# Vanishing Price of Anarchy in Large Coordinative Nonconvex Optimization

Mengdi Wang
mengdiw@princeton.edu

**Abstract**

We focus on a class of nonconvex cooperative optimization problems that involve multiple participants. We study the duality framework and provide geometric and analytic characterizations of the duality gap. The dual problem is related to a market setting in which each participant pursues self-interests at a given price of common goods. The duality gap is a form of price of anarchy. We prove that the nonconvex problem becomes increasingly convex as the problem scales up in dimension. In other words, the price of anarchy diminishes to zero as the number of participants grows. We prove the existence of a solution to the dual problem that is an approximate global optimum and achieves the minimal price of anarchy. We develop a coordination procedure to identify the solution from the set of all participants' best responses. Furthermore, we propose a globally convergent duality-based algorithm that relies on individual best responses to achieve the approximate social optimum. Convergence and rate of convergence analysis as well as numerical results are provided.

**Keywords:** nonconvex optimization, duality gap, price of anarchy, cooperative optimization, cutting plane method.

## 1 Introduction

Real-world social and engineering systems often involve a large number of participants with self-interests. The participants interact with one another by contributing to some common societal factors. For example, market participants make individual decisions on their consumption or production of common goods, which aggregate to the overall demand and supply in the market. As another example, factory owners make decisions about carbon emission, which aggregately influence the air quality of a city. The city as a whole pays a price for the carbon emissions generated by individuals. The overall social welfare is the aggregation of individual interests, as well as the utility or cost associated with the common factors. In this work, we study the social welfare optimization problem where the individual preferences are not necessarily convex.

Consider the optimization problem:

$$\begin{aligned} &\text{minimize } \left\{ F(x) = \sum_{i=1}^{N} p_i(x_i) + f\left(\sum_{i=1}^{N} g_i(x_i)\right)\right\}, \\ &\text{subject to } x_i \in \mathcal{X}_i, \qquad i = 1, \dots, N, \end{aligned} \tag{1}$$

where $p_i : \Re^{n_i} \mapsto \Re$, $g_i : \Re^{n} \mapsto \Re^M$ are functions that describe individual preferences and impacts on the common goods, $f : \Re^M \mapsto \Re$ is the social cost function of the common goods, $\mathcal{X}_i$ is a compact subset of $\Re^{n_i}$, and $n = \sum_{i=1}^{N} n_i$ is the dimension of all decisions. Here $N$ is the number of participants, and $M$ is the number of common factors. We focus on the case where $N \gg M$. Throughout this paper, we assume that the social cost function $f$ is convex, $p_i, g_i$ are arbitrary continuous functions, and $\mathcal{X}_i$ are arbitrary compact sets.

Type (1) problems are very common in cooperative optimization for multi-agent intelligence systems. Cooperative optimization is a critical challenge in practical decentralized systems, for example, aircraft or vehicle coordination [1, 30, 32], robot navigation [5, 18, 22], smart grid control [3, 31], communication and sensor networks [34, 35, 36]. Such a distributed system often involves a large number of agents that can communicate with one another subject to network communication constraints. These agents attempt to collaborate and to reach a global consensus or maximize a common objective. Meanwhile, each agent is likely to possess a private task or self-interests. Nonconvexity is inevitable in a majority of these applications. Due to the nonconvexity, it remains an open question whether the problem is tractable, let alone how to design efficient algorithms to achieve global coordination.

An important special case of problem (1) is the constrained problem:

$$\text{minimize} \sum_{i=1}^{N} p_i(x_i), \qquad \text{subject to} \sum_{i=1}^{N} g_i(x_i) \in \mathcal{A}, \quad x_i \in \mathcal{X}_i, \quad i = 1, \ldots, N. \tag{2}$$

It is a special case of problem (1) where $f(x) = 0$ if $x \in \mathcal{A}$, and $f(x) = +\infty$ if $x \notin \mathcal{A}$. Type (2) constrained problems find wide applications in resource allocation, where there is a budget on the total resources.

Let us consider a market setting. Suppose that each participant is charged at a price for making a negative impact on the public goods. Let $\mu \in \Re^M$ be the price vector. Each participant selfishly solves its own problem:

$$x_i(\mu) \in \text{argmin}\ \{p_i(x_i) + \mu^T g_i(x_i) \mid x_i \in \mathcal{X}_i\}, \quad i = 1, \ldots, N.$$

We ask the question:

> Is there a fair price vector $\mu^*$ such that the individual best responses $(x_1(\mu^*), \ldots, x_N(\mu^*))$ automatically achieve a social optimum, i.e., a global optimum of problem (1)?

In the case where problem (1) is convex, the answer is largely known to be yes. However, there is no simple answer in the lack of convexity. One may wonder why we care about the nonconvex case at all. In practical markets, each participant may face multiple alternative actions (discrete or continuous) to choose from. However, it is likely that the convex combination of two actions is either infeasible or suboptimal all the time. For example, a manufacturer seeks to reduce carbon emission by either installing a high-efficiency device or a low-efficiency one. It is likely that the convex combination of both is always suboptimal, because it requires two lump-sum payments. Indeed, nonconvexity is ubiquitous.

The price vector is naturally related to the multiplier of an appropriate dual problem. In the nonconvex case, strong duality fails to hold. There is a positive duality gap between the primal and dual problems. As a result, there is no guarantee that there exists a fair price at which social optimum can be automatically achieved via individual best responses. It is possible that, regardless of the price, the overall welfare is far from the optimal value due to lack of coordination among participants. We refer to this loss of efficiency as the *price of anarchy*. Indeed, we will illustrate with examples that a high price of anarchy is inevitable with neither convexity nor coordination.

In what follows, we aim to provide an answer to this question by analyzing the nonconvex duality of problem (1). We will show that the price of anarchy can be bounded by a form of duality gap between problem (1) and a suitable Fenchel dual problem. In order to quantify the duality gap, we study the dual geometry of problem (1). We discover that problem (1) exhibits a curious *convexification effect*, i.e., the problem becomes increasingly convex as the number of participants increases. This means that the duality gap can be made arbitrarily small when the number of participants becomes large. It also suggests that the highly nonconvex problem (1) can be approximately solved using an efficient duality-based approach. Under reasonable scaling of the problem, the price of anarchy vanishes to zero.

The convexification effect is due to an intuitive fact from convex geometry: *the sum of a large number of nonconvex sets tends to be convex.* The first result is the Shapley-Folkman Lemma, established by the Nobel prize-winning economist Lloyd Shapley [37]. It is used to derive an upper bound on the distance between the sum of many sets and its convex hull, and can be viewed as a discrete counterpart to the Lyapunov theorem on non-atomic measures [19]. Many others have looked into the convexification result from the analytical, geometric, and probabilistic perspective; see [2, 13, 14, 15, 29, 38, 49] for examples. The convexification effect has been widely studied in mathematical economy and game theory. It is used to show that central results of convex economic theory are good approximations to large economies with nonconvexities; see [11, 12, 17, 20, 37, 39, 46] for a selected few of these works. Most of the existing research focuses on proving the existence of quasi-equilibria in a variety of multi-person games and analyzing the economic impact.

In contrast to the economic literature, there exist only a handful of works that study the convexification effect from the optimization perspective. The pioneer work by Ekeland [16] studied a separable optimization problem:

$$\text{minimize } \sum_{i=1}^{N} p_i(x_i), \qquad \text{subject to} \quad \sum_{i=1}^{N} g_i(x_i) \le B, \tag{3}$$

and estimated an upper bound of the Langrangian duality gap. Later, Aubin and Ekeland [4], Bertsekas [9, 10], and Pappalardo [28] considered the same problem and proved sharper bounds on the duality gap. This idea of small duality gap has been used in spectral management in signal processing [24, 47], as well as in supply chain management [44]. Some earlier works have an emphasis on integer linear programming with special structures, e.g., [6, 43]. In addition, Lemaréchal and Renaud [23], Bertsekas [8], and Nedić and Asuman [27] have studied the Langrangian duality gap of generic nonconvex optimization from a geometric perspective. A recent work by Udell and Boyd [41] considers a separable problem with linear equality and inequality constraint. It proposes a randomized approximation method that minimizes the convex envelope of nonconvex cost functions. Another recent work by Fang et al. [45] in statistical learning proposes to estimate large sparse graphical models by solving an $\ell_0$-constrained optimization problem. It is shown that the duality gap is sufficiently smaller than the statistical error with high probability.

We note that the nonconvex problem (1) is closely related to integer programming. In fact, many integer linear programming problems can be seen as special cases of problem (1). An example is the Knapsack problem given by

$$\min c^T x, \qquad \text{subject to} \quad Ax \le b,\ x_i = \{0, 1\},\ i = 1, \ldots, n,$$

where $c \in \Re^n, A \in \Re^{m\times n}$, and $b \in \Re^m$. This can be seen to take the form (1) if we let $p_i, g_i$ be linear functions and let $\mathcal{X}_i$ be sets of discrete values. Although discrete constraint is not our focus in the current work, we keep in mind that our duality results do apply to these special cases. In contrast to the integer programming literature, our focus is to find an approximate solution by leveraging the duality. We are interested in developing algorithms that can be applied in multi-agent settings.

The main objective of this paper is to study the nonconvex duality of problem (1) from both the theoretical and algorithmic perspectives. By exploiting the convexification phenomenon of sums of nonconvex sets, we provide a mathematical characterization of the duality gap between the nonconvex problem (1) and its dual problem. We show that there exists a solution to the dual problem that is nearly primal optimal and attains the duality gap. In other words, there indeed exists a nearly fair price, which is the multiplier, such that the price of anarchy vanishes to 0 as the market size increases to infinity (assuming that the participants are willing to cooperate).

From an algorithmic perspective, we propose a coordinative algorithm that dynamically converges to an approximate global optimum. The algorithm relies on best responses from individual participants, who are required to optimize their own objective functions given a price vector/multiplier. The algorithm is based on a dual cutting-plane method and can be applied in a distributed multi-agent setting. Under mild conditions, it converges to the optimal multiplier at a rate of $\mathcal{O}(1/t)$ even if the dual function is nonsmooth. A key contribution is the algorithm's ability in identifying a near-optimal solution by coordinating individuals' alternative best responses. The coordination is cast into an approximate projection problem, which is related to finding the extreme point to a particular linear feasibility problem. We emphasize that coordination is the key to achieving approximate global optimum in nonconvex optimization. Without coordination, we find examples to show that the price of anarchy can be disastrously high.

To the best knowledge of the author, this is the first work that identifies the class of nonconvex problems (1) that bear a diminishing Fenchel duality gap (as the dimension of decision variable increases). This is also the first work that provides a tractable algorithmic solution to achieve the diminishing duality gap. It points out the necessity of enforcing coordination in multi-agent nonconvex optimization. The rest of the paper is summarized as follows:

**Outline** In Section 2, we introduce the duality framework of the nonconvex problem (1) and illustrate the geometry of the duality gap. In Section 3, we characterize the duality gap and show that it can be achieved by a particular solution to the dual problem. In Section 4, we study how to coordinate multiple best responses from the participants to achieve the approximate optimum, and we show by counter examples that the lack of coordination may result in an arbitrary high price of anarchy. In Section 5, we propose a coordinative algorithm that is based on individual best responses and show that it converges to the approximate global optimum at a favorable rate. In Section 6, we conduct numerical experiments, and in Section 7, we draw conclusions.

**Notation** All vectors are considered as column vectors. For a vector $x \in \Re^n$, we denote by $x^T$ its transpose, and denote by $\|x\| = \sqrt{x^T x}$ its Euclidean norm. For two sequences $\{a_k\}, \{b_k\}$, we denote by $a_k = \mathcal{O}(b_k)$ if there exists $c > 0$ such that $\|a_k\| \le c\|b_k\|$ for all $k$, and we denote by $a_k \to a$ if $\lim_{k\to\infty} a_k = a$. For a function $f(x)$, we denote by $\nabla f(x)$ its gradient at $x$ if $f$ is differentiable, and denote by $\partial f(x)$ its subdifferential (the set of subgradients) at $x$ if $f$ is nondifferentiable. For convenience, we denote by $\tilde{\nabla} f(x)$ a particular subgradient of $f$ at $x$, which will be specified in the context. For a set $\mathcal{A}$, we denote by $\mathrm{conv}(\mathcal{A})$ its convex hull, i.e., the set of all convex combinations of points in $\mathcal{A}$, and we denote by $|\mathcal{A}|$ its cardinality. For two sets $\mathcal{A}$ and $\mathcal{B}$, we denote by $\mathcal{A}/\mathcal{B}$ the set of their difference $\mathcal{A}/\mathcal{B} = \mathcal{A} \cap \mathcal{B}^c$, and we denote by $\mathcal{A} + \mathcal{B}$ their Minkowski sum (also known as vector sum), i.e., $\mathcal{A} + \mathcal{B} = \{a + b \mid a \in \mathcal{A}, b \in \mathcal{B}\}$. We denote by $\mathcal{X}$ the Cartesian product $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$, so the constraint of problem (1) can be written compactly as $x \in \mathcal{X}$.

## 2 Primal and Dual Problems

In this section, we introduce the duality framework for the nonconvex optimization problem (1). We also illustrate the geometric intuition of our analysis. Let $f^\star$ be the convex conjugate of the function $f$, which is defined as $f^\star(\mu) = \sup_{y\in\Re^M} \mu^T y - f(y)$. By using the convexity of $f$, we also have $f(y) = \sup_{\mu\in\Re^M} \mu^T y - f^\star(\mu)$.

