Random projections for conic programs

Leo Liberti, Pierre-Louis Poirion, Ky Vu

Random projections for conic programs

Leo Liberti[a], Pierre-Louis Poirion[b], Ky Vu[c]

[a] *LIX CNRS, Ecole Polytechnique, Institut Polytechnique de Paris, 91128 Palaiseau, France*
[b] *RIKEN Center for Advanced Intelligence Project, Tokyo, Japan*
[c] *Dept. of Mathematics, FPT University, Hanoi, Vietnam*

Abstract

We discuss the application of random projections to conic programming: notably linear, second-order and semidefinite programs. We prove general approximation results on feasibility and optimality using the framework of formally real Jordan algebras. We then discuss some computational experiments on randomly generated semidefinite programs in order to illustrate the practical applicability of our ideas.

Keywords: Mathematical Programming, Approximation, Johnson-Lindenstrauss Lemma, Jordan algebra.
2000 MSC: 90C22, 90C06

1. Introduction

In this paper we study Random Projection (RP) techniques for Symmetric Conic Programming (SCP): the class of optimization problems over symmetric cones. This class includes several convex optimization classes, such as Linear Programming (LP), Second-Order Cone Programming (SOCP), and Semidefinite Programming (SDP). It turns out that a symmetric cone can be represented as the cone of squares of a formally real Jordan algebra (FRJA) [2]. Using this framework it is possible to obtain theoretical results that apply to LP, SOCP and SDP alike [16]. We employ this framework in order to derive approximate feasibility and optimality results related to the application of RPs to SCPs. We then exhibit some computational experiments to show that our techniques can be used to approximately solve large scale SDPs in acceptable amounts of time.

RPs are random matrices that reduce the dimensions of the vectors in a given set while approximately preserving all pairwise Euclidean distances with high probability (whp), which means that the probability of failure decreases exponentially fast with increasing size of the reduced dimension. The normal application setting of RPs is to numerical data. This can speed up algorithms (at some cost in terms of the solution quality) which are essentially based on

Email addresses: liberti@lix.polytechnique.fr (Leo Liberti), pierre-louis.poirion@riken.jp (Pierre-Louis Poirion), vukhacky@gmail.com (Ky Vu)

Euclidean distances, such as, e.g., k-means or k-nearest neighbours [5, 11, 14]. The success of the application of RP to these algorithms is not surprising insofar as RPs give approximation guarantees for Euclidean distances by definition. This work continues the sequence of works [7, 20–22], which support the much more counter-intuitive statement that a Mathematical Programming (MP) formulation might be approximately invariant (as regards feasibiliy and optimality) w.r.t. randomly projecting the input parameters. In particular, in this work we generalize the results of [21] from the LP setting to all SCPs. Similarly to [21], our results assume that: (a) feasible instances are explicitly bounded; (b) strong SCP duality holds.

This work is motivated by the need to solve ever larger conic optimization problems. Its impact extends to LP, SOCP, SDP. Its importance is magnified by the fact that current SDP solution technology scales quite poorly in practice.

As far as we know, this work provides the first theoretical analysis of the application of RPs to the general class of SCPs. Notwithstanding, there exist previous works about the application of RPs to specific types of symmetric cones. The paper [21], for example, is about LPs. In [4], the authors reduce the dimension of SDPs using the RP matrix mapping $M \mapsto TMT^\top$, where T is an RP matrix. This approach is dual to ours, as we reduce the number of constraints rather than the dimension of the problem. We also note some literature results about fast low-rank approximations of PSD matrices using RPs [6, 17]. Based on these works, [8] exploits RPs to propose a bundle method for approximately solving SDPs without ever storing the whole constraint matrix and right hand side vector. While a small fraction of the theoretical results we present here bear some similarity to those in [6, 8, 17], the context and proofs are different. The main difference is that the results in [6, 8, 17] refer to the error yielded by some given algorithms deployed on randomly projected input. The results in this paper are about the formulation itself: we provide error guarantees on the optimality and feasibility of projected formulations w.r.t. the original ones independently of the solution method.

We note that even though this work generalizes the results of [21], the proof techniques are completely different, and justifiably so. A cornerstone of the results in [21] is that $T^\top Tx$ is approximately like x for a random projection matrix T (see below for a definition), whereas in the general conic case addressed in this paper, where x is a PSD matrix, the corresponding entity would not even be symmetric: we sidestep this difficulty by using a completely different framework, i.e. that of FRJA. Moreover, in [21] we extensively use the usual partial order on $\mathbb{R}^n$, which FRJA abstracts from by generalization to symmetric cones. Lastly, FRJA is essential in Section 4.2, where we estimate the infeasibility of the retrieved solution.

An RP is a random matrix sampled from a sub-Gaussian distribution [9]. Among other properties, applying an RP to a vector in a given set yields a vector with approximately the same norm. One famous example of the application of RPs is the Johnson-Lindenstrauss lemma (JLL) [12], which shows that a set of points can be embedded into a much lower dimension while keeping all pairwise distances approximately the same.

Definition 1.1 *A random matrix $T \in \mathbb{R}^{d\times m}$ is called a random projection if there are universal constants c and $\mathcal{C}$ such that, for any $x \in \mathbb{R}^m$ and $0 < \varepsilon < 1$ we have:*

$$\mathsf{Prob}\big[(1-\varepsilon)\|x\|_2^2 \leq \|Tx\|_2^2 \leq (1+\varepsilon)\|x\|_2^2\big] \geq 1 - ce^{-\mathcal{C}\varepsilon^2 d}.$$

An easy corollary of the JLL is the preservation of scalar products.

Proposition 1.2 *For $T \in \mathbb{R}^{d\times m}$ an RP with $\varepsilon \in (0,1)$ and $x, y \in \mathbb{R}^m$, we have:*

$$\mathsf{Prob}\big[\,\langle x,y\rangle - \varepsilon\|x\|_2\|y\|_2 \leq \langle Tx,Ty\rangle \leq \langle x,y\rangle + \varepsilon\|x\|_2\|y\|_2\,\big] \geq 1 - 2ce^{-\mathcal{C}\varepsilon^2 d}.$$

See [21, Proposition 1] for a proof. We also derive a variant of this result for later use.

Corollary 1.3 *For $T \in \mathbb{R}^{d\times m}$ an RP with $\varepsilon \in (0,1)$ and any $x \in \mathbb{R}^m$, we have:*

$$\mathsf{Prob}\big[\|(T^\top Tx)_i - x_i\|_\infty \leq \varepsilon\|x\|_2\big] \geq 1 - 2mce^{-\mathcal{C}\varepsilon^2 d}.$$

Proof. For every $i \leq m$, let $e_i = (0,\ldots,0,1_i,0,\ldots,0) \in \mathbb{R}^m$. From Proposition 1.2 applied to e_i and x, we have

$$e_i^\top x - \varepsilon\|x\|_2 \leq \langle Te_i, Tx\rangle = \langle e_i, T^\top Tx\rangle \leq e_i^\top x + \varepsilon\|x\|_2$$

for all $i \leq m$ whp, whence $-\varepsilon\|x\|_2 \leq \langle e_i, T^\top Tx\rangle - e_i^\top x \leq \varepsilon\|x\|_2$ and $-\varepsilon\|x\|_2 \leq (T^\top Tx - x)_i \leq \varepsilon\|x\|_2$. Thus, we have $|(T^\top Tx - x)_i| \leq \varepsilon\|x\|_2$ for any $i \leq m$. A union bound argument over all $i \leq m$ yields $\|T^\top Tx - x\|_\infty \leq \varepsilon\|x\|_2$ with the stated probability. □

In this paper, we shall prove that it is possible to find an approximately feasible solution of a given SCP (referred to as the *original* SCP), having an approximately optimal objective function value, by formulating another SCP (referred to as the *projected* SCP) subject to a random aggregation, obtained by means of an RP, of the original SCP constraints.

