$\pi_l(x,a) = \begin{cases} \frac{\mu(x,a)}{\tilde{p}_l(x)} & \text{if } \tilde{p}_l \neq 0 \\ \frac{1}{2} & \text{otherwise}\,. \end{cases}$$

Precisely, if some $\tilde{p}_l(x) = 0$, we may set the policy $\pi_l(a|x)$ arbitrarily. Since the MDP induced by $\tilde{p}$ is still weakly communicating, and in particular any policy is unichain, the optimal gain in this CMDP is not affected.

We now provide regret and representativeness guarantees.

Theorem 68. *With probability $\geq 1-\delta$, the regret of* `RL-CMDP` *satisfies:*

$$R(T) = O\big(\sqrt{|\mathcal{X}|T\log(|\mathcal{X}|T/\delta)}\big)$$
$$R^c(T) = O\big(\sqrt{|\mathcal{X}|T\log(|\mathcal{X}|T/\delta)}\big).$$

Moreover, with probability $1-\delta$, the representation loss of `RL-CMDP` *at horizon T satisfies:*

$$\|\lambda(C_T) - \rho\|_\infty = O\left(\frac{1}{g^*}\sqrt{\frac{|\mathcal{X}|\log\left(|\mathcal{X}|T/\delta\right)}{T}}\right).$$

The full proof is in Appendix E.2.2. It relies on decomposing regret over episodes, bounding the error on p which decreases over episodes as the confidence sets are refined, and leveraging martingale inequalities on the cumulative rewards.

Since $\frac{R(T)}{T} = g^* - \frac{N(T)}{T}$, it means that with high probability, the difference between the optimal selection rate and the selection rate of `RL-CMDP` decreases in $\sqrt{\log(T)/T}$ w.r.t. the horizon T. The representation loss decreases at the same speed, meaning that the agent should see enough candidates to accurately estimate p, and accept candidates at little cost for representativeness.

Compared to the bound from Proposition 66, the cost of not knowing p on representativeness is a $\sqrt{|\mathcal{X}|\log(|\mathcal{X}|)}$ factor. This is due to the estimation of p in the worst case, which is controlled by Lemma 67. As we show in our experiments (Sec. D.6), the impact of $|\mathcal{X}|$ on performance regret (and in turn on sample complexity) is not problematic in our typical citizens' assembly scenario: since there are only a handful of features, our algorithm selects candidates quickly in practice (though representativeness is weakened by not knowing p). For specific structures of p, we obtain bounds with better scaling in $|\mathcal{X}|$, by controlling each entry of p with Bernstein bounds Maurer and Pontil [2009], instead the ℓ^1-norm. For completeness, we describe this alternative in Appendix E.3.

Interestingly, the representation loss is also inversely proportional to g^*, the optimal selection rate in the true CMDP. The reason is that the CMDP constraints do not control the ratios $\lambda_j^i(C_T) = \frac{N_j^i(T)}{N(T)}$, but $N_j^i(T)$ instead (by definition of $R^c(T)$ and ξ_j^i). If $N(T)$ is small, i.e. due to a small selection rate g, then $R_j^i(T) = |N_j^i(T) - \rho_j^i N(T)|$ is small, but not necessarily $|\frac{N_j^i(T)}{N(T)} - \rho_j^i|$: the committee is too small to be representative.

D.6 Experiments

The goal of these experiments is to answer the following: **(Q1)** In practice, for which range of committee sizes do our strategies achieve satisfying sample complexity and representation loss? **(Q2)** What is the cost of not knowing the distribution p for the sample complexity and representation loss?

Experimental setting To answer these questions, we use summary data from the 2017 Citizens' Assembly on Brexit. The participants were recruited in an offline manner: volunteers could express interest in a survey, and then 53 citizens were drawn from the pool of volunteers using stratified sampling, in order to construct an assembly that reflects the diversity of the UK electorate. We use summary statistics published in the report Renwick et al. [2017] to simulate an online recruitment process.

There are $d = 6$ features: the organisers expressed target quotas for 2 ethnicity groups, 2 social classes, 3 age groups, 8 regions, 2 gender groups and 2 Brexit vote groups (remain, leave). The report also includes the number of people contacted per feature group (e.g., women, or people who voted to remain) and the volunteering rate for each feature group, which we use as probability of volunteering given a feature group. We use Bayes' rule to compute the probabilities of feature groups among volunteers, and use them as the marginal distributions $\Pr[x^i = j|\text{volunteers}]$ (since we only consider the population of volunteers). Since we only have access to the marginals, we compute the joint distribution as if the features were independent, although our model is agnostic to the dependence structure of the joint distribution. In Appendix E.4.2, we present additional experiments with non-independent features, using a real dataset containing demographic attributes. The results are qualitatively similar.

We study `Greedy` with tolerance $\epsilon = 0.02, 0.05$. We run experiments for $K = 50, 100, 150, 250, 500, 1000$, averaged over 50 simulations. More details are found in App. E.4.1.

(A1) We compare `Greedy` and `CMDP`, when the distribution p is *known*. Figure D.1 shows that the greedy strategy with $\epsilon = 0.05$ requires 10 times more samples than `CMDP`, and its representation loss is higher as soon as $K \geq 250$. `Greedy` with lower tolerance $\epsilon = 0.02$ achieves better representation than `CMDP` for smaller committees ($K \leq 100$), but the margin quickly decreases with K. However, even for small committees, it requires about 100 times more samples, which is prohibitively expensive. Figure D.1 shows that for `CMDP`, the sample complexity grows linearly in the committee size, with a reasonable slope (we need to find $\tau \approx 500$ volunteers for a committee of size $K \approx 200$).

(A2) To corroborate the previously discussed effect of $|\mathcal{X}|$ when p is *unknown*, we evaluate `RL-CMDP` on different configurations: (1) using only the features ethnicity, social class, and gender ($d = 3, |\mathcal{X}| = 8$), (2) using all features except regions ($d = 5, |\mathcal{X}| = 48$). Fig. D.2 shows that unlike `CMDP` which has full knowledge of p, it is for large committee sizes that `RL-CMDP` reaches low representation loss (below 0.05 for $K \geq 1500$ in the configuration(1)). This is because `RL-CMDP` needs to collect more samples to estimate p, as discussed in Th. 68. For known p, the CMDP approach achieves the same representativeness for middle-sized committees (repr. loss ≤ 0.05 for $K \approx 250$). Hence, comparing the cases of known (Fig. D.1) and unknown distribution p (Fig. D.2), the ignorance of p is not costly for sample complexity, but rather for the representation loss which decreases more slowly.

Consistently with Th. 68, we observe that the representation loss is higher when $\mathcal{X}$ is larger ($d = 5$). For small and middle-sized committees, the loss of `RL-CMDP` is much worse than `Greedy`'s which also works for unknown p. For large committees though, the margin is only 0.05 when $K \gtrsim 2000$ and $\tau \approx 3500$ for `RL-CMDP` (which is $\times 3$ more sample efficient than `Greedy`). In absolute terms, the theoretical regret bounds have a large constant $\sqrt{|\mathcal{X}|}$. This constant is likely unavoidable asymptotically because it comes from Lem. 67, but our experiments suggest that in the non-asymptotic regime, `RL-CMDP` performs better than the bound suggests.

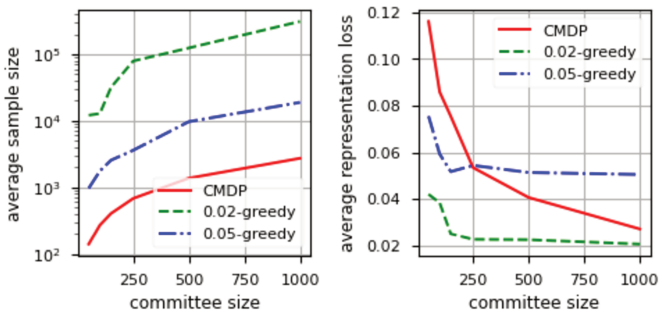


Figure D.1: Effect of committee size K on sample complexity and representation loss for different strategies, in the UK Brexit Assembly experiment, using all features. p is **known**.

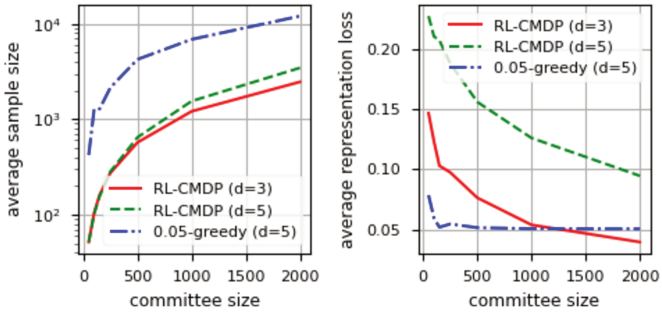


Figure D.2: Effect of committee size K on sample complexity and representation loss for `RL-CMDP`, on data simulated from the UK Brexit Assembly, using 3 and 5 features. p is **unknown**.

D.7 Conclusion

We formalised the problem of selecting a diverse committee with multi-attribute proportional representation in an online setting. We addressed the case of known candidate distributions with constrained MDPs, and leveraged exploration-exploitation techniques to address unknown distributions.

Appendix E

Appendix of Online selection of diverse committees

E.1 Details of the algorithms

For precision, we provide the pseudocode of `Greedy` in Alg. 11, and the `CMDP`-based strategy in Alg. 12.

We also prove the bound on the representation loss of `Greedy` from Proposition 63 in Section D.3.2.

Proof. For all i, j, we have by the if-condition and the termination condition:

$$\begin{aligned}\lambda_j^i(C_\tau) = \frac{N_j^i(\tau)}{K} &\leq \frac{\lceil \rho_j^i K \rceil}{K} + \frac{\epsilon}{D_i - 1} \\ &\leq \rho_j^i + \frac{1}{K} + \frac{\epsilon}{D_i - 1} \qquad \text{(E.1)} \\ &\leq \rho_j^i + \frac{D_i - 1}{K} + \epsilon. \qquad \text{(E.2)}\end{aligned}$$

For $i \in [\![d]\!]$, for $j_0 \in [\![D_i]\!]$, we have:

$$\rho_{j_0}^i = 1 - \sum_{j \neq j_0} \rho_j^i \,, \qquad \lambda_{j_0}^i(C_\tau) = 1 - \sum_{j \neq j_0} \lambda_j^i(C_\tau).$$

Combining these observations with (E.1):

$$\begin{aligned}\lambda_{j_0}^i(C_\tau) &\geq 1 - \sum_{j \neq j_0} \left(\rho_j^i + \frac{1}{K} + \frac{\epsilon}{D_i - 1}\right) \\ &= 1 - \sum_{j \neq j_0} \rho_j^i - \frac{D_i - 1}{K} - \epsilon \\ &= \rho_{j_0}^i - \frac{D_i - 1}{K} - \epsilon.\end{aligned}$$

Combining this lower bound with the upper bound (E.2), we have for all $i \in [\![d]\!], j_0 \in [\![D_i]\!]$, $\left|\lambda_{j_0}^i(C_\tau) - \rho_{j_0}^i\right| \leq \frac{D_i - 1}{K} + \epsilon$, which gives the result.

□

Algorithm 11: `Greedy` algorithm.

input : tolerance ϵ, committee size K, targets ρ
output : committee C_τ

$t \leftarrow 0$, $C_0 \leftarrow \emptyset$;
while $|C_t| < K$ **do**
  $t \leftarrow t+1$;
  Observe $x_t \sim p$;
  if $\forall i,j, N_j^i(t) + \mathbb{1}_{\{x_t^i=j\}} \leq \lceil \rho_j^i K \rceil + \frac{\epsilon K}{D_i - 1}$ **then**
    $C_t \leftarrow C_{t-1} \cup \{x_t\}$; // accept x_t
    $\forall i,j,\ N_j^i(t-1) \leftarrow N_j^i(t) + \mathbb{1}_{\{x_t^i=j\}}$
  end
end
return C_t

Algorithm 12: `CMDP`-based strategy.

input : optimal policy π^* of (D.1), committee size K
output : committee C_τ

$t \leftarrow 0$, $C_0 \leftarrow \emptyset$;
while $|C_t| < K$ **do**
  $t \leftarrow t+1$, observe $x_t \sim p$ and play $a_t \sim \pi^*(.|x_t)$;
  if $a_t = 1$ **then** $C_t \leftarrow C_t \cup \{x_t\}$;
end
return C_t

E.2 Proofs

E.2.1 Proofs of Section D.4

Proof of Lemma 64.

