Fair and Welfare-Efficient Constrained Multi-matchings under Uncertainty

Elita Lobo,[*] Justin Payan,[*] Cyrus Cousins, and Yair Zick
University of Massachusetts Amherst
{elobo, jpayan, cbcousins, yzick}@umass.edu

Abstract

We study fair allocation of constrained resources, where a market designer optimizes overall welfare while maintaining group fairness. In many large-scale settings, utilities are not known in advance, but are instead observed after realizing the allocation. We therefore *estimate* agent utilities using machine learning. Optimizing over estimates requires trading-off between mean utilities and their predictive variances. We discuss these trade-offs under two paradigms for preference modeling – in the *stochastic optimization* regime, the market designer has access to a probability distribution over utilities, and in the *robust optimization* regime they have access to an *uncertainty set* containing the true utilities with high probability. We discuss utilitarian and egalitarian welfare objectives, and we explore how to optimize for them under stochastic and robust paradigms. We demonstrate the efficacy of our approaches on three publicly available conference reviewer assignment datasets. The approaches presented enable scalable constrained resource allocation under uncertainty for many combinations of objectives and preference models.

1 Introduction

Constrained resource allocation without money underpins many important systems, including reviewer assignment for peer review (our primary example throughout the paper) [4, 16, 31, 45, 54], assigning resources to homeless populations [5, 34, 49], distributing emergency response resources [51, 56, 57], and more [1, 44, 53]. In these settings we assign *resources* to *agents*. Agents and resources are *constrained*; each agent has bounds on the minimum or maximum number of items they receive from different categories, and each item has required minimums and limited total capacity. Each agent has a valuation for every item, and we optimize a welfare function of the agent-item valuations. In the case of reviewer assignment, the reviewer-paper valuations measure the alignment between reviewers and papers, papers must receive a certain number of reviews from unique reviewers, reviewers have upper limits on the number of papers they can review, and conflicts of interest prevent some reviewers from being assigned to certain papers.

A crucial factor in all of the above settings is the presence of *uncertainty*. Uncertainty often stems from the fact that agents' valuations for resources depend on future outcomes. In reviewer assignment, a reviewer-paper pair's match quality is observed only after the reviewer submits his or her review. Uncertainty may also stem from our limited ability to collect data; for example, in deciding where to target lead pipe mitigation projects based on number of school-aged children per neighborhood, we may have access to imperfect school enrollment records, allowing only an approximate model of the impacts of mitigation on children in each neighborhood [53]. We adopt two possible stances towards uncertainty, depending on the information available. When we have access to *a*

[*] Authors contributed equally.

38th Conference on Neural Information Processing Systems (NeurIPS 2024).

probability distribution over preferences, we optimize the conditional expectation of the distribution at percentiles of interest [33, 50]. When we have access to *a set* of possible preferences, we adopt the *robust* approach, which is related to the minimax regret objective used in solving robust assignment problems [3, 10, 11, 32]. Uncertainty-aware optimization approaches can often result in significantly different allocations from the default of optimizing for welfare over a central estimate (see Example 2.1 for an intuitive explanation for this phenomenon).

Typically, we maximize the sum of agent utilities. However, in many of these settings, we are also concerned with *fairness* to individuals or groups of agents. Groups of agents may represent subject areas of papers in reviewer assignment, demographic groups in poverty alleviation campaigns, or regional groupings of computational resources in bandwidth allocation. Fairness to these groups may be legally required in some cases; in others it is an ethical choice by the decision maker. Although groups are often first-class objects worthy of receiving fair treatment, group fairness is often the smallest granularity of fairness achievable under uncertainty – in a large dataset uncertainty will always cause some individuals to have vanishing welfare, but group welfare can still be upheld. Although there is much literature on combinatorial optimization under uncertainty [3, 10, 11, 32, 33], to our knowledge it has not addressed the intersection of fairness and uncertainty in the constrained multi-matching problem.

1.1 Our Contributions

We study the broad problem of fair and efficient constrained multi-matchings under uncertainty about agents' valuations. We optimize for welfare while simultaneously accounting for the uncertainty inherent in real-world resource allocation problems. Specifically, we develop methods to efficiently optimize the utilitarian and egalitarian welfare objectives using the robust approach [7, 8, 26] and the CVaR approach [50]. Our results are summarized in Table 1.

For robust optimization, we construct an uncertainty set containing the true preferences with high probability (Section 3). This model is appropriate when building a predictor with statistical error bounds, but without making any assumptions on the full probability distribution over valuations. For utilitarian and egalitarian welfare functions, we robustly maximize welfare over such uncertainty sets. When the uncertainty sets are linear we can efficiently compute the exact optimal allocations for both utilitarian and egalitarian welfare in polynomial time (Corollaries 3.2 and 3.6). Under a single ellipsoidal uncertainty set, we can apply an *iterated quadratic programming* approach (Corollaries 3.3 and 3.7), while a projected subgradient ascent approach is needed when uncertainty sets consist of multiple ellipsoids (Propositions 3.1 and 3.5). Under general monotonic, concave welfare functions and arbitrary convex uncertainty sets, we apply the relatively expensive adversarial projected subgradient ascent algorithm of Cousins *et al.* [16].

When the market designer can construct a full probability distribution over preferences or sample from such a distribution, we consider *stochastic optimization* using the concept of *Conditional Value at Risk*, or CVaR [50]. This approach, laid out in Section 4, selects an allocation that maximizes the conditional expectation of welfare over the left tail of the welfare distribution. We often approximate CVaR objectives using sampling, then solve the resulting linear program or LP (as in Propositions 4.1 and H.3). However, in the case of utilitarian welfare and Gaussian-distributed valuations we present a simple reformulation of the CVaR objective (Proposition 4.3). Optimizing CVaR for general monotonic, concave welfare functions can require solving arbitrary concave optimization problems, even after sampling.

We also compare these optimization approaches empirically in Section 5 on reviewer assignment data from AAMAS 2015, 2016, and 2021.

1.2 Related work

We discuss the history of prior work on robust and CVaR optimization in Appendix A.

Some existing work applies stochastic or robust optimization to fair division problems. A line of work studies the minimax regret objective in combinatorial optimization problems, such as constrained resource allocation [3, 10, 11, 32]. This work does not explicitly consider multi-matching problems like those considered here, nor does it address the robust egalitarian welfare problem. Pujol *et al.* [48] study fair division problems with parameters noised for differential privacy, showing

Table 1: Summary of optimization algorithms for efficiently computing utilitarian and egalitarian welfare under different robustness concepts. Green highlights indicate problems which require solving a single linear program (low difficulty). Yellow highlights indicate solving a small number of linear or quadratic programs (medium difficulty). Red highlights indicate problems which require solving numerous quadratic programs or arbitrary concave programs.

	Robustness Concept				
	Robust			**CVaR**	
	Linear	**One Ellipsoid**	**ℓ Ellipsoids**	**Any (approx.)**	**Gaussian**
Utilitarian	LP Reduction (Coro. 3.2)	Iterated QP (Coro. 3.3)	Projected SGA (Prop. 3.1)	Sampling + LP (Prop. 4.1)	Projected GA (Prop. 4.3)
Egalitarian	LP Reduction (Coro. 3.6)	Iterated QP (Coro. 3.7)	Projected SGA (Prop. 3.5)	Sampling + LP (Prop. H.3)	
Monotone, Concave	Adversarial Projected SGA (Sec. 3.3)			Sampling + Concave Program (Sec. 4)	

that the noise can cause unfair allocations; they propose a Monte Carlo approach to mitigate the unfairness with high probability. Peters *et al.* [46] study envy-free rent division under probabilistic uncertainty. A central mechanism divides rooms and sets room prices for the items to minimize envy. We study a setting without money, both utilitarian and egalitarian objectives, and robust optimization in addition to stochastic optimization.

Cousins *et al.* [16] study robust optimization under the utilitarian objective. They propose an adversarial projected subgradient ascent method which requires solving two quadratic programs (one for the adversary and one for the projection) at each iteration for a large number of iterations. Our empirical analysis in Section 5 demonstrates the inefficiency of this method. Fair machine learning algorithms [17, 22, 23, 43, 59] often employ similar adversarial optimization techniques over an uncertainty set in a machine learning context. Other fair allocation research has studied the case where agent demand or item availability are uncertain but preferences are known [2, 14, 21, 27]. In our case demand and availability are known but preferences are not. Devic *et al.* [20] consider fair two-sided matching where the fairness constraint is defined with respect to unknown parameters; we assume knowledge of the parameters that define the fairness constraint (i.e., group identities).

2 Fair Resource Allocation under Uncertainty

We first introduce the problem of resource allocation without uncertainty, then lay out the two approaches we take to deal with the introduction of uncertainty. Our results are summarized in Table 1.

2.1 Fair Resource Allocation

We have a set of n agents $N = \{a_1, \dots a_n\}$, and m item types $I = \{i_1, \dots i_m\}$. Agents are partitioned into g groups $\mathcal{G} = \{G_1, \dots G_g\}$, with each $G \subseteq N$ and each agent i belonging to exactly one group.

For any $n \times m$ matrix $\boldsymbol{X}$ we use the same lower-case bold letter, i.e., $\boldsymbol{x}$ to denote the vector representing the vectorized form of the matrix $\boldsymbol{X}$, in row-major order. For any group of agents G, we use $\mathbf{x}_G \in \mathbb{R}^{|G|m}$ to denote the vector restricted to the agents in G. Given vectors $\mathbf{x}, \mathbf{y} \subseteq \mathbb{R}^{nm}$ and real number $c \in \mathbb{R}$, let $\mathbf{x} \succeq c$ denote that $\mathrm{x}_j \geq c$ for all j, and let $\mathbf{x} \succeq \mathbf{y}$ denote that $\mathbf{x} - \mathbf{y} \succeq 0$. The $\preceq$ operator is defined analogously.

We assume a valuation matrix $\boldsymbol{V}^* \in \mathbb{R}_{0+}^{n \times m}$, where $\boldsymbol{V}^*_{a,i}$ encodes the true value of assigning item type i to agent a. The values of $\boldsymbol{V}^*$ are typically unknown; we discuss our approaches to handle this problem in Section 2.2. We use tildes to denote random variables, for example, $\tilde{x} \sim \mathcal{D}_{\tilde{x}}$ denotes that $\tilde{x}$ is drawn from distribution $\mathcal{D}_{\tilde{x}}$.

Given some set of feasible assignments $\mathcal{A} \subseteq \mathbb{N}^{n \times m}$, we aim to find assignments $\boldsymbol{A} \in \mathcal{A}$ where $\boldsymbol{A}_{a,i}$ indicates the number of items of type i allocated to agent a. For each agent $a \in N$, we have

upper and lower bounds on assignments of the form $\underline{\kappa}_a \leq \sum_{i \in I} \boldsymbol{A}_{a,i} \leq \bar{\kappa}_a$. For each item i, we have lower and upper bounds on the total assignment of that item; $\underline{\psi}_i \leq \sum_{a \in N} \boldsymbol{A}_{a,i} \leq \bar{\psi}_i$. Finally, we have pairwise limits $\boldsymbol{C}_{a,i}$ for each agent a and item type i, requiring that $\boldsymbol{A}_{a,i} \leq \boldsymbol{C}_{a,i}$. It is always the case that the constraints define a finite set such that $|\mathcal{A}| \in \mathbb{N}$. In the example of reviewer assignment, these constraints reflect the review requirements per paper, load bounds for reviewers, conflicts of interest, and the constraint that no reviewer is assigned twice to any given paper.

Let $\mathbf{u} : \mathcal{A} \times \mathbb{R}_{0+}^{n \times m} \to \mathbb{R}^g$ be an affine function mapping from allocations to utilities for each group. $\mathrm{u}_G(\mathbf{a}, \mathbf{v})$ denotes the utility of the group G under allocation $\mathbf{a} \in \mathcal{A}$ (recall $\mathbf{a}$ is the vectorized version of the assignment $\boldsymbol{A}$). We write $\mathbf{u}$ instead of $\mathbf{u}(\mathbf{a}, \mathbf{v})$ when $\mathbf{a}$ and $\mathbf{v}$ are clear from context.

We assume that $\mathbf{u}$ is *additive* and normalized by group size, so $\mathrm{u}_G = \frac{\mathbf{a}_G^\intercal \mathbf{v}_G}{|G|}$. We define a *welfare function* $\mathrm{W} : \mathbb{R}^g \to \mathbb{R}$, where $\mathrm{W}(\mathbf{u}(\mathbf{a}, \mathbf{v}))$ denotes the overall welfare of allocation $\mathbf{a}$. The *weighted utilitarian social welfare* function is defined as $\mathbf{w} \cdot \mathbf{u}$, where $\mathbf{w} \in \mathbb{R}_{0+}^g$ denotes the weights on groups in $\mathcal{G}$. When $\mathrm{w}_G = |G|$ for all G, we call this function simply "utilitarian welfare" or "USW." The *group egalitarian social welfare* function (also "group egalitarian welfare" or "GESW") is defined as $\min_{G \in \mathcal{G}} \mathrm{u}_G$. We do not consider individual egalitarian welfare in this work; under robust and stochastic optimization the egalitarian welfare is zero when the number of items is proportional to the number of agents and uncertainty is non-trivial.

2.2 Optimizing Allocations under Uncertainty

We consider two main approaches to dealing with uncertainty: the *robust optimization* approach and the *Conditional Value at Risk* approach.

In the robust approach (Section 3), we obtain an *uncertainty set* $\mathcal{V}$ that contains the true agent valuations $\mathbf{v}^*$ with probability at least $1 - \alpha$ for some confidence parameter $\alpha \in [0, 1)$. We then optimize the welfare corresponding to the *worst* valuation matrix in the uncertainty set, i.e., $\max_{\mathbf{a} \in \mathcal{A}} \min_{\mathbf{v} \in \mathcal{V}} \mathrm{W}(\mathbf{u}(\mathbf{a}, \mathbf{v}))$. This approach applies when we do not have access to a full distribution over valuations but have some other way of describing $\mathcal{V}$ [16].

When we have access to a full distribution over the random variable $\tilde{\mathbf{v}} \in \mathbb{R}^{nm}$, we apply a stochastic approach instead. We compute the welfare distribution and optimize the conditional expectation over an α-percentile of the welfare or *Conditional Value at Risk at* α (CVaR_α). Suppose we have a random variable $\tilde{x} \sim \mathcal{D}_{\tilde{x}}$. For any risk level $\alpha \in (0, 0.5)$, $\mathrm{CVaR}_\alpha[\tilde{x}]$ is defined as $\mathbb{E}_{\tilde{x} \sim \mathcal{D}_{\tilde{x}}}[\tilde{x} \mid \tilde{x} \leq \nu_\alpha]$ where ν_α denotes the α-percentile of $\tilde{x}$. This approach is only appropriate when $\mathcal{D}_{\tilde{x}}$ is fully known, or when we can sample from it. We investigate this regime in Section 4, where we will compute and optimize $\mathrm{CVaR}_\alpha[\mathrm{W}(\mathbf{u}(\mathbf{a}, \tilde{\mathbf{v}}))]$ for a random variable $\tilde{\mathbf{v}} \sim \mathcal{D}_{\tilde{\mathbf{v}}}$.

Example 2.1 (The Importance of Considering Uncertainty)**.** Consider a simple two-agent, two-item instance, where each agent needs to get exactly one item, and either likes (utility 1) or dislikes it (utility 0). Agent preferences are Bernoulli random variables, where $\Pr[\tilde{\mathrm{v}}_{1,1} = 1] = 0.8, \Pr[\tilde{\mathrm{v}}_{1,2} = 1] = 0.9, \Pr[\tilde{\mathrm{v}}_{2,1} = 1] = 0.5$, and $\Pr[\tilde{\mathrm{v}}_{2,2} = 1] = 0.8$. If we maximize the expected USW, we would assign i_1 to a_1 and i_2 to a_2, for a total expected USW of 1.6. However, consider instead the $\mathrm{CVaR}_{0.3}$ of USW. When we make the expectation-maximizing assignment, then $\Pr[\mathrm{W}(\mathbf{u}) = 0] = 0.04$ and $\Pr[\mathrm{W}(\mathbf{u}) = 1] = 0.32$. However, if we assign i_2 to agent a_1 and item i_1 to agent a_2, we have that $\Pr[\mathrm{W}(\mathbf{u}) = 0] = 0.05$ and $\Pr[\mathrm{W}(\mathbf{u}) = 1] = 0.5$. This means that the conditional expectation of welfare at the 30th percentile is higher if we assign i_2 to a_1 and i_1 to a_2 (it is .32 in the first case and .5 in the second case). If we want to retain welfare in the face of uncertainty, we might well choose to maximize this quantity rather than the expectation of the welfare.

3 Robust Welfare Optimization

We construct the optimization problems for utilitarian and egalitarian welfare objectives with the robust approach. Many of these optimization problems are concave-convex max-min problems that can be directly solved using an adversarial projected subgradient ascent technique [16]: in each iteration of the algorithm, the inner minimization problem is solved to optimality, followed by a subgradient step on the allocation $\mathbf{a}$, followed by a projection onto the constraint space $\mathcal{A}$. However, this method does not exploit the structure of these problems and is often computationally expensive or intractable, as demonstrated empirically in Section 5. Despite the inherent complexities of

these problems, we show that under specific assumptions, these problems can be reduced to more manageable forms that are easier to optimize. We then discuss a range of algorithms for efficiently optimizing the simplified problems.

Scope: The robust approach detailed in Section 2 assumes the availability of an uncertainty set of the valuation matrix. For the sake of computational tractability, we focus on the class of uncertainty sets defined by linear and ellipsoidal constraints

$$\mathcal{V} = \left\{ \mathbf{v} \in \mathbb{R}^{nm} \mid \forall i \in [1, \ell],\, (\mathbf{v} - \bar{\mathbf{v}}_i)^\intercal \mathbf{S}_i^{-1} (\mathbf{v} - \bar{\mathbf{v}}_i) \leq r_i^2, \boldsymbol{Q}\mathbf{v} \succeq \mathbf{e}, \mathbf{v} \succeq 0 \right\} \ ,$$

where the ith ellipsoidal uncertainty set has center $\bar{\mathbf{v}}_i \in \mathbb{R}_{0+}^{nm}$, covariance matrix $\mathbf{S}_i \in \mathbb{R}^{nm \times nm}$, with radius $r_i \in \mathbb{R}$, $\boldsymbol{Q} \in \mathbb{R}^{k \times nm}$, and $\mathbf{e} \in \mathbb{R}^k$.

We will further assume that the covariance matrices corresponding to the ellipsoidal uncertainty sets are positive semi-definite. This limitation on the structure of uncertainty sets is not too restrictive; it is possible to construct such uncertainty sets for linear regression and logistic regression models using statistical bounds, as shown in Appendix D.

In all of our methods where obtaining an integer allocation is either not feasible or computationally tractable, we relax the set of feasible integer assignments $\mathcal{A} \subseteq \mathbb{N}^{n \times m}$ to a set of feasible continuous allocations $\mathcal{A} \subseteq \mathbb{R}_{0+}^{n \times m}$. One can obtain integer allocations satisfying all constraints by applying a randomized rounding technique that generalizes the Birkhoff-von Neumann decomposition [13]. The fractional solutions can thus be interpreted as randomized allocations.