The rest of this paper is organized as follows. In Section 2 we introduce FRJAs, the corresponding spectral theorem, and an abstract version of primal and dual SCP pairs. In Section 3 we derive projected reformulations of primal and dual SCP using RPs, and prove approximation results for feasibility and optimality. In Section 4 we discuss how to retrieve a solution of the original SCP from the projected SCP. In Section 5 we present a computational validation of the foregoing theory.

2. SCPs and Jordan Algebras

We consider the following SCP formulation in standard form:

$$\left.\begin{array}{lrcl} \min & c^\top x & & \\ & Ax & = & b \\ & x & \succeq_{\mathcal{K}} & 0, \end{array}\right\} \quad (\mathsf{SCP}) \tag{1}$$

where $c, x \in \mathbb{R}^n$, $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and $x \succeq_{\mathcal{K}} y$ means $x - y \in \mathcal{K}$, where $\mathcal{K}$ is a closed convex pointed cone with non-empty interior [3]. In general, $\mathcal{K}$ might be a cartesian product of linear, second-order, positive semidefinite (psd), and possibly other self-dual cones [18].

We are going to consider a generalization of (SCP) in the context of FRJAs [2, 10]. Using this approach, for example, complementarity results and interior point algorithms can be described in terms of algebraic statements, including proofs of polynomial time convergence of interior point methods [16].

The formalization of conic programming in FRJA calls for an algebraic interpretation of symbols standing for vectors and matrices (as abstract elements and operators of an algebra). On the other hand, since an algebra *is* a vector space, the abstract elements can always be realized as vectors. Accordingly, we shall define a formal abstract inner product $\langle a, b\rangle$ over a FRJA, which may be realized as a quadratic form $a^\top Qb$ for some square matrix Q if we see a, b as elements of the underlying vector space. Similarly, we shall denote Ax by $A \odot x$ (after formally defining $\odot$ by means of the inner product), but then we shall argue that the $A\odot$ operation can be represented by a linear operator $\mathbf{A}$ on vectors, which can be realized by a matrix.

2.1. Formally real Jordan algebras

We recall some basic facts about FRJAs [2, 10]. An algebra over $\mathbb{R}$ is a pair $(\mathbb{E}, \circ)$ where $\mathbb{E}$ is a vector space over $\mathbb{R}$ and $\circ : \mathbb{E} \times \mathbb{E} \to \mathbb{E}$ is distributive in both arguments, i.e., for all $a, b, c \in \mathbb{E}$ and $\lambda, \gamma \in \mathbb{R}$, we have (i) $a \circ (\lambda b + \gamma c) = \lambda(a \circ b) + \gamma(a \circ c)$ and (ii) $(\lambda b + \gamma c) \circ a = \lambda(b \circ a) + \gamma(c \circ a)$. For each $x \in \mathbb{E}$, we let $L(x)$ be the *left multiplication operator*: $L(x)(y) = x \circ y$ for all $y \in \mathbb{E}$.

An algebra $(\mathbb{E}, \circ)$ is called a *Jordan algebra* if it satisfies:

(a) $\forall a, b \in \mathbb{E} \quad a \circ b = b \circ a$ (commutativity);

(b) $\forall a, b \in \mathbb{E} \quad a^2 \circ (a \circ b) = a \circ (a^2 \circ b)$ (Jordan identity).

The Jordan identity implies that a Jordan algebra is power-associative, i.e. for all $p, q \in \mathbb{N}$ the matrices $L(x^p)$ and $L(x^q)$ commute. A Jordan algebra $(\mathbb{E}, \circ)$ is called *formally real* if for any integer $m \geq 1$ and any finite set $\{x_1, \ldots, x_m\} \subset \mathbb{E}$ we have:

$$\sum_{i\leq m} x_i^2 = 0 \quad \Rightarrow \quad x_1 = \ldots = x_m = 0. \tag{2}$$

Informally speaking, (2) states that square terms in FRJAs must be non-negative. This is a nontrivial statement insofar as elements of $\mathbb{E}$ are not totally ordered in general (so one cannot write $x^2 \geq 0$).

2.2. The spectral decomposition theorem

For each element $x \in \mathbb{E}$ the *degree* $\deg(x)$ of x is the largest integer d such that $e = x^0, x, \ldots, x^{d-1}$ (where e is the multiplicative unit of $\mathbb{E}$) are linearly independent. The degree of $\mathbb{E}$, denoted by $\deg(\mathbb{E})$, is the maximum degree of all elements over $\mathbb{E}$. This allows us to state the spectral decomposition theorem for FRJAs [10, Theorem III.1.1].

Theorem 2.1 ([10], Theorem III.1.1) *Let $\mathbb{E}$ be a FRJA of degree r. For every $x \in \mathbb{E}$ there are sets*

$$\Lambda(x) = \{\lambda_i \mid i \leq r\} \subset \mathbb{R} \quad \textit{and} \quad C(x) = \{c_i \mid 1 \leq i \leq r\} \subset \mathbb{E},$$

with $\Lambda(x)$ ordered so that $\lambda_i \leq \lambda_{i+1}$ for each $i < r$, and $C(x)$ satisfying the following properties:

(i) $\forall i \leq r \ (c_i^2 = c_i)$
(ii) $c_1 + \ldots + c_r = e$
(iii) $\forall i \neq j \ (c_i \circ c_j = 0)$

such that $x = \lambda_1 c_1 + \ldots + \lambda_r c_r$.

The elements of $\Lambda(x)$ are the *eigenvalues* of x. Theorem 2.1 is a generalization of the spectral decomposition theorem for real symmetric matrices.

Let $\mathbb{E}$ be a FRJA of degree r, and $x \in \mathbb{E}$. We define $\mathsf{tr}(L(x))$ as the trace of the matrix representation of the left multiplication operator $L(x)$. We then define the associative bilinear form

$$T(x, y) = \mathsf{tr}(L(x \circ y))$$

(for $y \in \mathbb{E}$). Since $(\mathbb{E}, \circ)$ is formally real, $T(x, y)$ defines a positive definite inner product denoted by $\langle x, y \rangle$. The element x of the algebra is *positive semidefinite* (psd) if there is $y \in \mathbb{E}$ s.t. $x = y^2$. The set $\mathcal{K}_{\mathbb{E}}$ of all psd elements of $\mathbb{E}$ is called *the cone of squares* of $\mathbb{E}$. We denote by $x \succeq 0$ the fact that x is psd in $\mathbb{E}$. By Theorem 2.1, $x \succeq 0$ is psd iff all of its eigenvalues are nonnegative; its smallest eigenvalue is

$$\lambda_{\min}(x) = \min_{u \neq 0} \frac{\langle u, x \circ u \rangle}{\|u\|_{\mathbb{E}}^2},$$

where $\| \cdot \|_{\mathbb{E}}$ is the norm induced by $\langle \cdot, \cdot \rangle$.

