Maximizing Expected Utility for Stochastic Combinatorial Optimization Problems [*]

Jian Li[†1] and Amol Deshpande[‡2]

[1]Institute for Interdisciplinary Information Sciences, Tsinghua University, Beijing, P.R.China
[2]Department of Computer Science, University of Maryland, College Park, USA

Abstract

We study the stochastic versions of a broad class of combinatorial problems where the weights of the elements in the input dataset are uncertain. The class of problems that we study includes shortest paths, minimum weight spanning trees, and minimum weight matchings, and other combinatorial problems like knapsack. We observe that the expected value is inadequate in capturing different types of *risk-averse* or *risk-prone* behaviors, and instead we consider a more general objective which is to maximize the *expected utility* of the solution for some given utility function, rather than the expected weight (expected weight becomes a special case). Under the assumption that there is a pseudopolynomial time algorithm for the *exact* version of the problem (This is true for the problems mentioned above), [1] we can obtain the following approximation results for several important classes of utility functions:

1. If the utility function μ is continuous, upper-bounded by a constant and $\lim_{x \to +\infty} \mu(x) = 0$, we show that we can obtain a polynomial time approximation algorithm with an *additive error* ϵ for any constant $\epsilon > 0$.

2. If the utility function μ is a concave increasing function, we can obtain a polynomial time approximation scheme (PTAS).

3. If the utility function μ is increasing and has a bounded derivative, we can obtain a polynomial time approximation scheme.

Our results recover or generalize several prior results on stochastic shortest path, stochastic spanning tree, and stochastic knapsack. Our algorithm for utility maximization makes use of the separability of exponential utility and a technique to decompose a general utility function into exponential utility functions, which may be useful in other stochastic optimization problems.

1 Introduction

The most common approach to deal with optimization problems in presence of uncertainty is to optimize the expected value of the solution. However, expected value is inadequate in expressing diverse people's

---

[1]Following the literature [55], we differentiate between *exact* version and *deterministic* version of a problem; in the exact version of the problem, we are given a target value and asked to find a solution (e.g., a path) with exactly that value (i.e., path length).

preferences towards decision-making under uncertain scenarios. In particular, it fails at capturing different *risk-averse* or *risk-prone* behaviors that are commonly observed. Consider the following simple example where we have two lotteries L_1 and L_2. In L_1, the player could win 1000 dollars with probability 1.0, while in L_2 the player could win 2000 dollars with probability 0.5 and 0 dollars otherwise. It is easy to see that both have the same expected payoff of 1000 dollars. However, many, if not most, people would treat L_1 and L_2 as two completely different choices. Specifically, a risk-averse player is likely to choose L_1 and a risk-prone player may prefer L_2 (Consider a gambler who would like to spend 1000 dollars to play double-or-nothing). A more involved but also more surprising example is the *St. Petersburg paradox* (see e.g., [45]) which has been widely used in the economics literature as a criticism of expected value. The paradox is named from Daniel Bernoulli's presentation of the problem, published in 1738 in the Commentaries of the Imperial Academy of Science of Saint Petersburg. Consider the following game: you pay a fixed fee X to enter the game. In the game, a fair coin is tossed repeatedly until a tail appears ending the game. The payoff of the game is 2^k where k is the number of heads that appear, i.e., you win 1 dollar if a tail appears on the first toss, 2 dollars if a head appears on the first toss and a tail on the second, 4 dollars if a head appears on the first two tosses and a tail on the third and so on. The question is what would be a fair fee X to enter the game? First, it is easy to see that the expected payoff is

$$\mathbb{E}[\text{payoff}] = \frac{1}{2} \cdot 1 + \frac{1}{4} \cdot 2 + \frac{1}{8} \cdot 4 + \frac{1}{16} \cdot 8 + \cdots = \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \frac{1}{2} + \cdots = \sum_{k=1}^{\infty} \frac{1}{2} = \infty$$

If we use the expected payoff as a criterion for decision making, we should therefore play the game at any finite price X (no matter how large X is) since the expected payoff is always larger. However, researchers have done extensive survey and found that not many people would pay even 25 dollars to play the game [45], which significantly deviates from what the expected value criterion predicts. In fact, the paradox can be resolved by expected utility theory with a logarithmic utility function, suggested by Bernoulli himself [7]. We refer interested reader to [59, 45] for more information. These observations and criticisms have led researchers, especially in Economics, to study the problem from a more fundamental perspective and to directly maximize user satisfaction, often called *utility*. The uncertainty present in the problem instance naturally leads us to optimize the *expected utility*.

Let $\mathcal{F}$ be the set of feasible solutions to an optimization problem. Each solution $S \in \mathcal{F}$ is associated with a random weight $w(S)$. For instance, $\mathcal{F}$ could be a set of lotteries and $w(S)$ is the (random) payoff of lottery S. We model the risk awareness of a user by a utility function $\mu : \mathbb{R} \to \mathbb{R}$: the user obtains $\mu(x)$ units of utility if the outcome is x, i.e., $w(S) = x$. Formally, the *expected utility maximization principle* is simply stated as follows: the most desirable solution S is the one that maximizes the expected utility, i.e.,

$$S = \arg \max_{S' \in \mathcal{F}} \mathbb{E}[\mu(w(S'))]$$

Indeed, expected utility theory is a branch of utility theory that studies "betting preferences" of people with regard to uncertain outcomes (gambles). The theory was formally initiated by von Neumann and Morgenstern in 1940s [65, 24] [2] who gave an axiomatization of the theory (known as *von Neumann-Morgenstern expected utility theorem*). The theory is well known to be versatile in expressing diverse risk-averse or risk-prone behaviors.

In this paper, we focus on the following broad class of combinatorial optimization problems. The deterministic version of the problem has the following form: we are given a ground set of elements $U =$

---

[2] Daniel Bernoulli also developed many ideas, such as risk aversion and utility, in his work *Specimen theoriae novae de mensura sortis (Exposition of a New Theory on the Measurement of Risk)* in 1738 [8].

$\{e_i\}_{i=1...n}$; each element e is associated with a weight w_e; each feasible solution is a subset of the elements satisfying some property. Let $\mathcal{F}$ denote the set of feasible solutions. The objective for the deterministic problem is to find a feasible solution S with the minimum (or maximum) total weight $w(S) = \sum_{e \in S} w_e$. We can see that many combinatorial problems such as shortest path, minimum spanning tree, and minimum weight matching belong to this class. In the stochastic version of the problem, the weight w_e of each element e is a nonnegative random variable. We assume all w_es are independent of each other. We use $p_e(.)$ to denote the probability density function for w_e (or probability mass function in the discrete case). We are also given a utility function $\mu : \mathbb{R}^+ \to \mathbb{R}^+$ which maps a weight value to a utility value. By the expected utility maximization principle, our goal here is to find a feasible solution $S \in \mathcal{F}$ that maximizes the expected utility, i.e., $\mathbb{E}[\mu(w(S))]$. We call this problem the *expected utility maximization (*EUM*)* problem.

Let us use the following toy example to illustrate the rationale behind EUM. There is a graph with two nodes s and t and two parallel links e_1 and e_2. Edge e_1 has a fixed length 1 while the length of e_2 is 0.9 with probability 0.9 and 1.9 with probability 0.1 (the expected value is also 1). We want to choose one edge to connect s and t. It is not hard to imagine that a risk-averse user would choose e_1 since e_2 may turn out to be a much larger value with a nontrivial probability. We can capture such behavior using the utility function (1) (defined in Section 1.1). Similarly, we can capture the risk-prone behavior by using, for example, the utility function $\mu(x) = \frac{1}{x+1}$. It is easy to see that e_1 maximizes the expected utility in the former case, and e_2 in the latter.

1.1 Our Contributions

In order to state our contribution, we first recall some standard terminologies. A polynomial time approximation scheme (PTAS) is an algorithm which takes an instance of a minimization problem (a maximization problem resp.) and a parameter $\epsilon > 0$ and produces a solution whose cost is at most $(1 + \epsilon)\mathsf{OPT}$ (at least $(1 - \epsilon)\mathsf{OPT}$ resp.), and the running time, for any fixed constant $\epsilon > 0$, is polynomial in the size of the input, where OPT is the optimal solution. We use $\mathfrak{A}$ to denote the deterministic combinatorial optimization problem under consideration. The *exact version* of $\mathfrak{A}$ asks the question whether there is a feasible solution of $\mathfrak{A}$ with weight exactly equal to a given integer K. We say an algorithm runs in *pseudopolynomial time* for the exact version of $\mathfrak{A}$ if the running time is polynomial in n and K. For many combinatorial problems, a pseudopolynomial algorithm for the exact version is known. Examples include shortest path, spanning tree, matching and knapsack.

We discuss in detail our results for EUM. We start with a theorem which underpins our other results. We denote $\|\mu\|_\infty = \sup_{x \geq 0} |\mu(x)|$. We say a function $\widetilde{\mu}(x)$ is an ϵ-*approximation* of $\mu(x)$ if $|\widetilde{\mu}(x) - \mu(x)| \leq \epsilon \|\mu\|_\infty$ for all $x \geq 0$. We allow $\widetilde{\mu}(x)$ to be a complex function (as we will see shortly, $\widetilde{\mu}(x)$ takes the form of a finite sum of complex exponentials).

Theorem 1 *Assume that there is a pseudopolynomial algorithm for the exact version of $\mathfrak{A}$. Further assume that given any constant $\epsilon > 0$, we can find an ϵ-approximation of the utility function μ as $\widetilde{\mu}(x) = \sum_{k=1}^{L} c_k \phi_k^x$, where $|\phi_k| \leq 1$ for all $1 \leq k \leq L$ (ϕ_k may be complex numbers). Let $\tau = \max_k |c_k| / \|\mu\|_\infty$. Then, there is an algorithm that runs in time $(n\tau/\epsilon)^{O(L)}$ and finds a feasible solution $S \in \mathcal{F}$ such that*

$$\mathbb{E}[\mu(w(S))] \geq \mathsf{OPT} - \epsilon \|\mu\|_\infty.$$

From the above theorem, we can see that if we can ϵ-approximate the utility function μ by a short sum of exponentials, we can obtain good approximation algorithms for EUM($\mathfrak{A}$). In this paper, we consider three important classes of utility functions.
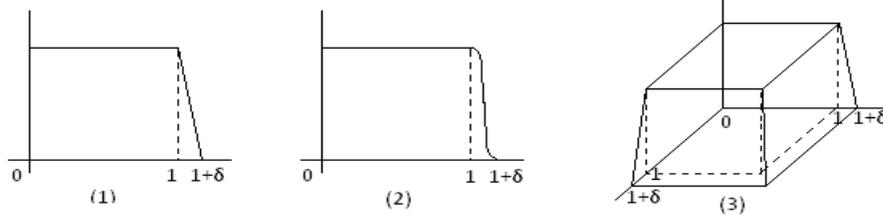
Figure 1: (1) The utility function $\widetilde{\chi}(x)$, a continuous variant of the threshold function $\chi(x)$; (2) A smoother variant of $\chi(x)$; (3) The utility function $\widetilde{\chi}_2(x)$, a continuous variant of the 2-d threshold function $\chi_2(x)$.

1. (Class $\mathfrak{C}_{\text{bounded}}$) Suppose $\mathfrak{A}$ is a minimization problem. In other words, we would like the cost of our solution to be as small as possible. In the corresponding stochastic version of $\mathfrak{A}$, we assume that $\mu(x)$ is bounded, continuous and $\lim_{x \to \infty} \mu(x) = 0$ (please see below for the detailed technical assumptions). The last condition captures the fact that if the cost of solution is too large, it becomes almost useless for us. We denote the class of such utility functions by $\mathfrak{C}_{\text{bounded}}$.

2. (Class $\mathfrak{C}_{\text{concave}}$) Suppose $\mathfrak{A}$ is a maximization problem. In other words, we want the value of our solution to be as large as possible. In the corresponding stochastic version of $\mathfrak{A}$, we assume that $\mu(x)$ is a positive, monotone nondecreasing and concave function. Note that concave functions are extensively used to model risk-averse behaviors in the economics literature. We denote the class of such utility functions by $\mathfrak{C}_{\text{concave}}$.

3. (Class $\mathfrak{C}_{\text{increasing}}$) Suppose $\mathfrak{A}$ is a maximization problem. In the corresponding stochastic version of $\mathfrak{A}$, we assume that $\mu(x)$ is a positive, differentiable and increasing function. We assume $\frac{\mathrm{d}}{\mathrm{d}x}\mu(x) \in [\mathsf{L}, \mathsf{U}]$ for $x \geq 0$, where $\mathsf{L}, \mathsf{U} > 0$ are constants. We denote the class of such utility functions by $\mathfrak{C}_{\text{increasing}}$. We can see that functions in $\mathfrak{C}_{\text{increasing}}$ can be concave, nonconcave, convex or nonconvex. Convex functions are often associated with risk-prone behaviors, while nonconvex-nonconcave utility functions have been also observed in various settings [36, 23].

Now, we state in details our assumptions and results for the above classes of utility functions.

Class $\mathfrak{C}_{\text{bounded}}$: Since μ is bounded, by scaling, without loss of generality, we can assume $\|\mu\|_\infty = 1$. Since $\lim_{x \to \infty} \mu(x) = 0$, for any $\epsilon > 0$, there exist a point T_ϵ such that $\mu(x) \leq \epsilon$ for $x > T_\epsilon$. We assume that T_ϵ is a constant only depending on ϵ. We further assume that the continuous utility function μ satisfies the α-Hölder condition, i.e., $|\mu(x) - \mu(y)| \leq C |x - y|^\alpha$, for some constant C and some constant $\alpha > 1/2$. We say f is C-*Lipschitz* if f satisfies 1-Hölder condition with coefficient C. Under the above conditions, we can prove Theorem 2.

Theorem 2 *If the utility function μ belongs to $\mathfrak{C}_{\text{bounded}}$, then, for any $\epsilon > 0$, we can obtain a function $\widetilde{\mu}(x) = \sum_{k=1}^{L} c_k \phi_k^x$, such that $|\widetilde{\mu}(x) - \mu(x)| \leq \epsilon, \text{for } x \geq 0, \text{where}$*

$$L = 2^{O(T_\epsilon)}\mathrm{poly}(1/\epsilon), \quad |c_k| \leq 2^{O(T_\epsilon)}\mathrm{poly}(1/\epsilon), \quad |\phi_k| \leq 1 \, \textit{for all } k = 1, \ldots, L,$$

To show the above theorem, we use the Fourier series technique. However, the technique cannot be used directly since it works only for periodic functions with bounded periodicities. In order to get a good approximation for $x \in [0, \infty)$, we leverage the fact that $\lim_{x \to \infty} \mu(x) = 0$ and develop a general framework that uses the Fourier series decomposition as a subroutine.