3.1 Robust Allocation for Utilitarian Welfare

We consider the problem of finding an allocation that optimizes the weighted utilitarian welfare under the worst valuation matrix in the uncertainty set. We formulate the problem as

$$\max_{\mathbf{a} \in \mathcal{A}} \min_{\mathbf{v} \in \mathcal{V}} \mathbf{w} \cdot \mathbf{u}(\mathbf{a}, \mathbf{v}) \ . \tag{1}$$

The objective and constraints of the inner-minimization problem described in (1) are convex. The problem is strictly feasible, which satisfies Slater's condition [12] for strong duality. Therefore, by taking the dual of the inner-minimization problem, we can simplify the problem in (1) into a single equivalent maximization problem. We provide the dual formation in Proposition 3.1.

To simplify notation in the results that follow, we assume, without loss of generality, that each group G has weight $\mathrm{w}_G = |G|$. In practice, if this assumption does not hold, the weights can be incorporated into the valuations $\mathbf{v}$ with a corresponding adjustment to the parameters of the valuation uncertainty set $\mathcal{V}$.

In the dual, let $\boldsymbol{\beta} \in \mathbb{R}_{0+}^k$ be the dual variable corresponding to the linear constraints $\boldsymbol{Q}\mathbf{v} \succeq \mathbf{e}$, $\boldsymbol{\lambda} \in \mathbb{R}_{0+}^\ell$ be the dual variable associated with the ellipsoidal constraints, and $\boldsymbol{\xi} \in \mathbb{R}^{nm}$ be the variable that combines the primal variable $\mathbf{a}$ with the dual variable of the non-negativity constraint on $\mathbf{v}$ for variable elimination. We define a set of feasible $\boldsymbol{\xi}$ as

$$\Lambda \doteq \mathcal{A} - \mathbb{R}_{0+}^{nm} = \left\{ \boldsymbol{\xi} \in \mathbb{R}^{nm} \,\middle|\, \forall a \in N : \sum_{i \in I} \xi_{am+i} \leq \bar{\kappa}_a, \forall i \in I : \sum_{a \in N} \xi_{am+i} \leq \bar{\psi}_i, \boldsymbol{\xi} \preceq \boldsymbol{c} \right\} \ , \tag{2}$$

which is Pareto-dominated by $\mathcal{A}$.

Proposition 3.1 (Robust Utilitarian Welfare Dual)**.** *The problem in* (1) *is equivalent to solving*

$$\max_{\substack{\boldsymbol{\xi} \in \Lambda, \boldsymbol{\lambda} \in \mathbb{R}_{0+}^l \\ \boldsymbol{\beta} \in \mathbb{R}_{0+}^k}} \mathbf{p}^\intercal \mathbf{T} \mathbf{q} + \boldsymbol{\beta}^\intercal \mathbf{e} - \frac{1}{4} \mathbf{p}^\intercal \mathbf{T} \mathbf{p} + \sum_{i=1}^{\ell} \left(\lambda_i \bar{\mathbf{v}}_i^\intercal \mathbf{S}_i \bar{\mathbf{v}}_i - \lambda_i r_i^2 \right) - \mathbf{q}^\intercal \mathbf{T} \mathbf{q} \ , \tag{3}$$

where $\mathbf{p} = \boldsymbol{\xi} - \boldsymbol{Q}^\intercal \boldsymbol{\beta}$, $\mathbf{q} = \sum_{i=1}^{\ell} \lambda_i \mathbf{S}_i^{-1} \bar{\mathbf{v}}_i$, *and* $\mathbf{T} = \left(\sum_{i=1}^{\ell} \boldsymbol{\lambda}_i \mathbf{S}_i^{-1} \right)^{-1}$. *Let* $\boldsymbol{\xi}^*$ *be the optimal* $\boldsymbol{\xi}$ *in* (3)*. Then the optimal allocation* $\mathbf{a}^*$ *can be derived from* $\boldsymbol{\xi}^*$ *by finding* $\mathbf{a} \in \mathcal{A}$ *such that* $\mathbf{a} \preceq \boldsymbol{\xi}^*$.

Proposition 3.1 shows that the optimal allocation for the problem in Equation (1) can be computed by first solving the concave program in Equation (3) to obtain $\boldsymbol{\xi}^*$ and then deriving the optimal allocation $\mathbf{a}^*$ from $\boldsymbol{\xi}^*$ by solving a system of equations. Notably, the problem in Equation (3)

is a single maximization problem with fewer variables and constraints as compared to the max-min problem in (1), making it simpler to solve. We can either solve it using off-the-shelf convex optimization tools, or by applying a projected subgradient ascent approach (without the previously required adversary).

When the valuation uncertainty set is polyhedral, the problem in (3) simplifies further into a linear program (LP) which can be solved efficiently using standard LP solvers like Gurobi [28].

Corollary 3.2 (Utilitarian Welfare with Polyhedral Uncertainty)**.** *In the case where the uncertainty set $\mathcal{V}$ is defined purely by linear constraints, i.e., $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^{nm} \mid \boldsymbol{Q}\mathbf{v} \succeq \mathbf{e}, \mathbf{v} \succeq 0\}$, the optimal allocation $\mathbf{a}^*$ for the problem in* (1) *can be computed by solving the linear program*

$$\max_{\mathbf{a}\in\mathcal{A},\boldsymbol{\beta}\in\mathbb{R}^k_{0+}} \boldsymbol{\beta}^\intercal\mathbf{e} \quad \textit{s.t. } \boldsymbol{Q}^\intercal\boldsymbol{\beta} \preceq \mathbf{a} \ .$$

When the valuation uncertainty set has a single ellipsoidal constraint with a non-negativity constraint, we compute the solution using iterated quadratic programming (Iterated QP).

Corollary 3.3 (Utilitarian Welfare with Ellipsoidal Uncertainty)**.** *Suppose that the set $\mathcal{V}$ in* (1) *is defined by a single truncated ellipsoidal constraint i.e., $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^{nm} \mid (\mathbf{v}-\bar{\mathbf{v}})^\intercal\mathbf{S}^{-1}(\mathbf{v}-\bar{\mathbf{v}}) \leq r^2, \mathbf{v} \succeq 0\}$. The problem in* (1) *is equivalent to solving*

$$\max_{\lambda\in\mathbb{R}_{0+},\boldsymbol{\xi}\in\Lambda} \boldsymbol{\xi}^\intercal\bar{\mathbf{v}} - \frac{\boldsymbol{\xi}^\intercal\mathbf{S}\boldsymbol{\xi}}{4\lambda} - \lambda r^2 \ . \tag{4}$$

The exact optimal solution $(\lambda^*, \boldsymbol{\xi}^*)$ to Equation (4) *can be computed by alternately performing two steps until convergence: first, fixing $\boldsymbol{\xi}$ and optimizing λ, i.e., $\lambda = {\boldsymbol{\xi}^\intercal\mathbf{S}\boldsymbol{\xi}}/{2r}$, and second, fixing λ and solving a concave quadratic program to optimize $\boldsymbol{\xi}$. The optimal allocation $\mathbf{a}^*$ can be computed from $\boldsymbol{\xi}^*$ as in Proposition 3.1.*

3.2 Robust Allocation for Group Egalitarian Welfare

We now consider the problem of maximizing egalitarian welfare under the robust approach. We can represent this problem as

$$\max_{\mathbf{a}\in\mathcal{A}} \min_{\mathbf{v}\in\mathcal{V}} \min_{G\in\mathcal{G}} \mathrm{u}_G(\mathbf{a},\mathbf{v}) \ . \tag{5}$$

This problem presents inherent challenges due to the non-smoothness of the inner-minimization problem and the joint constraint on the uncertainties of the valuation matrices of different groups. These factors make it difficult to compute the dual and reduce the problem or efficiently solve the problem using the quadratic program technique described in Corollary 3.3, though the generic adversarial subgradient ascent approach of Cousins *et al.* [16] can still be applied. For the remainder of this section, we assume that the uncertainty sets for each group $G \in \mathcal{G}$ are independent of each other. To simplify notation, we assume, without loss of generality, that the valuations $\mathbf{v}$ and the parameters of the valuation uncertainty set $\mathcal{V}$ are scaled to incorporate the factor $\frac{1}{|G|}$ in the representation of the utility of each group G in the corresponding group valuation $\mathbf{v}_G$.

Assumption 3.4 (Independence of Groups)**.** *The uncertainty set $\mathcal{V}$ is a Cartesian product of individual groups' uncertainty sets, $\mathcal{V} \doteq \bigotimes_{G\in\mathcal{G}} \mathcal{V}_G$.*

This assumption is not unreasonable in practical scenarios. For example, conferences often group papers into disjoint tracks or require paper authors to select a single primary subject area. Although papers may have multiple secondary subject areas, the top-level grouping remains independent. Assumption 3.4 allows us to reorder the two minimization problems without compromising generality:

$$\max_{\mathbf{a}\in\mathcal{A}} \min_{G\in\mathcal{G}} \min_{\mathbf{v}_G\in\mathcal{V}_G} \mathbf{a}_G^\intercal\mathbf{v}_G \ . \tag{6}$$

We take the dual of the inner minimization problem, then reorder the minimization over groups and the maximization over the dual variables to obtain a single, concave max-min problem. This can be solved with projected subgradient ascent in the general case, or with more efficient approaches in special cases. Proposition 3.5 expresses the general form of the result.

Proposition 3.5 (Robust Group Egalitarian Dual)**.** *The problem in* (5) *is equivalent to solving*

$$\max_{\substack{\boldsymbol{\xi}\in\Lambda \\ \boldsymbol{\lambda}\in\mathbb{R}^{g\times l}_{0+} \\ \boldsymbol{\beta}\in\mathbb{R}^{g\times k}_{0+}}} \min_{G\in\mathcal{G}} \boldsymbol{\beta}_G^\intercal\mathbf{e}_G + \mathbf{p}_G^\intercal\mathbf{T}\mathbf{q}_G - \frac{1}{4}\mathbf{p}_G^\intercal\mathbf{T}\mathbf{p}_G + \sum_{i=1}^{\ell}\left(\lambda_{G,i}\bar{\mathbf{v}}_{G,i}^\intercal\mathbf{T}\bar{\mathbf{v}}_{G,i} - \lambda_{G,i}r_{G,i}^2\right) - \mathbf{q}_G^\intercal\mathbf{T}\mathbf{q}_G \ , \tag{7}$$

where for each group $G \in \mathcal{G}$, $\mathbf{p}_G = \boldsymbol{\xi}_G - \boldsymbol{Q}_G^\intercal \boldsymbol{\beta}_G$, $\mathbf{q}_G = \sum_{i=1}^{\ell} \lambda_{G,i} \mathbf{S}_{G,i}^{-1} \bar{\mathbf{v}}_{G,i}$, $\mathbf{T} = \left(\sum_{i=1}^{\ell}, \lambda_{G,i} \mathbf{S}_{G,i}^{-1}\right)^{-1}$, *and* Λ *is defined as in Equation* (2). *The optimal allocation* $\mathbf{a}^*$ *can be computed from* $\boldsymbol{\xi}^*$ *as in Proposition 3.1.*

The dual variables $\boldsymbol{\lambda}_G, \boldsymbol{\beta}_G, \boldsymbol{\zeta}_G$ and $\boldsymbol{\xi}_G$ for each group G are interpreted as in Proposition 3.1. The optimization problem in (7) is concave with respect to the dual variables $\boldsymbol{\lambda}, \boldsymbol{\beta}$ and $\boldsymbol{\xi}$. We can solve it using an off-the-shelf convex programming library or by applying projected subgradient ascent.

Under polyhedral uncertainty sets, Equation (7) simplifies to a linear program. This is akin to what we observe in the robust utilitarian case (Corollary 3.2).

Corollary 3.6 (Group Egalitarian Welfare with Polyhedral Uncertainty)**.** *In the case where the uncertainty set* $\mathcal{V}$ *is defined only by linear constraints, i.e.,* $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^{nm} \mid \forall G \in \mathcal{G} : \forall G \in \mathcal{G} : \boldsymbol{Q}_G \mathbf{v}_G \succeq \mathbf{e}_G, \mathbf{v} \succeq 0\}$*, the max-min-min problem in* (5) *transforms into a linear program.*

When the valuation uncertainty set is defined by a single ellipsoidal constraint per group, we can employ the iterated quadratic programming (Iterated QP) approach used in Corollary 3.3, alternately fixing $\boldsymbol{\lambda}$ and optimizing the rest of the dual variables $(\boldsymbol{\beta}, \boldsymbol{\xi})$ until convergence.

Corollary 3.7 (Group Egalitarian Welfare with Ellipsoidal Uncertainty)**.** *Suppose that the set* $\mathcal{V}$ *in* (5) *is defined by a single truncated ellipsoidal constraint per group i.e.,* $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^{nm} \mid \forall G \in \mathcal{G} : (\mathbf{v}_G - \bar{\mathbf{v}}_G)\mathbf{S}_G^{-1}(\mathbf{v}_G - \bar{\mathbf{v}}_G) \leq r_G^2, \mathbf{v} \succeq 0\}$. *Then the problem in* (5) *is equivalent to solving*

$$\max_{\substack{\boldsymbol{\lambda} \in \mathbb{R}_{0+}^{g} \\ \boldsymbol{\xi} \in \Lambda}} \min_{G \in \mathcal{G}} \boldsymbol{\xi}_G^\intercal \bar{\mathbf{v}}_G - \frac{\boldsymbol{\xi}_G^\intercal \mathbf{S}_G \boldsymbol{\xi}_G}{4\lambda_G} - \lambda_G r_G^2 \ .$$

The exact optimal solution $(\lambda^*, \boldsymbol{\xi}^*)$ *to Equation* (4) *can be computed by alternately performing two steps until convergence: first, fixing* $\boldsymbol{\xi}$ *and optimizing* $\boldsymbol{\lambda}$*, i.e.,* $\forall G \in \mathcal{G}$, $\lambda_G = \boldsymbol{\xi}_G^\intercal \mathbf{S}_G \boldsymbol{\xi}_G / 2r_G$*, and second, fixing* $\boldsymbol{\lambda}$ *and solving a concave quadratic program to optimize* $\boldsymbol{\xi}$*. The optimal allocation* $\mathbf{a}^*$ *can be computed from* $\boldsymbol{\xi}^*$ *as in Proposition 3.1.*

3.3 Robust Allocation for Monotonic Welfare Functions

We now extend our findings to a broader class of monotonic welfare functions. Specifically, we show that when optimizing a monotonic welfare objective under Assumption 3.4, we can decompose the problem into sub-problems such that we independently determine the worst valuation in the uncertainty set of each group, while jointly optimizing the allocation over all groups.

Proposition 3.8 (Decomposition for Monotonic Welfare Functions)**.** *Consider an optimization problem of the form*

$$\max_{\mathbf{a} \in \mathcal{A}} \min_{\mathbf{v} \in \mathcal{V}} \ \mathrm{W_M}(\mathbf{u}(\mathbf{a}, \mathbf{v})) \ , \tag{8}$$

where the welfare function $\mathrm{W_M}$ *is monotonic in the utility of groups. If Assumption 3.4 holds, then* (8) *simplifies to*

$$\max_{\mathbf{a} \in \mathcal{A}} \ \mathrm{W_M}\left(\min_{\mathbf{v}_{G_1} \in \mathcal{V}_{G_1}} \mathrm{u}_{G_1}(\mathbf{a}_{G_1}, \mathbf{v}_{G_1}), \min_{\mathbf{v}_{G_2} \in \mathcal{V}_{G_2}} \mathrm{u}_{G_2}(\mathbf{a}_{G_2}, \mathbf{v}_{G_2}), \ldots, \min_{\mathbf{v}_{G_g} \in \mathcal{V}_{G_g}} \mathrm{u}_{G_g}(\mathbf{a}_{G_g}, \mathbf{v}_{G_g})\right) \ .$$

Proposition 3.8 helps us derive simplified versions of Equation (8), when Assumption 3.4 holds. The egalitarian problem in (5) is an instance of the class of optimization problem described in (8), hence Proposition 3.8 holds under Assumption 3.4 and allows us to derive a single maximization problem (Proposition 3.5). If the allocation and valuation uncertainty sets are convex and compact, the problem in (8) can be solved using constrained convex-concave minimax optimization algorithms [18, 25, 55], or adversarial projected gradient ascent [16]. These approaches do not depend on Assumption 3.4, though the optimization may be simplified if independence does hold.

4 Stochastic Welfare Optimization

In this section, we optimize the CVaR of utilitarian and egalitarian welfare. This approach works when the distribution $\mathcal{D}_{\tilde{\mathbf{v}}}$ over the valuation matrix is known, or when we can sample from $\mathcal{D}_{\tilde{\mathbf{v}}}$. We demonstrate that when the distribution follows a Gaussian distribution, the CVaR of the utilitarian

welfare has a simple representation that can be optimized without sample approximation using a projected gradient ascent method. In all other cases, we can approximately optimize CVaR using a sampling-based approach. In particular, when we have monotone, concave welfare functions, we can always approximate the CVaR objective using sampling. However, unlike in Propositions 4.1 and H.3, where the approximated problem becomes linear, with arbitrary monotone, concave welfare functions the problem may require general concave optimization.

4.1 CVaR Allocation for Utilitarian Welfare

We wish to find an allocation that maximizes the CVaR_α of the weighted utilitarian welfare. Let $\tilde{\mathbf{v}}$ represent the random valuation vector. For confidence level α, we formulate the problem as

$$\max_{\mathbf{a}\in\mathcal{A}} \mathrm{CVaR}_\alpha\left[\mathbf{w}\cdot\mathbf{u}(\mathbf{a},\tilde{\mathbf{v}})\right] \doteq \max_{\mathbf{a}\in\mathcal{A}, b\in\mathbb{R}} \left\{ b - \frac{1}{\alpha}\mathop{\mathbb{E}}_{\tilde{\mathbf{v}}\sim\mathcal{D}_{\tilde{\mathbf{v}}}}\left[\left(b-\mathbf{w}\cdot\mathbf{u}(\mathbf{a},\tilde{\mathbf{v}})\right)_+\right]\right\} \ , \tag{9}$$

where $(x)_+ = \max(x,0)$ represents the positive part of x [50]. Computing the exact expectation in this problem may not be feasible for every distribution $\mathcal{D}_{\tilde{\mathbf{v}}}$. Therefore, we adopt a sampling-based approach. We begin by drawing h i.i.d. samples of the valuation matrix from $\mathcal{D}_{\tilde{\mathbf{v}}}$ represented as $\mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3, \ldots, \mathbf{v}^h$. We then use these samples to solve the problem described in (9) by solving the linear program outlined in Proposition 4.1.

Proposition 4.1 (Approximate CVaR of USW)**.** *Given h i.i.d samples of $\tilde{\mathbf{v}}$, i.e., $\mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3, \ldots, \mathbf{v}^h$ from $\mathcal{D}_{\tilde{\mathbf{v}}}$, the optimal allocation for the problem in* (9) *can be approximately computed by solving*

$$\max_{\mathbf{a}\in\mathcal{A}} \max_{\substack{\boldsymbol{y}\in\mathbb{R}_{0+}^h \\ b\in\mathbb{R}}} \left(b - \frac{1}{\alpha}\sum_{j=1}^{h}\boldsymbol{y}_j\right) \qquad \forall j\in[1,h]: \ \boldsymbol{y}_j \geq \frac{1}{h}\left(b-\mathbf{w}\cdot\mathbf{u}(\mathbf{a},\mathbf{v}^j)\right) \ . \tag{10}$$

The CVaR estimator used in (10) is a strongly consistent estimator [29]. Therefore, the approximation error of the objective in (10) goes to 0 as $h\to\infty$. In Proposition 4.2, we bound the sample complexity of the problem in (10) when the valuation matrix is sub-Gaussian distributed.

