Approximation of Ellipsoids Using Bounded Uncertainty Sets *

André Chassein †

Fachbereich Mathematik, Technische Universität Kaiserslautern, Germany

Abstract

In this paper, we discuss the problem of approximating ellipsoid uncertainty sets with bounded (gamma) uncertainty sets. Robust linear programs with ellipsoid uncertainty lead to quadratically constrained programs, whereas robust linear programs with bounded uncertainty sets remain linear programs which are generally easier to solve.

We call a bounded uncertainty set an inner approximation of an ellipsoid if it is contained in it. We consider two different inner approximation problems. The first problem is to find a bounded uncertainty set which sticks close to the ellipsoid such that a shrank version of the ellipsoid is contained in it. The approximation is optimal if the required shrinking is minimal. In the second problem, we search for a bounded uncertainty set within the ellipsoid with maximum volume. We present how both problems can be solved analytically by stating explicit formulas for the optimal solutions of these problems.

Further, we present in a computational experiment how the derived approximation techniques can be used to approximate shortest path and network flow problems which are affected by ellipsoidal uncertainty.

Keywords Robust optimization; Ellipsoid Uncertainty Set; Bounded Uncertainty Sets; Approximation.

1 Introduction

The idea behind robust optimization is to replace uncertain parameter values with uncertainty sets. A robust solution needs to take into account all possible parameter realizations, which may arise from the uncertainty set. Different uncertainty sets have been proposed in the literature. Ben-Tal and Nemirovski propose in [2] to use ellipsoid uncertainty sets. Ellipsoids are natural to represent uncertainty, since many important probability distributions are elliptical. A probability distribution is elliptical if the superlevel sets of their density function

*Effort sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA8655-13-1-3066. The U.S Government is authorized to reproduce and distribute reprints for Governmental purpose notwithstanding any copyright notation thereon.

†Corresponding author. Email: chassein@mathematik.uni-kl.de

are ellipsoids, i.e., the most likely realizations of the corresponding random variable form an ellipsoid. A beneficial property of ellipsoid uncertainty sets is the tractability of the robust counter part for many important problem classes. For more information, we refer to the book of Ben-Tal, Nemirovski, and Ghaoui [1]. Another common uncertainty set is the bounded uncertainty set proposed by Bertsimas and Sim in [4]. These sets are frequently used in the robust optimization community since they own desirable properties. The most important benefit is the fact that the robust counterpart of a linear programming problem with respect to a bounded uncertainty set is a linear programming problem again.

1.1 Contribution

In this paper, we study the relation between ellipsoid and bounded uncertainty sets. The goal is to find a bounded uncertainty set such that the solution that is obtained by the robust counterpart with respect to the bounded uncertainty set is close to the solution of the robust counterpart with respect to ellipsoid uncertainty. To reach this goal we select a bounded uncertainty set that gives a *good* approximation of the ellipsoid. As it is a priori unclear what specifies a good approximation, we distinguish two different ways to approximate an ellipsoid. In both cases, we search for bounded uncertainty sets which are contained in the ellipsoid (so-called inner approximations).

The first way is to find a bounded uncertainty set such that a shrank version of the ellipsoid is contained in the bounded uncertainty set. The goal is to find a set such that the shrank ellipsoid is as large as possible. We derive an explicit formula for the optimal solution of this approximation problem. Note that a similar problem was already analyzed in [3]. Bertsimas, Pachamanova, and Sim define in [3] a norm such that the unit ball of this norm is is equal to a bounded uncertainty set. Further, they show how well this norm approximates the euclidean norm, i.e., how well bounded uncertainty sets can approximate the euclidean unit ball. We generalize their result by approximating general ellipsoids.

The second way is to find a bounded uncertainty set within the ellipsoid with maximum volume. First, we derive an explicit formula for the volume of a bounded uncertainty set. Next, we develop an explicit formula for the optimal solution of this problem which contains a constant that depends on the dimension of the underlying space. To avoid the exact computation of this constant, we derive an formula for its approximate value.

We define a shortest path and a network flow problem with ellipsoidal uncertainty to test the presented approximation techniques. The main insights of the experiments are: Replacing ellipsoid uncertainty sets with bounded uncertainty sets decreases the computation time, especially for large scale problems. Depending on the problem, the solution that is obtained by the approximate problem may be very close to the optimal solution.

1.2 Structure of the Paper

The paper is organized as follows. In Section 2, we introduce the notation used in the paper and define the two versions of the approximation problem. In Section 3 and 4, we derive explicit formulas for the optimal solutions of

the approximation problems for the case of axis-parallel ellipsoids which are centered at the origin. The topic of Section 5 is to also handle arbitrary rotated ellipsoids. To this end, we need to introduce the notion of a rotated bounded uncertainty set. In Section 6, we use shortest path and network flow problems which are affected by ellipsoidal uncertainty to test the presented approximation techniques. We summarize our findings and post future research questions in Section 7. Some technical lemmas which are used in Section 3 and 4 and their proofs are given in the Appendix.

2 Notation

We denote the index set $\{1, \ldots, n\}$ by $[n]$. The following notations are used to represent ellipsoids and bounded uncertainty sets: An *ellipsoid uncertainty set* is given by $\mathcal{E}(a^0, M) = \{x \mid (x - a^0)^T M(x - a^0) \leq 1\}$, where M is a positive semidefinite matrix and a^0 is called the center of the ellipsoid. A *bounded uncertainty set* is a polytope $\mathcal{U}(a^0, a, \Gamma) \subset \mathbb{R}^n$ that is characterized by $a^0, a \in \mathbb{R}^n$ and a budget parameter $\Gamma \in [0, n]$.

$$\mathcal{U}(a^0, a, \Gamma) = \left\{ x : \sum_{i=1}^{n} \frac{|x_i - a_i^0|}{a_i} \leq \Gamma, x_i \in [a_i^0 - a_i, a_i^0 + a_i] \ \forall i \in [n] \right\}$$

The goal is to approximate an ellipsoid $\mathcal{E}$ by a bounded uncertainty set. We start with an axis-parallel ellipsoid $\mathcal{E}(a^0, D)$, i.e., the matrix D is a diagonal matrix. Further, we assume that the center of the ellipsoid (and of the bounded uncertainty set) is 0. For short, we write $\mathcal{E}(D)$ and $\mathcal{U}(a, \Gamma)$. We present in Section 5 how both assumptions can be removed.

As already mentioned, we analyze two different types of approximation. In both problems, we try to approximate an ellipsoid uncertainty set $\mathcal{E}$ with a bounded uncertainty set which is completely contained in the ellipsoid. The first idea is to find a bounded uncertainty set which is shaped in such a way that a scaled version of $\mathcal{E}$ lies inside the bounded uncertainty set. The larger the scaling factor r, the better the approximation. The scaled version of $\mathcal{E}(D)$ is denoted by $r\mathcal{E}(D) = \{rx : x^T D x \leq 1\}$. We call this problem in the following *ratio approximation*. For a given matrix D, the ratio approximation problem is formally stated as

$$\begin{aligned} &\max_{r,a,\Gamma} \ r && \text{(AP-R)} \\ &\text{s.t. } r\mathcal{E}(D) \subset \mathcal{U}(a, \Gamma) \subset \mathcal{E}(D) \end{aligned}$$

The second problem tries to maximize the volume of the bounded uncertainty set, called *volume approximation*. The formal problem definition is given by

$$\begin{aligned} &\max_{a,\Gamma} \ \text{vol}(\mathcal{U}(a, \Gamma)) && \text{(AP-V)} \\ &\text{s.t. } \mathcal{U}(a, \Gamma) \subset \mathcal{E}(D) \end{aligned}$$

3 Ratio Approximation

In this section, we discuss the problem (AP-R). The next theorem gives an explicit formula to compute the optimal solution of the ratio approximation problem.