4

Now, we state some implications of the above results. Consider the utility function

$$\widetilde{\chi}(x) = \begin{cases} 1 & x \in [0, 1] \\ -\frac{x}{\delta} + \frac{1}{\delta} + 1 & x \in [1, 1 + \delta] \\ 0 & x > 1 + \delta \end{cases} \tag{1}$$

where $\delta > 0$ is a small constant (See Figure 1(1)). It is easy to verify that $\widetilde{\chi}$ is $1/\delta$-Lipschitz and $T_\epsilon = 2$ for any $\delta < 1$. Therefore, Theorem 2 is applicable. This example is interesting since $\widetilde{\chi}$ can be viewed as a continuous variant of the threshold function

$$\chi(x) = \begin{cases} 1 & x \in [0, 1] \\ 0 & x > 1 \end{cases}, \tag{2}$$

for which maximizing the expected utility is equivalent to maximizing $\Pr(w(S) \le 1)$. We first note that even the problem of computing the probability $\Pr(w(S) \le 1)$ exactly for a fixed set S is #P-hard [38] and there is an FPTAS [42]. Designing approximation algorithms for such special case has been considered several times in the literature for various combinatorial problems including stochastic shortest path [52], stochastic spanning tree [35, 26], stochastic knapsack [27] and some other stochastic problems [2, 50].

It is interesting to compare our result with the result for the stochastic shortest path problem considered by Nikolova et al. [52, 50]. In [52], they show that there is an exact $O(n^{\log n})$ time algorithm for maximizing the probability that the length of the path is at most 1, i.e., $\Pr(w(S) \le 1)$, assuming all edges are normally distributed and there is a path with its mean at most 1. Later, Nikolova [50] extends the result to an FPTAS for any problem under the same assumptions, if the deterministic version of the problem has a polynomial time exact algorithm. We can see that under such assumptions, the optimal probability is at least $1/2$. [3] Therefore, provided the same assumption and further assuming that $\Pr(w_e < 0)$ is miniscule, [4] our algorithm is a PTAS for maximizing $\mathbb{E}[\widetilde{\chi}(w(S))]$, which can be thought as a variant of the problem of maximizing $\mathbb{E}[\chi(w(S))]$. Indeed, we can translate this result to a bi-criterion approximation result of the following form: for any fixed constants $\delta, \epsilon > 0$, we can find in polynomial time a solution S such that

$$\Pr(w(S) \le 1 + \delta) \ge (1 - \epsilon) \Pr(w(S^*) \le 1).$$

where S^* is the optimal solution (Corollary 2). We note that such a bi-criterion approximation was only known for exponentially distributed edges before [52].

Let us consider another application of our results to the stochastic knapsack problem defined in [27]. Given a set U of independent random variables $\{x_1, \ldots, x_n\}$, with associated profits $\{v_1, \ldots, v_n\}$ and an overflow probability γ, we are asked to pick a subset S of U such that

$$\Pr(\sum_{i \in S} x_i \ge 1) \le \gamma$$

and the total profit $\sum_{i \in S} v_i$ is maximized. Goel and Indyk [27] showed that, for any constant $\epsilon > 0$, there is a polynomial time algorithm that can find a solution S with the profit as least the optimum and $\Pr(\sum_{i \in S} x_i \ge 1 + \epsilon) \le \gamma(1 + \epsilon)$ for exponentially distributed variables. They also gave a quasi-polynomial

---

[3]The sum of multiple Gaussians is also a Gaussian. Hence, if we assume the mean of the length of a path (which is a Gaussian) is at most 1, the probability that the length of the path is at most 1 is at least $1/2$.

[4]Our technique can only handle distributions with positive supports. Thus, we have to assume that the probability that a negative value appears is miniscule (e.g., less than $1/n^2$) and can be safely ignored (because the probability that there is any realized negative value is at most $1/n$).

5

time approximation scheme for Bernoulli distributed random variables. Quite recently, in parallel with our work, Bhalgat et al. [13] obtained the same result for arbitrary distributions under the assumption that $\gamma = \Theta(1)$. Their technique is based on discretizing the distributions and is quite involved. [5] Our result, applied to stochastic knapsack, matches that of Bhalgat et al. Our algorithm is arguably simpler and has a much better running time (Theorem 7).

Equally importantly, we can extend our basic approximation scheme to handle generalizations such as multiple utility functions and multidimensional weights. Interesting applications of these extensions include various generalizations of stochastic knapsack, such as *stochastic multiple knapsack* (Theorem 10) and *stochastic multidimensional knapsack (stochastic packing)* (Theorem 11).

Class $\mathfrak{C}_{\text{concave}}$: We assume the utility function $\mu : [0, \infty) \to [0, \infty)$ is a concave, monotone nondecreasing function. This is a popular class of utility functions used to model risk-averse behaviors. For this class of utility functions, we can obtain the following theorem in Section 5.

Theorem 3 *Assume the utility function μ belongs to $\mathfrak{C}_{\text{concave}}$, and there is a pseudopolynomial algorithm for the exact version of $\mathfrak{A}$. Then, there is a PTAS for $\mathsf{EUM}(\mathfrak{A})$.*

Theorem 3 is also obtained by an application of Theorem 1. However, instead of approximating the original utility function μ using a short sum of exponentials, which may not be possible in general, [6] we try approximate a truncated version of μ. Theorem 3 recovers the recent result of [14]. Finally, we remark the technique of [14] strongly relies on the concavity of μ, and seems difficult to extend to handle non-concave utility functions.

Class $\mathfrak{C}_{\text{increasing}}$: We assume the utility function $\mu : [0, \infty) \to [0, \infty)$ is a positive, differentiable, and increasing function. For technical reasons, we assume $\frac{\mathrm{d}}{\mathrm{d}x}\mu(x) \in [\mathsf{L}, \mathsf{U}]$ for some constants $\mathsf{L}, \mathsf{U} > 0$ and all $x \geq 0$. For this class of utility functions, we can obtain the following theorem in Section 6.

Theorem 4 *Assume the utility function μ belongs to $\mathfrak{C}_{\text{increasing}}$, and there is a pseudopolynomial algorithm for the exact version of $\mathfrak{A}$. Then, there is a PTAS for $\mathsf{EUM}(\mathfrak{A})$.*

Again, it may not be possible in general to approximate such an increasing function using a finite sum of exponentials. Instead, we approximate a truncated version of μ, similar to the concave case. We note this is the first such result for general increasing utility functions. Removing the bounded derivative assumption remains an interesting open problem.

We believe our technique can be used to handle other classes of utility functions or other stochastic optimization problems.

1.2 Related Work

In recent years stochastic optimization problems have drawn much attention from the computer science community and stochastic versions of many classical combinatorial optimization problems have been studied. In particular, a significant portion of the efforts has been devoted to the two-stage stochastic optimization problem. In such a problem, in a first stage, we are given probabilistic information about the input but the cost of selecting an item is low; in a second stage, the actual input is revealed but the costs for the elements

---

[5]Their discretization technique, combined with other ideas, can be used to handle the *adaptive version* of the stochastic knapsack problem as well [13].

[6]Suppose μ is a finite sum of exponentials. When x approaches to infinity, either $|\mu(x)|$ is periodic, or approaches to infinity, or approaches to 0.

are higher. We are asked to make decision after each stage and minimize the expected cost. Some general techniques have been developed [31, 60]. We refer interested reader to [64] for a comprehensive survey. Another widely studied type of problems considers designing *adaptive* probing policies for stochastic optimization problems where the existence or the exact weight of an element can be only known upon a probe. There is typically a budget for the number of probes (see e.g., [30, 19]), or we require an irrevocable decision whether to include the probed element in the solution right after the probe (see e.g., [22, 17, 4, 21, 13]). However, most of those works focus on optimizing the expected value of the solution. There is also sporadic work on optimizing the overflow probability or some other objectives subject to the overflow probability constraints. In particular, a few recent works have explicitly motivated such objectives as a way to capture the risk-averse type of behaviors [2, 50, 63]. Besides those works, there has been little work on optimizing more general utility functions for combinatorial stochastic optimization problems from an approximation algorithms perspective.

The most related work to ours is the stochastic shortest path problem (Stoch-SP), which was also the initial motivation for this work. The problem has been studied extensively for several special utility functions in operation research community. Sigal et al. [61] studied the problem of finding the path with greatest probability of being the shortest path. Loui [44] showed that Stoch-SP reduces to the shortest path (and sometimes longest path) problem if the utility function is linear or exponential. Nikolova et al. [51] identified more specific utility and distribution combinations that can be solved optimally in polynomial time. Much work considered dealing with more general utility functions, such as piecewise linear or concave functions, e.g., [48, 49, 6]. However, these algorithms are essentially heuristics and the worst case running times are still exponential. Nikolova et al. [52] studied the problem of maximizing the probability that the length of the chosen path is less than some given parameter. Besides the result we mentioned before, they also considered Poisson and exponential distributions. Despite much effort on this problem, no algorithm is known to run in polynomial time and have provable performance guarantees, especially for more general utility functions or more general distributions. This is perhaps because the hardness comes from different sources, as also noted in [52]: the shortest path selection per se is combinatorial; the distribution of the length of a path is the convolution of the distributions of its edges; the objective is nonlinear; to list a few.

Kleinberg et al. [38] first considered the stochastic knapsack problem with Bernoulli-type distributions and provided a polynomial-time $O(\log 1/\gamma)$ approximation where γ is the given overflow probability. For item sizes with exponential distributions, Goel and Indyk [27] provided a bi-criterion PTAS, and for Bernoulli-distributed items they gave a quasi-polynomial approximation scheme. Chekuri and Khanna [16] pointed out that a PTAS can be obtained for the Bernoulli case using their techniques for the multiple knapsack problem. Goyal and Ravi [29] showed a PTAS for Gaussian distributed sizes. Bhalgat, Goel and Khanna [13] developed a general discretizaton technique that reduces the distributions to a small number of equivalent classes which we can efficiently enumerate for both adaptive and nonadaptive versions of stochastic knapsack. They used this technique to obtain improved results for several variants of stochastic knapsack, notably a bi-criterion PTAS for the *adaptive version* of the problem. In a recent work [43], the bi-criterion PTAS was further simplified and extended to the more general case where the profit and size of an item can be correlated and an item can be cancelled in the middle. Dean at al. [22] gave the first constant approximation for the adaptive version of stochastic knapsack. The adaptive version of stochastic multidimensional knapsack (or equivalently stochastic packing) has been considered in [21, 13] where constant approximations and a bi-criterion PTAS were developed.

This work is partially inspired by our prior work on top-k and other queries over probabilistic datasets [39, 41]. In fact, we can show that both the *consensus answers* proposed in [39] and the *parameterized ranking functions* proposed in [41] follow the expected utility maximization principle where the utility functions

are materialized as distance metrics for the former and the weight functions for the latter. Our technique for approximating the utility functions is also similar to the approximation scheme used in [41] in spirit. However, no performance guarantees are provided in that work.

Recently, Li and Yuan [43] showed that an additive PTAS for $\mu \in \mathfrak{C}_{\text{bounded}}$ can be obtained using a completely different approach, called the Poisson approximation technique. Roughly speaking, the Poisson approximation technique allows us to extract a constant (depending on ϵ) number of features from each distribution (called signature in [43]) and reduce the stochastic problem to a constant dimensional deterministic optimization problem, which is similar to the algorithm presented in this paper. We suspect that besides this superficial similarity, there may be deeper connections between two different techniques.

There is a large volume of work on approximating functions using short exponential sums over a bounded domain, e.g., [54, 9, 10, 11]. Some works also consider using linear combinations of Gaussians or other *kernels* to approximate functions with finite support over the entire real axis $(-\infty, +\infty)$ [18]. This is however impossible using exponentials since α^x is either periodic (if $|\alpha| = 1$) or approaches to infinity when $x \to +\infty$ or $x \to -\infty$ (if $|\alpha| \neq 1$).

2   An Overview of Our Approach

The high level idea of our approach is very simple and consists of the following steps:

1. We first observe that the problem is easy if the utility function is an exponential function. Specifically, consider the exponential utility function $\mu(x) = \phi^x$ for some complex number $\phi \in \mathbb{C}$. Fix an arbitrary solution S. Due to independence of the elements, we can see that

$$\mathbb{E}[\phi^{w(S)}] = \mathbb{E}\big[\phi^{\sum_{e \in S} w_e}\big] = \mathbb{E}\Big[\prod_{e \in S} \phi^{w_e}\Big] = \prod_{e \in S} \mathbb{E}[\phi^{w_e}]$$

   Taking log on both sides, we get $\log \mathbb{E}[\phi^{w(S)}] = \sum_{e \in S} \log \mathbb{E}[\phi^{w_e}]$. If ϕ is a positive real number and $\mathbb{E}[\phi^{w_e}] \leq 1$ (or equivalently, $-\log \mathbb{E}[\phi^{w_e}] \geq 0$), this reduces to the deterministic optimization problem.

2. In light of the above observation, we ϵ-approximate the utility function $\mu(x)$ by a short exponential sum, i.e., $\sum_{i=1}^{L} c_i \phi_i^x$ with L being a small value (only depending on ϵ), where (c_i and ϕ_i may be complex numbers). Hence, $\mathbb{E}[\mu(w(S))]$ can be approximated by $\sum_{i=1}^{L} c_i \mathbb{E}[\phi_i^{w(S)}]$.

3. Consider the following multi-criterion version of the problem with L objectives $\{\mathbb{E}[\phi_i^{w(S)}]\}_{i=1,\ldots,L}$: given L complex numbers $v_1, \ldots, v_L$, we want to find a solution S such that $\mathbb{E}[\phi_i^{w(S)}] \approx v_i$ for $i = 1, \ldots, L$. We achieve this by utilizing the pseudopolynomial time algorithm for the exact version of the problem. We argue that we only need to consider a polynomial number of $v_1, \ldots, v_L$ combinations (which we call *configurations*) to find out the approximate optimum.

In Section 3, we show how to solve the multi-criterion problem provided that a short exponential sum approximation of μ is given. In particular, we prove Theorem 1. Then, we show how to approximate $\mu \in \mathfrak{C}_{\text{bounded}}$ by a short exponential sum by proving Theorem 2 in Section 4.1 and Section 4.2. For $\mu \in \mathfrak{C}_{\text{concave}}$ or $\mu \in \mathfrak{C}_{\text{increasing}}$, it may not be possible to approximate μ directly by an exponential sum, and some additional ideas are required. The details are provided in Section 5 and Section 6.

We still need to show how to compute $\mathbb{E}[\phi^{w_e}]$. If w_e is a discrete random variable with a polynomial size support, we can easily compute $\mathbb{E}[\phi^{w_e}]$ in polynomial time. If w_e has an infinite discrete or continuous

support, we can not compute $\mathbb{E}[\phi^{w_e}]$ directly and need to approximate it. We briefly discuss this issue and its implications in Appendix A.

3 Proof of Theorem 1

Now, we prove Theorem 1. We start with some notations. We use $|c|$ and $\arg(c)$ to denote the absolute value and the argument of the complex number $c \in \mathbb{C}$, respectively. In other words, $c = |c| \cdot (\cos(\arg(c)) + i\sin(\arg(c))) = |c|e^{i\arg(c)}$. We always require $\arg(c) \in [0, 2\pi)$ for any $c \in \mathbb{C}$. Recall that we say the exponential sum $\sum_{i=1}^{L} c_i \phi_i^x$ is an ϵ-*approximation* for $\mu(x)$ if the following holds:

$$|\mu(x) - \sum_{i=1}^{L} c_i \phi_i^x| \leq \epsilon \|\mu\|_\infty \quad \text{for } x \geq 0.$$

We first show that if the utility function can be decomposed exactly into a short exponential sum, we can approximate the optimal expected utility well.