For any allocation $\mathbf{a}$, let $\hat{c}_{h,\alpha}(\mathbf{a})$ represent the empirical estimate of $\mathrm{CVaR}_\alpha[\mathbf{w}\cdot\mathbf{u}(\mathbf{a},\tilde{\mathbf{v}})]$ computed from h samples and $c_{h,\alpha}(\mathbf{a})$ represent the corresponding true value. We will use $|\mathcal{A}|$ to denote the number of feasible allocations and $f_{\mathbf{a}}:\mathbb{R}\to\mathbb{R}_{0+}$ to denote the density function of the random welfare $\mathrm{W}(\mathbf{a},\tilde{\mathbf{v}})$. ν_α denotes the α percentile of $\mathrm{W}(\mathbf{a},\tilde{\mathbf{v}})$.

Proposition 4.2 (Sample Complexity of Approximate CVaR of USW)**.** *Suppose that $\tilde{\mathbf{v}}$ is a multivariate sub-Gaussian random variable with mean $\bar{\mathbf{v}}\in\mathbb{R}^{nm}$ and covariance proxy $\mathbf{S}\in\mathbb{R}^{nm\times nm}$, i.e., for all vectors $\boldsymbol{z}\in\mathbb{R}^{nm}$: $\mathbb{E}_{\mathbf{v}\sim\mathcal{D}_{\tilde{\mathbf{v}}}}\left[\exp((\mathbf{v}-\bar{\mathbf{v}})^\intercal\boldsymbol{z})\right] \leq \exp(\boldsymbol{z}^\intercal\mathbf{S}\boldsymbol{z}/2)$, and that, for any risk level $\alpha\in(0,\frac{1}{2})$ and allocation $\mathbf{a}\in\mathcal{A}$, there exists probability density threshold $\eta>0$ and radius $\gamma>0$, s.t., $f_{\mathbf{a}}(x)>\eta, \forall x\in[\nu_\alpha-\gamma,\nu_\alpha+\gamma]$. Set $\forall G\in\mathcal{G}: \mathbf{a}'_G = \frac{\mathbf{w}_G\cdot\mathbf{a}_G}{|G|}$. Then, for any confidence parameter $\delta\in(0,1)$ and error tolerance $\varepsilon>0$,*

$$\Pr\left[\sup_{\mathbf{a}\in\mathcal{A}}|\hat{c}_{h,\alpha}(\mathbf{a})-c_\alpha(\mathbf{a})|\leq\varepsilon\right]\geq 1-\delta \quad \textit{for} \quad h > \left\lceil \frac{8\max\left(\max_{\mathbf{a}\in\mathcal{A}}\mathbf{a}'^\intercal\mathbf{S}\mathbf{a}',8\right)\ln\left(\frac{6|\mathcal{A}|}{\delta}\right)}{\min(\varepsilon^2,16\gamma^2)\alpha^2\min(\eta^2,1)}\right\rceil \ .$$

Proposition 4.2 follows directly from Theorem 3.1 of L.A. *et al.* [35]. When $\tilde{\mathbf{v}}$ is Gaussian distributed, we can circumvent the sampling approach and instead solve an optimization problem (Proposition 4.3), which depends solely on the mean and covariance of $\tilde{\mathbf{v}}$.

Proposition 4.3 (CVaR of USW for Gaussian Distributions)**.** *If $\tilde{\mathbf{v}}$ is distributed as a multivariate Gaussian, i.e., $\tilde{\mathbf{v}}\sim\mathcal{N}(\bar{\mathbf{v}},\mathbf{S})$, then, the optimization problem in* (9) *simplifies to*

$$\max_{\mathbf{a}\in\mathcal{A}} \mathbf{a}^\intercal\bar{\mathbf{v}} - \frac{\phi(\Phi^{-1}(\alpha))}{\alpha}\sqrt{\mathbf{a}^\intercal\mathbf{S}\mathbf{a}} \ , \tag{11}$$

where ϕ and Φ denote the probability density function and cumulative density function, respectively, of the standard normal distribution $\mathcal{N}(0,1)$. The problem in (11) is concave and can be solved without sample approximation using the projected gradient ascent method.

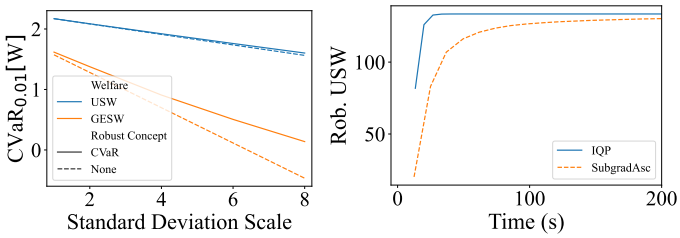


Figure 1: Left: CVaR as noise increases for AAMAS 2015. Right: Convergence behavior of the Iterated Quadratic Program (Iterated QP) vs. Adversarial Projected Subgradient Ascent approach on AAMAS 2015.

4.2 CVaR Allocation for Group Egalitarian Welfare

For our final objective, we wish to optimize egalitarian welfare under uncertainty using the CVaR approach. We formulate this optimization problem as

$$\max_{\mathbf{a}\in\mathcal{A}} \mathrm{CVaR}_\alpha\left[\min_{G\in\mathcal{G}} \mathbf{u}_G(\mathbf{a},\tilde{\mathbf{v}})\right] \doteq \max_{\mathbf{a}\in\mathcal{A}, b\in\mathbb{R}} \left\{ b - \frac{1}{\alpha} \mathop{\mathbb{E}}_{\tilde{\mathbf{v}}\sim\mathcal{D}_{\tilde{\mathbf{v}}}} \left[\left(b - \min_{G\in\mathcal{G}} \mathrm{u}_G(\mathbf{a},\tilde{\mathbf{v}}) \right)_+ \right] \right\} . \tag{12}$$

To optimize the problem described in the above equation, we solve a linear program similar to the one used for optimizing the CVaR utilitarian objective in (9). See Proposition H.3 for more details.

5 Experiments

We run experiments on three reviewer assignment datasets. The datasets contain bids from the International Conference on Autonomous Agents and Multiagent Systems (AAMAS) 2015, 2016, and 2021 [41, 42]. We consider the papers as the "agents" and the reviewers as the "items." This is a fairly standard assumption in most recent reviewer assignment approaches, reflecting the primary goal of peer review to assign qualified and interested reviewers to papers [16, 30, 31, 38, 45, 54].

Reviewers issue bids of `yes`, `maybe`, `no`, or `no response`. We run two experiments with this data. In one, we binarize the bids such that `yes` and `maybe` are considered affirmative and `no` is considered negative, while in the other we convert the bids to numerical scores such that `yes` is 1, `maybe` is .5, and `no` is 0.01. Under the binarized model, we fit a logistic matrix factorization model to predict whether the bid is affirmative or negative, and in the continuous model, we fit a Gaussian process matrix factorization model [36]. We derive probability distributions and uncertainty sets from these models. More details on prediction and uncertainty set construction are in Appendix E. These datasets do not contain groups of papers and reviewers, so we create 4 roughly balanced clusters of reviewers and papers for each dataset using the procedure outlined in Appendix F. We define our valid set of assignments $\mathcal{A}$ as follows. For each paper $a \in N$, we set $\underline{\kappa}_a = \bar{\kappa}_a = 3$ for all a in AAMAS 2015, and $\underline{\kappa}_a = \bar{\kappa}_a = 2$ for all a in AAMAS 2016 and 2021. For each reviewer i, we set $\underline{\psi}_i = 0$ and $\bar{\psi}_i = 15$ for 2015 and 2016 and 4 for 2021. We optimize and evaluate $\mathrm{CVaR}_{0.01}$; we take $4,000$ samples from the distribution to optimize for CVaR using the sampling-based approach, and we take $10,000$ samples to estimate the CVaR for evaluation. We optimize and evaluate the robust welfare at the $\alpha = 0.3$ level (there is a 70% chance the true values lie in the uncertainty set). We constrain the naïve and CVaR approaches to select integer allocations, while the robust approach selects fractional allocations without rounding.

All results are averaged over 5 runs of subsampling 20% of each dataset. For each run, we construct 6 allocations, maximizing the naïve central estimate, CVaR and robust statistics for USW and GESW respectively. We evaluate each allocation on each metric. For each run, we normalize each metric by the maximum value achieved for that metric by any allocation. We normalize in this manner to highlight that the allocation targeted for a given objective always returns the highest value on that objective, and because the absolute optimal values differ across runs.

All code is available at https://github.com/justinpayan/RAU2.

Table 2: Performance of different allocations across each metric on the AAMAS 2015 dataset.

Allocation	Evaluation Objective					
	USW	GESW	CVaR USW	CVaR GESW	Rob. USW	Rob. GESW
USW	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	0 ± 0	0 ± 0
GESW	0.97 ± 0.01	$\mathbf{1.00 \pm 0}$	0.97 ± 0.01	0.97 ± 0.02	0 ± 0	0 ± 0
CVaR USW	$\mathbf{1.00 \pm 0}$	0.99 ± 0	$\mathbf{1.00 \pm 0}$	0.99 ± 0	0 ± 0	0 ± 0
CVaR GESW	0.98 ± 0	0.99 ± 0	0.97 ± 0.01	$\mathbf{1.00 \pm 0}$	0 ± 0	0 ± 0
Rob. USW	0.92 ± 0.01	0.90 ± 0.02	0.92 ± 0.01	0.90 ± 0.02	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$
Rob. GESW	0.89 ± 0.04	0.85 ± 0.06	0.89 ± 0.04	0.86 ± 0.06	0.88 ± 0.02	$\mathbf{1.00 \pm 0}$

Overall Performance Table 2 shows the results for the binarized version of AAMAS 2015 bids. Similar tables for the 5 other settings are included in Appendix G. Each row shows the metrics for the allocation produced by the method which optimizes for the objective shown in the left-most column. All non-robust methods have 0 robust welfare, indicating that if robustness to adversarial noise is desired, it is very important to consider this objective explicitly. Relatively little noise is actually present in this dataset, as the $\mathrm{CVaR}_{0.01}$ is quite high for both the naïve USW-optimal and GESW-optimal in all cases.

Since the USW-optimal solution has very high GESW, we implement a simulated example to explore when the USW-optimal solution fails to have high GESW. We find that in a number of settings, the GESW of the USW-optimal solution is much lower than the GESW-optimal solution. Appendix I explains the details of the simulation setting and the results.

Robustness under Increasing Uncertainty Figure 1 shows the $\mathrm{CVaR}_{0.01}$ on the Gaussian version of all three datasets as we artificially increase the amount of noise. We multiply the standard deviations of the Gaussian distributions by a scalar and optimize for the CVaR or the naïvely-computed USW and GESW. We then plot $\mathrm{CVaR}_{0.01}$ as noise increases. Although the CVaR approach is less important at low noise levels, the CVaR of welfare decreases for both welfare measures as noise increases. GESW has a sharper decline than USW. We see that as the noise increases, the $\mathrm{CVaR}_{0.01}$ of the baseline USW and GESW maximizing allocations drops off relative to the same value for the CVaR-optimized allocation.

We also verify that when we model valuations using a negatively-skewed Gaussian distribution with the same means and variances, we see increasing importance of optimizing for CVaR relative to uncertainty-unaware USW and robust USW. The difference is sharper as the skew parameter gets more negative. Details of this experiment and its results are included in Appendix J.

Runtime For the robust optimization setting with ellipsoidal uncertainty sets (derived from confidence intervals over the Gaussian process matrix factorization), we compare the Iterated QP approach (Corollary 3.3) to adversarial projected subgradient ascent on the original max-min problem (as in [16]). We find that Iterated QP converges much faster than the adversarial projected subgradient ascent algorithm on both AAMAS 2015 (Figure 1) and 2016 (Figure 3). Adversarial projected subgradient ascent fails to converge in $1,000$ iterations for the robust GESW objective on all datasets and the USW objective on AAMAS 2021.

6 Conclusion

In conclusion, we explore the stochastic and robust optimization regimes for utilitarian and group egalitarian welfare objectives. The robust optimization algorithms depend on the form of the uncertainty set. We show that when the uncertainty set has linear constraints only, the resulting problem is an LP and can be solved efficiently. Under ellipsoidal constraints, we demonstrate an iterative quadratic programming approach converges much faster than adversarial projected subgradient ascent. In the stochastic regime, we lay out the sample complexity of CVaR for the utilitarian welfare objective. We demonstrate the feasibility of estimating probability distributions and uncertainty sets on three years of bid data from AAMAS, and show that the robust and CVaR approaches demonstrated in this paper combat the uncertainty present in these three datasets.

Acknowledgments and Disclosure of Funding

This work is generously supported by Army Research Lab DEVCOM Data and Analysis Center - Contract W911QX23D0009 and NSF grant IIS-2327057. Cousins was supported by the University of Massachusetts Center for Data Science Fellowship. This work was performed using high performance computing equipment obtained under a grant from the Collaborative R&D Fund managed by the Massachusetts Technology Collaborative.

References

[1] Emily L Aiken, Guadalupe Bedoya, Joshua E Blumenstock, and Aidan Coville. Program targeting with machine learning and mobile phone data: Evidence from an anti-poverty intervention in afghanistan. *Journal of Development Economics*, 161:103016, 2023.

[2] Amine Allouah, Christian Kroer, Xuan Zhang, Vashist Avadhanula, Anil Dania, Caner Gocmen, Sergey Pupyrev, Parikshit Shah, and Nicolas Stier. Robust and fair work allocation. *arXiv preprint arXiv:2202.05194*, 2022.

[3] Igor Averbakh. Minmax regret solutions for minimax optimization problems with uncertainty. *Operations Research Letters*, 27(2):57–65, 2000.

[4] Haris Aziz, Evi Micha, and Nisarg Shah. Group fairness in peer review. *Proceedings of the 37th Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 64885–64895, 2023.

[5] Mohammad Javad Azizi, Phebe Vayanos, Bryan Wilder, Eric Rice, and Milind Tambe. Designing fair, efficient, and interpretable policies for prioritizing homeless youth for housing resources. In *Proceedings of the 15th International Conference on the Integration of Constraint Programming, Artificial Intelligence, and Operations Research (CPAIOR)*, pages 35–51, 2018.

[6] Adelchi Azzalini and Antonella Capitanio. Statistical applications of the multivariate skew normal distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 61(3):579–602, 1999.

[7] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*. Princeton University Press, 2009.

[8] Dimitris Bertsimas, David B Brown, and Constantine Caramanis. Theory and applications of robust optimization. *SIAM Review*, 53(3):464–501, 2011.

[9] John R Birge and Francois Louveaux. *Introduction to stochastic programming*. Springer Science & Business Media, 2011.

[10] Craig Boutilier, Relu Patrascu, Pascal Poupart, and Dale Schuurmans. Constraint-based optimization and utility elicitation using the minimax decision criterion. *Artificial Intelligence*, 170(8-9):686–713, 2006.

[11] Craig Boutilier. Computational decision support: Regret-based models for optimization and preference elicitation. In Thomas Zentall and Philip Crowley, editors, *Comparative Decision Making*, chapter 14, pages 423–453. Oxford University Press, 2013.

[12] Stephen Boyd, Stephen P Boyd, and Lieven Vandenberghe. *Convex optimization*. Cambridge University Press, 2004.

[13] Eric Budish, Yeon-Koo Che, Fuhito Kojima, and Paul Milgrom. Implementing random assignments: A generalization of the Birkhoff-von Neumann theorem. In *Cowles Summer Conference*, 2009.

[14] Jan Buermann, Enrico H Gerding, and Baharak Rastegari. Fair allocation of resources with uncertain availability. In *Proceedings of the 19th International Conference on Autonomous Agents and Multi-Agent Systems (AAMAS)*, pages 204–212, 2020.

[15] Ioannis Caragiannis, David Kurokawa, Hervé Moulin, Ariel D Procaccia, Nisarg Shah, and Junxing Wang. The unreasonable fairness of maximum Nash welfare. *ACM Transactions on Economics and Computation (TEAC)*, 7(3):1–32, 2019.

[16] Cyrus Cousins, Justin Payan, and Yair Zick. Into the unknown: Assigning reviewers to papers with uncertain affinities. In *Proceedings of the 16th International Symposium on Algorithmic Game Theory (SAGT)*, pages 179–197, 2023.

[17] Cyrus Cousins. An axiomatic theory of provably-fair welfare-centric machine learning. In *Proceedings of the 35th Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 16610–16621, 2021.

[18] Constantinos Daskalakis, Stratis Skoulakis, and Manolis Zampetakis. The complexity of constrained min-max optimization. In *Proceedings of the 53rd Annual ACM Symposium on Theory of Computing (STOC)*, page 1466–1478, 2021.

[19] Anand Deo and Karthyek Murthy. Efficient black-box importance sampling for VaR and CVaR estimation. In *Proceedings of the 2021 Winter Simulation Conference (WSC)*, pages 1–12, 2021.

[20] Siddartha Devic, David Kempe, Vatsal Sharan, and Aleksandra Korolova. Fairness in matching under uncertainty. *Proceedings of the 40th International Conference on Machine Learning (ICML)*, pages 7775–7794, 2023.

[21] Kate Donahue and Jon Kleinberg. Fairness and utilization in allocating resources with uncertain demand. In *Proceedings of the 3rd ACM Conference on Fairness, Accountability and Transparency (FAT*)*, pages 658–668, 2020.

[22] Evan Dong and Cyrus Cousins. Decentering imputation: Fair learning at the margins of demographics. In *Proceedings of the 2022 ICML Queer in AI Workshop*, 2022.

[23] John C Duchi and Hongseok Namkoong. Learning models with uniform performance via distributionally robust optimization. *The Annals of Statistics*, 49(3):1378–1406, 2021.

[24] Corrado Gini. On the measure of concentration with special reference to income and statistics. *Colorado College Publication, General Series*, 208(1):73–79, 1936.

[25] Denizalp Goktas and Amy Greenwald. Convex-concave min-max Stackelberg games. *Proceedings of the 35th Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 2991–3003, 2021.

[26] Bram L Gorissen, İhsan Yanıkoğlu, and Dick Den Hertog. A practical guide to robust optimization. *Omega*, 53:124–137, 2015.

[27] Negin Gorlezaei, Patrick Jaillet, and Zijie Zhou. Online resource allocation with samples. *arXiv preprint arXiv:2210.04774*, 2022.

[28] Gurobi Optimization, LLC. Gurobi Optimizer Reference Manual, 2023. Accessed August 21, 2024.

[29] L. Jeff Hong and Guangwu Liu. Monte Carlo estimation of value-at-risk, conditional value-at-risk and their sensitivities. In *Proceedings of the 2011 Winter Simulation Conference (WSC)*, pages 95–107, 2011.

[30] Steven Jecmen, Hanrui Zhang, Ryan Liu, Nihar Shah, Vincent Conitzer, and Fei Fang. Mitigating manipulation in peer review via randomized reviewer assignments. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 12533–12545, 2020.

[31] Ari Kobren, Barna Saha, and Andrew McCallum. Paper matching with local fairness constraints. In *Proceedings of the 25th International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 1247–1257, 2019.

[32] Panos Kouvelis and Gang Yu. *Robust discrete optimization and its applications*. Springer Science & Business Media, 2013.

