Computer Science and Artificial Intelligence Laboratory
Technical Report

MIT-CSAIL-TR-2008-055 September 13, 2008

Stochastic Combinatorial Optimization with Risk

Evdokia Nikolova

massachusetts institute of technology, cambridge, ma 02139 usa — www.csail.mit.edu

Stochastic Combinatorial Optimization with Risk

Evdokia Nikolova
Massachusetts Institute of Technology, The Stata Center, Room 32-G596, 32 Vassar Street, Cambridge, MA 02139, USA,
{nikolova@mit.edu}

We consider general combinatorial optimization problems that can be formulated as minimizing the weight of a feasible solution $\mathbf{w}^T\mathbf{x}$ over an arbitrary feasible set. For these problems we describe a broad class of corresponding stochastic problems where the weight vector $\mathbf{W}$ has independent random components, unknown at the time of solution. A natural and important objective which incorporates risk in this stochastic setting, is to look for a feasible solution whose stochastic weight has a small tail or a small linear combination of mean and standard deviation. Our models can be equivalently reformulated as deterministic 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 combinatorial problem and thus immediately apply to arbitrary combinatorial problems. For example, from an available δ-approximation algorithm to the deterministic problem, we construct a $\delta(1+\epsilon)$-approximation algorithm that invokes the deterministic algorithm only a logarithmic number of times in the 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, combinatorial optimization, stochastic optimization, risk, nonconvex optimization

1. Introduction

Imagine driving to the airport through uncertain traffic. While we may not know specific travel times along different roads, we may have information on their distributions (for example their means and variances). We want to find a route that gets us to the airport on time. The route minimizing expected travel time may well cause us to be late. In contrast, arriving on time requires accounting for traffic variability and risk.

In this paper we consider general combinatorial optimization problems that can be formulated as minimizing the weight $\mathbf{w}^T\mathbf{x}$ of a feasible solution over a fixed feasible set. For these problems we describe a broad class of corresponding stochastic problems where the weight vector $\mathbf{W}$ has independent random components, unknown at the time of solution. A natural and important objective which incorporates risk in this stochastic setting, is to look for a feasible solution whose stochastic weight has a small tail (as in the example above, where we seek to minimize the probability that the random route length exceeds a given threshold) or a small linear combination of mean and standard deviation. Our models can be equivalently reformulated as deterministic nonconvex programs for which no efficient algorithms are known. In this paper, we make progress on these hard problems, and in particular our main contributions are as follows.

Our Results

1. Suppose we have an exact algorithm for the underlying deterministic combinatorial problem. Then, for all stochastic variants we consider, we obtain efficient $(1+\epsilon)$-approximation schemes, which make a logarithmic number of oracle calls to the deterministic algorithm (Theorem 4, Theorem 15).

2. Suppose we have a δ-approximate algorithm for the deterministic problem. Then, for the stochastic problem of minimizing the tail of the solution weight's distribution, we provide a $\sqrt{1-\left[\frac{\delta-(1-\epsilon^2/4)}{(2+\epsilon)\epsilon/4}\right]}$-approximation scheme, which as above makes a logarithmic number of oracle calls to the deterministic algorithm (Theorem 10). This result assumes normally distributed weights.

3. Suppose we have a δ-approximate algorithm for the deterministic problem. Then, for the stochastic (nonconvex) problem of minimizing a linear combination of mean and standard deviation of the solution weight, we give an $\delta(1+\epsilon)$-approximation scheme which makes a logarithmic number of oracle calls to the deterministic algorithm (Theorem 21). This result holds for arbitrary weight distributions, and only assumes knowledge of the mean and variance of the distributions.

To the best of our knowledge, this is the first treatment of stochastic combinatorial optimization that incorporates risk, together with providing general-purpose approximation techniques applicable to arbitrary combinatorial problems. In fact, since our algorithms are independent of the feasible set structure, they immediately apply to any *discrete* problems, and not just $\{0, 1\}$. Similarly, they continue to work in a continuous setting where the feasible set is compact and convex.

Our approximation schemes are based on a series of geometric lemmas analyzing the form of the objective function level sets, and on a novel construction of an approximate *non-linear separation oracle* from a *linear oracle* (the algorithm to the deterministic problem), in which the main technical lemma is that a logarithmic number of applications of the linear oracle suffice to get an arbitrarily good approximation.

Given the general-purpose nature of our algorithms and their near-optimal running time, our results constitute significant progress in both stochastic and nonconvex optimization. In particular, we believe that our approach and techniques would extend to give approximation algorithms for a wider class of nonconvex (and related stochastic) optimization problems, for which no efficient solutions are currently available.

Perhaps more importantly from a practical standpoint, as a by-product of our stochastic models we can approximate the *distribution* of the weight of the optimal solution: Applying our solution to the stochastic tail distribution objective for different threshold (tail) values will yield an approximate histogram of the optimal distribution. Consequently our models and algorithms provide a powerful tool for approximating arbitrary objectives and statistics from the distribution.

We defer to the next section a discussion of related work and contrast our results to potential other approaches. Our approximation algorithms are presented in the following four sections. Our algorithms for the tail (threshold) objective require some additional assumptions, and because of the form of that objective the analysis is more challenging and subtle. We present these in Sections 3 and 4 for the cases when we have an exact and an approximate oracle for solving the underlying deterministic problem respectively. Our algorithms for the mean-standard deviation objective are more general and somewhat easier to analyze: they are presented in Sections 5 and 6.

2. The Stochastic Framework

In this section, we formally define the classes of stochastic problems we consider. We then discuss related work and contrast our approach with other potential solution approaches.

Consider an arbitrary deterministic combinatorial problem which minimizes the weight of a feasible solution $\mathbf{w}^T\mathbf{x}$ over a fixed feasible set $\mathcal{F}$:

$$\text{minimize } \mathbf{w}^T\mathbf{x} \qquad \text{subject to } \mathbf{x} \in \mathcal{F}. \tag{1}$$

Notation We adopt the common standard of bold font for vectors and regular font for scalars, and denote the transpose of a vector, say $\mathbf{x}$, by $\mathbf{x}^T$. Define polytope$(\mathcal{F}) \in \mathbb{R}^n$ to be the convex hull of the feasible set $\mathcal{F}$. Let $\mathbf{W} = (W_1, ..., W_n)$ be a vector of independent stochastic weights, and $\boldsymbol{\mu} = (\mu_1, ..., \mu_n)$ and $\boldsymbol{\tau} = (\tau_1, ..., \tau_n)$ be the vectors of their means and variances respectively.

We consider the following broad classes of stochastic problems, summarized in the table below together with their equivalent reformulation as deterministic nonconvex problems.

Model Name	Stochastic Problem	Nonconvex Problem	
Threshold	maximize $\Pr(\mathbf{W}^T\mathbf{x} \le t)$ subject to $\mathbf{x} \in \mathcal{F}$	maximize $\dfrac{t - \boldsymbol{\mu}^T\mathbf{x}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}}}$ subject to $\mathbf{x} \in$ polytope$(\mathcal{F})$	(2)
Value-at-risk	minimize t subject to $\Pr(\mathbf{W}^T\mathbf{x} \le t) \ge p$ $\mathbf{x} \in \mathcal{F}$	minimize $\boldsymbol{\mu}^T\mathbf{x} + c\sqrt{\boldsymbol{\tau}^T\mathbf{x}}$ subject to $\mathbf{x} \in$ polytope$(\mathcal{F})$	(3)
Risk	minimize $\boldsymbol{\mu}^T\mathbf{x} + c\sqrt{\boldsymbol{\tau}^T\mathbf{x}}$ subject to $\mathbf{x} \in \mathcal{F}$		

We will give approximation algorithms for the two nonconvex problems above, using as an oracle the available solutions to the underlying problem (1). A δ-approximation algorithm for a minimization problem (with $\delta \geq 1$) is one which returns a solution with value at most δ times the optimum. We use the term *oracle-logarithmic time approximation scheme* (abbrev. oracle-LTAS) for a $(1+\epsilon)\delta$-approximation algorithm which makes a number of oracle queries that is logarithmic in the problem size and polynomial in $\frac{1}{\epsilon}$ (where δ is the multiplicative approximation factor of the oracle). We now briefly explain why solving the nonconvex problems will solve the corresponding stochastic problems.

Stochastic threshold objective When the weights come from independent normal distributions, a feasible solution $\mathbf{x}$ will have a normally distributed weight $\mathbf{W}^T\mathbf{x} \sim N(\boldsymbol{\mu}^T\mathbf{x}, \boldsymbol{\tau}^T\mathbf{x})$. Therefore

$$\Pr\left[\mathbf{W}^T\mathbf{x} \leq t\right] \quad = \quad \Pr\left[\frac{\mathbf{W}^T\mathbf{x} - \boldsymbol{\mu}^T\mathbf{x}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}}} \leq \frac{t - \boldsymbol{\mu}^T\mathbf{x}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}}}\right] = \Phi\Big(\frac{t - \boldsymbol{\mu}^T\mathbf{x}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}}}\Big),$$

where $\Phi(\cdot)$ is the cumulative distribution function of the standard normal random variable $N(0,1)$. Since $\Phi(\cdot)$ is monotone increasing, maximizing the stochastic threshold objective above is equivalent to maximizing the argument, namely it is equivalent to the nonconvex threshold problem (2). Furthermore, it can be easily shown that an approximation for the nonconvex problem (2) yields the same approximation factor for the stochastic problem.[1]

Stochastic risk and value-at-risk objectives When the weights come from arbitrary independent distributions, the mean and variance of a feasible solution $\mathbf{x}$ will be equal to the sum of means $\boldsymbol{\mu}^T\mathbf{x}$ and sum of variances $\boldsymbol{\tau}^T\mathbf{x}$ of the components of $\mathbf{x}$, hence the equivalent concave formulation (3). The value-at-risk objective also reduces to problem (3). For arbitrary distributions this follows from Chebyshev's bound, see Section 7.1 in the Appendix for details.