Theorem 5 *Assume that $\widetilde{\mu}(x) = \sum_{k=1}^{L} c_k \phi_k^x$ is the utility function where $|\phi_k| \leq 1$ for $1 \leq k \leq L$. Let $\tau = \max_k |c_k|/\|\mu\|_\infty$. We also assume that there is a pseudopolynomial algorithm for the exact version of $\mathfrak{A}$. Then, for any $\epsilon > 0$, there is an algorithm that runs in time $(n/\epsilon)^{O(L)}$ and finds a solution S such that*

$$|\mathbb{E}[\widetilde{\mu}(w(S))] - \mathbb{E}[\widetilde{\mu}(w(\widetilde{S}))]| < \epsilon \|\mu\|_\infty,$$

where $\widetilde{S} = \arg\max_{S'} |\mathbb{E}[\widetilde{\mu}(w(S'))]|$.

We use the scaling and rounding technique that has been used often in multi-criterion optimization problems (e.g., [58, 55]). Since our objective function is not additive and not monotone, the general results for multi-criterion optimization [55, 46, 58, 1] do not directly apply here. We provide the details of the algorithm here. We use the following parameters:

$$\gamma = \frac{\epsilon}{Ln\tau}, \quad J = \max\left(\left\lceil \frac{-\ln(\epsilon/L\tau)n}{\gamma} \right\rceil, \left\lceil \frac{2\pi n}{\gamma} \right\rceil\right).$$

Let $\mathbb{V}$ be the set of all $2L$-dimensional integer vectors of the form $\mathbf{v} = \langle x_1, y_1, \ldots, x_L, y_L \rangle$ where $1 \leq x_i \leq J$ and $1 \leq y_i \leq J$ for $i = 1, \ldots, L$.

For each element $e \in U$, we associate it with a $2L$-dimensional integer vector

$$\mathsf{Ft}(e) = \langle \alpha_1(e), \beta_1(e), \ldots, \alpha_L(e), \beta_L(e) \rangle,$$

$$\text{where} \quad \alpha_i(e) = \left\lfloor \min\left(\frac{-\ln|\mathbb{E}[\phi_i^{w_e}]|}{\gamma}, \frac{J}{n}\right) \right\rfloor \quad \text{and} \quad \beta_i(e) = \left\lfloor \frac{\arg(\mathbb{E}[\phi_i^{w_e}])}{\gamma} \right\rfloor. \tag{3}$$

We call $\mathsf{Ft}(e)$ the *feature vector* of e. Since $|\phi_i| \leq 1$, we can see that $\alpha_i(e) \geq 0$ for any $e \in U$. It is easy to see that $\mathsf{Ft}(e) \in \mathbb{V}$ for all $e \in U$ and $\sum_{e \in S} \mathsf{Ft}(e) \in \mathbb{V}$ for all $S \subseteq U$. Intuitively, $\alpha_i(e)$ and $\beta_i(e)$ can be thought as the scaled and rounded versions of $-\ln|\mathbb{E}[\phi_i^{w_e}]|$ and $\arg(\mathbb{E}[\phi_i^{w_e}])$, respectively.

We maintain $J^{2L} = (n/\epsilon)^{O(L)}$ *configurations* (a configuration is just like a state in a dynamic program). Each configuration $\mathsf{Cf}(\mathbf{v})$ is indexed by a $2L$-dimensional vector $\mathbf{v} \in \mathbb{V}$ and takes 0/1 value. In particular, the value of $\mathsf{Cf}(\mathbf{v})$ for each $\mathbf{v} \in \mathbb{V}$ is defined as follows: For each vector $\mathbf{v} \in \mathbb{V}$,

1. $\mathsf{Cf}(\mathbf{v}) = 1$ if and only if there is a feasible solution $S \in \mathcal{F}$ such that $\sum_{e \in S} \mathsf{Ft}(e) = \mathbf{v}$.

2. $\mathsf{Cf}(\mathbf{v}) = 0$ otherwise.

For any $\mathbf{v} = \langle x_1, y_1, \ldots, x_L, y_L \rangle$, define the *value* of $\mathbf{v}$ to be

$$\mathsf{Val}(\mathbf{v}) = \sum_{k=1}^{L} c_k e^{-x_k \gamma + i y_k \gamma}.$$

Lemma 1 tells us the value of a configuration is close to the expected utility of the corresponding solution. Lemma 2 shows we can compute those configurations in polynomial time.

Lemma 1 *Suppose $\widetilde{\mu}(x) = \sum_{k=1}^{L} c_k \phi_k^x$, where $|\phi_k| \leq 1$ for all $k = 1, \ldots, L$. Let $\tau = \max_k |c_k| / \|\mu\|_\infty$. For any vector $\mathbf{v} = \langle x_1, y_1, \ldots, x_L, y_L \rangle \in \mathbb{V}$, $\mathsf{Cf}_v(\mathbf{v}) = 1$ if and only if there is a feasible solution $S \in \mathcal{F}$ such that*

$$\left| \mathbb{E}[\widetilde{\mu}(w(S))] - \mathsf{Val}(\mathbf{v}) \right| = \left| \mathbb{E}[\widetilde{\mu}(w(S))] - \sum_{k=1}^{L} c_k e^{-x_k \gamma + i y_k \gamma} \right| \leq O(\epsilon \|\mu\|_\infty).$$

Proof: We first notice that $\mathbb{E}[\widetilde{\mu}(w(S))] = \mathbb{E}[\sum_{k=1}^{L} c_k \phi_k^{w(S)}] = \sum_{k=1}^{L} c_k \mathbb{E}[\phi_k^{w(S)}]$. Therefore, it suffices to show that for all $k = 1, \ldots, L$, $|\mathbb{E}[\phi_k^{w(S)}] - e^{-x_k \gamma + i y_k \gamma}| \leq O(\frac{\epsilon}{L\tau})$. Since $\mathsf{Cf}(\mathbf{v}) = 1$, we know that $\sum_{e \in S} \mathsf{Ft}(e) = \mathbf{v}$ for some feasible solution $S \in \mathcal{F}$. In other words, we have $\sum_{e \in S} \alpha_k(e) = x_k$ and $\sum_{e \in S} \beta_k(e) = y_k$ for all $1 \leq k \leq L$.

Fix an arbitrary $1 \leq k \leq L$. First, we can see that the arguments of $\mathbb{E}[\phi_k^{w(S)}]$ and $e^{-x_k \gamma + i y_k \gamma}$ are close:

$$\left| \arg(\mathbb{E}[\phi_k^{w(S)}]) - y_k \gamma \right| \leq \sum_{e \in S} \left| \arg(\mathbb{E}[\phi_k^{w_e}]) - \beta_k(e) \gamma \right| \leq \sum_{e \in S} \gamma \leq n\gamma = \frac{\epsilon}{L\tau},$$

where we use $\arg(c)$ to denote the argument of the complex number c. Now, we show the magnitude of $\mathbb{E}[\phi_k^{w(S)}]$ and $e^{-x_k \gamma + i y_k \gamma}$ are also close. We distinguish two cases:

1. Recall that $\alpha_i(e) = \left\lfloor \min\left(\frac{-\ln |\mathbb{E}[\phi_i^{w_e}]|}{\gamma}, \frac{J}{n} \right) \right\rfloor$. If there is some $e \in S$ such that $\frac{-\ln |\mathbb{E}[\phi_i^{w_e}]|}{\gamma} > \frac{J}{n}$ (which implies that $\alpha_k(e) = \lfloor \frac{J}{n} \rfloor$), we know that

$$-\ln(|\mathbb{E}[\phi_k^{w(S)}]|) = \sum_{e \in S}(-\ln(|\mathbb{E}[\phi_k^{w_e}]|)) > \frac{J\gamma}{n}.$$

In this case, we have $x_k = \sum_{e \in S} \alpha_k(e) \geq \frac{J}{n}$. Thus, we have that

$$\left| |\mathbb{E}[\phi_k^{w(S)}]| - |e^{-x_k \gamma}| \right| < e^{-J\gamma/n} \leq e^{\gamma \lceil \frac{n \ln(\epsilon/L\tau)}{\gamma} \rceil / n} < \frac{\epsilon}{L\tau}.$$

2. On the other hand, if $\alpha_k(e) = \left\lfloor \frac{-\ln |\mathbb{E}[\phi_k^{w_e}]|}{\gamma} \right\rfloor$ for all $e \in S$, we can see that

$$-\ln(|\mathbb{E}[\phi_k^{w(S)}]|) - x_k \gamma = \sum_{e \in S}(-\ln(|\mathbb{E}[\phi^{w_e}]|) - \alpha_k(e) \gamma) \leq \sum_{e \in S} \gamma \leq n\gamma \leq \frac{\epsilon}{L\tau}.$$

Since the derivative of e^x is less than 1 for $x < 0$, we can get that

$$\left| |\mathbb{E}[\phi_k^{w(S)}]| - |e^{-x_k \gamma}| \right| \leq |e^{-x_k \gamma - \epsilon/L\tau} - e^{-x_k \gamma}| \leq \frac{\epsilon}{L\tau}.$$

For any two complex numbers a, b with $|a| \leq 1$ and $|b| \leq 1$, if $\left||a| - |b|\right| < \epsilon$ and $|\angle ab| = |\arg(a) - \arg(b)| < \epsilon$, we can see that

$$\begin{aligned}
|a - b|^2 &= |a|^2 + |b|^2 - 2|a||b|\cos(\angle ab) \\
&= (|a| - |b|)^2 + 2|a||b|(1 - \cos(\angle ab)) \\
&\leq \epsilon^2 + 2(1 - \cos(\angle ab)^2) \\
&\leq \epsilon^2 + 2\sin(\angle ab)^2 \\
&\leq \epsilon^2 + 2|\arg(a) - \arg(b)|^2 \leq 3\epsilon^2.
\end{aligned}$$

In the third inequality, we use the fact that $\sin x < x$ for all $x > 0$. The proof is completed. $\qquad \square$

Lemma 2 *Suppose there is a pseudopolynomial time algorithm for the exact version of $\mathfrak{A}$, which runs in time polynomial in n and t (t is the maximum integer in the instance of $\mathfrak{A}$). Then, we can compute the values for all configurations $\{\mathsf{Cf}(\mathbf{v})\}_{\mathbf{v} \in \mathbb{V}}$ in time $(\frac{n\tau}{\epsilon})^{O(L)}$.*

Proof: For each vector $\mathbf{v} \in \mathbb{V}$, we can encode it as a nonnegative integer $I(\mathbf{v})$ upper bounded by $J^{2L} = (\frac{n}{\epsilon})^{O(L)}$. In particular, each coordinate of $\mathbf{v}$ takes the position of a specific digit in the integral representation, and the base is chosen to be J no carry can occur when we add at most n feature vectors. Then, determining the value of a configuration $\mathsf{Cf}(\mathbf{v})$ is equivalent to determining whether there is a feasible solution $S \in \mathcal{F}$ such that the total weight of S (i.e., $\sum_{e \in S} I(\mathsf{Ft}(e))$) is exactly the given value $I(\mathbf{v})$. Suppose the pseudopolynomial time algorithm for the exact version of $\mathfrak{A}$ runs in time $P_{\mathfrak{A}}(n, t)$ for some polynomial $P_{\mathfrak{A}}$. Therefore, the value of each such $\mathsf{Cf}(\mathbf{v})$ can be also computed in time $P_{\mathfrak{A}}(n, I(\mathbf{v})) = P_{\mathfrak{A}}(n, (\frac{n}{\epsilon})^{O(L)}) = (\frac{n}{\epsilon})^{O(L)}$. Since J are bounded by $(\frac{n\tau}{\epsilon})^{O(1)}$, the number of configuration is $(\frac{n\tau}{\epsilon})^{O(L)}$. The total running time is $(\frac{n\tau}{\epsilon})^{O(L)} \times (\frac{n\tau}{\epsilon})^{O(L)} = (\frac{n\tau}{\epsilon})^{O(L)}$. $\qquad \square$

Now, everything is ready to prove Theorem 5.

Proof of Theorem 5: We first use the algorithm in Lemma 2 to compute the values for all configurations. Then, we find the configuration $\mathsf{Cf}(\langle x_1, y_1, \ldots, x_L, y_L \rangle)$ that has value 1 and that maximizes the quantity $|\mathsf{Val}(\mathbf{v})| = |\sum_{k=1}^{L} c_k e^{-x_k\gamma + iy_k\gamma}|$. The feasible solution S corresponding to this configuration is our final solution. It is easy to see that the theorem follows from Lemma 1. $\qquad \square$

Theorem 1 can be readily obtained from Theorem 5 and the fact $\widetilde{\mu}$ is an ϵ-approximation of μ.

Proof of Theorem 1: Suppose S is our solution and S^* is the optimal solution for utility function μ. Recall $\widetilde{\mu}(x) = \sum_{k=1}^{L} c_k \phi_k^x$. From Theorem 5, we know that

$$|\mathbb{E}[\widetilde{\mu}(w(S))]| \geq |\mathbb{E}[\widetilde{\mu}(w(S^*))]| - O(\epsilon\|\mu\|_\infty).$$

Since $\widetilde{\mu}$ is an ϵ-approximation of μ, we can see that

$$\left|\mathbb{E}[\mu(w(S))] - \mathbb{E}[\widetilde{\mu}(w(S))]\right| = \left|\int (\mu(x) - \widetilde{\mu}(x))\mathrm{d}P_S(x)\right| \leq \left|\int \epsilon\|\mu\|_\infty \mathrm{d}P_S(x)\right| \leq \epsilon\|\mu\|_\infty.$$

for any solution S, where P_S is the probability measure of $w(S)$. Therefore, we have

$$\begin{aligned}
|\mathbb{E}[\mu(w(S))]| &\geq |\mathbb{E}[\widetilde{\mu}(w(S))]| - \epsilon\|\mu\|_\infty \geq |\mathbb{E}[\widetilde{\mu}(w(S^*))]| - O(\epsilon\|\mu\|_\infty) \\
&\geq |\mathbb{E}[\mu(w(S^*))]| - O(\epsilon\|\mu\|_\infty).
\end{aligned}$$

This completes the proof of Theorem 1. $\qquad \square$

4 Class $\mathfrak{C}_{\mathsf{bounded}}$

The main goal of this section is to prove Theorem 2. In Section 4.1, we develop a generic algorithm that takes as a subroutine an algorithm AP for approximating functions in a bounded interval domain, and approximates $\mu(x) \in \mathfrak{C}_{\mathsf{bounded}}$ in the infinite domain $[0, +\infty)$. In the Section 4.2, we use the Fourier series expansion as the choice of AP and show that important classes of utility functions can be approximated well.

4.1 Approximating the Utility Function

There are many works on approximating functions using short exponential sums, e.g., the Fourier decomposition approach [62], Prony's method [54], and many others [9, 10]. However, their approximations are done over a finite interval domain, say $[-\pi, \pi]$ or over a finite number of discrete points. No error bound can be guaranteed outside the domain. Our algorithm is a generic procedure that turns an algorithm that can approximate functions over $[-\pi, \pi]$ into one that can approximate our utility function μ over $[0, +\infty)$, by utilizing the fact that $\lim_{x \to \infty} \mu(x) = 0$.

