Analysis and Optimization of Multi-dimensional Percentile Mechanisms

Xin Sui and Craig Boutilier and Tuomas Sandholm

Abstract

We consider the mechanism design problem for agents with single-peaked preferences over multi-dimensional domains when multiple alternatives can be chosen. Facility location and committee selection are classic embodiments of this problem. We propose a class of *percentile mechanisms*, a form of generalized median mechanisms, that are (group) strategy-proof, and derive worst-case approximation ratios for social cost and maximum load for L_1 and L_2 cost models. More importantly, we propose a sample-based framework for optimizing the choice of percentiles relative to any prior distribution over preferences, while maintaining strategy-proofness. Our empirical investigations, using social cost and maximum load as objectives, demonstrate the viability of this approach and the value of such optimized mechanisms *vis-à-vis* mechanisms derived through worst-case analysis.

1 Introduction

Mechanism design deals with design of protocols to elicit the preferences of self-interested agents so as to achieve a certain social objective. An important property of mechanisms is *strategy-proofness*, which requires that agents have no incentive to misreport their preferences to the mechanism. While payments are often used to ensure that mechanisms are strategy-proof [23, 6, 11], in many settings payments are infeasible and restrictions on preferences are required. The simple but elegant class of *single-peaked preferences* is one such example: roughly speaking, each agent has a single, most-preferred point in the alternative space and alternatives become less preferred as they are moved away from that point. In such settings, choosing a single alternative can be accomplished in a strategy-proof fashion using the famous *median mechanism* [4] and its generalizations [18, 1]. Such models are used frequently for modeling political choice, facility location, and other problems. They also have potential applications in areas such as in the design of a family of products, customer segmentation, and related tasks, as we discuss below.

Unfortunately, such mechanisms are efficient (e.g., w.r.t. social cost) only in very limited circumstances. Furthermore, allowing the choice of multiple alternatives (e.g., multiple facilities) generally causes even these limited guarantees to evaporate. In response, authors have begun to address the question of *approximate mechanism design without money* [19], which focuses on the design of strategy-proof mechanisms for problems such as multi-facility location that are approximately efficient (i.e., have good approximation ratios) [19, 15, 10]. This work provides some positive results, but is generally restricted to settings involving two facilities (alternatives) and L_2 (Euclidean) preferences.

In this paper, we propose *percentile mechanisms*—a special case of generalized median mechanisms [2, 1], but in a more general fashion. Specifically: (a) we consider selection of multiple alternatives (e.g., multi-facility location) in a multi-dimensional alternative space; (b) we address both social cost and maximum load as performance metrics; and (c) we analyze our mechanisms relative to L_1 (Manhattan) and L_2 (Euclidean) preferences. Our first contribution is the analysis of the approximation ratios of various percentile mechanisms under various assumptions. The performance guarantees of such mechanisms under worst-case assumptions are quite discouraging (much like previous results above).

Indeed, designing mechanisms that have the best possible worst-case guarantees may lead to poor performance in practice. Our second contribution is the development of a sample-based *empirical framework for the optimization of percentile mechanisms* relative to a known preference distribution. In most realistic applications of mechanism design, such as facility location, product design, and many others, the designer will have *some* knowledge of the preferences of participating agents. Assuming this takes the form of a distribution over preference profiles, we use profiles sampled from this distribution to optimize the choice of percentiles. Since the result is a percentile mechanism, strategy-proofness is maintained. Our empirical results demonstrate that, by exploiting probabilistic domain knowledge, we obtain strategy-proof mechanisms that outperform mechanisms designed to guard against worst-case profiles. Our framework can be viewed, conceptually, as a form of *automated mechanism design (AMD)*, which advocates the use of preference (or type) distributions to optimize mechanisms [7, 20].

2 Preliminaries

In this section, we introduce our model along with required concepts, notation, and motivation, and then briefly discuss a selection of related work.

2.1 The Social Choice Problem

In a standard social choice setting, we must select an *outcome* o from an outcome set O, where each of agents $i \in N = \{1, 2, \ldots, n\}$ has a preference over O. Agent preferences are represented by (weak) total order over O, or in a more precise way by a *utility function.* In our setting, we focus on the m-dimensional, *q-facility location problem* (or (m, q)-problems): we must choose q points or *locations* in an m-dimensional space $\mathbb{R}^m$ (or some bounded subspace thereof) to place facilities. Outcomes are then *location vectors* of the form $\mathbf{x} = (x_1, \ldots, x_q)$, with $x_j \in \mathbb{R}^m$ (for $j \leq q$). Each agent i has a type t_i denoting the *cost* associated with any location $x \in \mathbb{R}^m$: we write $c_i(x, t_i)$ to denote this real-valued cost. Given an outcome $\mathbf{x}$, i will use the location that has least cost, hence $c_i(\mathbf{x}, t_i) = \min_{j \leq q} c_i(x_j, t_i)$.

Facility location can be interpreted literally, and naturally models the placement of q facilities (e.g., warehouses in a supply chain, public facilities such as parks, etc.) in some geographic space. Agents will then use the least cost (or "closest") facility. However, many other choice problems fit within this class. Voting is one example [4, 1]: we can think of political candidates as being ordered along several dimensions (e.g., stance on the environment, health care, fiscal policy)—voters have preferences over points in this space—and one must elect q representatives to a committee or legislative body. In product design, a vendor may launch a family q new, related products, each described by an m-dimensional feature vector, with consumer preferences over these options leading them to select their most preferred. This also can serve as a form of customer segmentation.