In order to leverage the separability with respect to $x_1, \ldots, x_N$, we rewrite problem (1) using the Fenchel representation as

$$\min_{x\in\mathcal{X}} \left\{ \sum_{i=1}^N p_i(x_i) + \sup_{\mu\in\Re^M} \left\{ -f^\star(\mu) + \mu^T \sum_{i=1}^N g_i(x_i) \right\} \right\}. \tag{P}$$

which we refer to as the *primal problem*. We define the *dual problem* as the problem obtained by exchanging the "min" and "sup":

$$\sup_{\mu\in\Re^M} \left\{ -f^\star(\mu) + \sum_{i=1}^N \min_{x\in\mathcal{X}} \left\{ \mu^T g_i(x_i) + p_i(x_i) \right\} \right\}. \tag{D}$$

We let $L(x, \mu)$ be the Lagrangian function

$$L(x, \mu) = -f^\star(\mu) + \mu^T \sum_{i=1}^N g_i(x_i) + \sum_{i=1}^N p_i(x_i),$$

which is decomposable with respect to $x_1, \ldots, x_N$. The primal function $F(x)$ can be represented using the Lagrangian function as

$$F(x) = \sup_{\mu\in\Re^M} L(x, \mu) = \sup_{\mu\in\Re^M} -f^\star(\mu) + \mu^T \sum_{i=1}^N g_i(x_i) + \sum_{i=1}^N p_i(x_i).$$

Similarly, we define the dual function as

$$Q(\mu) = \min_{x\in\mathcal{X}} L(x, \mu) = -f^\star(\mu) + \sum_{i=1}^N \min_{x_i\in\mathcal{X}_i} \left\{ \mu^T g_i(x_i) + p_i(x_i) \right\}.$$

We denote by $F^*$ and $Q^*$ the optimal values of problems (P) and (D), respectively. They satisfy

$$F^* = \min_{x\in\mathcal{X}} F(x) = \min_{x\in\mathcal{X}} \sup_{\mu\in\Re^M} L(x, \mu), \qquad Q^* = \sup_{\mu\in\Re^M} Q(\mu) = \sup_{\mu\in\Re^M} \min_{x\in\mathcal{X}} L(x, \mu).$$

By using the weak duality (see [33]), we have

$$F^* \geq Q^*.$$

In earlier literatures, the nonnegative difference $F^* - Q^*$ is often referred to as the duality gap of problem (1); see [7] for an example. We argue that this notion of duality gap is not useful enough. The difference between two optimal values does not quantify how the optimal solution to the dual problem performs in the primal problem.

We need a different notion of duality gap. Let us consider an optimal multiplier $\mu^*$ to the dual problem, which satisfies $\mu^* \in \text{argmax } Q(\mu)$. We let $\check{x} = (\check{x}_1, \ldots, \check{x}_N)$ be a corresponding optimal solution to the dual problem. It consists of individual best responses to the multiplier $\mu^*$, i.e., for $i = 1, \ldots, N$,

$$\check{x}_i \in \text{argmin } \{p_i(x_i) + {\mu^*}^T g_i(x_i) \mid x_i \in \mathcal{X}_i\}.$$

In lack of convexity, the dual solution $\check{x}$ is often a suboptimal solution to the primal problem. We are interested in the positive difference $F(\check{x}) - F^*$, which is exactly the loss of total efficiency by applying $\check{x}$ to the primal objective. Note that even if the optimal multiplier $\mu^*$ is unique, there may exist multiple choices of $\check{x}_i$ that are best responses to $\mu^*$ for the $i$-th participant. Due to the nonconvex nature of $p_i$ and $g_i$, the set of such best responses is also nonconvex. We define the *duality gap* to be the minimal loss of efficiency given by

$$\begin{aligned} &\text{minimize } F(\check{x}) - F^* \\ &\text{subject to } \check{x}_i \in \text{argmin } \left\{ p_i(x_i) + {\mu^*}^T g_i(x_i) \mid x \in \mathcal{X} \right\}, \\ &\qquad\qquad i = 1, \ldots, N. \end{aligned} \tag{4}$$

If we interpret $\mu^*$ as the price vector of common goods, the normalized duality gap $\dfrac{F(\check{x}) - F^*}{F^*}$ can be viewed as the *minimal price of anarchy* under price $\mu^*$.

Let us try to understand the nonconvexity of problem (1) and its duality gap from a geometric point of view. Related earlier works include [8, 23, 27], which have focused on the Lagrangian duality. We define $\mathcal{W}_i \subset \Re^{M+1}$, $i = 1, \ldots, N$, to be the set

$$\mathcal{W}_i = \{(z, y) \mid \exists\, x_i \in \mathcal{X}_i : z \geq p_i(x_i), y = g(x_i)\}.$$

If $p_i$, $\mathcal{X}_i$ are convex and $g_i$ is linear, the set $\mathcal{W}_i$ is convex; otherwise $\mathcal{W}_i$ is not necessarily convex. The set $\mathcal{W}_i$ provides a joint characterization of the functions $(p_i, g_i)$ and the set $\mathcal{X}_i$. Moreover, the set $\mathcal{W}_i$ provides an invariant representation of the $i$th participant's interest under change of variables. In subsequent analyses, we will use the convexity gap of $\mathcal{W}_1, \ldots, \mathcal{W}_N$ as a metric of the lack of convexity of problem (1).

We define $\mathcal{W}$ to be the *Minkowski sum* of the sets $\mathcal{W}_1, \ldots, \mathcal{W}_N$ given by

$$\mathcal{W} = \mathcal{W}_1 + \cdots + \mathcal{W}_N = \left\{ \sum_{i=1}^N w_i \;\middle|\; w_i \in \mathcal{W}_i, i = 1, \ldots, N \right\}.$$

Equivalently, we have

$$\mathcal{W} = \left\{ (z, y) \;\middle|\; \exists\, x \in \mathcal{X} : z \geq \sum_{i=1}^N p_i(x_i), y = \sum_{i=1}^N g_i(x_i), i = 1, \ldots, N \right\}.$$

Since $\mathcal{X}$ is compact and $p_i, g_i$ are continuous, we can verify that $\mathcal{W}_1, \ldots, \mathcal{W}_N, \mathcal{W}$ are nonempty and compact.

Next we show that the primal problem (P) and the dual problem (D) can be represented using the set $\mathcal{W}$ and its convex hull $\text{conv}(\mathcal{W})$, respectively. By using the definition of $\mathcal{W}$, we see that the primal problem (P) is equivalent to

$$F^* = \min_{w \in \mathcal{W}} \sup_{\mu \in \Re^M} [1, \mu^T] w - f^\star(\mu). \tag{5}$$

Similarly, we can rewrite the dual problem (D) as

$$Q^* = \sup_{\mu \in \Re^M} \min_{w \in \mathcal{W}} [1, \mu^T] w - f^\star(\mu).$$

Note that minimizing a linear function over a closed set is equivalent to minimizing over the convex hull. Therefore, we can replace $\mathcal{W}$ with its convex hull and apply the minimax theorem to obtain

$$\begin{aligned} Q^* &= \sup_{\mu \in \Re^M} \min_{w \in \text{conv}(\mathcal{W})} [1, \mu^T] w - f^\star(\mu) \\ &= \min_{w \in \text{conv}(\mathcal{W})} \sup_{\mu \in \Re^M} [1, \mu^T] w - f^\star(\mu). \end{aligned} \tag{6}$$

Now let us compare Eq. (5) and Eq. (6). Clearly, the dual problem can be viewed as a convexified version of the primal problem. It is the set $\mathcal{W}$ that is convexified in the dual problem, instead of the nonconvex function $p_i, g_i$ and sets $\mathcal{X}_i$. We emphasize that the dual problem is *not* obtained by replacing nonconvex functions in problem (1) with their convex envelops.

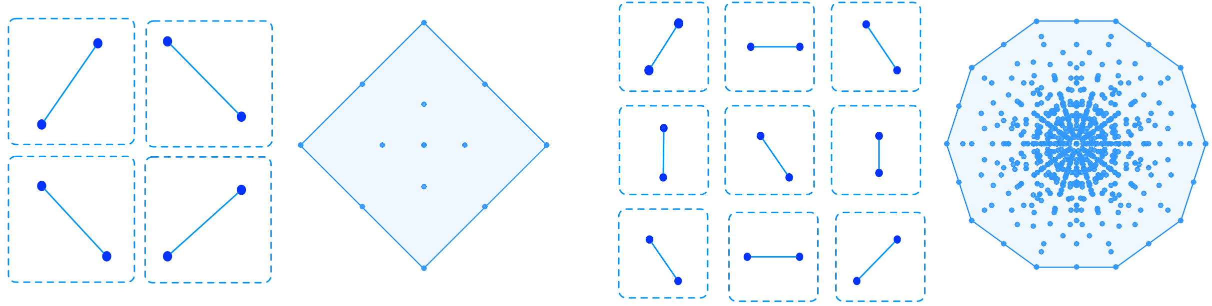

**Figure 1:** Illustration of the sums of nonconvex sets. On the left side, we plot four 2-point sets $\mathcal{W}_1, \ldots, \mathcal{W}_4$, their sum $\mathcal{W}_1 + \cdots + \mathcal{W}_4$, and the convex hulls of them. On the right side, we plot nine 2-point sets $\mathcal{W}_1, \ldots, \mathcal{W}_9$, their sum $\mathcal{W}_1 + \cdots + \mathcal{W}_9$, and the convex hulls of them. The convexity gap reduces as the number of sets increases.

The nonconvexity of the primal problem comes from the nonconvexity of the set $\mathcal{W}$. This implies that the duality gap between problems $(P)$ and $(D)$ can be estimated using the nonconvexity of $\mathcal{W}$. An important observation is that $\mathcal{W}$ is the sum of many sets $\mathcal{W} = \mathcal{W}_1 + \cdots + \mathcal{W}_N$. So the convexification phenomenon occurs (see Figure 1 for an illustration). When $N$ is a large number, we can show that the set $\mathcal{W}$ does not differ much from its convex hull $\text{conv}(\mathcal{W})$. This is the geometric motivation for our analysis on duality gap. In the rest of paper, we will leverage the duality framework and quantify the convexity gap of $\mathcal{W}$. Then we will be able to provide a complete mathematical characterization of the duality gap, and to show that the duality gap reduces to zero as the nonconvex problem scales up.

# 3 Bounding the Price of Anarchy by Duality Gap

We develop our main theoretical results in this section. First, we review some preliminaries about the convexity gap and the convexification of sums of sets. Second, we show that the primal optimality condition can be approximately satisfied by a solution to the dual problem. Then we derive upper bounds of the duality gap in two cases: the case with smooth penalty and the case with hard constraint. The minimal duality gap is achieved by a particular solution to the dual problem. Lastly, we study the asymptotic price of anarchy of the nonconvex problem under various scaling schemes.

## 3.1 Preliminaries

We state the Shapley-Folkman lemma which was first proved in [37]. It implies that the sum of many nonconvex sets does not differ much from its convex hull. To illustrate the key idea, we provide a simplified proof based on linear programming.

**Lemma 1** *(Shapley-Folkman lemma) Let $\mathcal{S}_1, \ldots, \mathcal{S}_n \subset \Re^m$, and let $x \in \text{conv}(\mathcal{S}_1 + \cdots + \mathcal{S}_n)$. There exists $x_1, .., x_n$ and $\mathcal{I} \subset \{1, \ldots, n\}$ such that $x = x_1 + \cdots + x_n$ and $|\mathcal{I}| \leq m$ with*

$$x_i \in \begin{cases} \mathcal{X}_i & \textit{if } i \notin \mathcal{I}, \\ \text{conv}(\mathcal{X}_i)/\mathcal{X}_i & \textit{if } i \in \mathcal{I}. \end{cases}$$

*Proof.* We start with the case where $\mathcal{S}_1 + \cdots + \mathcal{S}_n$ is a finite set, which implies that each $\mathcal{S}_i$ is also a finite set.

We let $A^{(i)}$ be the matrix whose columns are equal to vectors in $\mathcal{S}_i$, and let $e^{(i)}$ be the unit vector whose dimension is equal to the column dimension of $A^{(i)}$, where $i = 1, \ldots, n$. Consider the following linear feasibility problem for variables $z^{(i)}, i = 1, \ldots, n$,

$$\sum_{i=1}^{n} A^{(i)} z^{(i)} = x, \quad e^{(i)} z^{(i)} = 1, \quad z^{(i)} \geq 0, \quad i = 1, \ldots, n. \tag{7}$$

We see that (7) takes the standard form of linear programming and has $m + n$ equalities ($m$ from the row dimension of $A(i)$). We claim that there exists at least one feasible solution. Since $x \in \text{conv}(\mathcal{S}_1 + \cdots + \mathcal{S}_n)$, there exists $\lambda_1, \ldots, \lambda_m$ such that

$$x = \sum_{j=1}^{m} \lambda_j \left( \sum_{i=1}^{n} x_{ij} \right), \qquad \text{where } x_{ij} \in \mathcal{S}_i,$$

and $\sum_{j=1}^{m} \lambda_j = 1, \lambda_j \geq 0$, for all $i = 1, \ldots, n, j = 1, \ldots, m$. We let $z^{(i)}$ be the vector whose entry takes value $\lambda_j$ when the corresponding column in $A^{(i)}$ is equal to $x_{ij}$ and takes value zero otherwise. We can verify that the constructed solution $(z^{(1)}, \ldots, z^{(n)})$ is a feasible solution to (7).

Because (7) is feasible and has $m+n$ equality constraints, there exists a basic feasible solution $z = (z^{(1)}, \ldots, z^{(n)})$ to (7) with at most $m + n$ nonzero elements (see [42]). According to the constraint $e^{(i)} z^{(i)} = 1, z^{(i)} \geq 0$, each $z^{(i)}$ has at least one nonzero element. This implies that $z = (z^{(1)}, \ldots, z^{(n)})$ has at most $m$ components $z^{(i)}$ with more than one nonzero entries. We take $x_i = A^{(i)} z^{(i)}$ for each $i$, and $\mathcal{I} = \{i \mid \|z^{(i)}\|_0 > 1\}$. Clearly, we have $x_i \in \mathcal{X}_i$ if $i \notin \mathcal{I}$ and $x_i \in \text{conv}(\mathcal{X}_i)/\mathcal{X}_i$ if $i \in \mathcal{I}$. Moreover, the cardinality of the set $\mathcal{I}$ is at most $m$.

Finally, we consider the general case where $\mathcal{S}_1 + \cdots + \mathcal{S}_n$ is an infinite set. By Carathéodory's theorem, any point $x \in \text{conv}(\mathcal{S}_1 + \cdots + \mathcal{S}_n)$ can be represented by the convex combination of at most $m$ points in $\mathcal{S}_1 + \cdots + \mathcal{S}_n$. Then it is sufficient to focus on the remaining $m$ points and apply the preceding analysis. ■

The Shapley-Folkman lemma implies that any point in $\text{conv}(S_1 + \cdots + S_n)$ can be approximated by a point in $S_1 + \cdots + S_n$. In order to quantify the precision of approximation, we need a notion of convexity gap for sets. Let $\mathcal{S}$ be an arbitrary set. A notion *convexity gap* of $\mathcal{S}$ is defined as

$$\rho(\mathcal{S}) = \sup\{\|x - y\| \mid \lambda x + (1 - \lambda) y \in \text{conv}(\mathcal{S})/\mathcal{S}, \ \forall \lambda \in [0, 1]\},$$

which is the length of the longest line segment that lies in $\text{conv}(\mathcal{S})/\mathcal{S}$.

