An Active Set Algorithm for Robust Combinatorial Optimization Based on Separation Oracles

C. Buchheim[†] and M. De Santis[*]

[†]Fakultät für Mathematik
TU Dortmund
Vogelpothsweg 87 - 44227 Dortmund - Germany

[*] Dipartimento di Ingegneria Informatica, Automatica e Gestionale
Sapienza, Università di Roma
Via Ariosto, 25 - 00185 Roma -Italy

e-mail (Buchheim): christoph.buchheim@math.tu-dortmund.de
e-mail (De Santis): marianna.desantis@uniroma1.it

Abstract

We address combinatorial optimization problems with uncertain coefficients varying over ellipsoidal uncertainty sets. The robust counterpart of such a problem can be rewritten as a second-oder cone program (SOCP) with integrality constraints. We propose a branch-and-bound algorithm where dual bounds are computed by means of an active set algorithm. The latter is applied to the Lagrangian dual of the continuous relaxation, where the feasible set of the combinatorial problem is supposed to be given by a separation oracle. The method benefits from the closed form solution of the active set subproblems and from a smart update of pseudo-inverse matrices. We present numerical experiments on randomly generated instances and on instances from different combinatorial problems, including the shortest path and the traveling salesman problem, showing that our new algorithm consistently outperforms the state-of-the art mixed-integer SOCP solver of Gurobi.

Keywords. Robust Optimization, Active Set Methods, SOCP

1 Introduction

We address combinatorial optimization problems given in the general form

$$\min_{x \in P \cap \mathbb{Z}^n} \; c^\top x \tag{CP}$$

where $P \subseteq \mathbb{R}^n$ is a compact convex set, say $P \subseteq [l, u]$ with $l, u \in \mathbb{R}^n$, and the objective function vector $c \in \mathbb{R}^n$ is assumed to be uncertain. This setting appears in many applications where the feasible set is certain, but the objective function coefficients may have to be estimated or result from imprecise measurements. As an example, when searching for a shortest path in a road network, the topology of the network is usually considered fixed, but the travel times may vary depending on the traffic conditions.

A classical way of dealing with uncertain optimization problems is the strictly robust optimization approach, introduced in [3] for linear programming and in [2] for general convex programming; we also refer the reader to the book by Ben-Tal and Nemirovski [4]. In strictly robust optimization, we look for a worst-case solution, where the uncertain parameter c is assumed to belong to a bounded set $U \subseteq \mathbb{R}^n$, called the *uncertainty set*, and the goal of the robust counterpart is to compute the solution of the following min-max problem:

$$\min_{x \in P \cap \mathbb{Z}^n} \; \max_{c \in U} \; c^\top x \tag{RP}$$

A natural choice in this approach are ellipsoidal uncertainty sets, defined as

$$U = \{c \in \mathbb{R}^n \mid (c - \bar{c})^\top M(c - \bar{c}) \leq 1\},$$

where $M \in \mathbb{R}^{n \times n}$ is a symmetric positive definite matrix and $\bar{c} \in \mathbb{R}^n$ is the center of the ellipsoid. Assuming that the uncertain vector c in (CP), considered as a random variable, follows a normal distribution, we can interpret the ellipsoid U as a confidence set of c; in this case, M is the inverse covariance matrix of c and $\bar{c}$ is its expected value. Unfortunately, for ellipsoidal uncertainty sets, the robust counterpart (RP) is usually much harder to solve than the original problem (CP): it is known that Problem (RP) is NP-hard in this case for the shortest path problem, for the minimum spanning tree problem, and for the assignment problem [10] as well as for the unconstrained binary optimization problem [7].

Even in the case of a diagonal matrix M, i.e., when ignoring correlations and only taking variances into account, no polynomial time algorithm for the robust shortest path problem is known. There exists however an FPTAS for the diagonal case whenever the underlying problem (CP) admits an FPTAS [15], and polynomial time algorithms for the minimum spanning tree problem and the unconstrained binary problem have been devised for the diagonal case.

For general ellipsoids U, most exact solution approaches for (RP) are based on solving SOCPs. In fact, it is easy to see that the optimal solution of the inner maximization problem

$$\max_{c \in U} \; c^\top x$$

for fixed x is given by

$$\bar{c}^\top x + \sqrt{x^\top M^{-1} x}.$$

Therefore, Problem (RP) is equivalent to the integer non-linear problem

$$\begin{array}{ll} \min & f(x) = c^\top x + \sqrt{x^\top Q x} \\ \text{s.t.} & x \in P \cap \mathbb{Z}^n \end{array} \tag{P}$$

where $Q \in \mathbb{R}^{n \times n}$ is the symmetric and positive definite inverse of M and we replace $\bar{c}$ by c for ease of notation. Note that, when addressing so called value-at-risk models

$$\begin{array}{ll} \min & z \\ \text{s.t.} & \Pr(c^\top x \geq z) \leq \varepsilon \\ & x \in P \cap \mathbb{Z}^n \,, \end{array}$$

we arrive at essentially the same formulation (P), assuming normally distributed coefficients again; see, e.g., [15].

In the following, we assume that the convex set P is given by a separation algorithm, i.e., an algorithm that decides whether a given point $\bar{x} \in \mathbb{R}^n$ belongs to P or not, and, in the negative case, provides an inequality $a^\top x \leq b$ valid for P but violated by $\bar{x}$. Even in cases where the underlying problem (CP) is tractable, the polytope $\text{conv}(P \cap \mathbb{Z}^n)$ may have an exponential number of facets, so that a full linear description cannot be used efficiently. This is true, e.g., for the standard formulation of the spanning tree problem. However, we do not require that a complete linear description of $\text{conv}(P \cap \mathbb{Z}^n)$ be known; it suffices to have an integer linear description, i.e., we allow $P \neq \text{conv}(P \cap \mathbb{Z}^n)$. In particular, our approach can also be applied when the underlying problem is NP-hard, e.g., when (CP) models the traveling salesman problem.

As soon as P is given explicitly by linear constraints $Ax \leq b$ with $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$, the continuous relaxation of Problem (P) reduces to an SOCP of the form

$$\begin{array}{ll} \min & c^\top x + \sqrt{x^\top Q x} \\ \text{s.t.} & Ax \leq b \\ & x \in \mathbb{R}^n \,. \end{array} \tag{R2}$$

Such SOCPs can be solved efficiently using interior point algorithms [14] and popular solvers for SOCPs such as SeDuMi [17] or MOSEK [1] are based on interior point methods. However, in our branch-and-bound algorithm, we need to address a sequence of related SOCPs. Compared with interior point methods, active set methods have the advantage to allow warmstarting rules.

For this reason, in order to solve the SOCP relaxations of Problem (RP), we devised the active set algorithm `EllAS`. It is applied to the Lagrangian dual of (R2) and exploits the fact that the active set subproblems can be solved by closed form expressions. For this, the main ingredient is the pseudo-inverse of $AQ^{-\frac{1}{2}}$. Since the matrix A is updated in each iteration of the active set method, an incremental update of the pseudo-inverse is crucial for the running time of `EllAS`. Altogether, we can achieve a running time of $O(n^2)$ per iteration. Combined with an intelligent embedding into the branch-and-bound scheme, we obtain an algorithm that consistently outperforms the MISOCP solver of Gurobi 7.5.1, where the latter is either applied to a full linear description of P or, in case a compact linear description does not exist, uses the same separation oracle as `EllAS`.

The rest of the paper is organized as follows: the Lagrangian dual of (RP) is derived in Section 2. The closed-form solution of the resulting active set subproblems is developed in Section 3. The active set algorithm `EllAS` is detailed and analyzed in Sections 4 and 5. In Section 6, we discuss how to embed `EllAS` into a branch-and-bound algorithm. Numerical results for random integer instances as well as instances of different combinatorial optimization problems are reported in Section 7. Section 8 concludes.

