FULL LENGTH PAPER

Series B

Random projections for quadratic programs

Claudia D'Ambrosio[1] · Leo Liberti[1] (ORCID) **· Pierre-Louis Poirion[2] · Ky Vu[3]**

Abstract

Random projections map a set of points in a high dimensional space to a lower dimensional one while approximately preserving all pairwise Euclidean distances. Although random projections are usually applied to numerical data, we show in this paper that they can be successfully applied to quadratic programming formulations over a set of linear inequality constraints. Instead of solving the higher-dimensional original problem, we solve the projected problem more efficiently. This yields a feasible solution of the original problem. We prove lower and upper bounds of this feasible solution w.r.t. the optimal objective function value of the original problem. We then discuss some computational results on randomly generated instances, as well as a variant of Markowitz' portfolio problem. It turns out that our method can find good feasible solutions of very large instances.

✉ Leo Liberti
liberti@lix.polytechnique.fr

Claudia D'Ambrosio
dambrosio@lix.polytechnique.fr

Pierre-Louis Poirion
pierre-louis.poirion@riken.jp

Ky Vu
kyvk2@fe.edu.vn

[1] CNRS LIX École Polytechnique, Institut Polytechnique de Paris, Palaiseau, France

[2] RIKEN Center for Advanced Intelligence Project, Tokyo, Japan

[3] Department of Mathematics, FPT University, Hoa Lac Hi-Tech Park, Hanoi, Vietnam

1 Introduction

The goal of this paper is to show that Random Projections (RP) applied to Quadratic Programming (QP) problems subject to linear inequality constraints yield a QP with fewer variables, the solution of which is an approximate solution for the original QP. We consider the following QP formulation:

$$\mathsf{P} \equiv \left. \begin{array}{c} \max_x \ x^\top Q x + c^\top x \\ Ax \le b, \end{array} \right\} \tag{1}$$

where x is a vector of n decision variables, Q is a symmetric $n \times n$ matrix, $c \in \mathbb{R}^n$, A is $m \times n$ and $b \in \mathbb{R}^m$.

We make three assumptions about Eq. (1):

1. We assume that $Ax \le b$ defines a full dimensional polytope.
2. We assume we are given the radius R of a ball containing the polytope $Ax \le b$.
3. We assume that all the rows of A are unit vectors. This assumption, however, is without loss of generality (w.l.o.g.): we let $\mu_i = \|A_i\|$, where A_i is the i-th row of A; we then replace each A_i by $\frac{A_i}{\mu_i}$ and b_i by $\frac{b_i}{\mu_i}$; this yields $\forall i \le n \ \frac{A_i^\top x}{\mu_i} \le \frac{b_i}{\mu_i}$, which satisfies the assumption.

We remark that Assumption 3 might affect the problem input adversely, yielding instances that end up being difficult to solve for numerical instability reasons. No assumption is made on Q, which may be positive/negative semidefinite or indefinite, unless otherwise stated in some parts of the paper. In the following, all norm symbols $\|\cdot\|$ will be assumed to refer to the ℓ_2 norm $\|\cdot\|_2$, unless otherwise stated.

QP is now a ripe field with many applications (e.g. portfolio optimization, constrained linear regression, monopoly policy determination and many more [6,18]). If we assume that all the data are rational, then the decision version of Eq. (1) is **NP**-complete [19].

RPs are random matrices which are used to perform dimensionality reduction on a set of vectors while approximately preserving all pairwise Euclidean distances with high probability. The goal of this paper is the applicability of RPs to bounded QPs such as those of Eq. (1). Specifically, we will define a projected version of Eq. (1) and prove that it is likely to have optimal objective function value close to that of the original QP. We also perform a computational verification of our claim, and discuss the extent to which our theoretical results can be applied in practice. More precisely, our approach can find feasible points for QPs that are too large to be solved in a given computational set-up. For very large scale QPs that take very long to solve, these feasible solutions can be used as starting points for the original QPs. This yields solutions for these QPs in less time overall.

RPs are usually applied to numerical data in view of speeding up algorithms which are essentially based on Euclidean distances, such as k-means [3] or k-nearest neighbours [2,8,9]. Since, according to the Johnson–Lindenstrauss lemma [10], RPs provide good approximations of Euclidean distances, it is perhaps not so surprising that they should work well in those settings. The focus of the present work is the much more

suprising statement that a Mathematical Programming (MP) formulation is approximately invariant (as regards feasibility and optimality) to randomly projecting the input parameters. Similarly to our previous work on Linear Programming (LP) [24], but using different projection and proof techniques, the results of this paper are independent of the solution algorithm used to solve the formulations. While RPs have already been applied to some optimization problems, these are usually unconstrained minimizations of ℓ_2 norms and/or assume small Gaussian or doubling dimension of the feasible set [16,26]: two assumptions we do not make.

