Asymptotically Fair and Truthful Allocation of Public Goods

POUYA KANANIAN*, University of Toronto, Canada
ARNESH SUJANANI, University of Waterloo, Canada
SEYED MAJID ZAHEDI†, University of Waterloo, Canada

We study the fair and truthful allocation of m divisible public items among n agents, each with distinct preferences for the items. To aggregate agents' preferences fairly, we focus on finding a core solution. For divisible items, a core solution always exists and can be calculated by maximizing the Nash welfare objective. However, such a solution is easily manipulated; agents might have incentives to misreport their preferences. To mitigate this, the current state-of-the-art finds an approximate core solution with high probability while ensuring approximate truthfulness. However, this approach has two main limitations. First, due to several approximations, the approximation error in the core could grow with n, resulting in a non-asymptotic core solution. This limitation is particularly significant as public-good allocation mechanisms are frequently applied in scenarios involving a large number of agents, such as the allocation of public tax funds for municipal projects. Second, implementing the current approach for practical applications proves to be a highly nontrivial task. To address these limitations, we introduce PPGA, a (differentially) Private Public-Good Allocation algorithm, and show that it attains asymptotic truthfulness and finds an asymptotic core solution with high probability. Additionally, to demonstrate the practical applicability of our algorithm, we implement PPGA and empirically study its properties using municipal participatory budgeting data.

1 INTRODUCTION

Unlike the allocation of private goods, where each item goes to a single agent, public goods allow multiple agents to benefit from an allocated item. In this paper, we study the problem of fairly allocating m divisible public goods among n agents in a truthful manner. Different agents hold distinct preferences for the items. Each item has a size, and the total size of allocated items should not exceed the available capacity. The fair allocation of divisible public goods is a fundamental problem in social choice theory with many real-world applications. Examples include: (1) federal/state budget allocations between services such as healthcare, education, and defense or municipal budget allocations to improve utilities such as libraries, parks, and roads[1]; (2) shared memory allocations between files with different sizes; and (3) time allocations between activities during events.

An allocation mechanism produces outcomes based on reported preferences of all agents. Agents need not reveal their true preferences but strategically report them to maximize their utility. For instance, consider a setting where there are one or more commonly preferred items. Such items are highly likely to be allocated regardless of the reported preferences of a single agent. Given this and assuming that other agents report their preference truthfully, agents could be incentivized to *free-ride* by falsely claiming disinterest in commonly preferred items and reporting preferences only for their individually preferred items. By doing so, free riders increase the chances of their individually preferred items being allocated under a fair allocation mechanism.

To aggregate agents' preferences fairly, we focus on the classic game theoretic notion of the core [13, 25]. The core generalizes well-studied notions of proportionality and Pareto efficiency by ensuring group-wise fairness, providing fair outcomes to each agent subset relative to its size. The

*This research was conducted while the author was a graduate student at the University of Waterloo.
†Corresponding Author.
[1]Municipal budgets are more amenable to inclusive voting procedures like participetory budgeting, while federal/state budgets are more likely voted on by legislators

Authors' addresses: Pouya Kananian, pouya.kananian@mail.utoronto.ca, University of Toronto, Toronto, Canada; Arnesh Sujanani, arnesh.sujanani@uwaterloo.ca, University of Waterloo, Waterloo, Canada; Seyed Majid Zahedi, smzahedi@uwaterloo.ca, University of Waterloo, Waterloo, Canada.

notion of the core has been extensively studied in the context of public-good allocation [17, 18, 21, 42, 43]. For allocating divisible public goods, the core always exists, and it can be calculated by maximizing *Nash welfare (NW)* objective (i.e., the product of agents' utilities) [17]. However, the core is easy to manipulate; agents might be incentivized to free-ride.

To address this issue, Fain et al. [17] propose a method that aims to find an approximate core solution with high probability while also achieving approximate truthfulness. This method relies on the exponential mechanism derived from differential privacy [39]. The exponential mechanism uses a scoring function to assign a score to each outcome. Subsequently, a sample is drawn from a distribution that exponentially weights outcomes based on their scores. This guarantees that the selected outcome's score is approximately maximized with high probability.

Informally, differential privacy ensures that the output of a mechanism does not change significantly when any agent unilaterally modifies their data. This emphasis on unilateral deviations aligns closely with truthfulness in mechanism design, where a mechanism is truthful if agents have no incentive to misreport their types. As a result, differentially private mechanisms can be shown to be approximately dominant-strategy truthful [39]. For the exponential mechanism, the level of differential privacy–and consequently, truthfulness–is contingent on the sensitivity[2] of the scoring function to the reported input of any individual agent. Higher sensitivity corresponds to a lower quality of the guarantee.

The use of the exponential mechanism for public-good allocation faces two primary challenges. First, while the NW objective seems to be an ideal choice for the scoring function, its direct use is hindered by its high sensitivity to each agent's reported preferences. This limitation arises since the NW objective is not separable[3]. To address this, Fain et al. [17] propose using a proxy function to replace the NW objective in the scoring function.

The introduced proxy function strikes a balance between reducing the sensitivity of the scoring function–thereby improving truthfulness approximation–and retaining sufficient sensitivity to ensure an acceptable approximation to the core. However, the use of the proxy function, along with other approximations, introduces an approximation error in satisfying the core conditions. This approximation error can grow with the number of agents, potentially resulting in a solution that does not satisfy the asymptotic core. This limitation is particularly significant as public-good allocation mechanisms are frequently applied in scenarios involving a large number of agents, such as participatory budgeting elections for distributing municipal budgets.

Secondly, sampling an m-dimensional allocation from a distribution poses a significant practical challenge. To tackle this, Fain et al. [17] propose employing the hit-and-run method [53] to sample an allocation from an "approximately right distribution." However, implementing the hit-and-run method for practical applications proves to be a highly nontrivial task, as discussed in the conclusion of Sec. 2.2 by Lovász and Vempala [37]. Moreover, the implications of the extra approximation on the guarantees of truthfulness and core remain unclear.

1.1 Our Contributions

In Section 3, we introduce PPGA, a novel differentially private algorithm for public-good allocation. A key feature of PPGA is its approach to maximize the NW objective in a differentially private way without requiring a proxy objective. As previously discussed, the non-separable nature of the NW objective poses challenges in deploying differentially private mechanisms [17]. To tackle this challenge, we employ a key technique called *global variable consensus optimization* [7]. Consensus

[2]Informally, the sensitivity of a function is the maximum change in its output resulting from a change in its input (refer to Section 2.3 for a formal definition).

[3]$f(x)$ is separable with respect to a partition of x into n sub-vectors $x = (x_1, \ldots, x_n)$ if $f(x) = \sum f_i(x_i)$.

transforms the NW objective into a separable form that splits easily. Leveraging the *alternating direction method of multipliers (ADMM)* [24, 26] enables us to maximize the NW objective in a distributed manner. And this further allows us to employ the *Gaussian mechanism* [39] from differential privacy to achieve truthfulness. The application of differentially private ADMM to the global variable consensus problem for Nash welfare optimization is a novel contribution of this work.

In Section 4, we analytically study the properties of our proposed algorithm. Our primary technical contribution lies in showing that, for carefully chosen $\epsilon, \delta > 0$, PPGA achieves (ϵ, δ)-truthfulness and returns an $(\varepsilon_1, \varepsilon_2(1+\varepsilon_1))$-core outcome with probability at least $1 - 1/n - 1/n^m$, where

$$\varepsilon_1 = O\left(\sqrt{\frac{m\log(1/\delta)}{n\epsilon^2}}\right) \quad \text{and} \quad \varepsilon_2 = O\left(\sqrt[4]{\frac{m\log(1/\delta)}{n\epsilon^2}}\right).$$

Assuming that $m = o(\sqrt{n})$ and setting $\epsilon = \Theta(1/\log(n))$ and $\delta = \Theta(1/\sqrt{n})$, we further demonstrate that PPGA is asymptotically truthful (Theorem 12) and yields an asymptotic core solution with high probability (Theorem 17). To our knowledge, PPGA is the first polynomial-time algorithm (Theorem 19) that offers these guarantees.

In Section 5, we demonstrate that PPGA can be deployed in practice to solve large-scale public-good allocation problems. To this end, we implement PPGA and utilize our implementation to compare the outcome of PPGA with a core solution using data obtained from real-world participatory budgeting elections [19].

2 PRELIMINARIES

In this section, we first define the public-good allocation problem and its desired properties. We then provide an overview of differential privacy as a tool for designing truthful mechanisms. A summary of our notations is presented in Appendix A

2.1 Problem Formulation

We consider a public-good allocation problem with n agents and m divisible public items ($m \ll n$). The size of each item j is denoted by $s_j \in \mathbb{R}_{>0}$, and the size vector is denoted by $s = (s_1, \ldots, s_m)$. The total available capacity is $c \in \mathbb{R}_{>0}$. An allocation is a vector $z = (z_1, \ldots, z_m) \in [0,1]^m$, where z_j represents the fraction of the total capacity that is allocated to item j. The set of all feasible allocations is denoted by:

$$\mathcal{Z} \triangleq \{z \in [0,1]^m \mid \|z\|_1 \leq 1,\ cz \leq s\}.$$

Agent i's utility function for an allocation $z \in \mathcal{Z}$ is denoted by $U_i(z)$ and is parameterized by the utility vector $u_i = (u_{i1}, \ldots, u_{id})$, where d is a positive integer. For example, for a linear utility function of the form $U_i(z) = \sum_{j=1}^{m} u_{ij} z_j$, we have $d = m$, and each u_{ij} represents the relative value that agent i assigns to the fraction of the budget allocated to item j. In this paper, we focus on a subclass of utility functions defined on $\mathcal{Z}$ that are differentiable, strictly increasing, concave, and β-smooth, i.e., they have β-Lipschitz continuous gradients:

$$\|\nabla U_i(z) - \nabla U_i(z')\|_2 \leq \beta \|z - z'\|_2.$$

This subclass includes the common linear utility functions, which generalize additive utilities studied by [2, 4–6, 9, 18, 40, 45, 54]. Without loss of generality, we assume that $U_i \in [0,1]$ for all i, with $U_i(\mathbf{0}_m) = 0$ and $U_i(z) > 0$ for some $z \in \mathcal{Z}$. We further assume that $u_i \in \mathcal{U}$ for every i, where $\mathcal{U} \triangleq [0,1]^d$.

2.2 Mechanism Design for Public Goods

A randomized allocation mechanism M produces a probability distribution over feasible allocations given agents' reported utilities $u = (u_1, \ldots, u_n) \in \mathcal{U}^n$. We use $M(u)$ to denote the distribution produced by mechanism M for the reported utilities u, and at times, we also use $M(u)$ to represent a random allocation drawn from the distribution $M(u)$, slightly abusing the notation. Agents need not report their true utilities. They report strategically to optimize their total utility taking into account what (they think) other agents report. If agents are always incentivized to report their true utilities, no matter what others do, then the mechanism is *dominant-strategy truthful*:

Definition 1 (Dominant-strategy Truthfulness). *Let U_i be agent i's utility function parameterized by i's true utility vector u_i. A randomized mechanism M is (ϵ, δ)-truthful if $\mathbb{E}[U_i(M(u_i, u_{-i}))] \geq (1-\epsilon)\mathbb{E}[U_i(M(u_i', u_{-i}))] - \delta$ for every i, $u_i' \in \mathcal{U}$, and $u_{-i} \in \mathcal{U}^{(n-1)}$.*[4]

If $\epsilon, \delta = 0$, then M is *exactly truthful*. Approximate truthfulness is desirable in settings in which the approximation parameters ϵ and δ tend to 0 as the number of agents n grows large. This property is referred to as *asymptotic truthfulness*. Next, we formally define the classic notion of the *core*.

Definition 2 (Core). *For an allocation $z \in \mathcal{Z}$, a set of agents A form a blocking coalition if there exists another allocation $z' \in \mathcal{Z}$ such that $(|A|/n)U_i(z') \geq U_i(z)$ for every $i \in A$ with at least one strict inequality. An allocation is a core outcome if it admits no blocking coalitions.*

In this definition, when a subset A of agents deviates, they can choose any feasible allocation with the full capacity c. However, their utility is scaled down by a factor of $|A|/n$. An alternative way of defining a core solution is where a deviating coalition A could choose any allocation with a capacity of $(c|A|)/n$ instead of c, but their utilities would not be scaled down [21, 51]. For $|A| = n$, both notions capture Pareto efficiency. However, for $|A| = 1$, they provide different interpretations of proportionality–one based on utility and one based on capacity.

For divisible goods, the core coincides with the *max Nash welfare (MNW)* solution:[5]

Lemma 3. *If each U_i is differentiable and concave, then any allocation that maximizes $\sum_i \log(U_i(z))$ subject to $z \in \mathcal{Z}$ constitutes a core solution*[6].