Properties of the nonconvex objectives Objectives (2) and (3) are instances of quasi-convex maximization and concave minimization respectively; consequently they attain their optima at extreme points of the feasible set (Bertsekas et al., 2003; Nikolova et al., 2006).

2.1. Related Work

The stochastic threshold objective was previously considered in the special case of shortest paths (Nikolova et al., 2006). The authors showed that this objective has the property that its optimum is an extreme point of the feasible set, and gave an exact algorithm based on enumerating extreme points. The property that the optimum is an extreme point holds here as well, however this is where the similarity of our work to this prior work ends: For general combinatorial problems it is likely that the number of relevant extreme points is too

[1]We expect that under reasonable conditions, *e.g.,* if a feasible solution $\mathbf{x}$ has sufficiently many nonzero components, arbitrary weight distributions will lead to feasible solutions having approximately normal weight by the Central Limit Theorem. Thus our algorithms are likely to provide a good approximation in that general case as well.

high (or unknown) and enumerating them will yield very inefficient algorithms. The focus and results of this paper are instead on approximation algorithms, in particular ones which are guaranteed to be efficient for arbitrary combinatorial optimization problems.

Our nonconvex objectives fall in the class of constant-rank quasi-concave minimization problems considered by Kelner and Nikolova (2007) (in our case the objectives are of rank 2), who give approximation algorithms based on smoothed analysis for some of these problems. Their approximation algorithms do not apply to our setting, since they require the objective to have a *bounded gradient* and a *positive lower bound* for the optimum (so as to turn additive into multiplicative approximation), as is not the case here.

Perhaps closest in spirit to our oracle-approximation methods is the work on robust optimization by Bertsimas and Sim (2003). Although very different in terms of models and algorithmic solutions, they also show how to solve the robust combinatorial problems via a small number of oracle calls of the underlying deterministic problems.

A wealth of different models for stochastic combinatorial optimization have appeared in the literature, perhaps most commonly on two-stage and multi-stage stochastic optimization, see survey by Swamy and Shmoys (2006). Almost all such work considers linear objective functions (*i.e.*, minimizing the expected solution weight) and as such does not consider risk. Some of the models incorporate additional budget constraints (Srinivasan, 2007) or threshold (chance) constraints for specific problems such as knapsack, load balancing and others (Dean et al., 2004; Goel and Indyk, 1999; Kleinberg et al., 2000). A comprehensive survey on stochastic optimization with risk with a different focus (different solution concept and continuous settings) is provided by Rockafellar (2007). Similarly, the work on chance constraints (*e.g.,* Nemirovski and Shapiro (2006)) applies for linear and not discrete optimization problems. Additional related work includes research on multi-criteria optimization, *e.g.,* (Papadimitriou and Yannakakis, 2000; Ackermann et al., 2005; Safer et al., 2004; Warburton, 1987) and combinatorial optimization with a ratio of linear objectives (Megiddo, 1979; Radzik, 1992). In one multi-criteria setting, Safer et al. (2004) consider nonlinear objectives $f(x)$. However they assume that the statement "Is $f(x) \leq M$?" can be evaluated in polynomial time (that is a key technical challenge in our paper), and their functions $f(x)$ have a much simpler separable form.

2.2. Our results vs other potential approaches

Specific combinatorial problems under our framework can be solved with alternative approaches. For example, consider the NP-hard constrained optimization problem $\{\min \boldsymbol{\mu}^T\mathbf{x}$ subject to $\boldsymbol{\tau}^T\mathbf{x} \leq B,\ \mathbf{x} \in \mathcal{F}\}$. Suppose we can get an approximate solution $\mathbf{x}'$ to the latter, which satisfies $\boldsymbol{\mu}^T\mathbf{x}' \leq \boldsymbol{\mu}^T\mathbf{x}^*$ and $\boldsymbol{\tau}^T\mathbf{x}' \leq B(1+\epsilon)$, where $\mathbf{x}^*$ is the optimal solution to the constrained problem with budget B. Then we can derive a fully polynomial-time approximation scheme (FPTAS) to the nonconvex problems (2), (3) by considering a geometric progression of budgets B in the constraint above, and picking the solution with the best objective value (2) or (3). This approach can be used whenever we have the above type of FPTAS to the constrained problem, as is the case for shortest paths (Goel et al., 2001). However, since we do not have a black-box solution to the constrained problem in general, this approach does not seem to extend to arbitrary combinatorial problems.

Another approach similar to the constrained problem above would be to use the approximate Pareto boundary.[2] The latter consists of a polynomial set of feasible solutions, such that for any point $\mathbf{x}$ on the Pareto boundary, there is a point $\mathbf{x}'$ in the set that satisfies $\boldsymbol{\mu}^T\mathbf{x}' \leq (1+\epsilon)\boldsymbol{\mu}^T\mathbf{x}$ and $\boldsymbol{\tau}^T\mathbf{x}' \leq (1+\epsilon)\boldsymbol{\tau}^T\mathbf{x}$. *When available* (*e.g.*, for shortest paths, etc.), such a bicriteria approximation will translate into an FPTAS for the nonconvex risk objective (3). However it will *not* yield an approximation algorithm to the nonconvex threshold objective (2), because a multiplicative approximation of $\boldsymbol{\mu}^T\mathbf{x}$ does not translate into a multiplicative approximation of $(t - \boldsymbol{\mu}^T\mathbf{x})$.

Radzik gives a black-box solution for combinatorial optimization with rational objectives that are a ratio of two linear functions, by converting the rational objective into a linear constraint. A key property of the rational function that allows for an efficient algorithm is that it is *monotone* along the boundary of the feasible set; this is not the case for any of our objective functions and is one of the biggest challenges in working with nonconvex optimization problems: greedy, local search, interior point and other standard techniques do not work.

Our approach is conceptually very different from previous analyses of related problems. Common approximation techniques for hard instances of stochastic and multicriteria problems convert pseudopolynomial algorithms to FPTAS by scaling and rounding (Warburton, 1987; Safer et al., 2004), or they discretize the decision space and use a combination of dynamic programming and enumeration of possible feasible solutions over this cruder space (Goel and Indyk, 1999). In most cases the techniques are intimately intertwined with the structure of the underlying combinatorial problem and cannot extend to arbitrary problems. In contrast, the near-optimal efficiency of our algorithms is due to the fact that we carefully analyze the form of the objective function and use a "top-down" approach where our knowledge of the objective function level sets guides us to zoom down into the necessary portion of the feasible space.

[2]The Pareto boundary consists of all non-dominated feasible points $\mathbf{x}$, namely all points such that there is no other feasible point $\mathbf{x}'$ with smaller mean $\boldsymbol{\mu}^T\mathbf{x}' \leq \boldsymbol{\mu}^T\mathbf{x}$ and variance $\boldsymbol{\tau}^T\mathbf{x}' \leq \boldsymbol{\tau}^T\mathbf{x}$.

3. An oracle-LTAS for the nonconvex threshold objective with exact oracle

In this section, we give an oracle-logarithmic time approximation scheme (LTAS) for the nonconvex problem formulation (2) that uses access to an exact oracle for solving the underlying problem (1).

Our algorithms assume that the maximum of the objective is non-negative, in other words the feasible solution with smallest mean satisfies $\boldsymbol{\mu}^T\mathbf{x} \leq t$. Note, it is not clear a priori that that such a multiplicative approximation is even possible, since we still let the function have positive, zero or negative values on different feasible solutions. The case in which the maximum is negative is structurally very different (the objective on its negative range no longer attains optima at extreme points) and remains open. Even with this assumption, approximating the objective function is especially challenging due to its unbounded gradient and the form of its numerator.

We first note that if the optimal solution has variance 0, we can find it exactly with a single oracle query: Apply the linear oracle on the set of elements with zero variances to find the feasible solution with smallest mean. If the mean is no greater than t, output the solution, otherwise conclude that the optimal solution has positive variance and proceed with the approximation scheme below.

The main technical lemma that our algorithm is based on is an extension of the concept of separation and optimization: instead of deciding whether a line (hyperplane) is separating for a polytope, in the sense that the polytope lies entirely on one side of the line (hyperplane), we construct an approximate oracle which decides whether a non-linear curve (in our case, a parabola) is separating for the polytope.

From here on we will analyze the projections of the objective function and the feasible set onto the plane $span(\boldsymbol{\mu}, \boldsymbol{\tau})$ since the nonconvex problem (2) is equivalent in that space. 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 first 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 1. *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* $\epsilon \in (0,1)$*.*

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.

$$\begin{aligned} h'(\alpha) &= \frac{(m_2-m_1)\sqrt{\alpha(s_1-s_2)+s_2} - [t-\alpha(m_1-m_2)-m_2] * \frac{1}{2}(s_1-s_2)/\sqrt{\alpha(s_1-s_2)+s_2}}{\alpha(s_1-s_2)+s_2} \\ &= \frac{2(m_2-m_1)[\alpha(s_1-s_2)+s_2] - [t-\alpha(m_1-m_2)-m_2](s_1-s_2)}{2[\alpha(s_1-s_2)+s_2]^{3/2}} \\ &= \frac{\alpha(m_2-m_1)(s_1-s_2)+2(m_2-m_1)s_2-(t-m_2)(s_1-s_2)}{2[\alpha(s_1-s_2)+s_2]^{3/2}}. \end{aligned}$$

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 further down in the following expansion of $h(\alpha_{\min})$.