Proof. We have:

$$\sum_{x,a} \pi(x,a)p(x)r_j^i(x,a) = \mathbb{E}_{\substack{x\sim p \\ a\sim\pi(\cdot|x)}} \left[r_j^i(x,a)\right] = \mathbb{P}^{p,\pi}[a=1, x^i = j],$$

$$\text{and} \quad g^\pi = \sum_{x,a} \pi(x,a)p(x)r(x,a) = \mathbb{E}_{\substack{x\sim p \\ a\sim\pi(.|x)}} [r(x,a)] = \mathbb{P}^{p,\pi}[a=1].$$

The ratio of these two quantities is equal to ρ_j^i by the last constraint of (D.2). It is also equal to $\mathbb{P}[x^i = j | a = 1]$, which gives the result.

Note that it also holds true for $j = [\![D_i]\!]$, since

$$\mathbb{P}[x^i = D_i | a=1] = 1 - \sum_{j' \in [\![D_i - 1]\!]} \mathbb{P}[x^i = j' | a = 1] \quad \text{and}$$
$$\rho_{D_i}^i = 1 - \sum_{j' \in [\![D_i - 1]\!]} \rho_{j'}^i.$$

□

Proof of Proposition 66.

Proof. For any $t > 0$, we have

$$\lambda_j^i(C_t) = \frac{\sum_{s=1}^t \mathbb{1}_{\{x_s^i=j, a_s=1\}}}{\sum_{s=1}^t \mathbb{1}_{\{a_s=1\}}}.$$

and by Lemma 64, we have:

$$\mathbb{E}\left[\mathbb{1}_{\{x^i=j\}} | a=1\right] = \rho_j^i.$$

Let $\delta' > 0$. Conditionally on any $T \geq K, (a_1, ..., a_T) \in \{0,1\}^T$ s.t. $a_1 + ... + a_T = K$ and $a_T = 1$, the draws of $x_t^i | a_t = 1$ are independent and thus, by Hoeffding's inequality Hoeffding [1994], we have:

$$\begin{aligned}
&\mathbb{P}\left[|\lambda_j^i(C_T) - \rho_j^i| \geq \sqrt{\frac{\log(\frac{2}{\delta'})}{2N(T)}} \middle| a_1, ..., a_T\right] \geq 1 - \delta' \\
&= \mathbb{P}\left[|\lambda_j^i(C_T) - \rho_j^i| \geq \sqrt{\frac{\log(\frac{2}{\delta'})}{2K}} \middle| a_1, ..., a_T\right].
\end{aligned}$$

Summing up over all such sequences $(a_1, ..., a_T)$, we obtain that:

$$\mathbb{P}\left[|\lambda_j^i(C_\tau) - \rho_j^i| \geq \sqrt{\frac{\log(\frac{2}{\delta'})}{2K}}\right] \geq 1 - \delta'.$$

The result follows from applying a union bound over all $i \in [\![d]\!], j \in [\![D_i - 1]\!]$ (there are $\tilde{d}$ such (i,j) pairs) and choosing $\delta' = \delta/\tilde{d}$.

□

E.2.2 Proof of Theorem 68

The following lemma states a standard and useful inequality, which is similar to Lem. 19 in Jaksch et al. [2010].

Lemma 69. *Recall that L is the random number of episodes ran by `RL-CMDP` up until horizon T. We have:*

$$\sum_{l=1}^{L} \frac{|E_l|}{\sqrt{\tau_l - 1}} \leq 2\sqrt{T}.$$

Proof. The proof is similar to that of Lem. 13 in Zanette and Brunskill [2019]: we see E_l as the "derivative" of τ_l. Formally, let us define:

$$\begin{aligned}
F(x) &= \sum_{l=1}^{\lfloor x \rfloor} |E_l| + |E_{\lceil x \rceil}|(x - \lfloor x \rfloor) \\
f(x) &:= F'(x) = |E_{\lceil x \rceil}|.
\end{aligned}$$

We first observe that for any integer $l \in \mathbb{N}$, $f(l) = |E_l|$ and $F(l) = \tau_l$. Secondly, we have

$$F(x) \leq \sum_{l=1}^{\lfloor x \rfloor} |E_l| + |E_{\lceil x \rceil}| = \sum_{l=1}^{\lceil x \rceil} |E_l| = F(\lceil x \rceil),$$

and thus:

$$\frac{f(\lceil x \rceil)}{\sqrt{F(\lceil x \rceil) - 1}} \leq \frac{f(x)}{\sqrt{F(x) - 1}}.$$

We derive our bound as follows:

$$\begin{aligned}
\sum_{l=1}^{L} \frac{|E_l|}{\sqrt{\tau_l - 1}} &= \sum_{l=1}^{L} \frac{f(l)}{\sqrt{F(l) - 1}} = \int_1^L \frac{f(\lceil x \rceil)}{\sqrt{F(\lceil x \rceil) - 1}} \, dx \\
&\leq \int_1^L \frac{f(x)}{\sqrt{F(x) - 1}} \, dx = 2(\sqrt{F(L) - 1}) \\
&= 2(\sqrt{\tau_L - 1}) \leq 2\sqrt{T}.
\end{aligned}$$

□

We introduce the following notation: for $f : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$, let $f^\pi(x) := \sum_a f(x, a)\pi(x, a)$. For all $t > 0$, let l_t denote the episode number at time t. The following useful lemma is based on a martingale argument.

Lemma 70. *Let $f : \mathcal{X} \times \mathcal{A} \to \mathbb{R}$. Let $\delta' > 0$. We have:*

$$\mathbb{P}\left[\sum_{t=1}^{T} (\langle f^{\pi_{l_t}}, p\rangle - f(x_t, a_t)) \leq \sqrt{2T \log(1/\delta')}\right] \geq 1 - \delta'$$
$$\mathbb{P}\left[\sum_{t=1}^{T} \left|\langle f^{\pi_{l_t}}, p\rangle - f(x_t, a_t)\right| \leq \sqrt{2T \log(2/\delta')}\right] \geq 1 - \delta'.$$

Proof. We define the filtration $\mathcal{F}_t = \sigma(x_1, a_1, ..., x_t, a_t)$ and we first show that the sequence defined by $M_t = \langle f^{\pi_{l_t}}, p\rangle - f(x_t, a_t)$ is a martingale difference sequence w.r.t. $\mathcal{F}_t$. $\mathbb{E}[M_t] < \infty$ since the rewards are bounded. Next, the proof that $\mathbb{E}[M_t|\mathcal{F}_{t-1}] = 0$ relies on the fact that l_t, and in turn the stationary policy π_{l_t}, are $\mathcal{F}_{t-1}-$measurable.

Therefore,

$$\mathbb{E}\left[\langle f^{\pi_{l_t}}, p\rangle \middle| \mathcal{F}_{t-1}\right] = \langle f^{\pi_{l_t}}, p\rangle.$$

We also have:

$$\begin{aligned}
\mathbb{E}\left[f(x_t, a_t)\middle|\mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\sum_{x,a} f(x, a)\mathbb{1}_{\{(x_t,a_t)=(x,a)\}} \middle| \mathcal{F}_{t-1}\right] \\
&= \sum_{x,a} f(x, a)\pi_{l_t}(x, a) = \langle f^{\pi_{l_t}}, p\rangle.
\end{aligned}$$

Subtracting the two expressions above, we get $\mathbb{E}[M_t|\mathcal{F}_{t-1}] = 0$. $(M_t)_t$ is thus a Martingale difference sequence, such that $-1 \leq M_t \leq 1$. The result follows from Azuma-Hoeffding's inequality. □

We now prove Theorem 68.

Proof. We define $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$ to be the "good event", with:

$$\mathcal{E}_1 = \{\forall l \geq 1, \tilde{p}_l \in B_l\},$$
$$\mathcal{E}_2 = \{\sum_{t=1}^{T}(\langle r^{\pi_{l_t}}, p\rangle - r(x_t, a_t)) \leq \sqrt{2T\log(3/\delta)}\},$$
$$\mathcal{E}_3 = \left\{\forall i, j, \quad \sum_{t=1}^{T} \left|\langle \xi_j^{i\,\pi_{l_t}}, p\rangle - \xi_j^i(x_t, a_t)\right| \leq \sqrt{2T\log\left(\frac{6\tilde{d}}{\delta}\right)}\right\}.$$

By Lemma 67, we have

$$\mathbb{P}\left[\exists l \geq 1, \tilde{p}_l \in B_l\right] \geq 1 - \frac{\delta}{3}. \tag{E.3}$$

Combining (E.3) with Lemma 70 and using union bounds, $\mathbb{P}[\mathcal{E}] \geq 1 - \delta$. From now on, we assume that the good event $\mathcal{E}$ holds true.

Performance regret We start by upper bounding the performance regret $R(T)$. Let $\Delta_l = \sum_{t\in E_l}(g^* - r(x_t, a_t))$ be the regret of episode l. Let $(\pi_l, \tilde{p}_l)$ be the solution of the optimistic CMDP (D.3) at episode l. Since (π^*, p) is feasible for (D.3), then $g^* \leq g^{\tilde{p}_l, \pi_l}$. We also note that

$$g^{\tilde{p}_l,\pi_l} = \sum_{x,a} r(x,a)\tilde{p}_l(x)\pi_l(x,a) = \sum_x r^{\pi_l}(x)\tilde{p}_l(x).$$

Therefore, we have:

$$\begin{aligned}
\Delta_l &\leq \sum_{t\in E_l}(g^{\tilde{p}_l,\pi_l} - r(x_t,a_t)) \\
&= \sum_{t\in E_l}(\sum_x r^{\pi_l}(x)\tilde{p}_l(x) - r(x_t,a_t)) \\
&= \sum_{t\in E_l}\sum_x r^{\pi_l}(x)(\tilde{p}_l(x) - p(x)) \\
&\quad + \sum_{t\in E_l}(\sum_x r^{\pi_l}(x)p(x) - r(x_t,a_t))
\end{aligned}$$

Using Hölder's inequality and the fact that $\|r\|_\infty = 1$, the first term can be bounded by $|E_l|\|\tilde{p}_l - p\|_1$. By validity of the confidence intervals under event $\mathcal{E}$:

$$\|\tilde{p}_l - p\|_1 \leq 2\beta_l \leq \frac{2\sqrt{2|\mathcal{X}|\log\left(6|\mathcal{X}|T(T-1)/\delta\right)}}{\sqrt{\tau_l - 1}}$$

Summing up over episodes $l = 1, ..., L$:

$$\begin{aligned}
R(T) \leq 2\sqrt{2|\mathcal{X}|\log\left(\frac{6|\mathcal{X}|T(T-1)}{\delta}\right)}\sum_{l=1}^{L}\frac{|E_l|}{\sqrt{\tau_l - 1}} \\
+ \sum_{t=1}^{T}(\sum_x r^{\pi_{l_t}}(x)p(x) - r(x_t,a_t)).
\end{aligned}$$

We bound the first sum using Lemma 69. The second term can be bounded as in Lemma 70 because $\mathcal{E}_2$ holds true. This gives us the resulting bound which holds under $\mathcal{E}$:

$$R(T) \leq 4\sqrt{|\mathcal{X}|\log\left(\frac{6|\mathcal{X}|T(T-1)}{\delta}\right)T} + \sqrt{2T\log\left(\frac{3}{\delta}\right)}.$$

Cost of constraint violations The proof for the cost of constraint violations is very similar. Let us bound $R_j^i(T) := \sum_{t=1}^T |\xi_j^i(x_t, a_t)|$ for all i, j. We briefly drop the sub/superscripts i, j.