This lemma shows that the exact MNW solution is a core outcome. However, such a solution can be irrational even when all inputs are rational [1], potentially precluding the existence of an exact algorithm [18]. Therefore, we adopt an approximate notion of the core that still provides meaningful guarantees:

Definition 4 (Approximate core). *For $\epsilon, \delta \geq 0$, an allocation $z \in \mathcal{Z}$ is an (ϵ, δ)-core outcome if there exists no set of agents $A \subseteq N$ and no allocation $z' \in \mathcal{Z}$ such that $(|A|/n)U_i(z') \geq (1+\epsilon)U_i(z)+\delta$ for all $i \in A$ with at least one strict inequality.*

When ϵ and δ converge to zero asymptotically as n grows large, the allocation is said to be an *asymptotic core* solution. The following lemma shows that an approximate MNW solution implies an approximate core solution (see Appendix D.2 for the proof).

Lemma 5. *Let $\epsilon, \delta \geq 0$. Then, $z \in \mathcal{Z}$ is an (ϵ, δ)-core outcome if, for any $z' \in \mathcal{Z}$, we have:*

$$\frac{1}{n}\sum_i \frac{U_i(z')}{U_i(z) + \delta/(1+\epsilon)} \leq 1 + \epsilon \tag{1}$$

[4] Subscript $-i$ is used to refer to all agents other than agent i.

[5] Similar lemmas appear in [17, 18] for other classes of utility functions. For completeness, we provide the proof of Lemma 3 in Appendix D.1.

[6] In this paper, all logarithms are natural.

2.3 Mechanism Design via Differential Privacy

In this subsection, we provide some background on differential privacy as a tool for designing truthful mechanisms. Informally, a mechanism satisfies DP if its output is nearly equally likely to be observed for any pair of *adjacent* inputs. Inputs are considered adjacent if they differ in only one element. For allocation mechanisms, inputs correspond to agents' reported utilities. Thus, $u, u' \in \mathcal{U}^n$ are adjacent if they differ solely in the reported utility of a single agent. We now formally define DP [15]:

Definition 6 (DP). *A randomized mechanism M is (ϵ, δ)-DP if, for any two adjacent inputs $u, u' \in \mathcal{U}^n$ and any subset of outputs $O \subseteq \mathcal{Z}$, it satisfies $\mathbb{P}[M(u) \in O] \leq e^{\epsilon}\mathbb{P}[M(u') \in O] + \delta$.*[7]

In this definition, ϵ and δ control the desired level of privacy and are typically provided as inputs to the mechanism. In general, smaller values provide stronger privacy guarantees but result in higher levels of noise being required to be injected, which can adversely affect the quality of the output. A mechanism that satisfies (ϵ, δ)-DP is (ϵ, δ)-truthful:[8]

Lemma 7. *Let M be (ϵ, δ)-DP for some $\epsilon, \delta < 1$. Then, M is (ϵ, δ)-truthful.*

Proof. Consider any agent i, and let $U_i : \mathcal{Z} \mapsto [0, 1]$ be agent i's utility parameterized according to their true utility vector u_i. Define the set $S(t) = \{z \mid U_i(z) > t\}$. Since M is (ϵ, δ)-DP, for any $u = (u_i, u_{-i}) \in \mathcal{U}^n$ and $u'_i \in \mathcal{U}$, the following inequality holds:

$$\mathbb{P}[M(u) \in S(t)] \geq e^{-\epsilon}\mathbb{P}[M(u'_i, u_{-i}) \in S(t)] - \delta. \tag{2}$$

Given the definition of $S(t)$, we can rewrite (2) as:

$$\mathbb{P}[U_i(M(u)) > t] \geq e^{-\epsilon}\mathbb{P}[U_i(M(u'_i, u_{-i})) > t] - \delta. \tag{3}$$

Given that $\mathbb{E}[X] = \int_0^1 \mathbb{P}[X > t]dt$ for any random variable $X \in [0, 1]$, we obtain the following by integrating both sides of (3):

$$\mathbb{E}[U_i(M(u))] \geq e^{-\epsilon}\mathbb{E}[U_i(M(u'_i, u_{-i}))] - \delta \geq (1 - \epsilon)\mathbb{E}[U_i(M(u'_i, u_{-i}))] - \delta,$$

where the second inequality follows because $e^{-\epsilon} \geq 1 - \epsilon$. □

We next define *Rényi differential privacy (RDP)* as a relaxation of DP [41]:

Definition 8 (RDP). *A randomized mechanism M is (α, ϵ)-RDP with order $\alpha > 1$ if for any two adjacent inputs $u, u' \in \mathcal{U}^n$, it satisfies: $D_\alpha(M(u)\|M(u')) \leq \epsilon$, where D_α is the Rényi divergence of order α defined as:*

$$D_\alpha(P\|Q) \triangleq \frac{1}{\alpha - 1} \log\left(\mathbb{E}_{X\sim Q}\left[\left(\frac{P(X)}{Q(X)}\right)^\alpha\right]\right).$$

RDP provides strong guarantees regarding the concept of sequential *composition*. If M_1 and M_2 are (α, ϵ_1)-RDP and (α, ϵ_2)-RDP, respectively, then the mechanism $M_{1,2}$ defined as $M_{1,2}(x) \triangleq (M_1(x), M_2(x))$ is $(\alpha, \epsilon_1 + \epsilon_2)$-RDP [41, Proposition 1]. This property enables straightforward tracking of cumulative privacy loss for iterative mechanisms. If each iteration of an iterative mechanism is (α, ϵ)-RDP, then K iterations of the mechanism are $(\alpha, K\epsilon)$-RDP. We use this property to analyze our proposed mechanism in Section 4.1.

A common tool for achieving RDP is the *Gaussian mechanism*. The Gaussian mechanism evaluates a vector-valued function on the input and adds Gaussian noise independently to each coordinate of the output. The noise magnitude is calibrated to the function's ℓ_2 sensitivity.

[7] Symmetry of adjacency relation implies: $\mathbb{P}[M(u) \in O] \geq e^{-\epsilon}\mathbb{P}[M(u') \in O] - e^{-\epsilon}\delta \geq e^{-\epsilon}\mathbb{P}[M(u') \in O] - \delta$.

[8] When $\delta = 0$, McSherry and Talwar [39] show that mechanisms satisfying ϵ-differential privacy make truth-telling an $(\exp(\epsilon) - 1)$-approximately dominant strategy. However, we are not aware of any existing result for the case $\delta > 0$. Therefore, for completeness, we provide a proof of Lemma 7.

Definition 9 (L2 sensitivity). *The ℓ_2 sensitivity of $f : \mathcal{U}^n \mapsto \mathbb{R}^m$ is defined as:*

$$\Delta_2(f) \triangleq \max_{adj\, u,u' \in \mathcal{U}^n} \|f(u) - f(u')\|_2.$$

Given this definition, the Gaussian mechanism is formally defined as follows.

Definition 10 (Gaussian mechanism). *Let $\mathcal{N}(\mu, \Sigma)$ denote a multivariate normal distribution with mean vector μ and covariance matrix Σ. For $\alpha > 1$, $\epsilon > 0$, and function $f : \mathcal{U}^n \mapsto \mathbb{R}^m$ with an ℓ_2 sensitivity of $\Delta_2(f)$, the Gaussian mechanism $M^G_{f,\alpha,\epsilon}$ is defined as:*

$$M^G_{f,\alpha,\epsilon}(u) \triangleq \mathcal{N}(f(u), \sigma^2 I_m),$$

where I_m is the $m \times m$ identity matrix, and $\sigma^2 = \alpha\Delta_2^2(f)/2\epsilon$.

$M^G_{g,\alpha,\epsilon}$ is (α, ϵ)-RDP [41, Corollary 3]. Moreover, if a mechanism is (α, ϵ)-RDP, then it is $(\epsilon + \log(1/\delta)/(\alpha - 1), \delta)$-DP for any $0 < \delta < 1$ [41, Proposition 3].

3 ALGORITHM

In this section, we present PPGA (Algorithm 1), an algorithm that directly maximizes a *smoothed* version of the NW objective in a DP manner. Our approach involves a transformation of the objective into a *separable* form. Initially, we reframe the optimization problem of Lemma 3 into a consensus problem. Next, we convert the consensus problem into a distributed optimization using ADMM. Finally, to ensure truthfulness, we deploy the Gaussian mechanism.

3.1 Distributed Maximization of Nash Welfare

The NW objective function, $\sum_i \log(U_i(z))$, poses two challenges. First, it is undefined when any agent receives zero utility. Second, it is non-separable, as the shared variable z appears in all terms. To address the first issue, we use a smooth version of the NW objective: $\sum_i \log(U_i(z) + \upsilon)$, where $\upsilon > 0$ is a small smoothing parameter that vanishes asymptotically as the number of agents increases. To tackle the second issue, we introduce local variables x_i for each agent and a shared global variable z:

$$\begin{aligned} \text{Max.} \quad & \theta(x), && \\ \text{s.t.} \quad & z = x_i \quad && \forall i \in 1, \ldots, n, \\ & x_i \in \mathcal{Z} \quad && \forall i \in 1, \ldots, n, \end{aligned} \tag{4}$$

where $\theta(x)$ is defined for $x = (x_1, \ldots, x_n)$ as:

$$\theta(x) \triangleq \sum_i \theta_i(x_i) = \sum_i \log(U_i(x_i) + \upsilon).$$

This is referred to as the *global variable consensus problem*, as it requires all local variables to reach agreement by being equal. Consensus transforms the additive objective, which does not split, into a separable objective, which splits easily.

The partial augmented Lagrangian [28, 46] for (4) is defined as:

$$L^\rho(x, z, \gamma) \triangleq \sum_i L_i^\rho(x_i, z, \gamma_i) = \sum_i \left(\theta_i(x_i) - \gamma_i^T(x_i - z) - \frac{\rho}{2}\|x_i - z\|_2^2\right),$$

where γ_i is a dual variable corresponding to the constraint $z = x_i$, and $\rho > 0$ is a penalty parameter. Note that, similar to θ, the function L^ρ is separable in x and splits into separate components L_i^ρ for

each agent i. We next apply ADMM to solve (4) in a distributed way through the following iterative updates:

$$x_i^{(k)} := \underset{x_i \in \mathcal{Z}}{argmax}\ L_i^{\rho}(x_i, z^{(k-1)}, \gamma_i^{(k-1)}) \qquad \forall i \in 1, \ldots, n, \tag{5a}$$

$$z^{(k)} := \underset{z}{argmax}\ L^{\rho}(x^{(k)}, z, \gamma^{(k-1)}), \tag{5b}$$

$$\gamma_i^{(k)} := \gamma_i^{(k-1)} + \rho(x_i^{(k)} - z^{(k)}) \qquad \forall i \in 1, \ldots, n. \tag{5c}$$

In (5a), $x_i^{(k)}$'s can be computed independently for each agent i. Moreover, we can solve (5b) exactly by setting the gradient $\partial L^{\rho}/\partial z = \sum_i \left(\gamma_i^{(k-1)} + \rho(x_i^{(k)} - z^{(k)})\right)$ to zero, which leads to the following closed-form solution:

$$z^{(k)} = \frac{1}{n}\sum_i x_i^{(k)} + \frac{1}{n\rho}\sum_i \gamma_i^{(k-1)}. \tag{6}$$

We can find an optimal solution to (4) through ADMM's iterative updates. However, this procedure is not truthful. To address this limitation, we next incorporate DP into the process as a means of achieving truthfulness.

3.2 DP for Maximizing Nash Welfare

To illustrate our proposed mechanism, it might be beneficial to interpret ADMM as an interactive process. At iteration k, each agent i calculates the local variable $x_i^{(k)}$ autonomously. Given $z^{(k-1)}$ and $\gamma_i^{(k-1)}$, the value of $x_i^{(k)}$ depends solely on agent i's own utility. With $x_i^{(k)}$ and $z^{(k)}$ known, each agent i independently calculates $\gamma_i^{(k)}$. These local variables are then submitted by agents, aggregated by the algorithm, and used to compute the global variable $z^{(k)}$. This resultant global variable is broadcast back to the agents for the next iteration.

In the context of this interactive process, to ensure DP, it is imperative that the value of the global variable remains insensitive to any individual local variable. To achieve this, we employ the Gaussian mechanism, adding a normal random vector $q^{(k)}$ to $z^{(k)}$:

$$z^{(k)} = \frac{1}{n}\sum_i x_i^{(k)} + \frac{1}{n\rho}\sum_i \gamma_i^{(k-1)} + q^{(k)}. \tag{7}$$

According to (5c), we have:

$$\sum_i \gamma_i^{(k)} = \sum_i \left(\gamma_i^{(k-1)} + \rho(x_i^{(k)} - z^{(k)})\right). \tag{8}$$

Replacing $z^{(k)}$ from (7) into (8), we get $\sum_i \gamma_i^{(k)} = -\rho n q^{(k)}$, which is used to rewrite (7) as:

$$z^{(k)} = \frac{1}{n}\sum_i x_i^{(k)} - q^{(k-1)} + q^{(k)}. \tag{9}$$

This update rule shows how $z^{(k)}$ can be calculated by adding Gaussian noise to the average of $x_i^{(k)}$'s. The magnitude of the noise can be adjusted to achieve a desired DP guarantee.

Algorithm 1 shows the pseudocode of our proposed (differentially) private public-good allocation mechanism, PPGA. The algorithm takes as parameters K, υ, ϵ, δ, and α. K specifies the number of iterations. υ controls the smoothness of the objective function. ϵ, δ, and α together determine the desired level of privacy–and, consequently, the level of truthfulness. Specifically, ϵ and δ define the level of DP, while α controls the variance of the Gaussian noise (see Theorem 12).