[33] Pavlo Krokhmal, Jonas Palmquist, and Stanislav Uryasev. Portfolio optimization with conditional value-at-risk objective and constraints. *Journal of Risk*, 4:43–68, 2002.

[34] Amanda R Kube, Sanmay Das, and Patrick J Fowler. Fair and efficient allocation of scarce resources based on predicted outcomes: Implications for homeless service delivery. *Journal of Artificial Intelligence Research*, 76:1219–1245, 2023.

[35] Prashanth L.A., Krishna Jagannathan, and Ravi Kolla. Concentration bounds for CVaR estimation: The cases of light-tailed and heavy-tailed distributions. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, pages 5577–5586, 2020.

[36] Neil D Lawrence and Raquel Urtasun. Non-linear matrix factorization with Gaussian processes. In *Proceedings of the 26th International Conference on Machine Learning (ICML)*, pages 601–608, 2009.

[37] Daniel Levy, Yair Carmon, John C Duchi, and Aaron Sidford. Large-scale methods for distributionally robust optimization. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 8847–8860, 2020.

[38] Kevin Leyton-Brown, Mausam, Yatin Nandwani, Hedayat Zarkoob, Chris Cameron, Neil Newman, and Dinesh Raghu. Matching papers and reviewers at large conferences. *Artificial Intelligence*, 331:104119, 2024.

[39] Richard J Lipton, Evangelos Markakis, Elchanan Mossel, and Amin Saberi. On approximately fair allocations of indivisible goods. In *Proceedings of the 5th ACM Conference on Electronic Commerce (EC)*, pages 125–131, 2004.

[40] Elita A Lobo, Mohammad Ghavamzadeh, and Marek Petrik. Soft-robust algorithms for batch reinforcement learning. *arXiv preprint arXiv:2011.14495*, 2020.

[41] Nicholas Mattei and Toby Walsh. Preflib: A library of preference data. In *Proceedings of the 3rd International Conference on Algorithmic Decision Theory (ADT)*, pages 259–270, 2013.

[42] Nicholas Mattei and Toby Walsh. A preflib.org retrospective: Lessons learned and new directions. *Trends in Computational Social Choice. AI Access Foundation*, pages 289–309, 2017.

[43] Hongseok Namkoong and John C Duchi. Stochastic gradient methods for distributionally robust optimization with f-divergences. *Proceedings of the 30th Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 2208–2216, 2016.

[44] Włodzimierz Ogryczak, Adam Wierzbicki, and Marcin Milewski. A multi-criteria approach to fair and efficient bandwidth allocation. *Omega*, 36(3):451–463, 2008.

[45] Justin Payan and Yair Zick. I will have order! Optimizing orders for fair reviewer assignment. In *Proceedings of the 31st International Joint Conference on Artificial Intelligence (IJCAI)*, pages 440–446, 2022.

[46] Dominik Peters, Ariel D Procaccia, and David Zhu. Robust rent division. In *Proceedings of the 36th Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 13864–13876, 2022.

[47] András Prékopa. *Stochastic programming*, volume 324. Springer Science & Business Media, 2013.

[48] David Pujol, Ryan McKenna, Satya Kuppam, Michael Hay, Ashwin Machanavajjhala, and Gerome Miklau. Fair decision making using privacy-protected data. In *Proceedings of the 2020 ACM Conference on Fairness, Accountability and Transparency (FAT*)*, pages 189–199, 2020.

[49] Aida Rahmattalabi, Phebe Vayanos, Kathryn Dullerud, and Eric Rice. Learning resource allocation policies from observational data with an application to homeless services delivery. In *Proceedings of the 2022 ACM Conference on Fairness, Accountability and Transparency (FAccT)*, pages 1240–1256, 2022.

[50] R Tyrrell Rockafellar and Stanislav Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2:21–42, 2000.

[51] Erik Rolland, Raymond A Patterson, Keith Ward, and Bajis Dodin. Decision support for disaster management. *Operations Management Research*, 3:68–79, 2010.

[52] Andrzej Ruszczyński and Alexander Shapiro. Stochastic programming models. *Handbooks in Operations Research and Management Science*, 10:1–64, 2003.

[53] Isaac Slavitt. Prioritizing municipal lead mitigation projects as a relaxed knapsack optimization: a method and case study. *International Transactions in Operational Research*, 30(6):3719–3737, 2023.

[54] Ivan Stelmakh, Nihar B Shah, and Aarti Singh. PeerReview4All: Fair and accurate reviewer assignment in peer review. In *Proceedings of the 30th International Conference on Algorithmic Learning Theory (ALT)*, pages 828–856, 2019.

[55] Yuanhao Wang and Jian Li. Improved algorithms for convex-concave minimax optimization. In *Proceedings of the 34th Annual Conference on Neural Information Processing Systems (NeurIPS)*, pages 4800–4810, 2020.

[56] Felix Wex, Guido Schryen, and Dirk Neumann. A fuzzy decision support model for natural disaster response under informational uncertainty. *International Journal of Information Systems for Crisis Response and Management (IJISCRAM)*, 4(3):23–41, 2012.

[57] Felix Wex, Guido Schryen, Stefan Feuerriegel, and Dirk Neumann. Emergency response in natural disaster management: Allocation and scheduling of rescue units. *European Journal of Operational Research*, 235(3):697–708, 2014.

[58] Tony Sit Yue Xing and Hoi Ying Wong. Variance reduction for risk measures with importance sampling in nested simulation. *Quantitative Finance*, 22(4):657–673, 2022.

[59] Wenbin Zhang and Jeremy C Weiss. Fair decision-making under uncertainty. In *Proceedings of the 21st IEEE International Conference on Data Mining (ICDM)*, pages 886–895, 2021.

A Additional Related Work

Gorissen *et al.* [26] provide an excellent overview of optimization under uncertainty, including techniques used in this work, while Ben-Tal *et al.* [7], Bertsimas *et al.* [8] offer additional background on robust optimization. A standard approach in this regime is analyzing the dual of the uncertainty, as we generally do in this work. Stochastic optimization has a wide literature; the books by Birge and Louveaux [9], Levy *et al.* [37], Prékopa [47], Ruszczyński and Shapiro [52] present wide-ranging introductions to the topic. Conditional value at risk (CVaR) can often be approximately optimized by sampling and optimizing over an objective composing the different samples [35, 40, 50].

B Broader Impacts

We believe this work has the potential for a significant positive societal impact. Fair resource allocation algorithms are essential for various systems, including assigning reviewers in peer review processes, allocating resources to homeless and low-income populations, distributing emergency response resources during natural disasters, and resettling refugees. In this work, we develop methods for efficiently optimizing allocations of constrained resources under various fairness objectives while addressing uncertainty in resource preferences. These methods can be directly applied to the aforementioned problems. However, we advise users to conduct extensive testing on similar datasets before deploying these algorithms in real-world scenarios. We also encourage users to perform rigorous causal analysis and controlled experiments to validate their predictive models before applying our methods.

C Limitations

The CVaR approach requires solving linear programs with a large number of samples to be effective, which makes them computationally expensive. One potential solution is to leverage importance sampling methods to reduce the variance of the estimator [19, 58]. Future research could benefit from empirically and theoretically analyzing other fairness objectives like Nash welfare [15], Gini index [24], and envy-freeness [39].

D Constructing uncertainty sets

In this section we demonstrate a simple and natural approach to construct an uncertainty set using a logistic regression estimator. Logistic regression models with bounded cross-entropy loss result in polyhedral uncertainty sets. Replacing the logistic regression model with a model with bounded squared-error loss, or simply taking the confidence interval of a multivariate Gaussian, results in truncated ellipsoidal uncertainty sets. We construct uncertainty sets per group in all cases.

Assume we have a discrete set of c values $L \subseteq \mathbb{R}$, with $L = \{\ell_1, \dots \ell_c\}$. For each agent i and item type j we denote the true distribution over values $p^*(\ell|(i,j))$ and the distribution predicted by the logistic regression model is $\hat{p}(\ell|(i,j))$.

We estimate the cross-entropy loss of the model on a test set T, where $|T| = t$. This test set can be segmented by the group identities of the agents, such that we have $T_{G_1}, T_{G_2}, \dots T_{G_g}$ for each of the g groups (with sizes $t_{G_1}, \dots t_{G_g}$). We assume that the test set comes from the same distribution as the agent-item pairs of the assignment problem; this can be achieved either during dataset construction or by limiting the assignments (through the $\boldsymbol{C}$ constraints) to better reflect the test distribution. We can also apply likelihood reweighting in our uncertainty set construction, as in [16], though we do not do so here.

For an agent a and item type i, the cross-entropy loss of the distribution $\hat{p}$ with respect to the distribution p is defined as

$\mathbb{H}(p(\ell|(a,i)), \hat{p}(\ell|(a,i))) \doteq -\sum_{\ell \in L} p(\ell|(a,i)) \ln \hat{p}(\ell|(a,i))$. For each T_G, we compute the mean of the cross-entropy loss $\hat{\mu}_G = \frac{1}{t_G} \sum_{(a,i) \in T_G} \mathbb{H}(p(\ell|(a,i)), \hat{p}(\ell|(a,i)))$, as well as the standard error of the mean $\hat{\sigma}_G = \left(\frac{1}{t_G} \sum_{(i,j) \in T_G} (\mathbb{H}(p(\ell|(a,i)), \hat{p}(\ell|(a,i))) - \hat{\mu}_G)^2 \right)^{\frac{1}{2}}$. We model the distribution over cross-entropy losses for group G as $\mathcal{N}(\hat{\mu}_G, \hat{\sigma}_G)$. We want an uncertainty set $\mathcal{V}$ such that

the true values lie outside $\mathcal{V}$ with probability at most α. Thus, using a union bound, we require each uncertainty set $\mathcal{V}_G$ for individual groups to contain the true valuations with probability at least $1 - \frac{\alpha}{g}$. We can thus give the bound that the cross entropy loss is at most $\Phi^{-1}(1 - \frac{\alpha}{g}, \hat{\mu}_G, \hat{\sigma}_G)$, where $\Phi^{-1}(p, \mu, \sigma)$ denotes the p percentile of a Gaussian distribution with mean μ and standard deviation σ.

In our assignment problem, for each group G with agents N_G we obtain the uncertainty set

$$\frac{1}{t_G m} \sum_{a \in N_G, i \in I} \mathbb{H}(p(\ell|(a,i)), \hat{p}(\ell|(a,i))) \leq \Phi^{-1}(1 - \frac{\alpha}{g}, \hat{\mu}_G, \hat{\sigma}_G) \ .$$

The bound can be made tighter if we restrict some pairs using $\boldsymbol{C}$, in which case the cross-entropy term on the left side is only averaged over the pairs which are not restricted.

E Logistic and Gaussian Process Matrix Factorization

In this section we define matrix factorization models using either logistic regression or Gaussian processes. The logistic regression model can be used for predicting the missing elements of the binarized bid matrices, while the Gaussian process model is used for predicting missing real-valued elements. Both models define probability distributions over outcomes, which we use to compute and evaluate the CVaR of utilitarian and egalitarian welfare. For the logistic model, we build a polyhedral uncertainty set by estimating the cross-entropy loss on a held-out test set (as shown in Appendix D), and for the Gaussian process model we simply consider the confidence intervals of the resulting Gaussian distribution.

For the binarized bids, we first set aside some of the observed bids as a test set. We estimate the missing bids and the bids for the held-out test pairs using logistic matrix factorization. Setting a hidden dimension size d, we construct two matrices $\mathbf{X} \in \mathbb{R}^{n \times d}$ and $\mathbf{Y} \in \mathbb{R}^{m \times d}$. We set $d = 20$. Let $\mathbf{V}^*$ denote the true binarized bid matrix, where we observe entries for the training set pairs $(a, i) \in T$. We predict the probability of an affirmative bid as $f((\mathbf{X}\mathbf{Y}^\intercal)_{a,i})$ where f is the logistic sigmoid function. We select $\mathbf{X}$ and $\mathbf{Y}$ to minimize the loss function

$$\sum_{(a,i) \in T} -\mathbf{V}^*_{a,i} \ln\left(f((\mathbf{X}\mathbf{Y}^\intercal)_{a,i})\right) - (1 - \mathbf{V}^*_{a,i}) \ln\left(1 - f((\mathbf{X}\mathbf{Y}^\intercal)_{a,i})\right) \ .$$

For CVaR, we take samples from the distribution defined by $f(\mathbf{X}\mathbf{Y}^\intercal)$, assuming all pairs are independently-distributed. We also construct an uncertainty set as described in Appendix D using the cross-entropy loss on the test pairs.

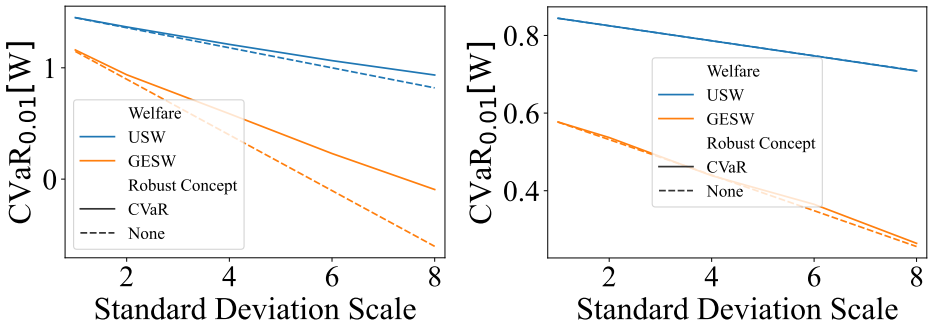
Under the Gaussian process matrix factorization model [36], we simply predict a mean and variance of a Gaussian distribution for each reviewer-paper pair. We can then sample values independently for each pair, or give a confidence interval for the joint Gaussian with $mn - 1$ degrees of freedom.

F Grouping Papers and Reviewers

We group papers and reviewers as follows: given the real-valued bids in the set $\{0.01, .5, 1\}$ we set unknown bids to be 0. We then construct a graph with all reviewers and papers as nodes, and the bid score between reviewers and papers is the edge weight. All inter-reviewer and inter-paper edges are set to 0 edge weight. We apply spectral embedding with 5 dimensions to transform the nodes into vectors, and cluster the resulting vectors into 4 clusters to obtain 4 groups containing both papers and reviewers. To ensure a balance of reviewers and papers across clusters, we employ Lloyd's algorithm for KMeans clustering with the modification that during each assignment step we enforce a lower bound on the number of papers and number of reviewers assigned to each cluster.

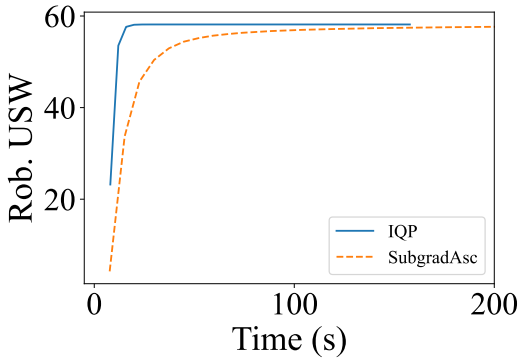
G Additional Experiments

For the binarized AAMAS 2016 and 2021 datasets, Tables 3 and 4 show the performance of the baseline USW and GESW maximizing allocations, the $\mathrm{CVaR}_{0.01}$ USW and GESW maximizing allocations, and the robust USW and GESW maximizing allocations at the $\alpha = 0.3$ level. Because



(a) AAMAS 2016 (b) AAMAS 2021

Figure 2: $\mathrm{CVaR}_{0.01}$ as noise increases for AAMAS 2016 and 2021.



(a) AAMAS 2016

Figure 3: Convergence of the Iterated QP vs. adversarial projected subgradient ascent on AAMAS 2016 dataset for the adversarial USW objective. The Iterated QP (in blue) converges much faster.

so many of the bids in AAMAS 2021 are recorded as `no`, since `no` is the default bid, we randomly select 90% of the `no` bids to be converted to `no response`.

Tables 5 to 7 show the same results for the Gaussian matrix factorization version of the 3 datasets, with the $\mathrm{CVaR}_{0.01}$ estimated by sampling from the estimated Gaussian distribution, and the adversarial welfare computed over the truncated ellipsoidal uncertainty set corresponding to the $1-\alpha$ confidence interval of the Gaussian.

Table 3: Performance of different allocations across each metric on the AAMAS 2016 dataset.

Allocation	Evaluation Objective					
	USW	GESW	CVaR USW	CVaR GESW	Rob. USW	Rob. GESW
USW	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	0 ± 0	0 ± 0
GESW	0.99 ± 0	$\mathbf{1.00 \pm 0}$	0.99 ± 0	0.99 ± 0.01	0 ± 0	0 ± 0
CVaR USW	0.99 ± 0	0.98 ± 0.01	0.99 ± 0	0.98 ± 0.01	0 ± 0	0 ± 0
CVaR GESW	0.99 ± 0.01	0.99 ± 0.01	0.98 ± 0.01	$\mathbf{1.00 \pm 0}$	0 ± 0	0 ± 0
Rob. USW	0.91 ± 0.02	0.87 ± 0.03	0.91 ± 0.02	0.90 ± 0.03	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$
Rob. GESW	0.76 ± 0.05	0.66 ± 0.04	0.76 ± 0.05	0.65 ± 0.05	0.74 ± 0.10	$\mathbf{1.00 \pm 0}$

Table 4: Performance of different allocations across each metric on the AAMAS 2021 dataset.

Allocation	Evaluation Objective					
	USW	GESW	CVaR USW	CVaR GESW	Rob. USW	Rob. GESW
USW	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	0 ± 0	0.40 ± 0.49
GESW	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	0 ± 0	0.40 ± 0.49
CVaR USW	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	0 ± 0	0.40 ± 0.49
CVaR GESW	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$	0.99 ± 0	$\mathbf{1.00 \pm 0}$	0 ± 0	0.40 ± 0.49
Rob. USW	0.85 ± 0.04	0.69 ± 0.14	0.84 ± 0.05	0.64 ± 0.19	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$
Rob. GESW	0.48 ± 0.09	0.32 ± 0.12	0.43 ± 0.09	0.20 ± 0.12	0.07 ± 0.08	$\mathbf{1.00 \pm 0}$

Table 5: Performance of different allocations across each metric on the Gaussian AAMAS 2015 dataset.

Allocation	Evaluation Objective					
	USW	GESW	CVaR USW	CVaR GESW	Rob. USW	Rob. GESW
USW	$\mathbf{1.00 \pm 0}$	0.95 ± 0.03	$\mathbf{1.00 \pm 0}$	0.94 ± 0.04	0.61 ± 0.19	0.34 ± 0.34
GESW	0.87 ± 0.08	$\mathbf{1.00 \pm 0}$	0.86 ± 0.09	0.98 ± 0.02	0.42 ± 0.30	0.32 ± 0.35
CVaR USW	$\mathbf{1.00 \pm 0}$	0.94 ± 0.03	$\mathbf{1.00 \pm 0}$	0.96 ± 0.04	0.63 ± 0.19	0.35 ± 0.34
CVaR GESW	0.90 ± 0.06	0.99 ± 0.01	0.90 ± 0.07	$\mathbf{1.00 \pm 0}$	0.51 ± 0.26	0.36 ± 0.33
Rob. USW	0.86 ± 0.07	0.76 ± 0.12	0.88 ± 0.06	0.80 ± 0.10	$\mathbf{1.00 \pm 0}$	0.99 ± 0.01
Rob. GESW	0.75 ± 0.13	0.77 ± 0.12	0.76 ± 0.13	0.82 ± 0.09	0.87 ± 0.09	$\mathbf{1.00 \pm 0}$

Table 6: Performance of different allocations across each metric on the Gaussian AAMAS 2016 dataset.