Let us focus on the optimization problem (1). In order to prove tight duality gap results, we need to customize the notion of convexity gap to our problem. Let

$$\rho_p(\mathcal{W}_i) = \sup\{\|z_1 - z_2\| \mid \lambda(z_1, y_1) + (1 - \lambda)(z_2, y_2) \in \text{conv}(\mathcal{W}_i)/\mathcal{W}_i, \forall \ \lambda \in [0, 1]\},$$

and

$$\rho_g(\mathcal{W}_i) = \sup\{\|y_1 - y_2\| \mid \lambda(z_1, y_1) + (1 - \lambda)(z_2, y_2) \in \text{conv}(\mathcal{W}_i)/\mathcal{W}_i, \forall \ \lambda \in [0, 1]\}.$$

They can be viewed as partial convexity gaps of set $\mathcal{W}_i$ with respect to the preference and impact functions $p_i$ and $g_i$, respectively. We easily see that $\rho_p(\mathcal{W}_i) \leq \rho(\mathcal{W}_i)$ and $\rho_g(\mathcal{W}_i) \leq \rho(\mathcal{W}_i)$ for all $i = 1, \ldots, N$. We further define

$$\delta_g = \max_{i=1,\ldots,N} \rho_g(\mathcal{W}_i), \qquad \delta_p = \max_{i=1,\ldots,N} \rho_p(\mathcal{W}_i).$$

They are the maximal nonconvex gaps of sets $\mathcal{W}_1, \ldots, \mathcal{W}_N$, with respect to the impact and preference, respectively. They will be used frequently in subsequent analyses.

We say a convex function $f$ is *$\beta$-strongly smooth* if it is continuously differentiable and satisfies for all $x, y$ that

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{\beta}{2} \|y - x\|^2.$$

We say $f$ is *$\sigma$-strongly convex* if for all $x, y$ and $\gamma \in [0, 1]$ that

$$\gamma f(x) + (1 - \gamma) f(y) \geq f(\gamma x + (1 - \gamma) y) + \frac{\sigma}{2} \gamma (1 - \gamma) \|y - x\|^2,$$

and if in addition $f$ is differentiable, we have for all $x, y$ that

$$f(y) \geq f(x) + \nabla f(x)^T (y - x) + \frac{\sigma}{2} \|y - x\|^2.$$

The next lemma states the duality between strong convexity and strong smoothness, which has been proved in [21] (see Theorem 6).

**Lemma 2** *Assume that $f$ is a closed and convex function. Then $f$ is $\beta$-strongly smooth if and only if its convex conjugate $f^\star$ is $1/\beta$-strongly convex.*

In addition, we need the following lemma that characterizes the subdifferential of the dual function.

**Lemma 3** *The subdifferential of the dual function is*

$$\partial Q(\mu) = \text{conv} \left\{ \sum_{i=1}^{N} g_i(x_i) - y \;\middle|\; x \in \check{\mathcal{X}}^*(\mu), y \in \partial f^\star(\mu) \right\},$$

*where*

$$\check{\mathcal{X}}^*(\mu) = \text{argmin} \left\{ (\mu)^T \left( \sum_{i=1}^{N} g_i(x_i) \right) - f^\star(\mu) + \sum_{i=1}^{N} p_i(x_i) \;\middle|\; x \in \mathcal{X} \right\}.$$

*Proof.* We have $Q(\mu) = -f^\star(\mu) + \min_{x \in \mathcal{X}} \left\{ \mu^T \sum_{i=1}^{N} g_i(x_i) + \sum_{i=1}^{N} p_i(x_i) \right\}$. Since $-Q(\mu)$ is the pointwise maximum of multiple convex functions, its subdifferential is the convex hull of all subgradients of functions that are currently active. ■

## 3.2 Main Results on Duality Gap

Our first main result establishes the existence of a "reasonable" solution to the dual problem. We show that one can construct a solution $\check{x}$ to the dual problem such that the strong duality condition is approximately achieved, i.e.,

$$\check{x} \in \text{argmin}\, L(x, \mu^*), \qquad \tilde{\nabla}_\mu L(\check{x}, \mu^*) \approx 0.$$

This implies that $\check{x}$ is an approximate saddle point of the Lagrangian function.

**Theorem 1 (Near-Optimality Condition)** *Suppose there exists an optimal dual multiplier $\mu^*$ to problem* (D). *Then there exists a solution $\check{x}$ to the dual problem* (D) *and a subgradient $\tilde{\nabla} f^\star(\mu^*) \in \partial f^\star(\mu^*)$ such that*

$$\left\| N \tilde{\nabla} f^\star(\mu^*) - \sum_{i=1}^{N} g_i(\check{x}_i) \right\| \leq M \delta_g, \qquad (\mu^*)^T \left( N \tilde{\nabla} f^\star(\mu^*) - \sum_{i=1}^{N} g_i(\check{x}_i) \right) \leq M \delta_p. \qquad (8)$$

*Proof.* Let $\mu^*$ be an optimal multiplier to the dual problem (D), i.e., $\mu^* \in \text{argmax } Q(\mu)$. Define $\check{\mathcal{X}}^*$ to be the set of optimal solutions to the dual problem given by

$$\check{\mathcal{X}}^* = \underset{x \in \mathcal{X}}{\text{argmin}} \left\{ (\mu^*)^T \left( \sum_{i=1}^N g_i(x_i) \right) - f^\star(\mu^*) + \sum_{i=1}^N p_i(x_i) \right\}.$$

Since the Langrangian function is decomposable with respect to $x_1, \dots, x_N$, we can rewrite $\check{\mathcal{X}}^*$ as the Cartesian product of many simpler sets, $\check{\mathcal{X}}^* = \prod_{i=1}^N \check{\mathcal{X}}_i^*$, where we define

$$\check{\mathcal{X}}_i^* = \underset{x_i \in \mathcal{X}_i}{\text{argmin}} \left\{ (\mu^*)^T g_i(x_i) + p_i(x_i) \right\}.$$

Since $\mathcal{X}_i$ is compact and $g_i, p_i$ are continuous functions, the set $\check{\mathcal{X}}_i^*$ is nonempty and compact (but not necessarily convex).

By using the definition of $\mu^*$ and the concavity of $Q(\mu)$, we have $0 \in \partial Q(\mu^*)$. Then, by applying Lemma 3, we obtain

$$0 \in \partial Q(\mu^*) = \text{conv} \left\{ \sum_{i=1}^N g_i(x_i) - y \;\middle|\; x \in \check{\mathcal{X}}^*, y \in \partial f^\star(\mu^*) \right\}.$$

Since $\partial f^\star$ is convex, it follows that there exists a subgradient $\tilde{\nabla} f^\star(\mu^*) \in \partial f^\star(\mu^*)$ such that

$$\begin{aligned} N\tilde{\nabla} f^\star(\mu^*) &\in \text{conv} \left\{ \sum_{i=1}^N g_i(x_i) \;\middle|\; x \in \check{\mathcal{X}}^* \right\} \\ &= \text{conv} \left\{ \sum_{i=1}^N g_i(x_i) \;\middle|\; x_i \in \check{\mathcal{X}}_i^*, i = 1, \dots, N \right\} \\ &= \text{conv} \left\{ \mathcal{G}_1 + \dots + \mathcal{G}_N \right\}, \end{aligned}$$

where we define

$$\mathcal{G}_i = \{ g_i(x_i) \mid x_i \in \check{\mathcal{X}}_i^* \}, \qquad i = 1, \dots, N.$$

By using the Shapley-Folkman Lemma 1, there exists $y_1, \dots, y_N$ such that

$$N\tilde{\nabla} f^\star(\mu^*) = y_1 + \dots + y_N,$$

where among $y_1, \dots, y_N$, at most $M$ out of them satisfy $y_i \in \text{conv}(\mathcal{G}_i)/\mathcal{G}_i$ and the rest satisfy $y_i \in \mathcal{G}_i$.

Now let us construct a solution $\check{x} \in \check{\mathcal{X}}^*$ that satisfies Eq. (8). The construction is index by index. There are two cases:

- Consider an index $i \in \{1, \dots, N\}$ such that $y_i \in \mathcal{G}_i$, we take $\check{x}_i \in \check{\mathcal{X}}_i^*$ and $z_i \in \Re$ to be such that

  $$y_i = g_i(\check{x}_i), \qquad z_i = p_i(\check{x}_i).$$

  This case happens at least $N - M$ times.

- Consider an index $i \in \{1, \dots, N\}$ such that $y_i \in \text{conv}(\mathcal{G}_i)/\mathcal{G}_i$. We let $z_i$ be such that $(z_i, y_i) \in \text{conv}(\check{\mathcal{W}}_i)/\check{\mathcal{W}}_i$, where $\check{\mathcal{W}}_i = \{(p_i(x), g_i(x)) \mid x \in \check{\mathcal{X}}_i^*\}$. Then we take $\check{x}_i$ to be

  $$\check{x}_i \in \underset{x_i \in \check{\mathcal{X}}_i^*}{\text{argmin}} \|y_i - g_i(x_i)\|^2, \qquad (9)$$

  so $g_i(\check{x}_i)$ and $y_i$ form a line segment in $\text{conv}(\mathcal{G}_i)/\mathcal{G}_i$. Such $\check{x}_i$ exists because $\check{\mathcal{X}}_i^*$ is nonempty and compact and $g_i, p_i$ are continuous. We see that the two points $(z_i, y_i)$ and

$(g_i(\check{x}_i), p_i(\check{x}_i))$ form a line segment in $\text{conv}(\check{\mathcal{W}}_i)/\check{\mathcal{W}}_i$. We claim that these two points also form a line segment in $\text{conv}(\mathcal{W}_i)/\mathcal{W}_i$. If it is not true, there would exist another $\hat{x}_i \in \mathcal{X}_i$ such that $(p_i(\hat{x}_i), g_i(\hat{x}_i))$ lies on the line segment between $(z_i, y_i)$ and $(g_i(\check{x}_i), p_i(\check{x}_i))$. Such a point $\hat{x}_i$ is also a minimizer of $(\mu^*)^T g_i(x_i) + p_i(x_i)$. So we have $\hat{x}_i \in \check{\mathcal{X}}_i^*$ and $(p_i(\hat{x}_i), g_i(\hat{x}_i)) \in \check{\mathcal{W}}_i$, which conflicts with the fact that $\big((z_i, y_i), (g_i(\check{x}_i), p_i(\check{x}_i))\big)$ is a line segment outside $\check{\mathcal{W}}_i$. Therefore, the two points $(z_i, y_i)$ and $(g_i(\check{x}_i), p_i(\check{x}_i))$ form a line segment in $\text{conv}(\mathcal{W}_i)/\mathcal{W}_i$. By using the definition of convexity gap, we have

$$\|y_i - g(\check{x}_i)\| \leq \rho_g(\mathcal{W}_i), \qquad \|z_i - p_i(\check{x}_i)\| \leq \rho_p(\mathcal{W}_i).$$

This case happens at most $M$ times.

So far we have constructed a solution $\check{x} \in \check{\mathcal{X}}^*$ to the dual problem (D). We can show that

$$\begin{aligned}
\|N\nabla f^\star(\mu^*) - \sum_{i=1}^N g_i(\check{x}_i)\| &= \|y_1 + \cdots + y_N - \sum_{i=1}^N g_i(\check{x}_i)\| \\
&\leq \|y_1 - g(\check{x}_1)\| + \cdots + \|y_N - g(\check{x}_N)\| \\
&\leq M \max_{i=1,\ldots,N} \rho_g(\mathcal{W}_i) \\
&= M\delta_g.
\end{aligned}$$

By using the definition of $\check{\mathcal{X}}_i^*$ and $\check{\mathcal{W}}_i = \{(g_i(x), p_i(x)) \mid x \in \check{\mathcal{X}}_i^*\}$, we see that ${\mu^*}^T g(x_i) + p(x_i)$ takes constant value for all $x_i \in \check{\mathcal{X}}_i^*$. Since $(z_i, y_i) \in \text{conv}(\check{\mathcal{W}}_i)/\check{\mathcal{W}}_i$ and $\check{x}_i \in \check{\mathcal{X}}_i^*$, we have

$${\mu^*}^T y_i + z_i = {\mu^*}^T g(\check{x}_i) + p(\check{x}_i), \qquad i = 1, \ldots, N.$$

We use the preceding equality and obtain that

$$\begin{aligned}
(\mu^*)^T \left( N\tilde{\nabla} f^\star(\mu^*) - \sum_{i=1}^N g_i(\check{x}_i) \right) &= (\mu^*)^T \left( \sum_{i=1}^N y_i - \sum_{i=1}^N g_i(\check{x}_i) \right) \\
&= \sum_{i=1}^N (p_i(\check{x}_i) - z_i) \\
&\leq M\delta_p.
\end{aligned}$$

■

Theorem 1 shows that the near-optimality condition (8) can be achieved by constructing a particular solution $\check{x}$ to the dual problem. We remark that the construction can be simplified to achieve a slightly relaxed error bound. For $i \in \{1, \ldots, N\}$ such that $z_i \in \text{conv}(\mathcal{G}_i)/\mathcal{G}_i$, we may omit the nonconvex projection step (9) and choose an arbitrary best response $\check{x}_i \in \check{\mathcal{X}}_i^*$ instead. As a result, we obtain a slightly relaxed near-optimality condition:

$$\left\| N\tilde{\nabla} f^\star(\mu^*) - \sum_{i=1}^N g_i(\check{x}_i) \right\| \leq M\gamma_g, \qquad (\mu^*)^T \left( N\tilde{\nabla} f^\star(\mu^*) - \sum_{i=1}^N g_i(\check{x}_i) \right) \leq M\gamma_p, \tag{10}$$

where $\gamma_g$ and $\gamma_p$ are the maximum radiuses of the sets $\{g_i(x) \mid x \in \mathcal{X}_i\}$ and $\{p_i(x) \mid x \in \mathcal{X}_i\}$, respectively. The relaxed bound uses the maximum radiuses of the sets $\mathcal{W}_1, \ldots, \mathcal{W}_N$ rather than the maximum convexity gap. In the case where $p_i, g_i, \mathcal{X}_i$ are uniformly bounded across all participants, the relaxed condition (10) has a similar asymptotic property as the one given in Theorem 1. If we normalize the error bounds by $1/F^* = \mathcal{O}(1/N)$, both error bounds diminish to zero as $N/M \to \infty$.

By leveraging the convexification phenomenon, Theorem 1 states that although there does not exist $x \in \check{\mathcal{X}}^*$ such that $\tilde{\nabla}_\mu L(x, \mu^*) = 0$, there exists a solution $\check{x} \in \check{\mathcal{X}}^*$ such that $\tilde{\nabla}_\mu L(\check{x}, \mu^*) \approx 0$. In other words, it is possible to find some $\check{x}$ such that strong duality is "nearly" satisfied. In what follows, we will use this fact to show that $\check{x}$ is indeed an approximate optimum to the primal problem. Our first duality gap result concerns the case where the penalty function $f$ is sufficiently smooth.