Algorithm 1: Private public-good allocation (PPGA)

Parameters: $K \in \mathbb{Z}$, $v, \epsilon, \delta \in (0,1)$, $\alpha > 1$

$\epsilon' \leftarrow (1/K)(\epsilon - \log(1/\delta)/(\alpha - 1))$;

$\sigma^2 \leftarrow \alpha / n^2 \epsilon'$;

$q^{(0)}, z^{(0)}, \gamma_i^{(0)}, x_i^{(0)} = \mathbf{0_m} \quad \forall i \in 1, \ldots, n$;

for $k = 1, \ldots, K$ **do**

  $x_i^{(k)} \leftarrow argmax_{x_i \in \mathcal{Z}}(L_i^{\rho}(x_i, z^{(k-1)}, \gamma_i^{(k-1)})) \quad \forall i \in 1, \ldots, n$;

  $q^{(k)} \sim \mathcal{N}(0, \sigma^2 I_m)$;

  $z^{(k)} \leftarrow (1/n) \sum_i x_i^{(k)} + q^{(k)} - q^{(k-1)}$;

  $\gamma_i^{(k)} \leftarrow \gamma_i^{(k-1)} + \rho(x_i^{(k)} - z^{(k)}) \quad \forall i \in 1, \ldots, n$;

end

$\bar{z} \leftarrow (1/K) \sum_{k=1}^{K} z^{(k)}$;

$\hat{z} \leftarrow \Pi_{\mathcal{Z}}(\bar{z})$;

Output: $\hat{z}$

At each iteration k, the optimal allocation $x_i^{(k)}$ is computed for each agent i, given $\gamma_i^{(k-1)}$ and $z^{(k-1)}$. This step can be executed in parallel for all agents. The algorithm then computes $z^{(k)}$ as a noisy average of the $x_i^{(k)}$'s. Given $z^{(k)}$ and $x_i^{(k)}$, the value $\gamma_i^{(k)}$ is then computed for each agent for the next iteration. After K iterations, the algorithm calculates $\bar{z}$, the time average of the $z^{(k)}$'s, and returns $\hat{z}$, the Euclidean projection of $\bar{z}$ onto $\mathcal{Z}$.[9]

3.3 Discussion

The integration of DP into ADMM inherently presents a trade-off between accuracy and privacy (truthfulness). Achieving a more accurate MNW solution requires a higher number of iterations. Fixing the amount of privacy loss per iteration, a higher number of iterations means a higher cumulative privacy loss, resulting in a weaker privacy guarantee. On the other hand, achieving a stronger privacy guarantee requires a lower cumulative privacy loss. Fixing the number of iterations, a lower cumulative privacy loss means a higher level of noise per iteration, resulting in diminished accuracy.

The expected value of the noise magnitude at each iteration of Algorithm 1 is:

$$\mathbb{E}\left[\|q^{(k)}\|_2^2\right] = m\sigma^2 = \frac{Km\alpha}{n^2(\epsilon - \log(1/\delta)/(\alpha - 1))}.$$

Assuming that $m = o(\sqrt{n})$, if we choose $K = \Theta(n)$, $\epsilon = \Theta(1/\log(n))$, and $\delta = \Theta(1/\sqrt{n})$, and set $\alpha = 2\log(1/\delta)/\epsilon + 1$, then the expected noise magnitude at each iteration converges to zero as n grows large–an essential property for achieving an asymptotic core outcome (see Section 4.2).

As a final remark, even though we described the algorithm as an interactive process in Section 3.2, we emphasize that our proposed algorithm is neither online nor interactive. All computations are carried out by the algorithm itself, rather than by the agents. Agents submit their private utility vectors and, at the end, observe a final allocation vector. As we show in Section 4, the algorithm satisfies DP, ensuring that agents' data remains private. Moreover, our mechanism guarantees asymptotic truthfulness, meaning that as n grows, agents have no incentive to misreport their utilities.

[9] $\Pi_{\mathcal{Z}}(z) = argmin_{z' \in \mathcal{Z}} \|z - z'\|_2^2$.

4 ANALYSIS

In this section, we first show that Algorithm 1 guarantees asymptotic truthfulness. We then demonstrate that it produces an asymptotic core solution with high probability. Finally, we analyze its computational complexity. All omitted proofs are provided in Appendix D.

4.1 Asymptotic Truthfulness

To analyze the end-to-end privacy guarantee of Algorithm 1, we separately analyze the DP guarantee of each iteration. Leveraging the properties of the Gaussian mechanism, we show that each iteration of the algorithm ensures (α, ϵ')-RDP. With the additivity property of RDP [41, Proposition 1], after K iterations, Algorithm 1 achieves $(\alpha, K\epsilon')$-RDP. It then follows from [41, Proposition 3] that Algorithm 1 is (ϵ, δ)-DP.

Lemma 11. *Algorithm 1 is (ϵ, δ)-DP.*

Proof. Algorithm 1 consists of K iterations. At each iteration k, the private data is $x^{(k)}$, while the publicly released data is $z^{(k)}$. Note that $\gamma^{(k)}$ is not publicly released, as each $\gamma_i^{(k)}$ is privately computed for each agent i. The z-update step at Line 8 of Algorithm 1 directly applies the Gaussian mechanism to the function $f(x) = \frac{1}{n}\sum_i x_i$. Let x and x' be two adjacent inputs that differ only in their i^{th} element, i.e., $x_i \neq x_i'$. Then, we have:

$$\|f(x) - f(x')\|_2 = \frac{1}{n}\|x_i - x_i'\|_2.$$

Since $x_i, x_i' \in [0,1]^m$ and $\|x_i\|_1, \|x_i'\|_1 \leq 1$, it follows that:

$$\|x_i - x_i'\|_2 \leq (\|x_i\|_2^2 + \|x_i'\|_2^2)^{1/2} \leq \sqrt{2}. \tag{10}$$

This implies $\Delta_2(f) \leq \sqrt{2}/n$. By [41, Corollary 3], each iteration k of the algorithm is (α, ϵ')-RDP. Consequently, by [41, Proposition 1], the composition of the K iterations satisfies $(\alpha, \bar{\epsilon})$-RDP, where $\bar{\epsilon} = K\epsilon' = \epsilon - \log(1/\delta)/(\alpha - 1)$. Finally, by [41, Proposition 3], the K iterations of Algorithm 1 satisfy (ϵ, δ)-DP. It is important to note that computing $\bar{z}$ after the K iterations and projecting it onto $\mathcal{Z}$ are merely post-processing steps. Since DP is immune to post-processing [16, Proposition 2.1], these steps do not affect the privacy guarantees[10]. □

We next establish our first technical result:

Theorem 12. *Algorithm 1 is asymptotically truthful.*

Proof. By Lemma 11, Algorithm 1 is (ϵ, δ)-DP. It then follows directly from Lemma 7 that it is also (ϵ, δ)-truthful. Setting $\delta = \Theta(1/\sqrt{n})$ and $\epsilon = \Theta(1/\log(n))$, we conclude that Algorithm 1 is asymptotically truthful. □

4.2 Asymptotic Core

Let $x = (x_1, \ldots, x_n)$ and $\gamma = (\gamma_1, \ldots, \gamma_n)$, and define $w \triangleq (x, z, \gamma) \in \mathcal{W} \triangleq (\mathcal{Z}^n, \mathbb{R}^m, \mathbb{R}^{mn})$. Let $w^{(k)} \triangleq (x^{(k)}, z^{(k)}, \gamma^{(k)})$, and define $G \triangleq (I_m, \ldots, I_m)$. To show that $\hat{z}$ is an approximate core solution, we aim to derive an upper bound on $\max_{z \in \mathcal{Z}} \sum_i \frac{U_i(z)}{U_i(\hat{z})+v}$. To this end, we proceed in three steps. First, we bound $\max_{z \in \mathcal{Z}} \sum_i \frac{U_i(z)}{U_i(\bar{x}_i)+v}$, where $\bar{x} = \frac{1}{K}\sum_k x^{(k)}$. Second, we establish a bound on the distance between $\hat{z}$ and each $\bar{x}$. Finally, using the smoothness of U_i's and applying Lemma 5, we conclude that $\hat{z}$ is an approximate core solution.

[10] If M is (ϵ, δ)-DP, then applying any randomized mapping f to $M(u)$ preserves the (ϵ, δ)-DP property.

Lemma 13. *Let $\{w^{(k)}\}$ and $\{q^{(k)}\}$ be sequences generated by Algorithm 1. Then, we have:*

$$\frac{1}{n}\sum_i \frac{U_i(z)}{U_i(\bar{x}_i)+v} \le 1 + \frac{\rho}{K}\sum_{k=1}^{K}(z^{(k)}-z)^T q^{(k)} + \frac{\rho}{2K} \qquad \forall z \in \mathcal{Z}. \tag{11}$$

The next lemma provides an upper bound on the distance between $\hat{z}$ and any $\bar{x}_i$.

Lemma 14. *Let $\{w^{(k)}\}$ and $\{q^{(k)}\}$ be sequences generated by Algorithm 1. Let $w^* = (x^*, z^*, \gamma^*)$ be an optimal solution to (4), with $x_i^* = z^*$ for all i. Then, we have:*

$$\|\bar{x} - G\hat{z}\|_2 \le \frac{2\sqrt{n}}{K}\left(\sum_{k=1}^{K}|(z^{(k)}-z^*)^T q^{(k)}|\right)^{\frac{1}{2}} + \frac{\sqrt{2n}}{K} + \frac{2}{\rho K}\|\gamma^*\|_2. \tag{12}$$

The right-hand side of (11) and (12) involves random variables–specifically, the sequences $\{z^{(k)}\}$ and $\{q^{(k)}\}$. The next lemma provides a bound on their tail behavior:

Lemma 15. *Let $\{w^{(k)}\}$ and $\{q^{(k)}\}$ be sequences generated by $K = \Theta(n)$ iterations of Algorithm 1. Suppose ϵ, δ, and α are chosen such that $m\sigma^2 < 1$. Then, for any $z \in \mathcal{Z}$ and some constant $C > 0$, with probability at least $1 - 1/n - 1/n^m$, we have:*

$$\frac{1}{K}\sum_{k=1}^{K}\left|(z^{(k)}-z)^T q^{(k)}\right| \le C\sqrt{m}\sigma. \tag{13}$$

Next, we establish that the outcome of Algorithm 1 is an approximate core solution:

Lemma 16. *Suppose that $m = o(\sqrt{n})$. Then, after $K = \Theta(n)$ iterations, Algorithm 1 returns an $(\varepsilon_1, \varepsilon_2(1+\varepsilon_1))$-core outcome with probability at least $1 - 1/n - 1/n^m$, where:*

$$\varepsilon_1 = O\left(\sqrt{\frac{m\log(1/\delta)}{n\epsilon^2}}\right) \quad \text{and} \quad \varepsilon_2 = O\left(\sqrt[4]{\frac{m\log(1/\delta)}{n\epsilon^2}}\right).$$

Finally, we establish Algorithm 1's asymptotic fairness.

Theorem 17. *Suppose that $m = o(\sqrt{n})$. Then, after $K = \Theta(n)$ iterations, the output of Algorithm 1 is an asymptotic core outcome with probability at least $1 - 1/n - 1/n^m$.*

Proof. If we choose $v = \Theta(1/\sqrt{n})$, $\delta = \Theta(1/\sqrt{n})$, and $\epsilon = \Theta(1/\log(n))$, and set $\alpha = 2\log(1/\delta)/\epsilon + 1$, then, by Lemma 16, Algorithm 1 returns an asymptotic core outcome. □

4.3 Computational Complexity

The main computation in each iteration k of Algorithm 1 is to compute $x_i^{(k)}$ for each agent i. This step involves solving a convex program. Several methods exist for solving broad classes of convex optimization problems with a number of operations that grow polynomially in the problem dimensions and logarithmically in $1/\xi$, where $\xi > 0$ denotes the desired accuracy [44]. Typically, such accuracy guarantees are provided with respect to the objective function value. For the subproblem in Line 6 of Algorithm 1, however, we require an accuracy guarantee on the first-order optimality condition (see (18)). Fortunately, due to the smoothness of U_i, such a guarantee can still be achieved in polynomial time using first-order methods–such as the one proposed by Lu and Mei [38].

Lemma 18. *For any $x_i \in \mathcal{Z}$ and $z, \gamma_i \in \mathbb{R}^m$, a point $x_i^* \in \mathcal{Z}$ that satisfies the inequality*

$$(x_i - x_i^*)^T(\nabla\theta_i(x_i^*) - \gamma_i - \rho(x_i^* - z)) \le \xi \tag{14}$$

can be computed in time $O(m\log(m)\sqrt{(L+\rho)/\xi}\log(1/\xi))$, where L is $\nabla\theta_i$'s Lipschitz constant.

We now establish our final technical result.