2.3. Primal and dual SCPs in FRJA notation

According to assumption (a) in Section 1, we introduce an explicit bound constraint $\langle e, x \rangle \leq \theta$ in the primal SCP of the following primal/dual pair.

$$\left.\begin{array}{lrcl} \min\limits_{x \in X} & \langle c, x \rangle & & \\ & A \odot x & = & b \\ & \langle e, x \rangle & \leq & \theta \\ & x & \succeq & 0 \end{array}\right\} \quad (P) \qquad \left.\begin{array}{lrcl} \max\limits_{\substack{y \in \mathbb{R}^m \\ \nu \geq 0}} & b^\top y - \theta\nu & & \\ & \sum\limits_{i \leq m} y_i A_i - \nu e & \preceq & c, \end{array}\right\} \quad (D)$$

where $A \odot x = (\langle A_i, x \rangle \mid i \leq m), b \in \mathbb{R}^m$.

In the rest of the paper we assume that either (P) is infeasible, or both (P) and (D) have optimal solutions x^* and y^* respectively. According to assumption (b) in Section 1, we also assume that strong duality holds.

3. Random projection of SCP constraints

Let $T = (T_{kj}) \in \mathbb{R}^{d\times m}$ be an RP. We consider the projected primal/dual pair:

$$\left.\begin{array}{rlcl} \min\limits_{x\in X} & \langle c, x\rangle & & \\ \forall k \le d & \langle \bar{A}_k, x\rangle & = & (Tb)_k \\ & \langle e, x\rangle & \le & \theta \\ & x & \succeq & 0 \end{array}\right\} (P_T) \qquad \left.\begin{array}{rlcl} \max\limits_{\substack{z\in\mathbb{R}^d \\ \nu\ge 0}} & (Tb)^\top z - \theta\nu & & \\ & \sum\limits_{k\le d} z_k\bar{A}_k - \nu e & \preceq & c, \end{array}\right\} (D_T)$$

where $\bar{A}_k = \sum_{i\le m} T_{ki}A_i$ for every $k \le d$. Note that (P_T) is a relaxation of (P), since its constraints are obtained by aggregating the constraints of (P). Hence $\mathsf{v}(P_T) \le \mathsf{v}(P)$. Moreover, for any feasible solution (z', ν') of (D_T), we have:

$$\sum_{k\le d} z'_k\bar{A}_k - \nu' e = \sum_{k\le d} z'_k \sum_{i\le m} T_{ki}A_i - \nu' e = \sum_{i\le m} (T^\top z')_i A_i - \nu' e \preceq c,$$

whence $(T^\top z', \nu')$ is a feasible solution of the dual (D) of the original problem.

3.1. Approximate feasibility

Since T acts linearly on left and right hand sides of the equality constraints of (P), if (P) is feasible then so is (P_T). The issue is that (P_T) may be feasible even if (P) is not. We therefore assume that (P) is infeasible, and prove that (P_T) is infeasible whp.

We first claim that (D) has non-empty interior: we can choose y_i small enough and ν large enough so that $\sum_i y_i A_i - v\nu \preceq c$ holds strictly. Since (P) is infeasible and (D) has non empty interior, we deduce by duality that (D) is unbounded. Hence there exists a dual feasible solution $(\hat{y}, \hat{\nu})$ such that

$$b^\top \hat{y} - \theta\hat{\nu} = 1 \qquad \sum_{i\le m} \hat{y}_i A_i - e\hat{\nu} \preceq 0. \tag{3}$$

We shall prove that (D_T) is also unbounded, by constructing (z', ν') such that $(Tb)^\top z' - \theta\nu' > 0$ and $\sum_{i\le m}(T^\top z')_i A_i - e\nu' \preceq 0$. Let $z' = T\hat{y}$. By Proposition 1.2, whp we have that

$$(Tb)^\top z' = (Tb)^\top T\hat{y} \ge b^\top \hat{y} - \varepsilon\|b\|_2\|\hat{y}\|_2. \tag{4}$$

Let $N = \sum_{i\le m} \hat{y}_i A_i - e\hat{\nu}$ and $N' = \sum_{i\le m}(T^\top z')_i A_i - e\hat{\nu}$. While $N \preceq 0$ by feasibility of (D), we derive an upper bound to $\lambda_{\max}(N')$ in terms of

$$\|A^\top\|_{\substack{\mathbb{E}\\ \infty}} = \sup_{\|y\|_\infty = 1} \|A^\top y\|_{\mathbb{E}}.$$

Lemma 3.1 *Let $T \in \mathbb{R}^{d\times m}$ be an RP with $\varepsilon \in (0,1)$. For $(\hat{y}, \hat{\nu})$ feasible (D) and $z' = T\hat{y}$, we have:*

$$\mathsf{Prob}\left[\lambda_{\max}(N') \le \varepsilon\|\hat{y}\|_2\|A^\top\|_{\substack{\mathbb{E}\\ \infty}}\right] \ge 1 - 2mce^{-\mathcal{C}\varepsilon^2 d}.$$

Proof. Let f, f' be the quadratic forms associated to N, N', namely $f(u) = \frac{\langle u, N \circ u\rangle}{\|u\|_{\mathbb{E}}^2}$ and $f'(u) = \frac{\langle u, N' \circ u\rangle}{\|u\|_{\mathbb{E}}^2}$. In particular, replacing z' with $T\hat{y}$, we have:

$$\|u\|_{\mathbb{E}}^2 f'(u) = \big\langle u, \big(\sum_{i\leq m}(T^\top z')_i A_i - e\hat{\nu}\big)\circ u\big\rangle = \sum_{i\leq m}(T^\top T\hat{y})_i\langle u, A_i\circ u\rangle - \|u\|_{\mathbb{E}}^2\hat{\nu}.$$

Since $f(u) = \frac{\sum_i \hat{y}_i\langle u, A_i\circ u\rangle}{\|u\|_{\mathbb{E}}^2} - \hat{\nu}$, we can write $f'(u) = f(u) + \frac{\sum_i \hat{z}_i\langle u, A_i\circ u\rangle}{\|u\|_{\mathbb{E}}^2}$, where $\hat{z} = T^\top T\hat{y} - \hat{y} \in \mathbb{R}^m$. This yields

$$f'(u) = f(u) + \frac{\langle u, \sum_i \hat{z}_i A_i \circ u\rangle}{\|u\|_{\mathbb{E}}^2} = f(u) + \frac{\langle u, (A^\top\hat{z})\circ u\rangle}{\|u\|_{\mathbb{E}}^2} \leq f(u) + \|A^\top\|_{\frac{\mathbb{E}}{\infty}}\|\hat{z}\|_\infty,$$

where the inequality on the right holds by definition of $\|A^\top\|_{\frac{\mathbb{E}}{\infty}}$. Note that $\lambda_{\max}(N') = \max_{u\neq 0} f'(u)$ and $\lambda_{\max}(N) = \max_{u\neq 0} f(u) \leq 0$ as $N \preceq 0$. Because Corollary 1.3 implies $\|\hat{z}\|_\infty \leq \varepsilon\|\hat{y}\|_2$ whp, we deduce that

$$\lambda_{\max}(N') \leq \lambda_{\max}(N) + \|A^\top\|_{\frac{\mathbb{E}}{\infty}}\|\hat{z}\|_\infty \leq \varepsilon\|\hat{y}\|_2\|A^\top\|_{\frac{\mathbb{E}}{\infty}}$$

with probability at least $1 - 2mce^{-\mathcal{C}\varepsilon^2 d}$, as claimed. □

Theorem 3.2 *Suppose (P) infeasible and $\varepsilon\|\hat{y}\|_2(\|b\|_2 + \|A^\top\|_{\frac{\mathbb{E}}{\infty}}) < 1$. With the same notation as above, we have* $\mathsf{Prob}[(P_T)$ *is infeasible*$] \geq 1 - 2(m+1)ce^{-\mathcal{C}\varepsilon^2 d}$.