**Theorem 2 (Duality Gap of Smooth Penalized Problems)** *Let $f$ be $\beta$-strongly smooth. Then there exists a solution $\check{x}$ to problem* (D) *that is a nearly optimal solution to problem* (1) *such that*

$$F^* \le F(\check{x}) \le F^* + 2\beta \left(M\delta_g\right)^2.$$

*Proof.* Given that $f$ is $\beta$-strongly smooth, we use the duality between conjugate functions and obtain that $f^\star$ is $\frac{1}{\beta}$-strongly convex (see Lemma 2). The dual function $Q(\mu)$ is the infimum of multiple $\frac{1}{\beta}$-strongly concave functions, therefore it is also $\frac{1}{\beta}$-strongly concave. It follows that there exists at least one optimal multiplier $\mu^* \in \text{argmin}\ Q(\mu)$. We apply Theorem 1 and obtain that there exists a solution $\check{x}$ to problem (D) such that $\left\| N\tilde{\nabla} f^\star(\mu^*) - \sum_{i=1}^N g_i(\check{x}_i) \right\| \le M\delta_g$, where $\tilde{\nabla} f^\star(\mu^*)$ is a subgradient of $f^\star$ at $\mu^*$.

In what follows, we show that $\check{x}$ satisfies the duality gap stated in the theorem. We apply $\check{x}$ to the primal problem (1) and obtain

$$F(\check{x}) = \sum_{i=1}^N p_i(\check{x}_i) + \sup_{\mu \in \Re^M} \left\{ \mu^T \left( \sum_{i=1}^N g_i(\check{x}_i) \right) - f^\star(\mu) \right\}.$$

We also apply $\check{x}$ to the dual problem (D) and obtain

$$Q^* = \sum_{i=1}^N p_i(\check{x}_i) + (\mu^*)^T \left( \sum_{i=1}^N g_i(\check{x}_i) \right) - f^\star(\mu^*).$$

Next, we compare the values of $F(\check{x})$ and $Q^*$.

Let us define the function $h(\cdot)$ and its maximizer $\hat{\mu}$ to be

$$h(\mu) = \mu^T \left( \sum_{i=1}^N g_i(\check{x}_i) \right) - f^\star(\mu), \qquad \hat{\mu} = \underset{\mu \in \Re^M}{\text{argmax}}\ h(\mu).$$

We can see that $F(\check{x}) - Q^* = h(\hat{\mu}) - h(\mu^*)$ and $\sum_{i=1}^N g_i(\check{x}_i) - \tilde{\nabla} f^\star(\mu^*) \in \partial h(\mu^*)$. On one hand, by using the concavity of $h$, we have

$$h(\hat{\mu}) - h(\mu^*) \le \left( \sum_{i=1}^N g_i(\check{x}_i) - \tilde{\nabla} f^\star(\mu^*) \right)^T (\hat{\mu} - \mu^*) \le \left\| \sum_{i=1}^N g_i(\check{x}_i) - \tilde{\nabla} f^\star(\mu^*) \right\| \|\hat{\mu} - \mu^*\|.$$

By applying Theorem 1, we obtain that $h(\hat{\mu}) - h(\mu^*) \le M\delta_g \|\hat{\mu} - \mu^*\|$. On the other hand, by using the $1/\beta$-strong convexity of $f^\star$ and $-h$ and using the optimality of $\hat{\mu}$, we further obtain $h(\hat{\mu}) - h(\mu^*) \ge \frac{1}{2\beta}\|\hat{\mu} - \mu^*\|^2$. Combining the preceding two relations, we obtain $\|\hat{\mu} - \mu^*\| \le 2\beta M\delta_g$. As a result, we have

$$F(\check{x}) - Q^* = h(\hat{\mu}) - h(\mu^*) \le M\delta_g \|\hat{\mu} - \mu^*\| \le 2\beta \left(M\delta_g\right)^2.$$

Finally, by using the weak duality, we have $F(\check{x}) - F^* \le F(\check{x}) - Q^* \le 2\beta \left(M\delta_g\right)^2$. ■

Next, we focus on the constrained version of problem (1) given by

$$\text{minimize } \sum_{i=1}^{N} p_i(x_i), \qquad \text{subject to } \sum_{i=1}^{N} g_i(x_i) \in \mathcal{A}, \quad x_i \in \mathcal{X}_i, \quad i = 1, \ldots, N. \tag{11}$$

Problem (11) can be viewed as a special case of problem (1) by letting $f(\cdot)$ be the indicator function of the set $\mathcal{A}$, i.e.,

$$f(y) = \begin{cases} 0 & y \in \mathcal{A}, \\ +\infty & y \notin \mathcal{A}. \end{cases}$$

When $\mathcal{A}$ is convex and nonempty, the convex conjugate of $f$ and its subdifferential is

$$f^\star(\mu) = \sup_{y\in\mathcal{A}} \mu^T y, \qquad \partial f^\star(\mu) = \operatorname*{argmax}_{y\in\mathcal{A}} \mu^T y.$$

It is easy to see that $\partial f^\star(\mu) \subset \mathcal{A}$ for all $\mu$. We have the following duality gap results.

**Theorem 3 (Duality Gap of Constrained Problems)** *Assume that $\mathcal{A}$ is a convex set and there exists $x \in \mathcal{X}$ such that $\sum_{i=1}^N g_i(x_i)$ is an interior point of $\mathcal{A}$. Then there exists a dual solution $\breve{x}$ to problem* (D) *that is a nearly feasible optimal solution to problem* (11) *such that*

$$\sum_{i=1}^{N} g_i(\breve{x}_i) \in \mathcal{A} + \mathcal{B}(0, M\delta_g), \qquad F^* \leq \sum_{i=1}^{N} p_i(\breve{x}_i) \leq F^* + M\delta_p,$$

*where $\mathcal{B}(0, \epsilon)$ denotes the closed ball centered at 0 with radius $\epsilon$.*

*Proof.* We claim that there exists at least one optimal multiplier to the dual problem. Since the primal problem is feasible, we have $F^* < \infty$. By using the weak duality, we have $Q^* \leq F^* < \infty$. Since there is at least one $\mu$ such that $Q(\mu) > -\infty$, we obtain that $Q^*$ is a finite value. Then there exists a sequence $\{\mu_k\}$ such that $\lim_{k\to\infty} Q(\mu_k) = Q^*$. We assume to the contrary that $\{\mu_k\}$ is unbounded. Let $\{\frac{\mu_k}{\|\mu_k\|}\}$ be the projections of $\{\mu_k\}$ on the unit ball, so it has at least one convergent subsequence. We assume without loss of generality that $\{\frac{\mu_k}{\|\mu_k\|}\}$ converges to some vector $d$ on the unit ball.

Let $\bar{\mathcal{A}}$ be a compact subset of $\mathcal{A}$ that contains some $\sum_{i=1}^N g_i(x_i)$, where $x \in \mathcal{X}$, as an interior point. On one hand, we have

$$\begin{aligned}
\lim_{k\to\infty} \frac{Q(\mu_k)}{\|\mu_k\|} &= \lim_{k\to\infty} \inf_{x\in\mathcal{X}, y\in\mathcal{A}} \frac{1}{\|\mu_k\|} \sum_{i=1}^{N} p_i(x_i) + \left( (\frac{\mu_k}{\|\mu_k\|} - d) + d \right)^T (\sum_{i=1}^{N} g_i(x_i) - y) \\
&\leq \lim_{k\to\infty} \inf_{x\in\mathcal{X}, y\in\bar{\mathcal{A}}} \frac{1}{\|\mu_k\|} \sum_{i=1}^{N} p_i(x_i) + \left( (\frac{\mu_k}{\|\mu_k\|} - d) + d \right)^T (\sum_{i=1}^{N} g_i(x_i) - y) \\
&= \inf_{x\in\mathcal{X}, y\in\bar{\mathcal{A}}} d^T (\sum_{i=1}^{N} g_i(x_i) - y),
\end{aligned}$$

where the inequality uses the fact $\bar{\mathcal{A}} \subset \mathcal{A}$, and last equality uses the boundedness of $\sum_{i=1}^N p_i(x_i)$ and $\sum_{i=1}^N g_i(x_i) - y$ over the compact set $\{x \in \mathcal{X}, y \in \bar{\mathcal{A}}\}$ and the facts $\frac{1}{\|\mu_k\|} \downarrow 0$ and $\frac{\mu_k}{\|\mu_k\|} - d \downarrow 0$. On the other hand, we have $\lim_{k\to\infty} \frac{Q(\mu_k)}{\|\mu_k\|} = 0$ because $Q(\mu^k) \to Q^*$. Now we have $\inf_{x\in\mathcal{X}, y\in\bar{\mathcal{A}}} d^T (\sum_{i=1}^N g_i(x_i) - y) \geq 0$. By using the interior point feasibility condition, we know that $(\sum_{i=1}^N g_i(x_i) - y)$ can take arbitrary value in a small open ball centered at the origin. This implies that $d = 0$, which contradicts the fact that $d$ is on the boundary of the unit ball. As a result, $\{\mu_k\}$ is a bounded sequence so it has at least one limit point $\mu^*$ attaining the optimal dual value $Q^*$.

Now that there exists an optimal multiplier $\mu^*$, we may apply Theorem 1. It follows that there exists $\check{x} \in \check{\mathcal{X}}^*$, a best response solution to $\mu^*$, such that

$$(\mu^*)^T \left( N\tilde{\nabla} f^\star(\mu^*) - \sum_{i=1}^N g_i(\check{x}_i) \right) \le M\delta_p, \qquad \|N\tilde{\nabla} f^\star(\mu^*) - \sum_{i=1}^N g_i(\check{x}_i)\| \le M\delta_g, \tag{12}$$

where $\tilde{\nabla} f^\star(\mu^*)$ is a subgradient of $f^\star$. Now it remains to show that $\check{x}$ is nearly feasible-optimal.

First, we prove the near-feasibility. Note that $\tilde{\nabla} f^\star(\mu^*) \in \mathcal{A}$ by using the property of $f^\star$. So we apply Eq. (12) and obtain

$$\sum_{i=1}^N g_i(\check{x}_i) \in \mathcal{A} + \mathcal{B}(0, M\rho_g).$$

Second, we prove the near-optimality. Note that since $f$ is the indicator function of $\mathcal{A}$, we have $f^\star(\mu) = \sup_y y^T\mu$ and $\tilde{\nabla} f^\star(\mu^*) \in \underset{y\in\mathcal{A}}{\operatorname{argmax}}\, y^T\mu^*$, therefore

$$f^\star(\mu^*) = (\mu^*)^T \tilde{\nabla} f^\star(\mu^*).$$

By applying the preceding equality and the inequality (12), we have

$$\begin{aligned} Q^* &= \sum_{i=1}^N p_i(\check{x}_i) + (\mu^*)^T \left( \sum_{i=1}^N g_i(\check{x}_i) \right) - f^\star(\mu^*) \\ &= \sum_{i=1}^N p_i(\check{x}_i) + (\mu^*)^T \left( \sum_{i=1}^N g_i(\check{x}_i) - \tilde{\nabla} f^\star(\mu^*) \right) \\ &\ge \sum_{i=1}^N p_i(\check{x}_i) - M\delta_p. \end{aligned}$$

Finally, we apply the weak duality and obtain $\sum_{i=1}^N p_i(\check{x}_i) \le Q^* + M\delta_p \le F^* + M\delta_p$. ■

We remark that the feasibility violation is often inevitable. When the radius of $\mathcal{A}$ is smaller than the duality gap, there is no guarantee that the constructed dual solution lies in the small feasible set. When $\mathcal{A}$ is sufficiently large, we may consider a modified problem in which the $\mathcal{A}$ is replaced with its subset. We let $\mathcal{A}_\epsilon$ be a subset set such that $\mathcal{A}_\epsilon + \mathcal{B}(0,\epsilon) \subset \mathcal{A}$, and solve the new problem instead. By choosing $\epsilon$ to be the duality gap and applying Theorem 3, we can show that the dual solution to the modified problem is a feasible solution to the original problem.

## 3.3 Asymptotic Price of Anarchy

We analyze the asymptotic price of anarchy as the cooperative optimization problem (1) scales up. Take the penalty case, for example, and let the smoothness parameter $\beta$ be a fixed constant value. We need additional assumption on the asymptotic behavior of the optimal value $F^*$. A reasonable assumption is that $\sum_{i=1}^N p_i(x_i)$ and $\sum_{i=1}^N g_i(x_i)$ scale up on the order of $\mathcal{O}(N)$, thus there exists $c > 0$ such that

$$F^* = \min_{x\in\mathcal{X}} \sum_{i=1}^N p_i(x_i) + f\left( \sum_{i=1}^N g_i(x_i) \right) \ge c \cdot (N + \beta N^2).$$

Suppose that the functions $p_i, g_i$ are uniformly bounded across all participants as the number of participants $N$ increases. In this case, the maximum nonconvexity gaps $\delta_p, \delta_g$ remain bounded as $N$ increases. The price of anarchy satisfies

$$\frac{F(\check{x}) - F^*}{F^*} \leq \mathcal{O}\left(\frac{\beta M^2}{N + \beta N^2}\right) \to 0, \qquad \text{as } N/M \to \infty.$$

Suppose that the preference of the participants are drawn independently from an unbounded distribution. Under some distributional assumptions, we conjecture that the maximum convexity gaps satisfy $\delta_p = \mathcal{O}(\log N)$ and $\delta_g = \mathcal{O}(\log N)$ with high probability. Then we can show that

$$\mathbf{P}\left(\frac{F(\check{x}) - F^*}{F^*} \leq \mathcal{O}\left(\frac{\beta M^2 \log N}{N + \beta N^2}\right)\right) \to 1, \qquad \text{as } N^2/(M^2 \log N) \to \infty.$$

The vanishing price of anarchy essentially requires that $N/M \to \infty$, i.e., the number of participants be substantially larger than the number of common factors. In contrast, if $N$ and $M$ are on the same order, the convexification effect becomes very minor, resulting in a potentially unbounded price of anarchy.

We have performed numerical experiments on the asymptotic price of anarchy. In Section 6, we present numerical results on randomly generated instances of problem (1). We compute and plot the samples of price of anarchy in Figure 3. The observed mean price of anarchy is inversely related to the number $N$, where $M, \beta$ are kept constant. This validates our theory. To summarize, the price of anarchy asymptotically converges to zero as the multi-agent problem scales up in proper ways. As long as the number of participants is much larger than the number of common goods, there exists a fair price that achieves a nearly zero price of anarchy.

