Approximation Algorithms for Reliable Stochastic Combinatorial Optimization

Evdokia Nikolova*
Massachusetts Institute of Technology
nikolova @mit.edu

September 17, 2010

Abstract

We consider optimization problems that can be formulated as minimizing the cost of a feasible solution $\mathbf{w}^T\mathbf{x}$ over an arbitrary combinatorial feasible set $\mathcal{F} \subset \{0,1\}^n$. For these problems we describe a broad class of corresponding stochastic problems where the cost vector $\mathbf{W}$ has independent random components, unknown at the time of solution. A natural and important objective that incorporates risk in this stochastic setting is to look for a feasible solution whose stochastic cost has a small tail or a small convex combination of mean and standard deviation. Our models can be equivalently reformulated as nonconvex programs for which no efficient algorithms are known. In this paper, we make progress on these hard problems.

Our results are several efficient general-purpose approximation schemes. They use as a black-box (exact or approximate) the solution to the underlying deterministic problem and thus immediately apply to arbitrary combinatorial problems. For example, from an available δ-approximation algorithm to the linear problem, we construct a $\delta(1+\epsilon)$-approximation algorithm for the stochastic problem, which invokes the linear algorithm only a logarithmic number of times in the problem input (and polynomial in $\frac{1}{\epsilon}$), for any desired accuracy level $\epsilon > 0$. The algorithms are based on a geometric analysis of the curvature and approximability of the nonlinear level sets of the objective functions.

Key words: approximation algorithms, reliable optimization, stochastic optimization, risk, mean-risk, nonlinear programming, nonconvex optimization

1 Introduction

In this paper, we consider generic combinatorial problems and ask what happens when their associated costs are stochastic. The most common approaches in stochastic optimization are to find the solution of minimum expected cost. However, in many applications reliability considerations are very important: risk-averse users need reassurance regarding the level of risk, and not just the expected cost of the provided solution. For example, the transportation community has recognized the importance of reliable route plans (*e.g.,* [7, 28, 25, 37, 9]), however the solutions offered are typically inefficient or heuristic with unknown approximation guarantee. Similarly, reliability is a key consideration in finance and other *continuous* optimization settings [34]. It has been noted that incorporating reliability [34, 29] transforms the problems into nonconvex ones for which there are no known efficient algorithms and rigorous approximative analysis is scarce. In this paper, we provide a rigorous treatment of reliable combinatorial optimization, offering fully-polynomial approximation schemes for a rich framework of reliability measures.

*This work was supported in part by the National Science Foundation under grant 0931550.

To illustrate our framework, consider an application such as driving to the airport in uncertain traffic. Our goal is to find a route that gets us to the airport on time. Clearly, the route which minimizes our expected travel time may not be an appropriate choice. In fact, the natural objectives may vary depending on when we are submitting the route query: ahead of time, when we are debating how much time to budget for our trip, or at the start of our trip, when we are optimizing our chance of ontime arrival. In the former setting, we would typically want to allocate enough time to ensure some confidence of ontime arrival, say 95%. In the latter, given a deadline to reach our destination, we need to find the route which will most likely reach by the deadline. Another natural objective, used for example by the Federal Highway Administration as a travel time reliability criterion, is given by the mean plus standard deviation of a route [10]. The latter reliability criterion has been considered in the context of stochastic minimum spanning treess as well [2], and this model is sometimes referred to as mean-risk optimization (*e.g.,* [2]).

We thus focus on a general framework for reliable stochastic combinatorial optimization, which includes the following problem settings:

1. minimize $(mean + c \cdot standard\, deviation)$ for a non-negative constant c which parametrizes the level of risk-aversion. [Call this the *Mean-risk model* or objective.]

2. maximize $\Pr(solution\ cost \leq budget)$ for a given *budget*. [*Probability tail model* / objective.]

3. minimize $budget$ such that $\Pr(solution\ cost \leq budget) \geq p$ for a given confidence probability p. [*Value-at-risk model.*]

In contrast with the diversity in model specifications above, we will show that the same approximation algorithm design can simultaneously address all. Throughout, we assume that the cost distributions are independent, although our algorithms also extend to the case of correlations of neighboring edges for example in shortest path problems (the graph with correlated edges is transformed into a slightly larger graph with independent edges and thus all our results here immediately carry through.)

Contributions. We start our discussion with the (relatively) simpler mean-risk model, which is equivalent to minimizing $\left(mean + c \cdot \sqrt{variance}\right)$. We provide strong results that apply to *arbitrary* cost distributions with given means and variances, and achieve essentially the same approximation factor as what is possible for the underlying deterministic problem. In particular, we provide general-purpose algorithms that use as a black-box an algorithm for the deterministic problem. We summarize our results for this setting below:

Theorem (See Theorems 1, 5)**.** *There is a fully-polynomial approximation scheme for the mean-risk stochastic model, when there is an exact or fully-polynomial approximation algorithm for the underlying deterministic problem.*

In addition, there is a $(1+\epsilon)\delta$*-approximation for the stochastic model running in time polynomial in* $\frac{1}{\epsilon}$*, when there is an available* δ*-approximation for the deterministic problem.*

A rigorous approximation-algorithmic analysis of the second and third models in the framework, which involve optimization of the probability tails, necessitates an assumption on the distribution: in the absence of any knowledge on the distributions, the best one can do is bound the tails, for example using Chernoff or Chebyshev bounds, and optimize those tail bounds instead—this will yield a conservative overestimate of the probability of exceeding the budget.

We provide strict approximation results under the commonly assumed Gaussian distributions; we then show how the same algorithmic techniques can apply to arbitrary distributions using tail bounds. In the former setting, minimizing the probability tail in the second model is equivalent to maximizing $\frac{budget - mean}{\sqrt{variance}}$ and we get the following approximations:

Theorem (See Theorems 1, 6)**.** *There is a fully-polynomial approximation scheme for the probability tail model, when there is an exact or fully-polynomial approximation algorithm for the underlying deterministic problem.*

In addition, when there is an available δ-approximation for the deterministic problem, there is a $\sqrt{1 - \left[\frac{\delta-(1-\epsilon^2/4)}{(2+\epsilon)\epsilon/4}\right]}$*-approximation for the stochastic model running in time polynomial in* $\frac{1}{\epsilon}$.