Theorem 19. *Algorithm 1 achieves asymptotic truthfulness and computes an asymptotic core solution with high probability in polynomial time.*

Proof. Consistent with the proof of Theorem 17, let $K = \Theta(n)$ and $v = \Theta(1/\sqrt{n})$. Then, $L \leq \frac{(1+\beta)^2}{2v^2} + \frac{\beta}{v} = \Theta(n)$. By Lemma 18, at each iteration k and for each agent i, we can compute a point $\hat{x}_i^{(k)}$ satisfying the following inequality for any $z \in \mathcal{Z}$:

$$(z - \hat{x}_i^{(k)})^T (\nabla \theta_i(\hat{x}_i^{(k)}) - \gamma_i^{(k-1)} - \rho(\hat{x}_i^{(k)} - z^{(k-1)})) \leq \xi.$$

Suppose we modify Line 6 of Algorithm 1 by replacing $x_i^{(k)}$ with $\hat{x}_i^{(k)}$. This modification does not affect the asymptotic truthfulness guarantee of the algorithm. However, it slightly alters the algorithm's approximation of the core. In particular, it can be verified that modified versions of Lemma 13 and Lemma 14 continue to hold, with additive error terms of ξ and $2\sqrt{n\xi}/(\sqrt{\rho}K)$ on the right-hand sides of (11) and (12), respectively.

Choosing $\xi = \Theta(1/\sqrt{n})$ ensures that Lemma 16, and therefore Theorem 17, continue to hold for the modified algorithm. Thus, the modified algorithm preserves the asymptotic guarantees of the original, while achieving a total running time of $O(n^{2.75} \log(n) m \log(m))$. □

5 EXPERIMENTS

In this section, we aim to show that PPGA can be deployed in practice to solve large-scale public-good allocation problems. To this end, we implement Algorithm 1 in Python using `CVXPY`, an open-source Python-embedded modeling language for convex optimization problems [14]. PPGA is highly parallelizable, particularly in the concurrent computation of x and γ for all agents. We leverage this feature in our implementation by distributing the computational workload across multiple processes using Python's `multiprocessing` package. The code for our implementation is provided at https://github.com/uwaterloo-mast/PPGA.

To conduct experiments, we leverage real-world data from Pabulib.org, an open participatory budgeting library [19]. Our experiments focus on 12 election instances, selected primarily based on the size of their voter population and the average number of approved projects per voter[11]. Each instance involves a collection of projects with associated costs and a designated total budget. Voters express their preferences for the projects by casting approval votes for one or more projects. We summarize the key characteristics of these election instances in Appendix B, and full details of each instance, such as project costs, are provided with our code (located in the `final_data` folder).

As just mentioned, the instances involve approval votes and indivisible projects. We utilized these instances to derive new ones wherein agents have cardinal utilities, and fractional allocations are deemed acceptable. Fractional budget allocations are inspired by the motivating examples in the introduction and various related works [10, 17, 22, 23]. We transform approval votes into cardinal utilities according to the *cost-utility* approach [19] using the following procedure: For each voter i and project j, we set $u_{ij} = 0$ if voter i does not approve project j, and $u_{ij} = 1$ otherwise. This ensures that voters' utilities are proportional to the budget allocated to the projects they support[12].

In the concluding remarks of Section 3.2, we provide guidelines for the DP parameters to guarantee our asymptotic properties. There are also established practical norms for acceptable ϵ and δ values. Following these norms, we set $\epsilon = c_\epsilon/\log(n)$, $\delta = c_\delta/\sqrt{n}$, and $K = c_K n$, where $c_\epsilon = 1.5$,

[11] We selected representative instances from about 60 instances that had at least 10k votes.

[12] Let P_i be the set of projects supported by voter i. Then, i's utility is given by $U_i(z) = \sum_{j \in P_i} z_j$, where $z_j \leq s_j/c$ represents the fraction of the total budget allocated to project j. This ensures that i's utility is proportional to the budget allocated to the projects they support.

Inst.	Core's PS		PPGA's PS		SD
	Min ($\times n$)	Avg	Min ($\times n$)	Avg	($\div m$)
1	90.7	0.27	111.9	0.27	0.00007
2	236.4	0.30	17.1	0.29	0.00016
3	235.5	0.18	191.1	0.18	0.00014
4	216.1	0.39	37.5	0.38	0.00023
5	15.0	0.33	14.3	0.33	0.00010
6	244.7	0.39	39.9	0.38	0.00030
7	11.0	0.29	11.1	0.29	0.00045
8	122.6	0.33	128.3	0.32	0.00008
9	163.5	0.34	168.0	0.34	0.00002
10	154.4	0.16	106.9	0.16	0.00034
11	519.8	0.45	513.4	0.45	0.00002
12	261.3	0.57	130.0	0.57	0.00003

Table 1. Proportionality score and statistical distance.

$c_\delta = 0.3$, and $c_K = 0.001$. We further set α such that $\log(1/\delta)/(\alpha - 1) = \epsilon/2$. This way, values for ϵ and δ approximate 0.3 and 0.001, respectively, keeping the noise magnitude, $\mathbb{E}\left[\|q^{(k)}\|_2^2\right]$, under 3e-4 for the majority of instances. We note that in our experiments, we set $\upsilon = 0$. The introduction of υ as a parameter was solely motivated by a technical requirement to ensure that θ_i is a smooth function. However, this smoothness condition has negligible practical significance.

We compare PPGA with the core[13] using the following metrics:

- **Social welfare (SW)** for an allocation z is defined as $\frac{1}{n}\sum_i U_i(z)$. SW serves as an indicator of the overall satisfaction achieved collectively by all agents from the allocation.
- **Proportionality score (PS)** of voter i for an allocation z is defined as the ratio of i's utility for z to i's maximum attainable utility, i.e., $\frac{U_i(z)}{\max_{z' \in \mathcal{Z}} U_i(z')}$. If the PS value is $\geq 1/n$ for all voters (or equivalently, if the minimum value of PS across voters multiplied by n is ≥ 1), then the allocation is *proportional* ($|A| = 1$ in Defenision 2). We report both the minimum (multiplied by n) and the average of PS values across all voters.
- **Statistical distance (SD)** between an allocation z and a core solution z^* is measured by their *total variation distance*, defined as $\frac{1}{2}\|z - z^*\|_1$. Two allocations over m items are considered statistically close if their total variation distance is a negligible function in m. To facilitate comparison, we normalize the total variation distance by dividing it by m.

For each metric, we report the average value over 50 runs.

Figure 1 illustrates the social welfare under PPGA normalized to that under the core solution, while Table 1 summarizes proportionality scores and statistical distances across all election instances. These results uncover several crucial insights. Firstly, the statistical distance between the budget allocation under PPGA and the core solution remains consistently close to zero in all instances, hovering below 0.00045 for all cases. Secondly, the observed discrepancy in social welfare values between PPGA and the core solution consistently falls below 3% across all election instances. Lastly, the minimum PS value $\times n$ exceeds 1 for all instance, indicating that PPGA satisfies the proportionality criteria for all instances. The average PS values tend to be slightly higher under the core solution for some instances, but the discrepancy between the average PS values under PPGA

[13]We find the core by solving the convex optimization of Lemma 3 using Algorithm 1 without adding noise.

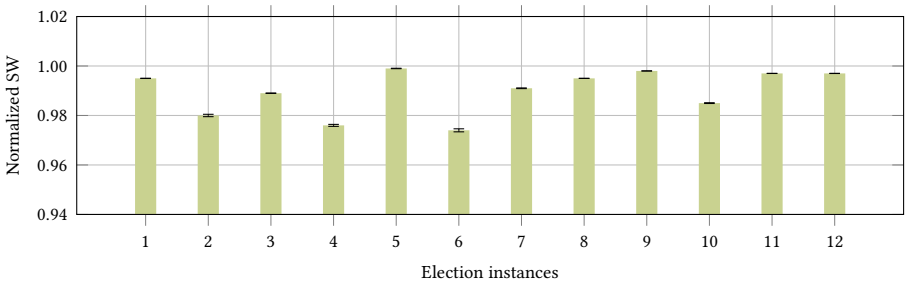

Fig. 1. Social welfare of PPGA normalized to that of core (w/ 95% confidence band).

and the core remains below 4% in all instances. Collectively, these findings strongly signify the high level of fairness achieved by PPGA.

Our empirical findings not only corroborate the theoretical results in Section 4 but also illustrate that PPGA yields solutions that are statistically close to the core solution for all election instances. We expect this result to hold for any instance with a large population and a linear utility model. This expectation is based on our proof in Theorem 17, wherein we demonstrate that the distance between $\bar{x}$ and $\hat{z}$ asymptotically approaches zero with high probability. For any linear utility model, it can be shown that the distance between $\bar{x}$ and z^* also asymptotically approaches zero with high probability. Consequently, the statistical distance between $\hat{z}$ and z^* is asymptotically negligible for any linear utility model. For other concave utility models, the statistical distance between $\hat{z}$ and z^* might be higher, depending on the curvature of the function. Nevertheless, one can demonstrate that the difference in the value of the NW objective for $\hat{z}$ and z^* asymptotically approaches zero with high probability, implying similar results for PS.

6 RELATED WORKS

Fair resource allocation without money (also known as cake cutting) has been extensively studied in the literature for private goods [47]. For public goods, the fair allocation problem has been studied in various contexts, including fair public decision-making [11], multi-agent knapsack problems [20], multi-winner elections [43], and participatory budgeting [45]. The truthful aggregation of agents' preferences has also been explored in public decision-making [10, 22, 23, 27, 48]. However, the settings in these works differ from ours, as they aim to maximize social welfare and focus on ℓ_1 preferences[14], whereas our focus is on concave preferences and maximizing Nash welfare.

The work most closely related to this paper is that of Fain et al. [17][15], which finds an approximate core solution with high probability while achieving approximate truthfulness. However, due to its reliance on several approximations, their approach fails to produce an asymptotic core solution. As the number of agents increases, the approximation error for fairness (core) may grow. In contrast, our approximation guarantee does not suffer from this issue. By combining the Gaussian mechanism with ADMM to directly optimize the NW objective, our method ensures asymptotic truthfulness and finds an asymptotic core solution with high probability.

Differentially private convex programming has been utilized in recent years to allocate private goods [12, 30, 31, 34, 35]. These methods often employ the dual ascent technique as a key tool [7]. The dual ascent method involves a sequence of two updates: the primal update, which

[14]An agent's disutility for an allocation is equal to the ℓ_1 distance between that allocation and the agent's most preferred allocation.

[15]Their notion of the core is based on capacity, where a blocking coalition receives a proportional share of the capacity rather than a proportional share of utility (see Defenision 2).

optimizes the Lagrangian while fixing the dual variable, and the dual update, which takes a gradient ascent step to update the dual variable given the optimized primal variable. However, the dual ascent method cannot be used for maximizing the NW objective, because, as we show in Section 3, the Lagrangian for the convex program is an affine function of some components of the primal variable. This causes the primal update to fail, as the dual problem is unbounded below for most values of the dual variable [7]. We avoid this by optimizing the augmented Lagrangian instead of the Lagrangian.

Differentially private ADMM methods have also been extensively studied [32, 33, 36, 52, 56, 57]. Although related, our work differentiates itself from these works in several aspects. Firstly, while previous studies focus on the convergence rate of the objective function, we study the convergence of a primal variable to an approximate core solution. To the best of our knowledge, our work is first to prove an asymptotic, game-theoretic property for a primal variable within differentially private ADMM. Secondly, unlike prior work that introduces noise to the local variables, PPGA adds noise to the global variable (as detailed in Section 3). Finally, many studies on differentially private ADMM rely on a restrictive assumption regarding the strong convexity of the objective function, which does not hold for the NW objective.

7 CONCLUSION

In this paper, we introduce PPGA, a mechanism designed for the fair and truthful allocation of divisible public goods. PPGA achieves fairness by directly maximizing the NW objective and ensures truthfulness by deploying the Gaussian mechanism from differential privacy. We showed that PPGA is asymptotically truthful and finds an asymptotic core solution with high probability. By conducting experiments using real-world data from participatory budgeting elections, we showcased the practical applicability of PPGA.

REFERENCES

[1] S Airiau, H Aziz, I Caragiannis, J Kruger, and J Lang. 2018. Positional social decision schemes: Fair and efficient portioning. In *Proceedings of the 7th International Workshop on Computational Social Choice (COMSOC)*.

[2] Haris Aziz, Barton E Lee, and Nimrod Talmon. 2018. Proportionally Representative Participatory Budgeting: Axioms and Algorithms. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. 23–31.

[3] Amir Beck. 2017. *First-order methods in optimization.* SIAM.

[4] Gerdus Benade, Swaprava Nath, Ariel D Procaccia, and Nisarg Shah. 2021. Preference elicitation for participatory budgeting. *Management Science* 67, 5 (2021), 2813–2827.

[5] Umang Bhaskar, Varsha Dani, and Abheek Ghosh. 2018. Truthful and near-optimal mechanisms for welfare maximization in multi-winner elections. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (AAAI)*. 925–932.

[6] Anna Bogomolnaia, Hervé Moulin, and Richard Stong. 2005. Collective choice under dichotomous preferences. *Journal of Economic Theory* 122, 2 (2005), 165–184.