In facility location problems and the other settings discussed above, it is natural to assume agent preferences are *single-peaked.* Intuitively, this means the agent has a single "ideal" location, and its cost for any chosen location increases as it "moves away from" this ideal. Formally, we don't need a distance metric, only a strict ordering on alternatives in each dimension, which is used to define a *betweenness relation.* Let $||\cdot||_1$ denote the L_1-norm.

Definition 1 *[2] An agent i's preference on m-dimensional space $\mathbb{R}^m$ is single-peaked if there exists a most preferred alternative $\tau(t_i)$ such that, $\forall \alpha, \beta \in \mathbb{R}^m$ satisfying $||\tau(t_i) - \beta||_1 = ||\tau(t_i) - \alpha||_1 + ||\alpha - \beta||_1$, we have $c_i(\alpha, t_i) \leq c_i(\beta, t_i)$.*

Single-peaked preferences require that if a point α lies within the "bounding box" of $\tau(t_i)$ and β, then α is at least as preferred as β. Intuitively, as we move farther away from i's ideal location $\tau(t_i)$ we can reach α via some path before we reach β. Note that this requirement does not restrict i's relative preference for α and β if neither lies within the other's bounding box (w.r.t. $\tau(t_i)$).

An agent's ideal location $\tau(t_i)$ does not fully determine its preference, even if it is single-peaked. Despite this, we will equate an agent's type t_i with its ideal location (for reasons that become clear below). However, within the class of single-peaked preferences, we can adopt specific cost functions that *are* fully determined by the ideal location t_i. Often *distance metrics* are used, and we consider both L_1 (Manhattan) and L_2 (Euclidean) distances below. Specifically, we define distance-based cost functions for i as follows:

$$c_i^p(\mathbf{x}, t_i) = \min_{j \leq q} ||t_i - x_j||_p \tag{1}$$

where $p \in \{1, 2\}$ reflects either L_1 or L_2 distance from i's nearest facility. We use $x^p[i; \mathbf{x}]$ to denote i's closest facility in the location vector $\mathbf{x}$ under the L_p-norm.

The aim in facility location is to select a set of q facilities that minimize some social objective. One natural objective is to minimize *social cost (SC)* given type profile $\mathbf{t}$, where social cost (relative to some norm p) is given by:

$$SC_p(\mathbf{x}, \mathbf{t}) = \sum_i c_i^p(\mathbf{x}, t_i) \tag{2}$$

Alternatively, we could try to balance the *load* by ensuring no facility is used by too many agents. Define the load on facility j given outcome $\mathbf{x}$ and type profile $\mathbf{t}$ as $l_j^p(\mathbf{x}, \mathbf{t}) = |\{i | x^p[i; \mathbf{x}] = j\}|$. We wish to minimize the *maximum load (ML)*, which is defined as:

$$ML_p(\mathbf{x}, \mathbf{t}) = \max_j l_j^p(\mathbf{x}, \mathbf{t}). \tag{3}$$

This objective makes sense, for instance, when a product designer launches a family of q new products, consumers purchase the product closest to their ideal product, but costs are minimized by balancing production; or when facility management costs increase superlinearly with load. Many other fundamental social objectives, such as fairness (e.g., maximum agent distance), and combinations thereof can be adopted depending on one's design goals.

2.2 Mechanisms

The goal of mechanism design is to construct mechanisms that (possibly indirectly) elicit information about agent preferences so that an outcome choice can be made that achieves some social objective. We consider *direct mechanisms* in which agents are asked to reveal their types, and an outcome is chosen based on the revealed types. In the facility location with single-peaked preferences, we consider mechanisms that ask agents to declare their ideal locations, then select an outcome $\mathbf{x}$: that is, a *mechanism M* is a function f that maps a declared type profile $\mathbf{t}$ into an outcome $f(\mathbf{t}) \in (\mathbb{R}^m)^q$ (i.e., q m-dimensional alternatives).

A mechanism f is *strategy-proof* (or truthful) if:[1]

$$c_i(f(t_i, \mathbf{t}_{-i}), t_i) \leq c_i(f(t'_i, \mathbf{t}_{-i}), t_i), \quad \forall i, t_i, t'_i, \mathbf{t}_{-i}$$

In other words, f is strategy-proof if no agent can obtain a better outcome by misreporting its true type (ideal location). *Group strategy-proofness* is defined similarly, but requires that

[1]We use *strategy-proof* to refer to dominant strategy incentive compatibility (participation is assured in our settings).

no group of agents $S \subseteq N$ can misreport their types, in a coordinated fashion, so that the outcome is better for at least one $i \in S$, and no worse for any $i \in S$.

While the ideal is to design strategy-proof mechanisms that achieve some social objective, such as minimizing social cost, this is not always feasible. In $(1,1)$-facility location problems, if agent preferences are single-peaked, the *median mechanism*, which selects the median of all reported ideal locations, is (group) strategy-proof [4, 18] and minimizes social cost if agent preferences are all determined under a suitable distance metric (such as L_1). However, when one moves to even just two facilities, strategy-proofness and efficiency are incompatible, as demonstrated by Procaccia and Tennenholtz [19]. They propose the study of *approximate mechanisms* to handle such situations: mechanisms that are strategy-proof and come as close as possible to achieving the social objective (e.g., minimizing social cost). Formally:

Definition 2 *A mechanism f has an* approximation ratio *ε w.r.t. social objective C if:*

$$C(f(\mathbf{t}), \mathbf{t}) \leq \varepsilon \cdot \min_{\mathbf{x}} C(\mathbf{x}, \mathbf{t}).$$

We refer to such a mechanism as *ε-optimal* w.r.t. objective C (or *ε-efficient* when considering social cost/welfare). When minimizing social cost, we assume the number of agents is greater than the number of facilities (otherwise, we can trivially locate facilities at each agent's ideal to obtain a (group) strategy-proof, efficient mechanism). Notice that our mechanisms are *non-imposing*: once facilities are selected, agents are free to choose their favourite (otherwise, one can trivially minimize ML by assigning agents to facilities in an arbitrary balanced way).