We remark that the above algorithms find the approximate solution, assuming there is a feasible solution with expected cost at most the budget, or $(1-\epsilon)$ times the budget in the exact and approximate deterministic settings respectively (in other words, the probability of exceeding the budget is at most $\frac{1}{2}$). Otherwise, if a given budget is so small that the probability of exceeding it is greater than $\frac{1}{2}$, we are in a risk-loving, rather than a risk-averse situation, which would be similar to minimizing a $(mean - standard\, deviation)$-type objective in model (1). In other words, we would prefer solutions with higher variances (for example, looking for longest paths).

The third (value-at-risk) model under Gaussian distributions is equivalent to the mean-risk model, with risk-aversion coefficient $c = \Phi^{-1}(p)$, where $\Phi^{-1}(\cdot)$ is the inverse cumulative distribution function of the standard normal $N(0, 1)$.

For arbitrary distributions, the third model again reduces to the mean-risk model, but with a more conservative risk-aversion coefficient $c = \sqrt{\frac{p}{1-p}}$, as a result of which our algorithms provide an overestimate of the true error probability of exceeding the budget. Optimizing a tail bound in the second model similarly provides an overestimate of the true probability, which is again the best one can hope to achieve in the absence of other distributional information.

Background and Challenges. Our algorithms build on the fact that the model formulations in our framework are all instances of concave minimization, for which it is known that the optimal solution is attained at an extreme point of the feasible set (see, *e.g.,* [4]). In particular, our objective functions depend only on the means and variances of feasible solutions. Thus, we can project the feasible set on the plane spanned by the mean and variance vectors and only consider extreme points on the projection (see Figure 1(a)). This greatly restricts the number of relevant extreme points. For example, in the case of minimum spanning trees and matroids there are only polynomially many such extreme points, which can be efficiently enumerated, hence the corresponding reliable spanning trees and matroids in a stochastic environment can be found with a straightforward polynomial-time algorithm. However, an arbitrary combinatorial problem would most likely have too many extreme points even on a two-dimensional projection (for example, shortest paths have $n^{\log n}$ such points [30]), hence our focus on approximation in this paper.

We can geometrically visualize the objective function in terms of its level sets on the mean-variance plane. These form parabolas, corresponding to higher objective function values at greater mean and variance values. The optimal solution is obtained at the lowest parabola touching the projected feasible set. Figure 1(a) depicts these parabolas and the challenge that arises with concave minimization problems: along the convex hull boundary of the feasible set, the objective function fluctuates and, in particular, many extreme points may be local optima and thus local search algorithms would fail to find a good approximation. What we do instead is follow the objective function levels to guide us into the relevant portion of the feasible set, as explained below.

Overview of Algorithms and Techniques. [For the case of easy deterministic problems.] The algorithm constructs a (non-linear) separation oracle for telling us whether, for a given function level set,[1] there is a feasible solution below the level set (with value less than the given function value) or else, whether the entire

[1]The level set of a function f for value λ is the subset of the domain on which the function equals λ, $L_\lambda = \{\mathbf{x} \mid f(\mathbf{x}) = \lambda\}$.

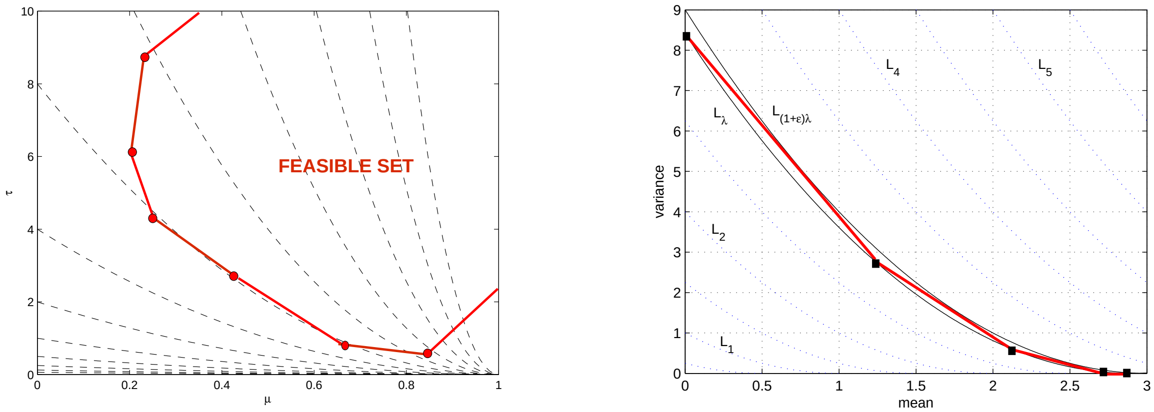

Figure 1: *(a)* Level sets of the probability tail objective function and the convex hull of the projected feasible set on the mean-variance plane. *(b)* Level sets and approximate separation oracle for the mean-risk objective on the mean-variance plane.

feasible set is above the given level. Afterwords, a binary search on the optimum objective function value combined with the separation oracle finds the desired approximate solution.

The separation oracle approximates a given level set curve by inscribing a (partial) polygon in it. Each side of the polygon induces a linear objective over the feasible set, which we minimize via a black-box call to the algorithm for the deterministic problem. If the resulting solution is below the current level set (more precisely, its associated original objective function value is smaller than $(1+\epsilon)$ times the given level), the separation oracle returns that solution. Else, if after minimizing with respect to all linear segments, we do not find any solutions below the level set, the separation oracle returns a negative answer that the entire feasible set is above it.

The subtlety arises in how to construct the polygonal segments to ensure a good and efficient approximation. To get an efficient algorithm, we need to approximate the level set curves with as few linear segments as possible. On the other hand, to get a good approximation factor, we need a finer polygon (with more and smaller sides), which is sandwiched between the desired level set with function value λ and the level set with function value $\lambda(1+\epsilon)$ (See Figure 1(b)). In particular, in the worst case when the level sets touch, as is the case for the probability tail objective, a polygon sandwiched between the two level sets will have infinitely many sides. We resolve this problem by carefully bounding the optimal solution so that we do not need all infinitely many linear segments from the polygon, and we prove that it suffices to consider only polynomially many such segments.