Proof. By Lemma 3.1, with probability at least $1 - 2mce^{-\mathcal{C}\varepsilon^2 d}$, we have $N' - \varepsilon\|\hat{y}\|_2\|A^\top\|_{\frac{\mathbb{E}}{\infty}} e \preceq 0$, whence $\sum_{i\leq m}(T^\top z')_i A_i - e(\hat{\nu} + \varepsilon\|\hat{y}\|_2\|A^\top\|_{\frac{\mathbb{E}}{\infty}}) \preceq 0$. Moreover, by (3) and (4), we obtain

$$\mathsf{Prob}\big[(Tb)^\top z' - \theta\hat{\nu} \geq 1 - \varepsilon\|b\|_2\|\hat{y}\|_2\big] \geq 2ce^{-\mathcal{C}\varepsilon^2 d}.$$

Let $\nu' = \hat{\nu} + \varepsilon\|\hat{y}\|_2\|A^\top\|_{\frac{\mathbb{E}}{\infty}}$. Then $\sum_{i\leq m}(T^\top z')_i A_i - e\nu' \preceq 0$. Furthermore,

$$(Tb)^\top z' - \theta\nu' = (Tb)^\top z' - \theta(\hat{\nu} + \varepsilon\|\hat{y}\|_2\|A^\top\|_{\frac{\mathbb{E}}{\infty}}) \geq 1 - \varepsilon\|\hat{y}\|_2(\|b\|_2 + \|A^\top\|_{\frac{\mathbb{E}}{\infty}}).$$

Hence, if $\varepsilon\|\hat{y}\|_2(\|b\|_2 + \|A^\top\|_{\frac{\mathbb{E}}{\infty}}) < 1$, (4) is satisfied, so (D_T) is unbounded, implying by duality that (P_T) is infeasible with probability given by union bound arguments. □

3.2. Approximate optimality

We show that any feasible solution $(\hat{y}, \hat{\nu})$ of (D) is close to a point of the form $(T^\top\hat{z}, \hat{\nu})$ whp, where $(\hat{z}, \hat{\nu})$ is a feasible solution of the following family of relaxations of (D_T) parametrized over a real number $\mu > 0$:

$$\left.\begin{array}{lll} \max\limits_{\substack{z\in\mathbb{R}^d \\ \nu\geq 0}} & (Tb)^\top z - \theta\nu & \\ \sum\limits_{i\leq m} & (T^\top z)_i A_i - \nu e & \preceq \quad c + \mu e. \end{array}\right\} \quad (\tilde{D}_T^\mu)$$

We use (D_T^μ) to show that the optimal objective function value of the original problem is approximately invariant w.r.t. RPs.

Let $z' = T\hat{y} \in \mathbb{R}^d$. We claim that there exists $\mu > 0$ s.t. $(z', \hat{\nu})$ is a feasible solution of $(\tilde{D}_T^\mu)$ whp. This will follow from $c + \mu e - \sum_{i\leq m}(T^\top z')_i A_i + e\hat{\nu} \succeq 0$ whp; or, equivalently, from $\lambda_{\min}(M') \geq -\mu$ for some suitable $\mu > 0$ whp, where $M' = c - \sum_{i\leq m}(T^\top z')_i A_i + e\hat{\nu}$. Let $M = c - \sum_{i\leq m}\hat{y}_i A_i + e\hat{\nu}$. Since $(\hat{y}, \hat{\nu})$ is a feasible dual solution of (D) we have $M \succeq 0$ and hence $\lambda_{\min}(M) \geq 0$. On the other hand M' might fail to be psd, as $\lambda_{\min}(M')$ might be negative. The following result provides a lower bound to $\lambda_{\min}(M')$ in terms of $\|A^\top\|_{\frac{\mathbb{E}}{\infty}}$.

Lemma 3.3 *Let $T \in \mathbb{R}^{d\times m}$ be an RP with $\varepsilon \in (0,1)$. For $(\hat{y}, \hat{\nu})$ feasible in (D) and $z' = T\hat{y}$, we have*

$$\mathsf{Prob}\big[\lambda_{\min}(M') \geq -\varepsilon\|\hat{y}\|_2\|A^\top\|_{\frac{\mathbb{E}}{\infty}}\big] \geq 1 - 2mce^{-\mathcal{C}\varepsilon^2 d}.$$

Proof. Let f, f' be the quadratic forms associated to M, M', namely $f(u) = \frac{\langle u, M\circ u\rangle}{\|u\|_{\mathbb{E}}^2}$ and $f'(u) = \frac{\langle u, M'\circ u\rangle}{\|u\|_{\mathbb{E}}^2}$. By bilinearity, we have

$$\begin{aligned}
\|u\|_{\mathbb{E}}^2 f'(u) &= \Big\langle u, \big(c - \sum_{i\leq m}(T^\top z')_i A_i + e\hat{\nu}\big)\circ u\Big\rangle \\
&= \Big\langle u, \big(c - \sum_{i\leq m}(T^\top T\hat{y})_i A_i + e\hat{\nu}\big)\circ u\Big\rangle \\
&= \langle u, (c + e\hat{\nu})\circ u\rangle - \sum_{i\leq m}(T^\top T\hat{y})_i\langle u, A_i\circ u\rangle.
\end{aligned}$$

Since $f(u) = \frac{\langle u,(c+e\hat{\nu})\circ u\rangle}{\|u\|_{\mathbb{E}}^2} - \sum_{i\leq m}\hat{y}_i\frac{\langle u, A_i\circ u\rangle}{\|u\|_{\mathbb{E}}^2}$, we have

$$f'(u) = f(u) - \sum_{i\leq m}\big((T^\top T\hat{y})_i - \hat{y}_i\big)\frac{\langle u, A_i\circ u\rangle}{\|u\|_{\mathbb{E}}^2} = f(u) - \sum_{i\leq m}\hat{z}_i\frac{\langle u, A_i\circ u\rangle}{\|u\|_{\mathbb{E}}^2},$$

where $\hat{z} = T^\top T\hat{y} - \hat{y} \in \mathbb{R}^m$. This yields

$$\begin{aligned}
f'(u) &= f(u) - \frac{\langle u, \sum_{i\leq m}\hat{z}_i A_i\circ u\rangle}{\|u\|_{\mathbb{E}}^2} \\
&= f(u) - \frac{\langle u, (A^\top\hat{z})\circ u\rangle}{\|u\|_{\mathbb{E}}^2} \geq f(u) - \|A^\top\|_{\frac{\mathbb{E}}{\infty}}\|\hat{z}\|_\infty
\end{aligned}$$

where the inequality on the right holds by definition of $\|A^\top\|_{\frac{\mathbb{E}}{\infty}}$ (note also that the probability does not depend on u). Now we have: (i) $\lambda_{\min}(M') = \min_{u\neq 0} f'(u)$; (ii) $\lambda_{\min}(M) = \min_{u\neq 0} f(u) \geq 0$ since $M \succeq 0$; (iii) Corollary 1.3 implies $\|\hat{z}\|_\infty \leq \varepsilon\|\hat{y}\|_2$ whp. From (i)-(iii) we deduce that

$$\lambda_{\min}(M') \geq \lambda_{\min}(M) - \|A^\top\|_{\frac{\mathbb{E}}{\infty}}\|\hat{z}\|_\infty \geq -\varepsilon\|\hat{y}\|_2\|A^\top\|_{\frac{\mathbb{E}}{\infty}}.$$

with probability at least $1 - 2mce^{-\mathcal{C}\varepsilon^2 d}$, as claimed. □