2 Dual problem

The algorithm we propose for solving Problem (RP) uses the Lagrangian dual of relaxations of the form (R2). Let $\mathscr{L}(x,\lambda) : \mathbb{R}^n \times \mathbb{R}^m \to \mathbb{R}$ be the Lagrangian function associated to (R2):

$$\mathscr{L}(x,\lambda) = c^\top x + \sqrt{x^\top Q x} + \lambda^\top (Ax - b) \; .$$

The Lagrangian dual of Problem (R2) is then

$$\max_{\lambda \in \mathbb{R}^m_+} \; \inf_{x \in \mathbb{R}^n} \; \mathscr{L}(x,\lambda) \; . \tag{1}$$

After applying the bijective transformation $z = Q^{\frac{1}{2}} x$, the inner minimization problem of (1) becomes

$$-b^\top \lambda + \inf_{z \in \mathbb{R}^n} \left(Q^{-\frac{1}{2}}(c + A^\top \lambda)\right)^\top z + \|z\|$$

for fixed $\lambda \in \mathbb{R}^m_+$. It is easy to see that

$$\inf_{z \in \mathbb{R}^n} \left(Q^{-\frac{1}{2}}(c + A^\top \lambda)\right)^\top z + \|z\| = \min_{z \in \mathbb{R}^n} \left(Q^{-\frac{1}{2}}(c + A^\top \lambda)\right)^\top z + \|z\| = 0$$

if $\|Q^{-\frac{1}{2}}(c + A^\top \lambda)\| \leq 1$ and $-\infty$ otherwise. Therefore, Problem (1) reduces to

$$\begin{array}{ll} \max & -b^\top \lambda \\ \text{s.t.} & (c + A^\top \lambda)^\top Q^{-1} (c + A^\top \lambda) \leq 1 \\ & \lambda \geq 0 \; . \end{array} \tag{D}$$

Theorem 1. *For the primal-dual pair of optimization problems* (R2) *and* (D)*, strong duality holds as soon as one of the two problems is feasible. Moreover, if one of the problems admits an optimal solution, the same holds for the other problem.*

Proof. This follows from the convexity of (R2) and from the fact that all constraints in (R2) are affine linear. □ □

In order to solve Problem (R2), we have devised the dual active set algorithm `EllAS` detailed in Section 4. Along its iterations, `EllAS` produces dual feasible solutions of Problem (D), converging to a KKT point of Problem (R2) and therefore producing also a primal optimal solution when terminating.

3 Solving the Active Set Subproblem

At every iteration, the active set algorithm `EllAS` presented in the subsequent sections fixes certain dual variables to zero while leaving unconstrained the remaining variables. In the primal problem, this corresponds to choosing a set of valid linear constraints $Ax \leq b$ for P and replacing inequalities by equations. We thus need to solve primal-dual pairs of problems of the following type:

$$\begin{aligned} \min \quad & f(x) = c^\top x + \sqrt{x^\top Q x} \\ \text{s.t.} \quad & \hat{A}x = \hat{b} \\ & x \in \mathbb{R}^n \end{aligned} \tag{P–AS}$$

$$\begin{aligned} \max \quad & -\hat{b}^\top \lambda \\ \text{s.t.} \quad & (c + \hat{A}^\top \lambda)^\top Q^{-1} (c + \hat{A}^\top \lambda) \leq 1 \\ & \lambda \in \mathbb{R}^{\hat{m}} \end{aligned} \tag{D–AS}$$

where $\hat{A} \in \mathbb{R}^{\hat{m} \times n}$, $b \in \mathbb{R}^{\hat{m}}$. For the efficiency of our algorithm, it is crucial that this pair of problems can be solved in closed form. For this, the pseudo-inverse $(\hat{A}Q^{-\frac{1}{2}})^+$ of $\hat{A}Q^{-\frac{1}{2}}$ will play an important role. It can be used to compute orthogonal projections onto the kernel and onto the range of $Q^{-\frac{1}{2}}\hat{A}^\top$ as follows: we have

$$\text{proj}_{\ker(Q^{-\frac{1}{2}}\hat{A}^\top)}(y) = y - \hat{A}Q^{-\frac{1}{2}}(\hat{A}Q^{-\frac{1}{2}})^+ y \tag{2}$$

and

$$\text{proj}_{\text{ran}(Q^{-\frac{1}{2}}\hat{A}^\top)}(y) = (\hat{A}Q^{-\frac{1}{2}})^+ \hat{A}Q^{-\frac{1}{2}} y \,, \tag{3}$$

see e.g. [11]. We later explain how to update the pseudo-inverse incrementally instead of computing it from scratch in every iteration, which would take $O(n^3)$ time; see Section 5.2.

In the following, we assume that the dual problem (D–AS) admits a feasible solution; this will be guaranteed in every iteration of our algorithm; see Lemma 1 below.

3.1 Dual Unbounded Case

If $\hat{b} \notin \text{ran}(\hat{A})$, or equivalently, if $\hat{b}$ is not orthogonal to $\ker(\hat{A}^\top) = \ker(Q^{-\frac{1}{2}}\hat{A}^\top)$, then the dual problem (D–AS) is unbounded, and the corresponding primal problem (P–AS) is infeasible. When this case occurs, `EllAS` uses an unbounded direction of (D–AS) to continue. The set of unbounded directions of (D–AS) is $\ker(Q^{-\frac{1}{2}}\hat{A}^\top)$. Consequently, the unbounded direction with steepest ascent can be obtained by projecting the gradient of the objective function $-\hat{b}$ to $\ker(Q^{-\frac{1}{2}}\hat{A}^\top)$. According to (2), this projection is

$$\text{proj}_{\ker(Q^{-\frac{1}{2}}\hat{A}^\top)}(-\hat{b}) = (\hat{A}Q^{-\frac{1}{2}})(\hat{A}Q^{-\frac{1}{2}})^+ \hat{b} - \hat{b} \,.$$

3.2 Bounded Case

If $\hat{b} \in \text{ran}(A)$, we first consider the special case $\hat{b} = 0$. As we assume (D–AS) to be feasible, its optimum value is thus 0. Therefore, the corresponding primal problem (P–AS) admits $x^* = 0$ as optimal solution. In the following, we may thus assume $\hat{b} \neq 0$. The feasible set of problem (D–AS) consists of all $\lambda \in \mathbb{R}^{\hat{m}}$ such that

$$||Q^{-\frac{1}{2}}(c + \hat{A}^\top \lambda)|| \leq 1 \,,$$

i.e., such that the image of λ under $-Q^{-\frac{1}{2}}\hat{A}^\top$ belongs to the ball $B_1(Q^{-\frac{1}{2}}c)$. Consider the orthogonal projection of $Q^{-\frac{1}{2}}c$ to the subspace $\text{ran}(Q^{-\frac{1}{2}}\hat{A}^\top)$, which by (3) is

$$p := \text{proj}_{\text{ran}(Q^{-\frac{1}{2}}\hat{A}^\top)}(Q^{-\frac{1}{2}}c) = (Q^{-\frac{1}{2}}\hat{A}^\top)(Q^{-\frac{1}{2}}\hat{A}^\top)^+ Q^{-\frac{1}{2}}c \,.$$

If $||p - Q^{-\frac{1}{2}}c|| > 1$, then the intersection $B_1(Q^{-\frac{1}{2}}c) \cap \operatorname{ran}(Q^{-\frac{1}{2}}\hat{A}^\top)$ is empty, so that Problem (D–AS) is infeasible, contradicting our assumption. Hence, we have that this intersection is a ball with center p and radius

$$r := \sqrt{1 - ||p - Q^{-\frac{1}{2}}c||^2}$$

and $\lambda \in \mathbb{R}^{\hat{m}}$ is feasible for (D–AS) if and only if $-Q^{-\frac{1}{2}}\hat{A}^\top\lambda \in B_r(p)$. Since $\hat{b} \in \operatorname{ran}(\hat{A}Q^{-\frac{1}{2}})$, we have $(\hat{A}Q^{-\frac{1}{2}})(\hat{A}Q^{-\frac{1}{2}})^+\hat{b} = \hat{b}$. This allows us to rewrite the objective function $-\hat{b}^\top\lambda$ of (D–AS) in terms of $Q^{-\frac{1}{2}}\hat{A}^\top\lambda$ only, as

$$-\hat{b}^\top\lambda = -\hat{b}^\top(Q^{-\frac{1}{2}}\hat{A}^\top)^+(Q^{-\frac{1}{2}}\hat{A}^\top)\lambda \,.$$

We can thus first compute the optimal solution $v^* \in \operatorname{ran}(Q^{-\frac{1}{2}}\hat{A}^\top)$ of

$$\begin{array}{ll} \max & \hat{b}^\top(Q^{-\frac{1}{2}}\hat{A}^\top)^+v \\ \text{s.t.} & ||Q^{-\frac{1}{2}}c - v|| \leq 1 \,, \end{array}$$

which is unique since $\hat{b} \neq 0$, and then solve $v^* = -(Q^{-\frac{1}{2}}\hat{A}^\top)\lambda$. We obtain

$$v^* = p + \frac{r}{||(\hat{A}Q^{-\frac{1}{2}})^+\hat{b}||}(\hat{A}Q^{-\frac{1}{2}})^+\hat{b} \,, \tag{4}$$

so that we can state the following

Proposition 1. *Let $\hat{b} \in ran(A) \setminus \{0\}$ and let v^* be defined as in* (4)*. Then, the unique optimal solution of* (D–AS) *with minimal norm is*

$$\lambda^* := -(Q^{-\frac{1}{2}}\hat{A}^\top)^+v^*.$$

From λ^*, it is possible to compute an optimal solution x^* of the primal problem (P–AS) as explained in the following result.