    Recall for $\mu \in \mathfrak{C}_{\mathsf{bounded}}$, we assume that for any constant $\epsilon > 0$, there exist a constant T_ϵ such that $\mu(x) \leq \epsilon$ for $x > T_\epsilon$. We also assume there is an algorithm AP that, for any function f (under some conditions specified later), can produce an exponential sum $\widehat{f}(x) = \sum_{i=1}^{L} c_i \phi_i^x$ which is an ϵ-approximation of $f(x)$ in $[-\pi, \pi]$ such that $|\phi_i| \leq 1$ and L depends only on ϵ and f. In fact, we can assume w.l.o.g. that AP can approximate $f(x)$ over $[-B, B]$ for any $B = O(1)$. This is because we can apply AP to the scaled version $g(x) = f(x \cdot \frac{B}{\pi})$ (which is defined on $[-\pi, \pi]$) and then scale the obtained approximation $\widehat{g}(x)$ back to $[-B, B]$, i.e., the final approximation is $\widehat{f}(x) = \widehat{g}(\frac{\pi}{B} \cdot x)$. Scaling a function by a constant factor $\frac{B}{\pi}$ typically does not affect the smoothness of f in any essential way and we can still apply AP. Recall that our goal is to produce an exponential sum that is an ϵ-approximation for $\mu(x)$ in $[0, +\infty)$. We denote this procedure by ESUM-APPROX.

Algorithm: ESUM-APPROX

1. Initially, we slightly change function $\mu(x)$ to a new function $\widehat{\mu}(x)$ as follows: We require $\widehat{\mu}(x)$ is a "smooth" function in $[-2T_\epsilon, 2T_\epsilon]$ such that $\widehat{\mu}(x) = \mu(x)$ for all $x \in [0, T_\epsilon]$; $\widehat{\mu}(x) = 0$ for $|x| \geq 2T_\epsilon$. We choose $\widehat{\mu}(x)$ in $[-2T_\epsilon, 0]$ and $[T_\epsilon, 2T_\epsilon]$ such that $\widehat{\mu}(x)$ is smooth. We do not specify the exact smoothness requirements now since they may depend on the choice of AP. Note that there may be many ways to interpolate μ such that the above conditions are satisfied (see Example 1 below). The only properties we need are: (1) $\widehat{\mu}$ is amenable to algorithm AP; (2) $|\widehat{\mu}(x) - \mu(x)| \leq \epsilon$ for $x \geq 0$.

2. We apply AP to $g(x) = \eta^x \widehat{\mu}(x)$ over domain $[-hT_\epsilon, hT_\epsilon]$ ($\eta \geq 1$ and $h \geq 2$ are constants to be determined later). Suppose the resulting exponential sum $\widehat{g}(x) = \sum_{i=1}^{L} c_i \phi_i^x$, such that $|\widehat{g}(x) - g(x)| \leq \epsilon$ for all $x \in [-hT_\epsilon, hT_\epsilon]$.

3. Let $\widetilde{\mu}(x) = \sum_{i=1}^{L} c_i (\frac{\phi_i}{\eta})^x$, which is our final approximation of $\mu(x)$ on $[0, \infty)$.

Example 1 *Consider the utility function $\mu(x) = 1/(x + 1)$. Let $T_\epsilon = \frac{1}{\epsilon} - 1$. So $\mu(x) < \epsilon$ for all $x > T_\epsilon$. Now we create function $\widehat{\mu}(x)$ according to the first step of* ESUM-APPROX. *If we only require $\widehat{\mu}(x)$ to be continuous, then we can use, for instance, the following piecewise function: $\widehat{\mu}(x) = \frac{1}{x+1}$ for $x \in [0, T_\epsilon]$; $\widehat{\mu}(x) = -\frac{x}{\epsilon T_\epsilon} + \frac{2}{\epsilon}$ for $x \in [T_\epsilon, 2T_\epsilon]$; $\widehat{\mu}(x) = 0$ for $x > 2T_\epsilon$; $\widehat{\mu}(x) = \widehat{\mu}(-x)$ for $x < 0$. It is easy to see that $\widehat{\mu}$ is continuous and ϵ-approximates μ.* $\square$

By setting $\eta = 2$ and

$$h \geq \max \left(2, \frac{\log(\sum_{i=1}^{L} |c_i|/\epsilon)}{T_\epsilon} \right), \tag{4}$$

we can show the following theorem.

Lemma 3 $\widetilde{\mu}(x)$ *is a 2ϵ-approximation of $\mu(x)$.*

Proof: We know that $|\widehat{g}(x) - g(x)| \leq \epsilon$ for $x \in [0, hT_\epsilon]$. Therefore, we have that

$$|\widetilde{\mu}(x) - \widehat{\mu}(x)| = \left| \frac{\widehat{g}(x)}{\eta^x} - \frac{g(x)}{\eta^x} \right| \leq \frac{\epsilon}{\eta^x} \leq \epsilon.$$

Combining with $|\widehat{\mu}(x) - \mu(x)| \leq \epsilon$, we obtain $|\widetilde{\mu}(x) - \mu(x)| \leq 2\epsilon$ for $x \in [0, hT_\epsilon]$. For $x > hT_\epsilon$, we can see that

$$|\widetilde{\mu}(x)| = \left| \sum_{i=1}^{L} c_i \left(\frac{\phi_i}{\eta} \right)^x \right| \leq \sum_{i=1}^{L} \left| c_i \left(\frac{\phi_i}{\eta} \right)^x \right| \leq \frac{1}{2^x} \sum_{i=1}^{L} |c_i| \leq \frac{1}{2^{hT_\epsilon}} \sum_{i=1}^{L} |c_i| \leq \epsilon$$

Since $\mu(x) < \epsilon$ for $x > hT_\epsilon$, the proof is complete. $\qquad\square$

Remark: Since we do not know c_i before applying AP, we need to set h to be a quantity (only depending on ϵ and T_ϵ) such that (4) is always satisfied. In particular, we need to provide an upper bound for $\sum_{i=1}^{L} |c_i|$. In the next subsection, we use the Fourier series decomposition as the choice for AP, which allows us to provide such a bound for a large class of functions.

4.2 A Particular Choice of AP: The Fourier Series Approach

Now, we discuss the choice of algorithm AP and the conditions that $f(x)$ needs to satisfy so that it is possible to approximate $f(x)$ by a short exponential sum in a bounded interval. In fact, if we know in advance that there is a short exponential sum that can approximate f, we can use the algorithms developed in [10, 11] (for continuous case) and [9] (for the discrete case). However, those works do not provide an easy characterization of the class of functions. From now on, we restrict ourselves to the classic Fourier series technique, which has been studied extensively and allows such characterizations.

Suppose $f(x)$ is a real periodic function defined on $[-\pi, \pi]$. Consider the partial sum of the Fourier series of the function $f(x)$:

$$(S_N f)(x) = \sum_{k=-N}^{N} c_k e^{ikx}$$

where the Fourier coefficient $c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} \mathrm{d}x$. It has $L = 2N + 1$ terms. Since $f(x)$ is a real function, we have $c_k = c_{-k}$ and the partial sum is also real. We are interested in the question under which conditions does the function $S_N f$ converge to f (as N increases) and what is convergence rate? Roughly speaking, the "smoother" f is, the faster $S_N f$ converges to f. However, this question in general is quite intricate and deep and is one of the central topics in harmonic analysis. In the following, we need one classic result about the convergence of Fourier series and show how to use it in our problem.

We say f satisfies the α-*Hölder condition* if $|f(x) - f(y)| \leq C |x - y|^\alpha$, for some constant C and $\alpha > 0$ and any x and y. The constant C is called the *Hölder coefficient* of f, also denoted as $|f|_{C^{0,\alpha}}$. We say f is *C-Lipschitz* if f satisfies 1-Hölder condition with coefficient C.

Example 2 *It is easy to check that the utility function μ in Example 1 is 1-Lipschitz since $|\frac{d\mu(x)}{dx}| \leq 1$ for $x \geq 0$. We can also see that $\widetilde{\chi}(x)$ (defined in (1)) is $\frac{1}{\delta}$-Lipschitz.*

We need the following classic result of Jackson.

Theorem 6 (See e.g., [56]) *If f satisfies the α-Hölder condition, it holds that*

$$|f(x) - (S_N f)(x)| \leq O\Big(\frac{|f|_{C^{0,\alpha}} \ln N}{N^\alpha}\Big).$$

We are ready to state the details of AP, which is simply as follows: Let $f(x) = g(xhT_\epsilon/\pi)$ where g is defined in ESUM-APPROX. Then, AP returns the following function $\widehat{g}$, which is a sum of exponential functions:

$$\widehat{g}(x) = S_N f\left(\frac{x\pi}{hT_\epsilon}\right).$$

Now, we show that $|\widehat{g}(x) - g(x)| \leq \epsilon$ for all $x \in [-hT_\epsilon, hT_\epsilon]$. For the later parts of the analysis, we need a few simple lemmas. The proofs of these lemmas are straightforward and thus omitted here.

Lemma 4 *Suppose $f : [a, c] \to \mathbb{R}$ is a continuous function which consists of two pieces $f_1 : [a, b] \to \mathbb{R}$ and $f_2 : [b, c] \to \mathbb{R}$. If both f_1 and f_2 satisfy the α-Hölder condition with Hölder coefficient C, then $|f|_{C^{0,\alpha}} \leq 2C$.*

Lemma 5 *Suppose $g : [a, c] \to \mathbb{R}$ is a continuous function satisfying the α-Hölder condition with Hölder coefficient C. Then, for $f(x) = g(tx)$ for some $t > 0$, we have $|f|_{C^{0,\alpha}} \leq Ct^\alpha$.*

By Lemma 4, we know that the piecewise function $\widehat{\mu}$ (defined in ESUM-APPROX) satisfies α-Hölder condition with coefficient $2C$. Therefore, we can easily see that $g(x) = \widehat{\mu}(x)\eta^x$ satisfies α-Hölder condition with coefficient at most $2^{1+2T_\epsilon}C$ on $[-hT_\epsilon, hT_\epsilon]$ (This is because $\widehat{\mu}$ is non-zero only in $[-2T_\epsilon, 2T_\epsilon]$). According to Lemma 5, we have $|f(x)|_{C^{0,\alpha}} = |g(xhT_\epsilon/\pi)|_{C^{0,\alpha}} \leq 2^{1+2T_\epsilon}(hT_\epsilon/\pi)^\alpha C$. Using Theorem 6, we obtain the following corollary.

Corollary 1 *Suppose $\mu \in \mathfrak{C}_{\text{bounded}}$ satisfies the α-Hölder condition with $|\mu|_{C^{0,\alpha}} = O(1)$. For*

$$N = 2^{O(T_\epsilon)}(h/\epsilon)^{1+1/\alpha},$$

it holds that $|g(x) - \widehat{g}(x)| \leq \epsilon$ for $x \in [-hT_\epsilon, hT_\epsilon]$.

How to Choose h: Now, we discuss the issue left in Section 4.1, that is how to choose h (the value should be independent of c_is and L) to satisfy (4), when μ satisfies the α-Hölder condition for some $\alpha > 1/2$. We need the following results about the absolute convergence of Fourier coefficients. If f satisfies the α-Hölder condition for some $\alpha > 1/2$, then $\sum_{i=-\infty}^{+\infty} |c_i| \leq |f|_{C^{0,\alpha}} \cdot c_\alpha$ where c_α only depends on α [62]. We can see that in order to ensure (4), it suffices to to set value h such that

$$hT_\epsilon \geq \log \frac{2^{1+2T_\epsilon}(hT_\epsilon/\pi)^\alpha Cc_\alpha}{\epsilon} = 2T_\epsilon + O\big(\log\big(hT_\epsilon/\epsilon\big)\big).$$

We can easily verify that the above condition can be satisfied by letting $h = \max(O(\frac{1}{T_\epsilon} \log \frac{1}{\epsilon}), 2)$.
Proof of Theorem 2: Everything is in place to prove Theorem 2. First, we bound L by Corollary 1:

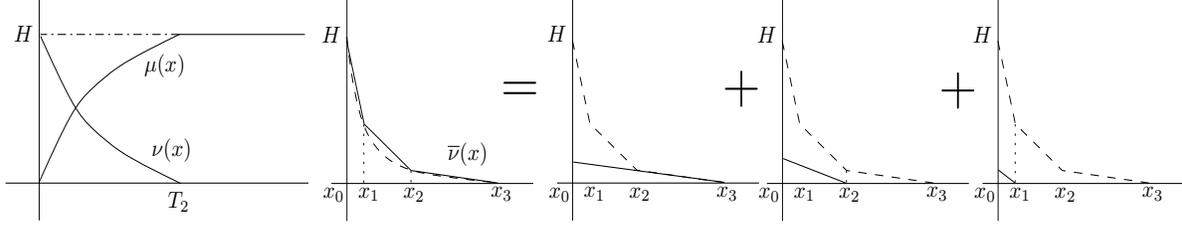$$L = 2N + 1 = 2^{O(T_\epsilon)}\text{poly}(1/\epsilon).$$

Figure 2: (1) The concave utility function $\mu(x)$ and $\nu(x) = H - \mu(x)$ (2) The piecewise linear function $\overline{\nu}(x)$. (3)-(5) Decomposing $\overline{\nu}(x)$ into three scaled copies of $\tau(x)$.

Next, we bound the magnitude of each c_k. Recall c_k is the Fourier coefficient: $c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} \mathrm{d}x$. where $f(x) = g(xhT_\epsilon/\pi) = \eta^{xhT_\epsilon/\pi} \widehat{\mu}(xhT_\epsilon/\pi)$ for $x \in [-\pi, \pi]$. Since $hT_\epsilon = \max(O(T_\epsilon), O(\log 1/\epsilon))$, we can see $|f(x)| \le 2^{O(T_\epsilon)} \mathrm{poly}(1/\epsilon)$ for $x \in [-\pi, \pi]$. Therefore,

$$|c_k| \le \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(x)| \mathrm{d}x \le 2^{O(T_\epsilon)} \mathrm{poly}(1/\epsilon).$$

Finally, combining Corollary 1 and Lemma 3, we complete the proof of Theorem 2. $\qquad\square$

5 Class $\mathfrak{C}_{\mathsf{concave}}$

In this section, we handle the case where the utility function $\mu : [0, \infty) \to [0, \infty)$ is a concave nondecreasing function and our goal is to prove Theorem 3.

We use OPT to denote the optimal value of our problem. We can assume without loss of generality that we know OPT, modulo a multiplicative factor of $(1 \pm \epsilon)$. This can be done by guessing all powers of $(1 + \epsilon)$ between $\max_{e \in U} \mathbb{E}[\mu(w_e)]$ and $\mathbb{E}[\mu(w(U))]$, [7] and run our algorithm for each guess. Let

$$H = \mathsf{OPT}/\epsilon^2, \quad T_1 = \mu^{-1}(\mathsf{OPT}/\epsilon) \quad \text{and} \quad T_2 = \mu^{-1}(\mathsf{OPT}/\epsilon^2). \tag{5}$$

We first make the following simplifying assumption and show how to remove it later:

S1. We assume $\mu(0) = 0$ and $\mu(x) = \mu(T_2) = H$ for all $x > T_2$.