# 4 The Role of Coordination

In this section, we aim to understand the role of coordination among multiple participants. From the computational perspective, we study how to find an approximate optimum $\check{x}$ when the optimal multiplier $\mu^*$ is known. Identifying such a solution out of many best responses can be viewed as coordinating the participants in a central manner. We discover that the coordination problem is essentially an approximate projection problem. Moreover, we show how to solve it by constructing a linear feasibility problem and finding an extreme point solution. We use examples to illustrate the necessity of coordination, without which the price of anarchy can be disastrously high.

## 4.1 Finding the Approximate Global Optimum When $\mu^*$ Is Given

We have proved the existence of a solution to the dual problem $\check{x} \in \check{\mathcal{X}}^*$ that is nearly primal optimal. However, the set of all solutions to the dual problem $\check{\mathcal{X}}^*$ may contain as many as $\mathcal{O}(2^N)$ number of solutions, each of them being a best response to the multiplier $\mu^*$. Now we consider how to identify a good solution $\check{x}$ out of the many candidates in $\check{\mathcal{X}}^*$. In other words, we need a coordination mechanism to select a decision $\check{x}_i$ for each participant $i$ from its best responses.

Suppose that the optimal multiplier $\mu^*$ is given; the remaining problem is to find a solution $\check{x}$ that satisfies the duality gap inequalities given in the former theory. Identifying an approximate optimum is equivalent to finding a point $\check{x} \in \check{\mathcal{X}}^*$ and a subgradient $\tilde{\nabla} f^\star(\mu^*) \in \partial f^\star(\mu^*)$ such that

$$\tilde{\nabla}_\mu L(\check{x}, \mu) = \sum_{i=1}^{N} g_i(\check{x}_i) - \tilde{\nabla} f^\star(\mu^*) \approx 0.$$

This can be viewed as a projection problem, in which we try to find $y \in \sum_{i=1}^N \{g_i(x_i) \mid x_i \in \check{\mathcal{X}}_i^*\}$ such that the distance between $y$ and the subdifferential $\partial f(\mu^*)$ is minimized.

Consider a more general nonconvex projection problem. Suppose that we are given the sets $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$ and $\mathcal{S}$. The projection problem is

$$\begin{aligned} &\text{minimize } \|\bar{y} - (y_1 + \cdots + y_n)\| \\ &\text{subject to } \bar{y} \in \mathcal{S}, y_1 \in \mathcal{Y}_1, \ldots, y_N \in \mathcal{Y}_N. \end{aligned} \tag{13}$$

In Algorithm 1, we develop a computational method to approximately solve problem (13). It has three steps: solving a quadratic optimization problem, finding the extreme point solution to a linear feasibility problem, and recovering the approximate projection from the extreme point.

---

**Algorithm 1** Approximate Projection onto Sums of Nonconvex Sets

---

**Input:** $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$, $S$

1: Let $A^{(i)}$ be the matrix consisting of column vectors in $\mathcal{Y}_i$, $i = 1, \ldots, N$.

2: The first step is to let

$$b^* \in \operatorname*{argmin}_{b \in S} \left\{ \|b - \sum_{i=1}^N A^{(i)} z^{(i)}\|^2 \;\middle|\; \sum_{i=1}^N A^{(i)} z^{(i)} = b,\ (e^{(i)})^T z^{(i)} = 1,\ z^{(i)} \geq 0 \right\},$$

where $e^{(i)}$ is the vector with all 1's whose dimension is equal to the column dimension of $A^{(i)}$.

3: The second step is to find a basic feasible solution $z$ to the linear feasibility problem

$$\sum_{i=1}^N A^{(i)} z^{(i)} = b^*, \qquad (e^{(i)})^T z^{(i)} = 1, \qquad z^{(i)} \geq 0, \qquad i = 1, \ldots, N.$$

4: The third step is to let

$$y_i = \begin{cases} A^{(i)} z^{(i)} & \text{if } z^{(i)} \text{ is an integer vector,} \\ \operatorname*{argmin}_{y \in \mathcal{Y}_i} \|y - A^{(i)} z^{(i)}\| & \text{otherwise.} \end{cases}$$

**Output:** $y = (y_1, \ldots, y_N)$

---

Next we show that Algorithm 1 indeed finds an approximate solution to the projection problem (13). Furthermore, we can apply the algorithm to the optimization problem (1) as a coordination procedure. When the optimal multiplier $\mu^*$ is given, we can use Algorithm 1 to identify a solution $\check{x}$ that achieves the small duality gap.

**Theorem 4** *(a) Let $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$ be nonconvex sets and let $\mathcal{S}$ be convex. Then the vector $y = (y_1, \ldots, y_N)$ generated by Algorithm 1 satisfies*

$$\begin{aligned} \min_{\bar{y} \in \mathcal{S}} \|\bar{y} - (y_1 + \cdots + y_n)\| &\leq \min_{\bar{y} \in S, y \in \operatorname{conv}(\mathcal{Y}_1 + \cdots + \mathcal{Y}_n)} \|\bar{y} - y\| + M \max_{i=1,\ldots,N} \rho(\mathcal{Y}_i) \\ &\leq \min_{\bar{y} \in S, y \in \mathcal{Y}_1 + \cdots + \mathcal{Y}_n} \|\bar{y} - y\| + M \max_{i=1,\ldots,N} \rho(\mathcal{Y}_i). \end{aligned}$$

*(b) Under the assumptions of Theorem 2 or 3, let $y$ be generated by Algorithm 1 with input*

$$\big(\{g_1(x_1) \mid x_1 \in \check{\mathcal{X}}_1^*\}, \ldots, \{g_N(x_N) \mid x_N \in \check{\mathcal{X}}_N^*\}, \partial f^\star(\mu^*)\big)$$

*and let $\check{x}$ be such that $g_i(\check{x}_i) = \hat{y}_i$ for $i = 1, \ldots, N$. Then $\check{x}$ is an approximate optimum to problem (1) that satisfies the inequalities given by Theorem 2 or 3.*

*Proof.* (a) According to the first step, $b^* \in \mathcal{S}$ achieves the minimal distance between $\mathcal{S}$ and $\sum_{i=1}^N \mathrm{conv}(\mathcal{Y}_i)$. Let $\bar{y}$ be the projection of $b^*$ on $\sum_{i=1}^N \mathrm{conv}(\mathcal{Y}_i)$. By following an analysis similar to that of Lemma 1 and Theorem 1, we see that the second and third steps generate $y$ such that $\|y - b^*\| \leq \|\bar{y} - b^*\| + M \max_{i=1,\ldots,N} \rho(\mathcal{Y}_i) = \mathrm{dist}(\mathcal{S}, \sum_{i=1}^N \mathrm{conv}(\mathcal{Y}_i)) + M \max_{i=1,\ldots,N} \rho(\mathcal{Y}_i) \leq \mathrm{dist}(\mathcal{S}, \mathrm{conv}(\sum_{i=1}^N \mathcal{Y}_i)) + M \max_{i=1,\ldots,N} \rho(\mathcal{Y}_i)$, where the last inequality is due to the fact $\sum_{i=1}^N \mathcal{Y}_i \subset \mathrm{conv}(\sum_{i=1}^N \mathcal{Y}_i) \subset \sum_{i=1}^N \mathrm{conv}(\mathcal{Y}_i)$.
(b) It follows from part (a) that the constructed solution $\check{x}$ satisfies the approximate optimality condition (8). Then the analysis of Theorem 2 and 3 follow directly. ■

Let us comment on the implementation and computational complexity of Algorithm 1. We suppose that $\mathcal{Y}_1, \ldots, \mathcal{Y}_N$ are finite sets. When they are not, we need to use finite discretization as an approximation. In the first step, computing $b^*$ requires solving a least square problem over linear inequality and convex constraints. This can be solved efficiently using convex optimization solvers. When we apply Algorithm 1 to the coordinative optimization problem, we have $\mathcal{S} = \partial f^\star(\mu^*)$. If $f$ is a strictly convex function, the conjugate function $f^\star$ is differentiable. Then we can omit this step and simply take $b^*$ to be the unique gradient $\nabla f^\star(\mu^*)$.

In the second step, we need to find a basic feasible solution to a linear programming problem. To do this, one approach is to apply the simplex method and directly obtain an optimal basic feasible solution. Another approach is to apply the interior point algorithm [26] that finds a basic feasible solution to any linear programming problem in polynomial time.

In the third step, we need to perform nonconvex projection $\underset{y \in \mathcal{Y}_i}{\mathrm{argmin}} \|y - A^{(i)} z^{(i)}\|$ for at most $M$ times. This step can be skipped if we are willing to relax the bounds. In particular, we can skip it and simply choose arbitrary $y_i \in \mathcal{Y}_i$ if $z^{(i)}$ is not an integer solution. Then we obtain analogous results where the maximal convexity gap is replaced with the maximal radius. When applying it to the coordinative optimization, we obtain duality gap results similar to that of Theorems 2 and 3 with $\delta_p, \delta_g$ replaced by $\gamma_p, \gamma_g$. In this way, the constructed solution is still an approximate optimum and the asymptotic price of anarchy remains on the same order.

We have presented a computational method for finding an approximate optimum out of many best response solutions, as long as $\mu^*, \check{\mathcal{X}}^*$ are given. It involves examining all candidate best responses in a central manner. The key step is to find an extreme point solution to a linear feasibility problem. It can be viewed as a form of coordination that selects a solution for each user in order to optimize the overall objective.

## 4.2 Price of Anarchy Without Coordination

Recall that our main motivation is to answer the question: Is there a fair price $\mu$ such that a global social optimum can be attained, as long as each participant reacts optimally to his own problem? So far, we have shown the dual optimal multiplier $\mu^*$ acts as a "nearly fair" price vector. We have shown the existence of a solution $\check{x} = (\check{x}_1, \ldots, \check{x}_N)$ to the dual problem that is nearly primal optimal. Such a solution $\check{x}$ consists of individual best responses to the nearly fair price $\mu^*$. As the number of participants increases, the normalized duality gap vanishes to zero, implying that the dual solution becomes asymptotically primal optimal. This provides a partial answer to the question raised earlier.

Due to the lack of convexity, individual users may have multiple optimal actions to choose from even if they face the fair price $\mu^*$. Any combination of individual decisions constitutes an optimal solution to the dual problem - there are many of them. It is not enough to know that one out of the large number of solutions achieves the small duality gap. We are also interested in how other solutions perform. Let us consider two simple examples.

**Example 1 (Efficiency loss without coordination)** Consider the unconstrained optimization problem:

$$\min_{x_1,\dots,x_N} (x_1 + \cdots + x_N)^2 + |x_1^2 - 1| + \cdots + |x_N^2 - 1|.$$

The optimal multiplier to the dual problem is $\mu^* = 0$ and the corresponding solution sets are $\check{\mathcal{X}}_i^* = \{-1, 1\}$. Each user is indifferent between two equally optimal decisions $-1$ and 1. In the case without coordination, the objective value can be as large as $\mathcal{O}(N^2)$ while the optimal value is 1 if $N$ is odd. In this case, without coordinating $x_1, \dots, x_N$ to achieve the duality gap, the worst-case price of anarchy can be as large as $\mathcal{O}(N^2)$ and increases to infinity as $N$ increases.

**Example 2 (Violation of feasibility without coordination)** Consider the constrained optimization problem:

$$\begin{aligned} &\min |x_1^2 - 1| + \cdots + |x_N^2 - 1| \\ &\text{subject to } |x_1 + \cdots + x_N| \le N/2. \end{aligned} \tag{14}$$

Similar to the earlier example, the optimal multiplier to the dual problem is $\mu^* = 0$ and the corresponding solution sets are $\check{\mathcal{X}}_i^* = \{-1, 1\}$. Again, each user is indifferent between two optimal decisions $-1$ and 1. Without coordination, the worst-case constraint violation can be as large as $N/2$. In this case, the price of anarchy does not improve as $N$ increases.

As demonstrated in the examples, the price of anarchy can be far greater than the small normalized duality gap, if we let users choose their best responses arbitrarily. Although there exists a nearly optimal solution $\check{x}$ consisting of best responses to $\mu^*$, there is no guarantee to reach it without any coordination. Without convexity, a fully autonomous system will not work to its best. In short, coordination is critical to the success of cooperative nonconvex optimization.

# 5 Dynamic Convergence to Approximate Social Optimum

In this section, we show how to achieve the approximate social optimum to problem (1) via a duality-based method. The method relies on best responses from individual participants. Even without convexity, the dynamic convergence to an approximate optimum can still be achieved, as long as the participants are cooperative and a central decision maker properly coordinates the individuals' behaviors. We provide a rigorous convergence and rate of convergence analysis, and we comment on the algorithm's complexity.

## 5.1 A Coordinative Best Response Algorithm

We propose a distributed coordinative algorithm for solving problem (1). The algorithm is based on a cutting plane method applied to the dual problem $\max_\mu Q(\mu)$. It is known that the cutting plane method works for the dual problem, because $Q(\cdot)$ is concave, regardless of the nonconvexity of the primal problem. However, even if an optimal multiplier $\mu^*$ is known, arbitrary best responses from individual participants may result in a high price of anarchy. This requires the algorithm to coordinate the participants in order to achieve an approximate social optimum. The coordination feature makes the proposed algorithm different from the traditional cutting plane method for convex optimization.

We present the coordinative dual algorithm in Algorithm 2. For simplicity, we use $\langle \cdot, \cdot \rangle$ to denote the inner product between two vectors, and we use $\epsilon\text{-argmin}_x f(x)$ to denote the set $\{x \mid f(x) \le \inf_z f(z) + \epsilon\}$.

Algorithm 2 involves a central decision maker that coordinates the behaviors of multiple participants. In each iteration, the coordinator updates the multiplier and an upper approximation of the dual function. Moreover, the coordinator collects the sets of selfish best responses from

**Algorithm 2** Distributed Coordinative Nonconvex Optimization via Dual Cutting Plane

**Input:** $\mu^0 \in \Re^M, R \in \Re, \beta \in \Re$.

1: **repeat**

2: Given the price vector $\mu^t$, each participant $i = 1, \dots, N$ finds a best response

$$x_i^t \in \operatorname{argmin}\{p_i(x_i) + \langle \mu^t, g_i(x_i)\rangle \mid x_i \in \mathcal{X}_i\}.$$

3: Update the approximate dual problem by

$$Q_{t+1}(\mu) = \min_{k \le t} \left\{ \langle \mu, \bar{g}^k \rangle + \bar{p}^k - f^\star(\mu) \right\},$$

where $\bar{p}^t = \sum_{i=1}^N p_i(x_i^t), \bar{g}^t = \sum_{i=1}^N g_i(x_i^t)$.