Allocation	Evaluation Objective					
	USW	GESW	CVaR USW	CVaR GESW	Rob. USW	Rob. GESW
USW	$\mathbf{1.00 \pm 0}$	0.99 ± 0.01	$\mathbf{1.00 \pm 0}$	0.99 ± 0.02	0.47 ± 0.27	0.25 ± 0.38
GESW	0.91 ± 0.06	$\mathbf{1.00 \pm 0}$	0.91 ± 0.07	0.98 ± 0.01	0.37 ± 0.32	0.24 ± 0.38
CVaR USW	$\mathbf{1.00 \pm 0}$	0.98 ± 0.02	$\mathbf{1.00 \pm 0}$	0.99 ± 0.01	0.52 ± 0.25	0.27 ± 0.37
CVaR GESW	0.92 ± 0.05	0.98 ± 0.02	0.92 ± 0.06	$\mathbf{1.00 \pm 0}$	0.41 ± 0.31	0.28 ± 0.37
Rob. USW	0.84 ± 0.08	0.77 ± 0.12	0.86 ± 0.07	0.84 ± 0.09	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0}$
Rob. GESW	0.73 ± 0.14	0.76 ± 0.13	0.74 ± 0.13	0.84 ± 0.09	0.85 ± 0.09	$\mathbf{1.00 \pm 0}$

Table 7: Performance of different allocations across each metric on the Gaussian AAMAS 2021 dataset.

Allocation	Evaluation Objective					
	USW	GESW	CVaR USW	CVaR GESW	Rob. USW	Rob. GESW
USW	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0.01}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0.01}$	0.53 ± 0.26	0.21 ± 0.40
GESW	0.80 ± 0.12	$\mathbf{1.00 \pm 0}$	0.79 ± 0.12	0.99 ± 0.01	0.24 ± 0.39	0.20 ± 0.40
CVaR USW	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0.01}$	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0.01}$	0.53 ± 0.26	0.21 ± 0.40
CVaR GESW	0.85 ± 0.08	$\mathbf{1.00 \pm 0}$	0.84 ± 0.08	$\mathbf{1.00 \pm 0}$	0.36 ± 0.34	0.20 ± 0.40
Rob. USW	0.81 ± 0.11	0.69 ± 0.16	0.81 ± 0.11	0.71 ± 0.16	$\mathbf{1.00 \pm 0}$	$\mathbf{1.00 \pm 0.01}$
Rob. GESW	0.70 ± 0.17	0.68 ± 0.17	0.71 ± 0.17	0.70 ± 0.16	0.88 ± 0.10	$\mathbf{1.00 \pm 0}$

H Proofs

H.1 Proof of Proposition 3.1

Proposition 3.1 (Robust Utilitarian Welfare Dual)**.** *The problem in* (1) *is equivalent to solving*

$$\max_{\substack{\boldsymbol{\xi}\in\Lambda,\boldsymbol{\lambda}\in\mathbb{R}^{l}_{0+}\\ \boldsymbol{\beta}\in\mathbb{R}^{k}_{0+}}} \mathbf{p}^{\intercal}\mathbf{T}\mathbf{q}+\boldsymbol{\beta}^{\intercal}\mathbf{e}-\frac{1}{4}\mathbf{p}^{\intercal}\mathbf{T}\mathbf{p}+\sum_{i=1}^{\ell}\Big(\lambda_i\bar{\mathbf{v}}_i^{\intercal}\mathbf{S}_i\bar{\mathbf{v}}_i-\lambda_i r_i^2\Big)-\mathbf{q}^{\intercal}\mathbf{T}\mathbf{q}\ , \tag{3}$$

where $\mathbf{p}=\boldsymbol{\xi}-\boldsymbol{Q}^{\intercal}\boldsymbol{\beta}$*,* $\mathbf{q}=\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i$*, and* $\mathbf{T}=\big(\sum_{i=1}^{\ell}\boldsymbol{\lambda}_i\mathbf{S}_i^{-1}\big)^{-1}$*. Let* $\boldsymbol{\xi}^*$ *be the optimal* $\boldsymbol{\xi}$ *in* (3)*. Then the optimal allocation* $\mathbf{a}^*$ *can be derived from* $\boldsymbol{\xi}^*$ *by finding* $\mathbf{a}\in\mathcal{A}$ *such that* $\mathbf{a}\preceq\boldsymbol{\xi}^*$*.*

Proof. Consider the inner-minimization problem

$$\begin{aligned}\min_{\mathbf{v}\in\mathbb{R}^{nm}}\ &\sum_{G\in\mathcal{G}}\mathbf{a}_G^{\intercal}\mathbf{v}_G\\ \forall i\in[1,l]:&(\mathbf{v}-\bar{\mathbf{v}}_i)^{\intercal}\mathbf{S}_i^{-1}(\mathbf{v}-\bar{\mathbf{v}}_i)\le r_i^2\\ &\boldsymbol{Q}\mathbf{v}\succeq\mathbf{e}\\ &\mathbf{v}\succeq 0\ .\end{aligned}$$

Note that the above optimization problem is convex as the objective is an affine combination of $\mathbf{v}$, which is convex, and the linear and quadratic constraints are also convex. Thus, from the theory of convex optimization (Section 5.1 in Boyd *et al.* [12]), we know that maximizing the dual of a convex optimization problem is equivalent to minimizing its primal counterpart. We will therefore use the Lagrangian method for computing the dual of the above problem.

We will use $\boldsymbol{\beta}\in\mathbb{R}^{k}_{0+}$ to represent the dual variable corresponding to the linear constraints $\boldsymbol{Q}\mathbf{v}\succeq\mathbf{e}$, $\boldsymbol{\lambda}\in\mathbb{R}^{\ell}_{0+}$ to represent the dual variable associated with the ellipsoidal constraints, and $\boldsymbol{\zeta}\in\mathbb{R}^{nm}$ to represent the dual variable corresponding to the non-negativity constraint on $\mathbf{v}$. The Lagrangian for the above problem is given by

$$L(\mathbf{v},\boldsymbol{\lambda},\boldsymbol{\beta},\boldsymbol{\zeta})=\mathbf{a}^{\intercal}\mathbf{v}+\sum_{i=1}^{\ell}\lambda_i\big((\mathbf{v}-\bar{\mathbf{v}}_i)^{\intercal}\mathbf{S}_i^{-1}(\mathbf{v}-\bar{\mathbf{v}}_i)-r_i^2\big)-\boldsymbol{\beta}^{\intercal}(Q\mathbf{v}-\mathbf{e})-\boldsymbol{\zeta}^{\intercal}\mathbf{v}\ . \tag{13}$$

From the first-order optimality conditions, we get

$$\begin{aligned}&\frac{\partial L(\mathbf{v},\boldsymbol{\lambda},\boldsymbol{\beta},\boldsymbol{\zeta})}{\partial\mathbf{v}}=0\\ &\mathbf{a}+\sum_{i=1}^{\ell}2\lambda_i\mathbf{S}_i^{-1}(\mathbf{v}-\bar{\mathbf{v}}_i)-\boldsymbol{Q}^{\intercal}\boldsymbol{\beta}-\boldsymbol{\zeta}=0\\ \implies\mathbf{v}=&\left(\sum_{i=1}^{\ell}2\lambda_i\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i-(\mathbf{a}-\boldsymbol{Q}^{\intercal}\boldsymbol{\beta}-\boldsymbol{\zeta})\right)\left(\sum_{i=1}^{\ell}2\lambda_i\mathbf{S}_i^{-1}\right)^{-1}\ .\end{aligned}$$

Substituting this value of $\mathbf{v}$ in (13), we get

$$\begin{aligned}\max_{\substack{\boldsymbol{\lambda}\in\mathbb{R}^{\ell}_{0+}\\ \boldsymbol{\beta}\in\mathbb{R}^{nm}_{0+}\\ \boldsymbol{\zeta}\in\mathbb{R}^{nm}_{0+}}}\ &-\frac{1}{4}\big((\mathbf{a}-\boldsymbol{Q}^{\intercal}\boldsymbol{\beta}-\boldsymbol{\zeta})^{\intercal}\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\big)^{-1}\big(\mathbf{a}-\boldsymbol{Q}^{\intercal}\boldsymbol{\beta}-\boldsymbol{\zeta}\big)\big)+\sum_{i=1}^{\ell}\lambda_i\bar{\mathbf{v}}_i^{\intercal}\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i\\ &-\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i\big)^{\intercal}\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\big)^{-1}\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i\big)\\ &+(\mathbf{a}-\boldsymbol{Q}^{\intercal}\boldsymbol{\beta}-\boldsymbol{\zeta})^{\intercal}\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\big)^{-1}\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i\big)-\sum_{i=1}^{\ell}\lambda_i r_i^2+\boldsymbol{\beta}^{\intercal}\mathbf{e}\ .\end{aligned}$$

Finally, we substitute the above dual problem back into the original problem in (1) to get

$$\begin{aligned}\max_{\substack{\mathbf{a}\in\mathcal{A}\\ \boldsymbol{\lambda}\in\mathbb{R}^{\ell}_{0+}\\ \boldsymbol{\beta}\in\mathbb{R}^{nm}_{0+}\\ \boldsymbol{\zeta}\in\mathbb{R}^{nm}_{0+}}} & -\frac{1}{4}\big((\mathbf{a}-\boldsymbol{Q}^{\intercal}\boldsymbol{\beta}-\boldsymbol{\zeta})^{\intercal}\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\big)^{-1}(\mathbf{a}-\boldsymbol{Q}^{\intercal}\boldsymbol{\beta}-\boldsymbol{\zeta})\big)+\sum_{i=1}^{\ell}\lambda_i\bar{\mathbf{v}}_i^{\intercal}\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i\\ & -\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i\big)^{\intercal}\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\big)^{-1}\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i\big)\\ & +(\mathbf{a}-\boldsymbol{Q}^{\intercal}\boldsymbol{\beta}-\boldsymbol{\zeta})^{\intercal}\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\big)^{-1}\big(\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i\big)-\sum_{i=1}^{\ell}\lambda_i r_i^2+\boldsymbol{\beta}^{\intercal}\mathbf{e}\ .\end{aligned}$$

Note that the dual problem is concave in $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, and $\boldsymbol{\zeta}$ (Section 5.1 in [12]). However, it is unclear if the dual is concave in allocation $\mathbf{a}$. In order to guarantee concavity, we use the change of variables $\boldsymbol{\xi}=\mathbf{a}-\boldsymbol{\zeta}$. From affine-composition rule in convex optimization (Section 3.2.2 in [12]), we know that if $f(\boldsymbol{x})$ $\boldsymbol{x}\in\mathbb{R}^n$ is convex, then $f(\boldsymbol{A}\boldsymbol{x}+\boldsymbol{b})$ $\boldsymbol{A}\in\mathbb{R}^{n\times n}, \boldsymbol{b}\in\mathbb{R}^n$ is also convex in $\boldsymbol{x}$. Thus, the variable change $\boldsymbol{\xi}=\mathbf{a}-\boldsymbol{\zeta}$ results in a objective that is concave in $\boldsymbol{\xi}$, $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$. The allocation variable $\mathbf{a}$ and the dual variable $\boldsymbol{\zeta}$ only appear in a linear constraint, which is also concave. Thus, the objective and the constraints are concave in $\mathbf{a}$, $\boldsymbol{\beta}$, $\boldsymbol{\lambda}$, and $\boldsymbol{\xi}$, and $\boldsymbol{\zeta}$.

$$\begin{aligned}\max_{\substack{\boldsymbol{\xi}\in\Lambda,\boldsymbol{\lambda}\in\mathbb{R}^{l}_{0+}\\ \boldsymbol{\beta}\in\mathbb{R}^{k}_{0+}\\ \mathbf{a}\in\mathcal{A},\boldsymbol{\zeta}\in\mathbb{R}^{mn}_{0+}}} & \ \mathbf{p}^{\intercal}\mathbf{T}\mathbf{q}+\boldsymbol{\beta}^{\intercal}\mathbf{e}-\frac{1}{4}\mathbf{p}^{\intercal}\mathbf{T}\mathbf{p}+\sum_{i=1}^{\ell}\Big(\lambda_i\bar{\mathbf{v}}_i^{\intercal}\mathbf{S}_i\bar{\mathbf{v}}_i-\lambda_i r_i^2\Big)-\mathbf{q}^{\intercal}\mathbf{T}\mathbf{q}\ ,\\ \text{s.t. } & \boldsymbol{\xi}=\mathbf{a}-\boldsymbol{\zeta}\ ,\end{aligned}\tag{14}$$

where $\mathbf{p}=\boldsymbol{Q}^{\intercal}\boldsymbol{\beta}+\boldsymbol{\xi}$, $\mathbf{T}=\big(\sum_{i=1}^{\ell}\boldsymbol{\lambda}_i\mathbf{S}_i^{-1}\big)^{-1}$, $\mathbf{q}=\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i$, and $\Lambda=\mathcal{A}-\mathbb{R}^{nm}_{0+}=\{\boldsymbol{\xi}\in\mathbb{R}^{nm}\mid\forall a\in N:\sum_{i\in I}\boldsymbol{\xi}_{am+i}\le\bar{\kappa}_a,\forall i\in I:\sum_{a\in N}\boldsymbol{\xi}_{am+i}\le\bar{\psi}_i,\boldsymbol{\xi}\preceq\boldsymbol{c}\}$.

We can solve the optimization problem in (14) using standard convex optimization techniques [12].

Alternatively, we can further simplify the problem by eliminating the allocation variable $\mathbf{a}$ and the dual variable $\boldsymbol{\zeta}$ and subsequently deriving them from the solution of the resultant problem.

Note that in (14), $\mathbf{a}-\boldsymbol{\zeta}=\boldsymbol{\xi}$. Let $(\mathbf{a}^*,\boldsymbol{\zeta}^*)$ represent an optimal $(\mathbf{a},\boldsymbol{\zeta})$ pair for the problem in (14). Now there can be multiple pairs of $(\mathbf{a},\boldsymbol{\zeta})$ that are optimal. To eliminate $\boldsymbol{\zeta}$ and $\mathbf{a}$, we need to first ensure that there exists a $\boldsymbol{\xi}'\in\Lambda$ such that $\boldsymbol{\xi}'=\mathbf{a}^*-\boldsymbol{\zeta}^*$ for at least one optimal pair $(\mathbf{a}^*,\boldsymbol{\zeta}^*)$. It is easy to see that if there exists such a $\boldsymbol{\xi}'\in\Lambda$, then, $\boldsymbol{\xi}'$ maximizes the objective in (14). Furthermore, since Λ is defined solely by upper-bound constraints on $\boldsymbol{\xi}$ imposed by $\mathcal{A}$, we can easily verify that it will contain at least one instance of $\boldsymbol{\xi}'$ that satisfies $\boldsymbol{\xi}'=\mathbf{a}^*-\boldsymbol{\zeta}^*$.

Thus, we can break down the problem in (14) into two sub-problems. In the first problem, we obtain the optimal value of $\boldsymbol{\lambda}$, $\boldsymbol{\xi}$ and $\boldsymbol{\beta}$ by solving

$$\begin{aligned}\boldsymbol{\zeta}^*,\boldsymbol{\beta}^*,\boldsymbol{\xi}^*=\operatorname*{arg\,max}_{\substack{\boldsymbol{\zeta}\in\mathbb{R}^{nm}_{0+}\\ \boldsymbol{\beta}\in\mathbb{R}^{k}_{0+}\\ \boldsymbol{\xi}\in\Lambda}} & \ \mathbf{p}^{\intercal}\mathbf{T}\mathbf{q}^{\intercal}+\boldsymbol{\beta}^{\intercal}\mathbf{e}-\frac{1}{4}\mathbf{p}^{\intercal}\mathbf{T}\mathbf{p}+\sum_{i=1}^{\ell}\Big(\lambda_i\bar{\mathbf{v}}_i^{\intercal}\mathbf{S}_i\bar{\mathbf{v}}_i-\lambda_i r_i^2\Big)-\mathbf{q}^{\intercal}\mathbf{T}\mathbf{q}+\mathbf{p}^{\intercal}\mathbf{T}^{-1}\mathbf{q}^{\intercal}\\ & -\sum_{i=1}^{\ell}\lambda_i r_i^2+\boldsymbol{\beta}^{\intercal}\mathbf{e}\ ,\end{aligned}$$

where $\mathbf{p}=-\boldsymbol{Q}^{\intercal}\boldsymbol{\beta}+\boldsymbol{\xi}$, $\mathbf{q}=\sum_{i=1}^{\ell}\lambda_i\mathbf{S}_i^{-1}\bar{\mathbf{v}}_i$, and $\mathbf{T}=\big(\sum_{i=1}^{\ell}\boldsymbol{\lambda}_i\mathbf{S}_i^{-1}\big)^{-1}$. Then, we can compute the set of optimal $(\mathbf{a},\boldsymbol{\zeta})$ pairs by solving the system of equations: $\{(\mathbf{a}^*,\boldsymbol{\zeta}^*)\mid\mathbf{a}\in\mathcal{A},G\in\mathcal{G}:\mathbf{a}_G-\boldsymbol{\zeta}_G=\boldsymbol{\xi}^*_G,\boldsymbol{\zeta}\in\mathbb{R}^{nm}_{0+}\}$. □

H.2 Proof of Corollary 3.2

Corollary 3.2 (Utilitarian Welfare with Polyhedral Uncertainty)**.** *In the case where the uncertainty set $\mathcal{V}$ is defined purely by linear constraints, i.e., $\mathcal{V}=\{\mathbf{v}\in\mathbb{R}^{nm}\mid\boldsymbol{Q}\mathbf{v}\succeq\mathbf{e},\mathbf{v}\succeq 0\}$, the optimal allocation $\mathbf{a}^*$ for the problem in* (1) *can be computed by solving the linear program*

$$\max_{\mathbf{a}\in\mathcal{A},\boldsymbol{\beta}\in\mathbb{R}^{k}_{0+}}\ \boldsymbol{\beta}^{\intercal}\mathbf{e}\quad \textit{s.t. } \boldsymbol{Q}^{\intercal}\boldsymbol{\beta}\preceq\mathbf{a}\ .$$

Proof. Now consider the inner-minimization problem:

$$\begin{aligned}\min_{\mathbf{v}\in\mathbb{R}^{nm}} \sum_{G\in\mathcal{G}} \mathbf{a}_G^\intercal \mathbf{v}_G \\ \boldsymbol{Q}\mathbf{v} \succeq \mathbf{e} \\ \mathbf{v} \succeq 0 \ .\end{aligned}$$

We compute the dual of the above problem using the Lagrangian method. We will use $\boldsymbol{\beta} \in \mathbb{R}^k_{0+}$ to represent the dual variable corresponding to the linear constraints $\boldsymbol{Q}\mathbf{v} \succeq \mathbf{e}$ and $\zeta \in \mathbb{R}^{nm}$ to represent the dual variable corresponding to the non-negativity constraint on $\mathbf{v}$.

$$\begin{aligned}L(\mathbf{v}, \boldsymbol{\beta}, \boldsymbol{\zeta}) &= \mathbf{a}^\intercal \mathbf{v} - \boldsymbol{\beta}^\intercal(\boldsymbol{Q}\mathbf{v} - \mathbf{e}) - \boldsymbol{\zeta}^\intercal \mathbf{v} \\ &= (\mathbf{a} - \boldsymbol{Q}^\intercal \boldsymbol{\beta} - \boldsymbol{\zeta})^\intercal \mathbf{v} + \boldsymbol{\beta}^\intercal \mathbf{e} \\ L(\boldsymbol{\beta}, \boldsymbol{\zeta}) &= \begin{cases} \boldsymbol{\beta}^\intercal \mathbf{e} & (\mathbf{a} - \boldsymbol{Q}^\intercal \boldsymbol{\beta} - \boldsymbol{\zeta}) \succeq 0 \\ -\infty & \text{otherwise} \end{cases} \ .\end{aligned}$$

Therefore, the dual is given by

$$\begin{aligned}\max_{\boldsymbol{\beta}\in\mathbb{R}^k_{0+}, \boldsymbol{\zeta}\in\mathbb{R}^{nm}_{0+}} \boldsymbol{\beta}^\intercal \mathbf{e} \\ \boldsymbol{Q}^\intercal \boldsymbol{\beta} - \boldsymbol{\zeta} \preceq \mathbf{a} \ .\end{aligned}$$

Since $\boldsymbol{\zeta}$ is non-negative, we can eliminate it to get

$$\begin{aligned}\max_{\boldsymbol{\beta}\in\mathbb{R}^k_{0+}} \boldsymbol{\beta}^\intercal \mathbf{e} \\ \boldsymbol{Q}^\intercal \boldsymbol{\beta} \preceq \mathbf{a} \ .\end{aligned}$$

By combining the dual with the outer-maximization problem in (1), we obtain the final result.