Lemma 6 *If the utility function $\mu \in \mathfrak{C}_{\mathsf{concave}}$ satisfies the additional assumption S1, then, for any $\epsilon > 0$, we can obtain an exponential sum $\widetilde{\mu}(x) = \sum_{k=1}^{L} c_k \phi_k^x$, such that $|\widetilde{\mu}(x) - \mu(x)| \le O(\epsilon \mathsf{OPT})$ for all $x > 0$, where $L = \mathrm{poly}(1/\epsilon)$, $|c_k| \le \mathrm{poly}(1/\epsilon) H$ and $|\phi_k| \le 1$ for all $k = 1, \ldots, L$.*

Proof: Consider the function $\nu(x) = H - \mu(x)$. We can see ν is a nonincreasing convex function and $\nu(x) = 0$ for all $x > T_2$. We first approximate ν by a piecewise linear function $\overline{\nu}$ as follows. Let $N = 1/\epsilon^3$. For all $0 \le i \le N$, let

$$x_i = \mu^{-1}\left(\frac{iH}{N}\right) = \nu^{-1}\left(\frac{(N-i)H}{N}\right) \quad \text{and} \quad x_{N+1} = \infty.$$

---

[7] We can assume every $e \in U$ is in at least one feasible solution $S \in \mathcal{F}$. Otherwise, we can simply remove those irrelevant elements. Then, OPT is at least $\max_{e \in U} \mathbb{E}[\mu(w_e)]$.

Let $h_i = \frac{\nu(x_{i+1}) - \nu(x_i)}{x_{i+1} - x_i}$ for $0 \le i \le N$. The piecewise linear function $\overline{\nu}$ is defined by $\overline{\nu}(x_i) = \nu(x_i)$ for all $0 \le i \le N$ and

$$\overline{\nu}(x) = \overline{\nu}(x_i) + (x - x_i)h_i, \quad \text{for } x \in [x_i, x_{i+1}].$$

It is easy to see $\overline{\nu}$ is also a convex function (see Figure 2) and $|\nu(x) - \overline{\nu}(x)| \le H/N \le \epsilon\mathsf{OPT}$.

Now we show $\overline{\nu}$ can be written as a linear sum of N scaled copies of the following function ρ:

$$\rho(x) = 1 - x \quad \text{for } 0 \le x \le 1, \text{ and } \rho(x) = 0 \quad \text{for } x > 1.$$

We let $\rho_{h,a}(x) = (-ha)\rho(x/a)$. It is easy to see that the first piece of $\rho_{h,a}$ has slope h and ends at $x = a$. Define

$$\rho_i(x) = \rho_{h_i - h_{i+1}, x_{i+1}}(x) = x_{i+1}(h_{i+1} - h_i)\rho\left(\frac{x}{x_{i+1}}\right) \quad \text{for } 0 \le i \le N.$$

It is not hard to verify that $\overline{\nu}(x) = \sum_{i=0}^{N-1} \rho_i(x)$ (see Figure 2).

By Theorem 2, we can find a function $\widetilde{\tau}(x) = \sum_{k=1}^{D} d_k \psi_k^x$ with $D = \mathrm{poly}(N/\epsilon^2) = \mathrm{poly}(1/\epsilon)$, $|d_k| = \mathrm{poly}(1/\epsilon)$ and $|\psi_k| \le 1$ for $k = 1, \ldots, D$, [8] such that $|\widetilde{\tau}(x) - \rho(x)| \le \epsilon^2/N$ for $x \ge 0$. Consider the function

$$\widetilde{\nu}(x) = \sum_{i=0}^{N-1} x_{i+1}(h_{i+1} - h_i)\widetilde{\tau}\left(\frac{x}{x_{i+1}}\right) = \sum_{i=0}^{N-1}\sum_{k=1}^{D} x_{i+1}(h_{i+1} - h_i)d_k(\psi_k^{1/x_{i+1}})^x.$$

Clearly, $\widetilde{\nu}$ is the summation of ND exponentials. It is not difficult to see the magnitude of each coefficient, $|x_{i+1}(h_{i+1} - h_i)d_k|$, is at most $-x_{i+1}h_i|d_k| \le \mathrm{poly}(1/\epsilon)H$. We can also see that

$$|\widetilde{\nu}(x) - \nu(x)| \le |\widetilde{\nu}(x) - \overline{\nu}(x)| + |\overline{\nu}(x) - \nu(x)| \le O(H\epsilon^2) \le O(\epsilon\mathsf{OPT}) \text{ for } x \ge 0.$$

Finally, letting $\widetilde{\mu}(x) = H - \widetilde{\nu}(x)$ finishes the proof. $\qquad\square$

Since $\|\mu\|_\infty = \mathsf{OPT}/\epsilon^2$, Lemma 6 implies that $\widetilde{\mu}$ is an ϵ^3-approximation of μ. Then, applying Theorem 1, we can immediately obtain a polynomial time algorithm that runs in time $(n/\epsilon)^{\mathrm{poly}(1/\epsilon)}$ and finds a solution $S \in \mathcal{F}$ such that $\mathsf{OPT} - \mathbb{E}[\mu(S)] \le \epsilon^2\|\mu\|_\infty \le \epsilon\mathsf{OPT}$, i.e., a PTAS.

Now, we show how to get rid of the assumption S1. From now on, the utility function μ is a general increasing concave utility function with $\mu(0) = 0$. [9] Let $\mu_H(x) = \min(\mu(x), H)$. We can see that $\mu_H(x)$ satisfies S1. We say a value $p \in \mathbb{R}^+$ is *huge* is if $p > T_2$. Otherwise, we call it *normal*. We use Huge to denote the set of huge values. For each element e, let w_e^{nm} be the random variable which has the same distribution as w_e in the normal value region, and zero probability elsewhere. For any $S \subseteq U$, let $w^{\mathrm{nm}}(S) = \sum_{e \in S} w_e^{\mathrm{nm}}$. In the following lemma, we show μ_H is a good approximation for μ for normal values.

Lemma 7 *For any $S \in \mathcal{F}$, we have that*

$$\mathbb{E}[\mu(w^{\mathrm{nm}}(S))] - O(\epsilon)\mathsf{OPT} \le \mathbb{E}[\mu_H(w^{\mathrm{nm}}(S))] \le \mathbb{E}[\mu(w^{\mathrm{nm}}(S))].$$

Proof: It is obvious that $\mathbb{E}[\mu_H(w^{\mathrm{nm}}(S))] \le \mathbb{E}[\mu(w^{\mathrm{nm}}(S))]$. So, we only need to prove the first inequality. For any $S \subseteq \mathcal{F}$, we have $\mathbb{E}[\mu(w(S))] \le \mathsf{OPT}$. By Markov inequality, $\Pr[w(S) \ge T_1] \le \epsilon$, which implies $\Pr[w^{\mathrm{nm}}(S) \ge T_1] \le \epsilon$. Now, we claim that for any integer $k \ge 1$,

$$\Pr\left[w^{\mathrm{nm}}(S) \ge (k+2)T_1\right] \le \epsilon \Pr\left[w^{\mathrm{nm}}(S) \ge kT_1\right]. \tag{6}$$

---

[8] It suffices to let $T_\epsilon = 1$ (i.e., $\rho(x) = 0$ for $x \ge 1$).

[9] The assumption that $\mu(0) = 0$ is without loss of generality. If $\mu(0) > 0$, we can solve the problem with the new utility function $\mu(x) - \mu(0)$. It is easy to verify a PTAS for the new problem is a PTAS for the original problem.

Consider the following stochastic process. Suppose the weights of the elements in S are realized one by one (say $w^{\mathsf{nm}}_{e_1}, \ldots, w^{\mathsf{nm}}_{e_n}$). Let Z_t be the sum of the first t realized values. Let t_1 be the first time such that $Z_{t_1+1} \geq kT_1$. If this never happens, let $t_1 = \infty$ and $Z_{t_1} = Z_n$. Let event E_1 be $t_1 \leq n$ and E_2 be $Z_n \geq (t+2)T_1$. Consider the random value $Z_n - Z_{t_1} = \sum_{t=t_1+1}^{n} w^{\mathsf{nm}}_{e_t}$. As $w^{\mathsf{nm}}(S) = Z_n = \sum_{t=1}^{n} w^{\mathsf{nm}}_{e_t}$ and all $w^{\mathsf{nm}}_{e_t}$ are nonnegative, we can see that

$$\Pr[Z_n - Z_{t_1} > T_1 \mid E_1] = \Pr\Big[\sum_{t=t_1+1}^{n} w^{\mathsf{nm}}_{e_t} > T_1 \mid t_1 \leq n \Big] \leq \Pr\Big[\sum_{t=1}^{n} w^{\mathsf{nm}}_{e_t} > T_1 \Big] \leq \epsilon.$$

Moreover, we can see that event $E_1 \wedge (Z_n - Z_{t_1} > T_1)$ is a necessary condition for event E_2. Hence, the claim holds because

$$\Pr[E_2] \leq \Pr[E_1] \Pr[Z_n - Z_{t_1} > T_1 \mid E_1] \leq \epsilon \Pr[E_1].$$

From (6), we can see that $\Pr[w^{\mathsf{nm}}(S) \geq 3T_1] \leq \epsilon^2$, $\Pr[w^{\mathsf{nm}}(S) \geq 5T_1] \leq \epsilon^3, \ldots$ so on and so forth. Furthermore, we can see that

$$\mathbb{E}[\mu(w^{\mathsf{nm}}(S))] - \mathbb{E}[\mu_H(w^{\mathsf{nm}}(S))] = \int_H^\infty \Pr[\mu(w^{\mathsf{nm}}(S)) \geq x] \mathrm{d}x = \int_H^\infty \Pr[w^{\mathsf{nm}}(S) \geq \mu^{-1}(x)] \mathrm{d}x$$

$$= \frac{\mathsf{OPT}}{\epsilon} \int_0^\infty \Pr[w^{\mathsf{nm}}(S) \geq \mu^{-1}(H + k\mathsf{OPT}/\epsilon)] \mathrm{d}k$$

$$\leq \frac{\mathsf{OPT}}{\epsilon} \int_0^\infty \Pr[w^{\mathsf{nm}}(S) \geq T_2 + kT_1] \mathrm{d}k$$

$$\leq \frac{2\mathsf{OPT}}{\epsilon} \sum_{k=2/\epsilon}^\infty \epsilon^k \leq O(\epsilon\mathsf{OPT}).$$

The first inequality holds due to the concavity of μ:

$$\mu^{-1}(H + k\mathsf{OPT}/\epsilon) \geq T_2 + \mu^{-1}(k\mathsf{OPT}/\epsilon) = T_2 + k\mu^{-1}(\mathsf{OPT}/\epsilon) = T_2 + kT_1 \text{ for } k \geq 0. \qquad \square$$

Now, we handle the contribution from huge values. Let $\mathsf{Huge} = \{p \mid p \geq T_2\}$ and

$$\mathsf{Hg}(e) = \sum_{p \in \mathsf{Huge}} \Pr[w_e = p] \mu(p).^{10}$$

$\mathsf{Hg}(e)$ can be thought as the expected contribution of huge values of e. We need the following observation in [14]: the contribution of the huge values can be essentially linearized and separated from the contribution of normal values, in the sense of the following lemma. We note that the simple insight has been used in a variety of contexts in stochastic optimization problems (e.g., [47, 33, 34]).

Lemma 8 (The first half of Theorem 2 in [14]) *For any $S \in \mathcal{F}$, we have that*

$$\mathbb{E}[\mu(w(S))] \in (1 \pm O(\epsilon))\Big(\mathbb{E}[\mu(w^{\mathsf{nm}}(S))] + \sum_{e \in S} \mathsf{Hg}(e)\Big).$$

---

[10] If w_e is continuously distributed, we let $\mathsf{Hg}(e) = \int_{T_2}^\infty \mu(x) p_e(x) \mathrm{d}x$. The algorithm and analysis for the continuous case are exactly the same as the discrete case.

Now, we are ready to state our algorithm, which is an extension of the algorithm in Section 3. Using Lemma 6, we first obtain a function $\widetilde{\mu}_H(x) = \sum_{k=1}^{L} c_k \phi_k^x$ such that $|\widetilde{\mu}_H(x) - \mu_H(x)| \leq \epsilon\mathsf{OPT}$. The feature vector $\mathsf{Ft}(e)$ is a $2L + 1$-dimensional integer vector

$$\mathsf{Ft}(e) = \langle \alpha_1(e), \beta_1(e), \ldots, \alpha_L(e), \beta_L(e), \lfloor n\mathsf{Hg}(e)/\epsilon\mathsf{OPT} \rfloor \rangle,$$

where $\alpha_i(e), \beta_i(e)$ are defined as in (3) with respect to w_e^{nm}. In other words, we extend the original feature vector by one more coordinate which represents the (scaled and rouned) contribution of huge values. Similarly, each configuration $\mathsf{Cf}(\mathbf{v})$ is indexed by such a $2L + 1$-dimensional vector $\mathbf{v}$. The last coordinate of $\mathbf{v}$ is at most n^2/ϵ. As before, we let $\mathsf{Cf}(\mathbf{v}) = 1$ if and only if there is a feasible solution $S \in \mathcal{F}$ such that $\sum_{e \in S} \mathsf{Ft}(e) = \mathbf{v}$. We slightly modify the definition of $\mathsf{Val}(\mathbf{v})$ to incorporate the contribution of huge values, as the following:

$$\mathsf{Val}(\langle x_1, y_1, \ldots, x_L, y_L, z \rangle)| = \sum_{k=1}^{L} c_k e^{-x_k \gamma + i y_k \delta} + z \cdot \frac{\epsilon\mathsf{OPT}}{n}.$$

Using the same technique as in Lemma 2 and the pseudopolynomail time algorithm for $\mathfrak{A}$, we can compute the values of all configurations in time $(n/\epsilon)^{\mathrm{poly}(1/\epsilon)}$. Then, we return the solution for which the corresponding configuration $\mathsf{Cf}(\langle x_1, y_1, \ldots, x_L, y_L, z \rangle)$ that takes value 1 and maximizes $|\mathsf{Val}(\mathbf{v})|$.