With respect to the paper [23], we remove the need for a ball constraint, thus providing a generalization to *any* QP satisfying the Assumptions 1–3 above. While the QP formulation considered here differs from the formulation of [23] by a single constraint, the bounds and proof methodologies are considerably different, e.g. Theorem 2. Moreover, Proposition 1 is more streamlined, the bounds in Lemma 3 are improved, the bounds derived for convex QPs in Sect. 4.1 are new. We propose new sparse random projections in Sect. 5.1, and the computational experiments are also new.

We note that, by considering the case $Q = 0$, our result also yields a RP technique for LPs in canonical form $\max\{c^\top x \mid Ax \leq b\}$. By considering the dual, one can easily show that one obtains a projected formulation for LPs in standard form. By inspection, this projected formulation turns out to be exactly equal to the one discussed in [24]. Thus, the restriction of the results in this paper to the LP case yields a new analysis for the projected formulation of [24], under somewhat different assumptions.

The rest of this paper is organized as follows. In Sect. 2 we define RPs, the projected QP, and the solution retrieval operation. In Sect. 3 we introduce some basic properties of RPs. In Sect. 4 we prove our main results about approximate optimality of the projected QP. In Sect. 5 we discuss computational results.

2 The projected problem

RPs are simple but powerful tools for dimension reduction [12,16,24–26]. They are often constructed as random matrices sampled from some given distribution classes. The simplest examples are suitably scaled matrices sampled componentwise from independently identically distributed (i.i.d.) random variables with sub-gaussian distributions [22]: e.g. $N(0, 1)$, uniform on $[-1, 1]$, or Rademacher ± 1 distributions. One of the most important features of a RP is that it approximately preserves the norm of any given vector with high probability [20]. In particular, let $P \in \mathbb{R}^{d \times n}$ be a RP every component of which is sampled from $N(0, 1/\sqrt{d})$ (where $1/d$ is the variance). Then, for any $x \in \mathbb{R}^n$ and $\varepsilon \in (0, 1)$, we have

$$\mathsf{Prob}\left[(1 - \varepsilon)\|x\|^2 \leq \|Px\|^2 \leq (1 + \varepsilon)\|x\|^2\right] \geq 1 - 2e^{-\mathcal{C}\varepsilon^2 d}, \tag{2}$$

where $\mathcal{C}$ is a *positive universal constant* (in fact a more precise statement should be existentially quantified by "there exists a constant $\mathcal{C}$ such that…").

Perhaps the most famous application of RPs is the Johnson–Lindenstrauss lemma [10]. It states that, for any $\varepsilon \in (0, 1)$ and for any finite set $X \subseteq \mathbb{R}^n$ there is a mapping

$F : \mathbb{R}^n \to \mathbb{R}^d$, where $d = O(\ln |X|/\varepsilon^2)$, such that

$$\forall x, y \in X \quad (1 - \varepsilon)\|x - y\|^2 \leq \|F(x) - F(y)\|^2 \leq (1 + \varepsilon)\|x - y\|^2. \quad (3)$$

Such a mapping F can be realized as the linear operator represented by the matrix P above. The existence of the correct mapping is shown (by the probabilistic method) using the union bound. Moreover, the probability of sampling a correct mapping can be made arbitrarily high. In practice, we found that there is often no need to re-sample P in algorithmic applications: even if a sampled P does not satisfy Eq. (3), the pairs x, y for which Eq. (3) is false are very few, which makes it unlikely that the resulting error will significantly impact the outcome.

2.1 The randomly projected QP

Let $P \in \mathbb{R}^{d \times n}$ be a RP. We want to "project" each vector $x \in \mathbb{R}^n$ to a lower dimensional vector $Px \in \mathbb{R}^d$. Consider the following *projected problem*:

$$\max \left\{ x^\top (P^\top P Q P^\top P)x + c^\top P^\top Px \mid A P^\top Px \leq b \right\}.$$

By setting $u = Px$, $\bar{c} = Pc$, $\bar{A} = A P^\top$, $\bar{Q} = P Q P^\top$, we can rewrite it as

$$\max_{u \in \mathsf{Im}(P)} \left\{ u^\top \bar{Q}u + \bar{c}^\top u \mid \bar{A}u \leq b \right\}, \quad (4)$$

where $\mathsf{Im}(P)$ is the image space of P. Since P is (randomly) generated with full rank with probability 1, it is very likely to be a surjective mapping. Therefore, we assume it is safe to remove the constraint $u \in \mathsf{Im}(P)$ and study the smaller dimensional problem:

$$\mathsf{RP} \equiv \max_{u \in \mathbb{R}^d} \left\{ u^\top \bar{Q}u + \bar{c}^\top u \mid \bar{A}u \leq b \right\}, \quad (5)$$

where u ranges in $\mathbb{R}^d$. As we will show later, Eq. (5) yields a good approximate solution of Eq. (1) with high probability.