Theorem 2. *Let $\hat{b} \in ran(A) \setminus \{0\}$. Let λ^* be an optimal solution of* (D–AS) *and $\bar{x} := Q^{-1}(c + \hat{A}^\top\lambda^*)$.*

(a) If $\hat{b}^\top\lambda^* \neq 0$, then the unique optimal solution of (P–AS) *is $x^* = \alpha\bar{x}$, with*

$$\alpha := -\frac{\hat{b}^\top\lambda^*}{c^\top\bar{x} - \sqrt{\bar{x}^\top Q\bar{x}}} \,.$$

(b) Otherwise, there exists a unique $\alpha < 0$ such that $\alpha\hat{A}\bar{x} = \hat{b}$. Then, $x^* = \alpha\bar{x}$ is the unique optimal solution of (P–AS)*.*

Proof. Let (x^*, λ^*) be a primal-dual optimal pair, which exists by Theorem 1. Since $\hat{b} \neq 0$ and $\hat{A}x^* = \hat{b}$, it follows that $x^* \neq 0$. The gradient equation yields

$$0 = \nabla_x\mathscr{L}(x^*, \lambda^*) = c + \frac{2Qx^*}{2\sqrt{(x^*)^\top Q(x^*)}} + \hat{A}^\top\lambda^*$$

which is equivalent to

$$\frac{Q^{\frac{1}{2}}x^*}{\|Q^{\frac{1}{2}}x^*\|} = -Q^{-\frac{1}{2}}(c + \hat{A}^\top\lambda^*)$$

and hence to

$$x^* = \alpha Q^{-1}(c + \hat{A}^\top\lambda^*) = \alpha\bar{x}$$

for some $\alpha \neq 0$. Since $\alpha = -\|Q^{\frac{1}{2}}x^*\|$, we have $\alpha < 0$. By strong duality, we then obtain

$$-\hat{b}^\top\lambda^* = c^\top x^* + \sqrt{(x^*)^\top Q(x^*)} = \alpha c^\top\bar{x} + |\alpha|\sqrt{\bar{x}^\top Q\bar{x}} = \alpha\big(c^\top\bar{x} - \sqrt{\bar{x}^\top Q\bar{x}}\big) \,.$$

Now if $\hat{b}^\top \lambda^* \neq 0$, also the right hand side of this equation is non-zero, and we obtain α as claimed. Otherwise, it still holds that there exists $\alpha < 0$ such that $\alpha\bar{x}$ is optimal. In particular, $\alpha\bar{x}$ is primal feasible and hence $\alpha\hat{A}\bar{x} = \hat{A}(\alpha\bar{x}) = \hat{b}$. As $\hat{b} \neq 0$, we derive $\hat{A}\bar{x} \neq 0$, as $\alpha < 0$. This in particular shows that α is uniquely defined by $\alpha\hat{A}\bar{x} = \hat{b}$. □ □

Note that the proof (and hence the statement) for case (b) in Theorem 2 are formally applicable also in case (a). However, in the much more relevant case (a), we are able to derive a closed formula for α in a more direct way.

4 The Dual Active Set Method `EllAS`

As all active set methods, our algorithm `EllAS` tries to forecast the set of constraints that are active at the optimal solution of the primal-dual pair (R2) and (D), adapting this forecast iteratively: starting from a subset of primal constraints $A^{(1)}x \leq b^{(1)}$, where $A^{(1)} \in \mathbb{R}^{m^{(1)} \times n}$ and $b^{(1)} \in \mathbb{R}^{m^{(1)}}$, one constraint is removed or added per iteration by performing a dual or a primal step; see Algorithm 1. We assume that a corresponding dual feasible solutions $\lambda^{(1)} \geq 0$ is given when starting the algorithm; we explain below how to obtain this initial solution.

Algorithm 1 Ellipsoidal Active SeT algorithm `EllAS`

Input: $Q \in \mathbb{R}^{n \times n}$, $c \in \mathbb{R}^n$, $A^{(1)} \in \mathbb{R}^{m^{(1)} \times n}$, $b^{(1)} \in \mathbb{R}^{m^{(1)}}$;
$\lambda^{(1)} \geq 0$ with $(c + (A^{(1)})^\top \lambda^{(1)})^\top Q^{-1}(c + (A^{(1)})^\top \lambda^{(1)}) \leq 1$;
pseudo-inverse $(A^{(1)}Q^{-\frac{1}{2}})^+$

Output: optimal solutions of (R2) and (D)

1: **for** $k = 1, 2, 3, \ldots$ **do**
2: **solve** (D-ASk) and obtain optimal $\tilde{\lambda}^{(k)}$ with minimal norm
3: **if** problem (D-ASk) is bounded and $\tilde{\lambda}^{(k)} \geq 0$ **then**
4: **set** $\lambda^{(k)} := \tilde{\lambda}^{(k)}$
5: **perform** the primal step (Algorithm 3) and **update** $x^{(k)}$, $A^{(k)}$, $b^{(k)}$
6: **else**
7: **perform** the dual step (Algorithm 2) and **update** $\lambda^{(k)}$, $A^{(k)}$, $b^{(k)}$
8: **end if**
9: **end for**

At every iteration k, in order to decide if performing the primal or the dual step, the dual subproblem is addressed, namely Problem (D) where only the subset of active constraints is taken into account. This leads to the following problem:

$$\begin{array}{ll} \max & -{b^{(k)}}^\top \lambda \\ \text{s.t.} & (c + {A^{(k)}}^\top \lambda)^\top Q^{-1}(c + {A^{(k)}}^\top \lambda) \leq 1 \\ & \lambda \in \mathbb{R}^{m^{(k)}} \end{array} \qquad \text{(D-ASk)}$$

The solution of Problem (D-ASk) has been explained in Section 3. Note that formally Problem (D-ASk) is defined in a smaller space with respect to Problem (D), but its solutions can also be considered as elements of $\mathbb{R}^m$ by setting the remaining variables to zero.

In case the dual step is performed, the solution of Problem (D-ASk) gives an ascent direction $p^{(k)}$ along which we move in order to produce a new dual feasible point with better objective function value.

We set

$$\lambda^{(k)} = \lambda^{(k-1)} + \alpha^{(k)} p^{(k)},$$

where the steplength $\alpha^{(k)}$ is chosen to be the largest value for which non-negativity is maintained at all entries. Note that the feasibility with respect to the ellipsoidal constraint in (D), i.e.,

$$(c + A^\top \lambda)^\top Q^{-1} (c + A^\top \lambda) \leq 1 \,,$$

is guaranteed from how p^k is computed, using convexity. Therefore, $\alpha^{(k)}$ can be derived by considering the negative entries of $p^{(k)}$. In order to maximize the increase of $-b^\top \lambda$, we ask $\alpha^{(k)}$ to be as large as possible subject to maintaining non-negativity; see Steps 9–10 in Algorithm 2.

Algorithm 2 Dual Step

1: **if** problem (D-ASk) is bounded **then**
2: **set** $p^{(k)} := \tilde{\lambda}^{(k)} - \lambda^{(k-1)}$
3: **else**
4: **let** $p^{(k)}$ be an unbounded direction of (D-ASk) with steepest ascent
5: **if** $p^{(k)} \geq 0$ **then**
6: **STOP:** primal problem is infeasible
7: **end if**
8: **end if**
9: **choose** $j \in \operatorname{argmin} \{-\lambda_i^{(k-1)}/p_i^{(k)} \mid i = 1, \ldots, m^{(k)},\ p_i^{(k)} < 0\}$
10: **set** $\alpha^{(k)} := -\lambda_j^{(k-1)}/p_j^{(k)}$
11: **set** $\lambda^{(k)} := \lambda^{(k-1)} + \alpha^{(k)} p^{(k)}$
12: **compute** $(A^{(k+1)}, b^{(k+1)})$ by removing row j in $(A^{(k)}, b^{(k)})$
13: **compute** $\lambda^{(k+1)}$ by removing entry j in $\lambda^{(k)}$
14: **set** $m^{(k+1)} := m^{(k)} - 1$
15: **update** $(A^{(k+1)} Q^{-\frac{1}{2}})^+$ from $(A^{(k)} Q^{-\frac{1}{2}})^+$

The constraint index j computed in Step 9 of Algorithm 2 corresponds to the primal constraint that needs to be released from the active set. The new iterate $\lambda^{(k+1)}$ is then obtained from $\lambda^{(k)}$, by dropping the j-th entry.