At each episode l, since $(\pi_l, \tilde{p}_l)$ is a solution of (D.3), we have $h^{\tilde{p}_l, \pi_l} = 0$, and thus $\sum_{x,a} \xi(x,a)\pi_l(x,a)\tilde{p}_l(x) = \sum_x \xi^{\pi_l}(x)\tilde{p}_l(x) = 0$. Therefore, we have:

$$\begin{aligned}
\left|\sum_{t=1}^T \xi(x_t, a_t)\right| &= \left|\sum_{l=1}^L \Big(\sum_{t\in E_l} \xi(x_t, a_t) - \sum_x \xi^{\pi_l}(x)\tilde{p}_l(x)\Big)\right| \\
&\le \Bigg|\sum_{l=1}^L \sum_{t\in E_l} \sum_x \xi^{\pi_l}(x)(p(x) - \tilde{p}_l(x)) \\
&\qquad + \sum_{l=1}^L \Big(\sum_{t\in E_l} \xi(x_t, a_t) - \sum_x \xi^{\pi_l}(x)p(x)\Big)\Bigg| \\
&\le \sum_{l=1}^L \sum_{t\in E_l} \left|\sum_x \xi^{\pi_l}(x)(p(x) - \tilde{p}_l(x))\right| \\
&\qquad + \left|\sum_{l=1}^L \Big(\sum_{t\in E_l} \xi(x_t, a_t) - \sum_x \xi^{\pi_l}(x)p(x)\Big)\right| \\
&\le \sum_{l=1}^L |E_l| \|\xi^{\pi_l}\|_\infty \|p - \tilde{p}_l\|_1 \\
&\qquad + \left|\sum_{t=1}^T \left(\xi(x_t, a_t) - \sum_x \xi^{\pi_{l_t}}(x)p(x)\right)\right|,
\end{aligned}$$

where the first part of the last inequality is again by Hölder's inequality. Similarly to the performance regret, the first term is bounded using the validity of confidence intervals under the good event $\mathcal{E}$ and Lemma 69, and the second term is bounded by the martingale argument using Lemma 70. Hence, under $\mathcal{E}$ we have for any i, j:

$$R_j^i(T) \le 4\sqrt{|\mathcal{X}| \log\big(\frac{6|\mathcal{X}|T(T-1)}{\delta}\big)T} + \sqrt{2T \log\big(\frac{6\tilde{d}}{\delta}\big)}.$$

And thus the same bounds holds for $R^c(T) = \max_{i,j} R_j^i(T)$.

Representation loss We may now derive the bound on representation loss.

Let $f(T) = O\big(\sqrt{|\mathcal{X}| \log(|\mathcal{X}|T/\delta)}\big)$. The regret bounds imply that with $1 - \delta$:

$$\begin{aligned}
R(T) =& g^*T - N(T) \le f(T) \Rightarrow N(T) \ge g^*T - f(T) \\
\frac{R^c(T)}{N(T)} =& \max_{i,j} \left|\frac{N_j^i(T)}{N(T)} - \rho_j^i \frac{N(T)}{N(T)}\right| \le \frac{f(T)}{N(T)} \\
\text{i.e., } & \|\lambda(C_T) - \rho\|_\infty \le \frac{f(T)}{N(T)}.
\end{aligned}$$

Therefore, using $N(T) \geq 1$, we have:

$$\begin{aligned}\|\lambda(C_T) - \rho\|_\infty &\leq \frac{f(T)}{\max(1, g^*T - f(T))} \\ &= O\left(\sqrt{\frac{|\mathcal{X}|\log(|\mathcal{X}|T/\delta)}{{g^*}^2 T}}\right).\end{aligned}$$

□

E.3 Alternative to RL-CMDP with Bernstein bounds

We present RL-CMDP-B, an alternative to RL-CMDP which uses Bernstein empirical bounds Maurer and Pontil [2009].

At each episode l, the algorithm estimates the distributions by $\hat{p}_l(x) = \frac{n_{\tau_l - 1}(x)}{\tau_l - 1}$ and maintains confidence intervals $[\underline{p}_l(x), \overline{p}_l(x)]$. These are built using Bernstein's empirical inequality Maurer and Pontil [2009], which implies that there exists constants B_1, B_2 such that with probability $\geq 1 - \frac{\delta}{3}$, for each $l \geq 1$ and $x \in \mathcal{X}$,

$$|p(x) - \hat{p}_l(x)| \leq B_1 \sqrt{\frac{\hat{\sigma}_l^2(x)\log(\frac{6|\mathcal{X}|\tau_l}{\delta})}{1 \wedge (\tau_l - 1)}} + B_2 \frac{\log(\frac{6|\mathcal{X}|\tau_l}{\delta})}{1 \wedge (\tau_l - 1)}, \tag{E.4}$$

where $\hat{\sigma}_l(x) = \sqrt{\hat{p}_l(x)(1 - \hat{p}_l(x))}$.

Following e.g. Efroni et al. [2020], we re-write (D.3) as an extended LP by introducing the state-action occupation measure $\mu(x, a) = \pi(x, a)p(x)$.

$$\begin{aligned}\max_{\mu \in \mathbb{R}^{\mathcal{X} \times \mathcal{A}}} \quad & \sum_{x,a} \mu(x,a) r(x,a) \\ \text{u.c.} \quad & \mu \geq 0, \sum_{x,a} \mu(x,a) = 1 \\ & \forall x, \sum_a \mu(x,a) \leq \overline{p}_l(x) \\ & \forall x, \sum_a \mu(x,a) \geq \underline{p}_l(x) \\ & \forall i, j, \sum_{x,a} \mu(x,a) \xi_j^i(x,a) = 0.\end{aligned}$$

The second to fourth constraints enforce the compatibility of μ with the confidence intervals. Controlling each entry of p with Bernstein bounds instead of the ℓ^1-norm allows for a simpler optimization problem than the extended LP (D.4). We get the following regret bound:

Theorem 71 (Regret guarantees)**.** *With probability $\geq 1 - \delta$, the regret of RL-CMDP-B satisfies:*

$$\begin{aligned}R(T) &= O\big(\sqrt{|\mathcal{X}|T\log(|\mathcal{X}|T/\delta)} + |\mathcal{X}|\log(|\mathcal{X}|T/\delta)^2\big) \\ R^c(T) &= O\big(\sqrt{|\mathcal{X}|T\log(|\mathcal{X}|T/\delta)} + |\mathcal{X}|\log(|\mathcal{X}|T/\delta)^2\big).\end{aligned}$$

With probability $\geq 1 - \delta$, the representation loss satisfies:

$$\begin{aligned}&\|\lambda(C_T) - \rho\|_\infty \\ &= O\left(\frac{1}{g^*}\sqrt{\frac{|\mathcal{X}|\log\left(|\mathcal{X}|T/\delta\right)}{T}} + \frac{|\mathcal{X}|\log(|\mathcal{X}|T/\delta)^2}{g^* T}\right).\end{aligned}$$

When using Bernstein bounds, the representation loss carries $O(|\mathcal{X}|\log(|\mathcal{X}|T/\delta)^2)$. This factor but has a bigger scaling with $|\mathcal{X}|$, but decreases rapidly in $\frac{\log(T)^2}{T}$.

The Bernstein version of `RL-CMDP` may be advantageous for some candidate distributions p. For example, if the support $\mathcal{S}$ of p is very small compared to $\mathcal{X}$, the first term in the Bernstein empirical inequality (E.4) is equal to zero for all x outside the support. Therefore, the representation loss scales as:

$$\begin{aligned}&\|\lambda(C_T) - \rho\|_\infty \\ &= O\left(\frac{1}{g^*}\sqrt{\frac{|\mathcal{S}|\log\left(|\mathcal{S}|T/\delta\right)}{T}} + \frac{|\mathcal{X}|\log(|\mathcal{X}|T/\delta)^2}{g^* T}\right),\end{aligned}$$

where $|\mathcal{S}| \ll |\mathcal{X}|$. Thus, the second term with fast decrease in $\frac{\log(T)^2}{T}$ controls the bound on representation loss.

E.3.1 Proofs

The following lemma states a useful inequality akin to Lemma 69.

Lemma 72. *We have:*

$$\sum_{l=1}^{L} \frac{|E_l|}{\tau_l - 1} \leq \log(T)$$

Proof. The proof is similar to Lem. 13 in Zanette and Brunskill [2019]. Using the same notation as in the proof of Lemma 69,

$$\begin{aligned}\sum_{l=1}^{L} \frac{|E_l|}{\tau_l - 1} = \sum_{l=1}^{L} \frac{f(l)}{F(l) - 1} &= \int_1^L \frac{f(\lceil x \rceil)}{F(\lceil x \rceil) - 1}\, dx \\ &\leq \int_1^L \frac{f(x)}{F(x) - 1}\, dx = \log(F(L) - 1) \\ &= \log(\tau_L - 1) \leq \log T.\end{aligned}$$

□

We now prove Theorem 71.

Proof. We re-use the same steps and notation as for the proof of Theorem 68.

Here instead, $\mathcal{E}_1$ is the event such that the confidence intervals are valid (E.4). Under the high-probability good event $\mathcal{E} = \mathcal{E}_1 \cap \mathcal{E}_2 \cap \mathcal{E}_3$, we thus have:

$|\pi_l(x) - p(x)| \lesssim \sqrt{\frac{\hat{p}_l(x)(1-\hat{p}_l(x))b_{\delta,T}}{\tau_l - 1}} + \frac{b_{\delta,T}}{\tau_l - 1}$

where $b_{\delta,T} = \log(\frac{6|\mathcal{X}|T}{\delta})$.

In the following, the first inequality is by validity of the Bernstein confidence intervals under $\mathcal{E}$,

and the second inequality is by Cauchy-Schwarz's inequality:

$$\begin{aligned}
&\sum_{t\in E_l}\sum_x r^{\pi_l}(x)(\tilde{p}_l(x)-p(x)) \\
&\leq \sum_{t\in E_l}\sum_x r^{\pi_l}(x)\sqrt{\frac{\hat{p}_l(x)(1-\hat{p}_l(x))b_{\delta,T}}{\tau_l-1}} \\
&\quad + \sum_{t\in E_l}\frac{b_{\delta,T}}{\tau_l-1}\underbrace{\sum_x r^{\pi_l}(x)}_{\leq|\mathcal{X}|} \\
&\leq \sum_{t\in E_l}\sqrt{\underbrace{(\sum_x 1-\hat{p}_l(x))}_{\leq|\mathcal{X}|}\underbrace{(\sum_x \hat{p}_l(x)r^{\pi_l}(x)b_{\delta,T})}_{\leq b_{\delta,T}}}\sqrt{\frac{1}{\tau_l-1}} \\
&\quad + \sum_{t\in E_l}\frac{|\mathcal{X}|b_{\delta,T}}{\tau_l-1}
\end{aligned} \tag{E.5}$$

By Lemmas 69 and 72, we have:

$$\begin{aligned}
\sqrt{|\mathcal{X}|b_{\delta,T}}\sum_{l=1}^{L}\frac{|E_l|}{\sqrt{\tau_l-1}} &\leq 2\sqrt{|\mathcal{X}|b_{\delta,T}T} \\
|\mathcal{X}|b_{\delta,T}\sum_{l=1}^{L}\frac{|E_l|}{\tau_l-1} &\leq |\mathcal{X}|b_{\delta,T}\log(T).
\end{aligned}$$

Summing up over episodes in inequality (E.5) and plugging in the above inequalities gives the desired bound by following the steps of the proof of Theorem 68. □

E.4 Experiments

E.4.1 Details on the Brexit experiments

We provide in Table E.1 the target vectors $(\rho_j^i)_{i,j}$ and marginal distributions $(\mathbb{P}^p[x^i=j])_{i,j}$ extracted from the Citizens' Assembly on Brexit report Renwick et al. [2017].[1] The report includes the volunteering rates for each feature group, i.e. $\Pr[\text{volunteer}|x^i=j]$. To compute the marginal distributions $(\Pr[x^i=j|\text{volunteer}])_{i,j}$, we thus use Bayes' rule to compute the probability of each feature group among the volunteer population[2], that is:

$$\begin{aligned}
\mathbb{P}^p[x^i=j] &= \Pr[x^i=j|\text{volunteer}] \\
&= \frac{\Pr[\text{volunteer}|x^i=j]\Pr[x^i=j]}{\Pr[\text{volunteer}]}.
\end{aligned}$$

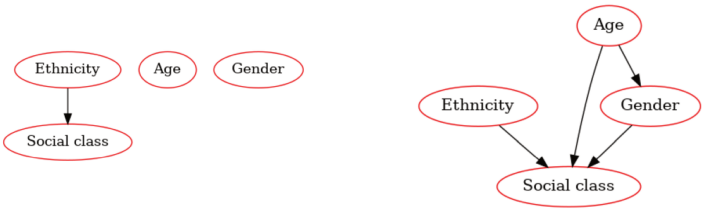
We often have $\rho_j^i \neq \mathbb{P}^p[x^i=j]$. For example, compared to the age target, we are less likely to find younger people (≤ 34 years old) among volunteers. For gender, while the target was gender parity, we are much less likely to find women than men in the volunteer population.