□

H.3 Proof of Corollary 3.3

Corollary 3.3 (Utilitarian Welfare with Ellipsoidal Uncertainty)**.** *Suppose that the set $\mathcal{V}$ in* (1) *is defined by a single truncated ellipsoidal constraint i.e., $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^{nm} \mid (\mathbf{v} - \bar{\mathbf{v}})^\intercal \mathbf{S}^{-1}(\mathbf{v} - \bar{\mathbf{v}}) \leq r^2, \mathbf{v} \succeq 0\}$. The problem in* (1) *is equivalent to solving*

$$\max_{\lambda\in\mathbb{R}_{0+}, \boldsymbol{\xi}\in\Lambda} \boldsymbol{\xi}^\intercal \bar{\mathbf{v}} - \frac{\boldsymbol{\xi}^\intercal \mathbf{S} \boldsymbol{\xi}}{4\lambda} - \lambda r^2 \ . \tag{4}$$

The exact optimal solution $(\lambda^*, \boldsymbol{\xi}^*)$ to Equation (4) *can be computed by alternately performing two steps until convergence: first, fixing $\boldsymbol{\xi}$ and optimizing λ, i.e., $\lambda = \boldsymbol{\xi}^\intercal \mathbf{S} \boldsymbol{\xi}/2r$, and second, fixing λ and solving a concave quadratic program to optimize $\boldsymbol{\xi}$. The optimal allocation $\mathbf{a}^*$ can be computed from $\boldsymbol{\xi}^*$ as in Proposition 3.1.*

Proof. Consider the inner-minimization problem

$$\begin{aligned}\min_{\mathbf{v}\in\mathbb{R}^{nm}} \sum_{G\in\mathcal{G}} \mathbf{a}_G^\intercal \mathbf{v}_G \\ (\mathbf{v} - \bar{\mathbf{v}})^\intercal \mathbf{S}^{-1}(\mathbf{v} - \bar{\mathbf{v}}) \leq r^2 \\ \mathbf{v} \succeq 0 \ .\end{aligned}$$

Similar to the approach in Proposition 3.1, we will use the Lagrangian method for computing the dual of the above problem. We will use $\lambda \in \mathbb{R}_{0+}$ to represent the dual variable associated with the ellipsoidal constraint, and $\zeta \in \mathbb{R}^{nm}$ to represent the dual variable corresponding to the non-negativity constraint on $\mathbf{v}$. The Lagrangian for the above problem is given by

$$L(\mathbf{v}, \lambda, \boldsymbol{\zeta}) = \mathbf{a}^\intercal \mathbf{v} + \lambda \left((\mathbf{v} - \bar{\mathbf{v}})^\intercal \mathbf{S}^{-1}(\mathbf{v} - \bar{\mathbf{v}}) - r^2 \right) - \boldsymbol{\zeta}^\intercal \mathbf{v} \ . \tag{15}$$

From the first-order optimality conditions, we get

$$\frac{\partial L(\mathbf{v}, \lambda, \boldsymbol{\zeta})}{\partial \mathbf{v}} = 0$$
$$\mathbf{a} + 2\lambda \mathbf{S}^{-1}(\mathbf{v} - \bar{\mathbf{v}}) - \boldsymbol{\zeta} = 0$$
$$\implies \mathbf{v} = \Big(2\lambda \mathbf{S}^{-1}\bar{\mathbf{v}} - (\mathbf{a} - \boldsymbol{\zeta})\Big)\left(\frac{\mathbf{S}}{2\lambda}\right) \ .$$

Substituting the value of $\mathbf{v}$ in the above equation in (15), we get

$$\max_{\substack{\lambda \in \mathbb{R}_{0+} \\ \boldsymbol{\zeta} \in \mathbb{R}_{0+}^{nm}}} -\frac{1}{4}\left(\left(\mathbf{a} - \boldsymbol{\zeta}\right)^{\intercal} \frac{\mathbf{S}}{\lambda}\left(\mathbf{a} - \boldsymbol{\zeta}\right)\right) + (\mathbf{a} - \boldsymbol{\zeta})^{\intercal}\bar{\mathbf{v}} - \boldsymbol{\lambda} r^2 \ .$$

From the theory of convex optimization [12], we know that the dual of a convex optimization problem is always concave, and therefore, the above optimization problem is concave in $\boldsymbol{\lambda}$ and $\boldsymbol{\zeta}$. Combining the dual with the outer-maximization problem in (1), we get

$$\max_{\mathbf{a} \in \mathcal{A}} \max_{\substack{\lambda \in \mathbb{R}_{0+}^{\ell} \\ \boldsymbol{\zeta} \in \mathbb{R}_{0+}^{nm}}} -\frac{1}{4}\left(\left(\mathbf{a} - \boldsymbol{\zeta}\right)^{\intercal} \frac{\mathbf{S}}{\lambda}\left(\mathbf{a} - \boldsymbol{\zeta}\right)\right) + (\mathbf{a} - \boldsymbol{\zeta})^{\intercal}\bar{\mathbf{v}} - \boldsymbol{\lambda} r^2 \ .$$

To further obtain concavity in allocation $\mathbf{a}$, we follow the same procedure as in Proposition 3.1 and use the change of variables $\boldsymbol{\xi} = \mathbf{a} - \boldsymbol{\zeta}$

$$\max_{\substack{\mathbf{a} \in \mathcal{A} \\ \lambda \in \mathbb{R}_{0+} \\ \boldsymbol{\xi} \in \Lambda}} \boldsymbol{\xi}^{\intercal}\bar{\mathbf{v}} - \frac{\boldsymbol{\xi}^{\intercal}\mathbf{S}\boldsymbol{\xi}}{4\lambda} - \lambda r^2 \ . \qquad \text{s.t. } \boldsymbol{\xi} = \mathbf{a} - \boldsymbol{\zeta} \ . \tag{16}$$

As a result of the change of variables, the allocation variable $\mathbf{a}$ and the dual variable $\boldsymbol{\zeta}$ now appear only in a linear constraint, which is convex. Furthermore, due to the affine composition property of convex functions (Section 3.2.2 in Boyd *et al.* [12]), the objective remains retains its concavity. Thus, the above optimization problem is concave in $\mathbf{a}$, $\boldsymbol{\lambda}$, $\boldsymbol{\zeta}$, and $\boldsymbol{\xi}$.

Similar to the approach used in the proof of Proposition 3.1, we further simplify the problem by eliminating the allocation variables $\mathbf{a}$ and the dual variable $\boldsymbol{\zeta}$ and subsequently deriving them from the solution of the resultant problem.

From (16), we know that that $\mathbf{a} - \boldsymbol{\zeta} = \boldsymbol{\xi}$. Let $(\mathbf{a}^*, \boldsymbol{\zeta}^*)$ represent an optimal $(\mathbf{a}, \boldsymbol{\zeta})$ pair for the problem in (14). Note that there can be multiple pairs of $(\mathbf{a}, \boldsymbol{\zeta})$ that are optimal. To eliminate $\boldsymbol{\zeta}$ and $\mathbf{a}$, we need to find a set of feasible $\boldsymbol{\xi}$, which we denote by Λ, such that there exists a $\boldsymbol{\xi}' \in \Lambda$ such that $\boldsymbol{\xi}' = \mathbf{a}^* - \boldsymbol{\zeta}^*$ for at least one optimal pair $(\mathbf{a}^*, \boldsymbol{\zeta}^*)$. It is easy to see that if there exists such a $\boldsymbol{\xi}' \in \Lambda$, then, $\boldsymbol{\xi}'$ maximizes the objective in (16). Furthermore, it easy to verify that Λ satisfies this criteria for optimality, i.e., it contains at least one $\boldsymbol{\xi}'$ that satisfies $\boldsymbol{\xi}' = \mathbf{a}^* - \boldsymbol{\zeta}^*$, for some optimal pair $(\mathbf{a}^*, \boldsymbol{\zeta}^*)$.

Thus, we can efficiently solve the problem in (16) in two steps. First, we obtain the optimal value of λ and $\boldsymbol{\xi}$ by solving the problem

$$\lambda^*, \boldsymbol{\xi}^* = \arg\max_{\substack{\lambda \in \mathbb{R}_{0+} \\ \boldsymbol{\xi} \in \Lambda}} \boldsymbol{\xi}^{\intercal}\bar{\mathbf{v}} - \frac{\boldsymbol{\xi}^{\intercal}\mathbf{S}\boldsymbol{\xi}}{4\lambda} - \lambda r^2 \ .$$

As in Proposition 3.1, we can compute the set of optimal $(\mathbf{a}, \boldsymbol{\zeta})$ pairs by solving the system of equations: $\{(\mathbf{a}^*, \boldsymbol{\zeta}^*) \mid \mathbf{a} \in \mathcal{A},\, G \in \mathcal{G} : \mathbf{a}_G - \boldsymbol{\zeta}_G = \boldsymbol{\xi}_G^*, \boldsymbol{\zeta} \in \mathbb{R}_{0+}^{nm}\}$.

□

H.4 Proof of Proposition 3.5

Proposition 3.5 (Robust Group Egalitarian Dual)**.** *The problem in* (5) *is equivalent to solving*

$$\max_{\substack{\boldsymbol{\xi} \in \Lambda \\ \boldsymbol{\lambda} \in \mathbb{R}_{0+}^{g \times l} \\ \boldsymbol{\beta} \in \mathbb{R}_{0+}^{g \times k}}} \min_{G \in \mathcal{G}} \boldsymbol{\beta}_G^{\intercal}\mathbf{e}_G + \mathbf{p}_G^{\intercal}\mathbf{T}\mathbf{q}_G - \frac{1}{4}\mathbf{p}_G^{\intercal}\mathbf{T}\mathbf{p}_G + \sum_{i=1}^{\ell}\left(\lambda_{G,i}\bar{\mathbf{v}}_{G,i}^{\intercal}\mathbf{T}\bar{\mathbf{v}}_{G,i} - \lambda_{G,i} r_{G,i}^2\right) - \mathbf{q}_G^{\intercal}\mathbf{T}\mathbf{q}_G \ , \tag{7}$$

where for each group $G \in \mathcal{G}$, $\mathbf{p}_G = \boldsymbol{\xi}_G - \boldsymbol{Q}_G^\intercal \boldsymbol{\beta}_G$, $\mathbf{q}_G = \sum_{i=1}^{\ell} \lambda_{G,i} \mathbf{S}_{G,i}^{-1} \bar{\mathbf{v}}_{G,i}$, $\mathbf{T} = \left(\sum_{i=1}^{\ell}, \lambda_{G,i} \mathbf{S}_{G,i}^{-1}\right)^{-1}$, *and* Λ *is defined as in Equation* (2). *The optimal allocation* $\mathbf{a}^*$ *can be computed from* $\boldsymbol{\xi}^*$ *as in Proposition 3.1.*

Proof. Consider the following optimization problem.

$$\begin{aligned}
\max_{\mathbf{a} \in \mathcal{A}} \min_{G \in \mathcal{G}} \min_{\mathbf{v}_G \in \mathcal{V}_G} \; & \mathbf{a}_G^\intercal \mathbf{v}_G \\
& \forall i \in [1, l] : \forall G \in \mathcal{G} : \; (\mathbf{v}_G - \bar{\mathbf{v}}_{i,G})^\intercal \mathbf{S}_{i,G}^{-1} (\mathbf{v}_G - \bar{\mathbf{v}}_{i,G}) \leq r_{i,G}^2 \\
& \forall G \in \mathcal{G} : \; Q_G \mathbf{v}_G \succeq \mathbf{e}_G \\
& \mathbf{v}_G \succeq 0 \;\; .
\end{aligned} \tag{17}$$

It is important to note that the inner-most minimization is a convex optimization problem and the outer-maximization is a concave maximization problem. This is due to the fact that affine functions are either concave or convex and minimum of concave objectives is concave.

Notice that the inner-most minimization problem for each group is independent of other groups. Thus, we can simply replace each of these minimization problems with their Lagrangian dual counterparts. Furthermore, we note that these duals are computed following the approach outlined in the proof of Proposition 3.1 and are exact equivalents of their respective primal problems [12]. The resultant optimization problem is given by

$$\begin{aligned}
\max_{\mathbf{a} \in \mathcal{A}} \min_{G \in \mathcal{G}} \max_{\substack{\boldsymbol{\lambda}_G \in \mathbb{R}_{0+}^{\ell} \\ \boldsymbol{\beta}_G \in \mathbb{R}_{0+}^{k} \\ \boldsymbol{\zeta}_G \in \mathbb{R}_{0+}^{|G|m}}} \; & -\frac{1}{4} \left(\left(\mathbf{a}_G - \boldsymbol{Q}_G^\intercal \boldsymbol{\beta}_G - \boldsymbol{\zeta}_G\right)^\intercal \left(\sum_{i=1}^{\ell} \lambda_i \mathbf{S}_{G,i}^{-1} \right)^{-1} \left(\mathbf{a}_G - \boldsymbol{Q}_G^\intercal \boldsymbol{\beta}_G - \boldsymbol{\zeta}_G\right) \right) + \\
& \sum_{i=1}^{\ell} \lambda_{G,i} \bar{\mathbf{v}}_{G,i}^\intercal \mathbf{S}_{G,i}^{-1} \bar{\mathbf{v}}_{G,i} - \left(\sum_{i=1}^{\ell} \lambda_{G,i} \mathbf{S}_{G,i}^{-1} \bar{\mathbf{v}}_{G,i} \right)^\intercal \left(\sum_{i=1}^{\ell} \lambda_{G,i} \mathbf{S}_{G,i}^{-1} \right)^{-1} \left(\sum_{i=1}^{\ell} \lambda_{G,i} \mathbf{S}_{G,i}^{-1} \bar{\mathbf{v}}_{G,i} \right) \\
& + \left(\mathbf{a}_G - \boldsymbol{Q}_G^\intercal \boldsymbol{\beta}_G - \boldsymbol{\zeta}_G\right)^\intercal \left(\sum_{i=1}^{\ell} \lambda_{G,i} \mathbf{S}_{G,i}^{-1} \right)^{-1} \left(\sum_{i=1}^{\ell} \lambda_{G,i} \mathbf{S}_{G,i}^{-1} \bar{\mathbf{v}}_{G,i} \right) \\
& - \sum_{i=1}^{\ell} \lambda_{G,i} r_{G,i}^2 + \boldsymbol{\beta}_G^\intercal \mathbf{e}_G \;\; .
\end{aligned}$$

Notice that the inner-most maximization problem is concave in $\boldsymbol{\lambda}$, $\boldsymbol{\beta}$, and $\boldsymbol{\zeta}$. We now apply the same procedure as in the proof of Proposition 3.1 to obtain concavity in allocation $\mathbf{a}$. Using the change of variables $\forall G \in \mathcal{G} : \boldsymbol{\xi}_G = \mathbf{a}_G - \boldsymbol{\zeta}_G$, we get

$$\begin{aligned}
\max_{\mathbf{a} \in \mathcal{A}} \min_{G \in \mathcal{G}} \max_{\substack{\boldsymbol{\lambda}_G \in \mathbb{R}^{l} \\ \boldsymbol{\beta}_G \in \mathbb{R}_{0+}^{k} \\ \boldsymbol{\zeta}_G \in \mathbb{R}_{0+}^{|G|m} \\ \boldsymbol{\xi}_G \in \mathbb{R}^{|G|m}}} \; & \boldsymbol{\beta}_G^\intercal \mathbf{e}_G + \mathbf{p}_G^\intercal \mathbf{T} \mathbf{q}_G - \frac{1}{4} \mathbf{p}_G^\intercal \mathbf{T} \mathbf{p}_G + \sum_{i=1}^{\ell} \left(\lambda_{G,i} \bar{\mathbf{v}}_{G,i}^\intercal \mathbf{T} \bar{\mathbf{v}}_{G,i} - \lambda_{G,i} r_{G,i}^2 \right) - \mathbf{q}_G^\intercal \mathbf{T} \mathbf{q}_G \\
& \text{s.t. } \boldsymbol{\xi}_G = \mathbf{a}_G - \boldsymbol{\zeta}_G \;\; ,
\end{aligned}$$

where for any $G \in \mathcal{G}, \mathbf{p}_G = \boldsymbol{Q}_G^\intercal \boldsymbol{\beta}_G + \boldsymbol{\xi}_G$, $\mathbf{T} = \left(\sum_{i=1}^{\ell} \lambda_{G,i} \mathbf{S}_{G,i}^{-1}\right)^{-1}$, and $\mathbf{q}_G = \sum_{i=1}^{\ell} \lambda_{G,i} \mathbf{S}_{G,i}^{-1} \bar{\mathbf{v}}_{G,i}$.

Since the inner maximization for each group is independent of the other groups, we can re-order the inner minimization over groups and the inner-maximization problem. Thus, without loss of generality, we can write the above optimization problem as

$$\begin{aligned}
\max_{\substack{\mathbf{a} \in \mathcal{A} \\ \boldsymbol{\zeta} \in \mathbb{R}^{nm} \\ \boldsymbol{\lambda} \in \mathbb{R}_{0+}^{g \times l} \\ \boldsymbol{\beta} \in \mathbb{R}_{0+}^{g \times k} \\ \boldsymbol{\xi} \in \mathbb{R}^{nm}}} \min_{G \in \mathcal{G}} \; & \boldsymbol{\beta}_G^\intercal \mathbf{e}_G + \mathbf{p}_G^\intercal \mathbf{T} \mathbf{q}_G - \frac{1}{4} \mathbf{p}_G^\intercal \mathbf{T} \mathbf{p}_G + \sum_{i=1}^{\ell} \left(\lambda_{G,i} \bar{\mathbf{v}}_{G,i}^\intercal \mathbf{T} \bar{\mathbf{v}}_{G,i} - \lambda_{G,i} r_{G,i}^2 \right) - \mathbf{q}_G^\intercal \mathbf{T} \mathbf{q}_G \\
& \text{s.t. } \boldsymbol{\xi}_G = \mathbf{a}_G - \boldsymbol{\zeta}_G \;\; .
\end{aligned}$$

Using the same technique as in Appendix H.1, we can simplify the problem by eliminating the variables $\mathbf{a}$ and $\boldsymbol{\zeta}$ in the above problem and then derive them from the optimal $\boldsymbol{\xi}$.