[7] Stephen Boyd, Neal Parikh, Eric Chu, Borja Peleato, Jonathan Eckstein, et al. 2011. Distributed optimization and statistical learning via the alternating direction method of multipliers. *Foundations and Trends® in Machine learning* 3, 1 (2011), 1–122.

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

[9] Florian Brandl, Felix Brandt, Dominik Peters, and Christian Stricker. 2021. Distribution rules under dichotomous preferences: Two out of three ain't bad. In *Proceedings of the 22nd ACM Conference on Economics and Computation (EC)*. 158–179.

[10] Ioannis Caragiannis, George Christodoulou, and Nicos Protopapas. 2022. Truthful aggregation of budget proposals with proportionality guarantees. In *Proceedings of the 36th AAAI Conference on Artificial Intelligence (AAAI)*. 4917–4924.

[11] Vincent Conitzer, Rupert Freeman, and Nisarg Shah. 2017. Fair public decision making. In *Proceedings of the 18th ACM Conference on Economics and Computation (EC)*. 629–646.

[12] Rachel Cummings, Michael Kearns, Aaron Roth, and Zhiwei Steven Wu. 2015. Privacy and truthful equilibrium selection for aggregative games. In *Proceedings of the International Conference on Web and Internet Economics (WINE)*.

286–299.

[13] Gerard Debreu and Herbert Scarf. 1963. A limit theorem on the core of an economy. *International Economic Review* 4, 3 (1963), 235–246.

[14] Steven Diamond and Stephen Boyd. 2016. CVXPY: A Python-embedded modeling language for convex optimization. *The Journal of Machine Learning Research* 17, 1 (2016), 2909–2913.

[15] Cynthia Dwork, Frank McSherry, Kobbi Nissim, and Adam Smith. 2006. Calibrating noise to sensitivity in private data analysis. In *Proceedings of the 3rd Conference on Theory of Cryptography (TCC)*. 265–284.

[16] Cynthia Dwork and Aaron Roth. 2014. The Algorithmic Foundations of Differential Privacy. *Theoretical Computer Science* 9, 3-4 (2014), 211–407.

[17] Brandon Fain, Ashish Goel, and Kamesh Munagala. 2016. The core of the participatory budgeting problem. In *Proceedings of the 12th International Conference on Web and Internet Economics (WINE)*. 384–399.

[18] Brandon Fain, Kamesh Munagala, and Nisarg Shah. 2018. Fair allocation of indivisible public goods. In *Proceedings of the 19th ACM Conference on Economics and Computation (EC)*. 575–592.

[19] Piotr Faliszewski, Jarosław Flis, Dominik Peters, Grzegorz Pierczyński, Piotr Skowron, Dariusz Stolicki, Stanisław Szufa, and Nimrod Talmon. 2023. Participatory budgeting: data, tools, and analysis. In *Proceedings of the 32nd International Joint Conference on Artificial Intelligence (IJCAI)*. 2667–2674.

[20] Till Fluschnik, Piotr Skowron, Mervin Triphaus, and Kai Wilker. 2019. Fair knapsack. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*. 1941–1948.

[21] Duncan K Foley. 1970. Lindahl's Solution and the Core of an Economy with Public Goods. *Econometrica: Journal of the Econometric Society* 38, 1 (1970), 66–72.

[22] Rupert Freeman, David M Pennock, Dominik Peters, and Jennifer Wortman Vaughan. 2019. Truthful aggregation of budget proposals. In *Proceedings of the 20th ACM Conference on Economics and Computation (EC)*. 751–752.

[23] Rupert Freeman and Ulrike Schmidt-Kraepelin. 2024. Project-fair and truthful mechanisms for budget aggregation. In *Proceedings of the 38th AAAI Conference on Artificial Intelligence (AAAI)*. 9704–9712.

[24] Daniel Gabay and Bertrand Mercier. 1976. A dual algorithm for the solution of nonlinear variational problems via finite element approximation. *Computers & Mathematics with Applications* 2, 1 (1976), 17–40.

[25] Donald Bruce Gillies. 1953. *Some theorems on n-person games*. Princeton University.

[26] Roland Glowinski and Americo Marroco. 1975. Sur l'approximation, par éléments finis d'ordre un, et la résolution, par pénalisation-dualité d'une classe de problèmes de Dirichlet non linéaires. *Revue française d'automatique, informatique, recherche opérationnelle. Analyse numérique* 9, R2 (1975), 41–76.

[27] Ashish Goel, Anilesh K Krishnaswamy, Sukolsak Sakshuwong, and Tanja Aitamurto. 2019. Knapsack voting for participatory budgeting. *ACM Transactions on Economics and Computation (TEAC)* 7, 2 (2019), 1–27.

[28] Magnus R Hestenes. 1969. Multiplier and gradient methods. *Journal of Optimization Theory and Applications* 4, 5 (1969), 303–320.

[29] Matthew Hough and Stephen A Vavasis. 2024. A primal-dual Frank-Wolfe algorithm for linear programming.

[30] Justin Hsu, Zhiyi Huang, Aaron Roth, Tim Roughgarden, and Zhiwei Steven Wu. 2014. Private matchings and allocations. In *Proceedings of the 46th Annual ACM Symposium on Theory of Computing (STOC)*. 21–30.

[31] Justin Hsu, Zhiyi Huang, Aaron Roth, and Zhiwei Steven Wu. 2016. Jointly private convex programming. In *Proceedings of the 27th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 580–599.

[32] Zonghao Huang, Rui Hu, Yuanxiong Guo, Eric Chan-Tin, and Yanmin Gong. 2019. DP-ADMM: ADMM-based distributed learning with differential privacy. *IEEE Transactions on Information Forensics and Security* 15 (2019), 1002–1012.

[33] Zhenqi Huang, Sayan Mitra, and Nitin Vaidya. 2015. Differentially private distributed optimization. In *Proceedings of the 16th International Conference on Distributed Computing and Networking*. 1–10.

[34] Zhiyi Huang and Xue Zhu. 2018. Near optimal jointly private packing algorithms via dual multiplicative weight update. In *Proceedings of the 29th Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 343–357.

[35] Zhiyi Huang and Xue Zhu. 2019. Scalable and Jointly Differentially Private Packing. In *Proceedings of the 46th International Colloquium on Automata, Languages, and Programming (ICALP)*.

[36] Roger Iyengar, Joseph P Near, Dawn Song, Om Thakkar, Abhradeep Thakurta, and Lun Wang. 2019. Towards practical differentially private convex optimization. In *2019 IEEE Symposium on Security and Privacy (SP)*. IEEE, 299–316.

[37] László Lovász and Santosh Vempala. 2007. The geometry of logconcave functions and sampling algorithms. *Random Structures & Algorithms* 30, 3 (2007), 307–358.

[38] Zhaosong Lu and Sanyou Mei. 2023. Accelerated first-order methods for convex optimization with locally Lipschitz continuous gradient. *SIAM Journal on Optimization* 33, 3 (2023), 2275–2310.

[39] Frank McSherry and Kunal Talwar. 2007. Mechanism design via differential privacy. In *Proceedings of the 48th Annual IEEE Symposium on Foundations of Computer Science (FOCS)*. 94–103.

[40] Marcin Michorzewski, Dominik Peters, and Piotr Skowron. 2020. Price of fairness in budget division and probabilistic social choice. In *Proceedings of the 34th AAAI Conference on Artificial Intelligence (AAAI)*. 2184–2191.

[41] Ilya Mironov. 2017. Rényi differential privacy. In *Proceedings of the 30th IEEE Computer Security Foundations Symposium (CSF)*. 263–275.
[42] Thomas J Muench. 1972. The core and the Lindahl equilibrium of an economy with a public good: An example. *Journal of Economic Theory* 4, 2 (1972), 241–255.
[43] Kamesh Munagala, Yiheng Shen, Kangning Wang, and Zhiyi Wang. 2022. Approximate Core for Committee Selection via Multilinear Extension and Market Clearing. In *Proceedings of the Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*. 2229–2252.
[44] Yurii Nesterov and Arkadii Nemirovskii. 1994. *Interior-point polynomial algorithms in convex programming*. SIAM.
[45] Dominik Peters, Grzegorz Pierczyński, and Piotr Skowron. 2021. Proportional participatory budgeting with additive utilities. *Proceedings of the 35th Conference on Neural Information Processing Systems (NeurIPS)*, 12726–12737.
[46] Michael JD Powell. 1969. A method for nonlinear constraints in minimization problems. , 283–298 pages.
[47] Ariel D Procaccia. 2013. Cake cutting: Not just child's play. *Commun. ACM* 56, 7 (2013), 78–87.
[48] Ariel D Procaccia and Moshe Tennenholtz. 2009. Approximate mechanism design without money. In *Proceedings of the 10th ACM Conference on Electronic Commerce (EC)*. 177–186.
[49] Walter Rudin. 1976. *Principles of Mathematical Analysis* (3rd ed.). McGraw-Hill.
[50] Ernest K Ryu and Stephen Boyd. 2016. Primer on monotone operator methods. *Appl. Math. Comput.* 15, 1 (2016), 3–43.
[51] Herbert E Scarf. 1967. The core of an N person game. *Econometrica: Journal of the Econometric Society* 35, 1 (1967), 50–69.
[52] Wei Shi, Qing Ling, Kun Yuan, Gang Wu, and Wotao Yin. 2014. On the linear convergence of the ADMM in decentralized consensus optimization. *IEEE Transactions on Signal Processing* 62, 7 (2014), 1750–1761.
[53] Robert L Smith. 1984. Efficient Monte Carlo procedures for generating points uniformly distributed over bounded regions. *Operations Research* 32, 6 (1984), 1296–1308.
[54] Nimrod Talmon and Piotr Faliszewski. 2019. A framework for approval-based budgeting methods. In *Proceedings of the 33rd AAAI Conference on Artificial Intelligence (AAAI)*. 2181–2188.
[55] Roman Vershynin. 2018. *High-dimensional probability: An introduction with applications in data science*. Vol. 47. Cambridge University Press.
[56] Tao Zhang and Quanyan Zhu. 2016. Dynamic differential privacy for ADMM-based distributed classification learning. *IEEE Transactions on Information Forensics and Security* 12, 1 (2016), 172–187.
[57] Xueru Zhang, Mohammad Mahdi Khalili, and Mingyan Liu. 2018. Improving the privacy and accuracy of ADMM-based distributed algorithms. In *Proceedings of the 35th International Conference on Machine Learning (ICML)*. 5796–5805.

A NOTATIONS

Notation	Description
n	Number of agents
m	Number of public items
s_j	Size of item j
s	Size vector, i.e., $(s_1, \ldots, s_m)$
c	Total capacity
z_j	Fraction of the total capacity that is allocated to item j
z	Allocation variable, i.e., $(z_1, \ldots, z_m)$
$\mathcal{Z}$	Set of all feasible allocations, i.e., $\{z \in [0,1]^m \mid \|z\|_1 \leq 1,\ cz \leq s\}$
$U_i(z)$	Agent i's utility function for allocation z
u_i	Agent i's utility vector, i.e., parameters of U_i: $(u_{i1}, \ldots, u_{id})$
$\mathcal{U}$	Set $[0,1]^d$
u	Utility vectors for all agents, i.e., $(u_1, \ldots, u_n)$
u_{-i}	Utility vector of all agents except agent i, i.e., $(u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_n)$
$M(u)$	Randomized mechanism that maps $u \in \mathcal{U}^n$ to probability distribution over $\mathcal{Z}$
x_i	Feasible allocation of agent i
x	Vector of allocations, i.e., $(x_1, \ldots, x_n) \in \mathcal{Z}^n$
$\theta_i(x_i)$	Smoothed logarithm of agent i's utility, i.e., $\log(U_i(x_i) + \epsilon)$
$\theta(x)$	Summation of θ_i's: $\sum_i \theta_i(x_i)$
L	Lipschitz parameter of $U_i(z)$'s
ϵ	Multiplicative approximation factor for truthfulness, core, and DP
δ	Additive approximation factor for truthfulness, core, and DP
α	Rényi divergence parameter
$\mathcal{N}(\mu, \Sigma)$	Multivariate normal distribution with mean vector μ and covariance matrix Σ
K	Total number of iterations in Algorithm 1
$z^{(k)}$	Global allocation variable at iteration k, i.e., $(z_{i1}^{(k)}, \ldots, z_{im}^{(k)})$
$x_i^{(k)}$	Agent i's local allocation variable at iteration k, i.e., $(x_{i1}^{(k)}, \ldots, x_{im}^{(k)})$
$x^{(k)}$	Vector of local allocations at iteration k, i.e., $(x_1^{(k)}, \ldots, x_n^{(k)})$
$\gamma_i^{(k)}$	Dual variable for $z = x_i$ constraint at iteration k, i.e., $(\gamma_{i1}^{(k)}, \ldots, \gamma_{im}^{(k)})$
$\gamma^{(k)}$	Vector of dual variables, i.e., $(\gamma_1^{(k)}, \ldots, \gamma_n^{(k)})$
$q^{(k)}$	Multivariate Gaussian noise added to $z^{(k)}$ at iteration k
σ^2	Variance of added noise to each dimension of z
ρ	Penalty parameter for the augmented Lagrangian
L_i^ρ	Agent i's partial augmented Lagrangian with parameter ρ
L^ρ	Summation of partial augmented Lagrangian functions, i.e., $sum_i L_i^\rho$
η	Regularization parameter for the linearized augmented Lagrangian
$L_i^{\rho,\eta}$	Agent i's linearized partial augmented Lagrangian with parameters ρ and η
$\Pi_{\mathcal{Z}}(z)$	Euclidean projection of z onto $\mathcal{Z}$, i.e., $argmin_{z' \in \mathcal{Z}} \|z - z'\|_2^2$
$\bar{z}$	Time average of $z^{(k)}$'s, i.e., $(1/K) \sum_{k=1}^K z^{(k)}$
$\hat{z}$	Euclidean projection of $\bar{z}$ onto $\mathcal{Z}$, i.e., $\Pi(\bar{z})$

Table . List of notations

B ELECTION INSTANCES

Inst.	Election	# Voters (n)	# Proj. (m)	Budget (c)	Avg. # votes per voter
1	Wroclaw'17	62,529	50	4,000,000	1.8
2	Warszawa'20 Praga Poludnie	14,897	134	5,900,907	9.1
3	Katowice'21	36,370	47	3,003,438	1.5
4	Warszawa'21 Mokotow	12,933	98	7,147,577	9.7
5	Wroclaw'16 Rejon NR 10-750	12,664	13	750,000	1
6	Warszawa'23 Mokotow	11,067	81	8,697,250	9.1
7	Wroclaw'16 Rejon NR 12-250	10,711	15	650,000	1
8	Wroclaw'16	67,103	52	4,500,000	1.8
9	Warszawa'22	81,234	129	28,072,528	7.9
10	Gdansk'20	30,237	28	3,600,000	1
11	Warszawa'21	95,899	106	24,933,409	8.3
12	Warszawa'20	86,721	101	24,933,409	7.2

Table 3. Characteristics of election instances.