Proof of Theorem 3: The proof is similar to that of Theorem 1. Let any $S \subseteq U$, let $\mathbf{v}_S = \sum_{e \in S} \mathsf{Ft}(e)$ Using the same proof of Lemma 1 and the fact that $|\widetilde{\mu}_H(x) - \mu_H(x)| \leq \epsilon\mathsf{OPT}$, we can see that for any $S \in \mathcal{F}$,

$$|\mathsf{Val}(\mathbf{v}_S)| = \mathbb{E}[\mu(w^{\mathsf{nm}}(S))] + \sum_{e \in S} \mathsf{Hg}(e) \pm O(\epsilon\mathsf{OPT}).$$

Combining with Lemma 7 and Lemma 8, we can further see that for any $S \in \mathcal{F}$,

$$|\mathsf{Val}(\mathbf{v}_S)| = (1 \pm O(\epsilon))\mathbb{E}[\mu(w(S))] \pm O(\epsilon\mathsf{OPT}).$$

Suppose S is our solution and S^* is the optimal solution for utility function μ. From our algorithm, we know that $|\mathsf{Val}(\mathbf{v}_S)| \geq |\mathsf{Val}(\mathbf{v}_{S^*})|$, which implies $\mathbb{E}[\mu(w(S))] \geq (1 - O(\epsilon))\mathsf{OPT}$ and completes the proof.
□

6   Class $\mathfrak{C}_{\mathsf{increasing}}$

Recall that $\mu(x) \in \mathfrak{C}_{\mathsf{increasing}}$ is a positive, differentiable and increasing function and $\frac{\mathrm{d}}{\mathrm{d}x}\mu(x) \in [\mathsf{L}, \mathsf{U}]$ for some constants $\mathsf{L}, \mathsf{U} > 0$ and all $x \geq 0$. By scaling, we can assume without loss of generality that $\mathsf{L} \leq 1 \leq \mathsf{U}$. Our algorithm is almost the same as the one in Section 5 except that we use a slightly different set of parameters:

$$H = \frac{\mathsf{OPT}}{\epsilon^2} \cdot \frac{\mathsf{L}}{\mathsf{U}}, \quad T_1 = \frac{\mathsf{OPT}}{\epsilon} \cdot \frac{\mathsf{L}}{\mathsf{U}}, \quad T_2 = \mu^{-1}\left(\frac{\mathsf{OPT}}{\epsilon^2} \cdot \frac{\mathsf{L}}{\mathsf{U}}\right), \quad \text{and} \quad \mathsf{Huge} = \{p \mid p \geq T_2\}$$

Let $\mu_H(x) = \min(\mu(x), H)$. So, μ_H satisfies assumption S1. However, we can not use Lemma 6 since it requires concavity. Nevertheless, we can still approximate μ_H by a short exponential sum, as in Lemma 9. The remaining algorithm is exactly the same as the one in Section 5. To prove the performance guarantee, we only need to prove analogues of Lemma 7 and Lemma 8. Now, we prove the aforementioned lemmas.

Lemma 9 *For any $\epsilon > 0$, we can obtain an exponential sum $\widetilde{\mu}_H(x) = \sum_{k=1}^{L} c_k \phi_k^x$, such that $|\widetilde{\mu}_H(x) - \mu_H(x)| \leq O(\epsilon\mathsf{OPT})$ for all $x > 0$, where $L = \mathrm{poly}(1/\epsilon)$, $|c_k| \leq \mathrm{poly}(1/\epsilon)H$ and $|\phi_k| \leq 1$ for all k.*

Proof: Since $\frac{\mathrm{d}}{\mathrm{d}x}\mu(x) \in [\mathsf{L}, \mathsf{U}]$, we can see that $H/\mathsf{U} \leq T_2 \leq H/\mathsf{L}$. Consider $\nu(x) = H - \mu_H(x)$. We can see ν is a decreasing, differentiable function and $\nu(x) = 0$ for all $x > T_2$. Consider the function $\overline{\nu}(x) = \frac{1}{H}\nu(xH)$. First, let $T_\epsilon = 1/\mathsf{L} \geq T_2/H$ and we can see $\overline{\nu}(x) = 0$ for $x > T_\epsilon$. Hence, $\overline{\nu} \in \mathfrak{C}_{\mathsf{bounded}}$ and satisfies U-Lipschitz condition. By Theorem 2, we can compute a function $\widetilde{\nu}(x) = \sum_{k=1}^{L} d_k \psi_k^x$, which is an ϵ^2-approximation of ν, with $L = \mathrm{poly}(1/\epsilon)$, $|\psi_k| \leq 1$ and $d_k = \mathrm{poly}(1/\epsilon)$ for all k. Therefore, $\widetilde{\mu}_H(x) = H - H\widetilde{\nu}(x/H) = H - \sum_{k=1}^{L}(Hd_k)(\psi_k^{1/H})^x$ is the desired approximation. $\qquad\square$

The following lemma is an analogue of Lemma 7.

Lemma 10 *For any $S \in \mathcal{F}$, we have that*

$$\mathbb{E}[\mu_H(w^{\mathsf{nm}}(S))] \in (1 \pm O(\epsilon))\mathbb{E}[\mu(w^{\mathsf{nm}}(S))].$$

Proof: The proof is almost the same as that of Lemma 7, except that the last line makes use of the bounded derivative assumption (instead of the concavity):

$$\mu^{-1}\Big(H + \frac{k\mathsf{OPT}}{\epsilon}\Big) \geq T_2 + \frac{k\mathsf{OPT}}{\epsilon\mathsf{U}} \geq T_2 + kT_1 \text{ for } k > 0. \qquad\square$$

We handle the contribution from huge values in the same way. Recall $\mathsf{Hg}(e) = \sum_{p\in\mathsf{Huge}}\Pr[w_e = p]\mu(p)$. The following lemma is an analogue of Lemma 8.

Lemma 11 *For any $S \in \mathcal{F}$, we have that*

$$\mathbb{E}[\mu(w(S))] \in (1 \pm O(\epsilon))\Big(\mathbb{E}[\mu(w^{\mathsf{nm}}(S))] + \sum_{e\in S}\mathsf{Hg}(e)\Big).$$

Proof: We can use exactly the same proof of Theorem 2 in [14] to show that $\mathbb{E}[\mu(w(S))] \geq (1 - O(\epsilon))(\mathbb{E}[\mu(w^{\mathsf{nm}}(S))] + \sum_{e\in S}\mathsf{Hg}(e))$, as the proof holds even without the concavity assumption. The other direction requires a different argument, which goes as follows. Let E_0 be the event that no w_e is realized to a huge value and $E_{e,p}$ be the event that w_e is realized to value $p \in \mathsf{Huge}$. By Markov inequality, we have $\Pr[E_0] \geq 1 - \epsilon\mathsf{L}/\mathsf{U}$. Moreover, using the fact that $e^{-x} \geq 1 - x$, we have that

$$\exp\Big(-\sum_{e\in S, p\in\mathsf{Huge}}\Pr[E_{e,p}]\Big) = \prod_{e\in S}\exp\Big(-\sum_{p\in\mathsf{Huge}}\Pr[E_{e,p}]\Big)$$
$$\geq \prod_{e\in S}\Big(1 - \sum_{p\in\mathsf{Huge}}\Pr[E_{e,p}]\Big)$$
$$= \Pr[E_0] \geq 1 - \epsilon\mathsf{L}/\mathsf{U} \geq 1 - \epsilon.$$

Hence, $\sum_{e\in S, p\in\mathsf{Huge}}\Pr[E_{e,p}] \leq -(\ln(1-\epsilon)) \leq 2\epsilon$ for $\epsilon < 1/2$.

Next, we can see that $\mathbb{E}[\mu(w(S)) \mid E_0]\Pr[E_0] \leq \mathbb{E}[\mu(w^{\mathsf{nm}}(S))]$ (for each realization of $\{w_e\}_{e\in S}$ satisfying E_0, there is a corresponding realization of $\{w_e^{\mathsf{nm}}\}_{e\in S}$). From the bounded derivative assumption,

we can also see that $\mathbb{E}[\mu(w(S)) \mid E_{e,p}] \leq \mu(p) + \mathsf{U} \cdot \mathbb{E}[w(S)]$ By inclusion-exclusion, we have that

$$\mathbb{E}[\mu(w(S))] \leq \mathbb{E}[\mu(w(S)) \mid E_0] \Pr[E_0] + \sum_{e \in S, p \in \mathsf{Huge}} \mathbb{E}[\mu(w(S)) \mid E_{e,p}] \Pr[E_{e,p}]$$

$$\leq \mathbb{E}[\mu(w^{\mathsf{nm}}(S))] + \sum_{e \in S, p \in \mathsf{Huge}} \Pr[E_{e,p}]\mu(p) + \sum_{e \in S, p \in \mathsf{Huge}} \Pr[E_{e,p}] \cdot \mathsf{U} \cdot \mathbb{E}[w(S)]$$

$$\leq \mathbb{E}[\mu(w^{\mathsf{nm}}(S))] + \sum_{e \in S} \mathsf{Hg}(e) + O(\epsilon)\mathbb{E}[\mu(w(S))].$$

The last inequality holds since $\mathbb{E}[w(S)] \leq \mathbb{E}[\mu(w(S))]/\mathsf{L} = O(\mathbb{E}[\mu(w(S))])$. $\qquad\square$

7   Applications

We first consider two utility functions $\chi(x)$ and $\widetilde{\chi}(x)$ presented in the introduction. Note that maximizing $\mathbb{E}[\chi(w(S))]$ is equivalent to maximizing $\Pr(w(S) \leq 1)$. The following lemma is straightforward.

Lemma 12 *For any solution S,*

$$\Pr(w(S) \leq 1) \leq \mathbb{E}[\widetilde{\chi}(w(S))] \leq \Pr(w(S) \leq 1 + \delta).$$

Corollary 2 *Suppose there is a pseudopolynomial time algorithm for the exact version of $\mathfrak{A}$. Then, for any fixed constants $\epsilon > 0$ and $\delta > 0$, there is an algorithm that runs in time $(\frac{n}{\epsilon})^{\mathrm{poly}(1/\epsilon)}$, and produces a solution $S \in \mathcal{F}$ such that*
$$\Pr(w(S) \leq 1 + \delta) + \epsilon \geq \max_{S' \in \mathcal{F}} \Pr(w(S') \leq 1)$$

Proof: By Theorem 1, Theorem 2 and Lemma 12, we can easily obtain the corollary. Note that we can choose $T_\epsilon = 2$ for any $\delta \in (0,1)$ and $\epsilon > 0$. Thus $\mathsf{L} = \mathrm{poly}(1/\epsilon)$. $\qquad\square$

Now, let us see some applications of our general results to specific problems.

Stochastic Shortest Path: Finding a path with the exact target length (we allow non-simple paths)[11] can be easily done in pseudopolynomial time by dynamic programming.

Stochastic Spanning Tree: We are given a graph G, where the weight of each edge e is an independent, nonnegative random variable. Our objective is to find a spanning tree T in G, such that $\Pr(w(T) \leq 1)$ is maximized. Polynomial time algorithms have been developed for Gaussian distributed edges [35, 26]. To the best of our knowledge, no approximation algorithm with provable guarantee is known for other distributions. Noticing there exists a pseudopolynomial time algorithm for the exact spanning tree problem [5], we can directly apply Corollary 2.

Stochastic k-Median on Trees: The problem asks for a set S of k nodes in the given probabilistic tree G such that $\Pr(\sum_{v \in V(G)} \mathsf{dis}(v, S) \leq 1)$ is maximized, where $\mathsf{dis}(v, S)$ is the minimum distance from v to any node in S in the tree metric. The k-median problem can be solved optimally in polynomial time on trees by dynamic programming [37]. It is straightforward to modify the dynamic program to get a pseudopolynomial time algorithm for the exact version.

---

[11]The exact version of *simple* path is NP-hard, since it includes the Hamiltonian path problem as a special case.

Stochastic Knapsack with Random Sizes: We are given a set U of n items. Each item i has a random size w_i and a deterministic profit v_i. We are also given a positive constant $0 \leq \gamma \leq 1$. The goal is to find a subset $S \subseteq U$ such that $\Pr(w(S) \leq 1) \geq \gamma$ and the total profit $v(S) = \sum_{i \in S} v_i$ is maximized.

If the profits of the items are polynomially bounded integers, we can see the optimal profit is also a polynomially bounded integer. We can first guess the optimal profit. For each guess g, we solve the following problem: find a subset S of items such that the total profit of S is exactly g and $\mathbb{E}[\widetilde{\chi}(w(S))]$ is maximized. The exact version of the deterministic problem is to find a solution S with a given total size and a given total profit, which can be easily solved in pseudopolynomial time by dynamic programming. Therefore, by Corollary 2, we can easily show that we can find in polynomial time a set S of items such that the total profit $v(S)$ is at least the optimum and $\Pr(w(S) \leq 1 + \epsilon) \geq (1 - \epsilon)\gamma$ for any constant ϵ and γ.

If the profits are general integers, we can use the standard scaling technique to get a $(1-\epsilon)$-approximation for the total profit. We first make a guess of the optimal profit, rounded down to the nearest power of $(1+\epsilon)$. There are at most $\log_{1+\epsilon} \frac{n \max_i v_i}{\min_i v_i}$ guesses. For each guess g, we solve the following problem. We discard all items with a profit larger than g. Let $\Delta = \frac{\epsilon g}{n^2}$. For each item with a profit smaller than $\frac{\epsilon g}{n}$, we set its new profit to be $\bar{v}_i = 0$. Then, we scale each of the rest profits v_i to $\bar{v}_i = \Delta \lfloor \frac{v_i}{\Delta} \rfloor$. Now, we define the feasible set

$$\mathcal{F}(g) = \{S \mid \sum_{i \in S} (1 - 2\epsilon)g \leq \sum_{i \in S} \bar{v}_i \leq (1 + 2\epsilon)g\}.$$

Since there are at most $\frac{n^2}{\epsilon}$ distinct $\bar{v}$ values, we can easily show that finding a solution S in $\mathcal{F}(g)$ with a given total *size* can be solved in pseudopolynomial time by dynamic programming.

Denote the optimal solution by S^* and the optimal profit by OPT. Suppose g is the right guess, i.e., $(\frac{1}{1+\epsilon})OPT \leq g \leq OPT$. We can easily see that for any solution S, we have that

$$(1 - \frac{1}{n}) \sum_{i \in S} v_i - \epsilon g \leq \sum_{i \in S} \bar{v}_i \leq \sum_{i \in S} v_i$$

where the first inequalities are due to $v_i \geq \frac{\epsilon g}{n}$ and we set at most ϵg profit to zero. Therefore, we can see $S^* \in \mathcal{F}(g)$. Applying Corollary 2, we obtain a solution S such that $\Pr(w(S) \leq 1 + \delta) + \epsilon \geq \Pr(w(S^*) \leq 1 + \delta)$. Moreover, the profit of this solution $v(S) = \sum_{i \in S} v_i \geq \sum_{i \in S} \bar{v}_i \geq (1 - 2\epsilon)g \geq (1 - O(\epsilon))OPT$.

In sum, we have obtained the following result.

Theorem 7 *For any constants $\epsilon > 0$ and $\gamma > 0$, there is a polynomial time algorithm to compute a set S of items such that the total profit $v(S)$ is within a $1-\epsilon$ factor of the optimum and $\Pr(w(S) \leq 1+\epsilon) \geq (1-\epsilon)\gamma$.*

Bhalgat et al. [13, Theorem 8.1] obtained the same result, with a running time $n^{2^{\mathrm{poly}(1/\epsilon)}}$, while our running time is $n^{\mathrm{poly}(1/\epsilon)}$.

Moreover, we can easily extend our algorithm to generalizations of the knapsack problem if the corresponding exact version has a pseudopolynomial time algorithm. For example, we can get the same result for the partial-ordered knapsack problem with tree constraints [25, 58]. In this problem, items must be chosen in accordance with specified precedence constraints and these precedence constraints form a partial order and the underlining undirected graph is a tree (or forest). A pseudopolynomial algorithm for this problem is presented in [58].