We remark in passing that we have a trivial upper bound $\|A^\top\|_{\mathbb{E}}^{\mathbb{E}} \leq \sum_{i\leq m} \rho(A_i)$, where, for all $i \leq m$, $\rho(A_i) = \max_{u\neq 0} |\langle u, A_i \circ u\rangle|/\|u\|_{\mathbb{E}}^2$ is the eigenvalue of A_i having the largest absolute value. We also note that this bound is not going to be used in the rest of the paper.

By Lemma 3.3, we can let $\mu = \varepsilon\|\hat{y}\|_2 \|A^\top\|_{\mathbb{E}}^{\mathbb{E}}$ in (D_T^μ), and obtain the relaxation:

$$\left.\begin{array}{lll} \max\limits_{\substack{z\in\mathbb{R}^d \\ \nu\geq 0}} & (Tb)^\top z - \theta\nu & \\ \sum\limits_{k\leq d} & z_k\big(\sum\limits_{i\leq m} T_{ki}A_i\big) - \nu e \preceq & c + \varepsilon\|\hat{y}\|_2\|A^\top\|_{\mathbb{E}}^{\mathbb{E}} e \end{array}\right\} \quad (\tilde{D}_T)$$

of (D_T). By construction, $(z', \hat{\nu})$ is a feasible solution of $(\tilde{D}_T)$ whp, as claimed.

Lemma 3.4 *Let (y^*, ν^*) be an optimum of (D). With the same notation as above, we have*

$$\mathsf{Prob}\big[\mathsf{v}(\tilde{D}_T) \geq (Tb)^\top Ty^* - \theta\nu^* \geq \mathsf{v}(D) - \varepsilon\|b\|_2\|y^*\|_2\big] \geq 1 - 2(m+1)ce^{-\mathcal{C}\varepsilon^2 d}.$$

Proof. By Proposition 1.2, we have $(Tb)^\top Ty^* \geq b^\top y^* - \varepsilon\|b\|_2\|y^*\|_2$ whp; since $\mathsf{v}(D) = b^\top y^* - \theta\nu^*$, we obtain $(Tb)^\top Ty^* - \theta\nu^* \geq \mathsf{v}(D) - \varepsilon\|b\|_2\|y^*\|_2$. On the other hand, (y^*, ν^*) is optimal in (D) so it is also feasible in (D), hence by Lemma 3.3 (Ty^*, ν^*) is feasible in $(\tilde{D}_T)$ whp. Thus $\mathsf{v}(\tilde{D}_T) \geq (Tb)^\top Ty^* - \theta\nu^*$, as $(\tilde{D}_T)$ is a maximization problem; the stated probability is given by union bound arguments. □

We now consider the dual of $(\tilde{D}_T)$:

$$\left.\begin{array}{rrcl} \min\limits_x & \big\langle c + \varepsilon\|\hat{y}\|_2\|A^\top\|_{\mathbb{E}}^{\mathbb{E}} e & , & x\big\rangle \\ \forall k \leq d & \langle \bar{A}_k, x\rangle & = & (Tb)_k \\ & \langle e, x\rangle & \leq & \theta \\ & x & \succeq & 0. \end{array}\right\} \quad (\tilde{P}_T)$$

Let x' be an optimal solution of (P_T). By the weak duality theorem of conic programming, with probability at least $1 - 2mce^{-\mathcal{C}\varepsilon^2 d}$ we have:

$$\mathsf{v}(\tilde{D}_T) \leq \mathsf{v}(\tilde{P}_T) \leq \mathsf{v}(P_T) + \varepsilon\|y^*\|_2\|A^\top\|_{\mathbb{E}}^{\mathbb{E}}\langle e, x_T\rangle \leq \mathsf{v}(P_T) + \varepsilon\|y^*\|_2\|A^\top\|_{\mathbb{E}}^{\mathbb{E}}\theta, \quad (5)$$

where (y^*, ν^*) is an optimal solution of (D). Hence, by combining Eq. (5) with Lemma 3.4 and the fact that the strong duality theorem holds for the original problem (P), we have proved the following result.

Theorem 3.5 *Assume strong duality between (P) and (D) holds. With the same notation as above,* $\mathsf{Prob}\big[\mathsf{v}(P_T) \geq \mathsf{v}(P) - \varepsilon\|y^*\|_2(\|A^\top\|_{\mathbb{E}}^{\mathbb{E}}\theta + \|b\|_2)\big] \geq 1 - (2m+1)ce^{-\mathcal{C}\varepsilon^2 d}$.

In the rest of the paper, d can be considered fixed to $\mathcal{C}_0 \log(m)/\varepsilon^2$, where $\mathcal{C}_0$ is chosen so that the probability in Theorem 3.5 is as high as desired.

4. Solution retrieval

Solution retrieval, i.e. obtaining a good solution $\tilde{x}$ for (P) from a solution x_T of (P_T), is an important part of most RP-based techniques. The error analysis of $\tilde{x}$ w.r.t. (P) turns out to depend on the feasibility error of x_T in (P).

4.1. Feasibility error of the projected solution

The *Gaussian width* of a bounded set $S \subset \mathbb{R}^n$ is $w(S) = \mathbb{E}_g(\sup_{x\in S} g \cdot x)$, where g is a standard normal vector in $\mathbb{R}^n$ and $\cdot$ is the usual scalar product on $\mathbb{R}^n$. Here are some of its basic properties [19]:

- for any $y \in \mathbb{R}^n$, $w(S + y) = w(S)$;
- $w(\mathsf{conv}(S)) = w(S)$
- for any $m \times n$ matrix A, $w(AS) \leq \|A\|_2 w(S)$, where $\|A\|_2$ is the matrix norm induced by the Euclidean norm;
- if S is finite, $w(S) \leq \tilde{C}\sqrt{\log(|S|)}\mathsf{diam}(S)$ for some universal constant $\tilde{C}$.

Next, we recall the M^* bound theorem, which bounds the diameter of the intersection of a subset $S \subset \mathbb{R}^n$ with $\ker(T)$, where T is an RP [19]. We let $\mathsf{rad}(S) = \max_{x\in S} \|x\|_2$ and $\mathsf{diam}(S) = \max_{x,y\in S} \|x - y\|_2$.