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

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

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

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

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:

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

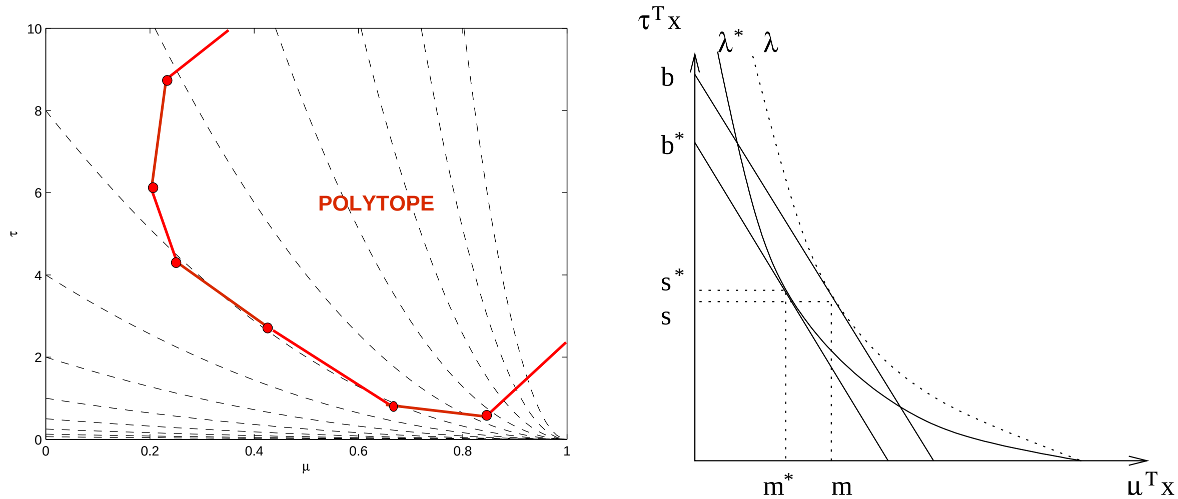

Figure 1: *(a)* Level sets of the objective function and the projected polytope on the $span(\boldsymbol{\mu}, \boldsymbol{\tau})$-plane. *(b)* Applying the approximate linear oracle on the optimal slope gives an approximate value b of the corresponding linear objective value b^*. The resulting solution has nonlinear objective function value of at least λ, which is an equally good approximation for the optimal value λ^*.

for all $0 < \epsilon < 1$. The inequality (4) 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. $\square$

Lemma 1 now makes it easy to show our main lemma, namely 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 component of the variance vector, and s_{max} the variance of the feasible solution with smallest mean.

Lemma 2. *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 the variance of the optimal solution ranges from $s_{\min}$ to $s_{\max}$. By Lemma 1, the segments connecting the points on L_λ with variances $s_{\max}, s_{\max}(1-\epsilon)^4, s_{\max}(1-\epsilon)^8, \ldots, s_{\min}$ 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 $\lceil \frac{1}{4} \log\left(\frac{s_{\max}}{s_{\min}}\right) / \log \frac{1}{1-\epsilon} \rceil$. $\square$

The above lemma yields an approximate separation oracle for the nonlinear level set L_λ and polytope$(\mathcal{F})$. The oracle takes as input the level λ and either returns a solution $\mathbf{x}$ with objective value $f(\mathbf{x}) \geq (1-\epsilon)\lambda$ from the feasible set, or guarantees that $f(\mathbf{x}) < \lambda$ for all $\mathbf{x} \in$ polytope$(\mathcal{F})$. Therefore, an exact oracle for solving problem (1) allows us to obtain an approximate nonlinear separation oracle, by applying the former to weight vectors $a\boldsymbol{\mu} + \boldsymbol{\tau}$, for all possible slopes $(-a)$ of the segments approximating the level set. We formalize this in the next theorem.

Theorem 3 (**Approximate Nonlinear Separation Oracle).** *Suppose we have an exact (linear) oracle for solving problem (1). Then, we can construct a nonlinear 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 $\mathbf{x} \in$ polytope$(\mathcal{F})$,*

and the number of linear oracle calls it makes is $\frac{1}{4}\log\left(\frac{s_{\max}}{s_{\min}}\right)/\log\frac{1}{1-\epsilon}$, that is $O(\frac{1}{\epsilon}\log\frac{s_{\max}}{s_{\min}})$.

We can now give an oracle-LTAS for the nonconvex problem (2), by applying the above nonlinear oracle on a geometric progression of possible values λ of the objective function f. We first need to bound the maximum value f_{opt} of the objective function f. A lower bound f_l is provided by the solution $\mathbf{x}_{mean}$ with smallest mean or the solution $\mathbf{x}_{var}$ with smallest positive variance, whichever has a higher objective value: $f_l = \max\{f(\mathbf{x}_{mean}), f(\mathbf{x}_{var})\}$ where $f(\mathbf{x}) = \frac{t-\boldsymbol{\mu}^T\mathbf{x}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}}}$. On the other hand, $\boldsymbol{\mu}^T\mathbf{x} \geq \boldsymbol{\mu}^T\mathbf{x}_{mean}$ and $\boldsymbol{\tau}^T\mathbf{x} \geq \boldsymbol{\tau}^T\mathbf{x}_{var}$ for all $\mathbf{x} \in$ polytope$(\mathcal{F})$, so an upper bound for the objective f is given by $f_u = \frac{t-\boldsymbol{\mu}^T\mathbf{x}_{mean}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}_{var}}}$ (recall that $t - \boldsymbol{\mu}^T\mathbf{x}_{mean} > 0$ by assumption).

Theorem 4. *Suppose we have an exact oracle for problem (1) and suppose the smallest mean feasible solution satisfies $\boldsymbol{\mu}^T\mathbf{x} \leq t$. Then for any $\epsilon \in (0,1)$, there is an algorithm for solving the nonconvex threshold problem (2), which returns a feasible solution $\mathbf{x} \in \mathcal{F}$ with value at least $(1-\epsilon)$ times the optimum, and makes $O\left(\log\left(\frac{s_{\max}}{s_{\min}}\right)\log\left(\frac{f_u}{f_l}\right)\frac{1}{\epsilon^2}\right)$ oracle calls.*

Proof. Now, apply the approximate separation oracle from Theorem 3 with $\epsilon' = 1 - \sqrt{1-\epsilon}$ successively on the levels $f_u, (1-\epsilon')f_u, (1-\epsilon')^2 f_u, \ldots$ until we reach a level $\lambda = (1-\epsilon')^i f_u \geq f_l$ for which the oracle returns a feasible solution $\mathbf{x}'$ with

$$f(\mathbf{x}') \geq (1-\epsilon')\lambda = (\sqrt{1-\epsilon})^{i+1} f_u.$$

From running the oracle on the previous level $f_u(1-\epsilon')^{i-1}$, we know that $f(\mathbf{x}) \leq f(\mathbf{x}_{opt}) < (\sqrt{1-\epsilon})^{i-1} f_u$ for all $\mathbf{x} \in$ polytope$(\mathcal{F})$, where $\mathbf{x}_{opt}$ denotes the optimal solution. Therefore,

$$(\sqrt{1-\epsilon})^{i+1} f_u \leq f(\mathbf{x}') \leq f(\mathbf{x}_{opt}) < (\sqrt{1-\epsilon})^{i-1} f_u, \qquad \text{and hence}$$

$$(1-\epsilon) f(\mathbf{x}_{opt}) < f(\mathbf{x}') \leq f(\mathbf{x}_{opt}).$$

So the solution $\mathbf{x}'$ gives a $(1-\epsilon)$-approximation to the optimum $\mathbf{x}_{opt}$. In the process, we run the approximate nonlinear separation oracle at most $\log\left(\frac{f_u}{f_l}\right)/\log\frac{1}{1-\epsilon'}$ times, and each run makes $\frac{1}{4}\log\left(\frac{s_{\max}}{s_{\min}}\right)/\log\frac{1}{1-\epsilon'}$ queries to the linear oracle, so the algorithm makes at most $\frac{1}{4}\log\left(\frac{s_{\max}}{s_{\min}}\right)\log\left(\frac{f_u}{f_l}\right)/\left(\frac{1}{2}\log\frac{1}{1-\epsilon}\right)^2$ queries to the oracle for the linear problem (1). Finally, since $\log\frac{1}{1-\epsilon} \geq \epsilon$ for $\epsilon \in [0,1)$, we get the desired bound for the total number of queries. □

4. The nonconvex threshold objective with approximate linear oracle

In this section, we show that a δ-approximate oracle to problem (1) yields an efficient approximation algorithm to the nonconvex problem (2). As in Section 3, we first check whether the optimal solution has zero variance and if not, proceed with the algorithm and analysis below.

We first prove several geometric lemmas that enable us to derive the approximation guarantees later.