Proposition 2. *The set considered in Step 9 of Algorithm 2 is non-empty.*

Proof. If Problem (D-ASk) is bounded, there is an index i such that $\tilde{\lambda}_i^{(k)} < 0$, since $\tilde{\lambda}^{(k)}$ is dual infeasible. As $\lambda^{(k-1)} \geq 0$, we derive $p_i^{(k)} = \tilde{\lambda}_i^{(k)} - \lambda_i^{(k-1)} < 0$. If Problem (D-ASk) is unbounded, we explicitly check whether $p^{(k)} \geq 0$ and only continue otherwise. □ □

The primal step is performed in case the solution of Problem (D-ASk) gives us a dual feasible solution. Starting from this dual feasible solution, we compute a corresponding primal solution $x^{(k)}$ according to the formula in Theorem 2. If $x^{(k)}$ belongs to P we are done: we have that $(x^{(k)}, \lambda^{(k)})$ is a KKT point of Problem (R2) and, by convexity of Problem (R2), $x^{(k)}$ is its global optimum. Otherwise, we compute a cutting plane violated by $x^{(k)}$ that can be considered active and will be then taken into account in defining the dual subproblem (D-ASk) at the next iteration. The new iterate $\lambda^{(k+1)}$ is obtained from $\lambda^{(k)}$ by adding an entry to $\lambda^{(k)}$ and setting this additional entry to zero.

Theorem 3. *Whenever Algorithm* `EllAS` *terminates, the result is correct.*

Algorithm 3 `Primal Step`

if $-(b^{(k)})^\top \lambda^{(k)} = 0$ **then**
  STOP: $(0, \lambda^{(k)})$ is an optimal primal-dual solution
else
  compute $x^{(k)}$ from $\lambda^{(k)}$ according to Theorem 2
  if $x^{(k)} \in P$ **then**
    STOP: $(x^{(k)}, \lambda^{(k)})$ is an optimal primal-dual solution
  else
    compute a cutting plane $a^\top x \leq b$ violated by $x^{(k)}$
    compute $(A^{(k+1)}, b^{(k+1)})$ by appending $(a^\top, b)$ to $(A^{(k)}, b^{(k)})$
    compute $\lambda^{(k+1)}$ by appending zero to $\lambda^{(k)}$
    set $m^{(k+1)} := m^{(k)} + 1$
    update $(A^{(k+1)}Q^{-\frac{1}{2}})^+$ from $(A^{(k)}Q^{-\frac{1}{2}})^+$
  end if
end if

Proof. If Algorithm `EllAS` stops at the primal step, the optimality of the resulting primal-dual pair follows from the discussion in Section 3. If Algorithm `EllAS` stops at the dual step, it means that the ascent direction $p^{(k)}$ computed is a feasible unbounded direction for Problem (D), so that Problem (D) is unbounded and hence Problem (R2) is infeasible. □ □

It remains to describe how to initialize `EllAS`. For this, we use the assumption of boundedness of P and construct $A^{(1)}$, $b^{(1)}$, and $\lambda^{(1)}$ as follows: for each $i = 1, \ldots, n$, we add the constraint $x_i \leq u_i$ if $c_i < 0$, with corresponding $\lambda_i := -c_i$, and the constraint $-x_i \leq -l_i$ otherwise, with $\lambda_i := c_i$. These constraints are valid since we assumed $P \subseteq [l, u]$ and it is easy to check that $(A^{(1)})^\top \lambda^{(1)} = -c$ by construction, so that $\lambda^{(1)}$ is dual feasible for (D). Moreover, we can easily compute $(A^{(1)}Q^{-\frac{1}{2}})^+$ in this case, as $A^{(1)}$ is a diagonal matrix with ± 1 entries: this implies $(A^{(1)}Q^{-\frac{1}{2}})^+ = Q^{\frac{1}{2}}A^{(1)}$.

5 Analysis of the Algorithm

In this section, we show that Algorithm `EllAS` converges in a finite number of steps if cycling is avoided. Moreover, we prove that the running time per iteration can be bounded by $O(n^2)$, if implemented properly.

5.1 Convergence Analysis

Our convergence analysis follows similar arguments to those used in [16] for the analysis of primal active set methods for strictly convex quadratic programming problems. In particular, as in [16], we assume that we can always take a nonzero steplength along the ascent direction. Under this assumption we will show that Algorithm `EllAS` does not undergo cycling, or, in other words, this assumption prevents from having $\lambda^{(k)} = \lambda^{(l)}$ and $(A^{(k)}, b^{(k)}) = (A^{(l)}, b^{(l)})$ in two different iterations k and l. As for other active set methods, it is very unlikely in practice to encounter a zero steplength. However, there are techniques to avoid cycling even theoretically, such as perturbation or lexicographic pivoting rules in Step 9 of Algorithm 2.

Lemma 1. *At every iteration k of Algorithm `EllAS`, Problem* (D-ASk) *admits a feasible solution.*

Proof. It suffices to show that the ellipsoidal constraint

$$(c + {A^{(k)}}^\top \lambda^{(k)})^\top Q^{-1} (c + {A^{(k)}}^\top \lambda^{(k)}) \leq 1 \tag{5}$$

is satisfied for each k. For $k = 1$, this is explicitely required for the input of Algorithm `EllAS`. Let $\lambda^{(k)}$ be computed from $\lambda^{(k-1)}$ by moving along the direction $p^{(k)}$. The feasibility of $\lambda^{(k)}$ with respect to (5) then follows from the definition of $p^{(k)}$ and from the convexity of the ellipsoid. □ □

Proposition 3. *At every iteration k of Algorithm `EllAS`, the vector $\lambda^{(k)}$ is feasible for* (D).

Proof. Taking into account the proof of Lemma 1, it remains to show nonnegativity of $\lambda^{(k)}$, which is guaranteed by the choice of the steplength $\alpha^{(k)}$. □ □

Proposition 4. *Assume that the steplength α^k is always non-zero in the dual step. If Algorithm `EllAS` does not stop at iteration k, then one of the following holds:*

(i) $-b^{(k+1)^\top}\lambda^{(k+1)} > -b^{(k)^\top}\lambda^{(k)}$;

(ii) $-b^{(k+1)^\top}\lambda^{(k+1)} = -b^{(k)^\top}\lambda^{(k)}$ *and* $\|\lambda^{(k+1)}\| < \|\lambda^{(k)}\|$.

Proof. In the primal step, suppose that $\tilde{\lambda}^{(k)} \geq 0$ solves Problem (D-ASk) and that the corresponding unique primal solution satisfies $x^{(k)} \notin P$. After adding a violated cutting plane, the optimal value of Problem (P–AS) strictly increases and the same is true for the optimal value of Problem (D–AS) by strong duality. Then,

$$p^{(k+1)} = \tilde{\lambda}^{(k+1)} - \lambda^{(k)} = \tilde{\lambda}^{(k+1)} - \tilde{\lambda}^{(k)}$$

is a strict ascent direction for $-b^\top\lambda$ and case (i) holds.

In the dual step, if $p^{(k+1)}$ is an unbounded direction, case (i) holds again. Otherwise, observe that $\lambda^{(k)} \neq \tilde{\lambda}^{(k+1)}$, as $\tilde{\lambda}^{(k+1)}$ is not feasible with respect to the nonnegativity constraints. Then, since $\tilde{\lambda}^{(k+1)}$ is the unique optimal solution for Problem (D-ASk) with minimal norm, $p^{(k+1)} = \tilde{\lambda}^{(k+1)} - \lambda^{(k)}$ is either a strict ascent direction for $-b^\top\lambda$, or $-b^\top p^{(k+1)} = 0$ and $p^{(k+1)}$ is a strict descent direction for $\|\lambda\|$, so that case (ii) holds. □ □

Lemma 2. *At every iteration k of Algorithm `EllAS`, we have $m^{(k)} \leq n + 1$. Furthermore, if Algorithm `EllAS` terminates at iteration k with an optimal primal-dual pair, then $m^{(k)} \leq n$.*

Proof. As only violated cuts are added, the primal constraints $A^{(k)}x = b^{(k)}$ either form an infeasible system or are linearly independent. If $m^{(k)} = n + 1$, the primal problem is hence infeasible. Thus Problem (D-ASk) is unbounded, so that at iteration k a dual step is performed and a dependent row of $(A^{(k)}, b^{(k)})$ is deleted, leading to an independent set of constraints again. □ □

Theorem 4. *Assume that whenever a dual step is performed, Algorithm `EllAS` takes a non-zero steplength α^k. Then, after at most $n2^m$ iterations, Algorithm `EllAS` terminates with a primal-dual pair of optimal solutions for* (R2) *and* (D).