Theorem 4.1 (M^* bound theorem, [19]) *For $S \subset \mathbb{R}^n$ and $u \geq 0$,*

$$\mathsf{Prob}\big[\mathsf{diam}(S \cap \ker(T)) \leq \frac{C_2 w(S) + u\,\mathsf{rad}(S)}{\sqrt{d}}\big] \geq 1 - 2e^{-u^2},$$

where C_2 is a universal constant.

Let $\mathbf{A}$ be the matrix representation of the linear mapping $x \mapsto A \odot x$ from $\mathbb{R}^n$ to $\mathbb{R}^m$, i.e. the $m \times n$ matrix such that for all $x \in \mathbb{R}^n$, $A \odot x = \mathbf{A}x$. Let $\mathbb{B} = \{x \in \mathcal{K}_{\mathbb{E}} \mid \langle e, x\rangle \leq 1\}$. Since $\mathcal{K}_{\mathbb{E}}$ is a pointed cone [10, Chapter I], $\mathbb{B}$ is a bounded region of $\mathbb{R}^n$ of given diameter Δ (as $\langle e, x\rangle > 0$ over $\mathcal{K}_{\mathbb{E}} \setminus \{0\}$), hence its Gaussian width $w(\mathbb{B})$ is finite.

Proposition 4.2 *Let x_T be an optimal solution of (P_T) and let $u > 0$. Then we have:*

$$\mathsf{Prob}\big[\|\mathbf{A}x_T - b\|_2 \leq \varepsilon\theta\|\mathbf{A}\|_2(C_2 w(\mathbb{B}) + u\Delta)/\sqrt{\log(n)}\big] \geq 1 - 2e^{-u^2}.$$

Proof. Consider $S = \{\mathbf{A}x - b \mid x \in \theta\mathbb{B}\} \subset \mathbb{R}^m$. By definition, $\mathbf{A}x_T - b \in S$; moreover, $\mathbf{A}x_T - b \in \ker(T)$ since x_T is a solution of (P_T), hence

$$\mathbf{A}x_T - b \in S \cap \ker(T).$$

Moreover,

$$w(S) = w(\theta\mathbf{A}\mathbb{B} - b) \leq \theta\|\mathbf{A}\|_2 w(\mathbb{B}).$$

Furthermore, using the fact that there exists $x \in \mathbb{R}^n$ such that $\mathbf{A}x = b$, we have

$$\mathsf{rad}(S) \leq \|\mathbf{A}\|_2 \mathsf{diam}(\theta\mathbb{B}) \leq \theta\|\mathbf{A}\|_2\Delta.$$

The result follows by Theorem 4.1. □

4.2. Error analysis of the retrieved solution

We define the retrieved solution $\tilde{x}$ of (P) as the orthogonal projection of x_T on the feasible set of (P), i.e. $\tilde{x} \in \arg\min_x\{\|x - x_T\|_2^2 \mid \mathbf{A}x = b\}$. By the KKT conditions, $\tilde{x}$ can be computed using the pseudoinverse of $\mathbf{A}$:

$$\tilde{x} = x_T + \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}(b - \mathbf{A}x_T). \tag{6}$$

Since $\tilde{x}$ satisfies $\mathbf{A}x = b$ by construction, we only bound the error of $\tilde{x}$ w.r.t. membership in the cone $\mathcal{K}_\mathbb{E}$ in terms of the negativity of $\lambda_{\min}(\tilde{x})$.

Lemma 4.3 *For $x, a, b \in \mathbb{E}$ such that $x = a + b$, $\lambda_{\min}(x) \geq \lambda_{\min}(a) + \lambda_{\min}(b)$.*

Proof. For each $y \in \mathbb{E}$ we have $\lambda_{\min}(y) = \min\limits_{u\neq 0} \frac{\langle u, y\circ u\rangle}{\|u\|_\mathbb{E}^2}$, so

$$\begin{aligned}\lambda_{\min}(x) &= \min_{u\neq 0} \frac{\langle u, (a+b)\circ u\rangle}{\|u\|_\mathbb{E}^2} = \min_{u\neq 0}\left(\frac{\langle u, a\circ u\rangle}{\|u\|_\mathbb{E}^2} + \frac{\langle u, b\circ u\rangle}{\|u\|_\mathbb{E}^2}\right)\\ &\geq \min_{u\neq 0} \frac{\langle u, a\circ u\rangle}{\|u\|_\mathbb{E}^2} + \min_{u\neq 0} + \frac{\langle u, b\circ u\rangle}{\|u\|_\mathbb{E}^2} = \lambda_{\min}(a) + \lambda_{\min}(b).\end{aligned}$$

□

Theorem 4.4 *Suppose the inner product in $\mathbb{E}$ is realized by a bilinear form based on the symmetric matrix Q. With the notation above, we have:*

$$\mathsf{Prob}\left[\lambda_{\min}(\tilde{x}) \geq \lambda_1 - \varepsilon\theta\kappa(\mathbf{A})\|Q^{\frac{1}{2}}\|_2 \frac{C_2 w(\mathbb{B}) + u\Delta}{\sqrt{\log(n)}}\right] \geq 1 - 2e^{-u^2}, \tag{7}$$

where $\kappa(\mathbf{A})$ is the condition number of $\mathbf{A}$.

Proof. Let $\sum\limits_{j\leq r} \lambda_j c_j$ be the spectral decomposition of x_T. Since x_T solves (P_T), $x_T \in \mathcal{K}_\mathbb{E}$. So its eigenvalues are $0 \leq \lambda_1 \leq \ldots \leq \lambda_r$, where r is the degree of $\mathbb{E}$. Suppose that the spectral decomposition of

$$\tilde{x} - x_T = \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}(b - \mathbf{A}x_T)$$

is $\sum\limits_{j\leq r} \mu_j c'_j$, with $\mu_1 \leq \ldots \leq \mu_r$. Then

$$\lambda_{\min}(\tilde{x}) \geq \lambda_1 + \mu_1 \tag{8}$$

by Lemma 4.3. We have $\|\tilde{x} - x_T\|_{\mathbb{E}} = \sqrt{\sum_{j \le r} \mu_j^2}$ by definition. Therefore

$$|\mu_1| = \sqrt{\mu_1^2} \le \sqrt{\sum_j \mu_j^2} = \|\tilde{x} - x_T\|_{\mathbb{E}}.$$

We are going to reason about $\|\tilde{x} - x_T\|_2$ instead, and then use the fact that $\langle a, b\rangle = a^\top Q y$, which will allow us to derive a bound on $\|\tilde{x} - x_T\|_{\mathbb{E}}$ from a bound on $\|\tilde{x} - x_T\|_2$. By (6), we have

$$\|\tilde{x} - x_T\|_2 \le \|\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\|_2 \|\mathbf{A}x_T - b\|_2.$$