[Hard deterministic problems.] We could use the same algorithm design as above, by appropriately modifying its analysis and approximation factors, when we have a δ-approximation rather than an exact algorithm for solving the underlying deterministic problem. It turns out that for this case, a cruder and simpler algorithm gives the same approximation factor. In particular, all we need to do here is apply the algorithm for the deterministic problem on a small sequence of linear cost functions of the form $mean + k \cdot variance$, for a geometric progression of coefficients k.

However, even if we know what single choice of k would find the optimal solution, the difficulty is to translate the approximation given by the deterministic black-box algorithm for the *linear function* into an approximation for the *original concave function*: the two functions have nothing in common (except that the former is a gradient of the latter at some point), and a priori it is not clear that an approximation of the former would at all yield a meaningful approximation factor for the original objective. Fortunately, all objective functions in our framework admit such an approximation (the probability tail objective is again more challenging due to the given budget and requires us to know that there is a feasible solution at least a small distance away from the budget).

Related Work. A rich body of work in stochastic combinatorial optimization focuses on two-stage and multistage optimization (*e.g.*, [36, 17, 21, 16, 18]). The models there typically look for solutions of minimum expected cost, and Swamy and Shmoys remark that "it would be interesting to explore stochastic models that incorporate risk" [39]. There are models that incorporate additional budget constraints [38] or threshold constraints for specific problems such as knapsack, load balancing and others [8, 13, 23].

At the other end of the spectrum is the paradigm of robust optimization (see survey [5]), which provides completely reliable (robust) solutions, though this is only possible when the uncertainty is bounded, namely the random variables have bounded support. Our framework for reliable optimization falls between stochastic optimization, which minimizes expected cost, and robust optimization, which minimizes the maximum cost. Interestingly, part of our framework (the mean-risk model) arises in robust discrete optimization under ellipsoidal uncertainty sets [6]. Bertsimas and Sim offer for it pseudopolynomial algorithms, assuming that the underlying deterministic problem can be solved exactly, in contrast with our fully polynomial approximation schemes that work with both exact and approximate algorithms for the deterministic problem.

Atamtürk and Narayanan [2] also consider mean-risk minimization in discrete optimization, giving a characterization in terms of submodular minimization. Our feasible set is an arbitrary subset of the hypercube vertices, on which it is not known how to do submodular minimization. As a curiosity, we mention here that the mean-risk objective is also supermodular via the Lovász extension [24]. However, supermodular minimization is even harder and this perspective does not help our problem at hand.

The probability tail objective was previously considered in the special context of stochastic shortest paths and an exact algorithm was given based on enumerating relevant extreme points from the path polytope [30]. The same type of algorithm extends to arbitrary combinatorial problems and its complexity is polynomial for minimum spanning trees and matroids. However, in general, it is superpolynomial or exponential, hence our focus on approximation algorithms in this paper.

A comprehensive survey of models that incorporate risk in *continuous* settings is provided by Rockafellar [34]. The solution concepts and continuous nature of the problems make this work very different from ours. Similarly, continuous optimization work with probability (chance) constraints (*e.g.*, [29]) applies for linear and *not* discrete optimization problems. Additional related work on the combinatorial optimization side includes research on multi-criteria optimization (*e.g.*, [32, 1, 35, 40]) and combinatorial optimization with a ratio of linear objectives [27, 33]. Our models can also be seen as instances of concave discrete minimization; however, the existing work in this area requires assumptions that do not hold in our framework, such as restrictive properties on the feasible set, strictly positive range of the objective function, or boundedness/positivity of the objective function gradient [31, 3, 22, 14].

2 An FPTAS for the reliable versions of easy combinatorial problems

In this section, we formally define the models in our reliable stochastic optimization framework and present a general-purpose FPTAS design for these problems. The FPTAS uses as a black-box an exact algorithm for the underlying deterministic problem and is based on a geometric analysis of the curvature and approximability of the level sets of the objective functions.

Suppose we have an arbitrary combinatorial set of feasible solutions $\mathcal{F} \subset \{0,1\}^n$, together with an oracle for optimizing linear objectives over the set. In addition, we are given nonnegative vectors of means $\boldsymbol{\mu} \in \mathbb{R}^n$ and variances $\boldsymbol{\tau} \in \mathbb{R}^n$ for the stochastic cost vector $\mathbf{W}$, coming from independent distributions so that the mean and variance of a solution $\mathbf{x} \in \mathcal{F}$ is $\boldsymbol{\mu}^T\mathbf{x}$ and $\boldsymbol{\tau}^T\mathbf{x} \geq 0$ respectively. We are interested in finding a feasible solution with optimal cost, where the notion of optimality incorporates risk.

1. *[Mean-risk model]* A family of objectives that has been analyzed in continuous optimization settings, mostly in the context of finance [11, 26], is the family of convex combinations of mean and standard deviation. Formally, this problem is to:

$$\begin{aligned} \text{minimize} \quad & \boldsymbol{\mu}^T\mathbf{x} + c\sqrt{\boldsymbol{\tau}^T\mathbf{x}} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{F}, \end{aligned} \tag{1}$$

where the constant c parametrizes the degree of the user's risk aversion.

2. *[Probability tail model]* An alternative natural model maximizes the probability that the stochastic solution cost is within a desired budget or threshold t: maximize $\Pr\left(\mathbf{W}^T\mathbf{x} \le t\right)$ subject to $\mathbf{x} \in \mathcal{F}$. When the stochastic costs $\mathbf{W}$ are Gaussian, subtracting the mean and dividing by the standard deviation transforms the problem into the following equivalent formulation (which is also approximation-preserving as we show in the extended version):

$$\begin{aligned} \text{maximize} \quad & \frac{t - \boldsymbol{\mu}^T\mathbf{x}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}}} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{F}. \end{aligned} \tag{2}$$