C SUPPLEMENTARY CLAIMS

Claim 20. *Let $f : \mathcal{D} \mapsto \mathbb{R}$ be strictly positive and concave. Then, $F(x) = \frac{1}{f(x)}$ is convex.*

Proof. First, not that since f is strictly positive, F is well-defined over D. To show that F is convex, we need to verify that for any $x, x' \in D$ and any $\lambda \in [0, 1]$:

$$F(\lambda x + (1-\lambda)x') \leq \lambda F(x) + (1-\lambda)F(x').$$

Since f is concave, it satisfies $f(\lambda x + (1-\lambda)x') \geq \lambda f(x) + (1-\lambda)f(x')$ for any $\lambda \geq 0$. Because f is strictly positive, all terms are positive, and taking reciprocals reverses the inequality:

$$\frac{1}{f(\lambda x + (1-\lambda)x')} \leq \frac{1}{\lambda f(x) + (1-\lambda)f(x')} \leq \lambda \frac{1}{f(x)} + (1-\lambda)\frac{1}{f(x')},$$

where the final inequality follows from Jensen's inequality applied to the convex function $t \mapsto 1/t$ on the positive reals. Thus, $F(x)$ is convex. □

Claim 21. *If $f : \mathcal{D} \mapsto \mathbb{R}$ is concave and L-Lipschitz continuous, then $\|\nabla f\|_2 \leq L$. Also, if f is concave and $\|\nabla f\|_2 \leq L$, then f is L-Lipschitz continuous.*

Proof. First, we show that if f is concave and L-Lipschitz continuous, then $\|\nabla f\|_2 \leq L$. Since f is concave, for all $x, x' \in D$, we have:

$$f(x') \leq f(x) + \nabla f(x)^T (x' - x).$$

Set $x' = x - \delta \frac{\nabla f(x)}{\|\nabla f(x)\|_2}$ for some $0 < \delta < D$, where $D = \sup_{x,x' \in \mathcal{D}} \|x - x'\|_2$. Then, we have:

$$\delta \|\nabla f(x)\|_2 \leq f(x) - f(x') \leq L\|x - x'\|_2 = L\delta,$$

where the last inequality is due to Lipschitz continuity of f. Dividing both sides by δ yields the desired result.

Next, we show that if f is concave and $\|\nabla f\|_2 \leq L$, then f is L-Lipschitz continuous. This follows directly from concavity and boundedness of the gradient:

$$f(x') - f(x) \leq \nabla f(x)^T (x' - x) \leq \|\nabla f(x)\|_2 \|x' - x\|_2 \leq L\|x' - x\|_2.$$

By switching x and x' in the above inequality, we can show:

$$f(x) - f(x') \leq L\|x - x'\|_2,$$

for the same x and x'. □

Claim 22. *Let $f : \mathcal{D} \mapsto \mathbb{R}$ be concave, β-smooth, and bounded between 0 and 1. Suppose that $\mathcal{D}$ is a bounded convex set, such that $\sup_{x,x' \in \mathcal{D}} \|x - x'\|_2 \leq D < \infty$. Then, f is L-Lipschitz, with $L \leq \frac{1}{D} + \frac{\beta D}{2}$.*

Proof. Given that f is concave and β-smooth, for all $x, x' \in \mathcal{D}$, we have:

$$f(x) + \nabla f(x)^T(x' - x) - f(x') \leq \frac{\beta}{2}\|x' - x\|_2^2.$$

Setting $x' = x + \delta \frac{\nabla f(x)}{\|f(x)\|_2}$, where $0 < \delta \leq D$, we have:

$$\delta\|\nabla f(x)\|_2 \leq f(x') - f(x) + \frac{\beta}{2}\delta^2 \leq 1 + \frac{\beta}{2}\delta^2.$$

Dividing by $\delta > 0$ and substituting $\delta = D$ gives us the desired upper bound on $\nabla f(x)$. □

Claim 23. *Let $f : \mathcal{D} \mapsto \mathbb{R}$ be concave and bounded between 0 and 1. Suppose that $\mathcal{D}$ is a bounded convex set, such that $\sup_{x,x' \in \mathcal{D}} \|x - x'\|_2 \leq D < \infty$. If f is β-smooth, then for $v > 0$, $h(x) = \log(f(x) + v)$ has L-Lipschitz gradient (i.e., is smooth), with $L \leq \frac{M^2}{v^2} + \frac{\beta}{v}$, where $M = \frac{1}{D} + \frac{\beta D}{2}$.*

Proof. For any $x, x' \in \mathcal{D}$, we have:

$$\begin{aligned}
\|\nabla h(x) - \nabla h(x')\|_2 &= \left\| \frac{\nabla f(x)}{f(x) + v} - \frac{\nabla f(x')}{f(x') + v} \right\|_2 \\
&= \left\| \left(\frac{1}{f(x) + v} - \frac{1}{f(x') + v} \right) \nabla f(x) + \frac{\nabla f(x) - \nabla f(x')}{f(x') + v} \right\|_2 \\
&\leq \left| \frac{1}{f(x) + v} - \frac{1}{f(x') + v} \right| \|\nabla f(x)\|_2 + \frac{\|\nabla f(x) - \nabla f(x')\|_2}{f(x') + v} \\
&= \frac{|f(x) - f(x')|}{(f(x) + v)(f(x') + v)} \|\nabla f(x)\|_2 + \frac{\|\nabla f(x) - \nabla f(x')\|_2}{f(x') + v} \\
&\leq \frac{M^2}{v^2}\|x - x'\|_2 + \frac{\beta}{v}\|x - x'\|_2 \\
&= \left(\frac{M^2}{v^2} + \frac{\beta}{v} \right) \|x - x'\|_2,
\end{aligned}$$

where the last inequality follows from f's M-Lipschitz continuity (Claim 22) and β-smoothness. □

D OMITTED PROOFS

D.1 Proof of Lemma 3

Proof. By concavity of U_i, for all $z, z' \in \mathcal{Z}$, we have:

$$U_i(z') - U_i(z) \leq \nabla U_i(z)^T(z' - z). \tag{15}$$

Let z^* be an MNW solution. The first-order optimality condition for z^* requires that the following inequality holds for all $z' \in \mathcal{Z}$:

$$\sum_i \frac{\nabla U_i(z^*)^T}{U_i(z^*)}(z' - z^*) \leq 0 \xRightarrow{\text{by (15)}} \frac{1}{n}\sum_i \frac{U_i(z')}{U_i(z^*)} \leq 1. \tag{16}$$

For contradiction, suppose that z^* is not a core outcome. Then, there exists a set of agents A and an allocation z' such that $(|A|/n)U_i(z') \geq U_i(z^*)$, and at least one inequality is tight. This implies $(1/n)\sum_{i\in A} U_i(z')/U_i(z^*) > 1$, which contradicts (16). □

D.2 Proof of Lemma 5

Proof. Suppose, for contradiction, that z is not a core solution. Then, there must exist a set A and some $z' \in \mathcal{Z}$ such that $(|A|/n)U_i(z') \geq (1+\epsilon)U_i(z) + \delta$ for all $i \in A$, with at least one strict inequality. This implies: $(1/n)\sum_{i\in A} U_i(z')/(U_i(z) + \delta/(1+\epsilon)) > 1+\epsilon$, contradicting (1). □

D.3 Proof of Lemma 13

To prove Lemma 13, we first present the following lemma, which relates $\tilde{w}^{(k)}$ to any $w \in \mathcal{W}_{\mathcal{Z}}$, where $\mathcal{W}_{\mathcal{Z}} \triangleq \{(x,z,\gamma) \in \mathcal{W} \mid x = Gz\}$:

Lemma 24. *Let $\{w^{(k)}\}$ and $\{q^{(k)}\}$ be sequences produced by Algorithm 1. Then, the following inequality holds for any $w \in \mathcal{W}_{\mathcal{Z}}$:*

$$(x-x^{(k)})^T\nabla\theta(x^{(k)}) + \gamma^T(x^{(k)} - Gz^{(k)}) \leq \rho(z^{(k)} - z)^T q^{(k)} \\ + \frac{n\rho}{2}\left(\|z - z^{(k-1)}\|_2^2 - \|z - z^{(k)}\|_2^2\right) + \frac{1}{2\rho}\left(\|\gamma - \gamma^{(k-1)}\|_2^2 - \|\gamma - \gamma^{(k)}\|_2^2\right). \tag{17}$$

Proof. The first-order optimality conditions corresponding to the update step in Line 6 of Algorithm 1 imply the following inequality for all i and $z \in \mathcal{Z}$.

$$(z - x_i^{(k)})^T(\nabla\theta_i(x_i^{(k-1)}) - \gamma_i^{(k-1)} - \rho(x_i^{(k)} - z^{(k-1)})) \leq 0. \tag{18}$$

Let $\tilde{\gamma}^{(k)} \triangleq \gamma^{(k-1)} + \rho(x^{(k)} - Gz^{(k-1)})$. Then, we can rewrite (18) as:

$$(z - x_i^{(k)})^T(\nabla\theta_i(x_i^{(k)}) - \tilde{\gamma}_i^{(k)}) \leq 0.$$

Summing this over all i, for any $z \in \mathcal{Z}$ and $x = Gz$, we have:

$$(x - x^{(k)})^T\nabla\theta(x^{(k)}) - (x - x^{(k)})^T\tilde{\gamma}^{(k)} \leq 0. \tag{19}$$

Next, given (7), Line 8 of Algorithm 1 implies that $z^{(k)}$ is a solution to:

$$\underset{z}{\text{maximize}} \quad \sum_i \left(-(\gamma_i^{(k-1)})^T(x_i^{(k)} - z + q^{(k)}) - \frac{\rho}{2}\|x_i^{(k)} - z + q^{(k)}\|_2^2\right).$$

The first-order optimality conditions for this optimization imply:

$$(z - z^{(k)})^T\left(\sum_i\left(\gamma_i^{(k-1)} + \rho(x_i^{(k)} - z^{(k)} + q^{(k)})\right)\right) \leq 0 \quad \text{for all } z \in \mathbb{R}^m. \tag{20}$$

Given the definition of $\tilde{\gamma}^{(k)}$, we can rewrite (20) for all $z \in \mathbb{R}^m$ as:

$$(z - z^{(k)})^T\left(\sum_i \tilde{\gamma}_i^{(k)} - n\rho(z^{(k)} - z^{(k-1)}) + n\rho q^{(k)}\right) \leq 0 \;\Rightarrow \\ (z - z^{(k)})^T\sum_i \tilde{\gamma}_i^{(k)} \leq n\rho(z - z^{(k)})^T(z^{(k)} - z^{(k-1)}) - n\rho(z - z^{(k)})^T q^{(k)}. \tag{21}$$

Next, given Line 9 of Algorithm 1, for all $\gamma \in \mathbb{R}^{mn}$ we have:

$$x^{(k)} - Gz^{(k)} = (\gamma^{(k)} - \gamma^{(k-1)})/\rho \;\Rightarrow \\ (\gamma - \tilde{\gamma}^{(k)})^T(x^{(k)} - Gz^{(k)}) = (\gamma - \tilde{\gamma}^{(k)})^T(\gamma^{(k)} - \gamma^{(k-1)})/\rho. \tag{22}$$