[1] https://citizensassembly.co.uk/wp-content/uploads/2017/12/Citizens-Assembly-on-Brexit-Report.pdf, pages 28-32.

[2] In doing so, we notice that the probability of finding non-voter volunteers is almost zero, hence we only consider "remain" and "leave" for the feature Brexit vote. Indeed, the report states "The only target that proved impossible to meet was that for non-voters in the 2016 referendum." p.28.

	Targets	Marginals
Ethnicity	0.860 / 0.140	0.863 / 0.136
Social class	0.550 / 0.450	0.556 / 0.444
Age	0.288 / 0.344 / 0.367	0.154 / 0.432 / 0.414
Region	0.233 / 0.160 / 0.093 / 0.134 / 0.222 / 0.047 / 0.082 / 0.028	0.179 / 0.155 / 0.090 / 0.117 / 0.211 / 0.073 / 0.154 / 0.021
Gender	0.507 / 0.493	0.384 / 0.616
Brexit vote	0.481 / 0.519	0.565 / 0.434

Table E.1: Target quotas (from the report) and marginal distribution (computed using Bayes' rule) for the Brexit experiment.



(a) Structure 1: weak dependence. (b) Structure 2: strong dependence.

Figure E.1: Bayesian network structures for the Census Income dataset.

For our experiments presented in Section D.6, we used Python and the CPLEX LP solver, and a machine with Intel Xeon Gold 6230 CPUs, 2.10 GHz, 1.3 MiB of cache.

E.4.2 Experiments with dependent features

The goal of these experiments is to answer the following: what is the impact of the dependence structure of the joint feature distribution p on the sample complexity and representation loss of our algorithms? Since we may only retrieve marginal distributions from the Citizen's Assembly on Brexit report, we keep the target quotas on each feature but simulate joint feature distributions from another dataset with demographic attributes, the standard Adult Census Income dataset Dua and Graff [2017].

The Adult dataset consists of approximately 49.000 entries of subjects in the US, each with 14 demographic features and a binary label indicating whether a subject's income is above or below 50K USD. We only keep features that can be mapped to our Brexit Citizen's Assembly example: gender, age, ethnicity and income, which we use in lieu of social class. We do not consider proportional representation for region and Brexit vote since there are no such features in the Adult dataset. In our preprocessing of the Adult dataset, we create the same three age categories (<35, 35-54, >54), the same two ethnicity groups (white / non-white) and we use the binary income variable as a proxy for social class, by assigning $>$ 50K to upper class and $\leq$ 50K to lower class. This leaves us with 4 features with $2, 2, 2, 3$ possible values.

To create dependencies between features, we consider two graphical structures shown in Figure E.1, and for each we fit a Bayesian network to the dataset to generate a model of the joint distribution $p(x)$. We consider one structure with little dependence, and one structure with strong dependence between features.

Figure E.2 shows that both when p is known, the sample complexity is higher when there is more dependence (Bayesian network (2)) between features, but the representation loss is the same.

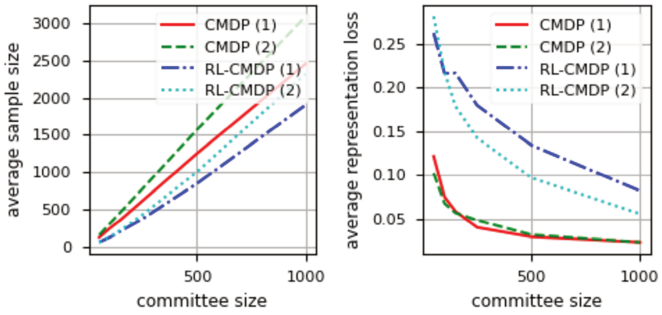


Figure E.2: Effect of committee size K on sample complexity and representation loss for `CMDP` (**known** p) and `RL-CMDP` (**unknown** p), on the two different Bayesian networks (1) and (2) fitted on the Census Income dataset.

gender \ age	S	J
M	$1/3$	$1/4$
F	$1/4$	$1/6$

When p is unknown, the representation loss is lower for structure (2) with more dependence, than structure (1) with little dependence, but the sample size is higher for (2). For structure (2), the representation loss is low (≈ 0.07) for $K = 1000$. Importantly, it implies that in practice, the representation loss is much lower than the worst case bound given by Theorem 68.

E.5 Detailed example for Section D.4

We take the same attributes and same distribution as in Table D.1, with $\epsilon' = 1/6$:

The target vectors are $\rho^{\text{gender}} = (1/2, 1/2)$ and $\rho^{\text{age}} = (1/2, 1/2)$, that is, an ideal committee contains as many women as men and as many senior than junior.

We solve the linear program

$$\max \quad \frac{\pi(MS,1)}{3} + \frac{\pi(FS,1)}{4} + \frac{\mu(MJ,1)}{4} + \frac{\mu(FJ,1)}{6}$$

$$\text{u.c.} \quad \frac{\pi(MS,1)}{3} + \frac{\pi(FS,1)}{4} = \frac{\mu(MJ,1)}{4} + \frac{\mu(FJ,1)}{6}$$
$$\frac{\pi(MS,1)}{3} + \frac{\mu(MJ,1)}{4} = \frac{\pi(FS,1)}{4} + \frac{\mu(FJ,1)}{6}$$

Its solution is

$$\pi^*(MS,1) = 1/2$$
$$\pi^*(FJ,1) = 1$$
$$\pi^*(MJ,1) = 1$$
$$\pi^*(FS,1) = 1$$

Thus, each time the current volunteer is a senior male, we select him with probability $1/2$; all other volunteers are selected with probability 1. The expected final composition of the pool is 30% of junior male, 30% of senior female, 20% of junior female and 20% of senior male. As the policy selects in average $5/6$ of the volunteers, the expected time until we select K candidates is $\mathbb{E}^{p,\pi^*}[\tau] = (6/5)K$.

Appendix F

Résumé de la thèse en français

Contents

Résumé

Les algorithmes d'apprentissage automatique (*machine learning*) sont largement utilisés dans les systèmes de recommandation qui alimentent les plateformes de streaming, de commerce et les réseaux sociaux. Leur principal objectif est de fournir aux utilisateurs des recommandations personnalisées en prédisant leurs préférences et en triant les contenus disponibles en fonction de ces prédictions. Cependant, en sélectionnant le contenu de certains producteurs plutôt que d'autres, les algorithmes de recommandation décident de qui est visible ou non. Ces décisions ont de réelles implications éthiques et sociales, comme les risques d'invisibilisation de groupes minoritaires ou défavorisés dans la suggestion de profils à des employeurs, ou les problèmes de sous- ou surreprésentation de certaines opinions et cultures sur les réseaux sociaux. Il est donc devenu crucial de garantir que ces décisions automatisées soient non biaisées et équitables envers les producteurs de contenu, en évitant de donner à certains groupes un avantage ou un désavantage excessif. En plus de décider quels producteurs sont visibles, les algorithmes de recommandation jouent également un rôle clé dans la décision de quels utilisateurs sont exposés à certains contenus, notamment les contenus associés à des opportunités économiques telles que les offres d'emploi et annonces immobilières. Par conséquent, des préoccupations se posent quant à l'équité d'accès à ces opportunités parmi les

utilisateurs des systèmes de recommandation.

Cette thèse vise à adresser les limites des algorithmes de recommandation actuels en développant des systèmes plus équitables qui tiennent compte à la fois des utilisateurs et des producteurs de contenu. Cependant, le développement d'algorithmes équitables présente plusieurs défis, notamment la définition de critères d'équité appropriés et l'implémentation efficace d'algorithmes de *ranking* qui satisfont ces critères. En nous appuyant sur la riche littérature de la théorie du choix social, nous proposons un cadre conceptuel pour évaluer l'équité des listes ordonnées de recommandations, à partir de concepts établis pour les problèmes de partage équitable qui ont été peu étudiés en *machine learning* et en recommandation. Dans ce cadre conceptuel, nous développons de nouvelles méthodes de recommandation qui suivent les principes du partage équitable et distribuent l'exposition plus équitablement entre les producteurs de contenu, sans compromettre la qualité des recommandations pour les utilisateurs. Ces méthodes sont soutenues par des résultats théoriques sur la satisfaction de propriétés d'équité, sur les garanties de convergence et l'efficacité algorithmique des algorithmes proposés, ainsi que par des évaluations expérimentales sur des jeux de données publics.

F.1 Les impacts sociétaux des systèmes de recommandation

Les systèmes de recommandation font partie intégrante des plateformes numériques modernes, desservant jusqu'à des milliards d'utilisateurs dans le monde entier. Ces systèmes sont présents sur les places de marché en ligne, les services de streaming, les plateformes de partage de contenu et les médias sociaux en ligne. Ils jouent un rôle crucial dans l'organisation de la vaste quantité d'informations disponibles en fournissant des recommandations personnalisées aux utilisateurs à diverses fins, comme la navigation d'articles d'actualité, la recherche de produits, d'emplois, de logements ou de personnes avec lesquelles se connecter.

À l'ère de l'apprentissage automatique et de son adoption croissante dans de nombreuses applications qui affectent notre vie quotidienne, les systèmes de recommandation se démarquent comme l'une des applications les plus réussies des algorithmes d'apprentissage automatique. L'apprentissage automatique a été instrumental pour exploiter les vastes quantités de données disponibles sur les plateformes en ligne pour personnaliser l'expérience utilisateur et faciliter la découverte de nouveaux items pertinents. Ces algorithmes analysent les modèles statistiques du comportement de navigation passé des utilisateurs, les interactions avec les items, les préférences exprimées et d'autres caractéristiques pour prédire leurs intérêts futurs. Ces prédictions permettent la récupération d'items à recommander dans le but de maximiser l'engagement des utilisateurs, comme l'augmentation du nombre de clics, de likes, de partages ou de temps passé sur la plateforme. L'apprentissage automatique offre la promesse de recommandations hautement personnalisées qui reflètent les goûts

et les préférences individuelles, conduisant à une plus grande satisfaction des utilisateurs et à une utilisation accrue de la plateforme.

Cependant, au-delà de la promesse d'une augmentation de l'engagement des utilisateurs, les algorithmes de recommandation ont des conséquences sociales profondes. Avec le pouvoir de décider qui est visible et qui ne l'est pas, ces algorithmes ont un impact significatif sur les producteurs d'items (Figure 1.1). Par exemple, les organes de presse dépendent de leur exposition sur les fils d'actualité pour générer des revenus auprès des lecteurs, tandis que les créateurs sur les plateformes de partage de contenu et les artistes sur les plateformes de streaming dépendent des spectateurs et des auditeurs pour rester durables. De même, l'attractivité des commerçants comme les restaurants et les magasins dépend largement de leur exposition aux clients potentiels dans les recommandations locales de Google Maps. Le succès d'un chercheur d'emploi sur les plateformes de recherche d'emploi comme LinkedIn dépend du recruteur qui arrive à voir son CV, et l'efficacité d'une application de rencontres dépend également des utilisateurs à qui le profil de quelqu'un est recommandé.