Lemma 5 (Geometric lemma). *Consider two objective function values $\lambda^* > \lambda$ and points $(m^*, s^*) \in L_{\lambda^*}$, $(m, s) \in L_\lambda$ in the positive orthant 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*

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

Proof. Suppose the slope of the tangents is $(-a)$, where $a > 0$. Then the y-intercepts of the two tangent lines satisfy

$$b = s + am, \qquad b^* = s^* + am^*.$$

In addition, since the points (m, s) and (m^*, s^*) lie on the level sets L_λ, L_{λ^*}, they satisfy

$$t - m = \lambda\sqrt{s}, \qquad t - m^* = \lambda^*\sqrt{s^*}.$$

Since the first line is tangent at (m, s) to the parabola $y = (\frac{t-x}{\lambda})^2$, the slope equals the first derivative at this point, $-\frac{2(t-x)}{\lambda^2}|_{x=m} = -\frac{2(t-m)}{\lambda^2} = -\frac{2\lambda\sqrt{s}}{\lambda^2} = -\frac{2\sqrt{s}}{\lambda}$, so the absolute value of the slope is $a = \frac{2\sqrt{s}}{\lambda}$. Similarly the absolute value of the slope also satisfies $a = \frac{2\sqrt{s^*}}{\lambda^*}$, therefore

$$\sqrt{s^*} = \frac{\lambda^*}{\lambda}\sqrt{s}.$$

Note that for $\lambda^* > \lambda$, this means that $s^* > s$. From here, we can represent the difference $m - m^*$ as

$$m - m^* \;=\; (t - m^*) - (t - m) = \lambda^*\sqrt{s^*} - \lambda\sqrt{s} = \frac{(\lambda^*)^2}{\lambda}\sqrt{s} - \lambda\sqrt{s} = \Big[\Big(\frac{\lambda^*}{\lambda}\Big)^2 - 1\Big]\lambda\sqrt{s}.$$

Substituting the slope $a = \frac{2\sqrt{s}}{\lambda}$ in the tangent line equations, we get

$$\begin{aligned}
b - b^* &= s + \frac{2\sqrt{s}}{\lambda}m - s^* - \frac{2\sqrt{s}}{\lambda}m^* \\
&= s - \Big(\frac{\lambda^*}{\lambda}\Big)^2 s + \frac{2\sqrt{s}}{\lambda}(m - m^*) \\
&= s - \Big(\frac{\lambda^*}{\lambda}\Big)^2 s + \frac{2\sqrt{s}}{\lambda}\lambda\sqrt{s}\Big[\Big(\frac{\lambda^*}{\lambda}\Big)^2 - 1\Big] \\
&= s - \Big(\frac{\lambda^*}{\lambda}\Big)^2 s + 2s\Big[\Big(\frac{\lambda^*}{\lambda}\Big)^2 - 1\Big] \\
&= s\Big[\Big(\frac{\lambda^*}{\lambda}\Big)^2 - 1\Big] = s^*\Big[1 - \Big(\frac{\lambda}{\lambda^*}\Big)^2\Big],
\end{aligned}$$

as desired. □

The next lemma builds intuition as well as helps in the analysis of the algorithm. It shows that we can approximate the optimal solution well if we know the optimal weight vector to use with the available approximate oracle for problem (1).

Lemma 6. *Suppose we have a δ-approximate linear oracle for optimizing over the feasible polytope$(\mathcal{F})$ 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.

Proof. Denote the projection of the optimal point $\mathbf{x}^*$ on the plane by $(m^*, s^*) = (\boldsymbol{\mu}^T\mathbf{x}^*, \boldsymbol{\tau}^T\mathbf{x}^*)$. We apply the linear oracle with respect to the slope $(-a)$ of the tangent to the level set L_{λ^*} at (m^*, s^*). The value of the linear objective at the optimum is $b^* = s^* + am^*$, which is the y-intercept of the tangent line. The linear oracle returns a δ-approximate solution, that is a solution on a parallel line with y-intercept $b \leq \delta b^*$. Suppose the original (nonlinear) objective value at the returned solution is lower-bounded by λ, that is it lies on a line tangent to L_λ (See Figure 1(b)). From Lemma 5, we have $b - b^* = s^*\left[1 - \left(\frac{\lambda}{\lambda^*}\right)^2\right]$, therefore

$$\left(\frac{\lambda}{\lambda^*}\right)^2 = 1 - \frac{b-b^*}{s^*} = 1 - \left(\frac{b-b^*}{b^*}\right)\frac{b^*}{s^*} \geq 1 - \delta\frac{b^*}{s^*}. \quad (5)$$

Recall that $b^* = s^* + m^*\frac{2\sqrt{s^*}}{\lambda^*}$ and $m^* \leq (1-\epsilon)t$, then

$$\begin{aligned} \frac{b^*}{s^*} &= 1 + \frac{2m^*}{\lambda^*\sqrt{s^*}} \\ &= 1 + \frac{2m^*}{t - m^*} \leq 1 + \frac{2m^*}{\frac{\epsilon}{1-\epsilon}m^*} \\ &= 1 + \frac{2(1-\epsilon)}{\epsilon} = \frac{2-\epsilon}{\epsilon}. \end{aligned}$$

Together with Eq. (5), this gives a $\sqrt{1-\delta\frac{2-\epsilon}{\epsilon}}$-approximation factor to the optimal.

On the other hand, setting $\epsilon = \sqrt{\delta}$ gives the approximation factor $\sqrt{1-\delta\frac{2-\sqrt{\delta}}{\sqrt{\delta}}} = 1 - \sqrt{\delta}$. □

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 7. *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 $\frac{b}{b_1} \leq \frac{1}{1-\xi^2}$.*

Proof. The equation of the level set L_λ is $y = (\frac{t-x}{\lambda})^2$ so the slope at a point $(m,s) \in L_\lambda$ is given by the derivative at $x = m$, that is $-\frac{2(t-m)}{\lambda^2} = -\frac{2\sqrt{s}}{\lambda}$. So, the slope of the tangent to the level set L_λ at point (m^*, s^*) is $-a = -\frac{2\sqrt{s^*}}{\lambda}$. Similarly the slope of the tangent at (m,s) is $-a(1+\xi) = -\frac{2\sqrt{s}}{\lambda}$. Therefore, $\sqrt{s} = (1+\xi)\sqrt{s^*}$, or equivalently $(t-m) = (1+\xi)(t-m^*)$.

Since b, b_1 are intercepts with the y-axis, of the lines with slopes $-a(1+\xi) = -\frac{2\sqrt{s}}{\lambda}$ containing the points (m^*, s^*), (m,s) respectively, we have

$$\begin{aligned} b_1 &= s + \frac{2\sqrt{s}}{\lambda} m = \frac{t^2 - m^2}{\lambda^2} \\ b &= s^* + (1+\xi)\frac{2\sqrt{s^*}}{\lambda} m^* = \frac{t - m^*}{\lambda^2}(t + m^* + 2\xi m^*). \end{aligned}$$

Therefore

$$\begin{aligned} \frac{b}{b_1} &= \frac{(t-m^*)(t+m^*+2\xi m^*)}{(t-m)(t+m)} = \frac{1}{1+\xi}\frac{t+m^*+2\xi m^*}{t+m} = \frac{1}{1+\xi}\frac{t+(1+2\xi)m^*}{(1-\xi)t+(1+\xi)m^*} \\ &\le \frac{1}{1+\xi}\left(\frac{1}{1-\xi}\right) = \frac{1}{1-\xi^2}, \end{aligned}$$

where we use $m = t - (1+\xi)(t-m^*)$ from above and the last inequality follows by Lemma 8. □

Lemma 8. *For any* $q, r > 0$, $\frac{q+(1+2\xi)r}{(1-\xi)q+(1+\xi)r} \le \frac{1}{1-\xi}$.

Proof. This follows from the fact that $\frac{1+2\xi}{1+\xi} \le \frac{1}{1-\xi}$ for $\xi \in [0,1)$. □

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

Lemma 9. *Suppose that we use an approximately optimal weight vector with a* δ*-approximate linear oracle (1) for solving the nonconvex threshold problem (2). In particular, suppose the weight vector (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 nonconvex threshold problem (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* $\boldsymbol{\mu}^T \mathbf{x}^* \le (1-\epsilon)t$.

Proof. Suppose the optimal solution is (m^*, s^*) and it lies on the optimal level set λ^*. Let the slope of the tangent to the level set boundary at the optimal solution be $(-a)$. We apply our δ-approximation linear oracle with respect to slope that is $(1+\xi)$ times the optimum slope $(-a)$. Suppose the resulting black box solution lies on the line with y-intercept b_2, and the true optimum lies on the line with y-intercept b'. We know $b' \in [b_1, b]$, where b_1 and b are the y-intercepts of the lines with slope $-(1+\xi)a$ that are tangent to L_{λ^*} and pass through (m^*, s^*) respectively. Then we have $\frac{b_2}{b} \le \frac{b_2}{b'} \le \delta$.

Furthermore, by Lemma 7 we have $\frac{b}{b_1} \le \frac{1}{1-\xi^2}$.

On the other hand, from Lemma 5, $b_2 - b_1 = s[1 - (\frac{\lambda_2}{\lambda^*})]$, where λ_2 is the smallest possible objective function value along the line with slope $-a(1+\xi)$ and y-intercept b_2, in other words the smallest possible objective function value that the solution returned by the approximate linear oracle may have; (m, s) is the tangent point of the line with slope $-(1+\xi)a$, tangent to L_{λ^*}.