To put everything together, we use the following identity:

$$(x^{(k)} - Gz)^T \tilde{\gamma}^{(k)} + (z - z^{(k)})^T \sum_i \tilde{\gamma}_i^{(k)} + (\gamma - \tilde{\gamma}^{(k)})^T (x^{(k)} - Gz^{(k)}) = \gamma^T (x^{(k)} - Gz^{(k)}).$$

With this, we can combine (19)–(22) to get the following inequality for any $w = \mathcal{W}_{\mathcal{Z}}$:

$$\begin{aligned}(x - x^{(k)})^T \nabla\theta(x^{(k)}) + \gamma^T (x^{(k)} - Gz^{(k)}) \leq n\rho(z^{(k)} - z)^T q^{(k)} \\ + n\rho(z - z^{(k)})^T (z^{(k)} - z^{(k-1)}) + (\gamma - \tilde{\gamma}^{(k)})^T (\gamma^{(k)} - \gamma^{(k-1)})/\rho.\end{aligned} \tag{23}$$

Next, we focus on the right-hand side of (23). Given the following identity:

$$2(a-b)^T(c-d) = \|a-d\|_2^2 - \|a-c\|_2^2 + \|b-c\|_2^2 - \|b-d\|_2^2,$$

we have:

$$2(z - z^{(k)})^T (z^{(k)} - z^{(k-1)}) = \|z - z^{(k-1)}\|_2^2 - \|z - z^{(k)}\|_2^2 - \|z^{(k)} - z^{(k-1)}\|_2^2, \tag{24}$$

$$\begin{aligned}2(\gamma - \tilde{\gamma}^{(k)})^T (\gamma^{(k)} - \gamma^{(k-1)}) = \|\gamma - \gamma^{(k-1)}\|_2^2 - \|\gamma - \gamma^{(k)}\|_2^2 - \|\tilde{\gamma}^{(k)} - \gamma^{(k-1)}\|_2^2 \\ + \|\tilde{\gamma}^{(k)} - \gamma^{(k)}\|_2^2.\end{aligned} \tag{25}$$

Given the definition of $\tilde{\gamma}^{(k)}$ and Line 9 of Algorithm 1, we have:

$$\begin{aligned}\|\tilde{\gamma}^{(k)} - \gamma^{(k)}\|_2^2 &= \|\rho(x^{(k)} - Gz^{(k-1)}) - (\gamma^{(k)} - \gamma^{(k-1)})\|_2^2 \\ &= \rho^2 \|x^{(k)} - Gz^{(k-1)} - x^{(k)} + Gz^{(k)}\|_2^2 \\ &= n\rho^2 \|z^{(k)} - z^{(k-1)}\|_2^2.\end{aligned} \tag{26}$$

Substituting (24)–(26) into (23) gives (17). □

We are now ready to prove Lemma 13:

Proof. We start by rewriting $(x - x^{(k-1)})^T \nabla\theta(x^{(k-1)})$ as:

$$(x - x^{(k)})^T \nabla\theta(x^{(k)}) = \sum_i \frac{(x_i - x_i^{(k)})^T \nabla U_i(x_i^{(k)})}{U_i(x_i^{(k)}) + v}.$$

Since $U_i(x)$ is concave, for any i and for any $x, x' \in \mathcal{Z}$, we have:

$$U_i(x') - U_i(x) \leq (x' - x)^T \nabla U_i(x). \tag{27}$$

Therefore, (17) implies:

$$\begin{aligned}\sum_i \frac{U_i(x_i) + v}{U_i(x_i^{(k)}) + v} + \gamma^T (x^{(k)} - Gz^{(k)}) \leq n + n\rho(z^{(k)} - z)^T q^{(k)} \\ + \frac{n\rho}{2} \left(\|z - z^{(k-1)}\|_2^2 - \|z - z^{(k)}\|_2^2 \right) + \frac{1}{2\rho} \left(\|\gamma - \gamma^{(k-1)}\|_2^2 - \|\gamma - \gamma^{(k)}\|_2^2 \right).\end{aligned} \tag{28}$$

Next, since (28) holds for any $w \in \mathcal{W}_{\mathcal{Z}}$, it in particular holds when $\gamma = \mathbf{0}_{mn}$, which yields:

$$\begin{aligned}\frac{1}{n} \sum_i \frac{U_i(z) + v}{U_i(x_i^{(k-1)}) + v} \leq 1 + \rho(z^{(k)} - z)^T q^{(k)} \\ + \frac{\rho}{2} \left(\|z - z^{(k-1)}\|_2^2 - \|z - z^{(k)}\|_2^2 \right) + \frac{1}{2n\rho} \left(\|\gamma^{(k-1)}\|_2^2 - \|\gamma^{(k)}\|_2^2 \right).\end{aligned} \tag{29}$$

For any $z \in \mathcal{Z}$, we have $\|z\|_2^2 \leq \|z\|_1^2 \leq 1$. Given this inequality, by summing (29) over $k = 1$ to K and dividing by K, we obtain the following for any $z \in \mathcal{Z}$:

$$\frac{1}{n}\sum_i \frac{1}{K}\sum_{k=1}^{K} \frac{U_i(z)+v}{U_i(x_i^{(k)})+v} \leq 1 + \frac{\rho}{K}\sum_{k=1}^{K}(z^{(k)}-z)^T q^{(k)} + \frac{\rho}{2K}. \tag{30}$$

Since $U_i(x_i) + v$ is strictly positive and concave, the function $1/(U_i(x_i)v)$ is convex (Claim 20). As a result, by Jensen's inequality, it follows that for any $z \in \mathcal{Z}$, we have:

$$\frac{1}{K}\sum_{k=1}^{K} \frac{U_i(z)+v}{U_i(x_i^{(k)})+v} \geq \frac{U_i(z)+v}{U_i\left(\frac{1}{K}\sum_{k=1}^{K} x_i^{(k)}\right)+v} = \frac{U_i(z)+v}{U_i(\bar{x})+v} \geq \frac{U_i(z)}{U_i(\bar{x})+v}.$$

Given the last inequality, (30) implies (11). □

D.4 Proof of Lemma 14

Proof. First, we note that, since the objective function of (4) is continuous on a compact set, the problem attains its bounded global maximum. Therefore, there exist optimal solutions $z^* \in \mathcal{Z}$ and $x^* \in \mathcal{Z}^n$ such that $x_i^* = z^*$ for all i, which achieve this maximum value [49, Theorem 4.16].
Second, *strong duality* holds for (4). This follows from three facts: (i) the objective function is concave, (ii) the constraints are affine, and (iii) *Slater's condition* is satisfied, i.e., there exists a *strictly feasible* point that lies in the relative interior of $\mathcal{Z}$ and satisfies all constraints (e.g., $x_i = z = s/(c\|s\|)$ for all i). As a result, the dual optimal value is bounded and attained, and there exist optimal Lagrange multipliers γ^* that achieves this value [8].
Next, Since z^* is a solution to (4), the first-order optimality conditions require $\sum_i \gamma_i^* = 0$. Therefore, by setting $w = w^*$ in (17), we have:

$$\begin{aligned}(x^*-x^{(k)})^T\nabla\theta(x^{(k)}) + \gamma^{*T}x^{(k)} &\leq n\rho(z^{(k)}-z)^T q^{(k)} \\ &+ \frac{n\rho}{2}\left(\|z^*-z^{(k-1)}\|_2^2 - \|z^*-z^{(k)}\|_2^2\right) + \frac{1}{2\rho}\left(\|\gamma^*-\gamma^{(k-1)}\|_2^2 - \|\gamma^*-\gamma^{(k)}\|_2^2\right).\end{aligned} \tag{31}$$

Since x^* is a solution to (4), the first-order optimality conditions require:

$$(x^{(k)}-x^*)^T(\nabla\theta(x^*)-\gamma^*) \leq 0. \tag{32}$$

Therefore, by summing (31) and (32), we obtain:

$$\begin{aligned}(x^*-x^{(k)})^T(\nabla\theta(x^{(k)}) - \nabla\theta(x^*)) &\leq n\rho(z^{(k)}-z^*)^T q^{(k)} \\ &+ \frac{n\rho}{2}\left(\|z^*-z^{(k-1)}\|_2^2 - \|z^*-z^{(k)}\|_2^2\right) + \frac{1}{2\rho}\left(\|\gamma^*-\gamma^{(k-1)}\|_2^2 - \|\gamma^*-\gamma^{(k)}\|_2^2\right).\end{aligned} \tag{33}$$

Since $\theta(x)$ is concave, we have $(x-x')^T(\nabla\theta(x) - \nabla\theta(x')) \leq 0$. Therefore, (33) implies:

$$\|\gamma^{(k)}-\gamma^*\|_2^2 - \|\gamma^{(k-1)}-\gamma^*\|_2^2 \leq n\rho^2\left(2(z^{(k)}-z^*)^T q^{(k)} + \|z^*-z^{(k-1)}\|_2^2 - \|z^*-z^{(k)}\|_2^2\right).$$

Given that $\|z^*\|_2^2 \leq 1$, by summing this last inequality over $k = 1$ to K, we get:

$$\|\gamma^{(K)}-\gamma^*\|_2^2 \leq 2n\rho^2\sum_{k=1}^{K}(z^{(k)}-z^*)^T q^{(k)} + n\rho^2 + \|\gamma^*\|_2^2. \tag{34}$$

Next, we have:

$$\begin{aligned}\|\rho K(\bar{x} - G\bar{z})\|_2^2 = \|\gamma^{(K)}\|_2^2 &= \|\gamma^{(K)} - \gamma^* + \gamma^*\|_2^2 \\ &\leq 2\|\gamma^*\|_2^2 + 2\|\gamma^{(K)} - \gamma^*\|_2^2 \\ &\leq 4n\rho^2 \sum_{k=1}^{K} (z^{(k)} - z^*)^T q^{(k)} + 2n\rho^2 + 4\|\gamma^*\|_2^2,\end{aligned}$$

which implies:

$$\|\bar{x} - G\bar{z}\|_2 \leq \frac{2\sqrt{n}}{K}\left(\sum_{k=1}^{K} |(z^{(k)} - z^*)^T q^{(k)}|\right)^{\frac{1}{2}} + \frac{\sqrt{2n}}{K} + \frac{2}{\rho K}\|\gamma^*\|_2. \tag{35}$$

The Euclidean projection onto $\mathcal{Z}$ is contractive. Therefore, since $\bar{x}, \hat{z} \in \mathcal{Z}$, we have:

$$\|\bar{x} - G\hat{z}\|_2 = \|\bar{x} - G\Pi_{\mathcal{Z}}(\bar{z})\|_2 \leq \|\bar{x} - G\bar{z}\|_2.$$

Given this inequality, (35) implies (12). □

D.5 Proof of Lemma 15

Proof. For any $z \in \mathcal{Z}$, we have:

$$\begin{aligned}|(z^{(k)} - z)^T q^{(k)}| &= |(\frac{1}{n}\sum_i x_i^{(k)} + q^{(k)} - q^{(k-1)} - z)^T q^{(k)}| \\ &\leq |(\frac{1}{n}\sum_i x_i^{(k)})^T q^{(k)}| + \|q^{(k)}\|_2^2 + |{q^{(k-1)}}^T q^{(k)}| + |z^T q^{(k)}| \\ &\leq \|\frac{1}{n}\sum_i x_i^{(k)}\|_1 \|q^{(k)}\|_2 + \|q^{(k)}\|_2^2 + |{q^{(k-1)}}^T q^{(k)}| + \|z\|_1 \|q^{(k)}\|_2 \\ &\leq 2\|q^{(k)}\|_2 + \|q^{(k)}\|_2^2 + |{q^{(k-1)}}^T q^{(k)}|. \end{aligned} \tag{36}$$

Here, the first inequality follows from the triangle inequality. The second inequality follows from the Cauchy–Schwarz inequality and the fact that $\|\cdot\|_2 \leq \|\cdot\|_1$–that is, for any vectors a and b in an inner product space, $|a^T b| \leq \|a\|_2\|b\|_2 \leq \|a\|_1\|b\|_2$. The third inequality holds because $\frac{1}{n}\sum_i x_i^{(k)} \in \mathcal{Z}$, and for any $z \in \mathcal{Z}$, we have $\|z\|_2 \leq \|z\|_1 \leq 1$. Finally, for the last term in (36), we apply Young's inequality to obtain:

$$|{q^{(k-1)}}^T q^{(k)}| \leq \frac{1}{2}\|q^{(k-1)}\|_2^2 + \frac{1}{2}\|q^{(k)}\|_2^2,$$

Substituting the last inequality into (36) and summing over k, we have:

$$\begin{aligned}|\sum_{k=1}^{K}(z^{(k)} - z)^T q^{(k)}| &\leq \sum_{k=1}^{K} |(z^{(k)} - z)^T q^{(k)}| \\ &\leq 2\sum_{k=1}^{K}\left(\|q^{(k)}\|_2^2 + \|q^{(k)}\|_2\right). \end{aligned} \tag{37}$$

We next focus on tail behavior of $\|q^{(k)}\|_2^2$ and $\|q^{(k)}\|_2$ separately. Starting with $\|q^{(k)}\|_2^2$, note that each $q_j^{(k)} \sim \mathcal{N}(0, \sigma^2)$ is a *sub-Gaussian* random variable[16]. Therefore, by [55, Lemma 2.7.6], each

[16]A real-valued random variable X is called sub-Gaussian if there exists a constant $\sigma > 0$ such that for all $t \in \mathbb{R}$, $\mathbb{E}[\exp(t(X - \mathbb{E}[X]))] \leq \exp((t^2\sigma^2)/2)$.