2.3 Related Work

Black [4] first proposed the median mechanism for $(1,1)$-facility location, showing it to be strategy-proof for single-peaked preferences. Moulin [18] proposed a generalized median scheme (allowing *phantom peaks*) that he proved to be the unique class of (anonymous) strategy-proof mechanisms for such preferences. Barberà *et al.* [2] later generalized this class of mechanisms further using *coalitional systems* and provided a characterization result for $(m,1)$-problems. We refer to this class as *m-dimensional generalized median* schemes. These schemes select a location by choosing its coordinates in each dimension independently (in a "median-like" fashion).

Some work considers strategy-proof mechanisms with even more restricted preferences and domain assumptions. Border and Jordan [5] characterize strategy-proof mechanisms in m-dimensional spaces assuming *separable star-shaped* preferences (which include quadratic preferences). As in [2], location coordinates are chosen in each dimension separately. Massó and Moreno de Barreda [17] consider symmetric, single-peaked preferences (of which L_1 and L_2 are instances), and show that a mechanism is strategy-proof iff it is a *disturbed* generalized median voter schemes (which allows discontinuities). Schummer and Vohra [21] consider the problem of choosing a location on a graph (e.g., a network) relative to an extended notion of single-peakedness, obtaining positive results for trees, and negative results for cyclic graphs.

Recent attention has been focused on algorithmic aspects and approximation in strategy-proof facility location when agents have L_2 preferences. Procaccia and Tennenholtz [19] study the one-dimensional problems, and provide upper and lower bounds on the approximation ratio for social cost. Of interest here is their deterministic *left-right mechanism*, which is $(n-1)$-efficient for $(1,2)$-problems. Lu *et al.* [15] define the (randomized) *proportional mechanism* with an approximation ratio of 4 for general distance metrics, but it cannot be applied for more than two facilities. Fotakis and Tzamos [10] show that a *winner-imposing* variant of the proportional mechanism is strategy-proof for any number of facilities, with an approximation ratio of $4q$. Escoffier *et al.* [8] define the first mechanism

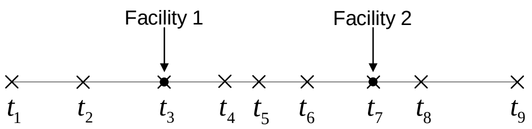


Figure 1: The $(0.25, 0.75)$-percentile mechanism for $n = 9$.

for general multi-dimensional location problems, a randomized mechanism with an approximation ratio of $n/2$, but only in the very restrictive setting where the number of agents is exactly one more than the number of facilities.

Work on load balancing games is somewhat related, but differs in that cost functions reflect the externalities agents impose on one another (by sharing a facility or some other resource). Considerable research has developed price of anarchy [12, 3] and related results. However, externalities give those models a very different character than ours.

3 Percentile Mechanisms

In this section, we introduce and analyze the class of *percentile mechanisms*, a special case of *m-dimensional generalized median mechanisms* [2, 1].

3.1 One-dimensional Percentile Mechanisms

We begin with one-dimensional facility location problems to develop intuitions. We wish to place q facilities, with each agent i having a single ideal location t_i and single-peaked preferences. Without loss of generality, we rename the agents so their ideal locations are ordered: $t_1 \leq t_2 \leq \ldots \leq t_n$. A *percentile mechanism* is specified by a vector $\mathbf{p} = (p_1, p_2, \ldots, p_q)$, where $0 \leq p_1 \leq p_2 \ldots \leq p_q \leq 1$: the $\mathbf{p}$-percentile mechanism locates the jth facility at the p_jth percentile of the reported ideal locations. In other words, the jth location is placed at $x_j = t_{i_j}$, where $i_j = \lfloor (n-1) \cdot p_j \rfloor + 1$.[2] Intuitively, we can decompose the mechanism into q independent rules, each locating one facility.

Example 1 *We illustrate the* $(0.25, 0.75)$*-percentile mechanism for a two-facility problem with* $n = 9$ *agents in Fig. 1. Ordering reported locations so that* $t_1 \leq \ldots \leq t_9$*, the mechanism locates the first facility at* $x_1 = t_3$ *(since* $\lfloor 8 \cdot 0.25 \rfloor + 1 = 3$*) and the second at* $x_2 = t_7$*.*

The following theorem shows an important property of the mechanism:

Theorem 1 *The* $\mathbf{p}$*-percentile mechanism is (group) strategyproof for any* $\mathbf{p}$*.*

Proof: We prove the theorem for the case of $q = 2$ (proofs for other cases are similar).