Stochastic Knapsack with Random Profits: We are given a set U of n items. Each item i has a deterministic size w_i and a random profit v_i. The goal is to find a subset of items that can be packed into a knapsack with capacity 1 and the probability that the profit is at least a given threshold T is maximized.

Henig [32] and Carraway et al. [15] studied this problem for normally distributed profits and presented dynamic programming and branch and bound heuristics to solve this problem optimally.

We can solve the equivalent problem of minimizing the probability that the profit is at most the given threshold, subject to the capacity constraint. We first show that relaxing the capacity constraint is necessary. Consider the following knapsack instance. The profit of each item is the same as its size. The given threshold is 1. We can see that the optimal probability is 1 if and only if there is a subset of items of total size exactly 1. Otherwise, the optimal probability is 0. Therefore, it is NP-hard to approximate the original problem within any additive error less than 1 without violating the capacity constraint.

The corresponding exact version of the deterministic problem is to find a set of items S such that $w(S) \leq 1$ and $v(S)$ is equal to a given target value. In fact, there is no pseudopolynomial time algorithm for this problem. Since otherwise we can get an ϵ additive approximation without violating the capacity constraint, contradicting the lower bound argument. Note that a pseudopolynomial time algorithm here should run in time polynomial in the profit value (not the size). However, if the sizes can be encoded in $O(\log n)$ bits (we only have a polynomial number of different sizes), we can solve the problem in time polynomial in n and the largest profit value by standard dynamic programming.

For general sizes, we can round the size of each item down to the nearest multiple of $\frac{\delta}{n}$. Then, we can solve the exact version in pseudopolynomial time $\mathrm{poly}(\max_i v_i, n, 1/\delta)$ by dynamic programming. It is easy to show that for any subset of items, its total size is at most the total rounded size plus δ. Therefore, the total size of our solution is at most $1 + \delta$. We summarize the above discussion in the following theorem.

Theorem 8 *If the optimal probability is $\Omega(1)$, we can find in time $(n/\epsilon\delta)^{\mathrm{poly}(1/\epsilon)}$ a subset S of items such that $\Pr(v(S) > (1 - \epsilon)T) \geq (1 - \epsilon)\mathsf{OPT}$ and $w(S) \leq 1 + \delta$, for any constant $\epsilon > 0$.*

8 Extensions

In this section, we discuss some extensions to our basic approximation scheme. We first consider optimizing a constant number of utility functions in Section 8.1. Then, we study the problem where the weight of each element is a random vector in Section 8.2.

8.1 Multiple Utility Functions

The problem we study in this section contains a set U of n elements. Each element e has a random weight w_e. We are also given d utility functions $\mu_1, \ldots, \mu_d$ and d positive numbers $\lambda_1, \ldots, \lambda_d$. We assume d is a constant. A feasible solution consists of d subsets of elements that satisfy some property. Our objective is to find a feasible solution $S_1, \ldots, S_d$ such that $\mathbb{E}[\mu_i(w(S_i))] \geq \lambda_i$ for all $1 \leq i \leq d$.

We can easily extend our basic approximation scheme to the multiple utility functions case as follows. We decompose these utility functions into short exponential sums using ESUM-APPROX as before. Then, for each utility function, we maintain $(n/\epsilon)^{O(L)}$ configurations. Therefore, we have $(n/\epsilon)^{O(dL)}$ configurations in total and we would like to compute the values for these configurations. We denote the deterministic version of the problem under consideration by $\mathfrak{A}$. The exact version of $\mathfrak{A}$ asks for a feasible solution $S_1, \ldots, S_d$ such that the total weight of S_i is exactly the given number t_i for all i. Following an argument similar to Lemma 2, we can easily get the following generalization of Theorem 1.

Theorem 9 *Assume that there is a pseudopolynomial algorithm for the exact version of $\mathfrak{A}$. Further assume that given any $\epsilon > 0$, we can ϵ-approximate each utility function by an exponential sum with at most L*

terms. Then, there is an algorithm that runs in time $(n/\epsilon)^{O(dL)}$ and finds a feasible solution $S_1, \ldots, S_d$ such that $\mathbb{E}[\mu_i(w(S_i)) \geq \lambda_i - \epsilon$ for $1 \leq i \leq d$, if there is a feasible solution for the original problem.

Now let us consider two simple applications of the above theorem.

Stochastic Multiple Knapsack: In this problem we are given a set U of n items, d knapsacks with capacity 1, and d constants $0 \leq \gamma_i \leq 1$. We assume d is a constant. Each item i has a random size w_i and a deterministic profit v_i. Our objective is to find d disjoint subsets $S_1, \ldots, S_d$ such that $\Pr(w(S_i) \leq 1) \geq \gamma_i$ for all $1 \leq i \leq d$ and $\sum_{i=1}^{d} v(S_i)$ is maximized. The exact version of the problem is to find a packing such that the load of each knapsack i is *exactly* the given value t_i. It is not hard to show this problem can be solved in pseudopolynomial time by standard dynamic programming. If the profits are general integers, we also need the scaling technique as in stochastic knapsack with random sizes. In sum, we can get the following generalization of Theorem 7.

Theorem 10 *For any constants $d \in \mathbb{N}$, $\epsilon > 0$ and $0 \leq \gamma_i \leq 1$ for $1 \leq i \leq d$, there is a polynomial time algorithm to compute d disjoint subsets $S_1, \ldots, S_d$ such that the total profit $\sum_{i=1}^{d} v(S_i)$ is within a $1 - \epsilon$ factor of the optimum and $\Pr(w(S_i) \leq 1 + \epsilon) \geq (1 - \epsilon)\gamma_i$ for $1 \leq i \leq d$.*

Stochastic Multidimensional Knapsack: In this problem we are given a set U of n items and a constant $0 \leq \gamma \leq 1$. Each item i has a deterministic profit v_i and a random size which is a random d-dimensional vector $\mathbf{w}_i = \{w_{i1}, \ldots, w_{id}\}$. We assume d is a constant. Our objective is to find a subset S of items such that $\Pr(\bigwedge_{j=1}^{d}(\sum_{i \in S} w_{ij} \leq 1)) \geq \gamma$ and total profit is maximized. This problem can be also thought as the fixed set version of the stochastic packing problem considered in [21, 13]. We first assume the components of each size vector are independent. The correlated case will be addressed in the next subsection.

For ease of presentation, we assume $d = 2$ from now on. Extension to general constant d is straightforward. We can solve the problem by casting it into a multiple utility problem as follows. For each item i, we create two copies i_1 and i_2. The copy i_j has a random weight w_{ij}. A feasible solution consists of two sets S_1 and S_2 such that S_1 (S_2) only contains the first (second) copies of the elements and S_1 and S_2 correspond to exactly the same subset of original elements. We enumerate all such pairs (γ_1, γ_2) such that $\gamma_1 \gamma_2 \geq \gamma$ and $\gamma_i \in [\gamma, 1]$ is a power of $1 - \epsilon$ for $i = 1, 2$. Clearly, there are a polynomial number of such pairs. For each pair (γ_1, γ_2), we solve the following problem: find a feasible solution S_1, S_2 such that $\Pr(\sum_{i \in S_j} w_{ij} \leq 1) \geq \gamma_j$ for all $j = 1, 2$ and total profit is maximized. Using the scaling technique and Theorem 9 for optimizing multiple utility functions, we can get a $(1 - \epsilon)$-approximation for the optimal profit and $\Pr(\bigwedge_{j=1}^{2}(\sum_{i \in S_j} w_{ij} \leq 1)) = \prod_{j=1}^{2} \Pr(\sum_{i \in S_j} w_{ij} \leq 1) \geq (1 - O(\epsilon))\gamma_1 \gamma_2 \geq (1 - O(\epsilon))\gamma$.

We note that the same result for independent components can be also obtained by using the discretization technique developed for the adaptive version of the problem in [13] [12]. If the components of each size vector are correlated, we can not decompose the problem into two 1-dimensional utilities as in the independent case. Now, we introduce a new technique to handle the correlated case.

8.2 Multidimensional Weight

The general problem we study contains a set U of n elements. Each element e has a random weight vector $w_i = (w_{i1}, \ldots, w_{id})$. We assume d is a constant. We are also given a utility function $\mu : \mathbb{R}^d \rightarrow \mathbb{R}^+$. A feasible solution is a subset of elements satisfying some property. We use $w(S)$ as a shorthand notation

---

[12]With some changes of the discretization technique, the correlated case can be also handled [12].

for vector $(\sum_{i \in S} w_{i1}, \ldots, \sum_{i \in S} w_{id})$. Our objective is to find a feasible solution S such that $\mathbb{E}[\mu(w(S)]$ is maximized.

From now on, x and k denote d-dimensional vectors and kx (or $k \cdot x$) denotes the inner product of k and x. As before, we assume $\mu(x) \in [0, 1]$ for all $x \geq 0$ and $\lim_{|x| \to +\infty} \mu(x) = 0$, where $|x| = \max(x_1, \ldots, x_d)$, Our algorithm is almost the same as in the one dimension case and we briefly sketch it here. We first notice that expected utilities decompose for exponential utility functions, i.e., $\mathbb{E}[\phi^{k \cdot w(S)}] = \prod_{i \in S} \mathbb{E}[\phi^{k \cdot w_i}]$. Then, we attempt to ϵ-approximate the utility function $\mu(x)$ by a short exponential sum $\sum_{|k| \leq N} c_k \phi_k^{kx}$ (there are $O(N^d)$ terms). If this can be done, $\mathbb{E}[\phi^{k \cdot w(S)}]$ can be approximated by $\sum_{|k| \leq N} c_k \mathbb{E}[\phi^{k \cdot w(S)}]$. Using the same argument as in Theorem 1, we can show that there is a polynomial time algorithm that can find a feasible solution S with $\mathbb{E}[\mu(w(S))] \geq OPT - \epsilon$ for any $\epsilon > 0$, provided that a pseudopolynomial algorithm exists for the exact version of the deterministic problem.

To approximate the utility function $\mu(x)$, we need the multidimensional Fourier series expansion of a function $f : \mathbb{C}^d \to \mathbb{C}$ (assuming f is 2π-periodic in each axis): $f(x) \sim \sum_{k \in \mathbb{Z}^d} c_k e^{ikx}$ where $c_k = \frac{1}{(2\pi)^d} \int_{x \in [-\pi, \pi]^d} f(x) e^{-ikx} \, dx$. The *rectangular partial sum* is defined to be

$$S_N f(x) = \sum_{|k_1| \leq N} \cdots \sum_{|k_d| \leq N} c_k e^{ikx}.$$

It is known that the rectangular partial sum $S_N f(x)$ converges uniformly to $f(x)$ in $[-\pi, \pi]^d$ for many function classes as n tends to infinity. In fact, a generalization of Theorem 6 to $[-\pi, \pi]^d$ also holds [3]: If f satisfies the α-Hölder condition, then

$$|f(x) - (S_N f)(x)| \leq O\Big(\frac{|f|_{C^{0,\alpha}} \ln^d N}{N^\alpha}\Big) \quad \text{for } x \in [-\pi, \pi]^d.$$

Now, we have an algorithm AP that can approximate a function in a bounded domain. It is also straightforward to extend ESUM-APPROX to the multidimensional case. Hence, we can ϵ-approximate μ by a short exponential sum in $[0, +\infty)^d$, thereby proving the multidimensional generalization of Theorem 2. Let us consider an application of our result.

Stochastic Multidimensional Knapsack (Revisited): We consider the case where the components of each weight vector can be correlated. Note that the utility function χ_2 corresponding to this problem is the two dimensional threshold function: $\chi_2(x, y) = 1$ if $x \leq 1$ and $y \leq 1$; $\chi_2(x, y) = 0$ otherwise. As in the one dimensional case, we need to consider a continuous version $\widetilde{\chi}_2$ of χ_2 (see Figure 1(3)). By the result in this section and a generalization of Lemma 12 to higher dimension, we can get the following.

Theorem 11 *For any constants $d \in \mathbb{N}$, $\epsilon > 0$ and $0 \leq \gamma \leq 1$, there is a polynomial time algorithm for finding a set S of items such that the total profit $v(S)$ is $1 - \epsilon$ factor of the optimum and $\Pr(\bigwedge_{j=1}^{d} (\sum_{i \in S} w_{ij} \leq 1 + \epsilon)) \geq (1 - \epsilon)\gamma$.*

9   A Few Remarks

Convergence of Fourier series: The convergence of the Fourier series of a function is a classic topic in harmonic analysis. Whether the Fourier series converges to the given function and the rate of the convergence typically depends on a variety of smoothness condition of the function. We refer the readers to [62] for a more comprehensive treatment of this topic. We note that we could obtain a smoother version of χ (e.g., see

Figure 1(2)), instead of the piecewise linear $\widetilde{\chi}$, and then use Theorem 6 to obtain a better bound for L. This would result in an even better running time. Our choice is simply for the ease of presentation.

Discontinuous utility functions: If the utility function μ is discontinuous, e.g., the threshold function, then the partial Fourier series behaves poorly around the discontinuity (this is known as the *Gibbs phenomenon*). However, informally speaking, as the number of Fourier terms increases, the poorly-behaved strip around the edge becomes narrower. Therefore, if the majority of the probability mass of our solution lies outside the strip, we can still guarantee a good approximation of the expected utility. There are also techniques to reduce the effects of the Gibbs phenomenon (See e.g., [28]). However, the techniques are not sufficient to handle discontinuous functions. We note that very recently, Daskalakis et al. [20] obtained a true additive PTAS (instead of a bi-criterion additive PTAS) for a closely related problem, called *the fault tolerant storage problem*, under certain technical assumptions. [13] It is not clear how to use their technique to obtain a true additive PTAS for our expected utility maximization problem. We leave this problem as an interesting open problem.

10   Conclusion

We consider the problem of maximizing expected utility for many stochastic combinatorial problems, such as shortest path, spanning tree and knapsack. We develop a polynomial time approximation scheme with additive error ϵ for any $\epsilon > 0$. A key ingredient in our algorithm is to decompose the utility function into a short exponential sum. In this paper, we use the Fourier series technique to fulfill the task. Exploring other decomposition approaches is an interesting future work. Our general approximation framework may be useful for other stochastic optimization problems. We leave the problems of obtaining a true additive PTAS, or approximations with reasonable multiplicative factors, or nontrivial inapproximability results, for class $\mathfrak{C}_{\mathsf{bounded}}$, as interesting open problems.

11   Acknowledgments

References

[1] H. Ackermann, A. Newman, H. Röglin, and B. Vöcking. Decision making based on approximate and smoothed pareto curves. *Algorithms and Computation*, pages 675–684, 2005.

[2] S. Agrawal, A. Saberi, and Y. Ye. Stochastic Combinatorial Optimization under Probabilistic Constraints. *Arxiv preprint arXiv:0809.0460*, 2008.

[3] S. Alimov, R. Ashurov, and A. Pulatov. Multiple fourier series and fourier integrals, in commutative harmonic analysis. IV: Harmonic analysis in $\mathbb{R}_n$. *Encyclopedia of Mathematical Science*, 42, 1992.