Proof. First note that, by Lemma 2, at most n dual steps can be performed in a row. Hence, it is enough to show that in any two iterations $k \neq l$ where a primal step is performed, we have $(A^{(k)}, b^{(k)}) \neq (A^{(l)}, b^{(l)})$. Otherwise, assuming $(A^{(k)}, b^{(k)}) = (A^{(l)}, b^{(l)})$, we obtain $\tilde{\lambda}^{(k)} = \tilde{\lambda}^{(l)}$ and hence $\lambda^{(k)} = \lambda^{(l)}$. This leads to a contradiction to Proposition 4. □ □

5.2 Running time per iteration

The running time in iteration k of `EllAS` is $O(m^{(k)}n)$ and hence linear in the size of the matrix $A^{(k)}$, if implemented properly. The main work is to keep the pseudo-inverse $(A^{(k)}Q^{-\frac{1}{2}})^+$ up-to-date. Since $A^{(k)}Q^{-\frac{1}{2}}$ is only extended or shrunk by one row in each iteration, an update of $(A^{(k)}Q^{-\frac{1}{2}})^+$ is possible in $O(m^{(k)}n)$ time by a generalization of the Sherman-Morrison-formula [12]. Exploiting the fact that the matrix $A^{(k)}$

has full row rank in most iterations, we can proceed as follows. If $A^{(k+1)}$ is obtained from $A^{(k)}$ by adding a new row a, we first compute the row vectors

$$h := aQ^{-\frac{1}{2}}(A^{(k)}Q^{-\frac{1}{2}})^+, \quad v := aQ^{-\frac{1}{2}} - hA^{(k)}Q^{-\frac{1}{2}} .$$

Now $v \neq 0$ if and only if $A^{(k+1)}$ has full row rank, and in the latter case

$$(A^{(k+1)}Q^{-\frac{1}{2}})^+ = \Big((A^{(k)}Q^{-\frac{1}{2}})^+ \mid 0\Big) - \tfrac{1}{||v||^2}v^\top(h \mid -1) .$$

Otherwise, if $v = 0$, we are adding a linearly dependent row to $A^{(k)}$, making the primal problem (P–AS) infeasible. In this case, an unbounded direction of steepest ascent of (D–AS) is given by $(-h \mid 1)^\top$ and the next step will be a dual step, meaning that a row will be removed from $A^{(k+1)}$ and the resulting matrix $A^{(k+2)}$ will have full row rank again. We can thus update $(A^{(k)}Q^{-\frac{1}{2}})^+$ to $(A^{(k+2)}Q^{-\frac{1}{2}})^+$ by first removing and then adding a row, in both cases having full row rank.

It thus remains to deal with the case of deleting the r-th row a of a matrix $A^{(k)}$ with full row rank. Here we obtain $(A^{(k+1)}Q^{-\frac{1}{2}})^+$ by deleting the r-th column in

$$(A^{(k)}Q^{-\frac{1}{2}})^+ - \tfrac{1}{||w||^2}ww^\top(A^{(k)}Q^{-\frac{1}{2}})^+ ,$$

where w is the r-th column of $(A^{(k)}Q^{-\frac{1}{2}})^+$.

Theorem 5. *The running time per iteration of Algorithm* `EllAS` *is* $O(n^2)$.

Proof. This follows directly from Lemma 2 and the discussion above. □ □

Clearly, the incremental update of the pseudo-inverse $(A^{(k)}Q^{-\frac{1}{2}})^+$ may cause numerical errors. This can be avoided by recomputing it from scratch after a certain number of incremental updates. Instead of a fixed number of iterations, we recompute $(A^{(k)}Q^{-\frac{1}{2}})^+$ whenever the primal solution computed in a primal step is infeasible, i.e., violates the current constraints, where we allow a small tolerance.

In order to avoid wrong solutions even when pseudo-inverses are not precise, we make sure in our implementation that the dual solution $\lambda^{(k)}$ remains feasible for (D–AS) in each iteration, no matter how big the error of $(A^{(k)}Q^{-\frac{1}{2}})^+$ is. For this, we slightly change the computation of $\tilde{\lambda}^{(k)}$: after computing $\tilde{\lambda}^{(k)}$ exactly as explained, we determine the largest $\delta \in \mathbb{R}$ such that $(1-\delta)\lambda^{(k-1)} + \delta\tilde{\lambda}^{(k)}$ is dual feasible. Such δ must exist since $\lambda^{(k-1)}$ is dual feasible, and it can easily be computed using the midnight formula. We then replace $\tilde{\lambda}^{(k)}$ by $(1-\delta)\lambda^{(k-1)} + \delta\tilde{\lambda}^{(k)}$ and go on as before.

6 Branch-and-Bound Algorithm

For solving the integer Problem (RP), the method presented in the previous sections must be embedded into a branch-and-bound scheme. The dual bounds are computed by Algorithm `EllAS` and the branching is done by splitting up the domain $[l_i, u_i]$ of some variable x_i. Several properties of Algorithm `EllAS` can be exploited to improve the performance of such a branch-and-bound approach.

Warm starts Clearly, as branching adds new constraints to the primal feasible region of the problem, while never extending it, all dual solutions remain feasible. In every node of the branch-and-bound-tree, the active set algorithm can thus be warm started with the optimal set of constraints of the parent node. As in [5, 6], this leads to a significant reduction of the number of iterations compared to a cold start. Moreover, the newly introduced bound constraint is always violated and can be directly added as a new active constraint, which avoids resolving the same dual problem and hence saves one more iteration per node. Finally, the data describing the problem can either be inherited without changes or updated quickly; this is particularly important for the pseudo-inverse $(AQ^{-\frac{1}{2}})^+$.

Early pruning Since we compute a valid dual bound for Problem (RP) in every iteration of Algorithm `EllAS`, we may prune a subproblem as soon as the current bound exceeds the value of the best known feasible solution.

Avoiding cycling or tailing off Last but not least, we may also stop Algorithm `EllAS` at every point without compromising the correctness of the branch-and-bound algorithm. In particular, we can stop as soon as an iteration of Algorithm `EllAS` does not give a strict (or a significant) improvement in the dual bound. In particular, this avoids cycling.

7 Numerical Results

To test the performance of our algorithm `EllAS`, we considered random binary instances with up to one million constraints (Section 7.1) as well as combinatorial instances of Problem (RP), where the underlying problem is the Shortest Path problem (Section 7.2), the Assignment problem (Section 7.3), the Spanning Tree problem (Section 7.4), and the Traveling Salesman problem (Section 7.5). Concerning our approach, these combinatorial problems have different characteristics: while the first two problems have compact and complete linear formulations, the standard models for the latter problems use an exponential number of constraints that can be separated efficiently. In the case of the Spanning Tree problem, this exponential set of constraints again yields a complete linear formulation, while this is not the case for the NP-hard Traveling Salesman problem. In the latter case, however, we still have a complete integer programming formulation, which suffices for the correctness of our approach.

For all problems, we consider instances where the positive definite matrix $Q \in \mathbb{R}^{n\times n}$ is randomly generated. For this, we chose n eigenvalues λ_i uniformly at random from $[0, 1]$ and orthonormalized n random vectors v_i, each entry of which was chosen uniformly at random from $[-1, 1]$, then we set $Q = \sum_{i=1}^{n} \lambda_i v_i v_i^\top$. For the random binary instances, the entries of c were chosen uniformly at random from $[-1, 1]$, while for all remaining instances the vector c was uniformly one.

In the following, we present a comparison of `BB-EllAS`, a C++ implementation of the branch-and-bound-algorithm based on `EllAS`, with the MISOCP solver of Gurobi 7.5.1 [9]. According to the latest benchmark results of Hans D. Mittelmann [13], Gurobi is currently the fastest solver for MISOCPs. We use Gurobi with standard settings, except that we use the same optimality tolerance as in `BB-EllAS`, setting the absolute optimality tolerance `MIPGapAbs` to 10^{-4}. All other standard parameters are unchanged. In particular, Gurobi uses presolve techniques that decrease the solution times significantly. In case of the Spanning Tree problem and the Traveling Salesman problem, we apply dynamic separation algorithms using a callback adding lazy constraints.

All our experiments were carried out on Intel Xeon processors running at 2.60 GHz. All running times were measured in CPU seconds and the time-limit was set to one CPU hour for each individual instance. All tables presented in this section include the following data for the comparison between `BB-EllAS` and Gurobi: the number of instances solved within the time limit, the average running time, and the average number of branch-and-bound nodes. For `BB-EllAS`, we also report the average total number of active set iterations and the average number of times the pseudo-inverse $(A^{(k)}Q^{-\frac{1}{2}})^+$ is recomputed from scratch, the latter in percentage with respect to the number of iterations. All averages are taken over the set of instances solved within the time limit. For all applications, we also present performance profiles, as proposed in [8]. Given our set of solvers $\mathcal{S}$=$\{$`BB-EllAS`, Gurobi$\}$ and a set of problems $\mathcal{P}$, we compare the performance of a solver $s \in \mathcal{S}$ on problem $p \in \mathcal{P}$ against the best performance obtained by any solver in $\mathcal{S}$ on the same problem. To this end we define the performance ratio $r_{p,s} = t_{p,s}/\min\{t_{p,s'} : s' \in \mathcal{S}\}$, where $t_{p,s}$ is the computational time, and we consider a cumulative distribution function

$$\rho_s(\tau) = |\{p \in \mathcal{P} : \ r_{p,s} \leq \tau\}|/|\mathcal{P}|.$$

The performance profile for $s \in S$ is the plot of the function ρ_s.