When the stochastic costs $\mathbf{W}$ come from arbitrary distributions, the maximum probability is lower-bounded by $\frac{(t-\boldsymbol{\mu}^T\mathbf{x})^2}{(t-\boldsymbol{\mu}^T\mathbf{x})^2+(\boldsymbol{\tau}^T\mathbf{x})}$ (by the one-sided Chebyshev bound, also known as Cantelli's inequality [15], $\Pr(X \le E[X] + k\sqrt{Var(X)}) \ge 1 - \frac{1}{1+k^2}$, with $k = \frac{t-\boldsymbol{\mu}^T\mathbf{x}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}}}$). While maximizing a lower-bound will not yield a strict approximation of the probability tail objective, it is the best one can achieve in the absence of other distributional information—and our techniques can strictly approximate this bound as well:

$$\begin{aligned} \text{maximize} \quad & \frac{(t - \boldsymbol{\mu}^T\mathbf{x})^2}{(t - \boldsymbol{\mu}^T\mathbf{x})^2 + \boldsymbol{\tau}^T\mathbf{x}} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{F}. \end{aligned} \tag{3}$$

3. *[Value-at-risk model]* Finally, we may wish to minimize the budget t such that the probability of not exceeding it is at least a given confidence level p:

$$\begin{aligned} \text{minimize} \quad & t \\ \text{subject to} \quad & \Pr(\mathbf{W}^T\mathbf{x} \le t) \ge p \\ & \mathbf{x} \in \mathcal{F}. \end{aligned} \tag{4}$$

Depending on whether we have Gaussian or arbitrary distributions, this problem is exactly equivalent to, or its solution can be upper-bounded using Chebyshev's bound by the mean-risk model (1) with $c = \Phi^{-1}(p)$ or $c = \sqrt{\frac{p}{1-p}}$ (See Ghaoui *et al.* [12]; more details are provided in the extended version of this paper).

We can obtain fully-polynomial approximation schemes (FPTAS) for all models above, with the same FPTAS template, which we explain below. All models are instances of concave minimization (equivalently, convex maximization) over $\mathbf{x} \in \mathcal{F}$. Our algorithms make black-box calls to an exact algorithm (sometimes referred to as the *linear oracle*) for solving the underlying deterministic (linear) problem:

$$\begin{aligned} \text{minimize} \quad & \mathbf{w}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{x} \in \mathcal{F}, \end{aligned} \tag{5}$$

for a carefully chosen *small* set of linear objectives $\mathbf{w} \geq 0$. We remark that in general such a set may not even exist; for example, the necessary number of linear objectives may be large or even infinite if the objective function has unbounded gradient (as is the case in the second model above). From a complexity perspective, minimizing a concave function over some feasible set may be hard to approximate even if minimizing a linear function over the same set can be done in polynomial time [22].

Thanks to the form of the objective functions, they can all be projected onto the mean-variance plane $span(\boldsymbol{\mu}, \boldsymbol{\tau})$ and can be thought of as functions on two dimensions. In that plane, the projected level sets of the objective functions are parabolas. We construct an approximate separation oracle, which tells us whether for a given function value λ there is a feasible solution below the $(1-\epsilon)\lambda$-level set or else if the entire feasible set is above the λ-level set. We do this by inscribing a (partial) polygon between these two level sets. Geometrically, the optimal polygon choice (with fewest sides) is such that its vertices are on one level set and its sides are tangent to the other, as shown in Figure 1(b). The FPTAS template for a maximization problem is described more formally in Figure 3 in the Appendix (it is analogous for a minimization problem).

Theorem 1. *There is an oracle fully-polynomial time approximation scheme for all problems in the reliable stochastic framework above, which uses as a black-box an exact algorithm for solving the underlying deterministic problem (5).*

In the rest of this section we prove this theorem. The crux of the proof is in establishing that the approximate separation oracle can be constructed from polynomially many linear segments, as described in the following main technical lemma. (The Lemma is stated for a stochastic maximization problem as in Eq. (2); the analogous statement holds for a stochastic minimization problem as in Eq. (1).) The argument for how the theorem follows from the Lemma is provided in the extended version.

Lemma 2 (**Approximate Separation Oracle).** *Suppose we have an exact algorithm for solving the deterministic problem (5). Then, we can construct an oracle which solves the following approximate separation problem: given a level λ and $\epsilon \in (0,1)$, the oracle returns*

1. *A solution $\mathbf{x} \in \mathcal{F}$ with $f(\mathbf{x}) \geq (1-\epsilon)\lambda$,* or
2. *An answer that $f(\mathbf{x}) < \lambda$ for all $\mathbf{x} \in \mathcal{F}$,*

and the number of linear oracle calls it makes is polynomial in $\frac{1}{\epsilon}$ and the size of the input.

The proof-construction of the Approximate Separation Oracle from Lemma 2 follows from a series of lemmas about bounding the size and number of the linear segments that approximate a level set and comprise the separation oracle. Since the level sets and their position with respect to each other is different for the different objectives, the actual computations of the size and number of linear segments differs. For lack of space we provide the proof for the probability tail formulation (2), which is more subtle due to the budget threshold and the fact the level sets are tangent to each other. The proofs for the remaining objectives are analogous.

Consider the lower level sets $\underline{L}_\lambda = \{\mathbf{z} \mid f(\mathbf{z}) \leq \lambda\}$ of the objective function $f(m,s) = \frac{t-m}{\sqrt{s}}$, where $m, s \in \mathbb{R}$. Denote $L_\lambda = \{\mathbf{z} \mid f(\mathbf{z}) = \lambda\}$. We will prove that any level set boundary can be approximated by a small number of linear segments. The main work here involves deriving a condition for a linear segment with endpoints on L_λ, to have objective function values within $(1-\epsilon)$ of λ.