En déterminant quels producteurs d'items sont visibles ou non, les systèmes de recommandation prennent des décisions qui posent de réelles préoccupations éthiques et sociales. Celles-ci incluent les risques de négliger ou de désavantager les chercheurs d'emploi de groupes sous-représentés [Geyik et al., 2019], d'amplifier les biais raciaux dans les applications de rencontres [Hutson et al., 2018] et de surreprésenter les groupes démographiques, culturels ou politiques sur les médias sociaux et les résultats de recherche. Par exemple, des recherches ont montré que les femmes sont systématiquement sous-représentées dans les résultats de recherche pour diverses professions [Kay et al., 2015]. Nous fournissons un nouvel exemple de cela à la Figure 1.2, où les résultats de la recherche pour le terme "DJ" montrent principalement des DJs masculins. D'autres recherches sur Twitter ont montré que leur algorithme de recommandation favorisait le contenu des politiciens et des médias d'extrême droite sur le contenu de gauche [Huszár et al., 2022], une découverte qui a reçu une grande couverture médiatique[1]. Les systèmes de recommandation ont également le potentiel de favoriser de manière disproportionnée les créateurs et les artistes établis sur les plateformes de partage de contenu, conduisant à la marginalisation et à l'éventuel déclin de ceux qui sont plus petits et qui ne reçoivent pas assez d'exposition pour réussir [Mehrotra et al., 2018]. Pour atténuer l'impact potentiellement négatif des systèmes de recommandation sur les producteurs d'items, il est crucial d'évaluer soigneusement leurs implications sociétales et de s'assurer qu'ils ne désavantagent pas injustement certains groupes.

Du côté des utilisateurs, les systèmes de recommandation sont traditionnellement conçus pour leur fournir les items les plus pertinents, un objectif qui semble en adéquation avec leurs intérêts. Cependant, des inquiétudes ont été soulevées ces dernières années quant à l'impact des algorithmes

[1] voir par exemple, The Guardian https://www.theguardian.com/technology/2021/oct/22/twitter-admits-bias-in-algorithm-for-rightwing-politicians-and-news-outlets

de recommandation sur les utilisateurs. Des audits des systèmes de recommandation ont révélé des disparités dans le contenu livré à divers groupes sociaux d'utilisateurs. Par exemple, Datta et al. [2015] a découvert que les femmes également qualifiées recevaient moins de publicités en ligne pour des emplois bien rémunérés que les hommes. Pour prévenir le risque de livraison inéquitable d'opportunités à travers les utilisateurs, des efforts significatifs ont été faits pour auditer les systèmes de recommandation pour des biais involontaires ou de la discrimination envers leurs utilisateurs. Ces efforts appellent au développement de nouveaux algorithmes de recommandation qui offrent un accès équitable à l'information et aux opportunités à leurs utilisateurs.

Étant donné les impacts réels des systèmes de recommandation sur leurs utilisateurs et les producteurs d'items, l'équité dans les systèmes de recommandation est devenue un sujet central dans la recherche en apprentissage automatique et en récupération de l'information. L'équité dans les systèmes de recommandation peut être examinée d'au moins deux côtés différents : le côté des items et le côté des utilisateurs. Du côté des items, l'objectif est de fournir aux producteurs d'items une part équitable de l'exposition dans les recommandations. Du côté des utilisateurs, il est nécessaire de s'assurer que les systèmes de recommandation ne créent pas ou n'amplifient pas de biais involontaires et fournissent des recommandations qui bénéficient à tous les utilisateurs. Il y a une demande croissante pour des systèmes de recommandation qui atteignent simultanément ces deux objectifs, afin de maintenir un écosystème de recommandation sain qui sert les intérêts de toutes les parties prenantes [Patro et al., 2020, Abdollahpouri et al., 2020]. L'impact sociétal des systèmes de recommandation est significatif, et assurer l'équité pour les utilisateurs et les producteurs d'items est crucial pour éviter de perpétuer ou d'amplifier les biais et les inégalités existants.

L'équité dans les systèmes de recommandation est un point focal dans un débat plus large et actif sur les impacts sociétaux des algorithmes d'apprentissage automatique. Alors que les modèles d'apprentissage automatique continuent de gagner en traction dans notre vie quotidienne, il y a eu une préoccupation publique croissante quant au potentiel des modèles d'apprentissage automatique à introduire des biais et de la discrimination dans les décisions algorithmiques [Buolamwini and Gebru, 2018, Barocas and Selbst, 2016]. Par conséquent, l'équité est devenue un sujet central dans la recherche en apprentissage automatique, en particulier dans le contexte de la classification et de l'apprentissage supervisé [Barocas et al., 2019]. Face au potentiel des algorithmes pour perpétuer les biais et la discrimination dans la prise de décision, les chercheurs ont proposé une série de métriques et de méthodes d'équité pour répondre à ces préoccupations dans diverses tâches d'apprentissage supervisé, y compris la prédiction de la récidive, l'embauche et la notation de crédit. Ces méthodes visent à garantir que les algorithmes ne perpétuent pas des pratiques injustes, telles que les différences de traitement ou de résultats basées sur le sexe, la race ou d'autres

caractéristiques protégées. Dans ce chapitre, nous nous pencherons sur le rôle clé de l'équité dans les systèmes de recommandation au sein du vaste et constamment évolutif champ de l'apprentissage automatique équitable, et nous présenterons nos contributions à ce domaine critique.

F.2 Problèmes d'équité dans les systèmes de recommandation

F.2.1 Sources d'inéquité dans les systèmes de recommandation

Aperçu des systèmes de recommandation La tâche d'un système de recommandation est de fournir à chacun de ses utilisateurs une liste d'items classés, qui sont sélectionnés à partir d'un vaste ensemble d'items candidats (par exemple, des vidéos) fournis par des producteurs (par exemple, des créateurs de vidéos). Le système de recommandation évalue la qualité des classements à l'aide de "vrais" scores de pertinence, qui mesurent la valeur d'un item pour un utilisateur. À un niveau élevé, les algorithmes de recommandation s'appuient sur deux étapes pour générer des recommandations classées :

1. **Apprentissage (*Learning*):** Estimer la valeur de chaque item pour chaque utilisateur. Cette opération est réalisée à l'aide d'un modèle d'apprentissage automatique qui s'appuie sur les interactions passées des utilisateurs avec les items, les caractéristiques des items (par exemple, la catégorie, la date de publication) et les caractéristiques de l'utilisateur (par exemple, l'âge, le pays).
2. **Classement (*Ranking*):** Choisir un classement des K meilleurs items pour chaque utilisateur en fonction des scores estimés. Il en résulte une politique de classement personnalisée, dans le cadre de laquelle des listes d'items différentes sont recommandées à différents utilisateurs en fonction de leurs préférences prédites.

Traditionnellement, l'étape de classement consiste simplement à trier les items par scores décroissants pour chaque utilisateur. Lorsque le système de recommandation dispose des véritables scores de pertinence, cette stratégie est optimale pour maximiser les mesures de classement standard, telles que le *discounted cumulated gain* (DCG) [Järvelin and Kekäläinen, 2002], qui mesure la qualité des classements du point de vue de l'utilisateur. Toutefois, il ne tient pas compte des producteurs d'items qui sont exposés dans les classements. Comme nous l'avons vu précédemment, il s'agit d'une question d'équité cruciale, car la visibilité accordée aux producteurs d'items (ou l'absence de visibilité) a des conséquences sociales réelles.

L'objectif principal de cette thèse est d'aborder la question de l'équité dans l'étape de classement des systèmes de recommandation. L'étape de classement est une étape cruciale au cours de laquelle

le système de recommandation décide quels items seront recommandés à quels utilisateurs, une fois que les préférences de ces derniers ont été estimées. Il s'agit d'une décision collective qui a un impact à la fois sur les utilisateurs et sur les producteurs d'items. Par conséquent, notre objectif est de veiller à ce que l'étape de classement tienne compte des intérêts des utilisateurs et des producteurs d'items, et les équilibre de manière équitable.

Sources d'inéquité. L'étape de classement, ou la combinaison des étapes d'apprentissage et de classement, peut avoir des conséquences involontaires et indésirables de différentes manières. L'étape de classement peut produire des effets de "winner-take-all", où certains groupes de producteurs d'items s'emparent de toute l'exposition disponible. Dans la solution de classement traditionnelle, qui consiste simplement à classer les items en fonction de leur score, même de petites différences de score entraînent de grandes différences d'exposition entre les producteurs d'items. Il en résulte un effet de longue traîne où quelques items populaires tendent à dominer les positions les plus élevées du classement, laissant de côté un grand nombre d'items moins populaires avec peu ou pas d'exposition (figure 1.3). Cet effet de longue traîne peut être problématique pour les petits producteurs, car ils luttent pour gagner en visibilité ou en reconnaissance, ce qui exacerbe encore la distribution en loi de puissance de l'exposition [Abdollahpouri et al., 2019b]. En outre, des biais systématiques dans l'estimation des préférences peuvent résulter de l'apprentissage de stéréotypes ou de biais de popularité [Mehrotra et al., 2018]. Ces biais à l'étape de l'apprentissage peuvent être amplifiés par l'étape du classement, où les items des groupes défavorisés dont les valeurs sont systématiquement sous-estimées ne sont finalement pas présentés aux utilisateurs (figure 1.2).

La combinaison de l'apprentissage et du classement peut également conduire à des résultats injustes du côté des utilisateurs. Au cours de l'étape d'estimation, les systèmes de recommandation s'appuient souvent sur des hypothèses de modélisation solides et sur l'apprentissage multitâche pour faire face à la rareté des données par utilisateur, avec des méthodes telles que la factorisation matricielle de faible rang [Koren et al., 2009]. La capacité limitée des modèles ou des hypothèses incorrectes peuvent laisser de côté les utilisateurs dont les schémas de préférence sont moins courants. De ce fait, le système peut apprendre à tort les goûts stéréotypés des utilisateurs, tels que les associations sexuées entre les préférences des utilisateurs et les catégories d'emploi. L'étape de classement amplifie alors ces biais en classant les items en fonction des valeurs estimées, ce qui se traduit par de mauvaises performances en matière de recommandation pour les utilisateurs ayant des goûts non stéréotypés [Ekstrand et al., 2018] ou par des biais dans la recommandation de certains contenus pour des groupes sensibles [Sweeney, 2013, Imana et al., 2021]. En outre, dans le cas des marchés publicitaires, des biais dans la diffusion des publicités apparaissent lorsque la décision de classement tient compte des résultats d'une vente aux enchères dans laquelle les

annonceurs sont en concurrence pour le même groupe d'utilisateurs [Ali et al., 2019]. Par exemple, les annonceurs d'offres d'emploi sont parfois en concurrence avec des annonces de produits ciblant les femmes, ce qui fait qu'elles sont diffusées à moins de femmes que d'hommes.

F.2.2 Recommandation équitable *vs.* classification équitable

Apprentissage et prise de décision en matière de classification L'équité dans les systèmes de recommandation est un domaine de recherche essentiel dans le cadre plus large de l'équité dans l'apprentissage automatique, qui a fait l'objet d'une attention particulière ces dernières années. Si les systèmes de recommandation peuvent être décomposés en une étape *apprentissage* et une étape *classement*, de nombreuses autres applications d'apprentissage automatique comportent également ces composantes *apprentissage* et *prise de décision* [Kleinberg et al., 2018b, Kilbertus et al., 2020, Corbett-Davies et al., 2017]. Le cadre le plus étudié est celui de la classification équitable (binaire), où l'objectif est de prédire une étiquette binaire pour chaque individu, par exemple s'il remboursera ou non un prêt, afin d'aider à la prise d'une décision, comme l'acceptation ou le refus d'une demande de prêt. D'autres exemples courants sont la prédiction de la récidive et l'embauche [Corbett-Davies and Goel, 2018, Barocas et al., 2019]. Nous examinons comment les considérations d'équité dans les étapes d'apprentissage et de décision des tâches de classification sont liées aux considérations d'équité dans les étapes d'apprentissage et de classement des systèmes de recommandation.