Let $S \subseteq N$ be a coalition of agents, $x = (x_1, x_2)$ be the location vector if agents truthfully report their ideals, and $x' = (x'_1, x'_2)$ be the location vector if agents in S jointly deviate from their peaks. In addition, let $\Delta_1 = x_1 - x'_1$ and $\Delta_2 = x'_2 - x_2$. An important observation is that, according to our mechanism, if either of Δ_1 or Δ_2 is greater or less than 0, some agent in S must be strictly worse off. We consider four cases:

I. $\Delta_1 \geq 0$ and $\Delta_2 \geq 0$. Note that we can ignore the case where both Δ_1 and Δ_2 are 0, since no agent in S gains by misreporting if neither facility moves. Assume, w.l.o.g., that $\Delta_1 > 0$ and $\Delta_2 \geq 0$. Recall that x_1 is the p_1th percentile among all reported

[2] We could equivalently use order statistics; but the percentile formulation removes dependence on the number of the agents in the mechanism's specification. It is well-known that, for any fixed n, Moulin's phantom peaks can easily be arranged to implement any order statistic.

peaks. Hence $\Delta_1 > 0$ implies that some agent $i \in S$, with $t_i \geq x_1$, reports a new ideal to the left of x_1. Agent i's cost is now:

$$c_i(x', t_i) = \min\{t_i - x'_1, x'_2 - t_i\} \geq \min\{t_i - x_1, x_2 - t_i\} = c_i(x, t_i)$$

II. $\Delta_1 \geq 0$ and $\Delta_2 < 0$. In this case, there must be an $i \in S$, with $t_i \geq x_2$, that reports a new ideal to the left of x_2; it's cost is:

$$c_i(x', t_i) = t_i - x'_2 \geq t_i - x_2 = c_i(x, t_i)$$

III. $\Delta_1 < 0$ and $\Delta_2 \geq 0$. This case is completely symmetric to Case II.

IV. $\Delta_1 < 0$ and $\Delta_2 < 0$. The case is similar to Case II: There must be an $i \in S$ whose ideal is to the right of x_2 but misreports to the left of x_2, increasing its cost.

We conclude that our percentile mechanism is (group) strategy-proof. ■

Since any percentile mechanism is strategy-proof for any class of single-peaked preferences, it prevents strategic manipulation even when applied to specific cost/preference functions. Unfortunately, percentile mechanisms can give rise to poor approximation ratios when we consider specific cost functions, specifically, L_2 or L_1 costs.[3]

Theorem 2 *Let agents have L_2 (equivalently, L_1) preferences. Let $\mathbf{p} = (p_1, p_2, \ldots, p_q)$ define a percentile mechanism M. If $q \geq 3$, the approximation ratio of M w.r.t. social cost is unbounded. The approximation ratio w.r.t maximum load is $q \cdot z$, where $z = \max_{1 \leq j \leq q}(p_{j+1} - p_{j-1})$ (defining $p_0 = 0$ and $p_{q+1} = 1$).*

The proof is provided in a longer version of the paper, but we sketch the intuitions here for the case of social cost.[4] The key point is that for any percentile vector, we can construct an ideal location profile in which the number of different peaks is exactly one more than the number of facilities, and two of the peaks are arbitrarily close. The percentile mechanism can locate one facility at each of the "close peaks," while the optimal solution will select only one of them. Since optimal social cost is arbitrarily small, an unbounded approximation ratio results.

Notice that the theorem does not hold for social cost with $q = 2$ facilities: the *left-right mechanism*, which in our terminology is the $(0, 1)$-percentile mechanism, has a bounded approximation ratio of $n - 1$ for social cost [19]. Indeed, it is not hard to show the $(0, 1)$-percentile mechanism is the *only* mechanism within the percentile family that has a bounded approximation ratio. We conjecture there is no other deterministic mechanism (even outside the percentile family) that has a bounded approximation ratio. This gives further motivation to the use of probabilistic priors to optimize the choice of percentiles (see Sec. 4).

With respect to maximum load, it is natural to ask which percentile vector $\mathbf{p}$ minimizes z in Thm. 2. We can show that the percentile mechanism that "evenly distributes" facilities is approximately optimal, and that it has the smallest approximation ratio within the family.

Proposition 1 *Let agents have L_2 (equiv. L_1) preferences. If q is odd, then the percentile mechanism with $p_j = \frac{j}{q+1}$, $\forall 1 \leq j \leq q$, is $\frac{2q}{q+1}$-optimal w.r.t. maximum load. If q is even, then the percentile mechanism with $p_j = p_{j+1} = \frac{j+1}{q+2}$, $\forall j = 2j' - 1, 1 \leq j' \leq q/2$, is $\frac{2q}{q+2}$-optimal w.r.t. maximum load.[5] In each case, the mechanism has the smallest approximation ratio within the percentile family.*

---

[3] Of course, other mechanisms, beyond simple generalized medians, depending on the preference class (e.g., disturbed median mechanisms for symmetric costs [17] like L_1 and L_2).

[4] Any omitted proofs of our main results can be found in the appendix of a longer version of this paper; see: `http://www.cs.toronto.edu/~cebly/papers.html`.

[5] For even q, the mechanism is partially imposing. We locate two facilities at each selected location, and

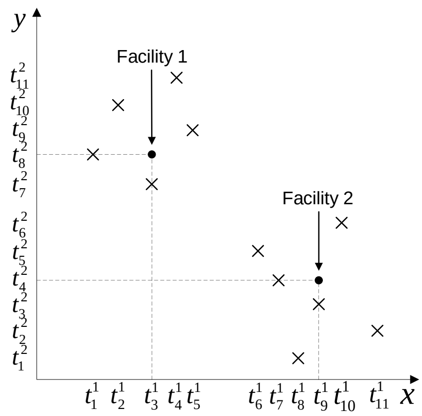


Figure 2: A percentile mechanism for a (2, 2)-problem when $n = 11$.

3.2 Multi-dimensional Percentile Mechanisms

As discussed above, many social choice problems can be interpreted as "facility location" problems when viewed as choice in a higher dimensional space, such as selection of political/committee representatives, product design, and the like. We now analyze a generalization of the percentile mechanism to multi-dimensional spaces.