Table 1: Comparison on random binary instances.

		BB-EllAS					Gurobi		
n	m	#sol	time	nodes	iter	%ps	#sol	time	nodes
25	10^3	10	0.00	3.9e+01	2.5e+02	0.12	10	0.76	1.3e+01
25	10^4	10	0.03	6.5e+01	4.9e+02	0.45	10	9.39	2.1e+01
25	10^5	10	0.60	1.0e+02	9.7e+02	1.22	10	156.63	2.8e+01
25	10^6	10	16.91	2.5e+02	2.5e+03	0.85	10	1973.87	8.6e+01
50	10^3	10	0.01	6.3e+01	3.5e+02	0.54	10	0.96	1.1e+01
50	10^4	10	0.05	6.7e+01	4.4e+02	0.62	10	11.93	1.8e+01
50	10^5	10	0.85	7.8e+01	6.8e+02	1.13	10	246.32	2.1e+01
50	10^6	10	18.84	1.7e+02	1.7e+03	1.22	0	—	—
100	10^3	10	0.40	1.3e+02	8.3e+02	6.79	10	2.35	2.5e+01
100	10^4	9	0.36	1.6e+02	1.1e+03	1.78	10	27.51	7.7e+01
100	10^5	9	4.26	2.6e+02	2.0e+03	1.60	10	761.07	2.3e+02
100	10^6	7	94.88	4.9e+02	4.8e+03	3.68	0	—	—
200	10^3	10	1.00	1.3e+02	7.6e+02	3.98	10	4.55	3.1e+01
200	10^4	8	2.04	1.6e+02	1.3e+03	3.92	10	23.63	4.1e+01
200	10^5	9	11.84	3.1e+02	2.6e+03	3.15	10	899.84	1.3e+02
200	10^6	7	61.25	1.9e+02	1.4e+03	14.11	0	—	—

7.1 Random Instances

For a first comparison, we consider instances of Problem (P) where the objective function vector $c \in \mathbb{R}^n$ and the positive definite matrix $Q \in \mathbb{R}^{n\times n}$ are randomly generated as described above. The set P is explicitely given as $\{x \in \mathbb{R}^n \mid Ax \leq b\}$, where $A \in \mathbb{R}^{m\times n}$ and $b \in \mathbb{R}^m$ are also randomly generated: the entries of A were chosen uniformly at random from the integers in the range $[0, 10]$ and b was defined by $b_i = \lfloor \frac{1}{2}\sum_{j=1}^{n} a_{ij} \rfloor$, $i = 1, \ldots, m$. Altogether, we generated 160 different problem instances for (P): for each combination of $n \in \{25, 50, 100, 200\}$ and $m \in \{10^3, 10^4, 10^5, 10^6\}$, we generated 10 instances. Since the set P is explicitely given here, the linear constraints are separated by enumeration in `BB-EllAS`. More precisely, at Step 8 of Algorithm 3, we pick the linear constraint most violated by $x^{(k)}$. We report our results in Table 1.

From the results in Table 1, note that the average number of branch-and-bound nodes enumerated by `BB-EllAS` is generally larger than the number of nodes needed by Gurobi, but always by less than a factor of 10 on average. However, in terms of running times, `BB-EllAS` outperforms Gurobi on all instance types except for the larger instances with a medium number of constraints, i.e., for $n \in \{100, 200\}$ and $m \in \{10^4, 10^5\}$. On all other instance classes, `BB-EllAS` either solves significantly more instances than Gurobi within the time limit or has a faster running time by many orders of magnitude. This in confirmed by the performance profiles presented in Figure 1. The low number of iterations performed by `EllAS` per node (less than 10 on average) highlights the benefits of using warmstarts.

7.2 Shortest Path Problem

Given a directed graph $G = (V, E)$, where V is the set of vertices and E is the set of edges, and weights associated with each edge, the Shortest Path problem is the problem of finding a path between two vertices s and t such that the sum of the weights of its constituent edges is minimized. Our approach

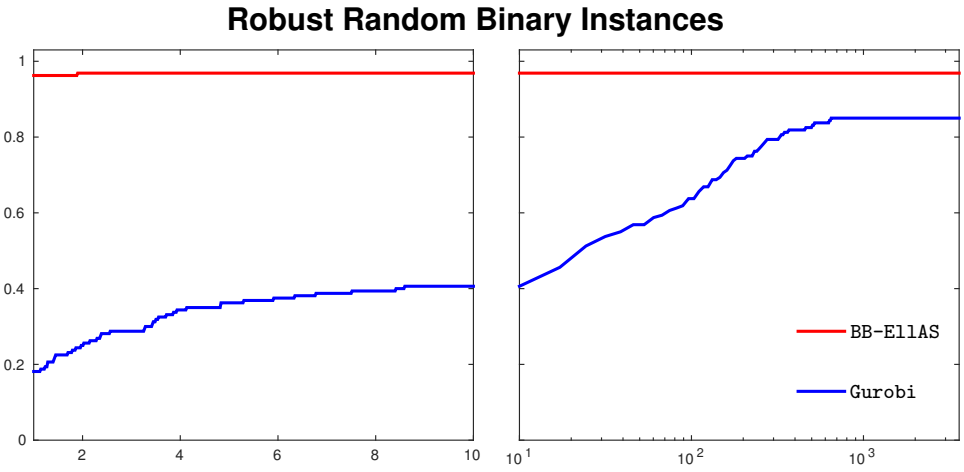

Figure 1: Performance profile with respect to running times for random binary instances.

uses the following flow based formulation of the Robust Shortest Path problem:

$$\begin{array}{lrcll} \min & c^\top x + \sqrt{x^\top Q x} & & & \\ \text{s.t.} & \sum_{e\in\delta^+(i)} x_e - \sum_{e\in\delta^-(i)} x_e & = & 0 & \forall i \in V \setminus \{s,t\} \\ & \sum_{e\in\delta^+(s)} x_e - \sum_{e\in\delta^-(s)} x_e & = & 1 & \\ & \sum_{e\in\delta^+(t)} x_e - \sum_{e\in\delta^-(t)} x_e & = & -1 & \\ & x & \in & \{0,1\}^E & \end{array} \tag{6}$$

In our test set, we produced squared grid graphs with r rows and columns, where all edges point from left to right and from top to bottom. In this way, we produced graphs with $|V| = r^2$ vertices and $|E| = 2r^2 - 2r$ edges. In the IP model (6), we thus have $n := 2r^2 - 2r$ variables and $m := |E| + 2|V| = 4r^2 - 2r$ many inequalities, taking into account also the box constraints $x_e \in [0,1]$ for $e \in E$. Since this number is polynomial in n, we can separate them by enumeration within `EllAS`, whereas we can pass the formulation (6) to Gurobi directly. Concerning the objective function of (6), we set all expected lenghts c_i to 1 and built the positive definite matrix Q as described above. Altogether, we generated 100 different problem instances for (6): for each $r \in \{5, \ldots, 14\}$ we generated 10 instances.

In Table 2, we report the comparison between `BB-EllAS` and the MISOCP solver of Gurobi. The average number of branch-and-bound nodes needed in `BB-EllAS` is in the same order of magnitude of that needed by Gurobi. However, `EllAS` is able to process the nodes very quickly, leading to a branch-and-bound scheme that outperforms Gurobi in terms of computational time. Note also that for graphs with $r = 13$, Gurobi does not solve any instance within one hour CPU time, while `BB-EllAS` is able to solve 3 of them. Both solvers fail for instances with $r \geq 14$. See Figure 2 for the performance profiles.

7.3 Assigment Problem

Given an undirected, bipartite and weighted graph $G = (V, E)$ with bipartition $V = V_1 \cup V_2$, the Assignment problem consists in finding a one-to-one assignment from the nodes in V_1 to the nodes in V_2 such that the sum of the weights of the edges used for the assignment is minimized. In other words, we search for a minimum-weight perfect matching in the bipartite graph G. Our approach uses the following

Table 2: Comparison on robust shortest path instances.