Therefore, applying the above inequalities, we get

$$\left(\frac{\lambda_2}{\lambda^*}\right)^2 = 1 - \frac{b_2 - b_1}{s} = 1 - \frac{b_2 - b_1}{b_1}\frac{b_1}{s} = 1 - \left(\frac{b_2}{b}\frac{b}{b_1} - 1\right)\frac{b_1}{s} \geq 1 - \left(\frac{\delta}{1-\xi^2} - 1\right)\frac{2-\epsilon}{\epsilon},$$

where $\frac{b_1}{s} \leq \frac{2-\epsilon}{\epsilon}$ follows as in the proof of Lemma 6. The result follows. □

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.

Theorem 10. *Suppose we have a δ-approximation linear oracle for problem (1). Then, the nonconvex threshold problem (2) has a $\sqrt{1 - \left[\frac{\delta-(1-\epsilon^2/4)}{(2+\epsilon)\epsilon/4}\right]}$ -approximation algorithm that calls the linear oracle a logarithmic in the input and polynomial in $\frac{1}{\epsilon}$ number of times, assuming the optimal solution to (2) satisfies $\boldsymbol{\mu}^T\mathbf{x}^* \leq (1-\epsilon)t$.*

Proof. The algorithm applies the linear approximation oracle with respect to a small number of linear functions, and chooses the best resulting solution. In particular, suppose the optimal slope (tangent to the corresponding level set at the optimal solution point) lies in the interval $[L, U]$ (for lower and upper bound). We find approximate solutions with respect to the slopes $L, L(1+\xi), L(1+\xi)^2, ..., L(1+\xi)^k \geq U$, namely we apply the approximate linear oracle $\frac{\log(U/L)}{\log(1+\xi)}$ times, where $\xi = \frac{\epsilon^3}{2(1+\epsilon^3)}$. With this, we are certain that the optimal slope will lie in some interval $[L(1+\xi)^i, L(1+\xi)^{i+1}]$ and by Lemma 9 the solution returned by the linear oracle with respect to slope $L(1+\xi)^{i+1}$ will give a $\sqrt{1 - \left[\frac{\delta}{1-\xi^2} - 1\right]\frac{2-\epsilon}{\epsilon}}$- approximation to our non-linear objective function value. Since we are free to choose ξ, setting it to $\xi = \epsilon/2$ gives the desired number of queries.

We conclude the proof by noting that we can take L to be the slope tangent to the corresponding level set at (m_L, s_L) where s_L is the minimum positive component of the variance vector and $m_L = t(1-\epsilon)$. Similarly let U be the slope tangent at (m_U, s_U) where $m_U = 0$ and s_U is the sum of components of the variance vector. □

Note that when $\delta = 1$, namely 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-logarithmic 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 t so it is

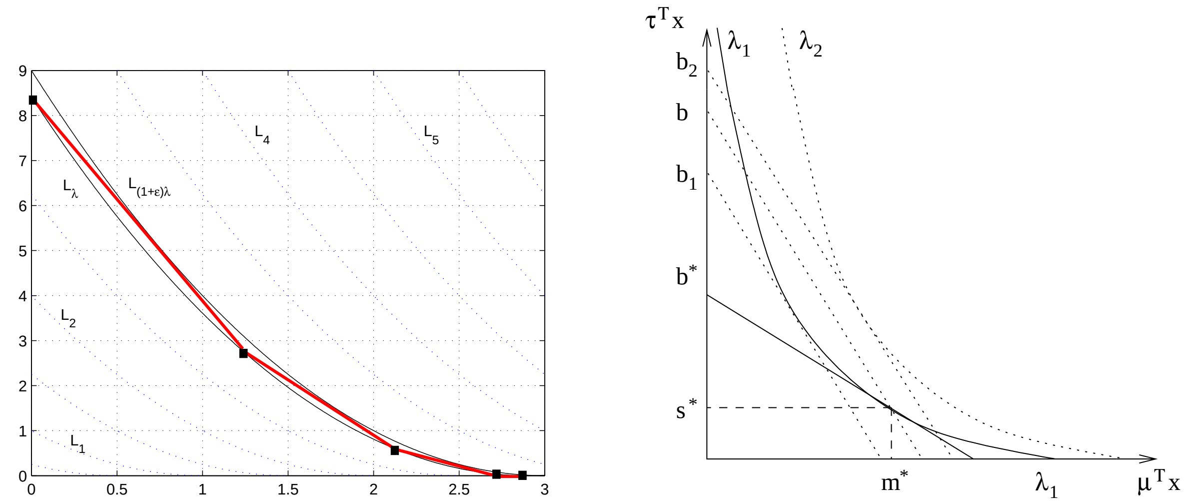

Figure 2: *(left)* Level sets and approximate nonlinear separation oracle for the projected non-convex (stochastic) objective $f(\mathbf{x}) = \boldsymbol{\mu}^T\mathbf{x} + c\sqrt{\boldsymbol{\tau}^T\mathbf{x}}$ on the $span(\boldsymbol{\mu}, \boldsymbol{\tau})$-plane. *(right)* Approximating the objective value λ_1 of the optimal solution (m^*, s^*).

weaker than our previous algorithm, which had no such requirement. This is expected, since of course this algorithm is cruder, taking a single geometric progression of linear functions rather than tailoring the linear oracle applications to the objective function value that it is searching for, as in the case of the nonlinear separation oracle that the previous algorithm from Section 3 is based on.

5. An oracle-LTAS for the nonconvex risk objective with an exact oracle

In this section we present an oracle-logarithmic time approximation scheme for the nonconvex problem (3), using an exact oracle for solving the underlying problem (1).

The projected level sets of the objective function $f(\mathbf{x}) = \boldsymbol{\mu}^T\mathbf{x} + c\sqrt{\boldsymbol{\tau}^T\mathbf{x}}$ onto the $span(\boldsymbol{\mu}, \boldsymbol{\tau})$ plane are again parabolas, though differently arranged and the analysis in the previous sections does not apply. Following the same techniques however, we can derive similar approximation algorithms, which construct an approximate nonlinear separation oracle from the linear one and apply it appropriately a small number of times.

To do this, we first need to decide which linear segments to approximate a level set with and how many they are. In particular we want to fit as few segments as possible with endpoints on the level set L_λ, entirely contained in the nonlinear band between L_λ and $L_{(1+\epsilon)\lambda}$ (over the range $m = \boldsymbol{\mu}^T\mathbf{x} \in [0, \lambda]$, $s = \boldsymbol{\tau}^T\mathbf{x} \in [0, \lambda^2]$). Geometrically, the optimal choice of segments starts from one endpoint of the level set L_λ and repeatedly draws tangents to the level set $L_{(1+\epsilon)\lambda}$, as shown in Figure 2.

We first show that the tangent-segments to $L_{(1+\epsilon)\lambda}$ starting at the endpoints of L_λ are sufficiently long.

Lemma 11. *Consider points* (m_1, s_1) *and* (m_2, s_2) *on* L_λ *with* $0 \leq m_1 < m_2 \leq \lambda$ *such that the segment with these endpoints is tangent to* $L_{(1+\epsilon)\lambda}$ *at point* $\alpha(m_1, s_1) + (1-\alpha)(m_2, s_2)$*. Then* $\alpha = \frac{c^2}{4}\frac{s_1 - s_2}{(m_2-m_1)^2} - \frac{s_2}{s_1-s_2}$ *and the objective value at the tangent point is* $\left[\frac{c^2}{4}\frac{s_1-s_2}{m_2-m_1} + s_2\frac{m_2-m_1}{s_1-s_2} + m_2\right]$.

Proof. Let $\bar{f} : \mathbb{R}^2 \to \mathbb{R}$, $\bar{f}(m, s) = m + c\sqrt{s}$ be the projection of the objective $f(\mathbf{x}) = \boldsymbol{\mu}^T\mathbf{x} + c\sqrt{\boldsymbol{\tau}^T\mathbf{x}}$. The objective values along the segment with endpoints (m_1, s_1), (m_2, s_2) are given by

$$h(\alpha) = \alpha\bar{f}(m_1, s_1) + (1-\alpha)\bar{f}(m_2, s_2) = \alpha(m_1 - m_2) + m_2 + c\sqrt{\alpha(s_1 - s_2) + s_2},$$

for $\alpha \in [0, 1]$. The point along the segment with maximum objective value (that is, the tangent point to the minimum level set bounding the segment) is found by setting the derivative $h'(\alpha) = m_1 - m_2 + c\frac{s_1-s_2}{2\sqrt{\alpha(s_1-s_2)+s_2}}$, to zero:

$$\begin{aligned}
& m_2 - m_1 = c\frac{s_1 - s_2}{2\sqrt{\alpha(s_1 - s_2) + s_2}} \\
\Leftrightarrow \quad & \sqrt{\alpha(s_1 - s_2) + s_2} = c\frac{s_1 - s_2}{2(m_2 - m_1)} \\
\Leftrightarrow \quad & \alpha(s_1 - s_2) + s_2 = c^2\frac{(s_1 - s_2)^2}{4(m_2 - m_1)^2} \\
\Leftrightarrow \quad & \alpha(s_1 - s_2) = c^2\frac{(s_1 - s_2)^2}{4(m_2 - m_1)^2} - s_2 \\
\Leftrightarrow \quad & \alpha = c^2\frac{s_1 - s_2}{4(m_2 - m_1)^2} - \frac{s_2}{s_1 - s_2}.
\end{aligned}$$