2.2 Solution retrieval

When we solve RP we obtain a solution in the projected space $\mathbb{R}^d$. In this section we discuss the issue of *solution retrieval*, i.e. how to exploit the projected solution in order to derive a reasonable solution for the original problem P. It turns out that feasibility is not an issue, as feasibility in the projected problem implies feasibility in the original problem.

Proposition 1 *Let $P \in \mathbb{R}^{d \times n}$ be a RP. For any feasible solution u of the projected problem* (5)*, $P^\top u$ is also feasible for the original problem in Eq. (1).*

Proof Let u be any feasible solution for the projected problem (5) and take $\hat{x} = P^\top u$. Then we have $A\hat{x} = AP^\top u \leq b$. $\qquad\square$

The approximate optimality of $P^\top u$ will be established in Sect. 4.

For a feasible and bounded MP formulation F we let $\mathsf{v}(\mathsf{F})$ be the optimal objective function value of F.

Proposition 2 $\mathsf{v}(\mathsf{RP}) \leq \mathsf{v}(\mathsf{P})$.

Proof Let u^* be an optimum of RP. Let $x' = P^\top u^*$. As remarked above, x' is feasible for P. Now observe that the objective function value of P at x' is

$$(x')^\top Qx + c^\top x' = u^{*\top} PQP^\top u^* + (Pc)^\top u^* = u^{*\top} \bar{Q} u^* + \bar{c}^\top u^* = \mathsf{v}(\mathsf{RP}).$$

Since P, RP are maximization problems, the result follows. $\qquad\square$

Since P is in maximization form, we remark that Proposition 2 does not prove that RP is a relaxation of P. We make this statement explicitly in order to mark a notable difference with our analysis of RPs and LPs in [24].

3 Basic properties of random projections

In this section we introduce three results that are going to be used later on in the paper. Lemma 1 shows that RPs approximately preserve scalar products. Lemma 2 shows that RPs approximately preserve double-sided inequalities. Lemma 3 shows that RPs approximately preserve quadratic forms. Throughout the paper, M_i denotes the i-th row of the matrix (or vector) M; for a product between matrices A, B, for example, we use the notation $(AB)_i$. Vectors (even when they are extracted from rows of matrices) may be used as row or column vectors depending on whether they multiply from the left or from the right.

Lemma 1 *Let* $P \in \mathbb{R}^{d \times n}$ *be a RP satisfying Eq. (2) and let* $0 < \varepsilon < 1$. *Then there is a universal constant* $\mathcal{C}$ *such that, for any* $x, y \in \mathbb{R}^n$,

$$\langle x, y \rangle - \varepsilon \|x\| \|y\| \leq \langle Px, Py \rangle \leq \langle x, y \rangle + \varepsilon \|x\| \|y\|$$

with probability at least $1 - 4e^{-\mathcal{C}\varepsilon^2 d}$.

A proof of this lemma can be found in [24, Eq. 7 and Prop. 1]; we reprove the result here to make the paper self-contained, and because it is often used in the rest of the paper.

Proof Let $\mathcal{C}$ be the same universal constant as in Eq. (2). By the property in Eq. (2), for any two vectors $u + v$, $u - v$ and using the union bound, we have

$$|\langle Pu, Pv \rangle - \langle u, v \rangle| = \frac{1}{4}\left|\|P(u+v)\|^2 - \|P(u-v)\|^2 - \|u+v\|^2 + \|u-v\|^2\right|$$

$$\leq \frac{1}{4}\left|\|P(u+v)\|^2 - \|u+v\|^2\right| + \frac{1}{4}\left|\|P(u-v)\|^2 - \|u-v\|^2\right|$$

$$\leq \frac{\varepsilon}{4}(\|u+v\|^2 + \|u-v\|^2) = \frac{\varepsilon}{2}(\|u\|^2 + \|v\|^2),$$

with probability at least $1 - 4e^{-\mathcal{C}\varepsilon^2 d}$. Apply this result to $u = \frac{x}{\|x\|}$ and $v = \frac{y}{\|y\|}$, to obtain the desired inequality. □

Lemma 2 *Let $P \in \mathbb{R}^{d \times n}$ be a RP satisfying Eq. (2), let $0 < \varepsilon < 1$, and let $\mathbf{1}$ be the all-one vector. Then there is a universal constant $\mathcal{C}$ such that, for any $x \in \mathbb{R}^n$ and $A \in \mathbb{R}^{m \times n}$ having unit row vectors, we have*

$$Ax - \varepsilon\|x\|\mathbf{1} \leq AP^\top Px \leq Ax + \varepsilon\|x\|\mathbf{1}$$

with probability at least $1 - 4\,m\,e^{-\mathcal{C}\varepsilon^2 d}$.