			BB-EllAS					Gurobi		
r	n	m	#sol	time	nodes	iter	%ps	#sol	time	nodes
5	40	90	10	0.01	4.4e+01	4.9e+02	0.00	10	0.17	5.0e+01
6	60	132	10	0.04	1.2e+02	1.5e+03	0.45	10	0.65	1.4e+02
7	84	182	10	1.26	3.8e+02	5.4e+03	5.51	10	2.29	3.5e+02
8	112	240	10	2.52	9.7e+02	1.6e+04	1.68	10	8.86	8.9e+02
9	144	306	10	5.71	2.5e+03	4.7e+04	0.42	10	92.12	3.0e+03
10	180	380	10	67.68	6.0e+03	1.3e+05	1.99	10	349.19	6.6e+03
11	220	462	10	214.08	1.9e+04	4.4e+05	0.93	10	1294.86	1.8e+04
12	264	552	10	659.91	4.8e+04	1.3e+06	0.47	1	1682.57	2.1e+04
13	312	650	3	2925.07	1.2e+05	3.4e+06	0.51	0	—	—
14	364	756	0	—	—	—	—	0	—	—

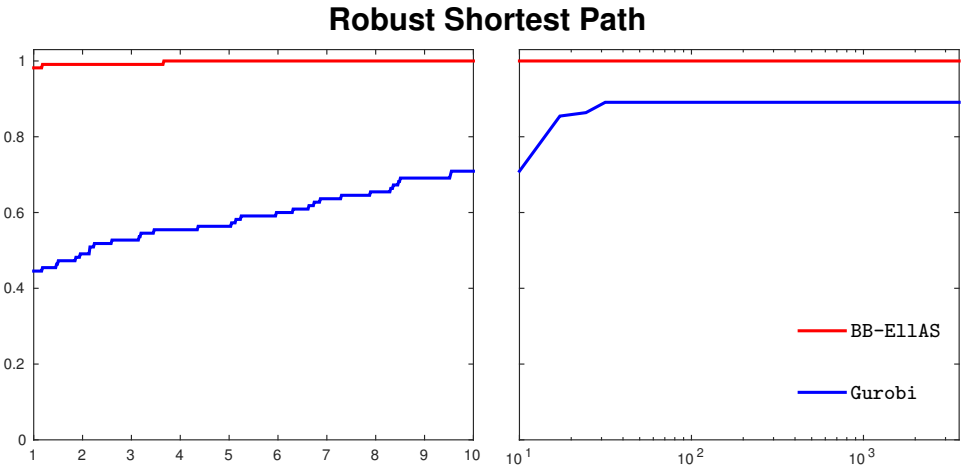

Figure 2: Performance profile with respect to running times for shortest path instances.

standard formulation of the Assignment problem:

$$\begin{array}{lll} \min & c^\top x + \sqrt{x^\top Q x} & \\ \text{s.t.} & \sum_{e\in\delta(i)} x_e = 1 & \forall i \in V \\ & x \in \{0,1\}^E & \end{array}$$

We consider complete bipartite graphs, so that the number of variables is $n = \frac{1}{4}|V|^2$. The number of constraints including $x \geq 0$ is $m = |V| + n$. Note that in the bipartite case the above formulation yields a complete description of $\text{conv}(P \cap \mathbb{Z}^n)$, which is not true in the case of general graphs. In our instances, we use expected weights 1 again, while the non-linear part of the objective function is generated as before. Altogether, we generated 80 different problem instances: for each $|V| \in \{10, 12, \ldots, 24\}$ we generated 10 different instances. Results are presented in Table 3 and Figure 3. The general picture is very similar to the one for the Shortest Path problem.

Table 3: Comparison on robust assignment instances.

			BB-EllAS					Gurobi		
$\|V\|$	n	m	#sol	time	nodes	iter	%ps	#sol	time	nodes
10	25	35	10	0.00	7.9e+01	6.2e+02	0.07	10	0.11	8.3e+01
12	36	48	10	0.01	2.6e+02	2.3e+03	0.07	10	0.48	2.8e+02
14	49	63	10	0.11	1.1e+03	9.9e+03	0.09	10	2.83	1.1e+03
16	64	80	10	1.34	4.5e+03	4.8e+04	0.73	10	29.98	4.4e+03
18	81	99	10	11.90	2.6e+04	3.0e+05	0.49	10	198.47	2.0e+04
20	100	120	10	112.16	1.2e+05	1.5e+06	0.85	10	1521.62	1.0e+05
22	121	143	9	669.68	6.6e+05	8.9e+06	0.36	0	—	—
24	144	168	0	—	—	—	—	0	—	—

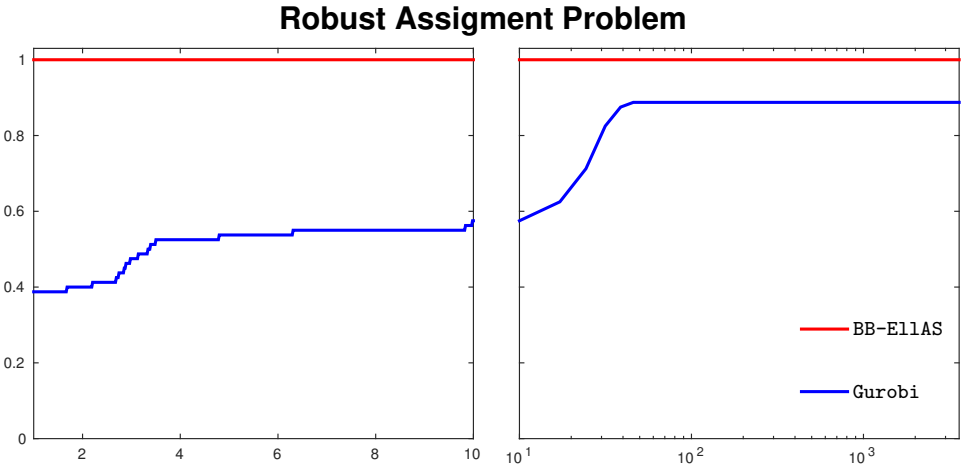

Figure 3: Performance profile with respect to running times for assignment instances.

7.4 Spanning Tree Problem

Given an undirected weighted graph $G = (V, E)$, a minimum spanning tree is a subset of edges that connects all vertices, without any cycles and with the minimum total edge weight. Our approach uses the following formulation of the Robust Spanning Tree problem:

$$\begin{array}{lrcll} \min & c^\top x + \sqrt{x^\top Q x} & & & \\ \text{s.t.} & \sum_{e \in E} x_e & = & |V| - 1 & \\ & \sum_{e \subseteq X} x_e & \leq & |X| - 1 & \forall \emptyset \neq X \subseteq V \\ & x & \in & \{0,1\}^E & \end{array} \tag{7}$$

In the above model, the number of constraints, taking into account also the non-negativity constraints, is $m = 2^{|V|} + n$. Since this number is exponential, we also have to use a separation algorithm for Gurobi. For both `BB-EllAS` and Gurobi, we essentially use the same simple implementation based on the Ford-Fulkerson algorithm.

For our experiments, we considered both complete graphs and grid graphs, the latter being produced as for the Shortest Path Problem. In both cases, expected edge weights are set to 1 again, while we

Table 4: Comparison on robust minimum spanning tree instances (complete graphs).

			BB-EllAS					Gurobi		
$\|V\|$	n	m	#sol	time	nodes	iter	%ps	#sol	time	nodes
10	45	1,069	10	2.93	1.4e+04	9.6e+04	1.35	10	78.59	1.6e+04
11	55	2,103	10	11.92	5.7e+04	4.3e+05	0.16	10	794.29	7.0e+04
12	66	4,162	10	120.84	4.4e+05	3.7e+06	0.06	1	2652.38	1.4e+05
13	78	8,270	10	1060.63	2.7e+06	2.4e+07	0.12	0	—	—
14	91	16,475	0	—	—	—	—	0	—	—

Table 5: Comparison on robust minimum spanning tree instances (grid graphs).

			BB-EllAS					Gurobi		
r	n	m	#sol	time	nodes	iter	%ps	#sol	time	nodes
5	40	3.4e+07	10	0.50	1.0e+03	8.7e+03	0.10	10	61.35	7.3e+03
6	60	6.9e+10	10	18.64	1.1e+04	1.2e+05	0.67	8	1805.72	9.2e+04
7	84	5.6e+14	9	1038.24	2.3e+05	3.3e+06	0.38	0	—	—
8	112	1.8e+19	0	—	—	—	—	0	—	—

built the positive definite matrix Q as above. Altogether, we generated 90 different problem instances: for each $|V| \in \{10, \ldots, 14\}$ we generated 10 different complete instances, while for each $r \in \{5, \ldots, 8\}$ we generated 10 different grid instances. As shown in Tables 4 and 5, `BB-EllAS` clearly outperforms the MISOCP solver of Gurobi on all the instances considered. For the performance profiles, see Figure 4.