4: Set the new test price and estimated gap as

$$\mu^{t+1} \in \operatorname*{argmax}_{\mu \in \Re^M} \; Q_{t+1}(\mu), \qquad \varepsilon_t = Q_t(\mu^t) - \max_{k \le t}\{\langle \mu^k, \bar{g}^k\rangle + \bar{p}^k - f^\star(\mu^k)\}.$$

5: **if** $\langle \mu^t, \bar{g}^t\rangle + \bar{p}^t - f^\star(\mu^t) > \max_{k \le t-1}\{\langle \mu^k, \bar{g}^k\rangle + \bar{p}^k - f^\star(\mu^k)\}$ **then**

6: Each participant $i = 1, \dots, N$ finds the $\xi_t$-optimal responses

$$\mathcal{X}_i^t = \xi_t\text{-}\operatorname{argmin}\{p_i(x_i) + \langle \mu^t, g_i(x_i)\rangle \mid x_i \in \mathcal{X}_i\}, \qquad \mathcal{G}_i^t = \{g_i(x_i) \mid x_i \in \mathcal{X}_i^t\},$$

where $\xi_t = R\sqrt{\beta \varepsilon_t}$.

7: Apply Algorithm 1 with input $(\mathcal{G}_1^t, \cdots, \mathcal{G}_N^t, \partial f^\star(\mu^t))$ and obtain $y^t = (y_1^t, \dots, y_n^t)$.

8: Set $\hat{\mu}^t = \mu^t$, $\hat{x}^t = (\hat{x}_1^t, \dots, \hat{x}_n^t)$ such that $g_i(\hat{x}_i^t) = y_i^t$ for $i = 1, \dots, N$.

9: **else**

10: Set $\hat{\mu}^t = \hat{\mu}^{t-1}$, $\hat{x}^t = \hat{x}^{t-1}$.

11: **end if**

12: **until** $\varepsilon_t \le \epsilon$.

all participants and selects a decision for each of them. The multiplier converges to the optimal multiplier, and the coordinated solution asymptotically converges into the set of approximate global optima. Steps 2-4 of Algorithm 2 are the dual updates and guarantee the convergence of the multiplier. Steps 5-10 are the coordination steps and select a nearly best solution out of the many candidates. The algorithm does not require any tuning stepsize, which makes it preferable to subgradient methods.

The proposed algorithm relies on individual participants to solve their own nonconvex problems. When the dimension $M$ is a fixed small value, we assume that each individual's nonconvex problem is small-scale and can be solved efficiently in constant time. Thus, it suffices to analyze the iteration complexity of Algorithm 2. We can prove the following convergence and rate of convergence results.

**Theorem 5** *Assume that $f$ is $\alpha$-strongly convex and $\beta$-strongly smooth. Let $R > 0$ be such that $\sup\{\|\sum_{i=1}^N g_i(x_i) - \sum_{i=1}^N g_i(y_i)\| \mid x, y \in \mathcal{X}\} \le R$. Then*

*(a) The multiplier $\hat{\mu}^t$ converges to $\mu^*$ as $t \to \infty$ and satisfies for all $t$ that*

$$Q^* - Q(\hat{\mu}^t) \le \frac{\beta R^2 + \alpha^{-2}\beta^2(Q^* - Q(\mu^0))}{t/2}.$$

*(b) Every limit point of $\hat{x}^t$ is an approximate global optimum and satisfies for all $t$ that*

$$F(\hat{x}^t) - F^* \leq \beta(M\delta_g)^2 + \beta R^2 \cdot \mathcal{O}\left(\frac{\alpha^{-2}\beta^2}{t} + \frac{N}{\sqrt{t}}\right).$$

*(c) With the stopping criterion $\varepsilon_t \leq \epsilon$, the algorithm terminates within $T = \frac{\beta R^2 + \alpha^{-2}\beta^2(Q^* - Q(\mu^0))}{\epsilon/2}$ number of iterations and satisfies*

$$F(\hat{x}^T) - F^* \leq \beta\left(M\delta_g\right)^2 + NR\sqrt{\beta\epsilon} + \frac{\beta^2}{\alpha^2}\epsilon.$$

Theorem 5 asserts that the nonconvex problem (1) can be solved efficiently up to a constant error using a dual method. The constant error is equal to the duality gap. Due to the nonconvexity, the primal convergence and dual convergence result from different mechanisms. We defer the formal proof to Section 5.2.

The dual convergence $\mu^t \to \mu^*$ is guaranteed by the concavity of the dual problem. Note that the dual problem is almost always nonsmooth, because it is the pointwise minimum of a number of functions. We remark that our dual convergence result applies to the more general nonsmooth convex optimization problem:

$$\min_{w}\left\{\rho(w) + \max_{\theta\in\Theta}\ell(w;\theta)\right\}, \tag{15}$$

where $\rho(\cdot)$ is a strongly convex regularization function and $\ell(w;\theta)$ is a convex loss function in $w$ for all $\theta$. Problem (15) is very common in empirical risk minimization, e.g., see [40] for an application in machine learning. In Theorem 5 part (a), we show that the cutting plane method converges at a rate of $\mathcal{O}(1/t)$ for problem (15). This result contains the earlier results in [40] as a special case. In fact, it has been show in [48] that the $\mathcal{O}(1/t)$ convergence rate of the cutting plane method is nonimprovable for problem (15).

The primal convergence is not automatically guaranteed by the dual convergence. Without the coordination steps, the primal functions may not converge at all, even if $\mu^t \to \mu^*$. There are two difficulties. First, even if $\mu^*$ is known, finding the approximate optimum requires the selection of a good solution out of a large number of best responses. Otherwise, as demonstrated in Section 4.2, the price of anarchy can be very high. Second, the set of best responses is not continuous with respect to the multiplier. When $\mu^t \to \mu^*$ but $\mu^t \neq \mu^*$ for all $t$, it is possible that many best responses to $\mu^*$ are never best responses to any $\mu^t$. This suggests that the primal solutions will never be close to optimal, even if $\mu^t \to \mu^*$. To induce convergence, we require that the participants submit their $\xi_t$-optimal responses. Here $\xi_t$ is a diminishing error tolerance that induces continuity in the best-response set with respect to the multiplier $\mu^t$. After a careful balance between the error tolerance and the convergence error, we can construct a sequence of $\hat{x}^t$ that converges into the approximate social optima.

Other than the cutting plane method, alternative dual methods may apply too. One example is the dual ascent method, which only requires the dual subgradient and does not require the dual function value. This means that participants only need to report their $g_i(x_i)$ values and they can keep their preference values $p_i(x_i)$ private. However, use of the dual ascent method may result in a slower rate of convergence or additional tuning stepsizes. Careful convergence analysis of alternative methods is a topic for future research.

## 5.2 Proof of Convergence

This subsection is devoted to the convergence analysis of Algorithm 2. For readers who are not interested, it can be safely skipped. The proof of Theorem 5 is developed through a series of lemmas.

**Lemma 4** *Let $f_1, f_2 : \Re^n \mapsto \Re$ be continous $\sigma$-strongly convex functions, and let $\gamma \in \Re, g \in \Re^n$ be such that $\min_w f_1(w) = f_1(0) = 0, f_2(0) = \gamma, g \in \partial f_2(0)$. Then*

$$\min_w \max\{f_1(w), f_2(w)\} \geq \max\left\{\gamma - \frac{\|g\|^2}{2\sigma}, \frac{2\gamma^2\sigma}{\|g\|^2}\right\}.$$

*Proof.* We use the minimax duality for convex-concave function to obtain

$$\begin{aligned}\min_w \max\{f_1(w), f_2(w)\} &= \min_w \max_{\lambda\in[0,1]} (1-\lambda)f_1(w) + \lambda f_2(w) \\ &= \max_{\lambda\in[0,1]} \min_w (1-\lambda)f_1(w) + \lambda f_2(w).\end{aligned} \tag{16}$$

By the assumptions, we obtain that $(1-\lambda)f_1(w) + \lambda f_2(w)$ is also $\sigma$-strongly convex, and moreover,

$$(1-\lambda)f_1(0) + \lambda f_2(0) = \lambda\gamma, \qquad \lambda g \in \partial\big((1-\lambda)f_1 + \lambda f_2\big)(0).$$

By using the definition of strongly convex functions, we obtain for all $w$ that

$$(1-\lambda)f_1(w) + \lambda f_2(w) \geq \frac{\sigma}{2}\|w\|^2 + \lambda g^T w + \lambda\gamma.$$

By applying the preceding relation to the minimax equality (16) and using basic calculations, we obtain

$$\begin{aligned}\min_w \max\{f_1(w), f_2(w)\} &\geq \max_{\lambda\in[0,1]} \min_w \frac{\sigma}{2}\|w\|^2 + \lambda g^T w + \lambda\gamma \\ &\geq \max_{\lambda\in[0,1]} \lambda\gamma - \frac{\lambda^2\|g\|^2}{2\sigma} \\ &\geq \max\left\{\gamma - \frac{\|g\|^2}{2\sigma}, \frac{2\gamma^2\sigma}{\|g\|^2}\right\}.\end{aligned}$$

■

**Lemma 5** *Let $\{a_t\}$ be a sequence of positive scalars and let $\eta$ be a positive scalar, such that $a_{t+1} \leq a_t - \eta a_t^2$ for all $t \geq 0$ and $a_0 \leq \frac{1}{2\eta}$. Then $a_t \leq \frac{a_0}{\eta a_0 t + 1}$ for all $t \geq 0$.*

*Proof.* We prove by induction. The statement is clearly true when $t = 0$. Now suppose that it is true for some $t > 0$. Then we have

$$\begin{aligned}a_{t+1} &\leq a_t - \eta a_t^2 \\ &\leq \frac{a_0}{\eta a_0 t + 1} - \eta\left(\frac{a_0}{\eta a_0 t + 1}\right)^2 \\ &= a_0 \cdot \frac{\eta a_0(t-1) + 1}{(\eta t a_0 + 1)^2} \\ &\leq a_0 \cdot \frac{1}{\eta(t+1)a_0 + 1},\end{aligned}$$

where the second inequality uses the monotonicity of $f(x) = x - \eta x^2$ when $x \leq \frac{1}{2\eta}$ and the fact $a_t \leq a_0 \leq \frac{1}{2\eta}$. Thus, we have shown by induction that the inequality holds for all $t \geq 0$. ■

The complete proof of Theorem 5 is divided into two parts: the dual convergence and the primal convergence. We remark that the dual convergence analysis of the cutting plane method applies to the more general risk minimization problem (15), where the objective is the sum of a strongly convex function and the maximum of a number of convex functions.

**Proof of Theorem 5(a)**

Since $f$ is $\alpha$-strongly convex and $\beta$-strongly smooth, it follows from Lemma 2 that $f^\star$ has $1/\alpha$-Lipschitz continuous gradient and is $1/\beta$-strongly convex. It follows that $Q(\mu), Q_t(\mu), L(x,\mu)$ are all $1/\beta$-strongly concave functions with respect to $\mu$.

According to the update rule of $Q_t$, we have

$$Q_{t+1}(\mu) = \min\{Q_t(\mu), \langle \mu, \bar{g}^t\rangle + \bar{p}^t - f^\star(\mu)\} = \min\{Q_t(\mu), L(x^t,\mu)\}.$$

Note that each $L(x^t,\mu)$ is an upper bound on the dual function $Q(\mu)$. Therefore, $Q_t$ is a decreasing sequence of functions that approximate $Q$ from above, i.e., $Q(\mu) \le Q_{t+1}(\mu) \le Q_t(\mu)$ for all $t$ and $\mu$. According to the update rule of $\mu^t$, we have $Q_t(\mu^t) = \max_\mu Q_t(\mu) \ge \max_\mu Q(\mu) = Q(\mu^*)$ for all $t$.

Recall the estimated gap at time $t$ is

$$\varepsilon_t = Q_t(\mu^t) - \max_{k\le t}\{\langle \mu^k, \bar{g}^k\rangle + \bar{p}^k - f^\star(\mu^k)\} = Q_t(\mu^t) - \max_{k\le t} Q(\mu^k).$$

We have

$$\begin{aligned} \varepsilon_{t+1} - \varepsilon_t &= Q_{t+1}(\mu^{t+1}) - \max_{k\le t+1} Q(\mu^k) - (Q_t(\mu^t) - \max_{k\le t} Q(\mu^k)) \\ &\le Q_{t+1}(\mu^{t+1}) - Q_t(\mu^t) \\ &= \max_\mu Q_{t+1}(\mu) - Q_t(\mu^t) \\ &= \max_\mu \min\{Q_t(\mu), L(x^t,\mu)\} - Q_t(\mu^t), \end{aligned} \tag{17}$$

where the inequality uses the fact $\max_{k\le t} Q(\mu^k) \le \max_{k\le t+1} Q(\mu^k)$, the second equality uses the definition of $\mu^t$, and the last relation uses the definition of $Q_{t+1}$.