As above, we assume that agents have single-peaked preferences (see Defn. 1). Reported types t_i are now points in $\mathbb{R}^m$. For any type profile $\mathbf{t}$, let $t^k_1 \leq t^k_2 \leq \ldots \leq t^k_n$ be the *ordered projection* of $\mathbf{t}$ in the kth dimension (for $k \leq m$). In other words, we simply order the reported coordinates in each dimension independently. An *m-dimensional percentile mechanism* is specified by a $q \times m$ matrix $\mathbf{P} = (\mathbf{p}_1; \mathbf{p}_2; \ldots; \mathbf{p}_q)$, where each $\mathbf{p}_j \in [0, 1]^m$ is an m-vector in the m-dimensional unit cube, with $\mathbf{p}_j = (p^1_j, p^2_j, \ldots, p^m_j)$. Given a reported profile $\mathbf{t}$, the $\mathbf{P}$-percentile mechanism locates the jth facility by selecting, for each dimension $k \leq m$, the p^k_jth percentile of the ordered projection of $\mathbf{t}$ in the kth dimension as the coordinate of facility j in that dimension. In other words:

$$x_j = (t^1_{\lfloor (n-1)\cdot p^1_j \rfloor + 1}, t^2_{\lfloor (n-1)\cdot p^2_j \rfloor + 1}, \ldots, t^m_{\lfloor (n-1)\cdot p^m_j \rfloor + 1}).$$

Example 2 *Fig. 2 illustrates a 2-D, two facility problem with 11 agents. With* $\mathbf{P} = (0.2, 0.7; 0.8, 0.3)$*, the* $\mathbf{P}$*-percentile mechanism locates the first facility at the x-coordinate of* t_3 *(since* $\lfloor 10 \cdot 0.2 \rfloor + 1 = 3$*) and at the y-coordinate of* t_8*; and the second facility is placed at the x-coordinate of* t_9 *and and the y-coordinate of* t_4*. Notice facilities need not be located at the ideal point of any agent.*

The following results generalize the corresponding one-dimensional results above.

Theorem 3 *The m-dimensional* $\mathbf{P}$*-percentile mechanism is (group) strategy-proof for any* $\mathbf{P}$*.*

Theorem 4 *Let agents have* L_1 *or* L_2 *preferences, and* $\mathbf{P}$ *define a percentile mechanism* M *for an* (m, q)*-facility location problem with* $m > 1$*. The approximation ratio of* M *is unbounded w.r.t. social cost for any* $\mathbf{P}$*. The approximation ratio of* M *is* $q \cdot z$ *w.r.t. maximum load, where* $z = \prod_{k=1}^m \max_{1 \leq j \leq q}(p^k_{j+1} - p^k_{j-1})$ *(where we define* $p^k_0 = 1$ *and* $p^k_{q+1} = 1$*).*

balance the agents choosing that location. Agents are indifferent to the "imposed" assignment, so this is unlike truly imposing mechanisms that remove choice from agents' hands [10]. We use this mechanism for convenience—one can define a strictly non-imposing mechanism with the same approximation ratio.

Notice that this result differs from the one-dimensional case, where the $(0,1)$-percentile (i.e., left-right) mechanism has a bounded approximation ratio for social cost. When $m > 1$, *no* percentile mechanism has this property—this holds because the mechanism may place no facility at the ideal location of any agent. As above, however, we can optimize the percentiles for maximum load, when $q = \tilde{q}^m$ for some $\tilde{q}$ by exploiting Prop. 1 in each dimension:

Proposition 2 *Let $q = \tilde{q}^m$. If $\tilde{q}$ is odd, the mechanism that locates one facility at each percentile of the form $\frac{1}{\tilde{q}+1}$ in each dimension is $\left(\frac{2\tilde{q}}{\tilde{q}+1}\right)^m$-optimal w.r.t. maximum load. If $\tilde{q}$ is even, the mechanism that locates two facilities at each percentile of the form $\left(\frac{2}{\tilde{q}+2}\right)$ in each dimension is $\left(\frac{2\tilde{q}}{\tilde{q}+2}\right)^m$-optimal w.r.t. maximum load. Moreover, these are the smallest approximation ratios possible within the family of percentile mechanisms.*

4 Optimizing Percentile Mechanisms

We've seen that percentile mechanisms are (group) strategy-proof for general (m, q)-facility location problems, and can offer bounded approximation ratios for L_1 and L_2 preferences (though only under restricted circumstances for social cost). Unfortunately, these guarantees require optimizing the choice of percentiles w.r.t. worst-case profiles, which can sometimes lead to poor performance in practice. For example, in a $(1,2)$-problem, decent approximation guarantees for social cost require using the $(0,1)$-percentile mechanism; but if agent preferences are uniformly distributed in one dimension, this will, in fact, perform quite poorly. Intuitively, the $(0.25, 0.75)$-percentile mechanism should have lower expected social cost by the (probabilistically) "suitable" placement of two facilities, each for use by half of the agents.

We consider a framework for empirical optimization of percentiles within the family of percentile mechanisms that should admit much better performance in practice. As in automated mechanism design [7, 20], we assume a prior distribution D over agent preference profiles. Hence agent preferences can be correlated in our model. One will often assume a prior model D (e.g., learned from observation) that renders individual agent preferences independent *given* that model, but this is not required. In many practical settings, such as facility location or product design, such distributional information will in fact be readily available. We sample preference profiles from this distribution, and use them to optimize the percentiles in the $\mathbf{P}$ matrix to ensure the best possible expected performance w.r.t. our social objective.