Proof Let $A_1, \ldots, A_m$ be (unit) row vectors of A. Then

$$AP^\top Px - Ax = \begin{pmatrix} A_1^\top P^\top Px - A_1^\top x \\ \ldots \\ A_m^\top P^\top Px - A_m^\top x \end{pmatrix} = \begin{pmatrix} \langle PA_1, Px \rangle - \langle A_1, x \rangle \\ \ldots \\ \langle PA_m, Px \rangle - \langle A_m, x \rangle \end{pmatrix}.$$

The claim follows by applying Lemma 1 and the union bound. □

Lemma 3 *Let $P \in \mathbb{R}^{d \times n}$ be a RP satisfying Eq. (2) and let $0 < \varepsilon < 1$. Then there is a universal constant $\mathcal{C}$ such that, for any two vectors $x, y \in \mathbb{R}^n$ and a square matrix $Q \in \mathbb{R}^{n \times n}$, with probability at least $1 - 8\,k\,e^{-\mathcal{C}\varepsilon^2 d}$, we have:*

$$x^\top Qy - 3\varepsilon\|x\|\,\|y\|\,\|Q\|_F \leq x^\top P^\top PQP^\top Py \leq x^\top Qy + 3\varepsilon\|x\|\,\|y\|\,\|Q\|_F,$$

where $\|Q\|_F$ is the Frobenius norm of Q and k is the rank of Q.

Proof Let $Q = U\Sigma V^\top$ be the singular value decomposition of Q. Here U, V are $(n \times k)$-real matrices with orthogonal unit column vectors $u_1, \ldots, u_k$ and $v_1, \ldots, v_k$, respectively and $\Sigma = \mathsf{diag}(\sigma_1, \ldots, \sigma_k)$ is a diagonal real matrix with positive entries. Denote by $\mathbf{1}_k = (1, \ldots, 1)^\top$ the k-dimensional column vector of all 1 entries. By adding and subtracting $x^\top U$ and $y^\top V$ to the left and (respectively) right terms around the matrix Σ, and transposing twice each term, we obtain:

$$x^\top P^\top PQP^\top Py = (U^\top P^\top Px)^\top \Sigma(V^\top P^\top Py)$$

$$= \left[U^\top x + U^\top(P^\top P - I_n)x\right]^\top \Sigma\left[V^\top y + V^\top(P^\top P - I_n)y\right].$$

Therefore,

$$
\begin{aligned}
|x^\top P^\top P Q P^\top P y - x^\top Q y| \leq\ & |(U^\top x)^\top \Sigma V^\top (P^\top P - I_n) y| \\
& + |(U^\top (P^\top P - I_n) x)^\top \Sigma V^\top y| \\
& + |(U^\top (P^\top P - I_n) x)^\top \Sigma V^\top (P^\top P - I_n) y|, \quad (6)
\end{aligned}
$$

since $U^\top x \Sigma V^\top x = x^\top Q y$. Noting that for every $i \leq k$, $(\Sigma V^\top P^\top P y)_i = \langle P(\Sigma V^\top)_i, P y \rangle$ and $(\Sigma V^\top y)_i = \langle (\Sigma V^\top)_i, y \rangle$, by Lemma 1 we have

$$
\forall i \leq k \quad |(\Sigma V^\top (P^\top P - I_n) y)_i| \leq \varepsilon \sigma_i \|y\|. \quad (7)
$$

Moreover, since Σ is symmetric, we also have

$$
\forall i \leq k \quad |((U^\top (P^\top P - I_n) x)^\top \Sigma)_i| \leq \varepsilon \sigma_i \|x\|. \quad (8)
$$

We now apply the Cauchy–Schwartz inequality to the terms on the right hand side of Eq. (6):

$$
\langle U^\top x, \Sigma V^\top (P^\top P - I_n) y \rangle \leq \|U^\top x\| \|\Sigma V^\top (P^\top P - I_n) y\| \quad (9)
$$

$$
\langle (U^\top (P^\top P - I_n) x)^\top \Sigma, V^\top y \rangle \leq \|(U^\top (P^\top P - I_n) x)^\top \Sigma\| \|V^\top y\| \quad (10)
$$

$$
\langle (U^\top (P^\top P - I_n) x)^\top \Sigma, V^\top (P^\top P - I_n) y \rangle \leq \|(U^\top (P^\top P - I_n) x)^\top \Sigma\| \|V^\top (P^\top P - I_n) y\|. \quad (11)
$$

We claim that

$$
\|U^\top x\| \leq \|x\| \quad (12)
$$

$$
\|V^\top y\| \leq \|y\|. \quad (13)
$$

Since the columns of U are orthonormal, $UU^\top x$ is the projection of x on the column space of U. Since projected vectors are always shorter, we have $\|UU^\top x\| \leq \|x\|$. Now $\|UU^\top x\|^2 = x^\top U U^\top U U^\top x$. Again by orthonormality of the columns of U, $U^\top U$ is the identity; hence $\|UU^\top x\|^2 = x^\top U U^\top x = \|U^\top x\|^2$, which yields $\|U^\top x\| = \|UU^\top x\| \leq \|x\|$ as claimed. The argument for V, y is the same. Next, we use Eqs. (12)–(13) and Eqs. (7)–(8) to bound Eqs. (9)–(11). We obtain:

$$
\langle U^\top x, \Sigma V^\top (P^\top P - I_n) y \rangle \leq \varepsilon \|x\| \|y\| \|\sigma\| \quad (14)
$$

$$
\langle (U^\top (P^\top P - I_n) x)^\top \Sigma, V^\top y \rangle \leq \varepsilon \|y\| \|x\| \|\sigma\| \quad (15)
$$

$$
\langle (U^\top (P^\top P - I_n) x)^\top \Sigma, V^\top (P^\top P - I_n) y \rangle \leq \varepsilon^2 \|x\| \|y\| \|\sigma\|, \quad (16)
$$

where $\sigma = (\sigma_1, \ldots, \sigma_k)$ and the bound for Eq. (16) was obtained from Eq. (11) by setting $\Sigma = I_k$ in Eqs. (7)–(8). Finally, we obtain

$$
|x^\top P^\top P Q P^\top P y - x^\top Q y| \leq 3\varepsilon \|x\| \|y\| \|\sigma\| \quad (17)
$$

with probability at least $1 - 8k e^{-C \varepsilon^2 d}$. □

4 Approximate optimality

We now prove that the objective of the quadratic problem in Eq. (1) is approximately preserved under RPs. To do so, we study the relationship between Eq. (5) and

$$\mathsf{RP}_\varepsilon \equiv \max_{u \in \mathbb{R}^d} \{u^\top \bar{Q}u + \bar{c}^\top u \mid \bar{A}u \leq b + R\varepsilon \}. \tag{18}$$

Let u^- be an optimal solution for RP (Eq. (5)) and u^+ be an optimal solution for RP_ε (Eq. (18)). Denote by $x^- = P^\top u^-$ and $x^+ = P^\top u^+$. Let x^* be an optimal solution for the original problem P (Eq. (1)). We will bound $\mathsf{v}(\mathsf{P}) = x^{*\top} Q x^* + c^\top x^*$ between $\mathsf{v}(\mathsf{RP}) = x^{-\top} Q x^- + c^\top x^-$ and $\mathsf{v}(\mathsf{RP}_\varepsilon) = x^{+\top} Q x^+ + c^\top x^+$, the two values that are expected to be approximately close to each other.

From here onwards, we introduce an arbitrary $\delta > 0$ which will appear in statements about arbitrarily high probability: namely, most of our results will hold with probability $1 - \delta$ (or, occasionally, $1 - 2\delta$). For this to make sense, we have to bound the projected dimension d from below, so that $1 - \delta$ is a lower bound to various expressions tending to 1 exponentially fast. The most stringent such expression, obtained by union bound arguments, turns out to be $1 - 4(m4k + 2k + 1)e^{-\mathcal{C}\varepsilon^2 d}$, where $\mathcal{C}$ is a (positive) universal constant, $k \leq n$ is the rank of Q, and $\varepsilon \in (0, 1)$. The corresponding bound on d is

$$d \geq \frac{\ln(4(m + 2k + 1)/\delta)}{\mathcal{C}\varepsilon^2}. \tag{19}$$

Theorem 1 *For any $\delta > 0$, let $d, \varepsilon, \mathcal{C}$ be as in Eq. (19), and $P \in \mathbb{R}^{d \times n}$ be a RP. Then the following holds with probability at least $1 - \delta$:*

$$\mathsf{v}(\mathsf{RP}) \leq \mathsf{v}(\mathsf{P}) \leq \mathsf{v}(\mathsf{RP}_\varepsilon) + 3R^2\varepsilon\|Q\|_F + R\varepsilon\|c\|.$$

Proof The constant $\mathcal{C}$ is chosen as in Eq. (2). Note that by Proposition 1 x^- is feasible in Eq. (1). By Proposition 2 we have $\mathsf{v}(\mathsf{RP}) \leq \mathsf{v}(\mathsf{P})$. Moreover, by Lemmas 3 and 1, with probability at least $1 - (8k + 4)e^{-\mathcal{C}\varepsilon^2 d}$, where k is the rank of Q, we have the following two double-sided inequalities:

$$x^{*\top} Q x^* \leq x^{*\top} P^\top P Q P^\top P x^* + 3\varepsilon\|x^*\|^2 \|Q\|_F \leq x^{*\top} P^\top P Q P^\top P x^* + 3R^2\varepsilon\|Q\|_F$$
$$c^\top x^* \leq c^\top P^\top P x^* + \varepsilon\|c\| \|x^*\| \leq c^\top P^\top P x^* + R\varepsilon\|c\|,$$
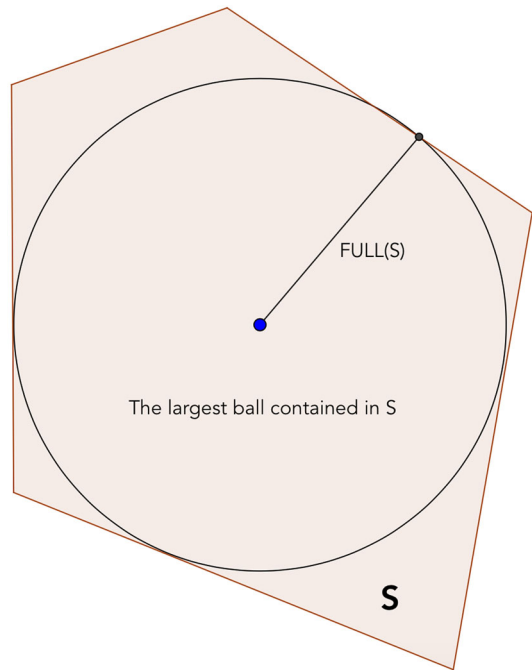
since $\|x^*\| \leq R$ (we note that the terms $8k$ and 4 in the probability estimation expression is due to the application of the union bound to Lemmas 3 and 1, respectively). Hence

$$\mathsf{v}(\mathsf{P}) = x^{*\top} Q x^* + c^\top x^* \leq x^{*\top} P^\top P Q P^\top P x^* + c^\top P^\top P x^* + R\varepsilon\|c\| + 3R^2\varepsilon\|Q\|_F.$$

On the other hand, let $\hat{u} = Px^*$; by Lemma 2, we have

$$A P^\top \hat{u} = A P^\top P x^* \leq A x^* + \varepsilon\|x^*\|\mathbf{1} \leq A x^* + R\varepsilon\mathbf{1} \leq b + R\varepsilon\mathbf{1}$$

Fig. 1 Fullness of a set

with probability at least $1 - 4\,m\,e^{-C\varepsilon^2 d}$. Therefore, $\hat{u}$ is a feasible solution for problem (18) with probability at least $1 - 4me^{-C\varepsilon^2 d}$. Due to the optimality of u^+ for problem (18), it follows that

$$
\begin{aligned}
\mathsf{v}(\mathrm{P}) = x^{*\top} Q x^* + c^\top x^* &\le x^{*\top} P^\top P Q P^\top P x^* + c^\top P^\top P x^* + R\varepsilon \|c\| + 3R^2 \varepsilon \|Q\|_F \\
&= \hat{u}^\top P Q P^\top \hat{u} + c^\top P^\top \hat{u} + R\varepsilon \|c\| + 3R^2 \varepsilon \|Q\|_F \\
&\le u^{+\top} P Q P^\top u^+ + (Pc)^\top u^+ + R\varepsilon \|c\| + 3R^2 \varepsilon \|Q\|_F \\
&= x^{+\top} Q x^+ + c^\top x^+ + R\varepsilon \|c\| + 3R^2 \varepsilon \|Q\|_F \\
&= \mathsf{v}(\mathrm{RP}_\varepsilon) + R\varepsilon \|c\| + 3R^2 \varepsilon \|Q\|_F
\end{aligned}
$$

with probability at least $1 - 4(m + 2k + 1)e^{-C\varepsilon^2 d}$, which is at least $1 - \delta$ for the chosen universal constant $\mathcal{C}$ (again, the term $4(m + 2k + 1)$ is due to the application of the union bound to Lemma 2 and to the result above). Hence $\mathsf{v}(\mathrm{P}) \le \mathsf{v}(\mathrm{RP}_\varepsilon) + 3R^2 \varepsilon \|Q\|_F + R\varepsilon \|c\|$, which concludes the proof. □

The above result implies that the value of $\mathsf{v}(\mathrm{P})$ lies between $\mathsf{v}(\mathrm{RP})$ and $\mathsf{v}(\mathrm{RP}_\varepsilon)$ (plus some error term). We will now prove that the latter two values are not so far from each other. In order to achieve this goal, we associate with each set S a "fullness measure" $\mathrm{FULL}(S) > 0$, defined as the maximum radius of any closed ball contained in S.

We now let S be the feasible region of Eq. (1). By Assumption 1 on page 2, S is full dimensional, which implies $\mathrm{FULL}(S) = r > 0$ (see Fig. 1).

We define S_ε^+ as the feasible region of Eq. (18). The following lemma characterizes the fullness of S_ε^+ with respect to r. Its proof extensively uses the fact that, for any row vector $a \in \mathbb{R}^n$, we have $\max_{\|u\| \le r} a^\top u = r\|a\|$ (the equality condition in the Cauchy–Schwartz inequality).