This is a maximum, since the derivative $h'(\alpha)$ is decreasing in α. The objective value at the maximum is

$$\begin{aligned}
h(\alpha_{\max}) &= \alpha_{\max}(m_1 - m_2) + m_2 + c\sqrt{\alpha_{\max}(s_1 - s_2) + s_2} \\
&= \left[c^2\frac{s_1 - s_2}{4(m_2 - m_1)^2} - \frac{s_2}{s_1 - s_2}\right](m_1 - m_2) + m_2 + c^2\frac{s_1 - s_2}{2(m_2 - m_1)} \\
&= -\frac{c^2}{4}\frac{s_1 - s_2}{m_2 - m_1} - s_2\frac{m_1 - m_2}{s_1 - s_2} + m_2 + \frac{c^2}{2}\frac{s_1 - s_2}{m_2 - m_1} \\
&= \frac{c^2}{4}\frac{s_1 - s_2}{m_2 - m_1} + s_2\frac{m_2 - m_1}{s_1 - s_2} + m_2.
\end{aligned}$$

Further, since $s_1 = (\frac{\lambda - m_1}{c})^2$ and $s_2 = (\frac{\lambda - m_2}{c})^2$, their difference satisfies $s_1 - s_2 = \frac{1}{c^2}(m_2 - m_1)(2\lambda - m_1 - m_2)$, so $\frac{s_1-s_2}{m_2-m_1} = \frac{2\lambda - m_1 - m_2}{c^2}$ and the above expression for the maximum function value on the segment becomes

$$h(\alpha_{\max}) \quad = \quad \frac{c^2}{4}\frac{2\lambda - m_1 - m_2}{c^2} + \frac{c^2 s_2}{2\lambda - m_1 - m_2} + m_2 = \frac{2\lambda - m_1 - m_2}{4} + \frac{(\lambda - m_2)^2}{2\lambda - m_1 - m_2} + m_2.$$

□

Now we can show that the tangent segments at the ends of the level set L_λ are long.

Lemma 12. *Consider the endpoint $(m_2, s_2) = (\lambda, 0)$ of L_λ. Then either the single segment connecting the two endpoints of L_λ is entirely below the level set $L_{(1+\epsilon)\lambda}$, or the other endpoint of the segment tangent to $L_{(1+\epsilon)\lambda}$ is $(m_1, s_1) = (\lambda(1-4\epsilon), (\frac{4\epsilon\lambda}{c})^2)$.*

Proof. Since $0 \le m_1 < \lambda$, we can write $m_1 = \beta\lambda$ for some $\beta \in [0, 1)$. Consequently, $s_1 = (\frac{\lambda - m_1}{c})^2 = \frac{\lambda^2(1-\beta)^2}{c^2}$ and $\frac{s_1 - s_2}{m_2 - m_1} = \frac{\lambda^2(1-\beta)^2}{c^2\lambda(1-\beta)} = \frac{\lambda(1-\beta)}{c^2}$. By Lemma 11, the objective value at the tangent point is

$$\frac{c^2}{4}\frac{\lambda(1-\beta)}{c^2} + \lambda = \lambda\left(\frac{1-\beta}{4} + 1\right) = (1+\epsilon)\lambda.$$

The last equality follows by our assumption that the tangent point lies on the $L_{(1+\epsilon)\lambda}$ level set. Hence, $\beta = 1 - 4\epsilon$, so $m_1 = (1-4\epsilon)\lambda$ and $s_1 = (\frac{\lambda - m_1}{c})^2 = (\frac{4\epsilon\lambda}{c})^2$. □

Next, we characterize the segments with endpoints on L_λ that are tangent to the level set $L_{\lambda(1+\epsilon)}$.

Lemma 13. *Consider two points (m_1, s_1), (m_2, m_2) on L_λ with $0 \le m_1 < m_2 \le \lambda$ and such that the segment connecting the two points is tangent to $L_{(1+\epsilon)\lambda}$. Then the ratio $\frac{s_1}{s_2} \ge (1+2\epsilon)^2$.*

Proof. Let point (m, s) on the segment with endpoints (m_1, s_1), (m_2, m_2) be the tangent point to the level set $L_{(1+\epsilon)\lambda}$. Then the slope $\frac{s_1 - s_2}{m_1 - m_2}$ of the segment is equal to the derivative of the function $y = (\frac{(1+\epsilon)\lambda - x}{c})^2$ at $x = m$, which is $-2\frac{(1+\epsilon)\lambda - m}{c^2} = -\frac{2\sqrt{s}}{c}$. Since $\frac{s_1 - s_2}{m_1 - m_2} = \frac{s_1 - s_2}{(\lambda - m_2) - (\lambda - m_1)} = \frac{s_1 - s_2}{c(\sqrt{s_2} - \sqrt{s_1})} = -\frac{\sqrt{s_2} + \sqrt{s_1}}{c}$, equating the two expressions for the slope we get $2\sqrt{s} = \sqrt{s_2} + \sqrt{s_1}$.

On the other hand, since $(m, s) \in L_{(1+\epsilon)\lambda}$, we have

$$m = (1+\epsilon)\lambda - c\sqrt{s} = (1+\epsilon)\lambda - \frac{c\sqrt{s_2} + c\sqrt{s_1}}{2} = (1+\epsilon)\lambda - \frac{\lambda - m_2 + \lambda - m_1}{2} = \epsilon\lambda + \frac{m_1 + m_2}{2}$$

and $m = \alpha(m_1 - m_2) + m_2$ for some $\alpha \in (0, 1)$. Therefore $\alpha = \frac{1}{2} - \frac{\epsilon\lambda}{m_2 - m_1} = \frac{1}{2} - \frac{\epsilon\lambda}{c(\sqrt{s_1} - \sqrt{s_2})}$.

Next,

$$\begin{aligned} s \quad &= \quad \alpha(s_1 - s_2) + s_2 = \left[\frac{1}{2} - \frac{\epsilon\lambda}{c(\sqrt{s_1} - \sqrt{s_2})}\right](s_1 - s_2) + s_2 = \frac{s_1 - s_2}{2} - \frac{\epsilon\lambda}{c}(\sqrt{s_1} + \sqrt{s_2}) + s_2 \\ &= \quad \frac{s_1 + s_2}{2} - \frac{\epsilon\lambda}{c}(\sqrt{s_1} + \sqrt{s_2}) \end{aligned}$$

therefore using $2\sqrt{s} = \sqrt{s_2} + \sqrt{s_1}$ from above, we get two equivalent expressions for $4s$:

$$
\begin{aligned}
& 2(s_1 + s_2) - \frac{4\epsilon\lambda}{c}(\sqrt{s_1} + \sqrt{s_2}) = s_1 + s_2 + 2\sqrt{s_1 s_2} \\
\Leftrightarrow\quad & s_1 + s_2 - \frac{4\epsilon\lambda}{c}(\sqrt{s_1} + \sqrt{s_2}) - 2\sqrt{s_1 s_2} = 0 \\
\Leftrightarrow\quad & \frac{s_1}{s_2} + 1 - \frac{4\epsilon\lambda}{c\sqrt{s_2}}(\sqrt{\frac{s_1}{s_2}} + 1) - 2\sqrt{\frac{s_1}{s_2}} = 0
\end{aligned}
$$

Denote for simplicity $z = \sqrt{\frac{s_1}{s_2}}$ and $w = \frac{2\epsilon\lambda}{c\sqrt{s_2}}$, then we have to solve the following quadratic equation for z in terms of w:

$$
\begin{aligned}
& z^2 + 1 - 2w(z+1) - 2z = 0 \\
\Leftrightarrow\quad & z^2 - 2z(w+1) + 1 - 2w = 0.
\end{aligned}
$$

The discriminant of this quadratic expression is $D = (w+1)^2 - 1 + 2w = w^2 + 4w$ and its roots are $z_{1,2} = 1 + w \pm \sqrt{w^2 + 4w}$. Since $\frac{s_1}{s_2} > 1$, we choose the bigger root $z_2 = 1 + w + \sqrt{w^2 + 4w}$. Therefore since $w = \frac{2\epsilon\lambda}{c\sqrt{s_2}} \geq 0$ we have

$$
\sqrt{\frac{s_1}{s_2}} = 1 + w + \sqrt{w^2 + 4w} \geq 1 + w = 1 + \frac{2\epsilon\lambda}{c\sqrt{s_2}} \geq 1 + \frac{2\epsilon\lambda}{c\frac{\lambda}{c}} = 1 + 2\epsilon,
$$

where the last inequality follows from the fact that $\sqrt{s_2} < \sqrt{s_1} \leq \frac{\lambda}{c}$. This concludes the proof. □

The last lemma shows that each segment is sufficiently long so that overall the number of tangent segments approximating the level set L_λ is small, in particular it is polynomial in $\frac{1}{\epsilon}$ (and does not depend on the input size of the problem!). This gives us the desired approximate nonlinear separation oracle for the level sets of the objective function.

Theorem 14. *A nonlinear $(1+\epsilon)$-approximate separation oracle to any level set of the nonconvex objective $f(\mathbf{x})$ in problem (3) can be found with $(1 + \frac{\log(\frac{1}{16\epsilon^2})}{2\log(1+2\epsilon)})$ queries to the available linear oracle for solving problem (1).*

The nonlinear oracle takes as inputs λ, ϵ and returns either a feasible solution $\mathbf{x} \in \mathcal{F}$ with $f(\mathbf{x}) \leq (1+\epsilon)\lambda$ or an answer that $f(\mathbf{x}) > \lambda$ for all $\mathbf{x}$ in the polytope$(\mathcal{F})$.