Unlike classic AMD, we restrict ourselves to the specific family of percentile mechanisms. While this limits the space of mechanisms, we do this for several reasons. First, it provides a much more compact mechanism parameterization over which to optimize than in typical AMD settings.[6] Second, since the resulting mechanism is (group) strategy-proof no matter which percentiles are chosen, the optimization need not account for incentive constraints (unlike standard AMD). Of course, when considering specific classes of single-peaked preferences, such as L_1 or L_2 costs as we do here, a wider class of strategy-proof mechanisms could be used (e.g., disturbed median mechanisms [17]); but these have more parameters, and as we will see below empirically, they are unlikely to offer any better performance—since our optimized percentile mechanisms achieve near-optimal social cost. In addition, errors due to sampling, or even misestimation of the prior D, have no impact on the strategyproofness of the mechanism. Third, unlike Bayesian optimization—in other words, methods that choose optimal facility placement relative to the prior with *no elicitation* of ideal locations—optimized percentile mechanisms are *responsive* to the specific preferences of the agents.

[6] AMD has been explored in a parameterized mechanism space, e.g., in combinatorial auctions [13, 14].

Distribution		$q = 2$	$q = 3$	$q = 4$
D_u	SC	$(0.25, 0.75)$	$(0.16, 0.5, 0.84)$	$(0.12, 0.37, 0.63, 0.88)$
	ML	$(0.49, 0.50)$	$(0.33, 0.35, 0.98)$	$(0.25, 0.26, 0.74, 0.75)$
D_g	SC	$(0.25, 0.75)$	$(0.15, 0.5, 0.85)$	$(0.1, 0.35, 0.65, 0.9)$
	ML	$(0.49, 0.50)$	$(0.33, 0.35, 0.9)$	$(0.25, 0.26, 0.74, 0.75)$
D_{gm}	SC	$(0.17, 0.68)$	$(0.16, 0.59, 0.93)$	$(0.12, 0.37, 0.68, 0.94)$
	ML	$(0.49, 0.50)$	$(0.14, 0.65, 0.66)$	$(0.17, 0.34, 0.73, 0.74)$

Table 1: Optimal percentiles for different distributions, objectives, and numbers of facilities.

Let agent type profiles $\mathbf{t} = (t_1, t_2, \ldots, t_n)$ be drawn from distribution D. Given a $\mathbf{P}$-percentile mechanism, let $f_{\mathbf{P}}(\mathbf{t})$ denote the chosen locations when the agent type profile is $\mathbf{t}$. The goal is to select $\mathbf{P}$ to minimize the expected social cost or maximum load:

$$\min_{\mathbf{P}} \mathbb{E}_D \left[SC_p(f_{\mathbf{P}}(\mathbf{t}), \mathbf{t}) \right]; \quad \text{or} \quad \min_{\mathbf{P}} \mathbb{E}_D \left[ML_p(f_{\mathbf{P}}(\mathbf{t}), \mathbf{t}) \right]$$

Naturally, other objectives can be modelled in this way too.

Given Y sampled preference profiles, we optimize percentile selection relative to the Y sampled profiles. In our experiments below, we use simple numerical optimization for this purpose. Specifically, we consider all possible values for the percentile matrix $\mathbf{P}$. For each of them, we compute the average social cost (maximum load) over Y sample profiles, and select the one that has the minimum objective value. Alternatively, one can formulate the minimization problem as a mixed integer programming (MIP) for L_1 costs, or a mixed integer quadratically constained program (MIQCP) for L_2 costs, and use standard optimization tools, e.g., CPLEX, to solve the problem. However, determining concise formulations is non-trivial and effective use of these formulations is left to future research.[7]

In the following experiments, we consider problems with $n = 101$ agents, with agent preferences drawn independently from three classes of distributions: uniform D_u, Gaussian D_g and mixture of Gaussians D_{gm} with 2 or 3 components. Each distribution reflects rather different assumptions about agent preferences: that they are spread evenly (D_u); that they are biased toward one specific location (D_g); or that they partitioned into 2 or 3 loose clusters (D_{gm}). In all cases, $T = 500$ sampled profiles are used for optimization. We examine results for both social cost and maximum load.

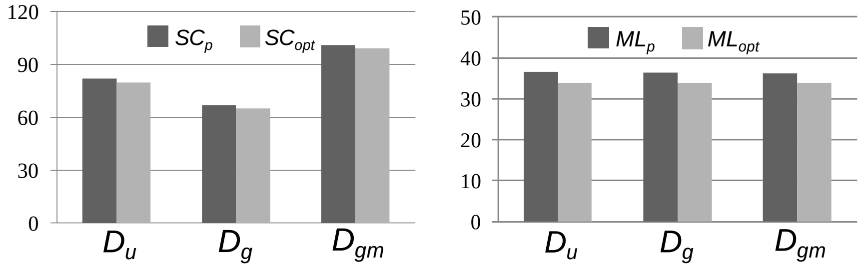
One-dimensional mechanisms

We begin with simple one-dimensional problems with $q = 2$, 3 or 4. Table 1 shows the percentiles resulting from our optimization for both SC and ML under each of the three distributions.[8] For example, when agent ideal locations are uniformly distributed, the $(0.25, 0.75)$-percentile mechanism minimizes the expected social cost for two facilities. This is expected, since the uniform (and Gaussian) distribution partitions agents into two groups of roughly equal size, and facilities should be located at the median positions of each group.