Lemma 4 *For any $\delta > 0$, let $d, \varepsilon, \mathcal{C}$ be as in Eq. (19), and $P \in \mathbb{R}^{d \times n}$ be a RP. Then S_ε^+ is full-dimensional with $\mathrm{FULL}(S_\varepsilon^+) \ge (1 - \varepsilon)r$ with probability at least $1 - \delta$.*

Proof By Assumption 3, every row A_i of A has unit norm. For some $x_0 \in S$, let $B(x_0, r) \subset S$ be a closed ball in S of maximum radius r. Then for any $x \in \mathbb{R}^n$ with $\|x\| \le r$, we have $x_0 + x \in S$, and hence $A(x_0 + x) = Ax_0 + Ax \le b,$. This implies that

$$\forall i \le m \quad b_i \ge (Ax_0)_i + \max_{\|x\| \le r} A_i\, x = (Ax_0)_i + r\|A_i\| = (Ax_0)_i + r. \quad (20)$$

Hence $Ax_0 + r\mathbf{1} \le b$ or, equivalently, $Ax_0 \le b - r\mathbf{1}$. By Lemma 2, with probability at least $1 - \delta$, we have $AP^\top Px_0 \le Ax_0 + R\varepsilon\mathbf{1} \le b + (R\varepsilon - r)\mathbf{1}$, where R is the radius in Assumption 2. Let $u \in \mathbb{R}^n$ with $\|u\| \le (1 - \varepsilon)r$, then, by the above inequality, we have:

$$\forall i \le m \quad (AP^\top(Px_0 + u))_i = (AP^\top Px_0)_i + (AP^\top u)_i \le b_i + R\varepsilon - r + (AP^\top)_i u$$
$$= b_i + R\varepsilon - r + A_i P^\top u.$$

By the Cauchy–Schwartz inequality we have

$$\forall i \le m \quad (AP^\top(Px_0 + u))_i \le b_i + R\varepsilon - r + (1 - \varepsilon)r\|A_i P^\top\|.$$

By applying Eq. (2) m times and the union bound, we can see that, with probability at least $1 - 2me^{-\mathcal{C}\varepsilon^2 d}$ (which, by definition of δ, is also greater than $1 - \delta$), we have

$$\forall i \le m \quad (1 - \varepsilon)\|A_i\| \le \|A_i P^\top\| \le (1 + \varepsilon)\|A_i\| = (1 + \varepsilon). \quad (21)$$

We note that $-r + (1 - \varepsilon)r(1 + \varepsilon) = r(1 - \varepsilon^2) - r = -\varepsilon^2 r < 0$. Hence

$$AP^\top(Px_0 + u) \le b + (R\varepsilon - r)\mathbf{1} + (1 - \varepsilon)(1 + \varepsilon)r\mathbf{1} \le b + R\varepsilon\mathbf{1}$$

with probability at least $1 - \delta$. In other words, with probability at least $1 - \delta$, the closed ball $\bar{B}$ centered at Px_0 with radius $(1 - \varepsilon)r$ is contained in $\{u \mid AP^\top u \le b + R\varepsilon\mathbf{1}\}$. Therefore, by definition of S_ε^+ we have $B\big(Px_0, (1 - \varepsilon)r\big) \subseteq S_\varepsilon^+$, which implies that the fullness of S_ε^+ is at least $(1 - \varepsilon)r$ with probability $\ge 1 - \delta$. $\qquad\square$

Now we will estimate the gap between the two objective functions of the problems Eqs. (5) and (18) using the fullness measure. The next theorem states that, as long as the fullness of the original polyhedron is large enough, the gap between them is $O(\varepsilon)$. This ensures that the bounds around the objective function value of Eq. (1) derived in Theorem 1 are not themselves unbounded above and below.

Theorem 2 *For any $\delta > 0$, let $\varepsilon \in (0, \frac{r}{R})$ and $d, \mathcal{C}$ be as in Eq.* (19). *Then with probability at least $1 - \delta$ we have*

$$\sigma^2 \mathsf{v}(RP_\varepsilon) \leq \mathsf{v}(RP) \leq \mathsf{v}(RP_\varepsilon),$$

where $\sigma = 1 - \frac{R}{r} \frac{\varepsilon}{(1-\varepsilon)^2}$ is assumed positive, $r = $ FULL(S), and S is the feasible region of P.

Proof The fact that $\mathsf{v}(RP) \leq \mathsf{v}(RP_\varepsilon)$ follows easily because RP_ε is a relaxation of RP. We now tackle the nontrivial part of the proof. Let v be the optimal solution of RP and w of RP_ε. We are going to prove the theorem by means of a scaling technique which uses w to derive a solution $\bar{w}$ feasible in RP. We shall first assume that $b > 0$ and later we shall prove that this assumption loses no generality. Because w is feasible in RP_ε (see Eq. (18)),

$$\forall i \leq m \quad \bar{A}_i w \leq b_i + R\varepsilon.$$

Assuming $b > 0$, this is equivalent to

$$\forall i \leq m \quad \frac{1}{b_i} \bar{A}_i w \leq 1 + \frac{R\varepsilon}{b_i}. \tag{22}$$

We divide both sides of the above inequality by $1 + R\varepsilon / \min_i b_i$ (i.e., the maximum value of $1 + \frac{R\varepsilon}{b_i}$ over $i \leq m$). This allows us to define $\bar{w} = \rho w$, where $\rho = \frac{1}{1+(R\varepsilon / \min_i b_i)}$ (we note that $0 < \rho < 1$). From Eq. (22) we deduce that $\frac{1}{b_i} \bar{A}_i \bar{w} \leq 1 \Rightarrow \bar{A}_i \bar{w} \leq b_i$ for each $i \leq m$. In other words, $\bar{w}$ is feasible in RP (see Eq. (5)). This yields

$$\rho^2 \mathsf{v}(RP_\varepsilon) \leq \rho^2 w^\top \bar{Q} w + \rho \bar{c}^\top w = \bar{w}^\top \bar{Q} \bar{w} + \bar{c}^\top \bar{w} \leq \mathsf{v}(RP) \tag{23}$$

by optimality.