Considérons un exemple classique dans la littérature sur l'apprentissage automatique équitable, où un prêteur utilise un algorithme pour déterminer s'il doit ou non approuver une demande de prêt [Hardt et al., 2016b, Liu et al., 2018]. Dans l'étape *apprentissage*, un algorithme d'apprentissage supervisé produit un score pour chaque individu en estimant la probabilité qu'il appartienne à la classe positive (c'est-à-dire la probabilité qu'il rembourse le prêt). Ce score est prédit par un classificateur probabiliste formé sur la base de données historiques. Des injustices peuvent survenir lors de l'étape d'apprentissage lorsque les données utilisées pour former le modèle ne sont pas représentatives de la population à laquelle il est appliqué. Le modèle qui en résulte peut ne pas être performant sur des données inédites provenant d'une population différente, ou il peut apprendre des associations problématiques entre des attributs sensibles et des résultats. Dans l'exemple du prêt, si les données d'apprentissage contiennent une majorité de demandes de prêt infructueuses émanant de personnes d'une certaine race ou d'un certain milieu socio-économique, le modèle résultant peut produire des estimations biaisées en défaveur de ces groupes. Cela peut conduire à des résultats injustes où certains groupes se voient systématiquement refuser l'accès aux prêts en raison d'une sous-estimation systématique de leur solvabilité. Comme nous l'avons vu dans la section précédente, les algorithmes d'apprentissage visant à prédire la valeur des items dans les systèmes de recommandation peuvent également surestimer la valeur des items populaires en raison

de l'absence de retour d'information de la part des utilisateurs pour les items moins populaires dans les données historiques.

Dans l'étape de prise de décision, les individus sont classés comme positifs ou négatifs sur la base de leurs scores prédits. Dans l'exemple du prêt, la décision d'accepter une demande de prêt est basée sur le fait que le demandeur est prédit comme étant solvable, ce qui est fait en appliquant un seuil à la probabilité estimée de remboursement. Le seuil de décision peut avoir des implications importantes en termes d'équité, car il détermine quels individus sont considérés comme éligibles pour certaines opportunités ou services de la vie. En particulier, lors du choix de politiques de seuil spécifiques à un groupe (ou agnostiques), la distribution des résultats positifs qui en résulte peut ou non conduire à des gains de bien-être pour les groupes défavorisés [Kleinberg et al., 2018b, Corbett-Davies et al., 2017].

Dans les systèmes de recommandation, les algorithmes de classement prennent également une décision sur les bénéficiaires des résultats positifs. Cette décision est plus complexe que les seuils de classification binaires, et ce pour au moins deux raisons. Premièrement, elle consiste à produire un classement des items pour chaque utilisateur, au lieu d'un simple seuil par utilisateur. Deuxièmement, elle implique des compromis complexes entre les intérêts des différentes parties prenantes qui apprécient différemment les recommandations : les utilisateurs recherchent les classements qui correspondent le mieux à leurs préférences, tandis que les items recherchent une forte exposition - par conséquent, la notion de résultat positif n'est pas absolue.

Dans cette thèse, nous nous concentrons sur l'équité de la décision qui se produit à l'étape du classement des systèmes de recommandation, plus précisément sur le problème de planification sociale qui consiste à choisir un compromis entre les utilités des utilisateurs et des items (nous clarifions plus tard les définitions des utilités dans la section 1.3). Cette position est similaire à celle de Kleinberg et al. [2018b] qui affirme que les considérations d'équité devraient affecter la manière dont le planificateur social utilise les scores appris pour prendre une décision, plutôt que le choix de l'algorithme d'apprentissage, dans le contexte des problèmes de décision binaires (c'est-à-dire les admissions à l'université).

Critères d'équité dans la classification. Des critères d'équité ont été proposés pour les étapes d'apprentissage et de décision. L'équité des résultats obtenus lors de l'étape d'apprentissage a fait l'objet d'études approfondies dans le domaine de la classification. Les critères comprennent *calibration* entre les groupes et *parité*[2] des scores prédits [Kleinberg et al., 2016, Pleiss et al., 2017]. Dans l'exemple du prêt, la parité exige que le score de crédit moyen soit le même pour tous les groupes, tandis que l'étalonnage exige que la probabilité de remboursement d'un prêt pour un score

[2]Dans la classification, les critères de parité sont plus souvent considérés au niveau des résultats, c'est-à-dire des décisions.

de crédit donné soit la même pour tous les groupes. Dans la littérature sur la recommandation équitable, quelques critères d'équité des scores ont été proposés [Yao and Huang, 2017, Islam et al., 2021], mais plusieurs auteurs ont souligné l'insuffisance de considérer les scores indépendamment de la décision finale, c'est-à-dire les classements [Beutel et al., 2019a, Singh and Joachims, 2018]. En particulier, l'étalonnage des scores ne s'étend pas trivialement aux systèmes de recommandation [Steck, 2018], car l'impact du score d'un item n'est significatif que par rapport aux scores d'autres items [Beutel et al., 2019a].

Les critères d'équité dans l'étape de décision des tâches de classification visent généralement à égaliser les résultats entre les groupes sensibles. *La parité démographique* exige une probabilité égale de résultats positifs parmi les groupes sensibles [Feldman et al., 2015, Zliobaite, 2015] et *l'égalité des chances* [Hardt et al., 2016b] (ou égalité des taux d'erreur [Zafar et al., 2017a, 2019]) visent à égaliser les probabilités de résultats positifs pour la classe positive à travers les groupes. Geyik et al. [2019] proposent une mise en correspondance de la parité démographique et de l'égalité des chances dans le cadre du classement. Lorsque les items sont répartis en groupes sensibles, la parité démographique exige que les groupes d'items bénéficient d'une exposition égale dans les classements, tandis que l'égalité des chances est similaire à un critère populaire basé sur le mérite pour les classements que nous présentons dans la section 1.3.5.

Corbett-Davies et al. [2017], Hu and Chen [2020] insistent sur le coût pour le bien-être social de la recherche de la parité des résultats dans les problèmes de classification, car il est possible d'égaliser les résultats entre les groupes en privant les individus de résultats positifs sans les redistribuer aux individus défavorisés. Dans cette thèse, nous démontrons également les conséquences indésirables de l'application de contraintes d'équité sur l'exposition aux items (chapitre 3). Toutefois, nous soutenons que la réduction des inégalités dans la distribution des résultats est raisonnable dans le cas du classement, où la décision est *allocative*, parce qu'elle peut conduire à des changements positifs dans le bien-être social. En revanche, les décisions prises dans la plupart des problèmes de classification équitable ne sont pas allocatives, car il n'y a pas de budget sur le nombre de classifications positives [Zafar et al., 2019, 2017a, Hardt et al., 2016b, Agarwal et al., 2018]. En d'autres termes, ces travaux portent sur des problèmes de classification stricte plutôt que sur des problèmes de sélection. En pratique cependant, les décisions binaires d'acceptation/refus sont souvent budgétisées : il y a typiquement un budget fixe à dépenser dans les problèmes de prêt, et un nombre fixe de places dans les admissions à l'université. Les considérations budgétaires comme dans [Kleinberg et al., 2018b, Emelianov et al., 2022] rapprochent les problèmes de classification des systèmes de recommandation où il y a un nombre fixe de créneaux de recommandation à allouer. Dans ces contextes budgétisés, il est souhaitable de redistribuer les résultats, car un résultat positif qui est retiré à quelqu'un est nécessairement *transféré* à quelqu'un d'autre. Nous présentons dans la

section suivante, la principale contribution de cette thèse, qui est un cadre pour guider la décision d'attribution du classement dans les systèmes de recommandation, ancré dans les principes de justice distributive du choix social.

F.3 Le choix social pour la recommandation équitable

Cette section présente une contribution clé de cette thèse : un cadre conceptuel pour l'équité dans les systèmes de recommandation qui est ancré dans la théorie du choix social.

F.3.1 Répartition équitable de l'exposition dans les systèmes de recommandation

Comme nous l'avons déjà discuté, à l'étape du classement, les systèmes de recommandation prennent une décision *allocative* collective sur quels items reçoivent de l'exposition, et à quels utilisateurs ils sont exposés. L'équité dans les problèmes d'*allocation*, ou la *division équitable*, a une longue histoire dans la théorie du choix social, qui est une branche de l'économie qui étudie les processus de prise de décision collective basés sur les préférences hétérogènes de plusieurs agents [Arrow et al., 2010, Moulin, 2003]. Dans cette thèse, nous abordons l'équité dans les systèmes de recommandation comme un nouveau problème de division équitable, où la ressource rare à distribuer est la quantité de contenu que le système peut afficher à ses utilisateurs, c'est-à-dire l'exposition totale disponible. Différents producteurs d'items sont en compétition pour une part de cette ressource limitée. Notre point de vue est que le recommandeur est un planificateur social dont le but est de fournir des recommandations classées aux utilisateurs en répartissant équitablement le budget d'exposition parmi les producteurs d'items, tout en tenant compte de l'impact du mécanisme d'allocation sur la satisfaction de l'utilisateur. Nous nous appuyons sur les nombreuses recherches sur la division équitable qui ont été menées dans le passé dans la théorie du choix social et l'économie du bien-être cardinal.

Nous utilisons le terme *utilité* dans son sens large dans l'économie du bien-être cardinal comme une "*mesure de la caractéristique de haut niveau qui est pertinente pour le problème particulier de justice distributive en question*" [Moulin, 2003]. Dans notre problème d'allocation, il y a deux types d'agents – les utilisateurs et les producteurs d'items – qui bénéficient différemment des classements. Les utilisateurs apprécient les classements de haute qualité qui correspondent le mieux à leurs préférences, et les items bénéficient d'un grand nombre de vues. Par conséquent, nous définissons l'utilité de l'utilisateur comme une mesure de performance du classement, et l'utilité de l'item comme le nombre de vues attendu. Le problème d'allocation consiste à choisir des classements en faisant des compromis entre les utilités des utilisateurs et celles des items. Nous nous référons à ce

problème d'allocation comme le problème de la *répartition équitable de l'exposition*.

Comme nous l'avons déjà discuté, l'approche traditionnelle dans les systèmes de recommandation est de maximiser uniquement l'utilité moyenne de l'utilisateur, en classant les items par pertinence décroissante pour chaque utilisateur. Cependant, cette approche peut avoir des effets indésirables, tels que des effets injustes de type "winner-take-all" et l'amplification des biais dans les scores estimés, comme décrit dans la Section 1.2.1. Par conséquent, notre motivation à prendre en compte l'équité de la répartition de l'exposition envers les utilisateurs et les producteurs d'items est d'atténuer et empêcher ces conséquences négatives.

F.4 Plan détaillé et contributions

Nous présentons maintenant le plan de cette thèse et résumons les contributions par chapitre, qui correspondent chacun à un article publié pendant la thèse.

Les deux premiers chapitres, le chapitre 3 et le chapitre 4, se concentrent sur l'équité de l'étape de classement, dans un modèle de recommandation statique.

Chapitre 3 : L'équité dans les classements avec des fonctions de bien-être concaves additives. Nous proposons d'évaluer l'équité des classements pour les utilisateurs et les items dans les systèmes de recommandation sur la base des principes fondamentaux de justice distributive de l'économie du bien-être, fondés sur l'efficacité de Pareto et le principe de transfert de Pigou-Dalton. Nous montrons que certaines approches populaires du classement équitable ne satisfont pas à ces principes. Par exemple, les contraintes d'équité basées sur le mérite peuvent diminuer l'utilité de l'utilisateur tout en augmentant les inégalités d'exposition entre les producteurs d'items, ce qui va à l'encontre du principe de transfert qui vise à réduire les inégalités. Pour surmonter les limites des approches existantes, nous proposons une nouvelle approche pour générer des classements équitables qui est fondée sur l'économie du bien-être. Elle consiste à maximiser les fonctions de bien-être concaves additives, qui sont une famille de fonctions de bien-être lisses. Ces fonctions de bien-être peuvent être interprétées comme des sommes d'utilités d'agents qui ont des rendements décroissants. La propriété des rendements décroissants pour l'exposition signifie qu'*"une vue supplémentaire compte plus pour les items qui ont 10 vues que pour ceux qui ont 10 millions de vues"*, ce qui est particulièrement pertinent pour les systèmes de recommandation. Les classements obtenus en maximisant ces fonctions de bien-être satisfont à l'efficacité de Pareto et au principe de transfert de Pigou-Dalton.