Theorem 3.1. *Let the ellipsoid $\mathcal{E}(D) \subset \mathbb{R}^n$ be defined by the diagonal matrix $D = diag\left(\frac{1}{d_1^2}, \ldots, \frac{1}{d_n^2}\right)$. The optimal solution (r^*, Γ^*, a^*) of problem* (AP-R) *is given by*

$$\begin{aligned} r^* &= \frac{1}{\sqrt{\lfloor\sqrt{n}\rfloor + (\sqrt{n} - \lfloor\sqrt{n}\rfloor)^2}} \\ \Gamma^* &= \sqrt{n} \\ a_i^* &= r^* d_i \ \forall i \in [n] \end{aligned}$$

Proof. Symmetry allows us to reduce our analysis to the positive orthant $\mathbb{R}^n_+$, since

$$r\mathcal{E}(D) \subset \mathcal{U}(a,\Gamma) \subset \mathcal{E}(D) \Leftrightarrow (r\mathcal{E}(D) \cap \mathbb{R}^n_+) \subset (\mathcal{U}(a,\Gamma) \cap \mathbb{R}^n_+) \subset (\mathcal{E}(D) \cap \mathbb{R}^n_+).$$

Hence, we assume in the following that all elements of $\mathcal{E}(D)$ or $\mathcal{U}(a,\Gamma)$ are non-negative. We begin by rewriting the constraint $r\mathcal{E}(D) \subset \mathcal{U}(a,\Gamma)$. Note that $x \in \mathcal{U}(a,\Gamma)$ is equivalent to the following constraints

$$\begin{aligned} &0 \leq x_i \leq a_i \ \forall i \in [n] \\ &\sum_{i=1}^n \frac{x_i}{a_i} \leq \Gamma. \end{aligned}$$

Using this equivalence, we obtain

$$\begin{aligned} r\mathcal{E}(D) \subset \mathcal{U}(a,\Gamma) \Leftrightarrow &\max_{x \in r\mathcal{E}(D)} x_i \leq a_i \ \forall i \in [n] \\ &\max_{x \in r\mathcal{E}(D)} \sum_{i=1}^n \frac{x_i}{a_i} \leq \Gamma. \end{aligned}$$

Lemma A.1 and A.2 show that the optimal values of the inner maximization problems are rd_i for $i \in [n]$ and $r\sqrt{\sum_{i=1}^n \frac{d_i^2}{a_i^2}}$, respectively. Inserting these values and rearranging yields

$$\begin{aligned} r\mathcal{E}(D) \subset \mathcal{U}(a,\Gamma) \Leftrightarrow &r \leq \frac{a_i}{d_i} \ \forall i \in [n] \\ &r \leq \frac{\Gamma}{\sqrt{\sum_{i=1}^n \frac{d_i^2}{a_i^2}}}. \end{aligned}$$

Next, we analyze the constraint $\mathcal{U}(a,\Gamma) \subset \mathcal{E}(D)$. Using the definitions of both sets yields to

$$\mathcal{U}(a,\Gamma) \subset \mathcal{E}(D) \Leftrightarrow \max \sum_{i=1}^n \frac{x_i^2}{d_i^2} \leq 1 \qquad \Leftrightarrow \max \sum_{i=1}^n \frac{a_i^2}{d_i^2} y_i^2 \leq 1$$

$$\text{s.t. } 0 \le x_i \le a_i \ \forall i \in [n] \qquad \text{s.t. } 0 \le y_i \le 1 \ \forall i \in [n]$$
$$\sum_{i=1}^{n} \frac{x_i}{a_i} \le \Gamma \qquad \sum_{i=1}^{n} y_i \le \Gamma$$

We denote by $v_i = \frac{a_i^2}{d_i^2}$ the entries of vector $v \in \mathbb{R}^n$. Lemma A.3 shows that the value of the last maximization problem is equal to $||v||^{(\lfloor \Gamma \rfloor)} \left(1 - (\Gamma - \lfloor \Gamma \rfloor)^2\right) + ||v||^{(\lfloor \Gamma \rfloor + 1)}(\Gamma - \lfloor \Gamma \rfloor)^2$ where $||v||^{(k)}$ denotes the sum of the k largest entries of v. Summarizing both results, we obtain

$$\begin{aligned} r\mathcal{E}(D) \subset \mathcal{U}(a,\Gamma) &\Leftrightarrow r^2 \le v_i \ \forall i \in [n] \\ r^2 &\le \frac{\Gamma}{\sqrt{\sum_{i=1}^{n} v_i}} \\ \mathcal{U}(a,\Gamma) \subset \mathcal{E}(D) &\Leftrightarrow ||v||^{(\lfloor \Gamma \rfloor)} \left(1 - (\Gamma - \lfloor \Gamma \rfloor)^2\right) + ||v||^{(\lfloor \Gamma \rfloor + 1)}(\Gamma - \lfloor \Gamma \rfloor)^2 \le 1. \end{aligned}$$

Hence, we can rewrite problem (AP-R) by the following optimization problem

$$\begin{aligned} &\text{maximize } r \\ &\quad \text{s.t. } r^2 \le v_i \ \forall i \in [n] \\ &\qquad r^2 \le \frac{\Gamma^2}{\sum_{i=1}^{n} \frac{1}{v_i}} \\ &\qquad ||v||^{(\lfloor \Gamma \rfloor)} \left(1 - (\Gamma - \lfloor \Gamma \rfloor)^2\right) + ||v||^{(\lfloor \Gamma \rfloor + 1)}(\Gamma - \lfloor \Gamma \rfloor)^2 \le 1 \end{aligned}$$

We claim that an optimal solution exists for this problem with $v_i = v_j$ for all $i, j \in [n]$. Assume that $(\tilde{r}, \tilde{\Gamma}, \tilde{v})$ is an optimal solution of this problem. We denote by $\overline{v} = \frac{1}{n}\sum_{i=1}^{n} v_i$ the average of the components of $\tilde{v}$ and by $\hat{v} = (\overline{v}, \ldots, \overline{v})^T$ the vector containing $\overline{v}$ in every component. We claim that $(\tilde{r}, \tilde{\Gamma}, \hat{v})$ is also feasible for this problem and, hence, optimal. Note that $\tilde{r}^2 \le \min_{i \in [n]} \tilde{v}_i$ directly implies $\tilde{r}^2 \le \overline{v}$. Further, Lemma A.4 implies that

$$\frac{\tilde{\Gamma}^2}{\sum_{i=1}^{n} \frac{1}{v_i}} \le \frac{\tilde{\Gamma}^2}{\sum_{i=1}^{n} \frac{1}{\overline{v}}}.$$

Finally, using Lemma A.5, we obtain that

$$\begin{aligned} &||\hat{v}||^{(\lfloor \Gamma \rfloor)} \left(1 - (\Gamma - \lfloor \Gamma \rfloor)^2\right) + ||\hat{v}||^{(\lfloor \Gamma \rfloor + 1)}(\Gamma - \lfloor \Gamma \rfloor)^2 \le \\ &||v||^{(\lfloor \Gamma \rfloor)} \left(1 - (\Gamma - \lfloor \Gamma \rfloor)^2\right) + ||v||^{(\lfloor \Gamma \rfloor + 1)}(\Gamma - \lfloor \Gamma \rfloor)^2. \end{aligned}$$

This shows the claim that $(\tilde{r}, \tilde{\Gamma}, \hat{v})$ is feasible for the problem. Hence, adding the constraint that $\gamma = v_i$ for all $i \in [n]$ does not change the optimal value of this problem. Adding this constraint simplifies the problem to

$$\begin{aligned} &\text{maximize } r \\ &\quad \text{s.t. } r^2 \le \gamma \\ &\qquad r^2 \le \frac{\Gamma^2}{n}\gamma \\ &\qquad \gamma(\lfloor \Gamma \rfloor) \left(1 - (\Gamma - \lfloor \Gamma \rfloor)^2\right) + \gamma(\lfloor \Gamma \rfloor + 1)(\Gamma - \lfloor \Gamma \rfloor)^2 \le 1 \end{aligned}$$

Rearranging the last inequality yields

$$\begin{aligned} &\gamma(\lfloor\Gamma\rfloor)\left(1-(\Gamma-\lfloor\Gamma\rfloor)^2\right)+\gamma(\lfloor\Gamma\rfloor+1)(\Gamma-\lfloor\Gamma\rfloor)^2\leq 1 \\ \Leftrightarrow\ &\gamma\leq\frac{1}{\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2}. \end{aligned}$$