Note that $\max_\mu Q_t(\mu) = Q_t(\mu^t)$ and $\bar{g}_t - \nabla f^\star(\mu^t) \in \partial_\mu L(x^t,\mu^t)$. Moreover, $L(x^t,\mu)$ and $Q_t(\mu)$ are $1/\beta$-strong concave with respect to $\mu$. Therefore we are able to apply Lemma 4, yielding

$$\min\{Q_t(\mu), L(x^t,\mu)\} - Q_t(\mu^t) \le -\frac{2(L(x^t,\mu^t) - Q_t(\mu^t))^2}{\beta\|\bar{g}_t - \nabla f^\star(\mu^t)\|^2}. \tag{18}$$

Note that $L(x^t,\mu^t) - Q(\mu^t) = Q(\mu^t) - Q_t(\mu^t) \le \max_{k\le t} Q(\mu^k) - Q_t(\mu^t) = -\varepsilon_t$. Also note that

$$\|\bar{g}_t - \nabla f^\star(\mu^t)\| \le \|\bar{g}_t - \nabla f^\star(\mu^*)\| + \|\nabla f^\star(\mu^*) - \nabla f^\star(\mu^t)\| \le R + \frac{\sqrt{\beta\epsilon_t}}{\alpha},$$

where $\|\bar{g}_t - \nabla f^\star(\mu^*)\| \le R$ because $\nabla f^\star(\mu^*)$ is a convex combination of $\sum_{i=1}^N g_i(x_i)$ the set $\{\sum_{i=1}^N g_i(x_i) \mid x \in \mathcal{X}\}$ has a radius smaller than $R$, and $\|\nabla f^\star(\mu^*) - \nabla f^\star(\mu^t)\| \le \frac{\sqrt{\beta\epsilon_t}}{\alpha}$ because of the strong convexity and strong smoothness assumptions. So it follows from Eq. (18) that

$$\min\{Q_t(\mu), L(x^t,\mu)\} - Q_t(\mu^t) \le -\frac{2\varepsilon_t^2}{\beta(R + \frac{\sqrt{\beta\epsilon_t}}{\alpha})^2} \le -\frac{\varepsilon_t^2}{2(\beta R^2 + \alpha^{-2}\beta^2\varepsilon_t)}. \tag{19}$$

By combining Eqs. (17) and (19), we obtain for all $t \ge 0$,

$$\varepsilon_{t+1} \le \varepsilon_t - \frac{\varepsilon_t^2}{2(\beta R^2 + \alpha^{-2}\beta^2\varepsilon_t)} \le \varepsilon_t - \frac{\varepsilon_t^2}{2(\beta R^2 + \alpha^{-2}\beta^2\varepsilon_0)}.$$

Note that $\varepsilon_t \le \varepsilon_0 \le \alpha^{-2}\beta^2\varepsilon_0$. So we apply Lemma 5 and obtain $\varepsilon_t \le \frac{\beta R^2 + \alpha^{-2}\beta^2\varepsilon_0}{t/2 + (\beta R^2 + \alpha^{-2}\beta^2\varepsilon_0)\beta\varepsilon_0^{-1}} \le \frac{\beta R^2 + \alpha^{-2}\beta^2\varepsilon_0}{t/2}$ for all $t \ge 0$. Note that $Q^* = \max_\mu Q(\mu) \le \max_\mu Q_t(\mu) = Q_t(\mu^t)$, and $\max_{k\le t} Q(\mu^k) =$

$Q(\hat{\mu}^t)$. So we have

$$Q^* - Q(\hat{\mu}^t) = Q^* - \max_{k \le t} Q(\mu^k) \le Q_t(\mu^t) - \max_{k \le t} Q(\mu^k) = \varepsilon_t \le \frac{\beta R^2 + \alpha^{-2}\beta^2\varepsilon_0}{t/2}.$$

Finally, we use the definition of $\hat{\mu}^t$ and the $1/\beta$-strong concavity of $Q$ to obtain

$$\|\mu^* - \hat{\mu}^t\|^2 \le \beta\left(Q^* - Q(\hat{\mu}^t)\right) \le \beta\varepsilon_t \le \frac{\beta^2 R^2 + \alpha^{-2}\beta^3\varepsilon_0}{t/2}.$$

Thus, we have shown that $\hat{\mu}^t$ converges to $\mu^*$ as $t \to \infty$ at a rate of $\|\hat{\mu}^t - \mu^*\|^2 = \mathcal{O}(R^2\beta^2/t)$. ∎

Now we consider the convergence of the primal function values.

**Proof of Theorem 5(b),(c)**

Note that $L(x, \mu)$ is $1/\beta$-strongly concave with respect to $\mu$ and that $\sum_{i=1}^N g_i(x_i) - \nabla f^\star(\hat{\mu}^t) \in \partial_\mu L(x, \hat{\mu}^t)$ for all $x$. So we have for all $x, \mu$ that

$$L(x, \mu) \le L(x, \hat{\mu}^t) + \langle \sum_{i=1}^N g_i(x_i) - \nabla f^\star(\hat{\mu}^t), \mu - \hat{\mu}^t \rangle - \frac{1}{2\beta}\|\mu - \hat{\mu}^t\|^2.$$

Letting $x = \hat{x}^t$ and taking the supreme over $\mu$, we have

$$F(\hat{x}^t) = \sup_\mu L(\hat{x}^t, \mu) \le L(\hat{x}^t, \hat{\mu}^t) + \sup_\mu \left\{ \langle \sum_{i=1}^N g_i(\hat{x}_i^t) - \nabla f^\star(\hat{\mu}^t), \mu - \mu^t \rangle - \frac{1}{2\beta}\|\mu - \hat{\mu}^t\|^2 \right\}.$$

By the definition of $\hat{x}^t$ in Step 6 of Algorithm 2, we have

$$L(\hat{x}^t, \hat{\mu}^t) \le \inf_{x \in \mathcal{X}} L(x, \hat{\mu}^t) + N\xi_t = Q(\hat{\mu}^t) + N\xi_t.$$

We also have $Q(\hat{\mu}^t) \le \max_\mu Q(\mu) = Q^* \le F^*$, and $\langle \sum_{i=1}^N g_i(\hat{x}_i^t) - \nabla f^\star(\hat{\mu}^t), \mu - \mu^t \rangle - \frac{1}{2\beta}\|\mu - \hat{\mu}^t\|^2 \le \frac{\beta}{2}\|\sum_{i=1}^N g_i(\hat{x}_i^t) - \nabla f^\star(\hat{\mu}^t)\|^2$ for all $\mu$. It follows from the preceding relations that

$$F(\hat{x}^t) \le F^* + N\xi_t + \frac{\beta}{2}\|\sum_{i=1}^N g_i(\hat{x}_i^t) - \nabla f^\star(\hat{\mu}^t)\|^2. \tag{20}$$

It remains to show that $\|\sum_{i=1}^N g_i(\hat{x}_i^t) - \nabla f^\star(\hat{\mu}^t)\|^2$ diminishes to zero at a suitable rate.

We claim that $\check{\mathcal{X}}_i^* \subset \mathcal{X}_i^t$ for all $i = 1, \ldots, N, t \ge 0$. To see this, we let $x_i^*$ be any best response by the $i$th participant to the optimal multiplier $\mu^*$ and let $x_i^t$ be any best response by the $i$th participant to the optimal multiplier $\hat{\mu}^t$. We have

$$\begin{aligned}
p_i(x_i^*) + \langle \hat{\mu}^t, g_i(x_i^*) \rangle &= p_i(x_i^*) + \langle \mu^*, g_i(x_i^*) \rangle + \langle \hat{\mu}^t - \mu^*, g_i(x_i^*) \rangle \\
&\le p_i(x_i^t) + \langle \mu^*, g_i(x_i^t) \rangle + \langle \hat{\mu}^t - \mu^*, g_i(x_i^*) \rangle \\
&= p_i(x_i^t) + \langle \hat{\mu}^t, g_i(x_i^t) \rangle + \langle \hat{\mu}^t - \mu^*, g_i(x_i^*) - g_i(x_i^t) \rangle \\
&= \min_x \{p_i(x) + \langle \hat{\mu}^t, g_i(x) \rangle\} + \langle \hat{\mu}^t - \mu^*, g_i(x_i^*) - g_i(x_i^t) \rangle,
\end{aligned}$$

where the inequality uses the optimality of $x_i^*$, and the last relation uses the optimality of $x_i^t$. By using the bounded radius $R$ of $\{g_i(x_i) \mid x_i \in \mathcal{X}_i\}$, we also have

$$\langle \hat{\mu}^t - \mu^*, g_i(x_i^*) - g_i(x_i^t) \rangle \le \|\hat{\mu}^t - \mu^*\| \cdot \|g_i(x_i^*) - g_i(x_i^t)\| \le \sqrt{\beta\varepsilon_t} \cdot R = \xi_t.$$

According to the definition of $\mathcal{X}_i^t$, it provides sufficient error tolerance so that $x_i^* \in \mathcal{X}_i^t$ for all $t$. Thus we have proved the claim.

Because $\check{\mathcal{X}}_i^* \subset \mathcal{X}_i^t$, we have

$$\mathcal{G}_i^* = \{g_i(x_i) \mid x_i \in \check{\mathcal{X}}_i^*\} \subset \mathcal{G}_i^t, \qquad i = 1, \dots, N.$$

So we have $\nabla f^\star(\mu^*) \in \text{conv}(\mathcal{G}_1^* + \cdots + \mathcal{G}_N^*) \subset \text{conv}(\mathcal{G}_1^t + \cdots + \mathcal{G}_N^t)$. According to the construction of $\hat{x}^t$, we use Theorem 4 to obtain

$$\begin{aligned} \|\sum_{i=1}^N g_i(\hat{x}_i^t) - \nabla f^\star(\hat{\mu}^t)\| &\le \min_{y \in \text{conv}(\mathcal{G}_1^t + \cdots + \mathcal{G}_N^t)} \|y - \nabla f^\star(\hat{\mu}^t)\| + M\delta_g \\ &\le \|\nabla f^\star(\mu^*) - \nabla f^\star(\hat{\mu}^t)\| + M\delta_g. \end{aligned}$$

By using the $1/\alpha$-Lipschitz continuity of $\nabla f^\star$ and the $1/\beta$-strong convexity of $f$, we have $\|\nabla f^\star(\mu^*) - \nabla f^\star(\hat{\mu}^t)\| \le \frac{1}{\alpha}\|\hat{\mu}^t - \mu^*\| \le \frac{\sqrt{\beta\epsilon_t}}{\alpha}$. Applying the preceding two inequalities to Eq. (20), we obtain that

$$F(\hat{x}^t) - F^* \ \le N\xi_t + \frac{\beta}{2}\|\sum_{i=1}^N g_i(\hat{x}_i^t) - \nabla f^\star(\hat{\mu}^t)\|^2 \le NR\sqrt{\beta\varepsilon_t} + \beta\left(M\delta_g\right)^2 + \frac{\beta^2}{\alpha^2}\varepsilon_t.$$

Since $F$ is continuous and $\varepsilon_t \to 0$, any limit point of $\hat{x}^t$ is an approximate optimum. Finally, we apply part (a) and parts (b) and (c) follow immediately. ■

## 5.3 Remarks on Complexity

Let us comment on the complexity of the nonconvex problem (1). Although problem (1) is continuous, it bears strong connection to the multi-row knapsack problem given by

$$\min c^T x, \qquad \text{subject to} \quad Ax \le b,\ x_i = \{0, 1\},\ i = 1, \dots, n,$$

where $c \in \Re^n, A \in \Re^{m \times n}$, and $b \in \Re^m$. In fact, we believe that they admit polynomial-time reductions to each other. Let us describe the main ideas without getting into technical details.

On one hand, the knapsack problem is a special case of problem (1). To see this, we let $p_i$ and $g_i$ be linear functions determined by $c_i$ and the $i$-th column of $A$, respectively. Also we let $\mathcal{X}_i = \{0, 1\}$ and $\mathcal{A} = \{y \le b\}$. In this way, we have constructed an instance of the constrained problem (2) that is equivalent to the knapsack problem. Note that problem (2) is a special case of problem (1). So we have obtained an immediate reduction from the knapsack problem to problem (1).

On the other hand, problem (1) can be approximated by the knapsack problem via discretization. Given a problem instance with functions $p_i, g_i, \mathcal{X}_i, f$, we add a dummy variable to get a constrained problem

$$\begin{aligned} &\text{minimize} \ \sum_{i=1}^N p_i(x_i) + f(z), \\ &\text{subject to} \ \sum_{i=1}^N g_i(x_i) = z, \quad x_i \in \mathcal{X}_i, \qquad i = 1, \dots, N. \end{aligned}$$

We discretize the functions $p_i, f$ into vectors, the mappings $g_i$ into matrices, and the sets $\mathcal{X}_i$ into grid points. Then we change the variables from $x_i$ to indice of the discrete vectors. The resulting problem is a multiple-choice multi-row knapsack problem, which can be further reduced to the multi-row $\{0, 1\}$-knapsack problem. Under additional continuity assumptions on the functions

$p_i, g_i, f$, we conjecture that the reduction is polynomial-time. Note that the discretization disregards any structure of the continuous functions and results in a huge-dimensional discrete problem. So it is not a practical computation method.

The two-way reduction implies that problem (1) and the knapsack problem belong to the same complexity class (requiring a rigorous definition of complexity of continuous problem). The knapsack problem is known to be $\mathcal{NP}$-complete. When it has at least two row constraints, the knapsack problem admits no fully polynomial-time approximation scheme unless $\mathcal{P} = \mathcal{NP}$ (see [25]). Because it is a special case of problem (1), the more generic nonconvex optimization problem (1) is at least as hard. In other words, we cannot develop an efficient algorithm to achieve an $\epsilon$-optimal solution to problem (1) for arbitrarily small $\epsilon > 0$. In contrast, we have developed Algorithm 2, which finds a constant-error approximate solution in a polynomial number of iterations. The constant error is the duality gap that diminishes to zero as $N/M$ increases. A rigorous complexity analysis and proof of reduction is beyond the scope of this work. It is an interesting topic for future investigation.

# 6 Numerical Experiments

We study the asymptotic price of anarchy by generating random instances of problem (1). We let $p_i, g_i$ be piecewise constant functions defined on $[0, 1]$, where the jump points are independent random vectors generated from the multivariate uniform distribution, for all $i = 1, \ldots, N$. We let $f(\cdot)$ be a quadratic function, where the Hessian is a random symmetric and positive definite matrix with mean $I_{N\times N}$. According to our random generation, the maximal convexity gaps $\delta_p, \delta_g$ are bounded for all $N$ with probability 1.

For each sample instance of problem (1), we apply Algorithm 2 to find an approximate optimal value $F(\tilde{x})$. The primal and dual trajectories are plotted in Figure 2. We observe that the dual variables converge very fast, usually within 20-50 iterations. In addition, the dual convergence seems to be at a geometric rate in initial iterations and slows down afterwards. This is consistent with our analysis that the dual error diminishes according to $\varepsilon_{t+1} \leq \varepsilon_t - \frac{\varepsilon_t^2}{\beta R^2 + \alpha^{-2}\beta^2\varepsilon_t}$, which is a contraction when the error is large. On the other hand, the primal convergence is much slower and has a constant error due to the positive duality gap. By comparing the trajectories for different values of $N, M$, we see that the normalized duality gap becomes small when the ratio $N/M$ is large. We also notice that the convergence is faster when $M$ is small. This is due to the random generation of the problem, because problems with small $M$ are more likely to be well-conditioned (resulting in small $\beta/\alpha$).

For comparison, we compute the global optimum $x^*$ and the optimal value $F^*$ by discretizing the optimization problem and using an exhaustive search. We compute the price of anarchy $\frac{F(\tilde{x})-F^*}{F^*}$ for each sampled problem, where $N$ varies and $M$ is fixed. The samples and means of the price of anarchy are illustrated in Figure 3. We observe that the price of anarchy is very close to zero with high probability. Moreover, the inverse of the mean price of anarchy is linearly related to the dimension $N$. This validates our theory that the price of anarchy vanishes to 0 at a rate of $\mathcal{O}(1/N)$ as the number of participants increases.