$(q_j^{(k)})^2$ is *sub-exponential*[17], with $\mathbb{E}[(q_j^{(k)})^2] = \sigma^2$ and

$$\|(q_j^{(k)})^2 - \sigma^2\|_{\psi_1} \leq C_1\sigma^2,$$

where C_1 is a constant, and $\|X\|_{\psi_1} = \inf\{t > 0 \mid \mathbb{E}[\exp(|X|/t)] \leq 2\}$ denotes the *sub-exponential norm* of a real-valued random variable X. Since $q_j^{(k)}$'s are i.i.d. across all k and j, for any $t \geq 0$ and some constant c_1, *Bernstein's inequality* [55, Theorem 2.8.1] implies:

$$\mathbb{P}\left[\sum_{k=1}^{K} \|q^{(k)}\|_2^2 - Km\sigma^2 \geq t\right] \leq \exp\left(-c_1 \min\left(\frac{t^2}{Km\sigma^4}, \frac{t}{\sigma^2}\right)\right). \tag{38}$$

Next, by [55, Theorem 3.1.1 and Lemma 2.6.8], $\|q^{(k)}\|_2$ is a sub-Gaussian random variable with

$$\left\|\|q^{(k)}\|_2 - \mathbb{E}\left[\|q^{(k)}\|_2\right]\right\|_{\psi_2} \leq C_2\sigma^2,$$

where C_2 is a constant, and $\|X\|_{\psi_2} = \inf\{t > 0 \mid \mathbb{E}[\exp(X^2/t^2)] \leq 2\}$ denotes the *sub-Gaussian norm* of a real-valued random variable X. Since $q^{(k)}$'s are independent, by the *general Hoeffding's inequality* [55, Theorem 2.6.2], for any $t \geq 0$ and some constant c_2, we have:

$$\mathbb{P}\left[\sum_{k=1}^{K}\left(\|q^{(k)}\|_2 - \mathbb{E}\left[\|q^{(k)}\|_2\right]\right) \geq t\right] \leq \exp(-\frac{c_2t^2}{K\sigma^4}).$$

We next provide an upper bound on $\mathbb{E}\left[\|q^{(k)}\|_2\right]$. Consider the inequality $\sqrt{u} \leq (1+u)/2$ which holds for any $u \geq 0$. By setting $u = \frac{1}{m\sigma^2}\|q^{(k)}\|_2^2$, we get:

$$\frac{\|q^{(k)}\|_2}{\sqrt{m}\sigma} \leq \frac{1 + (1/m\sigma^2)\|q^{(k)}\|_2^2}{2}.$$

Taking expectations on both sides of the inequality, we obtain:

$$\mathbb{E}\left[\|q^{(k)}\|_2\right] \leq \sqrt{m}\sigma\,\frac{1+1}{2} = \sqrt{m}\sigma.$$

Therefore, we have:

$$\mathbb{P}\left[\sum_{k=1}^{K}\|q^{(k)}\|_2 - K\sqrt{m}\sigma \geq t\right] \leq \mathbb{P}\left[\sum_{k=1}^{K}\left(\|q^{(k)}\|_2 - \mathbb{E}\left[\|q^{(k)}\|_2\right]\right) \geq t\right] \leq \exp(-\frac{c_2t^2}{K\sigma^4}). \tag{39}$$

Given (37)–(39) and the union bound, for $t' = 4t + 2Km\sigma^2 + 2K\sqrt{m}\sigma$, we have:

$$\begin{aligned}
\mathbb{P}\left[|\sum_{k=1}^{K}(z^{(k)} - z)^T q^{(k)}| \geq t'\right] &\leq \mathbb{P}\left[2\sum_{k=1}^{K}\left(\|q^{(k)}\|_2^2 + \|q^{(k)}\|_2\right) \geq t'\right] \\
&\leq \mathbb{P}\left[\sum_{k=1}^{K}\|q^{(k)}\|_2^2 - Km\sigma^2 \geq t\right] + \mathbb{P}\left[\sum_{k=1}^{K}\|q^{(k)}\|_2 - K\sqrt{m}\sigma \geq t\right] \\
&\leq \exp\left(-c_1\min\left(\frac{t^2}{Km\sigma^4}, \frac{t}{\sigma^2}\right)\right) + \exp(-\frac{c_2t^2}{K\sigma^4}).
\end{aligned}$$

[17] A real-valued random variable X is called sub-exponential if there exist constants $\nu, \alpha > 0$ such that for all $|t| < 1/\alpha$, $\mathbb{E}[\exp(t(X - \mathbb{E}[X]))] \leq \exp((t^2\nu^2)/2)$.

For some constant c, setting $t = c\sqrt{Km}\sigma^2 \log^{1/2}(n)$ in the previous inequality implies that the following inequality holds with probability at least $1 - 1/n - 1/n^m$ for any $z \in \mathcal{Z}$.

$$\frac{1}{K}\sum_{k=1}^{K} |(z^{(k)} - z)^T q^{(k)}| \leq 4c\sqrt{m}\sigma^2 \frac{\log^{1/2}(n)}{\sqrt{K}} + 2m\sigma^2 + 2\sqrt{m}\sigma. \tag{40}$$

Since $m\sigma^2 < 1$, (40) implies (13). □

D.6 Proof of Lemma 16

Proof. Let $\{w^{(k)}\}$ and $\{q^{(k)}\}$ be sequences generated by Algorithm 1. Let $w^* = (x^*, z^*, \gamma^*)$ be a solution to (4). Define $\gamma^*_{\max}$ and $\dot{z}$ as:

$$\gamma^*_{\max} = \max_{i,j} |\gamma^*_{i,j}| \quad \text{and} \quad \dot{z} = \underset{z \in \mathcal{Z}}{argmax} \sum_i \frac{U_i(z) + v}{U_i(\bar{x}_i) + v}.$$

Further, define ε_1 and ε_2 as:

$$\varepsilon_1 = \frac{\rho}{K}\sum_{k=1}^{K} |(z^{(k)} - \dot{z})^T q^{(k)}| + \frac{\rho}{2K},$$

and

$$\varepsilon_2 = \frac{2L\sqrt{n}}{K}\left(\sum_{k=1}^{K} |(z^{(k)} - z^*)^T q^{(k)}|\right)^{1/2} + \frac{L\sqrt{2n}}{K} + \frac{2L\sqrt{nm}}{\rho K}\gamma^*_{\max} + v,$$

where $L = (1+\beta)/\sqrt{2}$. By Lemma 13 and the definition of $\dot{z}$, for any $z \in \mathcal{Z}$, we have:

$$\frac{1}{n}\sum_i \frac{U_i(z) + v}{U_i(\bar{x}_i) + v} \leq \frac{1}{n}\sum_i \frac{U_i(\dot{z}) + v}{U_i(\bar{x}_i) + v} \leq 1 + \varepsilon_1. \tag{41}$$

Each U_i is concave, β-smooth, and bounded between 0 and 1. The domain $\mathcal{Z}$ is also bounded (see (10)). Therefore, each U_i is L-Lipschitz (Claim 22). Given this, and the fact that $\|\bar{x}_i - \hat{z}\|_2 \leq \|\bar{x} - G\hat{z}\|_2$ for any i, Lemma 14 implies the following:

$$U_i(\bar{x}) \leq U_i(\hat{z}) + \varepsilon_2 - v \qquad \forall i. \tag{42}$$

Combining (41) and (42), for any $z \in \mathcal{Z}$, we have:

$$\frac{1}{n}\sum_i \frac{U_i(z)}{U_i(\hat{z}) + \varepsilon_2} \leq \frac{1}{n}\sum_i \frac{U_i(z) + v}{U_i(\hat{z}) + \varepsilon_2} \leq 1 + \varepsilon_1. \tag{43}$$

Given (43), Lemma 5 implies that $\hat{z}$ is an $(\varepsilon_1, \varepsilon_2(1+\varepsilon_1))$-core outcome. Assuming $K = \Theta(n)$, and setting $\alpha = 2\log(1/\delta)/\epsilon + 1$, we have:

$$m\sigma^2 = \frac{Km\alpha}{n^2(\epsilon - \log(1/\delta)/(\alpha - 1))} = O\left(\frac{m\log(1/\delta)}{n\epsilon^2}\right).$$

Since $m = o(\sqrt{n})$, if we set $v = \Theta(1/\sqrt{n})$ and choose $\epsilon, \delta > 0$ such that $m\sigma^2 < 1$, then we can apply Lemma 15 to establish the lemma. □

D.7 Proof of Lemma 18

Proof. For any $z, \gamma_i \in \mathbb{R}^m$, let $h(x_i) \triangleq -L_i^{\rho}(x_i, z, \gamma_i)$. To prove the lemma, we rely on [38, Theorem 4]. To apply this theorem, the following four conditions must hold: (1) h is convex; (2) the diameter of $\mathcal{Z}$ is bounded; (3) there exists a point in the interior of $\mathcal{Z}$; and (4) $h(x_i)$ has a Lipschitz continuous gradient over $\mathcal{Z}$. Condition (1) is straightforward to verify. Condition (2) is established in (10). For (3), we note that a point such as $z = s/(c\|s\|)$ lies in the relative interior of $\mathcal{Z}$. Regarding (4), by Claim 23, θ_i has an L-Lipschitz continuous gradient over $\mathcal{Z}$ with $L \leq \frac{M^2}{v^2} + \frac{\beta}{v}$, where $M = \frac{1+\beta}{\sqrt{2}}$. It then follows from the definition of h that it has an $(L+\rho)$-Lipschitz continuous gradient over $\mathcal{Z}$. Since these four conditions are satisfied, by [38, Theorem 4], we can find a point $x_i^* \in \mathcal{Z}$ and a residual $v \in \mathbb{R}^m$ such that:

$$v \in \partial\left(h(x_i^*) + \mathbb{1}_{\mathcal{Z}}(x_i^*)\right), \quad \text{and} \quad \|v\|_2 \leq \frac{\xi}{2}, \tag{44}$$

where $\mathbb{1}_S$ denotes the indicator function[18] of the set S, and ∂f denotes the subdifferential[19] of the convex function f. This guarantee can be achieved using at most

$$O\left(\sqrt{\frac{L+\rho}{\xi}}\log(1/\xi)\right) \tag{45}$$

projections onto $\mathcal{Z}$, a convergence rate that is optimal up to a logarithmic factor.
Next, we observe that since $\mathcal{Z}$ has nonempty interior, the subdifferential of the sum of convex functions equals the sum of their subdifferentials [3, Theorem 3.40]. Therefore, we can rewrite (44) as:

$$v \in \nabla h(x_i^*) + \partial \mathbb{1}_{\mathcal{Z}}(x_i^*), \quad \text{and} \quad \|v\|_2 \leq \frac{\xi}{2}. \tag{46}$$

It is also straightforward to verify that $\partial \mathbb{1}_{\mathcal{Z}}(x_i^*) = N_{\mathcal{Z}}(x_i^*)$, where $N_{\mathcal{Z}}$ denotes the normal cone of $\mathcal{Z}$, defined as: $N_S(x) \triangleq \{y \in \mathbb{R}^m : y^T(u-x) \leq 0, \forall u \in S\}$ [50]. Using this identity, we can further rewrite (46) as:

$$v - \nabla h(x_i^*) \in N_{\mathcal{Z}}(x_i^*), \quad \text{and} \quad \|v\|_2 \leq \frac{\xi}{2}. \tag{47}$$

Finally, since $\sup_{x,x' \in \mathcal{Z}} \|x - x'\|_2 \leq \sqrt{2} \leq 2$, we can apply the Cauchy–Schwarz inequality to conclude that, for any $z \in \mathcal{Z}$, the following holds.

$$-(z - x_i^*)^T \nabla h(x_i^*) \leq -(z - x_i^{(k)})^T v \leq \|v\|_2 \|z - x_i^{(k)}\|_2 \leq \frac{\xi}{2}(2) = \xi. \tag{48}$$

Given that $-\nabla h(x_i^*) = \nabla \theta_i(x^*) - \gamma_i - \rho(x_i^* - z)$, it follows that (48) implies (14).
It is well-known that projecting a vector in $\mathbb{R}^m$ onto $\mathcal{Z}$ can be done in $O(m\log(m))$ time using a sorting-based algorithm (e.g., see [29, Algorithm 5.1]). Consequently, a point x_i^* satisfying (14) can be computed in total time

$$O\left(m\log(m) \cdot \sqrt{\frac{L+\rho}{\xi}}\log(1/\xi)\right).$$

□

[18] $\mathbb{1}_S(x)$ equals 0 if $x \in S$, and $+\infty$ otherwise.
[19] $\partial f(x) \triangleq \{u \in \mathbb{R}^m : f(x') \geq f(x) + u^T(x' - x), \forall x' \in \mathrm{dom}(f)\}$