Nous introduisons également l'outil connexe des courbes de Lorenz généralisées, issu de l'économie du bien-être, pour évaluer l'équité des classements. Les courbes de Lorenz généralisées sont une

représentation graphique qui permet de visualiser la distribution des utilités entre les utilisateurs et les items, et en particulier l'utilité des individus les plus mal lotis, que nous cherchons à améliorer. Grâce à cette représentation, nous pouvons observer la quantité d'utilité prélevée sur les individus les mieux lotis pour augmenter l'utilité des individus les moins bien lotis, lorsque l'on fait varier les paramètres de la fonction de bien-être additive.

Notre cadre conceptuel est également le premier à aborder simultanément l'équité dans les problèmes de recommandation réciproque et non réciproque. La recommandation réciproque est un cadre spécifique qui a été relativement négligé par la littérature sur l'équité et dans lequel les utilisateurs sont également des items. Leur utilité est donc double : ils bénéficient des recommandations qu'ils reçoivent et du fait d'être recommandés à d'autres utilisateurs. Nous montrons que l'approche de la fonction de bien-être pour la recommandation non réciproque peut être étendue au cas réciproque en utilisant notre nouvelle notion d'utilité bilatérale, afin de mieux servir les utilisateurs les plus mal lotis.

Du point de vue algorithmique, il est difficile d'optimiser les fonctions de bien-être globales qui tiennent compte de l'exposition des items, car l'exposition d'un item dépend des classements de tous les utilisateurs. Avant notre travail, les méthodes existantes relevaient ce défi avec des approches heuristiques sans aucune garantie ou contrôle sur les compromis réalisables. Nous proposons un algorithme efficace en termes de calcul pour un classement équitable basé sur la méthode Frank-Wolfe [Frank and Wolfe, 1956]. L'algorithme génère une politique de classement stochastique sous la forme d'une somme pondérée de politiques de classement déterministes. Cela élimine la nécessité d'une étape supplémentaire de décomposition de Birkhoff-von-Neumann [Birkhoff, 1940], qui était nécessaire dans les travaux antérieurs utilisant des classements stochastiques [Singh and Joachims, 2018, Wang and Joachims, 2021]. Notre algorithme est capable d'optimiser toute fonction concave des utilités des classements, ce qui englobe nos fonctions de bien-être additives, mais aussi les critères de classement équitable existants.

Nous simulons une tâche de recommandation musicale basée sur les données de Last.fm pour évaluer la performance de notre algorithme. Nos expériences confirment que les approches d'équité basées sur le mérite sont incapables de réduire l'inégalité entre les items et peuvent exacerber les effets "winner-take-all" lorsque les items populaires capturent une grande partie de l'exposition totale. En revanche, notre approche basée sur la maximisation des fonctions de bien-être additives permet d'obtenir de meilleurs compromis entre l'utilité totale de l'utilisateur et l'inégalité des utilités entre les items (mesurée par l'indice de Gini ou l'écart-type). En outre, en faisant varier les paramètres de la fonction de bien-être, nous sommes en mesure de rapprocher l'inégalité entre les items de zéro. Enfin, en ce qui concerne l'équité bilatérale, notre approche est capable de générer un large éventail de compromis entre l'équité pour les items et l'équité pour les utilisateurs, mesurée

par l'utilité des utilisateurs les plus mal lotis (10 % et 25 %).

Comme notre cadre englobe les problèmes de recommandation réciproque, nous fournissons également une évaluation expérimentale sur une tâche de recommandation sociale basée sur des données Twitter. En maximisant une fonction de bien-être concave additive de l'utilité bilatérale des utilisateurs, nous sommes en mesure de générer un large éventail de compromis entre l'utilité totale et l'utilité des 10% les plus défavorisés.

Chapitre 4 : Équité dans les classements avec les fonctions de bien-être de Gini généralisées. Nous proposons une autre approche de classement équitable basée sur les fonctions de bien-être de Gini généralisées (GGF), qui constituent une classe de fonctions de bien-être plus expressive que les fonctions de bien-être additives précédentes. L'un des inconvénients des fonctions de bien-être de Gini généralisées par rapport aux fonctions de bien-être additives est qu'elles ne peuvent pas être exprimées comme une somme d'utilités d'agents à rendements décroissants. Bien que nous perdions cette interprétation intuitive, nous gagnons en expressivité puisque les GGF sont capables d'exprimer directement des critères d'équité basés sur les quantiles d'utilité (par exemple, "maximiser l'utilité des 10 % les moins bien lotis"). Les GGF couvrent également des mesures d'inégalité plus classiques telles que l'indice de Gini, qui est largement utilisé dans la mesure des inégalités et, plus récemment, dans la littérature sur l'équité dans les systèmes de recommandation. Bien que les GGF n'aient pas d'interprétation intuitive en tant que somme d'utilités à rendements décroissants, leur principal avantage est qu'ils généralisent divers critères d'équité existants pour le classement. En soulignant la généralité des GGF, nous prouvons également que tous les classements Lorenz efficaces peuvent être générés en maximisant les GGF.

Le défi algorithmique de l'optimisation des GGF est qu'ils sont non différentiables et ne peuvent donc pas faire l'objet d'une optimisation par l'algorithme de Frank-Wolfe classique. Nous proposons d'adapter une variante de Frank-Wolfe pour les problèmes non lisses [Lan, 2013] qui utilise l'enveloppe de Moreau-Yosida comme technique de lissage [Moreau, 1962, Yosida et al., 1965], et nous présentons une procédure efficace sur le plan du calcul pour calculer l'approximation lisse des GGF.

Nous menons des expériences sur des tâches de recommandation de films et de musique et comparons notre algorithme qui optimise les GGF aux méthodes de recommandation précédentes, y compris notre propre approche basée sur les fonctions de bien-être concaves additives du chapitre 3. Comme prévu, nous constatons que notre approche basée sur les GGF permet d'obtenir de meilleurs compromis entre l'utilité totale de l'utilisateur et l'inégalité des items mesurée par l'indice de Gini. En effet, les GGF peuvent être instanciés en fonction de l'indice de Gini et notre variante de Frank-Wolfe permet une optimisation directe de cette mesure non différentiable. Pour l'équité bilatérale, nous obtenons également des compromis supérieurs entre l'utilité des 25 % d'utilisateurs

les plus mal lotis et l'indice de Gini des utilités des items, lorsque les GGF côté utilisateur et côté item sont instanciés en fonction de ces critères. Des expériences sur une tâche de recommandation réciproque basée sur des données Twitter montrent des résultats similaires lors de l'optimisation des compromis entre l'utilité des 25 % d'utilisateurs les plus mal lotis et l'utilité totale de l'utilisateur.

Les deux chapitres précédents se concentrent sur le problème de classement pour analyser ses propriétés depuis une perspective d'allocation équitable, en découplant les biais potentiels qui peuvent survenir lors de l'apprentissage. Cependant, il y a des limitations pratiques à la configuration batch précédente, dans laquelle l'apprentissage et la prise de décisions sont découplés, et une seule décision globale est prise pour tous les utilisateurs en même temps. Les systèmes de recommandation modernes interagissent avec les utilisateurs de manière en ligne : ils apprennent les valeurs personnalisées des items à partir des commentaires des utilisateurs, tout en décidant en temps réel ce contenu à afficher au utilisateur actuel. Les bandits contextuels sont un paradigme populaire pour modéliser ce contexte d'apprentissage et de prise de décisions personnalisées dans les systèmes de recommandation [Li et al., 2010].

Chapitre 5 : Classement équitable dans le contexte du bandit contextuel. Nous traitons le problème de classement équitable dans le contexte du bandit contextuel, qui est le paradigme choisi pour les systèmes de recommandation personnalisés en ligne qui apprennent à générer des recommandations à partir des commentaires des utilisateurs. Nous présentons un algorithme générique qui fonctionne pour de nombreux objectifs de classement équitable, y compris les fonctions de bien-être lisses de Chapitre 3 et les fonctions de bien-être non lisses de Chapitre 4. Ceci est le premier algorithme avec des garanties de regret pour le classement équitable dans le contexte du bandit contextuel. De plus, l'algorithme est rapide et a une interprétation intuitive : à chaque étape, l'algorithme donne un bonus adaptatif aux items qui ont reçu une faible exposition dans les recommandations précédentes, et le bonus dépend du gradient de l'objectif de classement équitable.

En fait, nous fournissons un traitement approfondi du problème plus général des bandits contextuels avec des récompenses convexes (CBCR) [Agrawal et al., 2016], qui est un problème de bandit multi-objectif. Dans CBCR, il y a un vecteur de plusieurs récompenses qui dépend d'un contexte aléatoire, et le trade-off entre les récompenses est défini par une fonction concave. Ce setting de CBCR couvre une variété de problèmes au-delà du classement équitable, notamment l'optimisation de plusieurs métriques d'engagement utilisateur (par exemple, clics, temps de streaming) dans les systèmes de recommandation. Des works théoriques antérieurs ont traité des versions simplifiées de CBCR avec des espaces de politiques simples : Agrawal and Devanur [2014], Busa-Fekete et al. [2017] se concentrent sur le setting non contextuel où les politiques sont des distributions sur les

actions, et Agrawal et al. [2016] traitent une restriction de CBCR à un espace de politiques fini. Nous supprimons ces restrictions et présentons des garanties de regret pour le problème général de CBCR en prouvant une réduction de CBCR à des bandits contextuels classiques avec des récompenses scalaires. Notre preuve repose sur une interprétation géométrique de CBCR comme un problème d'optimisation sur l'ensemble convexe de tous les récompenses espérées atteignables, et utilise des techniques d' analyse théorique de Frank-Wolfe en optimisation convexe avec contraintes.

Sur le plan expérimental, nous simulons une tâche de classement en ligne basée sur des données de recommandation de musique. Nous observons que par rapport à des algorithmes de bandit contextuel heuristiques pour le classement équitable, les algorithmes utilisant notre réduction atteignent la plus haute valeur de l'objectif de classement équitable à mesure que le nombre d'étapes augmente. Cela montre l'avantage d'un algorithme de bandit avec garanties comparé à des héuristiques sans garanties théoriques. Lorsque l'objectif de classement équitable est un trade-off entre l'utilité moyenne de l'utilisateur et l'inégalité entre les items, notre algorithme de bandit obtenu par réduction obtient une utilité moyenne supérieure à celle des algorithmes de bandit existants à tous les niveaux d'inégalité entre les items.

Dans les chapitres 3 et 4, nous avons étudié le problème de la planification sociale dans les systèmes de recommandation, où l'on cherche à faire trade-off entre les utilités des utilisateurs et des items pour concevoir des classements équitables à deux faces. Dans le chapitre 6, nous prenons une perspective différente : nous nous concentrons sur la certification de ces systèmes, et nous nous focalisons sur l'équité pour les utilisateurs. Ce travail a été principalement mené au début du programme de doctorat, motivé par la grande résonance des audits pour l'équité côté utilisateurs dans les systèmes de publicité en ligne. Par exemple, Datta et al. [2015] ont constaté que les femmes recevaient moins de publicités en ligne pour des emplois bien rémunérés par rapport à des hommes également qualifiés, tandis que Imana et al. [2021] ont observé des disparités liées au genre dans les taux de livraison de publicités pour différentes entreprises proposant des emplois similaires. Notre contribution à cette recherche est un complément aux audits existants de l'équité côté utilisateurs. Nous partons de l'observation que les audits existants ne contrôlent pas les disparités qui sont en accord avec les préférences des utilisateurs. Pour renforcer les conclusions de ces audits, nous proposons de tester un critère basé sur les préférences, l'*envy-freeness*, qui stipule qu'aucun utilisateur ne devrait préférer les recommandations de ses pairs aux siennes. L'envy-freeness est un critère d'équité qui a été étudié pour la première fois dans la théorie du partage équitable de ressources [Foley, 1967], et il a des racines similaires à celles du cadre conceptuel principal de cette thèse. Cependant, il mène à une évaluation différente dans le contexte des systèmes de recommandation, et le choix du critère d'équité dépend de l'application.