Lemma 3. *Consider the points $(m_1, s_1), (m_2, s_2) \in L_\lambda$ with $s_1 > s_2 > 0$. The segment connecting these two points is contained in the level set region $\underline{L}_\lambda \backslash \underline{L}_{\lambda(1-\epsilon)}$ whenever $s_2 \geq (1-\epsilon)^4 s_1$, for every $\epsilon \in (0,1)$ (See Fig. 2(a)).*

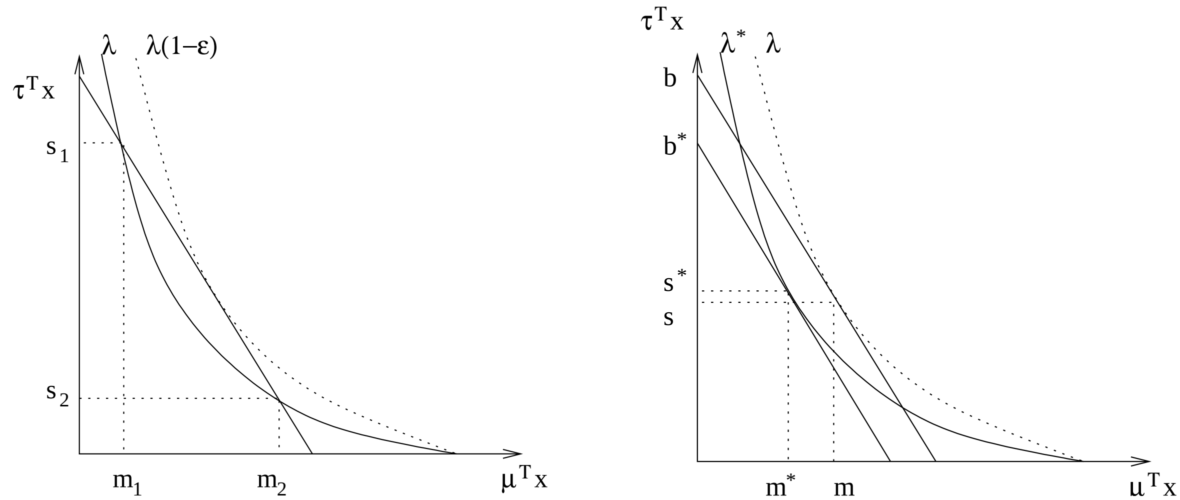

Figure 2: *(a)* The objective value along a segment is not too far from the objective value at the endpoints of the segment, provided s_1 and s_2 are not too far. λ and $\lambda(1-\epsilon)$ are the objective function values along the parabolic level sets. *(b)* Applying the approximate linear oracle on the optimal linear objective gives an approximate value b to the optimal linear objective value b^*. The challenge is to relate the linear oracle approximation factor $\frac{b}{b^*}$ to an approximation guarantee $\frac{\lambda}{\lambda^*}$ for the original nonlinear objective.

Proof. Any point on the segment $[(m_1, s_1), (m_2, s_2)]$ can be written as a convex combination of its endpoints, $(\alpha m_1 + (1-\alpha)m_2, \alpha s_1 + (1-\alpha)s_2)$, where $\alpha \in [0, 1]$. Consider the function $h(\alpha) = f(\alpha m_1 + (1-\alpha)m_2, \alpha s_1 + (1-\alpha)s_2)$. We have,

$$h(\alpha) \quad = \quad \frac{t - \alpha m_1 - (1-\alpha)m_2}{\sqrt{\alpha s_1 + (1-\alpha)s_2}} = \frac{t - \alpha(m_1 - m_2) - m_2}{\sqrt{\alpha(s_1 - s_2) + s_2}}$$

We want to find the point on the segment with smallest objective value, so we minimize with respect to α.

$$h'(\alpha) \quad = \quad \frac{\alpha(m_2 - m_1)(s_1 - s_2) + 2(m_2 - m_1)s_2 - (t - m_2)(s_1 - s_2)}{2[\alpha(s_1 - s_2) + s_2]^{3/2}}.$$

Setting the derivative to 0 is equivalent to setting the numerator above to 0, thus we get:

$$\alpha_{\min} = \frac{(t - m_2)(s_1 - s_2) - 2(m_2 - m_1)s_2}{(m_2 - m_1)(s_1 - s_2)} = \frac{t - m_2}{m_2 - m_1} - \frac{2s_2}{s_1 - s_2}.$$

Note that the denominator of $h'(\alpha)$ is positive and its numerator is linear in α, with a positive slope, therefore the derivative is negative for $\alpha < \alpha_{\min}$ and positive otherwise, so $\alpha_{\min}$ is indeed a global minimum as desired.

It remains to verify that $h(\alpha_{\min}) \geq (1-\epsilon)\lambda$. Note that $t - m_i = \lambda\sqrt{s_i}$ for $i = 1, 2$ since $(m_i, s_i) \in L_\lambda$ and consequently, $m_2 - m_1 = \lambda(\sqrt{s_1} - \sqrt{s_2})$. We use this in the following expansion of $h(\alpha_{\min})$.

$$\begin{aligned} h(\alpha_{\min}) \quad &= \quad \frac{t + \alpha_{\min}(m_2 - m_1) - m_2}{\sqrt{\alpha_{\min}(s_1 - s_2) + s_2}} = \frac{t + (\frac{t-m_2}{m_2-m_1} - \frac{2s_2}{s_1-s_2})(m_2 - m_1) - m_2}{\sqrt{(\frac{t-m_2}{m_2-m_1} - \frac{2s_2}{s_1-s_2})(s_1 - s_2) + s_2}} \\ &= \quad \frac{t + t - m_2 - 2s_2\frac{m_2-m_1}{s_1-s_2} - m_2}{\sqrt{(t - m_2)\frac{s_1-s_2}{m_2-m_1} - 2s_2 + s_2}} = \frac{2(t - m_2) - 2s_2\frac{\lambda(\sqrt{s_1}-\sqrt{s_2})}{s_1-s_2}}{\sqrt{\lambda\sqrt{s_2}\frac{s_1-s_2}{\lambda(\sqrt{s_1}-\sqrt{s_2})} - s_2}} = 2\lambda\frac{(s_1 s_2)^{1/4}}{\sqrt{s_1} + \sqrt{s_2}}. \end{aligned}$$

We need to show that when the ratio s_1/s_2 is sufficiently close to 1, $h(\alpha_{\min}) \geq (1-\epsilon)\lambda$, or equivalently

$$\frac{2(s_1 s_2)^{1/4}}{\sqrt{s_1}+\sqrt{s_2}} \geq 1-\epsilon \qquad \Leftrightarrow \qquad 2(s_1 s_2)^{1/4} \geq (1-\epsilon)(s_1^{1/2}+s_2^{1/2})$$

$$\Leftrightarrow \quad (1-\epsilon)\Big(\frac{s_1}{s_2}\Big)^{1/2} - 2\Big(\frac{s_1}{s_2}\Big)^{1/4} + (1-\epsilon) \leq 0 \tag{6}$$