If $U\Sigma V^\top$ is the Singular Value Decomposition (SVD) of $\mathbf{A}$, a straightforward computation yields $V\Sigma^\top(\Sigma\Sigma^\top)^{-1}U^\top$ as the SVD of $\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}$, which implies that $\|\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\|_2 = \frac{1}{\sigma_{\min}}$ where $\sigma_{\min}$ is the smallest singular value of $\mathbf{A}$. Moreover, by Proposition 4.2, we have

$$\|\mathbf{A}x_T - b\|_2 \le \varepsilon\theta\|\mathbf{A}\|_2(C_2 w(\mathbb{B}) + u\Delta)/\sqrt{\log(n)}$$

whp. Hence we obtain

$$\|\tilde{x} - x_T\|_2 \le \varepsilon\theta\kappa(\mathbf{A})(C_2 w(\mathbb{B}) + u\Delta)/\sqrt{\log(n)}, \tag{9}$$

where $\kappa(\mathbf{A}) = \frac{\|\mathbf{A}\|_2}{\sigma_{\min}}$, the condition number of $\mathbf{A}$. Furthermore,

$$\|z\|_{\mathbb{E}} = \langle z, z\rangle = z^\top Q z = (zQ^{\frac{1}{2}})^\top Q^{\frac{1}{2}} z = \|Q^{\frac{1}{2}} z\|_2$$

(since the inner product in $\mathbb{E}$ is positive definite), whence $\|z\|_{\mathbb{E}} \le \|Q^{\frac{1}{2}}\|_2 \|z\|_2$. Finally, by (8) and (9), we get (7) as claimed. □

As regards optimality, $\langle c, \tilde{x}\rangle = \langle c, x_T\rangle + \langle c, \mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}(b - \mathbf{A}x_T)\rangle$, so by Cauchy-Schwartz we get

$$|\langle c, \tilde{x}\rangle - \langle c, x_T\rangle| \le \|c\|_2 \|\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\|_2 \|b - \mathbf{A}x_T\|_2.$$

By the proof of Theorem 4.4 $\|\mathbf{A}^\top(\mathbf{A}\mathbf{A}^\top)^{-1}\|_2 = \frac{1}{\sigma_{\min}}$, and we can bound $\|b - \mathbf{A}x_T\|_2$ by Proposition 4.2, yielding (with the above notation):

$$\mathsf{Prob}\left[|\langle c, \tilde{x}\rangle - \langle c, x_T\rangle| \le \varepsilon\theta\kappa(\mathbf{A})\|c\|_2 \frac{C_2 w(\mathbb{B}) + u\Delta}{\sqrt{\log(n)}}\right] \ge 1 - 2e^{-u^2}.$$

5. Computational validation

All our tests were carried out on a core of a quad-core Intel Xeon 2.1GHz CPU of a 8-CPU system with 64GB RAM running CentOS Linux, using Julia 0.6.1 and the Mosek 9.1.5 [15] SDP solver. The general structure of our algorithm is very simple:

1. solve the original SCP and reduce it to standard form;
2. sample an RP and project the equality constraints of the SCP
3. solve the projected SCP and retrieve its approximate solution
4. compare the original and retrieved solutions on errors and CPU time.

We employed Achlioptas' RP matrices [1] in the "sparser" variant of [13] with density 0.1.

In order to validate our approach, we need to establish that: (a) projected versions of infeasible instances do not turn out to be feasible excessively often; and (b) the optimality and feasibility errors of retrieved solutions are not too large. We consider a set of randomly generated SDP instances. The ε parameter in Definition 1.1 was set to 0.2 for the feasible set, and to 0.13 for the infeasible set. All random coefficients in the instance data were sampled from a uniform distribution on $[0, 1]$.

5.1. Infeasible instances

We generated two groups of 5 infeasible instances. The details of the generation process (number m of equality constraints, number n of variables, number d of equality constraints in the projected problem, density dens of generated matrices) are reported in the first five columns of Table 1. The results for are

m	n	d	dens	P	P_T	cpu	cpu$_T$
1000	820	716	0.5	Infeas	100%	5.44	2.92
1000	1275	763	0.5	Infeas	100%	6.19	4.73

Table 1: Infeasible instances. See Table 3 for detailed results.

presented in Table 1. We report the solution status obtained by the solver on P (always infeasible over all instances in all groups), the percentage of success in detecting infeasibility when solving P_T, and the CPU time taken to solve P and P_T. We remark that, by "solving P_T", we mean: (i) sample the RP, (ii) project the problem, and (iii) actually solve it. The success score was perfect. Results worsened when increasing ε: with $\varepsilon = 0.14$ the success rate was 80%; with $\varepsilon = 0.2$ the success rate was 0%.

5.2. Feasible instances

We generated three groups of 10 feasible instances. As above, $m, n, d,$ dens are reported in Table 2. The first and second groups both have identity cost matrices, but different generation densities; the third group has a random cost matrix. The results for the feasible set are presented in Table 2 (each line reports averages over the corresponding group). We report the optimal objective

m	n	d	dens	$\mathsf{v}(P)$	$\mathsf{v}(P_T)$	err	$\langle c, \tilde{x} \rangle$	err	cpu	cpu_T
2000	1540	332	0.2	1540.630	461.289	0.7	1540.630	0.0	8.55	6.76
2000	1540	332	0.5	1540.557	468.083	0.7	1540.557	0.0	17.28	8.31
4000	1830	340	0.1	87.133	-3126.177	1.0	87.133	0.0	23.68	11.11

Table 2: Feasible instances. See Table 4 for detailed results.

function value $\mathsf{v}(P)$ of the original problem, the optimal objective function value $\mathsf{v}(P_T)$ of the projected problem and the relative error

$$\frac{\mathsf{v}(P) - \mathsf{v}(P_T)}{\max(|\mathsf{v}(P)|, |\mathsf{v}(P_T)|)},$$

the objective function value $\langle c, \tilde{x} \rangle$ of the retrieved solution $\tilde{x}$ and the relative error

$$\frac{\mathsf{v}(P) - \langle c, \tilde{x} \rangle}{\max(|\mathsf{v}(P)|, |\langle c, \tilde{x} \rangle|)},$$

the CPU time taken to solve P and the CPU time taken to solve P_T. By "solving P_T" we mean: (i) sample the RP, (ii) project the problem, (iii) actually solve it, and (iv) retrieve the solution for P. Standard deviations on CPU time are between 1% and 5% of the averages.

In each case, our methodology was able to find the optimal solution of the original problem exactly, namely $\tilde{x} = x^*$, in roughly half the time. Accordingly, $\mathsf{v}(P) = \langle c, \tilde{x} \rangle$ for all instances; moreover, infeasibility and negative definiteness errors were always zero.

As is typical for RP-based methods, the ratio between projected and original CPU time decreases for increasing problem size: this is due to the fact that the number of constraints increases only logarithmically in the number of variables. Moreover, denser instances also gain a CPU advantage, due to the fact that denser problem take more time to solve, whereas projected problems are already typically denser than their original counterparts. The perfect score in solution quality is much less typical [20, 21]. While it certainly validates the approach, this type of success rate may well fail to hold for SDPs and other conic programs from real applications.

References

[1] D. Achlioptas. Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences*, 66:671–687, 2003.

[2] F. Alizadeh. An introduction to Formally Real Jordan Algebras and their applications in optimization. In M. Anjos and J. Lasserre, editors, *Handbook on Semidefinite, Conic and Polynomial Optimization*, volume 166 of *Operations Research & Management Science*. Springer, Boston, MA, 2012.