The optimal value of γ is readily given by $\frac{1}{\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2}$. Hence, the optimization problem can be further simplified to

$$\max\ \min\left(\frac{1}{\sqrt{\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2}},\frac{\Gamma}{\sqrt{n}\sqrt{\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2}}\right).$$

Since $\frac{1}{\sqrt{\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2}}$ is strictly decreasing in Γ and $\frac{\Gamma}{\sqrt{n}\sqrt{\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2}}$ strictly increasing, the minimum of both functions is maximized if they are equal. This yields to the final equation

$$\frac{1}{\sqrt{\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2}}=\frac{\Gamma}{\sqrt{n}\sqrt{\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2}}\Leftrightarrow\Gamma=\sqrt{n}$$

Therefore, the optimal value of r^* is given by $r^*=\frac{1}{\sqrt{\lfloor\Gamma^*\rfloor+(\Gamma^*-\lfloor\Gamma^*\rfloor)^2}}$ and, since $a_i=\sqrt{v_i}d_i$, we obtain that $a_i^*=r^*d_i$. This concludes the proof. □

Note from the proof of Theorem 3.1 can be observed that if Γ is fixed, it is still optimal to set a_i to $\frac{1}{\sqrt{\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2}}$. We will show in the next section that the same holds true in the case of volume approximation.

4 Volume Approximation

In this section, we discuss problem (AP-V). Before we consider the optimization problem, we derive a formula for the volume of a bounded uncertainty set.

Lemma 4.1. *The volume of a bounded uncertainty set $\mathcal{U}(a,\Gamma)$ is given by*

$$\operatorname{vol}(\mathcal{U}(a,\Gamma))=\left(\prod_{i=1}^{n}a_i\right)\frac{2^n}{n!}\sum_{k=0}^{\lfloor\Gamma\rfloor}(-1)^k\binom{n}{k}(\Gamma-k)^n.$$

Proof. We restrict again our analysis to the positive orthant, since $\mathcal{U}(a,\Gamma)$ is symmetric with respect to the origin. We denote by $\mathcal{U}^+(a,\Gamma)$ the non-negative part of $\mathcal{U}(a,\Gamma)$, i.e., $\mathcal{U}^+(a,\Gamma):=\mathcal{U}(a,\Gamma)\cap\mathbb{R}^n_+$. Note that $\operatorname{vol}(\mathcal{U}(a,\Gamma))=2^n\operatorname{vol}(\mathcal{U}^+(a,\Gamma))$. Clearly, $\mathcal{U}^+(a,\Gamma)\subset\times_{i=1}^n[0,a_i]$. We compute $\operatorname{vol}(\mathcal{U}^+(a,\Gamma))$ using the idea of Monte Carlo sampling. Sample n random variables $X_i\sim U[0,a_i]$. The experiment is successful if $X=(X_1,\ldots,X_n)^T\in\mathcal{U}^+(a,\Gamma)$. Denote by p the success probability of this experiment. We use the theory of Monte Carlo sampling [6] to conclude that $\operatorname{vol}(\mathcal{U}^+(a,\Gamma))=p\cdot\operatorname{vol}(\times_{i=1}^n[0,a_i])$.

Let $Y_i\sim U[0,1]$ for $i\in[n]$, and define $Y^n:=\sum_{i=1}^n Y_i$. The distribution of random variable Y^n is known as the Irwin-Hall distribution [7]. The distribution

function F_{Y^n} is given by the following formula

$$F_{Y^n}(\Gamma) = \frac{1}{n!}\sum_{k=0}^{\lfloor\Gamma\rfloor}(-1)^k \binom{n}{k} (\Gamma - k)^n.$$

The success probability p is equal to $F_{Y^n}(\Gamma)$, since

$$P(\text{“success”}) = P\left(\sum_{i=1}^{n} \frac{X_i}{a_i} \leq \Gamma\right) = P\left(\sum_{i=1}^{n} Y_i \leq \Gamma\right) = P(Y^n \leq \Gamma) = F_{Y^n}(\Gamma).$$

Combining these formulas concludes the theorem:

$$\begin{aligned}\operatorname{vol}(\mathcal{U}(a,\Gamma)) &= 2^n \operatorname{vol}(\mathcal{U}^+(a,\Gamma)) = 2^n \cdot p \cdot \operatorname{vol}(\bigtimes_{i=1}^{n}[0,a_i]) \\ &= \left(\prod_{i=1}^{n} a_i\right) \frac{2^n}{n!}\sum_{k=0}^{\lfloor\Gamma\rfloor}(-1)^k \binom{n}{k} (\Gamma - k)^n\end{aligned}$$

□

Next, we consider the optimization problem for a fixed value of Γ.

$$\begin{aligned}&\max_{a} \ \operatorname{vol}(\mathcal{U}(a,\Gamma)) \qquad \text{(AP-V}(\Gamma)\text{)} \\ &\text{s.t. } \mathcal{U}(a,\Gamma) \subset \mathcal{E}(D)\end{aligned}$$

Lemma 4.2. *Let the ellipsoid $\mathcal{E}(D) \subset \mathbb{R}^n$ be defined by the diagonal matrix $D = diag\left(\frac{1}{d_1^2}, \ldots, \frac{1}{d_n^2}\right)$. The optimal solution a^* of problem* (AP-V(Γ)) *is given by*

$$a_i^* = \frac{d_i}{\sqrt{\lfloor\Gamma\rfloor + (\Gamma - \lfloor\Gamma\rfloor)^2}}, i \in [n]$$

Proof. Since the logarithm is a strictly increasing function, we can maximize $\log(\operatorname{vol}(\mathcal{U}(a,\Gamma)))$ instead of $\operatorname{vol}(\mathcal{U}(a,\Gamma))$. Like in the proof of Theorem 3.1, we define $v_i = \frac{a_i^2}{d_i^2}$.

$$\begin{aligned}\log(\operatorname{vol}(\mathcal{U}(a,\Gamma))) &= \log\left(\prod_{i=1}^{n} a_i \frac{2^n}{n!}\sum_{k=0}^{\lfloor\Gamma\rfloor}(-1)^k \binom{n}{k} (\Gamma - k)^n\right) \\ &= \log\left(\prod_{i=1}^{n} a_i\right) + \log\left(\frac{2^n}{n!}\sum_{k=0}^{\lfloor\Gamma\rfloor}(-1)^k \binom{n}{k} (\Gamma - k)^n\right) \\ &= \sum_{i=1}^{n}\frac{1}{2}\log(v_i d_i^2) + \log\left(\frac{2^n}{n!}\sum_{k=0}^{\lfloor\Gamma\rfloor}(-1)^k \binom{n}{k} (\Gamma - k)^n\right) \\ &= \frac{1}{2}\sum_{i=1}^{n}\log(v_i) + \sum_{i=1}^{n}\log(d_i) + \log\left(\frac{2^n}{n!}\sum_{k=0}^{\lfloor\Gamma\rfloor}(-1)^k \binom{n}{k} (\Gamma - k)^n\right)\end{aligned}$$

Since Γ is fixed in (AP-V(Γ)), it is sufficient to maximize $\sum_{i=1}^n \log(v_i)$. Using Lemma A.3, the constraint $\mathcal{U}(a,\Gamma) \subset \mathcal{E}(D)$ can be represented by $||v||^{(\lfloor\Gamma\rfloor)}\left(1-(\Gamma-\lfloor\Gamma\rfloor)^2\right)+||v||^{(\lfloor\Gamma\rfloor+1)}(\Gamma-\lfloor\Gamma\rfloor)^2 \leq 1$. Hence, the optimization problem can be written as

$$\begin{aligned}\text{maximize } & \sum_{i=1}^n \log(v_i) \\ \text{s.t. } & ||v||^{(\lfloor\Gamma\rfloor)}\left(1-(\Gamma-\lfloor\Gamma\rfloor)^2\right)+||v||^{(\lfloor\Gamma\rfloor+1)}(\Gamma-\lfloor\Gamma\rfloor)^2 \leq 1\end{aligned}$$