The minimum of the last quadratic function above is attained at $\left(\frac{s_1}{s_2}\right)^{1/4} = \frac{1}{1-\epsilon}$ and we can check that at this minimum the quadratic function is indeed negative:

$$(1-\epsilon)\Big(\frac{1}{1-\epsilon}\Big)^2 - 2\Big(\frac{1}{1-\epsilon}\Big) + (1-\epsilon) = (1-\epsilon) - \frac{1}{1-\epsilon} < 0,$$

for all $0 < \epsilon < 1$. The inequality (6) is satisfied at $\frac{s_1}{s_2} = 1$, therefore it holds for all $\left(\frac{s_1}{s_2}\right) \in [1, \frac{1}{(1-\epsilon)^4}]$. Hence, a sufficient condition for $h(\alpha_{\min}) \leq (1-\epsilon)\lambda$ is $s_2 \geq (1-\epsilon)^4 s_1$, and we are done. □

Using Lemma 3, we next show that any level set L_λ can be approximated within a multiplicative factor of $(1-\epsilon)$ via a small number of segments. Let s_{min} and s_{max} be a lower and upper bound respectively for the variance of the optimal solution. For example, take s_{min} to be the smallest positive coordinate of the variance vector, and s_{max} the variance of the feasible solution with smallest mean.

Lemma 4. *The level set $L_\lambda = \{(m,s) \in \mathbb{R}^2 \mid \frac{t-m}{\sqrt{s}} = \lambda\}$ can be approximated within a factor of $(1-\epsilon)$ by $\left\lceil \frac{1}{4}\log\left(\frac{s_{\max}}{s_{\min}}\right) / \log\frac{1}{1-\epsilon}\right\rceil$ linear segments.*

Proof. By definition of s_{min} and s_{max}, the variance of the optimal solution ranges from $s_{\min}$ to $s_{\max}$. By Lemma 3, the segments connecting the points on L_λ with variances $s_{\max}, s_{\max}(1-\epsilon)^4, s_{\max}(1-\epsilon)^8, ..., s_{\min}$ all lie in the level set region $\underline{L}_\lambda \backslash \underline{L}_{\lambda(1-\epsilon)}$, that is they underestimate and approximate the level set L_λ within a factor of $(1-\epsilon)$. The number of these segments is $\lceil \frac{1}{4}\log\left(\frac{s_{\max}}{s_{\min}}\right) / \log\frac{1}{1-\epsilon}\rceil$. □

The above lemma yields the approximate separation oracle for the level set L_λ and the feasible set $\mathcal{F}$, by applying the black-box algorithm for the deterministic problem to cost vectors $a\boldsymbol{\mu}+\boldsymbol{\tau}$, for all possible slopes $(-a)$ of the segments approximating the level set. This concludes the proof-construction for the separation oracle in Lemma 2.

3 Approximating the reliable versions of hard combinatorial problems

In this section, we show that a δ-approximate oracle to the deterministic problem (5), also called the linear oracle, can be used to construct efficient approximation algorithms for the reliable stochastic models. As in the approximative analysis for easy combinatorial problems, we first check whether the optimal solution has zero variance and if not, proceed with the algorithm and analysis below.

We can use the same approximation algorithm template that constructs a separation oracle as in the previous section, but it turns out that a cruder algorithm which simply tests a geometric progression of mean-variance tradeoffs provides the same approximation guarantees. The main technical challenge in the algorithm analysis is that even if we know the optimal mean-variance tradeoff to query from the black-box algorithm for the deterministic problem, it is not obvious and not intuitive what approximation factor one can get for the reliable objectives from the δ-approximation factor for the deterministic one.

We obtain a very strong result for the relatively simpler mean-risk objective—we can get essentially the same approximation factor as the available one for the deterministic problem:

Theorem 5. *Suppose we have a δ-approximation oracle for solving the deterministic combinatorial problem (5). The mean-risk model (1) can be approximated to a multiplicative factor of $\delta(1+\epsilon)$ by calling the oracle for the deterministic problem polynomially many times in the input size and $\frac{1}{\epsilon}$.*

We can also get the following approximation for the probability tail formulation (2):

Theorem 6. *Suppose we have a δ-approximation oracle for solving the deterministic combinatorial problem (5). The probability tail model (2) has a $\sqrt{1-\left[\frac{\delta-(1-\epsilon^2/4)}{(2+\epsilon)\epsilon/4}\right]}$ -approximation algorithm that calls the algorithm for the deterministic problem polynomially many times in $\frac{1}{\epsilon}$ and the input size, assuming the optimal solution to (2) satisfies $\boldsymbol{\mu}^T\mathbf{x}^* \leq (1-\epsilon)t$.*

The high-level analysis for these approximation algorithms is the same; it differs in the computation of the approximation factors. For lack of space, we only offer an overview of the proof of Theorem 6; the remaining details for both theorems are in the extended version.

We first prove several geometric lemmas that enable us to derive the approximation factor. The first lemma is key for the transition from approximating a linear objective (by the algorithm for the deterministic problem) to approximating the probability tail objective. See Figure 2(b) for visualizing the notation.

Lemma 7 (Geometric lemma)**.** *Consider two objective function values $\lambda^* > \lambda$ and points $(m^*, s^*) \in L_{\lambda^*}$, $(m, s) \in L_\lambda$ with positive coordinates, such that the tangents to the points at the corresponding level sets are parallel. Then, the y-intercepts b^*, b of the two tangent lines satisfy*

$$b - b^* = s^*\Big[1 - \Big(\frac{\lambda}{\lambda^*}\Big)^2\Big].$$

The next lemma shows that if we know the optimal linear objective to use with the available δ-approximate algorithm for the deterministic problem (5), then we can approximate the optimal solution well.

Lemma 8 (Optimal Linear Objective Lemma)**.** *Suppose we have a δ-approximate linear oracle for optimizing over the feasible set $\mathcal{F}$ and suppose that the optimal solution satisfies $\boldsymbol{\mu}^T\mathbf{x}^* \leq (1-\epsilon)t$. If we can guess the slope of the tangent to the corresponding level set at the optimal point $\mathbf{x}^*$, then we can find a $\sqrt{1-\delta\frac{2-\epsilon}{\epsilon}}$-approximate solution to the nonconvex problem (2).*

In particular, setting $\epsilon = \sqrt{\delta}$ gives a $(1-\sqrt{\delta})$-approximate solution.