The performance of the optimized percentile mechanisms is extremely good. Fig. 3 compares the expected social cost and maximum load of our mechanisms with those given by *optimal placement* of facilities (results for $q = 3$ are shown, but others are similar). Recognize however that optimal placement is not realizable with any strategy-proof mechanism. Despite this, optimized percentile mechanisms perform nearly as well in expectation in all three cases. Contrast this with the performance of the mechanisms with provable approximation ratios. When $q = 2$, the $(0, 1)$-percentile mechanism has an average social cost

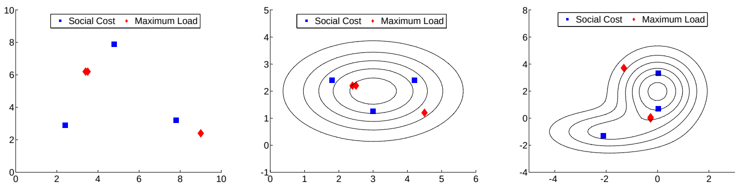
[7] We describe preliminary formulations of the MIP and MIQCP, which do not scale well, in the appendix of a longer version of this paper; see: `http://www.cs.toronto.edu/~cebly/papers.html`.

[8] D_u is uniform on $[0, 10]$. D_g is Gaussian $\mathcal{N}(0, 2)$ with $\mu = 0, \sigma^2 = 2$. D_{gm} is a Gaussian mixture with 3 components: $\mathcal{N}(-4, 4)$ (weight 0.4), $\mathcal{N}(0, 1)$ (weight 0.45), and $\mathcal{N}(5, 2)$ (weight 0.15).



(a) Social Cost (b) Maximum Load

Figure 3: Comparison of optimized percentile mechanism and optimal value ($q = 3$).



Distribution		Percentile/OPT
D_u	SC	871.2/832.4
	ML	51/51
D_g	SC	424.8/417.38
	ML	51/51
D_{gm}	SC	696.5/681.46
	ML	51/51

Figure 4: Optimized Percentiles for (a) **2D**: Uniform, (b) **2D**: Gaussian, (c) **2D**: Gaussian mixture, and (d) **4D**.

of 242.4, 340.9 and 523.2 for D_u, D_g and D_{gm}, respectively; but the social cost of our mechanisms are only 123.7, 76.5, and 165.1, respectively. When $q = 3$, the $(0.25, 0.5, 0.75)$-percentile mechanism has the best approximation ratio for ML (see Prop. 1). Its average maximum loads are 39.5, 38.7 and 38.3, which are close to (but not as good as) the loads of the optimized percentile mechanisms (36.5, 36.5, and 36.2).

Multi-dimensional mechanisms

We also experimented with two additional problems. **2D** is a $(2, 3)$-problem where agents have L_2 preferences, capturing, say, the placement of three public projects like libraries, or warehouses. **4D** is a $(4, 2)$-problem with L_1 preferences, which might model the selection of 2 products for launch, each with four attributes that predict consumer demand.[9]

For the problem **2D** we show the *expected placement* of facilities given the selected percentiles in Fig. 4(a)-(c), for both SC and ML, for each of the three distributions. (*Actual* facility placement will shift to match the reported type profile in each instance.) Placement for SC tends to be distributed appropriately, while ML places two facilities adjacent to one another. For **4D**, we measure performances rather than visualizing locations. Fig. 4(d) compares expected SC and ML of our optimized percentile mechanisms to those using true optimal facility placements: the percentile mechanisms are always optimal for ML;[10] and for SC, non-strategy-proof optimal placements are only 1.75%-4.45% better than placements using our optimized, strategy-proof mechanisms.[11] This strongly suggests that percentile mechanisms, optimized using priors over preferences, are well-suited to multi-dimensional, single-peaked domains.

[9] For **2D**, D_u is uniform over $[0, 10]$ in each dimension. D_g is normal with mean $\boldsymbol{\mu} = [3, 2]$ and covariance $\boldsymbol{\Sigma} = [2, 1]\mathbf{I}$. D_{gm} is a 2 component mixture: $\mathcal{N}([-2, -1], [2, 1]\mathbf{I})$ (weight 0.3) and $\mathcal{N}([0, 2], [1, 3]\mathbf{I})$ (weight 0.7). For **4D**, D_u is uniform over $[0, 10]$ in each dimension. D_g is $\mathcal{N}([3, 2, 1, 2], [2, 3, 4, 1]\mathbf{I})$. D_{gm} is a 2 component mixture: $\mathcal{N}([2, 1, 0, 1], [4, 6, 8, 5]\mathbf{I})$ (weight 0.4) and $\mathcal{N}([1, 2, 1, 0], [7, 4, 5, 8]\mathbf{I})$ (weight 0.6).

[10] This comes from the fact that the mechanism always locates two facilities at almost the same position, and achieves optimal maximum load. However, this is not always possible for more than two facilities.

[11] Computing the optimal solution in the multi-dimensional problem is NP-hard, so we use K-means clustering algorithms as approximations.

5 Conclusion and Future Research

We proposed a family of *percentile mechanisms* for multi-dimensional, multi-facility location problems, designed to be (group) strategy-proof when preferences are single-peaked. Using different costs measures, we derived several approximation ratios. We also developed a sample-based framework for optimizing percentile mechanisms that, much like automated mechanism design, exploits priors over preferences. Our empirical results demonstrate the power of this approach, showing social objectives can be optimized much more effectively than is possible using mechanisms with tight worst-case performance guarantees (indeed, our mechanisms provide close to optimal results in practice).

This work is a starting point for the design of optimized mechanisms for single-peaked domains, and can be extended in a number of ways. Obviously one can consider mechanisms for other classes of (single-peaked) preferences (e.g., quadratic [5] or symmetric single-peaked [17]). Other social objectives should be explored, including those that combine various desiderata (such as SC and ML), and those that trade off facility cost with benefit (e.g., additional facilities decrease social cost, but the expense must be factored in as well [16]). Additional development of the optimization models needed for percentile mechanisms (e.g., our MIP or MIQCP formulations) are needed to make our approach more practical; preliminary experiments suggest that local search techniques may be very promising in this respect. Sample complexity results are also of interest. Finally, incremental (or multi-stage) mechanisms that trade off social cost, communication costs, and agent privacy [9, 22] would be extremely valuable in our setting.