---

[13]In the fault tolerant storage problem, we are given n real numbers $0 < p_1 \leq \ldots \leq p_n < 1$, and an addition number $0 < \theta < 1$. Our goal is to partition 1 into n positive values $x_1, \ldots, x_n$ (i.e., $\sum_{i=1}^{n} x_i = 1$), such that $\Pr[\sum_{i=1}^{n} X_i \geq \theta]$ is maximized, where X_i is the Bernoulli random variable which takes value x_i with probability p_i. In order to obtained an additive PTAS, Daskalakis et al. [20] assumed that all p_is are bounded below by a constant.

[4] N. Bansal, A. Gupta, J. Li, J. Mestre, V. Nagarajan, and A. Rudra. When LP is the Cure for Your Matching Woes: Improved Bounds for Stochastic Matchings. *European Symposium on Algorithms*, pages 218–229, 2010.

[5] F. Barahona and W. Pulleyblank. Exact arborescences, matchings and cycles. *Discrete Applied Mathematics*, 16(2):91–99, 1987.

[6] J. Bard and J. Bennett. Arc reduction and path preference in stochastic acyclic networks. *Management Science*, 37(2):198–215, 1991.

[7] D. Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica: Journal of the Econometric Society*, pages 23–36, 1954. Originally published in 1738; translated by Dr. Louise Sommer.

[8] D. Bernoulli. Exposition of a new theory on the measurement of risk. *Econometrica*, 22(1):22–36, 1954. Originally published in 1738; translated by Dr. Lousie Sommer.

[9] G. Beylkin and L. Monzón. On Generalized Gaussian Quadratures for Exponentials and Their Applications* 1. *Applied and Computational Harmonic Analysis*, 12(3):332–373, 2002.

[10] G. Beylkin and L. Monzón. On approximation of functions by exponential sums. *Applied and Computational Harmonic Analysis*, 19(1):17–48, 2005.

[11] G. Beylkin and L. Monzón. Approximation by exponential sums revisited. *Applied and Computational Harmonic Analysis*, 28(2):131–149, 2010.

[12] A. Bhalgat, 2011. Personal Communication.

[13] A. Bhalgat, A. Goel, and S. Khanna. Improved approximation results for stochastic knapsack problems. In *ACM-SIAM symposium on Discrete algorithms*, 2011.

[14] A. Bhalgat and S. Khanna. A utility equivalence theorem for concave functions. In *Integer Programming and Combinatorial Optimization*, pages 126–137. Springer, 2014.

[15] R. Carraway, R. Schmidt, and L. Weatherford. An algorithm for maximizing target achievement in the stochastic knapsack problem with normal returns. *Naval research logistics*, 40(2):161–173, 1993.

[16] C. Chekuri and S. Khanna. A PTAS for the multiple knapsack problem. In *ACM-SIAM symposium on Discrete algorithms*, pages 213–222, 2000.

[17] N. Chen, N. Immorlica, A. Karlin, M. Mahdian, and A. Rudra. Approximating matches made in heaven. *International Colloquium on Automata, Languages and Programming*, pages 266–278, 2009.

[18] W. Cheney and W. Light. *A Course in Approximation Theory*. Brook/Cole Publishing Company, 2000.

[19] R. Cheng, J. Chen, and X. Xie. Cleaning uncertain data with quality guarantees. *Proceedings of the VLDB Endowment*, 1(1):722–735, 2008.

[20] C. Daskalakis, A. De, I. Diakonikolas, A. Moitra, and R. A. Servedio. A polynomial-time approximation scheme for fault-tolerant distributed storage. In *SODA*, pages 628–644. SIAM, 2014.

[21] B. Dean, M. Goemans, and J. Vondrák. Adaptivity and approximation for stochastic packing problems. In *ACM-SIAM symposium on Discrete algorithms*, pages 395–404, 2005.

[22] B. Dean, M. Goemans, and J. Vondrak. Approximating the Stochastic Knapsack Problem: The Benefit of Adaptivity. *Mathematics of Operations Research*, 33(4):945, 2008.

[23] M. Fazel and M. Chiang. Network utility maximization with nonconcave utilities using sum-of-squares method. In *Decision and Control, 2005 and 2005 European Control Conference. CDC-ECC'05. 44th IEEE Conference on*, pages 1867–1874. IEEE, 2005.

[24] P. Fishburn. *Utility Theory and Decision Making*. John Wiley & Sons, Inc, 1970.

[25] M. Garey and D. Johnson. *"Computers and Intractability: A Guide to the Theory of NP-Completeness"*. W.H. Freeman, 1979.

[26] S. Geetha and K. Nair. On stochastic spanning tree problem. *Networks*, 23(8):675–679, 1993.

[27] A. Goel and P. Indyk. Stochastic load balancing and related problems. In *Annual Symposium on Foundations of Computer Science*, page 579, 1999.

[28] D. Gottlieb and C. Shu. On the Gibbs phenomenon and its resolution. *SIAM review*, 39(4):644–668, 1997.

[29] V. Goyal and R. Ravi. Chance constrained knapsack problem with random item sizes. *To appear in Operation Research Letter*, 2009.

[30] S. Guha and K. Munagala. Adaptive Uncertainty Resolution in Bayesian Combinatorial Optimization Problems. *To appear in ACM Transactions on Algorithms*, 2008.

[31] A. Gupta, M. Pál, R. Ravi, and A. Sinha. Boosted sampling: approximation algorithms for stochastic optimization. In *ACM Symposium on Theory of Computing*, pages 417–426. ACM, 2004.

[32] M. Henig. Risk criteria in a stochastic knapsack problem. *Operations Research*, 38(5):820–825, 1990.

[33] L. Huang and J. Li. Approximating the expected values for combinatorial optimization problems over stochastic points. In *Automata, Languages, and Programming*, pages 910–921. Springer, 2015.

[34] L. Huang, J. Li, J. M. Phillips, and H. Wang. ϵ-kernel coresets for stochastic points. *arXiv preprint arXiv:1411.0194*, 2014.

[35] H. Ishii, S. Shiode, and T. Nishida Yoshikazu. Stochastic spanning tree problem. *Discrete Applied Mathematics*, 3(4):263–273, 1981.

[36] D. Kahneman and A. Tversky. Prospect theory: An analysis of decision under risk. *Econometrica: Journal of the Econometric Society*, pages 263–291, 1979.

[37] O. Kariv and S. Hakimi. An algorithmic approach to network location problems. II: The p-medians. *SIAM Journal on Applied Mathematics*, 37(3):539–560, 1979.

[38] J. Kleinberg, Y. Rabani, and É. Tardos. Allocating bandwidth for bursty connections. In *ACM Symposium on Theory of Computing*, page 673, 1997.

[39] J. Li and A. Deshpande. Consensus answers for queries over probabilistic databases. In *ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems*, 2009.

[40] J. Li and A. Deshpande. Ranking continuous probabilistic datasets. *Proceedings of the VLDB Endowment*, 3(1), 2010.

[41] J. Li, B. Saha, and A. Deshpande. A unified approach to ranking in probabilistic databases. In *Proceedings of the VLDB Endowment*, 2009.

[42] J. Li and T. Shi. A fully polynomial-time approximation scheme for approximating a sum of random variables. *Operations Research Letters*, 42(3):197–202, 2014.

[43] J. Li and W. Yuan. Stochastic combinatorial optimization via poisson approximation. In *ACM Symposium on Theory of Computing*, 2013.

[44] R. Loui. Optimal paths in graphs with stochastic or multidimensional weights. *Communications of the ACM*, 26(9):670–676, 1983.

[45] R. Martin. The St. Petersburg Paradox. *The Stanford Encyclopedia of Philosophy*, 2004. `http://plato.stanford.edu/archives/fall2004/entries/paradox-stpetersburg`.

[46] S. Mittal and A. Schulz. A general framework for designing approximation schemes for combinatorial optimization problems with many objectives combined into one. *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 179–192, 2008.

[47] A. Munteanu, C. Sohler, and D. Feldman. Smallest enclosing ball for probabilistic data. In *Proceedings of the thirtieth annual symposium on Computational geometry*, page 214. ACM, 2014.

[48] I. Murthy and S. Sarkar. Exact algorithms for the stochastic shortest path problem with a decreasing deadline utility function. *European Journal of Operational Research*, 103(1):209–229, 1997.

[49] I. Murthy and S. Sarkar. Stochastic shortest path problems with piecewise-linear concave utility functions. *Management Science*, 44(11):125–136, 1998.

[50] E. Nikolova. Approximation Algorithms for Reliable Stochastic Combinatorial Optimization. *International Workshop on Approximation Algorithms for Combinatorial Optimization Problems*, pages 338–351, 2010.

[51] E. Nikolova, M. Brand, and D. Karger. Optimal route planning under uncertainty. In *Proceedings of International Conference on Automated Planning and Scheduling*, 2006.

[52] E. Nikolova, J. Kelner, M. Brand, and M. Mitzenmacher. Stochastic shortest paths via quasi-convex maximization. In *European Symposium on Algorithms*, pages 552–563, 2006.

[53] F. Oberhettinger. *Fourier transforms of distributions and their inverses: a collection of tables*. Academic press, 1973.

[54] M. R. Osborne and G. K. Smyth. A modified prony algorithm for fitting sums of exponential functions. *SIAM Journal of Scientific Computing*, 1995.

[55] C. Papadimitriou and M. Yannakakis. On the approximability of trade-offs and optimal access of web sources. In *Annual Symposium on Foundations of Computer Science*, 2000.

[56] M. J. D. Powell. *Approximation theory and methods*. Cambridge University Press, 1981.

[57] A. Ralston and R. Rabinowitz. *A First Course in Numerical Analysis*. 2001.

[58] H. Safer, J. B. Orlin, and M. Dror. Fully polynomial approximation in multi-criteria combinatorial optimization, 2004. MIT Working Paper.

[59] P. A. Samuelson. St. petersburg paradoxes: Defanged, dissected, and historically described. *Journal of Economic Literature*, 15(1):24–55, 1977.

[60] D. Shmoys and C. Swamy. An approximation scheme for stochastic linear programming and its application to stochastic integer programs. *J. ACM*, 53(6):1012, 2006.

[61] C. Sigal, A. Pritsker, and J. Solberg. The stochastic shortest route problem. *Operations Research*, 28(5):1122–1129, 1980.

[62] E. Stein and R. Shakarchi. *Fourier analysis: an introduction*. Princeton University Press, 2003.

[63] C. Swamy. Risk-Averse Stochastic Optimization: Probabilistically-Constrained Models and Algorithms for Black-Box Distributions. *ACM-SIAM symposium on Discrete algorithms*, 2010.

[64] C. Swamy and D. Shmoys. Approximation algorithms for 2-stage stochastic optimization problems. *ACM SIGACT News*, 37(1):46, 2006.

[65] J. von Neumann and O. Morgenstern. *Theory of Games and Economic Behavior*. Princeton Univ. Press, 2nd edition, 1947.

A Computing $\mathbb{E}[\phi^{w_e}]$

If X is a random variable, then the *characteristic function* of X is defined as

$$G(z) = \mathbb{E}[e^{izX}].$$

We can see $\mathbb{E}[\phi^{w_e}]$ is nothing but the value of the characteristic function of w_e evaluated at $-i \ln \phi$ (here $\ln$ is the complex logarithm function). For many important distributions, including negative binomial, Poisson, exponential, Gaussian, Chi-square and Gamma, a closed-form characteristic function is known. See [53] for a more comprehensive list.

Example 3 *Consider the Poisson distributed w_e with mean λ, i.e., $\Pr(w_e = k) = \lambda^k e^{-\lambda}/k!$. Its characteristic function is known to be $G(z) = e^{\lambda(e^{iz}-1)}$. Therefore,*

$$\mathbb{E}[\phi^{w_e}] = G(-i \ln \phi) = e^{\lambda(\phi-1)}.$$

Example 4 *For Gaussian distribution $N(\mu, \sigma^2)$, we know its characteristic function is $G(z) = e^{iz\mu - \frac{1}{2}\sigma^2 z^2}$. Therefore,*

$$\mathbb{E}[\phi^{w_e}] = G(-i \ln \phi) = \phi^{u + \frac{1}{2}\sigma^2 \ln \phi}.$$

For some continuous distributions, no closed-form characteristic function is known and we need proper numerical approximation method.

If the support of the distribution is bounded, we can use for example Gauss-Legendre quadrature [57]. If the support is infinite, we can truncate the distribution and approximate the integral over the remaining finite interval; Generally speaking a quadrature method approximates $\int_a^b f(x)\mathrm{d}x$ by a linear sum $\sum_{i=1}^k c_i f(x_i)$ where c_i and x_i are some constants independent of the function f. A typical practice is to use *composite rule*, that is to partition $[a,b]$ into N subintervals and approximate the integral using some quadrature formula over each subinterval. For the example of Gauss-Laguerre quadrature, assuming continuity of the $2k$th derivative of $f(x)$ for some constant k, if we partition $[a,b]$ into M subintervals and apply Gauss-Legendre quadrature of degree k to each subinterval, the approximation error is

$$\mathsf{Error} = \frac{(b-a)^{2k+1}}{M^{2k}} \frac{(k!)^4}{(2k+1)[(2k)!]^3} f^{(2k)}(\xi)$$

where ξ is some point in (a,b) [57, pp.116]. Let $\Delta = \frac{b-a}{M}$. If we treat k as a constant, the behavior of the error (in terms of Δ) is $\mathsf{Error}(\Delta) = O(\Delta^{2k} \max_\xi f^{(2k)}(\xi))$. Therefore, if the support and $\max_\xi f^{(2k)}(\xi)$ are bounded by a polynomial, we can approximate the integral, in polynomial time, such that the error is $O(1/n^\beta)$ for any fixed integer β.

The next lemma shows that we do not lose too much even though we can only get an approximation of $\mathbb{E}[\phi^{w_e}]$.

Lemma 13 *Suppose in Theorem 5, we can only compute an approximate value of $\mathbb{E}[\phi_i^{w_e}]$, denoted by $E_{e,i}$, for each e and i, such that $|\mathbb{E}[\phi_i^{w_e}] - E_{e,i}| \leq O(n^{-\beta})$ for some positive integer β. Denote $E(S) = \sum_{k=1}^L c_k \prod_{e \in S} E_{e,i}$. For any solution S, we have that*

$$|\mathbb{E}[\widetilde{\mu}(w(S))] - E(S)| \leq O(n^{1-\beta}).$$

Proof: We need the following simple result (see [40] for a proof): $a_1, \ldots, a_n$ and $e_1, \ldots, e_n$ are complex numbers such that $|a_i| \leq 1$ and $|e_i| \leq n^{-\beta}$ for all i and some $\beta > 1$. Then, we have

$$\left| \prod_{i=1}^n (a_i + e_i) - \prod_{i=1}^n E_i \right| \leq O(n^{1-\beta}).$$

Since $|\phi_i| \leq 1$, we can see that

$$|\mathbb{E}[\phi_i^{w_e}]| = |\int_{x \geq 0} \phi_i^x p_e(x)\mathrm{d}x| \leq 1.$$

The lemma simply follows by applying the above result and noticing that L and all c_ks are constants. $\square$

We can show that Theorem 1 still holds even though we only have the approximations of the $\mathbb{E}[\phi^{w_e}]$ values. The proof is straightforward and omitted.