Let $\tilde{v}$ be an optimal solution of the optimization problem. We denote by $\sigma(n)$ all permutations of $[n]$ and by $\tilde{v}^\pi$ the vector $\tilde{v}$ whose entries are permuted by $\pi \in \sigma(n)$. Similar to the proof of Lemma A.5, we consider $\hat{v} = \frac{1}{n!}\sum_{\pi\in\sigma(n)} \tilde{v}^\pi$ the average over all permutations of $\tilde{v}$. Note that $\hat{v}$ is feasible for the problem. Since the objective function $f(v) = \sum_{i=1}^n \log(v_i)$ is concave, we obtain that

$$f(\hat{v}) = f\left(\frac{1}{n!}\sum_{\pi\in\sigma(n)} \tilde{v}^\pi\right) \geq \frac{1}{n!}\sum_{\pi\in\sigma(n)} f(\tilde{v}^\pi) = f(\tilde{v})$$

Hence, $\hat{v}$ is an optimal solution. Therefore, we may add the constraint $\gamma = v_i \ \forall i \in [n]$ without changing the optimal value of the problem. This simplifies the problem to

$$\begin{aligned}\text{maximize } & n\cdot\log(\gamma) \\ \text{s.t. } & \gamma(\lfloor\Gamma\rfloor)\left(1-(\Gamma-\lfloor\Gamma\rfloor)^2\right)+\gamma(\lfloor\Gamma\rfloor+1)(\Gamma-\lfloor\Gamma\rfloor)^2 \leq 1\end{aligned}$$

Since the objective function is strictly increasing, the optimal value of the problem is given by $\gamma^* = \frac{1}{\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2}$. Using that $a_i = \sqrt{v_i}d_i$, we obtain the claimed value of $a_i^* = \frac{d_i}{\sqrt{\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2}}$. □

Lemma 4.3. *The optimal objective value of problem* (AP-V(Γ)) *is given by*

$$\left(\prod_{i=1}^n d_i\right)(\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2)^{-\frac{n}{2}}\frac{2^n}{n!}\sum_{k=0}^{\lfloor\Gamma\rfloor}(-1)^k\binom{n}{k}(\Gamma-k)^n.$$

Proof. Immediate consequence of Lemma 4.1 and 4.2. □

Lemma 4.3 shows that the optimal Γ^* which is a minimizer of (AP-V) must also be a minimizer of

$$f_n(\Gamma) = (\lfloor\Gamma\rfloor+(\Gamma-\lfloor\Gamma\rfloor)^2)^{-\frac{n}{2}}\frac{1}{n!}\sum_{k=0}^{\lfloor\Gamma\rfloor}(-1)^k\binom{n}{k}(\Gamma-k)^n$$

Note that this function is independent of the ellipsoid $\mathcal{E}(D)$ (and only depends on n). Denote by $\Gamma^*(n)$ the minimizer of $f_n(\Gamma)$.

Theorem 4.4. *Let the ellipsoid $\mathcal{E}(D) \subset \mathbb{R}^n$ be defined by the diagonal matrix $D = diag\left(\frac{1}{d_1^2}, \ldots, \frac{1}{d_n^2}\right)$. The optimal solution (Γ^*, a^*) of problem* (AP-V) *is given by*

$$\Gamma^* = \Gamma^*(n)$$
$$a_i^* = \frac{d_i}{\sqrt{\lfloor \Gamma^*(n) \rfloor + (\Gamma^*(n) - \lfloor \Gamma^*(n) \rfloor)^2}}, i \in [n]$$

4.1 Approximation of $\Gamma^*(n)$

Since the exact computation of the minimum of $f_n(\Gamma)$ is computationally very expensive, we use an approximate function $\tilde{f}_n$. First, we replace $\lfloor \Gamma \rfloor + (\Gamma - \lfloor \Gamma \rfloor)^2$ with Γ. Second, we use the central limit theorem to replace the sum resulting from the Irwan-Hall distribution. Denote by $Y_i \sim U[0,1]$ random variables which are independent uniform distributed in $[0,1]$ and by $Z \sim \mathcal{N}(0,1)$ a standard normal distributed random variable. Then, we have

$$\begin{aligned}
\frac{1}{n!} \sum_{k=0}^{\lfloor \Gamma \rfloor} (-1)^k \binom{n}{k} (\Gamma - k)^n &= P\left(\sum_{i=1}^{n} Y_i \leq \Gamma \right) \\
&= P\left(\sum_{i=1}^{n} Y_i - \frac{n}{2} \leq \Gamma - \frac{n}{2} \right) \\
&= P\left(\frac{\sum_{i=1}^{n} Y_i - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \leq \frac{\Gamma - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \right) \\
&\approx P\left(Z \leq \frac{\Gamma - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \right) \\
&= \Phi\left(\frac{\Gamma - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \right)
\end{aligned}$$

where we use that $E\left(\sum_{i=1}^{n} Y_i\right) = \frac{n}{2}$ and $\sigma\left(\sum_{i=1}^{n} Y_i\right) = \sqrt{\frac{n}{12}}$ [7]. We can replace the standardized sum with a standard normal distributed random variable due to the central limit theorem. Using both replacements, we obtain

$$\tilde{f}_n(\Gamma) = \Gamma^{-\frac{n}{2}} \Phi\left(\frac{\Gamma - \frac{n}{2}}{\sqrt{\frac{n}{12}}} \right)$$

Denote by $\Gamma^*(n)$ ($\tilde{\Gamma}^*(n)$) the minimizer of $f_n(\Gamma)$ ($\tilde{f}_n(\Gamma)$). In Figure 1, we compare the different values. Interestingly, the linear function $h(n) = 0.3947n + 0.98$ seems to yield a good approximation of $\tilde{\Gamma}^*(n)$ at least for $n \leq 200$. Note that $\tilde{\Gamma}^*(n)$ cannot be given exactly by some linear function since this linear function could be used to derive an analytical expression of Φ, which is impossible. The experiments indicate that the true function is sublinear. Note that the central limit theorem is only applicable if n is large enough. We start using the approximate function $\tilde{f}_n$ for $n = 30$. We ran into numerical problems computing $\Gamma^*(n)$ for $n > 43$.

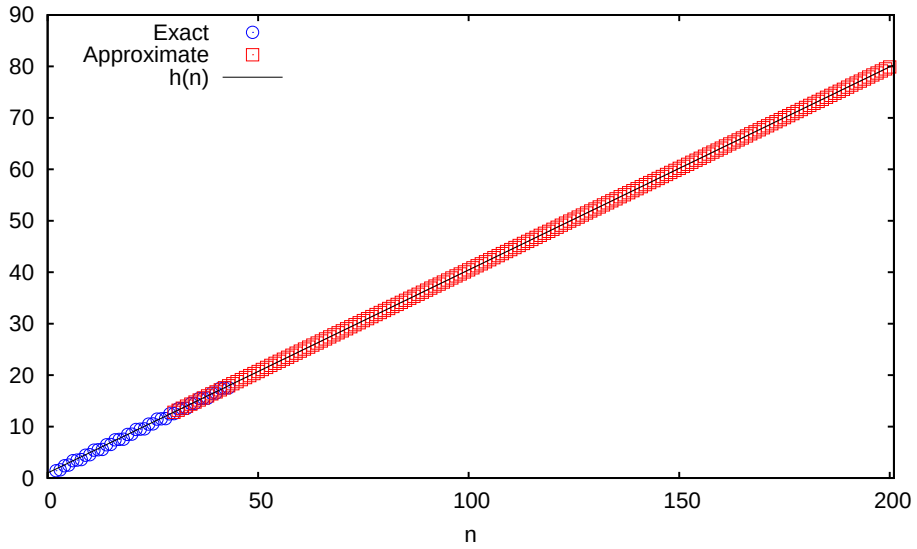

Figure 1: The values $\Gamma^*(n)$ $[\tilde{\Gamma}^*(n)]$ are shown as blue cycles [red boxes] on the y-axis. The function $\tilde{\Gamma}^*(n)$ can be approximated by the linear function $h(n) = 0.3947n + 0.98$.