ID	m	n	d	dens	P	P_T	cpu	cpu_T
1	1000	820	716	0.5	Infeas	Infeas	5.39	2.94
2	1000	820	716	0.5	Infeas	Infeas	5.45	2.91
3	1000	820	716	0.5	Infeas	Infeas	5.43	2.90
4	1000	820	716	0.5	Infeas	Infeas	5.48	2.94
5	1000	820	716	0.5	Infeas	Infeas	5.46	2.91
1	1000	1275	763	0.5	Infeas	Infeas	6.06	4.70
2	1000	1275	763	0.5	Infeas	Infeas	6.11	4.76
3	1000	1275	763	0.5	Infeas	Infeas	6.15	4.70
4	1000	1275	763	0.5	Infeas	Infeas	6.53	4.80
5	1000	1275	763	0.5	Infeas	Infeas	6.09	4.71

Table 3: Detailed results on infeasible instances.

ID	m	n	d	dens	$\mathsf{v}(P)$	$\mathsf{v}(P_T)$	$\langle c, \tilde{x} \rangle$	cpu	cpu_T
1	2000	1540	332	0.2	1562.983	479.373	1562.983	8.53	6.55
2	2000	1540	332	0.2	1508.010	435.141	1508.010	8.47	6.79
3	2000	1540	332	0.2	1512.362	467.159	1512.362	8.70	6.68
4	2000	1540	332	0.2	1554.651	482.174	1554.651	8.61	6.76
5	2000	1540	332	0.2	1536.599	451.946	1536.599	8.59	6.79
6	2000	1540	332	0.2	1570.673	475.775	1570.673	8.53	6.77
7	2000	1540	332	0.2	1556.245	473.629	1556.245	8.52	6.70
8	2000	1540	332	0.2	1505.491	463.241	1505.491	8.52	6.87
9	2000	1540	332	0.2	1564.126	463.902	1564.126	8.51	6.94
10	2000	1540	332	0.2	1535.160	420.553	1535.160	8.55	6.70
1	2000	1540	332	0.5	1547.108	449.995	1547.108	16.84	8.16
2	2000	1540	332	0.5	1535.504	458.891	1535.504	17.12	8.43
3	2000	1540	332	0.5	1514.214	478.169	1514.214	17.45	8.02
4	2000	1540	332	0.5	1571.305	457.359	1571.305	17.32	8.31
5	2000	1540	332	0.5	1529.055	471.201	1529.055	17.37	8.28
6	2000	1540	332	0.5	1548.015	492.389	1548.015	17.49	8.90
7	2000	1540	332	0.5	1527.581	445.974	1527.581	17.24	8.09
8	2000	1540	332	0.5	1518.795	445.112	1518.795	17.23	8.65
9	2000	1540	332	0.5	1550.371	474.559	1550.371	17.35	8.05
10	2000	1540	332	0.5	1563.617	507.185	1563.617	17.36	8.20
1	4000	1830	340	0.1	45.489	-3080.881	45.489	22.19	10.97
2	4000	1830	340	0.1	128.385	-3181.674	128.385	24.27	10.72
3	4000	1830	340	0.1	100.892	-3300.381	100.892	24.50	10.99
4	4000	1830	340	0.1	88.365	-2941.452	88.365	24.30	11.09
5	4000	1830	340	0.1	155.619	-3005.528	155.619	22.14	11.21
6	4000	1830	340	0.1	77.195	-3353.770	77.195	24.01	11.62
7	4000	1830	340	0.1	38.441	-3021.306	38.441	24.40	11.08
8	4000	1830	340	0.1	30.648	-3020.162	30.648	24.87	11.08
9	4000	1830	340	0.1	135.078	-2985.326	135.078	24.39	11.27
10	4000	1830	340	0.1	71.217	-3371.290	71.217	21.69	11.09
							Average	*16.50*	***8.73***

Table 4: Detailed results on feasible instances.

[3] M. Andersen, J. Dahl, and L. Vandenberghe. Implementation of nonsymmetric interior-point methods for linear optimization over sparse matrix cones. *Mathematical Programming Computation*, 2:167–201, 2010.

[4] Andreas Bluhm and Daniel Stilck França. Dimensionality reduction of SDPs through sketching. *Linear Algebra and its Applications*, 563:461–475, 2019.

[5] C. Boutsidis, A. Zouzias, and P. Drineas. Random projections for k-means clustering. In *Advances in Neural Information Processing Systems*, NIPS, pages 298–306, La Jolla, 2010. NIPS Foundation.

[6] K. Clarkson and D. Woodruff. Low-rank PSD approximation in input-sparsity time. In *Proceedings of the 28th annual ACM Symposium on Discrete Algorithms*, volume 28 of *SODA*, page 20612072, Philadelphia, 2017. ACM.

[7] C. D'Ambrosio, L. Liberti, P.-L. Poirion, and K. Vu. Random projections for quadratic programs. *Mathematical Programming B*, 183:619–647, 2020.

[8] L. Ding and B. Grimmer. Bundle method sketching for low rank semidefinite programming. In A. D'Aspremont, M. Schmidt, S. Sra, and S. Stich, editors, *Workshop on Optimization and Machine Learning*, Vancouver, 2019.

[9] S. Dirksen. Dimensionality reduction with subgaussian matrices: A unified theory. *Foundations of Computational Mathematics*, 16:1367–1396, 2016.

[10] J. Faraut and A. Korányi. *Analysis on Symmetric Cones*. Clarendon Press, Oxford, 1994.

[11] P. Indyk and A. Naor. Nearest neighbor preserving embeddings. *ACM Transactions on Algorithms*, 3(3):Art. 31, 2007.

[12] W. Johnson and J. Lindenstrauss. Extensions of Lipschitz mappings into a Hilbert space. In G. Hedlund, editor, *Conference in Modern Analysis and Probability*, volume 26 of *Contemporary Mathematics*, pages 189–206, Providence, RI, 1984. AMS.

[13] D. Kane and J. Nelson. Sparser Johnson-Lindenstrauss transforms. *Journal of the ACM*, 61(1):4, 2014.

[14] L. Liberti and B. Manca. Side-constrained minimum sum-of-squares clustering: Mathematical Programming and random projections. *Journal of Global Optimization*, accepted.

[15] Mosek ApS. *The `mosek` manual, Version 9*, 2019.

[16] S. Schmieta and F. Alizadeh. Associative and Jordan algebras, and polynomial-time interior point algorithms for symmetric cones. *Mathematics of Operations Research*, 26(3):543–564, 2001.

[17] J. Tropp, A. Yurtsever, M. Udell, and V. Cevher. Practical sketching algorithms for low-rank matrix approximation. *SIAM Journal of Matrix Analysis and Applications*, 38(4):1454–1485, 2017.

[18] L. Vandenberghe. *The CVXOPT linear and quadratic cone program solvers*. University of California at Los Angeles, 2010.

[19] R. Vershynin. *High-dimensional probability*. CUP, Cambridge, 2018.

[20] K. Vu, P.-L. Poirion, C. D'Ambrosio, and L. Liberti. Random projections for quadratic programs over a Euclidean ball. In A. Lodi and *et al.*, editors, *Integer Programming and Combinatorial Optimization (IPCO)*, volume 11480 of *LNCS*, pages 442–452, New York, 2019. Springer.

[21] K. Vu, P.-L. Poirion, and L. Liberti. Random projections for linear programming. *Mathematics of Operations Research*, 43(4):1051–1071, 2018.

[22] K. Vu, P.-L. Poirion, and L. Liberti. Gaussian random projections for Euclidean membership problems. *Discrete Applied Mathematics*, 253:93–102, 2019.