# 7 Concluding Remarks

We have studied a nonconvex cooperative optimization problem where $N$ participants minimize a common objective. We consider a Fenchel dual of the nonconvex problem in order to decompose it with respect to the individual decisions. We show that the dual problem becomes increasingly convex in a geometric sense as $N$ increases. We have mathematically characterized the duality gap of the nonconvex problem. The duality gap is a form of price of anarchy when

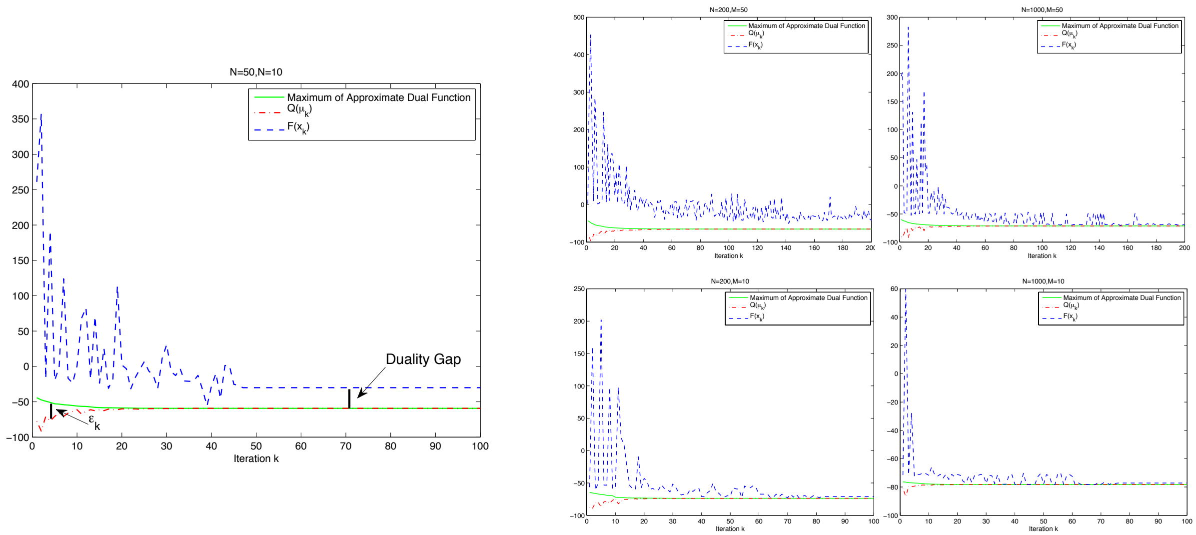


**Figure 2: Primal and Dual Convergence of Algorithm 2.** We plot trajectories of the primal objective values, dual objective values, and upper bounds of dual objective values generated by Algorithm 2. The estimated dual convergence error $\epsilon_t$ diminishes t o 0 very fast, while the primal objective value converges to a constant error that is equal to the duality gap.

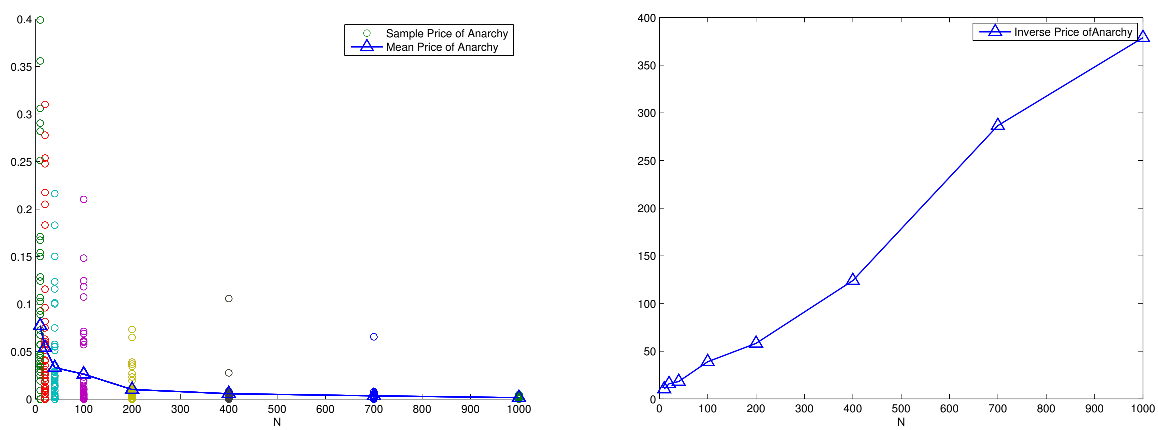


**Figure 3: Asymptotic Price of Anarchy.** We generate 50 random instances of problem (1) for each value of $N$ and find the approximate optimum using Algorithm 2. In the left figure, we plot the samples and means of the price of anarchy $\frac{F(\tilde{x})-F^*}{F^*}$ for different values of $N$. In the right figure, we plot the inverse of the mean price of anarchy against the values of $N$.

each participant optimizes for self-interests. We prove that the price of anarchy, which is due to the nonconvexity, asymptotically diminishes to zero as $N$ grows.

Algorithmically, we show how to achieve the approximate social optimum using a coordination procedure. Without coordination, we show by examples that the price of anarchy can be arbitrarily high. We propose a distributed duality-based algorithm that relies on individual best responses to find the approximate optimum. The algorithm has an important coordination step, which involves finding an extreme point of a linear feasibility problem. It ensures that both the primal and dual variables are convergent. We analyze the convergence and rate of convergence of the proposed algorithm, and we test the algorithm on randomly generated examples.

Finally, we are able to answer the question raised in Section 1: Is there a fair price such that individual best responses automatically achieve the social optimum? The answer is largely yes. Indeed, there exists a nearly fair price at which a best response solution is an approximate social optimum. However, such an approximate optimum can only be obtained via careful coordination of individuals' decisions. Coordination plays a critical role in solving the nonconvex optimization problem.

In future work, a theoretical challenge is to generalize the duality gap results to more abstract settings, such as the minimax problems and variational inequalities. Another potential work is to design and study more efficient algorithms with better error-complexity guarantees. From the practical perspective, an important future work is to identify practical instances of the cooperative problem and tailor the analysis and algorithms to specific applications.

# References


[1] Pramod Abichandani, Hande Y Benson, and Moshe Kam. Decentralized multi-vehicle path coordination under communication constraints. In *Intelligent Robots and Systems (IROS), 2011 IEEE/RSJ International Conference on*, pages 2306–2313. IEEE, 2011.

[2] Kenneth Joseph Arrow, Frank Hahn, et al. *General Competitive Analysis*. Holden-Day, 1971.

[3] Italo Atzeni, Luis Garcia Ordóñez, Gesualdo Scutari, Daniel P Palomar, and Javier Rodríguez Fonollosa. Noncooperative and cooperative optimization of distributed energy generation and storage in the demand-side of the smart grid. *Signal Processing, IEEE Transactions on*, 61(10):2454–2472, 2013.

[4] Jean-Pierre Aubin and Ivar Ekeland. Estimates of the duality gap in nonconvex optimization. *Mathematics of Operations Research*, 1(3):225–245, 1976.

[5] José Bento, Nate Derbinsky, Javier Alonso-Mora, and Jonathan S Yedidia. A message-passing algorithm for multi-agent trajectory planning. In *Advances in Neural Information Processing Systems*, pages 521–529, 2013.

[6] Dimitri Bertsekas, G Lauer, N Sandell Jr, and Thomas A Posbergh. Optimal short-term scheduling of large-scale power systems. *Automatic Control, IEEE Transactions on*, 28(1):1–11, 1983.

[7] Dimitri P Bertsekas. *Convex Optimization Theory*. Athena Scientific Belmont, MA, 2009.

[8] Dimitri P Bertsekas, Angelia Nedic, and A Ozdaglar. Min common/max crossing duality: A simple geometric framework for convex optimization and minimax theory. *Rep. LIDS-P-2536*, 2002.

[9] Dimitri P Bertsekas and NR Sandell. Estimates of the duality gap for large-scale separable nonconvex optimization problems. pages 782–785, 1982.

[10] DP Bertsekas. Convexification procedures and decomposition methods for nonconvex optimization problems. *Journal of Optimization Theory and Applications*, 29(2):169–197, 1979.

[11] Eric Budish. The combinatorial assignment problem: Approximate competitive equilibrium from equal incomes. *Journal of Political Economy*, 119(6):1061–1103, 2011.

[12] Guilherme Carmona and Konrad Podczeck. On the existence of pure-strategy equilibria in large games. *Journal of Economic Theory*, 144(3):1300–1319, 2009.

[13] JWS Cassels. Measures of the non-convexity of sets and the Shapley–Folkman–Starr theorem. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 78, pages 433–436. Cambridge Univ Press, 1975.

[14] Raphaël Cerf. Large deviations for sums of i.i.d. random compact sets. *Proceedings of the American Mathematical Society*, 127(8):2431–2436, 1999.

[15] Nira Dyn and Elza Farkhi. Set-valued approximations with Minkowski averages–convergence and convexification rates. *Numerical Functional Analysis and Optimization*, 25(3-4):363–377, 2005.

[16] Ivar Ekeland and Roger Temam. *Convex Analysis and 9 Variational Problems.* SIAM, 1976.

[17] Igor V Evstigneev and SD Flam. Sharing nonconvex costs. *Journal of Global Optimization*, 20(3-4):257–271, 2001.

[18] Paolo Fiorini and Zvi Shiller. Motion planning in dynamic environments using velocity obstacles. *The International Journal of Robotics Research*, 17(7):760–772, 1998.

[19] Birgit Grodal. The equivalence principle. *Optimization and Operation Research, Encyclopedia of Life Support Systems (EOLSS)*, 2002.

[20] Werner Hildenbrand, David Schmeidler, and Shmuel Zamir. Existence of approximate equilibria and cores. *Econometrica: Journal of the Econometric Society*, pages 1159–1166, 1973.

[21] Sham Kakade, Shai Shalev-Shwartz, and Ambuj Tewari. On the duality of strong convexity and strong smoothness: Learning applications and matrix regularization. *Unpublished Manuscript, http://ttic. uchicago. edu/shai/papers/KakadeShalevTewari09. pdf*, 2009.

[22] Oussama Khatib. Real-time obstacle avoidance for manipulators and mobile robots. *The International Journal of Robotics Research*, 5(1):90–98, 1986.

[23] Claude Lemaréchal and Arnaud Renaud. A geometric study of duality gaps, with applications. *Mathematical Programming*, 90(3):399–427, 2001.

[24] Zhi-Quan Luo and Shuzhong Zhang. Duality gap estimation and polynomial time approximation for optimal spectrum management. *Signal Processing, IEEE Transactions on*, 57(7):2675–2689, 2009.

[25] Michael J Magazine and Maw-Sheng Chern. A note on approximation schemes for multidimensional knapsack problems. *Mathematics of Operations Research*, 9(2):244–247, 1984.

[26] Nimrod Megiddo. On finding primal and dual optimal bases. *ORSA Journal on Computing*, 3(1):63–65, 1991.

[27] Angelia Nedich and Asuman Ozdaglar. A geometric framework for nonconvex optimization duality using augmented Lagrangian functions. *Journal of Global Optimization*, 40(4):545–573, 2008.

[28] M Pappalardo. On the duality gap in nonconvex optimization. *Mathematics of Operations Research*, 11(1):30–35, 1986.

[29] Madan L Puri and Dan A Ralescu. Limit theorems for random compact sets in Banach space. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 97, pages 151–158. Cambridge Univ Press, 1985.

[30] Robin L Raffard, Claire J Tomlin, and Stephen P Boyd. Distributed optimization for cooperative agents: Application to formation flight. In *Decision and Control, 2004. CDC. 43rd IEEE Conference on*, volume 3, pages 2453–2459. IEEE, 2004.

[31] Katayoun Rahbar, Chin Choy Chai, and Rui Zhang. Real-time energy management for cooperative microgrids with renewable energy integration. In *Smart Grid Communications (SmartGridComm), 2014 IEEE International Conference on*, pages 25–30. IEEE, 2014.

[32] Arthur Richards, John Bellingham, Michael Tillerson, and Jonathan How. Coordination and control of multiple UAVs. In *AIAA Guidance, Navigation, and Control Conference, Monterey, CA*, 2002.

[33] RT Rockefeller. *Convex Analysis*. Princeton University Press, Princeton, NJ, 1970.

[34] Gesualdo Scutari, Francisco Facchinei, and Jong-Shi Pang. Equilibrium selection in MIMO communication games. In *Signal Processing Advances in Wireless Communications (SPAWC), 2012 IEEE 13th International Workshop on*, pages 80–84. IEEE, 2012.

[35] Andrea Simonetto and Geert Leus. Distributed maximum likelihood sensor network localization. *arXiv preprint arXiv:1309.2502*, 2013.

[36] Seshan Srirangarajan, Ahmed H Tewfik, and Zhi-Quan Luo. Distributed sensor network localization using SOCP relaxation. *Wireless Communications, IEEE Transactions on*, 7(12):4886–4895, 2008.

[37] Ross M Starr. Quasi-equilibria in markets with non-convex preferences. *Econometrica: Journal of the Econometric Society*, pages 25–38, 1969.

[38] Ross M Starr. Approximation of points of the convex hull of a sum of sets by points of the sum: an elementary approach. *Journal of Economic Theory*, 25(2):314–317, 1981.

[39] Ross M Starr. *General Equilibrium Theory: An Introduction*. Cambridge University Press, 2011.

[40] Choon Hui Teo, SVN Vishwanthan, Alex J Smola, and Quoc V Le. Bundle methods for regularized risk minimization. *The Journal of Machine Learning Research*, 11:311–365, 2010.

[41] Madeleine Udell and Stephen Boyd. Bounding duality gap for problems with separable objective. *arXiv preprint arXiv:1410.4158*, 2014.

[42] Robert J Vanderbei. *Linear Programming*, volume 37. Springer, 2001.

[43] R. Vujanic, P. Mohajerin Esfahani, P.J. Goulart, S. Mariéthoz, and M. Morari. Vanishing Duality Gap in Large Scale Mixed-Integer Optimization: a Solution Method with Power System Applications. *Automatica*, March 2015.

[44] Robin Vujanic, Peyman Mohajerin Esfahani, Paul Goulart, and Manfred Morari. Large scale mixed-integer optimization: A solution method with supply chain applications. In *Control and Automation (MED), 2014 22nd Mediterranean Conference of*, pages 804–809. IEEE, 2014.

[45] Fang Ethan X., Liu Han, and Wang Mengdi. Blessing of massive scale: Spatial graphical model inference with a total cardinality constraint. *Working paper*, 2015.

[46] Akira Yamazaki. An equilibrium existence theorem without convexity assumptions. *Econometrica: Journal of the Econometric Society*, pages 541–555, 1978.

[47] Wei Yu and Raymond Lui. Dual methods for nonconvex spectrum optimization of multicarrier systems. *Communications, IEEE Transactions on*, 54(7):1310–1322, 2006.

[48] Xinhua Zhang, Ankan Saha, and SVN Vishwanathan. Lower bounds on rate of convergence of cutting plane methods. In *Advances in Neural Information Processing Systems*, pages 2541–2549, 2010.

[49] Lin Zhou. A simple proof of the Shapley-Folkman theorem. *Economic Theory*, 3(2):371–372, 1993.