5 General Ellipsoids

In this section, we shortly discuss how arbitrary ellipsoids $\mathcal{E}(a^0, M)$ which are not axis-parallel and not centered at the origin, i.e., M is not a diagonal matrix and $a^0 \neq 0$, may be approximated by a polyhedron. Shifting the center is straightforward, just use a^0 as center for the bounded uncertainty set $\mathcal{U}(a^0, a, \Gamma)$. However, in the case of ellipsoids which are not axis-parallel the standard bounded uncertainty set is not able to cover the orientation of the ellipsoid correctly. The idea is to first shift the center of the ellipsoid to the origin, then to rotate the ellipsoid such that it becomes axis-parallel. The next step is to fit a bounded uncertainty set to the rotated ellipsoid and to rotate the fitted uncertainty set back. Lastly, the polyhedral set is shifted by a^0 to approximate the original ellipsoid. The so obtained polyhedron is defined as *rotated bounded uncertainty set*.
The rotation which converts a general ellipsoid to one that is axis parallel is obtained by the eigendecomposition of the matrix M:

$$M = RDR^T$$

where D is a diagonal and R an orthogonal matrix. Rotating $\mathcal{E}(M)$ with matrix R^T yields the axis parallel ellipsoid $\mathcal{E}(D)$ [9], i.e.,

$$R^T(\mathcal{E}(M)) = \{R^T x : x \in \mathcal{E}(M)\} = \mathcal{E}(D)$$

The procedure of shifting, rotating, fitting, re-rotating, and re-shifting combined with Theorem 3.1 and 4.4 results in the following theorem.

Theorem 5.1. *Given a general ellipsoid* $\mathcal{E}(a^0, M) \subset \mathbb{R}^n$ *with* $M = RDR^T$ *and* $D = diag\left(\frac{1}{d_1^2}, \ldots, \frac{1}{d_n^2}\right)$. *The rotated bounded uncertainty set which gives the* best *approximation of this ellipsoid is described by*

$$\begin{aligned} R_i^T(x - a^0) &\leq z_i \ \forall i \in [n] \\ -R_i^T(x - a^0) &\leq z_i \ \forall i \in [n] \\ z_i\sqrt{\lfloor\Gamma^*\rfloor + (\Gamma^* - \lfloor\Gamma^*\rfloor)^2} &\leq d_i \ \forall i \in [n] \\ \sum_{i=1}^{n} z_i \frac{\sqrt{\lfloor\Gamma^*\rfloor + (\Gamma^* - \lfloor\Gamma^*\rfloor)^2}}{d_i} &\leq \Gamma^*. \end{aligned}$$

For ratio approximation Γ^* *is set to* $\sqrt{n}$.
For volume approximation Γ^* *is set to* $\Gamma^*(n)$.

Proof. Let $\mathcal{U}(a, \Gamma)$ be the optimal approximation of the rotated ellipsoid $R^T(\mathcal{E}(M))$. Then, $R(\mathcal{U}(a, \Gamma))$ is the optimal approximation of $\mathcal{E}(M)$. Further, $\{a^0\} + R(\mathcal{U}(a, \Gamma))$ approximates the original ellipsoid $\mathcal{E}(a^0, M)$. The linear description of the rotated bounded uncertainty set is an immediate consequence of the simple observation that

$$x \in \{a^0\} + R(\mathcal{U}(a, \Gamma)) \Leftrightarrow x - a^0 \in R(\mathcal{U}(a, \Gamma)) \Leftrightarrow R^T(x - a^0) \in \mathcal{U}(a, \Gamma).$$

The choices of Γ^* are given by Theorem 3.1 and 4.4. □

Note that the robust counterpart of a linear program with a rotated bounded uncertainty set is a linear program again. We present some ellipsoids in $\mathbb{R}^2$ and their best possible inner approximations by rotated bounded uncertainty sets in Figure 2. Note that for $n = 2$, we have that $\Gamma^* = \sqrt{2}$ is optimal for ratio and volume approximation.

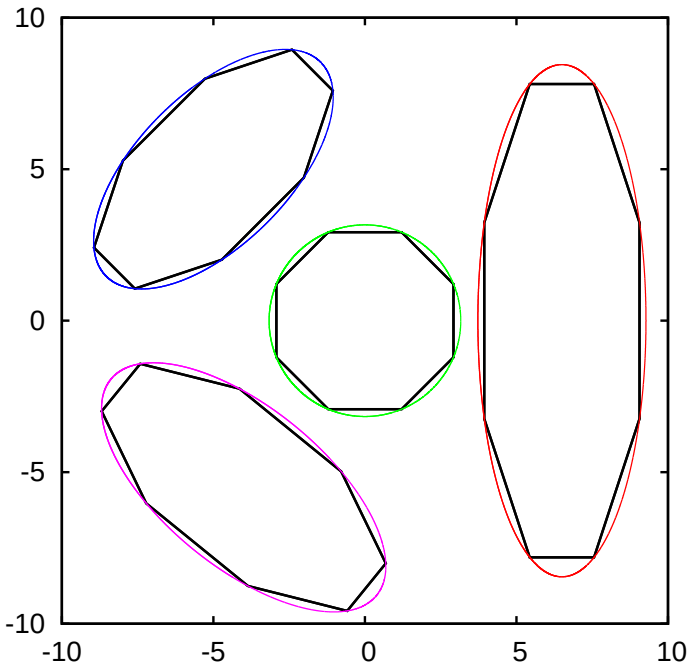

Figure 2: Some ellipsoids and the corresponding approximations by rotated bounded uncertainty sets. We choose $\Gamma^* = \sqrt{2}$ which is optimal for ratio and volume approximation for $n = 2$.

6 Computational Experiment

In this section, we use a shortest path and network flow problem with ellipsoidal uncertainty to evaluate the robust counterpart resulting from ellipsoid and bounded uncertainty sets. For the shortest path problem, we compare the solution times of the different robust counterparts. The network flow problem is used to evaluate the quality of the approximation for different values of Γ. The experiments are solved on a Intel Core i5-3470 processor, running at 3.20 GHz, 8 GB RAM under Windows 7. We used CPLEX version 12.6 to solve the optimization problems.

For both problems, the underlying networks are layered graphs. A layered graph $G = (V, A)$ is described by two parameters, the width $w \in \mathbb{N}$ and the number of layers $l \in \mathbb{N}$. The node set $V = \{s\} \cup V_1 \cup V_2 \cup \cdots \cup V_l \cup \{t\}$, where $V_i = \{v_{i1}, \ldots, v_{iw}\}$ is the i^{th} layer consisting of w nodes. The arcs go from s to V_1, from each node of V_i to each node of V_{i+1} for $i \in [l-1]$, and from each node of V_l to t (see Figure 3). We denote by n the number of arcs in G.

The costs of each arc a are given by c_a. We assume that the cost vector c is affected by ellipsoidal uncertainty, i.e., $c \in \mathcal{E}(\bar{c}, M)$ where $\bar{c}$ is the average cost vector and M a symmetric and positive definite matrix. The eigendecomposition of M is given by $M = RDR^T$, where R is a orthogonal matrix and D a diagonal matrix. We use the following procedure to sample the cost vector c and the matrix M. Each entry of c and each diagonal entry of D is chosen uniform at

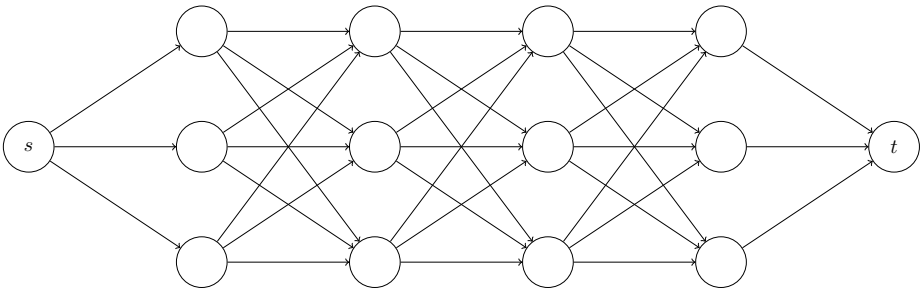

Figure 3: A layered graph G with $l = 4$ and $w = 3$.

random from the interval $[0, 10]$. The rotation matrix R is generated iteratively. At the beginning, we set $R = I$, where $I \in \mathbb{R}^{n\times n}$ is the unit matrix. We repeat the following procedure r times: We choose uniform at random two different indices from $[n]$ and an angle ϕ uniform from $[0, 2\pi]$. Denote by $H(i, j, \phi) \in \mathbb{R}^{n\times n}$ the two dimensional rotation matrix that is lifted to the $\mathbb{R}^{n\times n}$:

$$(H(i,j,\phi))_{y,z} = \begin{cases} 1 & \text{, if } y = z, y \neq i, y \neq j \\ cos(\phi) & \text{, if } y = z, y \in i, j \\ -sin(\phi) & \text{, if } y = i, z = j \\ sin(\phi) & \text{, if } y = j, z = i \\ 0 & \text{, else} \end{cases}$$

We update R, by $R := H(i, j, \phi)R$. Finally, after r rotations, we compute $M = RDR^T$.