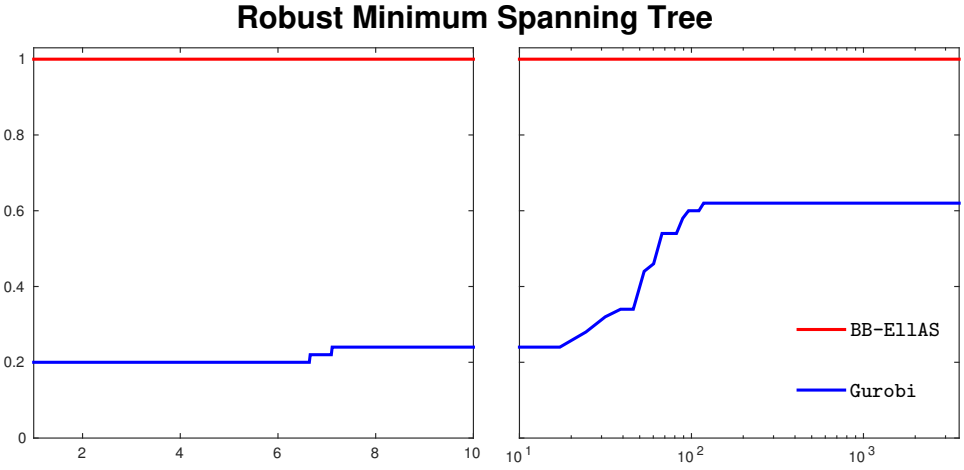

Figure 4: Performance profile with respect to running times for spanning tree instances.

7.5 Traveling Salesman Problem

Given an undirected, complete and weighted graph $G = (V, E)$, the Traveling Salesman problem consists in finding a path starting and ending at a given vertex $v \in V$ such that all the vertices in the graph are

Table 6: Comparison on robust traveling salesman instances.

			BB-EllAS					Gurobi		
$\|V\|$	n	m	#sol	time	nodes	iter	%ps	#sol	time	nodes
10	45	1,157	10	0.70	3.5e+03	2.5e+04	1.73	10	15.50	3.3e+03
11	55	2,211	10	2.37	1.6e+04	1.3e+05	0.18	10	69.96	1.2e+04
12	66	4,292	10	17.59	9.6e+04	8.2e+05	0.05	10	637.47	7.1e+04
13	78	8,424	10	150.00	5.4e+05	4.8e+06	0.14	3	2324.42	2.3e+05
14	91	16,655	10	1087.76	2.5e+06	2.4e+07	0.25	0	—	—
15	105	33,081	1	2966.10	6.2e+06	6.0e+07	0.06	0	—	—
16	120	65,894	0	—	—	—	—	0	—	—

visited exactly once and the sum of the weights of its constituent edges is minimized. Our approach uses the following formulation of the Traveling Salesman problem:

$$\begin{array}{lrcll} \min & c^\top x + \sqrt{x^\top Q x} & & & \\ \text{s.t.} & \sum_{e\in\delta(i)} x_e & = & 2 & \forall i \in V \\ & \sum_{e\in\delta(X)} x_e & \geq & 2 & \forall \emptyset \neq X \subsetneq V \\ & x & \in & \{0,1\}^E & \end{array} \tag{8}$$

Again, we consider complete graphs. The number of constraints including the bounds $x \in [0,1]^E$ is $m = 2^{|V|} + 3n - 2$ and hence again exponential. For both `BB-EllAS` and Gurobi, we basically use the same separation algorithm as for the Spanning Tree problem; see Section 7.4. Instances are identical to those generated for the Spanning Tree problem, but we can consider slightly larger graphs, namely graphs with $|V| \in \{10, \ldots, 16\}$. See Table 6 and Figure 5 for the results.

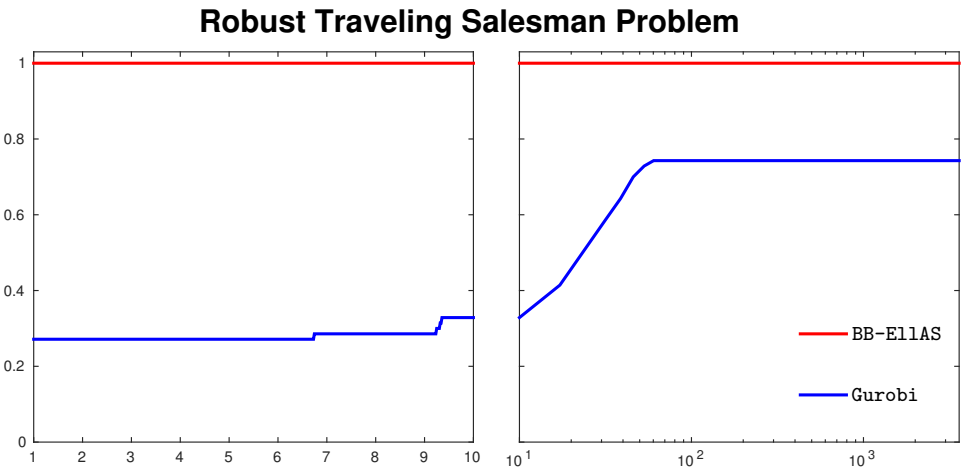

Figure 5: Performance profile with respect to running times for traveling salesman instances.

8 Conclusions

We presented a new branch-and-bound algorithm for robust combinatorial optimization problems under ellipsoidal uncertainty. We assume that the set of feasible solutions is given by a separation algorithm that decides whether a given point belongs to the convex hull of the feasible set or not, and, in the negative case, provides a valid but violated inequality. The branch-and-bound algorithm is based on the use of an active set method for the computation of dual bounds. Dealing with the Lagrangian dual of the continuous relaxation has the advantage of allowing an early pruning of the node. The closed form solution of the active set subproblems, the smart update of pseudo-inverse matrices, as well as the possibility of using warmstarts, leads to an algorithm that clearly outperforms the mixed-integer SOCP solver of Gurobi on the problem instances considered, including the robust counterpart of the shortest path and the traveling salesman problem.

9 Acknowledgments

The first author acknowledges support within the project "Mixed-Integer Non Linear Optimisation: Algorithms and Applications", which has received funding from the Europeans Union's EU Framework Programme for Research and Innovation Horizon 2020 under the Marie Skłodowska-Curie Actions Grant Agreement No 764759. The second author acknowledges support within the project "Nonlinear Approaches for the Solution of Hard Optimization Problems with Integer Variables" (No RP11715C7D8537BA) which has received funding from Sapienza, University of Rome.

References

[1] MOSEK ApS. *The MOSEK optimization toolbox for MATLAB manual. Version 8.0.0.81*, 2017.

[2] A. Ben-Tal and A. Nemirovski. Robust convex optimization. *Math. Oper. Res.*, 23(4):769–805, 1998.

[3] A. Ben-Tal and A. Nemirovski. Robust solutions of uncertain linear programs. *Oper. Res. Lett.*, 25:1–13, 1999.

[4] A. Ben-Tal and A. Nemirovski. *Lectures on Modern Convex Optimization*. SIAM, Philadelphia., 2001.

[5] C. Buchheim, M. De Santis, S. Lucidi, F. Rinaldi, and L. Trieu. A Feasible Active Set Method with Reoptimization for Convex Quadratic Mixed-Integer Programming. *SIAM Journal on Optimimization*, 26(3):1695–1714, 2016.

[6] C. Buchheim, M. De Santis, F. Rinaldi, and L. Trieu. A Frank-Wolfe Based Branch-and-Bound Algorithm for Mean-Risk Optimization. *Journal of Global Optimization*, 70(3):625–644, 2018.

[7] C. Buchheim and J. Kurtz. Robust combinatorial optimization under convex and discrete cost uncertainty. Technical report, Optimization Online, 2017.

[8] E. Dolan and J.Moré. Benchmarking optimization software with performance profiles. *Mathematical Programming*, 91:201–213, 2002.

[9] Gurobi Optimization, Inc. Gurobi optimizer reference manual, 2016.

[10] P. Kouvelis and G. Yu. *Robust Discrete Optimization and Its Applications*. Springer, 1996.

[11] C. D. Meyer. *Matrix Analysis and Applied Linear Algebra*. SIAM, Philadelphia, 2000.

[12] C. D. Meyer Jr. Generalized inversion of modified matrices. *SIAM Journal on Applied Mathematics*, 24(3):315–323, 1973.

[13] Hans D. Mittelmann. Mixed-integer SOCP Benchmark. http://plato.asu.edu/ftp/misocp.html (accessed March 1st, 2018; results from December 26th, 2017).

[14] Y. Nesterov and A. Nemirovski. *Interior-point polynomial algorithms in convex programming.* SIAM, Philadelphia, 1993.

[15] E. Nikolova. Approximation algorithms for offline risk-averse combinatorial optimization. Technical report, 2010.

[16] J. Nocedal and S. Wright. *Numerical Optimization (Second Edition).* Springer-Verlag, New York, 2006.

[17] J.F. Sturm. Using SeDuMi 1.02, a MATLAB toolbox for optimization over symmetric cones. *Optimization Methods and Software*, 11–12:625–653, 1999. Version 1.05 available from http://fewcal.kub.nl/sturm.