References

[1] Salvador Barberà. Strategy-proof social choice. In K. J. Arrow, A. K. Sen, and K. Suzumura, editors, *Handbook of Social Choice and Welfare*, volume 2, pages 731–832. North-Holland, Amsterdam, 2010.

[2] Salvador Barberà, Faruk Gul, and Ennio Stacchetti. Generalized median voter schemes and committees. *Journal of Economic Theory*, 61(2):262–289, 1993.

[3] Vöcking Berthold. Selfish load balancing. In Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V. Vazirani, editors, *Algorithmic Game Theory*, pages 517–542. Cambridge University Press, 2007.

[4] Duncan Black. On the rationale of group decision-making. *Journal of Political Economy*, 56(1):23–34, 1948.

[5] Kim C. Border and J. S. Jordan. Straightforward elections, unanimity and phantom voters. *The Review of Economic Studies*, 50(1):153–170, 1983.

[6] Edward H. Clarke. Multipart pricing of public goods. *Public Choice*, 11(1):17–33, 1971.

[7] Vincent Conitzer and Tuomas Sandholm. Complexity of mechanism design. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence (UAI-02)*, pages 103–110, Edmonton, 2002.

[8] Bruno Escoffier, Laurent Gourvès, Thang Nguyen Kim, Fanny Pascual, and Olivier Spanjaard. Strategy-proof mechanisms for facility location games with many facilities. In *Proceedings of the Second International Conference on Algorithmic Decision Theory (ADT-11)*, pages 67–81, Piscataway, NJ, 2011.

[9] Joan Feigenbaum, Aaron D. Jaggard, and Michael Schapira. Approximate privacy: Foundations and quantification (extended abstract). In *Proceedings of the Eleventh*

ACM Conference on Electronic Commerce (EC'10), pages 167–178, Cambridge, MA, 2010.

[10] Dimitris Fotakis and Christos Tzamos. Winner-imposing strategyproof mechanisms for multiple facility location games. In *Proceedings of the Sixth International Workshop on Internet and Network Economics (WINE-10)*, pages 234–245, Stanford, CA, 2010.

[11] Theodore Groves. Incentives in teams. *Econometrica*, 41:617–631, 1973.

[12] Elias Koutsoupias and Christos Papadimitriou. Worst-case equilibria. In *Proceedings of the 16th Annual Conference on Theoretical Aspects of Computer Science (STACS-99)*, pages 404–413, Trier, DE, 1999.

[13] Anton Likhodedov and Tuomas Sandholm. Methods for boosting revenue in combinatorial auctions. In *Proceedings of the Nineteenth National Conference on Artificial Intelligence (AAAI-04)*, pages 232–237, Pittsburgh, PA, 2004.

[14] Anton Likhodedov and Tuomas Sandholm. Approximating revenue-maximizing combinatorial auctions. In *Proceedings of the Twentieth National Conference on Artificial Intelligence (AAAI-05)*, pages 267–274, Pittsburgh, PA, 2005.

[15] Pinyan Lu, Xiaorui Sun, Yajun Wang, and Zeyuan A. Zhu. Asymptotically optimal strategy-proof mechanisms for two-facility games. In *Proceedings of the Eleventh ACM Conference on Electronic Commerce (EC'10)*, pages 315–324, Cambridge, MA, 2010.

[16] Tyler Lu and Craig Boutilier. Budgeted social choice: From consensus to personalized decision making. In *Proceedings of the Twenty-second International Joint Conference on Artificial Intelligence (IJCAI-11)*, pages 280–286, Barcelona, Spain, 2011.

[17] Jordi Massó and Inés Moreno de Barreda. On strategy-proofness and symmetric single-peakedness. *Games and Economic Behavior*, 72(2):467–484, 2011.

[18] Hervé Moulin. On strategy-proofness and single peakedness. *Public Choice*, 35(4):437–455, 1980.

[19] Ariel D. Procaccia and Moshe Tennenholtz. Approximate mechanism design without money. In *Proceedings of the Tenth ACM Conference on Electronic Commerce (EC'09)*, pages 177–186, Stanford, CA, 2009.

[20] Tuomas Sandholm. Automated mechanism design: A new application area for search algorithms. In *Proceedings of the International Conference on Principles and Practice of Constraint Programming (CP-03)*, Kinsale, Ireland, 2003.

[21] James Schummer and Rakesh V. Vohra. Strategy-proof location on a network. *Journal of Economic Theory*, 104(2):405–428, 2002.

[22] Xin Sui and Craig Boutilier. Efficiency and privacy tradeoffs in mechanism design. In *Proceedings of the Twenty-fifth AAAI Conference on Artificial Intelligence (AAAI-11)*, pages 738–744, San Francisco, CA, 2011.

[23] William Vickrey. Counterspeculation, auctions, and competitive sealed tenders. *Journal of Finance*, 16(1):8–37, 1961.

Xin Sui
Dept. of Computer Science
University of Toronto
Toronto, ON, CANADA
Email: `xsui@cs.toronto.edu`

Craig Boutilier
Dept. of Computer Science
University of Toronto
Toronto, ON, CANADA
Email: `cebly@cs.toronto.edu`

Tuomas Sandholm
Computer Science Dept.
Carnegie Mellon University
Pittsburgh, PA, USA
Email: `sandholm@cs.cmu.edu`