Dans les chapitres précédents 3 et 4, notre travail était motivé par des applications où les producteurs d'items ne sont pas des annonceurs payant pour l'attention des utilisateurs, mais plutôt des créateurs de contenu revendiquant une juste part de l'exposition sur la plateforme. Les exemples typiques sont les plateformes de partage de vidéos en ligne et les services de streaming de musique. Dans ces applications, l'équité côté items est une préoccupation clé, et nous avons conçu des algorithmes qui améliorent les expositions des petits items à travers les classements de tous les utilisateurs. Nous avons également cherché à garantir que les utilisateurs dont les classements sont impactés par la promotion des petits items ne sont pas ceux pour lesquels les boosts sont les plus coûteux. Nous avons abordé l'équité bilatérale en améliorant l'exposition des items les plus mal lotis, tout en priorisant les utilités des utilisateurs les plus mal lotis. Les politiques de classement que nous avons développées dans les chapitres précédents ne sont pas garanties d'être *envy-free* pour les utilisateurs en général, car les politiques de classement optimales pour des objectifs qui comprennent un terme concave de fairness pour les items ne sont pas *envy-free* pour les utilisateurs en général. Par exemple, si les utilisateurs Alice et Bob ont tous les deux envie de recevoir des publicités d'une entreprise populaire, mais que le concepteur promeut les employeurs moins populaires en boostant leurs publicités dans les recommandations de Bob, alors Bob sera envieux des recommandations d'Alice. En pratique, nous recommandons que notre certification de l'envy-freeness est utilisée dans les applications où l'équité côté utilisateur est la principale préoccupation et où l'équité pour les items n'est pas une priorité, comme dans les systèmes de publicité.

Globalement, notre perspective dans le Chapitre 6 est celle d'un auditeur qui se concentre uniquement sur l'évaluation de l'équité pour les utilisateurs, sans tenir compte de la question de savoir si l'inéquité du côté des utilisateur est une conséquence d'autres objectifs. Nous soutenons que la perspective de l'audit est aussi importante que celle du concepteur, compte tenu du rôle significatif joué par les audits d'équité pour les utilisateurs pour sensibiliser à la nécessité d'équité dans les systèmes de recommandation. De plus, les concepteurs peuvent utiliser les évaluations produites par les auditeurs comme des diagnostics supplémentaires pour améliorer leurs systèmes. En fait, les audits existants ont conduit à des accords qui ont poussé les plateformes en ligne à modifier leurs algorithmes de recommandation pour se conformer aux nouvelles exigences d'équité des utilisateurs [Bogen et al., 2023].

Chapitre 6: L'équité pour les utilisateurs comme *absence d'envie*. Dans le Chapitre 6, nous proposons d'évaluer l'équité des systèmes de recommandation pour leurs utilisateurs avec le critère de *envy-freeness* (absence d'envie) issu de la division équitable en théorie du choix social. Transposé au contexte de la recommandation, le critère *envy-freeness* stipule que chaque utilisateur devrait préférer ses recommandations à celles des autres utilisateurs. Par exemple, dans un système

de recommandation d'emploi où deux utilisateurs, Alice et Bob, cherchent des postes de chauffeur de taxi [Ali et al., 2019], si Bob est le seul à recevoir des annonces pour des emplois de chauffeur, alors le système est jugé inéquitable par le critère *envy-freeness*. Comparé à notre précédente approche de fonction de bien-être qui reposait sur la comparaison des utilités entre les utilisateurs, le critère *envy-freeness* évite l'hypothèse difficile des comparaisons interpersonnelles des utilités. En effet, dans le critère *envy-freeness*, différentes recommandations sont comparées du point de vue du même utilisateur (par exemple, Alice).

Nous présentons une analyse formelle des propriétés du *envy-freeness* comme critère d'équité du côté des utilisateurs pour les systèmes de recommandation, et nous montrons sa compatibilité avec les recommandations optimales. Nous montrons également son incompatibilité avec les contraintes d'équité basées sur le mérite du côté de l'objet.[3] Nous présentons également une relaxation probabiliste du *envy-freeness*, afin de supprimer la dépendance quadratique sur le nombre d'utilisateurs et de rendre la certification du *envy-freeness* réalisable.

Par rapport aux chapitres précédents où nous avons pris la perspective du concepteur du système de recommandation en tant que planificateur social, dans ce chapitre, nous adoptons la perspective d'un auditeur interne du système de recommandation. Auditer *envy-freeness* dans les systèmes de recommandation est techniquement difficile, car cela nécessite de sonder les préférences des utilisateurs pour les recommandations des autres, afin de répondre de manière fiable aux questions contrefactuelles : "est-ce que l'utilisateur Alice préférerait les recommandations de Bob?". Notre contribution algorithmique est que nous formulons le problème de la certification de l'envie, ou de son absence, comme une nouvelle forme de problème de bandit d'exploration pure, avec des contraintes d'exploration conservatrices. Les contraintes d'exploration conservatrices empêchent l'audit de détériorer de manière significative les performances de recommandation pour les utilisateurs, lors de l'échange de leurs recommandations avec celles des autres utilisateurs. Nous présentons OCEF, un algorithme d'audit avec des garanties théoriques sur sa complexité d'échantillonnage et le respect des contraintes d'exploration conservatrices. Nous confirmons expérimentalement que l'algorithme OCEF est capable de certifier envy-freeness sur deux tâches de recommandation, tout en maintenant une performance proche de celle du système de recommandation audité.

Nous concluons la thèse dans le Chapitre 7, où nous récapitulons les principales contributions de cette thèse et présentons des contributions supplémentaires. Ce chapitre comprend également un examen critique des limites imposées par nos choix de modélisation, ainsi qu'une discussion sur les enseignements tirés et les questions qui restent ouvertes.

[3] Dans le Chapitre 6, le résultat est prouvé pour les contraintes d'équité basées sur le mérite appliquées au niveau de chaque utilisateur, plutôt qu'entre les utilisateurs.

F.5 Conclusion

Cette thèse apporte à la fois des contributions conceptuelles et algorithmiques.

Dans cette thèse, nous avons développé un cadre conceptuel basé sur les principes de justice distributive de la théorie du choix social pour évaluer l'équité des recommandations classées. Nous abordons la recommandation comme un problème d'allocation équitable où le concepteur fait des compromis entre les utilités des utilisateurs et des items. Dans ce cadre, nous avons proposé une approche fondée pour générer des classements équitables en maximisant les fonctions de bien-être concaves des utilités des utilisateurs et des items. Dans le Chapitre 3, nous avons commencé par des fonctions de bien-être concaves additives, qui codent l'intuition de l'utilité marginale décroissante, puis nous avons traité dans le Chapitre 4 le cas des fonctions de bien-être de Gini généralisées, qui ont une forme plus complexe mais sont plus expressives. La perspective du choix social permet également de mieux comprendre les approches de classement existantes, où nous montrons que les approches méritoires populaires peuvent conduire à une inéquité distributive indésirable (Chapitre 3).

En plus du cadre conceptuel de cette thèse, nous avons apporté plusieurs contributions algorithmiques, basées sur les méthodes de Frank-Wolfe. Nous avons relevé le défi d'optimiser les fonctions concaves des politiques de classement stochastiques, qui peuvent être utilisées pour exprimer de nombreux objectifs pour la recommandation équitable et multi-objective. Nous avons d'abord montré comment utiliser efficacement les méthodes de Frank-Wolfe dans le cadre de la batch, pour le classement dans le modèle basé sur la position au Chapitre 3. Ensuite, nous avons montré comment étendre cette approche au cas des GGF non différentiables au Chapitre 4. Au Chapitre 5, nous avons abordé le problème du classement équitable dans le cadre du bandit contextuel, et présenté le premier algorithme de bandit avec des garanties de regret pour le problème. Tous les algorithmes développés dans cette thèse sont soutenus par des garanties théoriques sur leur convergence et leur complexité. Nous avons également évalué nos algorithmes par rapport à des benchmarks pertinents sur des environnements simulés basés sur des jeux de données publics tels que MovieLens, Last.fm et Twitter, qui comprennent jusqu'à $15k$ utilisateurs et items.

En plus de proposer de nouvelles méthodes pour *concevoir* des systèmes de recommandation qui sont équitables envers les utilisateurs et les items, nous avons également abordé un problème différent d'*audit*, qui est axé sur l'*équité côté utilisateur* au Chapitre 6. Motivés par des audits importants pour l'égalité dans la distribution des annonces d'emploi, nous proposons un audit pour l'absence d'envie, qui fournit des conclusions plus affinées mais qui est plus techniquement difficile. Nous relevons ce défi technique en développant un algorithme de bandit d'exploration pure efficient en termes d'échantillons pour la tâche, qui ne dégrade pas de manière significative les performances de recommandation pour les utilisateurs échantillonnés pour l'audit.

Notre recherche laisse plusieurs questions ouvertes. Celles-ci comprennent un traitement plus détaillé de l'équité bilatérale au niveau du groupe, une modélisation plus générale des utilités des utilisateurs et des items, et l'intégration de dynamiques réelles qui affectent les préférences et les comportements des utilisateurs et des items. De plus, bien que notre travail se concentre sur la perspective de la division équitable, le domaine du choix social offre des perspectives précieuses pour la communauté des recommandations qui méritent une exploration plus approfondie. Répondre à ces questions difficiles en conjonction avec nos contributions peut ouvrir des voies de recherche passionnantes. Malgré les questions restées ouvertes, notre recherche a fait des progrès significatifs dans l'amélioration de l'état de l'art actuel en matière d'équité pour les systèmes de recommandation. Nous avons mieux compris les limites des contraintes d'égalité et de mérite sur l'exposition, ainsi que la manière de concevoir des objectifs de classement principiels. Nos résultats ont conduit au développement d'algorithmes efficaces qui peuvent être mis en œuvre pratiquement, servant de tremplin pour le développement d'approches principielles de l'équité dans les systèmes de recommandation dans des contextes plus complexes. Nous espérons que notre travail inspirera de nouveaux progrès dans ce domaine.

RÉSUMÉ

Les algorithmes d'apprentissage automatique sont largement utilisés dans les systèmes de recommandation pour personnaliser les suggestions de contenu en fonction des préférences des utilisateurs. Cependant, ces algorithmes peuvent involontairement favoriser certains producteurs tout en négligeant d'autres, soulevant des préoccupations éthiques et sociétales concernant l'équité et la représentation. Ce travail vise à créer des algorithmes de recommandation plus équitables qui équilibrent les préférences des utilisateurs avec le bien-être des producteurs de contenu. En abordant les défis liés à la définition de l'équité et à l'efficacité computationnelle, nous proposons un cadre ancré dans la théorie du choix social. Ce cadre guide le développement de nouvelles méthodes de recommandation qui distribuent équitablement l'exposition aux producteurs de contenu sans compromettre la qualité des recommandations. Nos méthodes sont étayées par des garanties théoriques concernant l'équité, la convergence et l'efficacité computationnelle, ainsi que par des évaluations sur des jeux de données.

MOTS CLÉS

Equité, systèmes de recommandation, classement, apprentissage automatique, choix social, économie normative, optimisation convexe, algorithmes de bandit.

ABSTRACT

Machine learning algorithms are widely used in recommender systems for personalizing content recommendations based on user preferences. However, these algorithms can inadvertently promote certain producers while overlooking others, raising ethical and societal concerns about fairness and representation. This work aims to create fairer recommendation algorithms that balance user preferences with content producer welfare. Addressing the challenges of defining fairness for ranked recommendations and preserving computational efficiency, we propose a framework grounded in social choice theory. This framework guides the development of new recommendation methods that fairly distribute exposure to content producers without compromising recommendation quality. Our methods are substantiated by theoretical guarantees of fairness, convergence and computational efficiency, as well as real-world dataset evaluations.

KEYWORDS

Fairness, recommender systems, ranking, machine learning, social choice, welfare economics, convex optimization, bandit algorithms.