6.1 Runtime

First, we consider the problem of finding the cheapest path from s to t under the worst scenario of $\mathcal{E}(\bar{c}, M)$. Note that this problem is equivalent to the reliable shortest path problem if the arc costs $\mathbf{c}$ are assumed to be normally distributed [5]. The goal of the reliable shortest path problem is to find a path P that minimizes t with the constraint that $\mathbb{P}(\mathbf{c}(P) \leq t) \geq \alpha$ for some $\alpha \in [0.5, 1)$.

Denote by $\mathcal{X}$ the set of all binary vectors that represent an s-t path, i.e., $x \in \mathcal{X}$ if and only if $\{a \in A : x_a = 1\}$ is the arc set of a path from s to t. The resulting optimization problem is given by

$$\min_{x\in\mathcal{X}} \max_{c\in\mathcal{E}(\bar{c},M)} c^T x.$$

This is equivalent to the following quadratically constraint integer program since $\max_{\{c:c^T Mc\leq 1\}} c^T x = \sqrt{x^T M^{-1} x}$

$$\begin{aligned} &\min\ \bar{c}^T x + z && (\mathcal{P}) \\ &\text{s.t. } x^T M^{-1} x \leq z^2 \\ &\quad x \in \mathcal{X}. \end{aligned}$$

We use the presented techniques to approximate $\mathcal{E}(M)$ with a rotated bounded uncertainty set $R(\mathcal{U}(a,\Gamma))$. The resulting problem is given by

$$\min_{x\in\mathcal{X}}\left(\overline{c}^T x + \max_{c\in R(\mathcal{U}(a,\Gamma))} c^T x\right).$$

This is equivalent to

$$\min_{x\in\mathcal{X}}\left(\overline{c}^T x + \max_{c\in \mathcal{U}(a,\Gamma)} c^T R^T x\right).$$

Dualizing the inner maximization problem, we obtain a compact mixed integer programming formulation

$$\begin{aligned}
&\min\ \overline{c}^T x + a^T p + \Gamma\eta && (\mathcal{P}')\\
&\text{s.t. } \pi_i - p_i - \frac{\eta}{a_i} \le 0 \quad \forall i\in[n]\\
&\qquad -\pi_i \le R_i^T x \le \pi_i \quad \forall i\in[n]\\
&\qquad p,\eta \ge 0\\
&\qquad x\in\mathcal{X}
\end{aligned}$$

The focus of the first experiment is to compare the time that is necessary to solve $\mathcal{P}$ and $\mathcal{P}'$. We did two different runs of the experiment. In the first run, we fix the number of rotations r to 1000 and vary the number of layers l from 2 to 101. In the second run, we fix the number of layers l to 50 and vary the number of rotations from 0 to 3000. In both runs, we set the width w of the underlying graph to 4. For each run, we solve 1000 different shortest path problems. The time limit for a single shortest path problem is set to 30 seconds. We choose a and Γ as suggested by Theorem 3.1, i.e., $\Gamma = \sqrt{n}$. The results of both runs are shown in in Figure 4 and 5.

For a small number of layers, less than 20, problems $\mathcal{P}$ and $\mathcal{P}'$ are basically solved at the same time. Further increasing the number of layers increases not only the average but also the variance of the solution time of $\mathcal{P}$. Note that 100 layers are not enough to impact the computation time of $\mathcal{P}'$.

The solution time of $\mathcal{P}$ scales similarly with the number of rotations: An increasing number of rotations leads to an increase of the average and of the variance of the computation time. But, increasing the number of rotations shows also an effect on the solution time of $\mathcal{P}'$. Nevertheless, $\mathcal{P}'$ is still solved considerably faster than $\mathcal{P}$.

For both runs, the solutions found by $\mathcal{P}'$ are almost optimal. In particular, the solutions found by $\mathcal{P}'$ are on average 0.05% worse than the optimal solution. Indeed in 1160 of 2000 instances the optimal solutions of both problems are equal. Note that 123 instances could not be solved to optimality within the time limit. For these instances, we assume that the optimal solution is equal to the best solution found within the time limit or to the solution of $\mathcal{P}'$ if this solution is better (this happened 68 times).

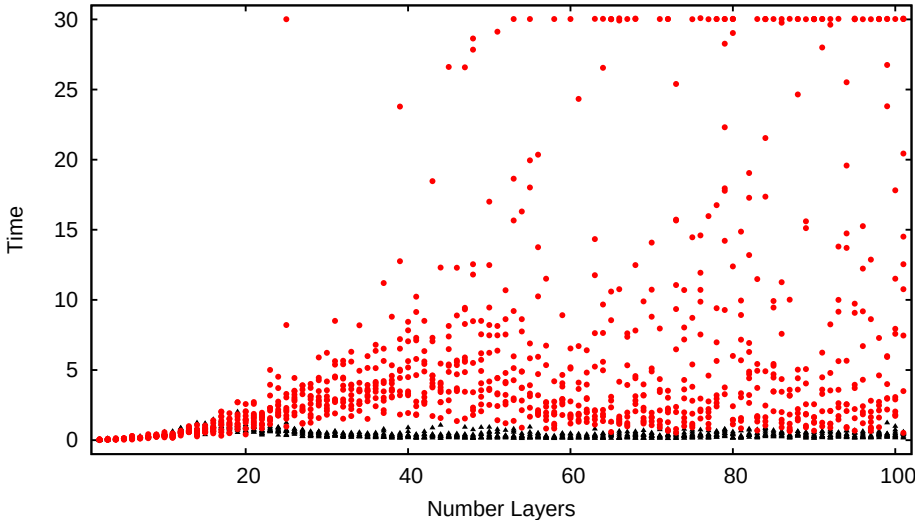

Figure 4: The red dots represent the time to solve $\mathcal{P}$ and the black triangles the time to solve $\mathcal{P}'$. The computation time is given in seconds. The time limit is set to 30 seconds.

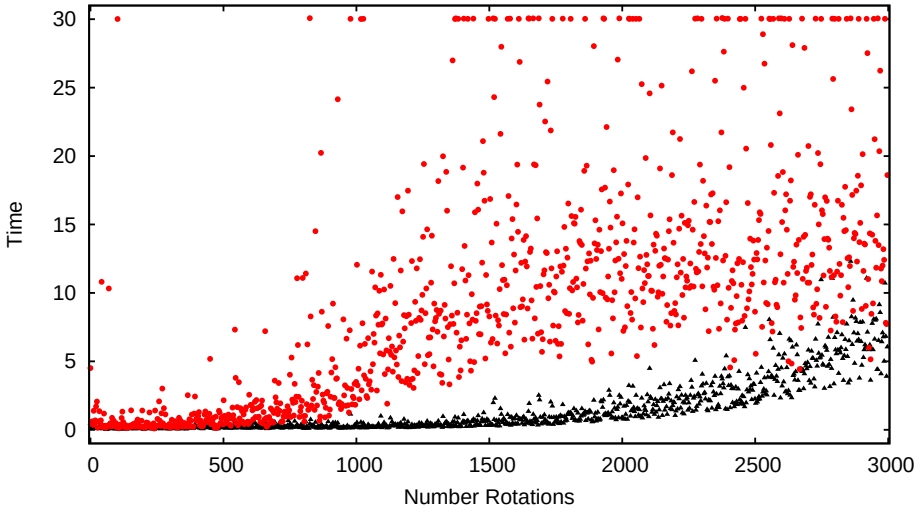

Figure 5: The red dots represent the time to solve $\mathcal{P}$ and the black triangles the time to solve $\mathcal{P}'$. The computation time is given in seconds. The time limit is set to 30 seconds.