Eliminating $\boldsymbol{\zeta}$ and $\mathbf{a}$ in the above equation, we get

$$\boldsymbol{\lambda}^*, \boldsymbol{\beta}^*, \boldsymbol{\xi}^* = \underset{\substack{,\boldsymbol{\lambda}\in\mathbb{R}^{g\times l}_{0+},\\ \boldsymbol{\beta}\in\mathbb{R}^{g\times k}_{0+}\\ \boldsymbol{\xi}\in\Lambda}}{\arg\max} \min_{G\in\mathcal{G}} \boldsymbol{\beta}_G^\intercal \mathbf{e}_G + \mathbf{p}_G^\intercal \mathbf{T}\mathbf{q}_G - \frac{1}{4}\mathbf{p}_G^\intercal \mathbf{T}\mathbf{p}_G + \sum_{i=1}^{\ell}\left(\lambda_{G,i}\bar{\mathbf{v}}_{G,i}^\intercal \mathbf{T}\bar{\mathbf{v}}_{G,i} - \lambda_{G,i} r_{G,i}^2\right) - \mathbf{q}_G^\intercal \mathbf{T}\mathbf{q}_G \ ,$$

where for each group $G \in \mathcal{G}$, $\mathbf{p}_G = (\boldsymbol{\xi}_G - \boldsymbol{Q}_G^\intercal \boldsymbol{\beta}_G)$, $\mathbf{q}_G = \sum_{i=1}^{\ell} \lambda_{G,i}\mathbf{S}_{G,i}^{-1}\bar{\mathbf{v}}_{G,i}$, $\mathbf{T} = \left(\sum_{i=1}^{\ell}, \lambda_{G,i}\mathbf{S}_{G,i}^{-1}\right)^{-1}$, and $\Lambda = \mathcal{A} - \mathbb{R}^{nm}_{0+} = \{\boldsymbol{\xi} \in \mathbb{R}^{nm} \mid \forall a \in N : \sum_{i\in I} \boldsymbol{\xi}_{am+i} \leq \bar{\kappa}_a, \forall i \in I : \sum_{a\in N} \boldsymbol{\xi}_{am+i} \leq \bar{\psi}_i, \boldsymbol{\xi} \preceq \mathbf{c}\}$.

As in Proposition 3.1, we can determine the set of optimal $(\mathbf{a}, \boldsymbol{\zeta})$ pairs by solving the system of equations: $\{(\mathbf{a}^*, \boldsymbol{\zeta}^*) \mid \mathbf{a} \in \mathcal{A}, G \in \mathcal{G} : \mathbf{a}_G - \boldsymbol{\zeta}_G = \boldsymbol{\xi}_G^*, \boldsymbol{\zeta} \in \mathbb{R}^{nm}_{0+}\}$.

□

H.5 Proof of Corollary 3.6

Corollary 3.6 (Group Egalitarian Welfare with Polyhedral Uncertainty)**.** *In the case where the uncertainty set $\mathcal{V}$ is defined only by linear constraints, i.e., $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^{nm} \mid \forall G \in \mathcal{G} : \forall G \in \mathcal{G} : \boldsymbol{Q}_G\mathbf{v}_G \succeq \mathbf{e}_G, \mathbf{v} \succeq 0\}$, the max-min-min problem in* (5) *transforms into a linear program.*

Proof. Now consider the inner-minimization problem:

$$\max_{\mathbf{a}\in\mathcal{A}} \min_{G\in\mathcal{G}} \min_{\mathbf{v}_G\in\mathbb{R}^{|G|m}_{0+}} \mathbf{a}_G{}^\intercal \mathbf{v}_G$$
$$\forall G \in \mathcal{G} : \ Q_G\mathbf{v}_G \succeq \mathbf{e}_G \ .$$

We can compute the Lagrangian dual of the inner-most minimization problem for each group independently by following steps outlined in the proof of Corollary 3.2. Note that since these minimization problems are simple linear programs, their corresponding duals are exact equivalents of their primal counterparts (Section 5.1 in Boyd *et al.* [12]). By substituting the duals in the above problem and reordering the inner-maximization and minimization problems, we obtain

$$\max_{\mathbf{a}\in\mathcal{A},\boldsymbol{\beta}\in\mathbb{R}^{g\times k}} \min_{G\in\mathcal{G}} \ \boldsymbol{\beta}_G^\intercal \mathbf{e}_G$$
$$\boldsymbol{Q}_G^\intercal \boldsymbol{\beta}_G \preceq \mathbf{a}_G \ .$$

Using simple algebraic manipulations, we can further simplify the above optimization problem as

$$\max_{\mathbf{a}\in\mathcal{A},\boldsymbol{\beta}\in\mathbb{R}^{g\times k},t\in\mathbb{R}} \ t$$
$$\forall G \in \mathcal{G} : t \leq \boldsymbol{\beta}_G^\intercal \mathbf{e}_G$$
$$\forall G \in \mathcal{G} : \boldsymbol{Q}_G^\intercal \boldsymbol{\beta}_G \preceq \mathbf{a}_G \ .$$

□

H.6 Proof of Corollary 3.7

Corollary 3.7 (Group Egalitarian Welfare with Ellipsoidal Uncertainty)**.** *Suppose that the set $\mathcal{V}$ in* (5) *is defined by a single truncated ellipsoidal constraint per group i.e., $\mathcal{V} = \{\mathbf{v} \in \mathbb{R}^{nm} \mid \forall G \in \mathcal{G} : (\mathbf{v}_G - \bar{\mathbf{v}}_G)\mathbf{S}_G^{-1}(\mathbf{v}_G - \bar{\mathbf{v}}_G) \leq r_G^2, \mathbf{v} \succeq 0\}$. Then the problem in* (5) *is equivalent to solving*

$$\max_{\substack{\boldsymbol{\lambda}\in\mathbb{R}^{g}_{0+}\\ \boldsymbol{\xi}\in\Lambda}} \min_{G\in\mathcal{G}} \boldsymbol{\xi}_G^\intercal \bar{\mathbf{v}}_G - \frac{\boldsymbol{\xi}_G^\intercal \mathbf{S}_G \boldsymbol{\xi}_G}{4\lambda_G} - \lambda_G r_G^2 \ .$$

The exact optimal solution $(\lambda^*, \boldsymbol{\xi}^*)$ to Equation (4) *can be computed by alternately performing two steps until convergence: first, fixing $\boldsymbol{\xi}$ and optimizing $\boldsymbol{\lambda}$, i.e., $\forall G \in \mathcal{G}$, $\lambda_G = \boldsymbol{\xi}_G^\intercal \mathbf{S}_G \boldsymbol{\xi}_G / 2r_G$, and second, fixing $\boldsymbol{\lambda}$ and solving a concave quadratic program to optimize $\boldsymbol{\xi}$. The optimal allocation $\mathbf{a}^*$ can be computed from $\boldsymbol{\xi}^*$ as in Proposition 3.1.*

Proof. Consider the following optimization problem.

$$\max_{\mathbf{a}\in\mathcal{A}} \min_{G\in\mathcal{G}} \min_{\mathbf{v}_G\in\mathbb{R}_{0+}^{|G|m}} \mathbf{a}_G^\intercal \mathbf{v}_G$$
$$\forall G\in\mathcal{G}:\ (\mathbf{v}_G-\bar{\mathbf{v}}_G)^\intercal \mathbf{S}_G^{-1}(\mathbf{v}_G-\bar{\mathbf{v}}_G)\le r_G^2$$
$$\mathbf{v}_G\succeq 0\ .$$

Similar to the general version of the problem in (17), the inner-most minimization is a convex optimization problem and the outer-maximization is a concave maximization problem. This is again due to the fact that affine functions are either concave or convex and minimum of concave objectives is concave.

The inner-most minimization over the uncertainty set of valuation matrices is independent for each group. Therefore, by simply replacing each of these minimization problems with their respective Lagrangian duals, as computed in Corollary 3.3, we obtain

$$\max_{\mathbf{a}\in\mathcal{A}} \min_{G\in\mathcal{G}} \max_{\substack{\lambda_G\in\mathbb{R}_{0+}\\ \boldsymbol{\zeta}_G\in\mathbb{R}_{0+}^{|G|m}}} -\frac{1}{4\lambda_G}\Big(\big(\mathbf{a}_G-\boldsymbol{\zeta}_G\big)^\intercal \mathbf{S}_G\big(\mathbf{a}_G-\boldsymbol{\zeta}_G\big)\Big)+(\mathbf{a}_G-\boldsymbol{\zeta}_G)^\intercal\bar{\mathbf{v}}_G-\lambda_G r_G^2\ .$$

Note that the dual is computed following the approach outlined in the proof of Proposition 3.1.

Using the change of variables $\forall G\in\mathcal{G}:\ \boldsymbol{\xi}_G=\mathbf{a}_G-\boldsymbol{\zeta}_G$, we get

$$\max_{\mathbf{a}\in\mathcal{A}} \min_{G\in\mathcal{G}} \max_{\substack{\lambda_G\in\mathbb{R}_{0+}\\ \boldsymbol{\zeta}_G\in\mathbb{R}_{0+}^{|G|m}\\ \boldsymbol{\xi}_G\in\mathbb{R}^{|G|m}}} \boldsymbol{\xi}_G^\intercal\bar{\mathbf{v}}_G-\frac{\boldsymbol{\xi}_G^\intercal\mathbf{S}_G\boldsymbol{\xi}_G}{4\lambda_G}-\lambda_G r_G^2 \quad \text{s.t. } \boldsymbol{\xi}_G=\mathbf{a}_G-\boldsymbol{\zeta}_G\ .$$

Since the inner maximization for each group is independent of the other groups, we can re-order the inner minimization over groups and the inner-maximization problem. Thus, without loss of generality, we can write the above optimization problem as

$$\max_{\substack{\mathbf{a}\in\mathcal{A}\\ \boldsymbol{\lambda}\in\mathbb{R}_{0+}^{g}\\ \boldsymbol{\zeta}\in\mathbb{R}_{0+}^{nm}\\ \boldsymbol{\xi}\in\mathbb{R}^{nm}}} \min_{G\in\mathcal{G}} \boldsymbol{\xi}_G^\intercal\bar{\mathbf{v}}_G-\frac{\boldsymbol{\xi}_G^\intercal\mathbf{S}_G\boldsymbol{\xi}_G}{4\lambda_G}-\lambda_G r_G^2$$
$$\text{s.t. } \boldsymbol{\xi}_G=\mathbf{a}_G-\boldsymbol{\zeta}_G\ .$$

Using the same technique as in Appendix H.1, we can simplify the problem by eliminating the variables $\mathbf{a}$ and $\boldsymbol{\zeta}$ in the above problem and then derive them from the optimal $\boldsymbol{\xi}$.

Eliminating $\boldsymbol{\zeta}$ and $\mathbf{a}$ in the above optimization problem, we get

$$\boldsymbol{\lambda}^*,\boldsymbol{\xi}^*=\arg\max_{\substack{\boldsymbol{\lambda}\in\mathbb{R}_{0+}^{g}\\ \boldsymbol{\xi}\in\Lambda}} \min_{G\in\mathcal{G}} \boldsymbol{\xi}_G^\intercal\bar{\mathbf{v}}_G-\frac{\boldsymbol{\xi}_G^\intercal\mathbf{S}_G\boldsymbol{\xi}_G}{4\lambda_G}-\lambda_G r_G^2$$
$$\text{s.t. } \boldsymbol{\xi}_G=\mathbf{a}_G-\boldsymbol{\zeta}_G\ ,$$

where $\Lambda=\mathcal{A}-\mathbb{R}_{0+}^{nm}=\{\boldsymbol{\xi}\in\mathbb{R}^{nm}\mid \forall a\in N:\ \sum_{i\in I}\boldsymbol{\xi}_{am+i}\le\bar{\kappa}_a,\forall i\in I:\ \sum_{a\in N}\boldsymbol{\xi}_{am+i}\le\bar{\psi}_i,\boldsymbol{\xi}\preceq\boldsymbol{c}\}$.

As in Proposition 3.1, we can now determine the set of optimal $(\mathbf{a},\boldsymbol{\zeta})$ pairs by solving the system of equations: $\{(\mathbf{a}^*,\boldsymbol{\zeta}^*)\mid \mathbf{a}\in\mathcal{A},\ G\in\mathcal{G}:\mathbf{a}_G-\boldsymbol{\zeta}_G=\boldsymbol{\xi}_G^*,\boldsymbol{\zeta}\in\mathbb{R}_{0+}^{nm}\}$.

□

H.7 Proof of Proposition 3.8

Proposition 3.8 (Decomposition for Monotonic Welfare Functions)**.** *Consider an optimization problem of the form*

$$\max_{\mathbf{a}\in\mathcal{A}} \min_{\mathbf{v}\in\mathcal{V}} \mathrm{W_M}(\mathbf{u}(\mathbf{a},\mathbf{v}))\ , \tag{8}$$

where the welfare function $\mathrm{W_M}$ *is monotonic in the utility of groups. If Assumption 3.4 holds, then* (8) *simplifies to*

$$\max_{\mathbf{a}\in\mathcal{A}} \mathrm{W_M}\left(\min_{\mathbf{v}_{G_1}\in\mathcal{V}_{G_1}} \mathrm{u}_{G_1}(\mathbf{a}_{G_1},\mathbf{v}_{G_1}),\min_{\mathbf{v}_{G_2}\in\mathcal{V}_{G_2}} \mathrm{u}_{G_2}(\mathbf{a}_{G_2},\mathbf{v}_{G_2}),\ldots,\min_{\mathbf{v}_{G_g}\in\mathcal{V}_{G_g}} \mathrm{u}_{G_g}(\mathbf{a}_{G_g},\mathbf{v}_{G_g})\right)\ .$$

Proof. The result directly follows from the monotonic property of the welfare function and the independence of the uncertainty sets across groups. □

H.8 Proof of Proposition 4.1

Proposition 4.1 (Approximate CVaR of USW)**.** *Given h i.i.d samples of $\tilde{\mathbf{v}}$, i.e., $\mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3, \ldots, \mathbf{v}^h$ from $\mathcal{D}_{\tilde{\mathbf{v}}}$, the optimal allocation for the problem in* (9) *can be approximately computed by solving*

$$\max_{\mathbf{a}\in\mathcal{A}} \max_{\substack{\boldsymbol{y}\in\mathbb{R}^h_{0+}\\ b\in\mathbb{R}}} \left(b - \frac{1}{\alpha}\sum_{j=1}^{h} \boldsymbol{y}_j \right) \qquad \forall j \in [1,h]: \ \boldsymbol{y}_j \geq \frac{1}{h}\left(b - \mathbf{w}\cdot\mathbf{u}(\mathbf{a},\mathbf{v}^j)\right) \ . \tag{10}$$

Proof. Consider the CVaR of utilitarian welfare optimization problem, given by

$$\max_{\mathbf{a}\in\mathcal{A}} \mathrm{CVaR}_\alpha\left[\mathbf{w}\cdot\mathbf{u}(\mathbf{a},\tilde{\mathbf{v}})\right] \doteq \max_{\mathbf{a}\in\mathcal{A}, b\in\mathbb{R}} \left\{ b - \frac{1}{\alpha}\mathop{\mathbb{E}}_{\tilde{\mathbf{v}}\sim\mathcal{D}_{\tilde{\mathbf{v}}}}\left[\left(b - \mathbf{w}\cdot\mathbf{u}(\mathbf{a},\tilde{\mathbf{v}})\right)_+\right]\right\} \ ,$$

Substituting the expectation with the empirical expectation computed from the h samples of random valuation $\tilde{v}$, and using a slack variable $\boldsymbol{y} \in \mathbb{R}^h_{0+}$ to convert $\left(b - \mathbf{w}\cdot\mathbf{u}(\mathbf{a},\tilde{\mathbf{v}})\right)_+$ term in the expectation to linear constraints, we get

$$\begin{aligned}
&\max_{\mathbf{a}\mathcal{A}} \max_{\boldsymbol{y}\in\mathbb{R}^h, b\in\mathbb{R}} \left(b - \frac{1}{\alpha}\sum_{j=1}^{h} y_j \right)\\
&\forall j \in [1,h]: \ y_j \geq 0\\
&\forall j \in [1,h]: \ y_j \geq \frac{1}{h}\left(b - \sum_{G\in\mathcal{G}} \frac{\mathbf{w}_G}{|G|}\cdot \mathbf{a}_G^\intercal \mathbf{v}_G^j \right)\\
&\mathbf{a} \in \mathcal{A} \ .
\end{aligned}$$

□

Assumption H.1 (L.A. *et al.* [35])**.** *The random variable $\tilde{x}$ is continuous with probability density function f that satisfies the following condition: There exists $\eta, \gamma > 0$ such that $\forall y \in [v_\alpha - \gamma, v_\alpha + \gamma]: \ f(y) > \eta$, where $v_\alpha = F^{-1}(\alpha)$.*

Theorem H.2 (Theorem 3.1 in [35])**.** *Let $(\tilde{x}_i)_{i=1}^h$ be a sequence of $i.i.d$ random variables. Suppose that $\tilde{x}_i, i = 1, \ldots, n$ are $\sigma-$sub-Gaussian and Assumption H.1 holds. If c_α and $\hat{c}_{h,\alpha}$ represent the true* CVaR *and the empirical* CVaR *of random variable $\tilde{x}$ estimated from h samples at confidence level α respectively, then for any $\varepsilon > 0$, we have*

$$\Pr\left[|\hat{c}_{h,\alpha} - c_\alpha| > \varepsilon\right] \leq 6\exp\left(\frac{-h\alpha^2 \min(\varepsilon^2, 16\gamma^2)\min(\eta^2, 1)}{8\max(8,\sigma^2)} \right) \ .$$

H.9 Proof of Proposition 4.2

Proposition 4.2 (Sample Complexity of Approximate CVaR of USW)**.** *Suppose that $\tilde{\mathbf{v}}$ is a multivariate sub-Gaussian random variable with mean $\bar{\mathbf{v}} \in \mathbb{R}^{nm}$ and covariance proxy $\mathbf{S} \in \mathbb{R}^{nm\times nm}$, i.e., for all vectors $\boldsymbol{z} \in \mathbb{R}^{nm}$: $\mathbb{E}_{\mathbf{v}\sim\mathcal{D}_{\tilde{\mathbf{v}}}}\left[\exp((\mathbf{v}-\bar{\mathbf{v}})^\intercal \boldsymbol{z}))\right] \leq \exp(\boldsymbol{z}^\intercal \mathbf{S}\boldsymbol{z}/2)$, and that, for any risk level $\alpha \in (0, \frac{1}{2})$ and allocation $\mathbf{a} \in \mathcal{A}$, there exists probability density threshold $\eta > 0$ and radius $\gamma > 0$, s.t., $f_{\mathbf{a}}(x) > \eta, \forall x \in [\nu_\alpha - \gamma, \nu_\alpha + \gamma]$. Set $\forall G \in \mathcal{G} : \mathbf{a}'_G = \frac{\mathbf{w}_G \cdot \mathbf{a}_G}{|G|}$. Then, for any confidence parameter $\delta \in (0,1)$ and error tolerance $\varepsilon > 0$,*

$$\Pr\left[\sup_{\mathbf{a}\in\mathcal{A}} |\hat{c}_{h,\alpha}(\mathbf{a}) - c_\alpha(\mathbf{a})| \leq \varepsilon\right] \geq 1 - \delta \quad \textit{for} \quad h > \left\lceil \frac{8\max\left(\max_{\mathbf{a}\in\mathcal{A}} \mathbf{a}'^\intercal \mathbf{S}\mathbf{a}', 8\right)\ln\left(\frac{6|\mathcal{A}|}{\delta}\right)}{\min(\varepsilon^2, 16\gamma^2)\alpha^2\min(\eta^2,1)} \right\rceil \ .$$

Proof. From the assumption, we know that the valuation vector $\mathbf{v}$ is a sub-Gaussian that satisfies the following condition

$$\forall \boldsymbol{z} \in \mathbb{R}^{nm} : \mathbb{E}\left[\exp((\mathbf{v}-\bar{\mathbf{v}})^\intercal \boldsymbol{z}))\right] \leq \exp(\boldsymbol{z}^\intercal \mathbf{S}\boldsymbol{z}/2) \ .$$

Using the above properties of sub-Gaussian, we can establish that the utilitarian welfare for a given allocation $\mathbf{a}$ is also a sub-Gaussian with variance-proxy = $\mathbf{a}'^\intercal \mathbf{S}\mathbf{a}'$ where $\forall G \in \mathcal{G} : \mathbf{a}'_G = \frac{\mathbf{w}_G \cdot \mathbf{a}}{|G|}$.