Next, we prove a geometric lemma that will be needed to analyze the approximation factor we get when applying the linear oracle on an approximately optimal slope.

Lemma 9. *Consider the level set L_λ and points (m^*, s^*) and (m, s) on it, at which the tangents to L_λ have slopes $-a$ and $-a(1+\xi)$ respectively. Let the y-intercepts of the tangent line at (m, s) and the line parallel to it through (m^*, s^*) be b_1 and b respectively. Then $\frac{b}{b_1} \leq \frac{1}{1-\xi^2}$.*

We now show that we get a good approximation even when we use an approximately optimal linear objective with our linear oracle.

Lemma 10. *Suppose that we use an approximately optimal linear objective with a δ-approximate linear oracle for solving the probability tail model (2). In particular, suppose the linear objective (slope) that we use is within $(1+\xi)$ of the slope of the tangent at the optimal solution. Then this will give a solution to the probability tail model (2) with value at least $\sqrt{1-\left[\frac{\delta}{1-\xi^2}-1\right]\frac{2-\epsilon}{\epsilon}}$ times the optimal, provided the optimal solution satisfies $\boldsymbol{\mu}^T\mathbf{x}^* \leq (1-\epsilon)t$.*

Consequently, we can approximate the optimal solution by applying the approximate linear oracle on a small number of appropriately chosen linear functions and picking the best resulting solution, as explained in the proof of Theorem 6 in the extended version.

When $\delta = 1$, that is when we can solve the underlying linear problem exactly in polynomial time, the above algorithm gives an approximation factor of $\sqrt{\frac{1}{1+\epsilon/2}}$, or equivalently $1-\epsilon'$, where $\epsilon = 2[\frac{1}{(1-\epsilon')^2} - 1]$. While this algorithm is still an oracle-fully polynomial time approximation scheme, it gives a bi-criteria approximation: it requires that there is a small gap between the mean of the optimal solution and the budget t so it is weaker than our previous algorithm, which had no such requirement. This is expected since, of course, this algorithm is cruder, simply taking a geometric progression of linear functions rather than tailoring the black-box algorithm calls for the deterministic problem to the objective function value that it is searching for, as in the approximate separation oracle that the FPTAS from the previous section is based on.

4 Conclusion

We have presented a framework for reliable stochastic combinatorial optimization that includes mean-risk minimization and models involving the probability tail of the stochastic cost of a solution. Our algorithms are independent of the feasible set structure and use solutions for the underlying linear (deterministic) problems as oracles for solving the corresponding stochastic models. As such, they apply to very general combinatorial settings for which *exact* or *approximate* linear oracles are available.

Our primary motivation for this work was to design an approximation algorithm for finding the most reliable route in a network with uncertain edge delays (in the sense that the route maximizes the probability of arriving on time under a given deadline), which consequently extended to the rich class of problems and reliability models considered here. An implementation of our approximation algorithm in the context of reliable routes reveals that they are also very practical: for example, they achieve 99.9%-accuracy with only up to 6 iterations of an algorithm for the deterministic problem.

In future work, it would be interesting to extend our offline stochastic models to online models, as has previously been done with offline linear to online linear problems [20, 19]. It would be also useful to consider adaptive stochastic reliability models, building on the framework of multistage stochastic optimization.

Acknowledgments. The author thanks Hari Balakrishnan, Dimitris Bertsimas, Shuchi Chawla, Costis Daskalakis, Nick Harvey, Michel Goemans, David Karger, David Kempe, Asu Ozdaglar, Christos Papadimitriou, Pablo Parrilo, Andreas Schulz, Nicolás Stier Moses, Madhu Sudan and John Tsitsiklis for valuable suggestions at various stages of this work.

References

[1] H. Ackermann, A. Newman, H. Röglin, and B. Vöcking. Decision making based on approximate and smoothed pareto curves. In *Proc. of 16th ISAAC*, pages 675–684, 2005.

[2] A. Atamtürk and V. Narayanan. Polymatroids and risk minimization in discrete optimization. *Operations Research Letters*, 36:618–622, 2008.

[3] Y. Berstein, J. Lee, S. Onn, and R. Weismantel. Nonlinear optimization for matroid intersection and extensions. *Manuscript at arXiv:0807.3907*, 2008.

[4] D. Bertsekas, A. Nedić, and A. Ozdaglar. *Convex Analysis and Optimization*. Athena Scientific, Belmont, MA, 2003.

[5] D. Bertsimas, D. Brown, and C. Caramanis. Theory and applications of robust optimization. *Manuscript*, 2007.

[6] D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Manuscript*, 2004.

[7] A. Chen and Z. Ji. Path finding under uncertainty. *Journal of advanced transportation*, 39(1):19–37, 2005.

[8] B. Dean, M. X. Goemans, and J. Vondrák. Approximating the stochastic knapsack: the benefit of adaptivity. In *Proceedings of the 45th Annual Symposium on Foundations of Computer Science*, pages 208–217, 2004.

[9] Y. Fan, R. Kalaba, and I. J. E. Moore. Arriving on time. *Journal of Optimization Theory and Applications*, 127(3):497–513, 2005.

[10] Federal Highway Administration. *Traffic Congestion and Reliability: Trends and advanced strategies for congestion mitigation*. Cambridge Systematics Inc. and Texas Transportation Institute, 2005.

[11] H. Föllmer and A. Schied. *Stochastic Finance: An Introduction in Discrete Time*. Walter de Gruyter, 2004.

[12] L. E. Ghaoui, M. Oks, and F. Oustry. Worst-case value-at-risk and robust portfolio optimization: A conic programming approach. *Oper. Res.*, 51(4):543–556, 2003.

[13] A. Goel and P. Indyk. Stochastic load balancing and related problems. In *Proceedings of the 40th Symposium on Foundations of Computer Science*, 1999.

[14] V. Goyal. An FPTAS for minimizing a class of quasi-concave functions over a convex set. Technical Report Tepper WP 2008-E24, Carnegie Mellon University Tepper School of Business, 2008.

[15] G. Grimmett and D. Stirzaker. *Probability and Random Processes, 3rd ed.* Oxford Univ. Press, 2001.