Proof. Apply the available linear oracle to the slopes of the segments with endpoints on the specified level set, say L_λ, and which are tangent to the level set $L_{(1+\epsilon)\lambda}$. By Lemma 13 and Lemma 12, the y-coordinates of endpoints of these segments are given by $s_1 = (\frac{\lambda}{c})^2$, $s_2 \leq \frac{s_1}{(1+2\epsilon)^2}$, $s_3 \leq \frac{s_1}{(1+2\epsilon)^4}$,... , $s_k \leq \frac{s_1}{(1+2\epsilon)^{2(k-1)}}$, $s_{k+1} = 0$, where $s_k = (\frac{4\epsilon\lambda}{c})^2$, so $k = 1 + \log(\frac{1}{16\epsilon^2})/2\log(1+2\epsilon)$, which is precisely the number of segments we use and the result follows. □

Finally, based on the approximate non-linear separation oracle from Theorem 14, we can now easily solve the nonconvex problem (3) running the oracle on a geometric progression from the objective function range. We again need to bound the optimal value of the objective function. For a lower bound, we can use $f_l = s_{min}$, the smallest positive variance component, and for an upper bound take $f_u = nm_{max} + c\sqrt{ns_{max}}$, where m_{max} and s_{max} are the largest components of the mean and variance vectors respectively.

Theorem 15. *There is a oracle-logarithmic time approximation scheme for the nonconvex problem (3), which uses an exact oracle for solving the underlying problem (1). This algorithm returns a* $(1+\epsilon)$*-approximate solution and makes* $(1 + \frac{2}{\epsilon}\log(\frac{f_u}{f_l}))(1 + \frac{\log(\frac{1}{16\epsilon^2})}{2\log(1+2\epsilon)})$ *oracle queries, namely logarithmic in the size of input and polynomial in* $\frac{1}{\epsilon}$.

Proof. The proof is analogous to that of Theorem 4, applying the nonlinear oracle with approximation factor $\sqrt{1+\epsilon}$ for the objective function values $f_l, f_l\sqrt{1+\epsilon}, f_l\sqrt{1+\epsilon}^2, ..., f_u = f_l\sqrt{1+\epsilon}^j$, namely applying it at most $1 + 2\log(\frac{f_u}{f_l})/\log(1+\epsilon) \leq \frac{2}{\epsilon}\log(\frac{f_u}{f_l})$ times. In addition, we run the linear oracle once with weight vector equal to the vector of means, over the subset of components with zero variances and return that solution if it is better than the above. □

6. An oracle-LTAS for the nonconvex risk objective with approximate linear oracle

Suppose we have a δ-approximate linear oracle for solving problem (1). We will provide an algorithm for the nonconvex problem (3) with approximation factor $\delta(1+\epsilon)$, which invokes the linear oracle a small number of times that is logarithmic in the input-size and polynomial in $\frac{1}{\epsilon}$.

We employ the same technique of designing the algorithm and analyzing it as in Section 4 for the threshold objective function, however again due to the different objective the previous analysis does not carry through directly.

First, we show that if we can guess the optimal linear objective, given by the slope of the tangent to the corresponding level set at the optimal solution, then applying the approximate linear oracle returns an approximate solution with the same multiplicative approximation factor δ. The above statement reduces to showing the following geometric fact.

Lemma 16. *Consider levels* $0 \leq \lambda_1 < \lambda_2$ *and two parallel lines tangent to the corresponding level sets* L_{λ_1} *and* L_{λ_2} *at points* (m_1, s_1) *and* (m_2, s_2) *respectively. Further, suppose the corresponding* y*-intercepts of these lines are* b_1 *and* b_2. *Then* $\frac{b_2}{b_1} = \frac{\lambda_2 + m_2}{\lambda_1 + m_1} \geq \frac{\lambda_2}{\lambda_1}$.

Proof. The function defining a level set L_λ has the form $y = \frac{(\lambda-x)^2}{c^2}$, and thus the slope of the tangent to the level set at a point $(m, s) \in L_\lambda$ is given by the first derivative at the point, $-\frac{2(\lambda-x)}{c^2}|_{x=m} = -\frac{2(\lambda-m)}{c^2} = -\frac{2\sqrt{s}}{c}$. Therefore the equation of the tangent line is $y = -\frac{2\sqrt{s}}{c}x + b$, where

$$b = s + \frac{2\sqrt{s}}{c}m = \sqrt{s}(\sqrt{s} + \frac{2m}{c}) = \sqrt{s}(\frac{\lambda - m}{c} + \frac{2m}{c}) = \sqrt{s}(\frac{\lambda + m}{c})$$

Since the two tangents from the lemma are parallel, their slopes are equal: $-\frac{2\sqrt{s_1}}{c} = -\frac{2\sqrt{s_2}}{c}$, therefore $s_1 = s_2$ and equivalently $(\lambda_1 - m_1) = (\lambda_2 - m_2)$.

Therefore the y-intercepts of the two tangents satisfy

$$\frac{b_2}{b_1} = \frac{\sqrt{s_2}(\frac{\lambda_2+m_2}{c})}{\sqrt{s_1}(\frac{\lambda_1+m_1}{c})} = \frac{\lambda_2 + m_2}{\lambda_1 + m_1} \geq \frac{\lambda_1}{\lambda_2}.$$

The last inequality follows from the fact that $\lambda_2 > \lambda_1$ and $\lambda_1 - m_1 = \lambda_2 - m_2$ (and equality is achieved when $m_1 = \lambda_1$ and $m_2 = \lambda_2$). □

Corollary 17. *Suppose the optimal solution to the nonconvex problem (3) is (m_1, s_1) with objective value λ_1. If we can guess the slope $-a$ of the tangent to the level set L_{λ_1} at the optimal solution, then applying the approximate linear oracle for solving problem (1) with respect to that slope will give a δ-approximate solution to problem (3).*

Proof. The approximate linear oracle will return a solution (m', s') with value $b_2 = s' + am' \leq \delta b_1$, where $b_1 = s_1 + am_1$. The objective function value of (m', s') is at most λ_2, which is the value at the level set tangent to the line $y = -ax + b_2$. By Lemma 16, $\frac{\lambda_2}{\lambda_1} \leq \frac{b_2}{b_1} \leq \delta$, therefore the approximation solution has objective function value at most δ times the optimal value, QED. □

If we cannot guess the slope at the optimal solution, we would have to approximate it. The next lemma proves that if we apply the approximate linear oracle to slope that is within $(1+\sqrt{\frac{\epsilon}{1+\epsilon}})$ of the optimal slope, we would still get a good approximate solution with approximation factor $\delta(1 + \epsilon)$.

Lemma 18. *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 $-a(1 + \sqrt{\frac{\epsilon}{1+\epsilon}})$ 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 $\frac{b}{b_1} \leq 1 + \epsilon$.*

Proof. Let $\xi = \sqrt{\frac{\epsilon}{1+\epsilon}}$. As established in the proof of Lemma 16, the slope of the tangent to the level set L_λ at point (m^*, s^*) is $-a = -\frac{2\sqrt{s^*}}{c}$. Similarly the slope of the tangent at (m, s) is $-a(1 + \xi) = -\frac{2\sqrt{s}}{c}$. Therefore, $\sqrt{s} = (1 + \xi)\sqrt{s^*}$, or equivalently $(\lambda - m) = (1 + \xi)(\lambda - m^*)$.

Since b, b_1 are intercepts with the y-axis, of the lines with slopes $-a(1+\xi) = -\frac{2\sqrt{s}}{c}$ containing the points (m^*, s^*), (m, s) respectively, we have

$$\begin{aligned} b_1 &= s + \frac{2\sqrt{s}}{c}m = \frac{\lambda^2 - m^2}{c^2} \\ b &= s^* + (1+\xi)\frac{2\sqrt{s^*}}{c}m^* = \frac{\lambda - m^*}{c^2}(\lambda + m^* + 2\xi m^*). \end{aligned}$$

Therefore

$$\frac{b}{b_1} = \frac{(\lambda - m^*)(\lambda + m^* + 2\xi m^*)}{(\lambda - m)(\lambda + m)} = \frac{1}{1+\xi}\frac{\lambda + m^* + 2\xi m^*}{\lambda + m} \leq \frac{1}{1+\xi}\left(\frac{1}{1-\xi}\right) = \frac{1}{1-\xi^2} = 1 + \epsilon,$$

where the last inequality follows by Lemma 19. □

Lemma 19. *Following the notation of Lemma 18,* $\frac{\lambda + m^* + 2\xi m^*}{\lambda + m} \leq \frac{1}{1-\xi}$.

Proof. Recall from the proof of Lemma 18 that $(\lambda - m) = (1+\xi)(\lambda - m^*)$, therefore $m = \lambda - (1+\xi)(\lambda - m^*) = -\xi\lambda + (1+\xi)m^*$. Hence,

$$\frac{\lambda + m^* + 2\xi m^*}{\lambda + m} = \frac{\lambda + (1+2\xi)m^*}{(1-\xi)\lambda + (1+\xi)m^*} = \frac{\frac{\lambda}{m^*} + (1+2\xi)}{(1-\xi)\frac{\lambda}{m^*} + (1+\xi)} \leq \frac{1}{1-\xi},$$

since $\frac{1+2\xi}{1+\xi} \leq \frac{1}{1-\xi}$ for $\xi \in [0, 1)$. □

A corollary from Lemma 16 and Lemma 18 is that applying the linear oracle with respect to slope that is within $(1 + \sqrt{\frac{\epsilon}{1+\epsilon}})$ times of the optimal slope yields an approximate solution with objective value within $(1+\epsilon)\delta$ times of the optimal.