6.2 Quality of Approximation

In the second experiment, we do some modifications of the experiment setup to generate problems for which the solutions of $\mathcal{P}'$ are not almost optimal.

First, we set the average cost vector $\bar{c} = 0$. Note that the resulting problem is equivalent to the quadratic shortest path problem (if the objective function is squared), which is a very challenging problem as presented in [8]. Second, we replace $\mathcal{X}$ with $conv(\mathcal{X})$, i.e., we shift the feasible space from s-t paths to s-t flows with flow value 1. This increases the solutions space and allows for a bigger gap between the exact and the approximate problem. As a result, the uncertainty set and its shape become more important for the value of the objective function. Note that under this setup scaling of $\mathcal{E}(M)$ and $\mathcal{U}(a, \Gamma)$ by an arbitrary factor $\lambda > 0$ has no impact on the optimal solution, since

$$\max_{c \in \lambda\mathcal{E}(M)} c^T x = \max_{\lambda\tilde{c} \in \lambda\mathcal{E}(M)} \lambda\tilde{c}^T x = \lambda \max_{\tilde{c} \in \mathcal{E}(M)} \tilde{c}^T x \qquad \text{and}$$
$$\max_{c \in \lambda\mathcal{U}(a,\Gamma)} c^T x = \max_{\lambda\tilde{c} \in \lambda\mathcal{U}(a,\Gamma)} \lambda\tilde{c}^T x = \lambda \max_{\tilde{c} \in \mathcal{U}(a,\Gamma)} \tilde{c}^T x.$$

In one run of the experiment, we first generate a random matrix M, as described above, with $r = 1000$ for the layered graph with $w = l = 5$. Note that the number of arcs is equal to $n = 110$. Next, we compute the optimal solution of the problem

$$\begin{aligned} &\min \sqrt{x^T M^{-1} x} \\ &\text{s.t. } x \in conv(\mathcal{X}) \end{aligned} \tag{$\mathcal{F}$}$$

For each value of Γ from 1 to 110, we solve the approximate problem

$$\min_{x \in conv(\mathcal{X})} \max_{c \in \mathcal{U}(a,\Gamma)} c^T R^T x \tag{$\mathcal{F}(\Gamma)$}$$

where the bounded uncertainty set $\mathcal{U}(a, \Gamma)$ is computed according to Lemma 4.2. We denote by x^Γ the optimal solutions of $\mathcal{F}(\Gamma)$ and by x^* the optimal solution of $\mathcal{F}$. Finally, we compute $\sqrt{\frac{x^\Gamma M^{-1} x^\Gamma}{x^* M^{-1} x^*}}$ to compare the performance of x^Γ and x^*. We execute 1000 runs of this experiment. The averaged results are shown in Figure 6.

The best approximation is achieved for $\Gamma = 41$. Interestingly, this is really close to $\Gamma^*(110) \approx 44$, the optimal Γ with respect to (AP-V). Note that we have tried to make the gap between $\mathcal{F}$ and $\mathcal{F}'$ as large as possible by setting $\bar{c} = 0$ and relaxing the integrality constraints. However, the gap for the optimal Γ is still below 10%.

We summarize the findings of both experiments. The first experiment underlines the expectation that $\mathcal{P}'$ is better suited for large scale problems than $\mathcal{P}$. The second experiment shows that the right choice of Γ is important to get a good approximation of the original problem.

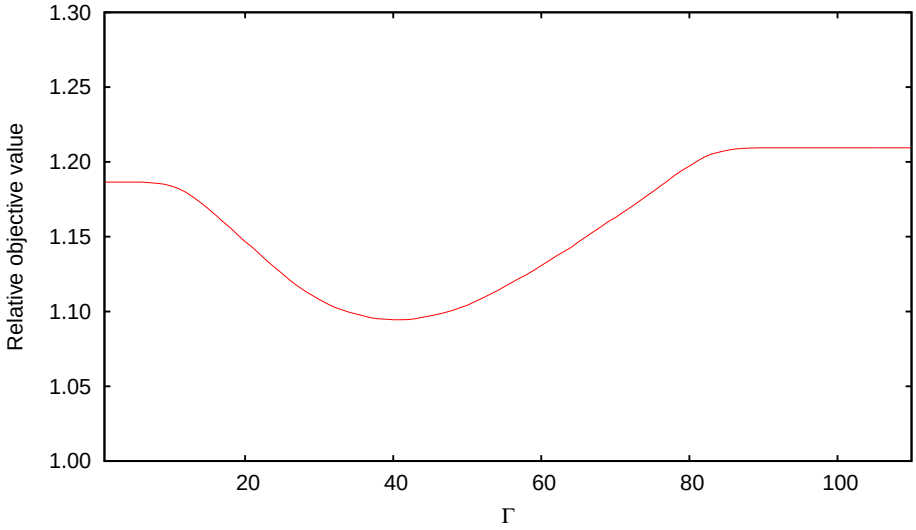

Figure 6: We evaluate x^Γ using the objective function of problem $\mathcal{F}$ and compare it to the objective value of x^* by computing $\sqrt{\frac{x^\Gamma M^{-1} x^\Gamma}{x^* M^{-1} x^*}}$ (shown on the y-axis). The values are averaged over 1000 runs.

7 Conclusions

We derived explicit formulas for the approximation of ellipsoid uncertainty sets by bounded uncertainty sets. To approximate ellipsoid uncertainty sets that are not axis-parallel, we introduced the rotated bounded uncertainty set as a generalization of the classic bounded uncertainty set. We proved that for a fixed budget parameter Γ, the optimal bounded uncertainty sets for the ratio and volume approximation problem are identical. The optimal solutions differ only in the choice of Γ. For ratio approximation, the optimal choice of Γ is equal to $\sqrt{n}$, whereas for volume approximation, we have $\Gamma \approx 0.4n$ (for $n \leq 200$), where n is the dimension of the underlying space.

In the experimental result section, we test the derived approximation techniques. In the first experiment, we consider a shortest path problem with an ellipsoid uncertainty set. For large scale problems, the approximate problem can be solved considerably faster than the original one and yields solutions which are almost optimal. In the second experiment, we use a relaxed version of the former problem to show the influence of the choice of Γ on the quality of the approximation. For the presented problem, the Γ maximizing the volume of the bounded uncertainty set proves to be a good choice.

The derived approximations give rise to interesting research questions: Are there other tractable polytope families beside the bounded uncertainty sets which can be used to approximate ellipsoids with the same or even better approximation properties? Which problem classes are well suited for ratio or volume approximation? How does scaling of the bounded uncertainty set influences the quality of the approximation?

References

[1] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust optimization*. Princeton University Press, 2009.

[2] Aharon Ben-Tal and Arkadi Nemirovski. Robust solutions of uncertain linear programs. *Operations research letters*, 25(1):1–13, 1999.

[3] Dimitris Bertsimas, Dessislava Pachamanova, and Melvyn Sim. Robust linear optimization under general norms. *Operations Research Letters*, 32(6):510–516, 2004.

[4] Dimitris Bertsimas and Melvyn Sim. The price of robustness. *Operations research*, 52(1):35–53, 2004.

[5] Bi Yu Chen, William HK Lam, Agachai Sumalee, Qingquan Li, Hu Shao, and Zhixiang Fang. Finding reliable shortest paths in road networks under uncertainty. *Networks and spatial economics*, 13(2):123–148, 2013.

[6] John Hammersley. *Monte carlo methods*. Springer Science & Business Media, 2013.