For any allocation $\mathbf{a}$, let $\hat{c}_{h,\alpha}(\mathbf{a})$ represent the empirical estimate of CVaR of utilitarian welfare and $c_{h,\alpha}(\mathbf{a})$ represent the corresponding true value.

Then, using Theorem H.2, we can bound the error of approximating the CVaR of the utilitarian welfare for allocation $\mathbf{a}$ as

$$\Pr\left[|\hat{c}_{h,\alpha}(\mathbf{a}) - c_\alpha(\mathbf{a})| > \varepsilon\right] \leq 6\exp\left(\frac{-h\alpha^2 \min(\varepsilon^2, 16\gamma^2)\min(\eta^2, 1)}{8\max(8, \mathbf{a}'^\intercal \mathbf{S}\mathbf{a}')}\right) \ .$$

The approximation error for all allocations can be upper-bounded as

$$\begin{aligned}\Pr\left[\sup_{\mathbf{a}\in\mathcal{A}} |\hat{c}_{h,\alpha}(\mathbf{a}) - c_\alpha(\mathbf{a})| > \varepsilon\right] &\leq \sum_{\mathbf{a}\in\mathcal{A}} \Pr\left[|\hat{c}_{h,\alpha}(\mathbf{a}) - c_\alpha(\mathbf{a})| > \varepsilon\right] \\ &\leq |\mathcal{A}| 6\exp\left(\frac{-h\alpha^2 \min(\varepsilon^2, 4\gamma^2)\min(\eta^2, 1)}{8\max(8, \mathbf{a}'^\intercal \mathbf{S}\mathbf{a}')}\right) \ .\end{aligned} \tag{18}$$

To obtain confidence guarantee $\Pr[\forall \mathbf{a} \in \mathcal{A} : |\hat{c}_{h,\alpha}(\mathbf{a}) - c_\alpha(\mathbf{a})| \leq \varepsilon] \geq 1-\delta$, we set the R.H.S of (18) to be less than δ, i.e.,

$$|\mathcal{A}| 6\exp\left(\frac{-h\alpha^2 \min(\varepsilon^2, 4\gamma^2)\min(\eta^2, 1)}{8\max(8, \mathbf{a}'^\intercal \mathbf{S}\mathbf{a}')}\right) < \delta \ .$$

Solving for h, we get

$$h > \left\lceil \left(\frac{8\max(\max_{\mathbf{a}\in\mathcal{A}} \mathbf{a}'^\intercal \mathbf{S}\mathbf{a}', 8)\ln\left(\frac{6|\mathcal{A}|}{\delta}\right)}{\min(\varepsilon^2, 4\gamma^2)\alpha^2 \min(\eta^2, 1)}\right)\right\rceil \ .$$

□

H.10 Proof of Proposition 4.3

Proposition 4.3 (CVaR of USW for Gaussian Distributions)**.** *If $\tilde{\mathbf{v}}$ is distributed as a multivariate Gaussian, i.e., $\tilde{\mathbf{v}} \sim \mathcal{N}(\bar{\mathbf{v}}, \mathbf{S})$, then, the optimization problem in* (9) *simplifies to*

$$\max_{\mathbf{a}\in\mathcal{A}} \mathbf{a}^\intercal \bar{\mathbf{v}} - \frac{\phi(\Phi^{-1}(\alpha))}{\alpha}\sqrt{\mathbf{a}^\intercal \mathbf{S}\mathbf{a}} \ , \tag{11}$$

Proof. The proof simply follows from the fact that for any Gaussian distributed random variable $\tilde{x} \sim \mathcal{N}(\mu, \sigma^2)$ with mean $\mu \in \mathbb{R}$ and standard deviation $\sigma \in \mathbb{R}_{0+}$, $\mathrm{CVaR}[\tilde{x}] = \mu - \frac{\phi(\Phi^{-1}(\alpha))}{\alpha}\sigma$. The mean of the utilitarian welfare under the assumption that $\mathbf{v}$ is Gaussian-distributed is given by $\mathbf{a}'^\intercal \mathbf{v}$ and variance is given by $\mathbf{a}'^\intercal \mathbf{S}\mathbf{a}'$, where for any $G \in \mathcal{G}, \mathbf{a}'_G = \frac{\mathbf{w}_G \cdot \mathbf{a}}{|G|}$. Substituting these values in the previously mentioned expression of CVaR of a Gaussian random variable, we obtain the results stated above. □

H.11 Linear Program for CVaR of Egalitarian Welfare

Proposition H.3 (Approximate CVaR of GESW)**.** *Given h samples of $\tilde{\mathbf{v}}$, i.e., $\mathbf{v}^1, \mathbf{v}^2, \mathbf{v}^3, \ldots \mathbf{v}^h$ sampled from $\mathcal{D}_{\tilde{\mathbf{v}}}$, the optimal allocation for the problem in* (12) *can be approximately computed by solving*

$$\begin{aligned}&\max_{\mathbf{a}\in\mathcal{A}} \max_{\boldsymbol{y}\in\mathbb{R}^h_{0+}, b\in\mathbb{R}} \left(b - \frac{1}{\alpha}\sum_{j=1}^{h} \boldsymbol{y}_j\right) \\ &\forall j \in [1, h] : \forall G \in \mathcal{G} : \ \boldsymbol{y}_j \geq \frac{1}{h}\left(b - \frac{1}{|G|}\cdot \mathbf{a}_G^\intercal \mathbf{v}_G^j\right) \ .\end{aligned} \tag{19}$$

Proof. Consider the CVaR of egalitarian welfare optimization problem, given by

$$\max_{\mathbf{a}\in\mathcal{A}} \mathrm{CVaR}_\alpha\left[\min_{G\in\mathcal{G}} \frac{1}{|G|}\cdot \mathbf{a}_G^\intercal \tilde{\mathbf{v}}_G\right] \quad \doteq \max_{b\in\mathbb{R},\mathbf{a}\in\mathcal{A}} \left\{ b - \frac{1}{\alpha} \mathop{\mathbb{E}}_{\tilde{\mathbf{v}}\sim\mathcal{D}_{\tilde{\mathbf{v}}}} \left[\left(b - \min_{G\in\mathcal{G}} \frac{1}{|G|}\cdot \mathbf{a}_G^\intercal \tilde{\mathbf{v}}_G \right)_+ \right] \right\} \quad .$$

Substituting the expectation in the above problem with the empirical expectation computed from the h samples of the valuation matrices, we get

$$\approx \max_{b\in\mathbb{R},\mathbf{a}\in\mathcal{A}} \left\{ b - \frac{1}{\alpha h} \sum_{i=1}^{h} \left(b - \min_{G\in\mathcal{G}} \frac{1}{|G|}\cdot \mathbf{a}_G^\intercal \mathbf{v}_G^i \right)_+ \right\} \quad .$$

Introducing slack variables $\boldsymbol{y}\in\mathbb{R}^h$, we can write the above problem as

$$\begin{aligned} &\max_{\mathbf{a}\in\mathcal{A}} \max_{\boldsymbol{y}\in\mathbb{R}^h, b\in\mathbb{R}} \left(b - \frac{1}{\alpha}\sum_{j=1}^{h} \boldsymbol{y}_j \right) \\ &\forall j\in[1,h]:\ \boldsymbol{y}_j \geq 0 \\ &\forall j\in[1,h]:\ \boldsymbol{y}_j \geq \frac{1}{h}\left(b - \min_{G\in\mathcal{G}} \frac{1}{|G|}\cdot \mathbf{a}_G^\intercal \mathbf{v}_G^j \right) \quad . \end{aligned} \tag{20}$$

Without loss of generality, we can represent the above problem as

$$\begin{aligned} &\max_{\mathbf{a}\in\mathcal{A}} \max_{\boldsymbol{y}\in\mathbb{R}^h, b\in\mathbb{R}} \left(b - \frac{1}{\alpha}\sum_{j=1}^{h} \boldsymbol{y}_j \right) \\ &\forall j\in[1,h]:\ \boldsymbol{y}_j \geq 0 \\ &\forall j\in[1,h]:\forall G\in\mathcal{G}:\ \boldsymbol{y}_j \geq \frac{1}{h}\left(b - \frac{1}{|G|}\cdot \mathbf{a}_G^\intercal \mathbf{v}_G^j \right) \quad . \end{aligned}$$

□

I GESW vs. USW under Different Scenarios

We run an experiment to discover scenarios where the USW-optimal solution has very sub-optimal GESW. Using the AAMAS 2015 dataset, we set all of the papers to be group 1, and we create a second, synthetic group of papers by copying and modifying a random subset of the papers. For these papers, we divide the copied valuations by some number, and set to zero all but the top valuations per paper. For each setting we compute the percentage of relative loss in GESW incurred by the maximum USW solution. Figure 4 shows the effects of varying the percentage of papers in the minority group, the number of non-zero entries, and the divisor. We use default values of 2 for the divisor, 5 for the number of nonzero entries, and 150 for the size of the minority group (corresponds to a ratio of roughly 20%). Taken together, the results suggest that if one group of papers has overall lower bids than another group, this has a very strong negative effect on the GESW of the USW-optimal solution.

J CVaR Performance under Skewed Gaussians

We use the AAMAS 2015 dataset and sample each valuation independently from a skewed-Gaussian distribution with varying skew parameter. We use the means and variances estimated by the Gaussian Process matrix factorization model described in Appendix E. The univariate skewed-Gaussian distribution with mean μ, variance σ^2, and skew α is defined by the probability density function

$$p(\tilde{\mathrm{x}};\alpha) = 2\phi(\tilde{\mathrm{x}}-\mu)\Phi(\alpha(\tilde{\mathrm{x}}-\mu)/\sigma),$$

where ϕ and Φ are the probability density function and cumulative density function of a standard Gaussian distribution [6].

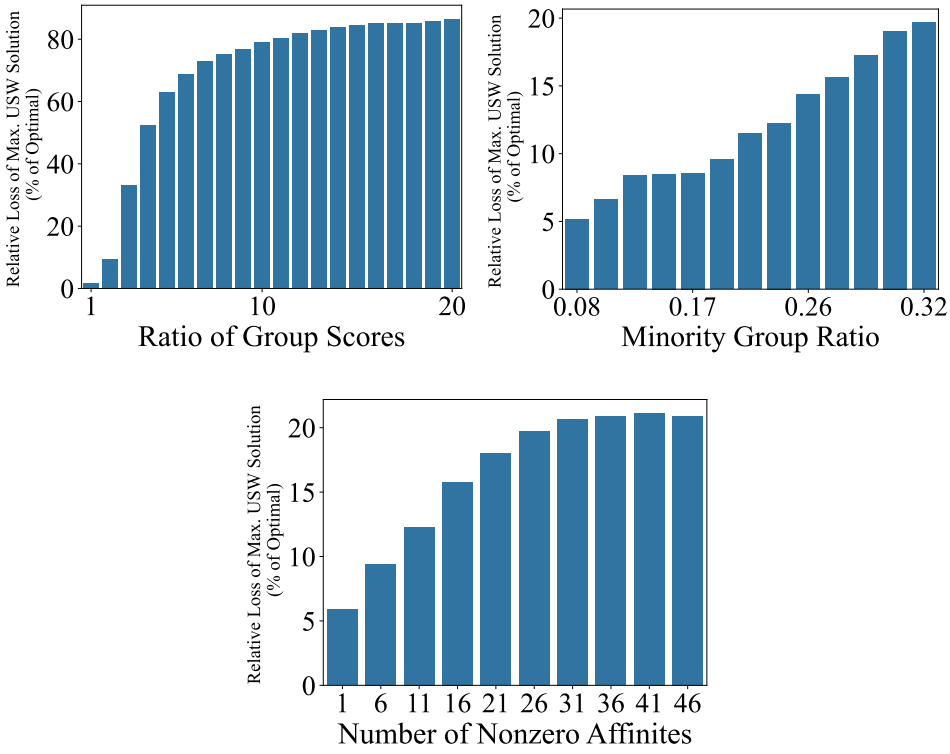


Figure 4: Relative loss (in GESW) of the maximum USW solution, compared to the optimal GESW solution. Results are reported for a synthetic 2-group example, varying 1) the divisor applied to artificially scale the minority group's valuations, 2) the ratio of the minority group to the overall number of papers, and 3) the number of valuations per paper that are artifically set to 0.

Table 8: Performance of different allocations for CVaR on the AAMAS 2015 with skewed Gaussians.

Skew	CVaR USW	USW	Rob. USW
-0.5	1.64	1.56	1.01
-1	1.45	1.21	0.86
-2	1.33	0.96	0.75
-5	1.29	0.84	0.70
-10	1.28	0.82	0.52

Table 9: Performance of different allocations for CVaR of USW on toy example with skewed Gaussians.

Approach	Solution	$\mathrm{CVaR}_{0.04}$[USW]
Naïve	$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$	0.947
Robust	$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$	0.972
CVaR	$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$	**0.985**

Table 10: Performance of different allocations for CVaR of GESW on toy example with skewed Gaussians.

Approach	Solution	$\mathrm{CVaR}_{0.04}$[GESW]
Naïve	$\begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{bmatrix}$	0.45
Robust	$\begin{bmatrix} 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$	0.46
CVaR	$\begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \end{bmatrix}$	**0.47**

We compare against the uncertainty-unaware and robust USW solution (using the ellipsoid derived from the Gaussian distribution, which is an optimistic uncertainty set). We optimize and evaluate for $\mathrm{CVaR}_{0.3}$. Table 8 displays the results. As the skew parameter α gets more negative, the $\mathrm{CVaR}_{0.3}$ of welfare of the naïve and robust solutions get much worse compared to the CVaR-optimized solution.

We also consider a constructed scenario with two agents and four items, where each agent must be assigned one item. Each agent's valuation for an item is independent and follows a skewed Gaussian distribution. The mean valuations for the items are represented by the matrix

$$\begin{bmatrix} 0.39 & 0.49 & 0.51 & 0.53 \\ 0.52 & 0.51 & 0.53 & 0.54 \end{bmatrix}.$$

The item values have standard deviations $\begin{bmatrix} 0.01 & 0.04 & 0.05 & 0.09 \end{bmatrix}$ respectively for both agents, and the skew parameter is $\alpha = 5$ for all item-agent pairs. We sample $20,000$ samples from the skewed Gaussian distibutions. We compute the naïve USW- or GESW-optimal allocations using the mean valuations, and we enumerate all possible allocations to identify the CVaR and robust allocations at a confidence level of 0.04. To estimate the robust statistic at a 0.04 confidence level, we assume independence and reject any samples that fall outside the $0.04/8$ confidence interval upper/lower bounds for any individual entry. We then compute the minimum welfare on the remaining samples. The computed solutions and the $\mathrm{CVaR}_{0.04}$ of each are displayed in Tables 9 and 10.

In both the USW and GESW cases, the naïve approach selects items 3 and 4, as they have the highest mean values, but it does not account for the uncertainty in the preferences. The robust approach, being more conservative, chooses items 1 and 2 due to their lower uncertainty. The CVaR approach strikes a balance between these two methods, selecting items 1 and 3. Item 1 has a lower mean value with low uncertainty, while item 3 has a higher mean value but with slightly higher uncertainty than item 2.

K Machine Specification

All experiments were run on Xeon E5-2680 v4 @ 2.40GHz machines with 128GB RAM with each experiment consuming at most 32 GB of memory. We ran 1500 experiments in total and each experiment took 3-4 hours.

NeurIPS Paper Checklist

1. **Claims**

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

   Answer: [Yes]

   Justification: The abstract describes the CVaR and robust optimization approaches taken in the paper, the utilitarian and egalitarian objectives studied, and mentions our theoretical and empirical results.

   Guidelines:

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

2. **Limitations**

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

   Answer: [Yes]

   Justification: See Appendix C.

   Guidelines:

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

3. **Theory Assumptions and Proofs**

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

Answer: [Yes]

Justification: All our formal statements appear with proof in the appendix.

Guidelines:

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

4. **Experimental Result Reproducibility**

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

Answer: [Yes]

Justification: Description of experiment provided in Section 5 and Appendix D.

Guidelines:

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

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

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

Answer: [Yes]

Justification: Link provided to code repository on Github in Section 5, link and recommended citation provided for data.

Guidelines:

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

6. **Experimental Setting/Details**

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

   Answer: [Yes]

   Justification: See Section 5 and appendix D, and code.

   Guidelines:

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

7. **Experiment Statistical Significance**

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

   Answer: [Yes] ,

   Justification: All results in the tables Section 5 and Appendix G are averaged over 5 runs of sub-sampling 20% of each dataset.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).

- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments Compute Resources**

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

   Answer: [Yes] ,

   Justification: The details of the compute resources used for the experiments is reported in Appendix K.

   Guidelines:

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

9. **Code Of Ethics**

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

   Answer: [Yes]

   Justification: We reviewed and abide by the code of ethics.

   Guidelines:

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

10. **Broader Impacts**

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

    Answer: [Yes] ,

    Justification: See Appendix B for details.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.

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

11. **Safeguards**

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

    Answer: [NA] .

    Justification: We have used publicly available datasets in our experiments.

    Guidelines:

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

12. **Licenses for existing assets**

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

    Answer: [Yes]

    Justification: We have only used Gurboi Solver for solving our optimization problems and have properly cited it.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, paperswithcode.com/datasets has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New Assets**

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

    Answer: [Yes]

    Justification: We linked to code in Section 5.

    Guidelines:

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

14. **Crowdsourcing and Research with Human Subjects**

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

    Answer: [NA]

    Justification: No human subjects.

    Guidelines:

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

15. **Institutional Review Board (IRB) Approvals or Equivalent for Research with Human Subjects**

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

    Answer: [NA]

    Justification: No human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.