Lemma 20. *Suppose the optimal solution to the nonconvex problem (3) is* (m^*, s^*) *with objective value* λ *and the slope of the tangent to the level set* L_λ *at it is* $-a$*. Then running the* δ*-approximate oracle for solving problem (1) with respect to slope that is in* $[-a, -a(1 + \sqrt{\frac{\epsilon}{1+\epsilon}})]$ *returns a solution to (3) with objective function value no greater than* $(1+\epsilon)\delta\lambda$*.*

Proof. Suppose the optimal solution with respect to the linear objective specified by slope $-a(1 + \sqrt{\frac{\epsilon}{1+\epsilon}})$ has value $b' \in [b_1, b]$, where b_1, b are the y-intercepts of the lines with that slope, tangent to L_λ and passing through (m^*, s^*) respectively (See Figure 2-right). Then applying the δ-approximate linear oracle to the same linear objective returns solution with value $b_2 \geq \delta b'$. Hence $\frac{b_2}{b} \leq \frac{b_2}{b'} \leq \delta$.

On the other hand, the approximate solution returned by the linear oracle has value of our original objective function equal to at most λ_2, where L_{λ_2} is the level set tangent to the line on which the approximate solution lies. By Lemma 16, $\frac{\lambda_2}{\lambda} \leq \frac{b_2}{b_1} = \frac{b_2}{b}\frac{b}{b_1} \leq \delta(1+\epsilon)$, where the last inequality follows by Lemma 18 and the above bound on $\frac{b_2}{b}$. □

Finally, we are ready to state our theorem for solving the nonconvex problem (3). The theorem says that there is an algorithm for this problem with essentially the same approximation factor as the underlying problem (1), which makes only logarithmically many calls to the latter.

Theorem 21. *Suppose we have a δ-approximate oracle for solving problem (1). The nonconvex problem (3) can be approximated to a multiplicative factor of $\delta(1+\epsilon)$ by calling the above oracle logarithmically many times in the input size and polynomially many times in $\frac{1}{\epsilon}$.*

Proof. We use the same type of algorithm as in Theorem 10: apply the available approximate linear oracle on a geometric progression of weight vectors (slopes), determined by the lemmas above. In particular, apply it to slopes $U, (1+\xi)U, ..., (1+\xi)^i U = L$, where $\xi = \sqrt{\frac{\epsilon}{1+\epsilon}}$, L is a lower bound for the optimal slope and U is an upper bound for it. For each approximate feasible solution obtained, compute its objective function value and return the solution with minimum objective function value. By Lemma 20, the value of the returned solution would be within $\delta(1+\epsilon)$ of the optimal.

Note that it is possible for the optimal slope to be 0: this would happen when the optimal solution satisfies $m^* = \lambda$ and $s^* = 0$. We have to handle this case differently: run the linear oracle just over the subset of components with zero variance-values, to find the approximate solution with smallest m. Return this solution if its value is better than the best solution among the above.

It remains to bound the values L and U. We established earlier that the optimal slope is given by $\frac{2\sqrt{s^*}}{c}$, where s^* is the variance of the optimal solution. Among the solutions with nonzero variance, the variance of a feasible solution is at least s_{min}, the smallest possible nonzero variance of a single element, and at most $(\lambda_{max})^2 \leq (nm_{max} + c\sqrt{ns_{max}})^2$, where m_{max} is the largest possible mean of a single element and s_{max} is the largest possible variance of a single element (assuming that a feasible solution uses each element in the ground set at most once). Thus, set $U = -\frac{2\sqrt{s_{min}}}{c}$ and $L = -\frac{2(nm_{max}+c\sqrt{ns_{max}})}{c}$ □

7. Conclusion

We have presented efficient approximation schemes for a broad class of stochastic problems that incorporate risk. Our algorithms are independent of the fixed feasible set and use solutions for the underlying deterministic problems as oracles for solving the stochastic counterparts. As such they apply to very general combinatorial and discrete, as well as continuous settings.

From a practical point of view, it is of interest to consider correlations between the components of the stochastic weight vector. We remark that in graph problems, our results can immediately extend to some realistic partial correlation models. We leave a study of correlations for future work.

Our models in this paper assume that the stochastic weight vector is unknown at the time of the solution. An interesting direction would be to extend this work to an online setting where one gains information on the weights with time, and is prompted to find an adaptive solution or optimal policy.

Appendix

7.1. Stochastic value-at-risk objective

In this section we show how the value-at-risk objective reduces to the problem of minimizing a linear combination of mean and standard deviation. We first establish the equivalence under normal distributions, and then show a reduction for arbitrary distributions using Chebyshev's bound.

Lemma 22. *The stochastic value-at-risk problem*

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

for a given probability p *is equivalent to the nonconvex problem*

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

with $c = \Phi^{-1}(p)$*, when the element weights come from independent normal distributions.*

Proof. As before $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal random variable $N(0,1)$, and $\Phi^{-1}(\cdot)$ denotes its inverse. For normally distributed weights $\mathbf{W}$ we have

$$\begin{aligned} & \Pr(\mathbf{W}^T\mathbf{x} \le t) \ge p \\ \Leftrightarrow \quad & \Pr\left(\frac{\mathbf{W}^T\mathbf{x} - \boldsymbol{\mu}^T\mathbf{x}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}}} \le \frac{t - \boldsymbol{\mu}^T\mathbf{x}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}}}\right) \ge p \\ \Leftrightarrow \quad & \Phi(\frac{t - \boldsymbol{\mu}^T\mathbf{x}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}}}) \ge p \\ \Leftrightarrow \quad & \frac{t - \boldsymbol{\mu}^T\mathbf{x}}{\sqrt{\boldsymbol{\tau}^T\mathbf{x}}} \ge \Phi^{-1}(p) \\ \Leftrightarrow \quad & t \ge \boldsymbol{\mu}^T\mathbf{x} + \Phi^{-1}(p)\sqrt{\boldsymbol{\tau}^T\mathbf{x}}. \end{aligned}$$

Note, the stochastic value-at-risk problem is minimizing over both t and $\mathbf{x}$. Therefore the smallest threshold t is equal to the minimum of $\boldsymbol{\mu}^T\mathbf{x}+c\sqrt{\boldsymbol{\tau}^T\mathbf{x}}$ over the feasible set $\mathbf{x} \in \mathcal{F}$, where the constant $c = \Phi^{-1}(p)$. □

For arbitrary distributions, we can apply Chebyshev's bound $\Pr(\mathbf{W}^T\mathbf{x} \geq \boldsymbol{\mu}^T\mathbf{x} + c\sqrt{\boldsymbol{\tau}^T\mathbf{x}}) \leq \frac{1}{c^2}$, or equivalently $\Pr(\mathbf{W}^T\mathbf{x} < \boldsymbol{\mu}^T\mathbf{x} + c\sqrt{\boldsymbol{\tau}^T\mathbf{x}}) > 1 - \frac{1}{c^2}$. Taking $c = \frac{1}{\sqrt{1-p}}$ gives the inequality $\Pr(\mathbf{W}^T\mathbf{x} < \boldsymbol{\mu}^T\mathbf{x} + c\sqrt{\boldsymbol{\tau}^T\mathbf{x}}) > p$. This shows the following lemma:

Lemma 23. *The stochastic value-at-risk problem with arbitrary distributions reduces to:*

$$\begin{array}{ll} \textit{minimize} & \boldsymbol{\mu}^T\mathbf{x} + \dfrac{1}{\sqrt{1-p}}\sqrt{\boldsymbol{\tau}^T\mathbf{x}} \\ \textit{subject to} & \mathbf{x} \in \mathcal{F} \end{array}$$

In particular, the optimal value of the above concave minimization problem will provide an upper bound of the minimum threshold t in the value-at-risk problem with given probability p.

We remark that in the absence of more information on the distributions, other than their means and standard deviations, this is the best one can do to solve the value-at-risk problem.

For an illustration of the difference between the above lemmas, consider again the following shortest path problem.

Example 24. *Suppose we need to reach the airport by a certain time. We want to find the minimum time (and route) that we need to allocate for our trip so as to arrive on time with probability at least $p = .95$. (That is, how close can we cut it to the deadline and not be late?) If we know that the travel times on the edges are normally distributed, the minimum time equals $\min_{\mathbf{x}\in\mathcal{F}} \boldsymbol{\mu}^T\mathbf{x} + 1.645\sqrt{\boldsymbol{\tau}^T\mathbf{x}}$, since $\Phi^{-1}(.95) = 1.645$. On the other hand, if we had no information about the distributions, we should instead allocate the upper bound $\min_{\mathbf{x}\in\mathcal{F}} \boldsymbol{\mu}^T\mathbf{x} + 4.5\sqrt{\boldsymbol{\tau}^T\mathbf{x}}$, since $\frac{1}{\sqrt{1-0.95}} \approx 4.5$ (which still guarantees that we would arrive with probability at least 95%).*

Acknowledgments

The author thanks Dimitris Bertsimas, Shuchi Chawla, Nick Harvey, Michel Goemans, David Karger, David Kempe, Christos Papadimitriou, Madhu Sudan and John Tsitsiklis for valuable suggestions.

References

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.

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

D. Bertsimas and M. Sim. Robust discrete optimization and network flows. *Mathematical Programming*, 98:49–71, 2003.

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.

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

A. Goel, K. Ramakrishnan, D. Kataria, and D. Logothetis. Efficient computation of delay-sensitive routes from one source to all destinations. In *Proceedings of IEEE Infocom*, 2001.

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.

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

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

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

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.

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.

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.

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

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

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. ISBN 978-0-898716-24-5.

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

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