[7] Norman L Johnson, Samuel Kotz, and N Balakrishnan. *Continuous Multivariate Distributions, Volume 1, Models and Applications*, volume 59. New York: John Wiley & Sons, 2002.

[8] Borzou Rostami, André Chassein, Michael Hopf, Davide Frey, Christoph Buchheim, Federico Malucelli, and Marc Goerigk. The quadratic shortest path problem: Complexity, approximability, and solution methods. *Technical Report available at www.optimization-online.org*, 2016.

[9] Gilbert Strang and Wellesley-Cambridge Press. *Introduction to linear algebra*, volume 3. Wellesley-Cambridge Press Wellesley, MA, 1993.

A Appendix

Lemma A.1. $\max_{x \in r\mathcal{E}(D)} x_i = rd_i$

Proof. The constraint defining the scaled ellipsoid $r\mathcal{E}(D)$ is $\sum_{i=1}^n \frac{x_i^2}{d_i^2} \leq r^2$. The i^{th} component is maximized by setting $x_j = 0$ for $j \neq i$ and $x_i = rd_i$. □

Lemma A.2. $\max_{x \in r\mathcal{E}(D)} \sum_{i=1}^n \frac{x_i}{a_i} = r\sqrt{\sum_{i=1}^n \frac{d_i^2}{a_i^2}}$

Proof. The maximum of a linear function over an ellipsoid is attained at the boundary of the ellipsoid. Hence, the optimization problem is equivalent to

$$\begin{aligned} \text{maximize} \quad & \sum_{i=1}^n \frac{x_i}{a_i} \\ \text{s.t.} \quad & \sum_{i=1}^n \frac{x_i^2}{d_i^2} = r^2. \end{aligned}$$

Using the Karush-Kuhn-Tucker theorem leads to the following equations

$$\frac{1}{a_i} + 2\lambda \frac{x_i^*}{d_i^2} = 0 \ \forall i \in [n]$$
$$\sum_{i=1}^{n} \frac{x_i^{*2}}{d_i^2} = r^2.$$

Under the assumption that $x^* \geq 0$, the system is uniquely solved by

$$x_i^* = r \frac{d_i^2}{a_i \sqrt{\sum_{i=1}^{n} \frac{d_i^2}{a_i^2}}}.$$

Computing the objective function at the point x^* completes the proof

$$\sum_{i=1}^{n} \frac{x_i^*}{a_i} = r \sqrt{\sum_{i=1}^{n} \frac{d_i^2}{a_i^2}}.$$

□

Lemma A.3. *The optimal value of the optimization problem*

$$\max \sum_{i=1}^{n} \frac{a_i^2}{d_i^2} y_i^2$$
$$s.t.\ 0 \leq y_i \leq 1 \ \forall i \in [n]$$
$$\sum_{i=1}^{n} y_i \leq \Gamma$$

is given by $||v||^{(\lfloor \Gamma \rfloor)} \left(1 - (\Gamma - \lfloor \Gamma \rfloor)^2\right) + ||v||^{(\lfloor \Gamma \rfloor + 1)} (\Gamma - \lfloor \Gamma \rfloor)^2$ *where* $v = (v_1, \ldots, v_n)^T$ *and* $v_i = \frac{a_i^2}{d_i^2}$ *for* $i \in [n]$.

Proof. With out loss of generality, we assume that $v_1 \geq v_2 \geq \cdots \geq v_n$. We claim that a optimal solution y^* of the maximization problem is given by

$$y_i^* = \begin{cases} 1, & 0 \leq i \leq \lfloor \Gamma \rfloor \\ \Gamma - \lfloor \Gamma \rfloor, & i = \lfloor \Gamma \rfloor + 1 \\ 0, & else \end{cases}$$

Let $i < j$. Assume that a vector $\tilde{y}$ is optimal and $\tilde{y}_i < \tilde{y}_j$. Switching the i^{th} and j^{th} entry of $\tilde{y}$ does not decrease the objective value, since $v_i^2 \tilde{y}_i^2 + v_j^2 \tilde{y}_j^2 \leq v_i^2 \tilde{y}_j^2 + v_j^2 \tilde{y}_i^2$. Hence, we can assume without loss of generality that $y_i \geq y_j$. Note that every vector y can be seen as a weight distribution of Γ units of weight to n different components. Transferring $\epsilon > 0$ units of weight from the j^{th} component to the i^{th} component does not decrease the objective function, since

$$v_i^2 (y_i + \epsilon)^2 + v_j^2 (y_j - \epsilon)^2 - v_i^2 y_i^2 - v_j^2 y_j^2$$
$$= v_i^2 (y_i^2 + 2\epsilon y_i + \epsilon^2 - y_i^2) + v_j^2 (y_j^2 - 2\epsilon y_j + \epsilon^2 - y_j^2)$$

$$
\begin{aligned}
&= (v_i^2 + v_j^2)\epsilon^2 + 2\epsilon(v_i^2 y_i - v_j^2 y_j) \\
&\geq (v_i^2 + v_j^2)\epsilon^2 + 2\epsilon(v_j^2 y_i - v_j^2 y_i) \\
&= (v_i^2 + v_j^2)\epsilon^2 \geq 0.
\end{aligned}
$$

This argument proves that y^* is an optimal solution of the maximization problem. Therefore, the optimal value of the problem is given by $\sum_{i=1}^{\lfloor\Gamma\rfloor} v_i + (\Gamma - \lfloor\Gamma\rfloor)^2 v_{\lfloor\Gamma\rfloor+1}$. Rearranging yields the claimed value

$$
\begin{aligned}
\sum_{i=1}^{\lfloor\Gamma\rfloor} v_i + (\Gamma - \lfloor\Gamma\rfloor)^2 v_{\lfloor\Gamma\rfloor+1} &= ||v||^{(\lfloor\Gamma\rfloor)} + (\Gamma - \lfloor\Gamma\rfloor)^2 (||v||^{(\lfloor\Gamma\rfloor+1)} - ||v||^{(\lfloor\Gamma\rfloor)}) \\
&= ||v||^{(\lfloor\Gamma\rfloor)}(1 - (\Gamma - \lfloor\Gamma\rfloor)^2) + ||v||^{(\lfloor\Gamma\rfloor+1)}(\Gamma - \lfloor\Gamma\rfloor)^2.
\end{aligned}
$$

□

Lemma A.4. *Let $v_i > 0$ for $i \in [n]$ and $\overline{v} = \frac{1}{n}\sum_{i=1}^n v_i$. Then,*

$$
\frac{n}{\overline{v}} \leq \sum_{i=1}^n \frac{1}{v_i}.
$$

Proof. We have to show that $n^2 \leq \sum_{i=1}^n v_i \sum_{i=1}^n \frac{1}{v_i}$. This holds since

$$
\sum_{i=1}^n v_i \sum_{i=1}^n \frac{1}{v_i} = \sum_{i,j=1}^n \frac{v_j}{v_i} = \frac{1}{2}\sum_{i,j=1}^n \left(\frac{v_j}{v_i} + \frac{v_i}{v_j}\right) \geq \frac{1}{2}\sum_{i,j=1}^n 2 = n^2.
$$

□

Lemma A.5. *Let $v_i > 0$ for $i \in [n]$, $\overline{v} = \frac{1}{n}\sum_{i=1}^n v_i$, and $\hat{v} = (\overline{v}, \ldots, \overline{v})^T$. Then, for any $k \in [n]$,*

$$
||\hat{v}||^{(k)} < ||v||^{(k)}.
$$

Proof. Denote a reordered version of v by $v^\pi = (v_{\pi(1)}, \ldots, v_{\pi(n)})^T$ where $\pi \in \sigma(n)$ is a permutation of $\{1, \ldots, n\}$. Observe that the function $f(v) = ||v||^{(k)}$ is convex and that $f(v^\pi) = f(v)$ for all $\pi \in \sigma(n)$. Using these two properties concludes the proof, since

$$
f(\hat{v}) = f\left(\frac{1}{n!}\sum_{\pi\in\sigma(n)} v^\pi\right) \leq \frac{1}{n!}\sum_{\pi\in\sigma(n)} f(v^\pi) = \frac{n!}{n!} f(v) = f(v).
$$

□