[16] A. Gupta, M. Pál, R. Ravi, and A. Sinha. Boosted sampling: approximation algorithms for stochastic optimization. In *Proceedings of the 36th Annual ACM Symposium on Theory of Computing*, pages 365–372, Chicago, IL, USA, 2004.

[17] A. Gupta, M. Pál, R. Ravi, and A. Sinha. What about Wednesday? Approximation algorithms for multistage stochastic optimization. In *Proceedings of 8th APPROX*, pages 86–98, 2005.

[18] N. Immorlica, D. Karger, M. Minkoff, and V. S. Mirrokni. On the costs and benefits of procrastination: approximation algorithms for stochastic combinatorial optimization problems. In *Proceedings of the fifteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 691–700, 2004.

[19] S. M. Kakade, A. T. Kalai, and K. Ligett. Playing games with approximation algorithms. In *STOC '07: Proceedings of the thirty-ninth annual ACM symposium on Theory of computing*, pages 546–555, New York, NY, USA, 2007. ACM.

[20] A. Kalai and S. Vempala. Efficient algorithms for on-line optimization. *Journal of Computer and System Sciences*, 71:291–307, 2005.

[21] I. Katriel, C. Kenyon-Mathieu, and E. Upfal. Commitment under uncertainty: Two-stage stochastic matching problems. *Theor. Comput. Sci.*, 408(2-3):213–223, 2008.

[22] J. A. Kelner and E. Nikolova. On the hardness and smoothed complexity of quasi-concave minimization. In *Proceedings of the 48st Annual Symposium on Foundations of Computer Science*, Providence, RI, USA, 2007.

[23] J. Kleinberg, Y. Rabani, and É. Tardos. Allocating bandwidth for bursty connections. *SIAM Journal on Computing*, 30(1):191–217, 2000.

[24] L. Lovász. Submodular functions and convexity. In *Mathematical Programming—The State of the Art, A. Bachem, M, Grőtschel and B. Korte, eds.*, pages 235–257. Springer-Verlag, 1983.

[25] Y. (Marco), Nie and X. Wu. Shortest path problem considering on-time arrival probability. *Transportation Research Part B: Methodological*, 43(6):597–613, 2009.

[26] H. M. Markowitz. *Mean-Variance Analysis in Portfolio Choice and Capital Markets*. Basil Blackwell, Cambridge, MA, 1987.

[27] N. Megiddo. Combinatorial optimization with rational objective functions. *Mathematics of Operations Research*, 4:414–424, 1979.

[28] E. Miller-Hooks and H. Mahmassani. Path comparisons for a priori and time-adaptive decisions in stochastic, time-varying networks. *European Journal of Operational Research*, 146(1):67–82, April 2003.

[29] A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2006.

[30] E. Nikolova, J. A. Kelner, M. Brand, and M. Mitzenmacher. Stochastic shortest paths via quasi-convex maximization. In *Lecture Notes in Computer Science 4168 (ESA 2006)*, pages 552–563, Springer-Verlag, 2006.

[31] S. Onn. Convex discrete optimization. *Encyclopedia of Optimization*, pages 513–550, 2009.

[32] C. H. Papadimitriou and M. Yannakakis. On the approximability of trade-offs and optimal access of web sources. In *Proceedings of the 41st Annual Symposium on Foundations of Computer Science*, pages 86–92, Washington, DC, USA, 2000.

[33] T. Radzik. Newton's method for fractional combinatorial optimization. In *Proceedings of the 33rd Annual Symposium on Foundations of Computer Science*, pages 659–669, 1992.

[34] R. T. Rockafellar. Coherent approaches to risk in optimization under uncertainty. In *Tutorials in Operations Research INFORMS*, pages 38–61, 2007.

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

[36] D. B. Shmoys and C. Swamy. An approximation scheme for stochastic linear programming and its application to stochastic integer programs. *Journal of the ACM*, 53(6):978–1012, 2006.

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

[38] A. Srinivasan. Approximation algorithms for stochastic and risk-averse optimization. In *SODA '07: Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 1305–1313, Philadelphia, PA, USA, 2007.

[39] C. Swamy and D. B. Shmoys. Approximation algorithms for 2-stage stochastic optimization problems. *ACM SIGACT News*, 37(1):33–46, 2006.

[40] A. Warburton. Approximation of pareto optima in multiple-objective, shortest-path problems. *Oper. Res.*, 35(1):70–79, 1987.

Appendix

Problem: Maximize $f(\mathbf{x})$ over $\mathbf{x} \in \mathcal{F}$.
Output: Solution $\mathbf{x}'$ such that $f(\mathbf{x}') \geq (1-\epsilon)f_{max}(\mathbf{x})$
Algorithm:

1. For appropriate lower and upper bounds of $f(\cdot)$, denoted f_l and f_u respectively, apply *approximate separation oracle* below with $\epsilon' = 1 - \sqrt{1-\epsilon}$ successively on the function values $f_u, (1-\epsilon')f_u, (1-\epsilon')^2 f_u, ...$ until we find a value, for which the separation oracle returns a feasible solution $\mathbf{x}'$.
2. Run the available black-box algorithm for the deterministic problem on subset of elements with zero mean, to find the smallest-variance solution among the solutions with mean zero. Compare with the solution above and return the solution with better objective function value.

Approximate Separation Oracle.
Input: Function value λ, approximation factor $\epsilon' > 0$; black-box access to algorithm for minimizing linear functions over $\mathbf{x} \in \mathcal{F}$.
Output:

(a) A solution $\mathbf{x}' \in \mathcal{F}$ with $f(\mathbf{x}') \geq (1-\epsilon')\lambda$, *or*

(b) An answer that $f(\mathbf{x}) < \lambda$ for all $\mathbf{x} \in \mathcal{F}$.

Algorithm:

1. Inscribe a polygon between the level sets corresponding to function values λ and $(1-\epsilon')\lambda$.
2. For each side of the polygon, minimize the induced linear objective.
3. If a resulting solution $\mathbf{x}'$ satisfies $f(\mathbf{x}') \geq (1-\epsilon)\lambda$, return $\mathbf{x}'$. Else return that $f(\mathbf{x}) < \lambda$ for all $\mathbf{x} \in \mathcal{F}$.

Figure 3: FPTAS template for solving reliable stochastic models.