We now remove the assumption $b > 0$ by a change of coordinates which will yield an equivalent system with positive right hand side $\hat{b}$. By Lemma 4, with probability $1 - \delta$ there is a ball $B(u_0, (1 - \varepsilon)r)$ contained in S_ε^+, the feasible region of RP_ε. We define new variables $u' = u - u_0$. Replacing u by $u' + u_0$ in Eq. (18) yields

$$\forall i \leq m \quad \bar{A}_i u' \leq b_i + R\varepsilon - \bar{A}_i u_0.$$

We let $\hat{b} = b - \bar{A} u_0$. Because $B(u_0, (1 - \varepsilon)r)$ is contained in S_ε^+, we have:

$$\forall i \leq m \quad \max\{\bar{A}_i u' \mid \|u'\| \leq (1 - \varepsilon)r\} \leq \hat{b}_i + R\varepsilon. \tag{24}$$

We now observe that $\max\{\bar{A}_i u' \mid \|u'\| \leq (1 - \varepsilon)r\} = \|\bar{A}_i\|(1 - \varepsilon)r$ for every $i \leq m$, since the maximum of $\bar{A}_i u'$ is achieved when u' is parallel to $\bar{A}_i$ and on the boundary of $B(u_0, (1 - \varepsilon)r)$. By Eq. (2), for all $i \leq m$ we have $\|\bar{A}_i\| = \|A_i P^\top\| = \|P A_i\| \geq (1 - \varepsilon)\|A_i\|$ (note that the probability of this event was already accounted for in

Lemma 4, Eq. (21); so, conditionally to that assumption, the probability of its holding in the context of this proof is 1). Since $\|A_i\| = 1$ by Assumption 3, we conclude that $\|\bar{A}_i\| \geq 1 - \varepsilon$. By Eq. (24) we have $\forall i \leq m \ (1 - \varepsilon)^2 r - R\varepsilon \leq \hat{b}_i$. This yields

$$\min_i \hat{b}_i \geq (1 - \varepsilon)^2 r - R\varepsilon. \tag{25}$$

After the change of variables, ρ becomes $\frac{1}{1+(R\varepsilon/\min_i \hat{b}_i)}$. We can now use the lower bound in Eq. (25) on $\min_i \hat{b}_i$ and derive $\sigma = 1 - \frac{R}{r} \frac{\varepsilon}{(1-\varepsilon)^2}$. Note that the condition $\sigma > 0$ is equivalent to $(1 - \varepsilon)^2 r - R\varepsilon > 0$, which ensures that $\hat{b}_i > 0$ for all $i \leq m$. Moreover, if we consider that $(1 - \varepsilon)^2 < 1$ for all $\varepsilon > 0$, we can obtain from $\sigma > 0$ an upper bound on $\varepsilon < \frac{r}{R}$.

Lastly, we take care of the effect of the variable change on Eq. (23). We let w' be the optimal solution of RP_ε after the variable change, so that $w = w' + u_0$. For any $i \leq m$, we have $r(1 - \varepsilon)^2 - R\varepsilon \leq \hat{b}_i$. By definition of σ, we have $r(1 - \varepsilon)^2 = \frac{R\varepsilon}{(1-\sigma)}$; therefore

$$\frac{R\varepsilon}{(1 - \sigma)} - R\varepsilon \leq \hat{b}_i \Leftrightarrow \sigma R\varepsilon \leq (1 - \sigma)\hat{b}_i \Leftrightarrow \sigma(\hat{b}_i + R\varepsilon) \leq \hat{b}_i.$$

Again by definition, $\bar{A}_i w' \leq \hat{b}_i + R\varepsilon$ for any $w' = w - u_0$ and w feasible for Eq. (18). Therefore

$$\bar{A}_i \sigma w' \leq \sigma(\hat{b}_i + R\varepsilon) \leq \hat{b}_i.$$

We define $\tilde{w} = \sigma w' + u_0$ and note that it is feasible since $\bar{A}_i \tilde{w} = \sigma \bar{A}_i w' + \bar{A}_i u_0 \leq \hat{b}_i + \bar{A}_i u_0 = b_i$. We therefore replace w in Eq. (23) by $w' + u_0$ and ρ by σ (since $\sigma \leq \rho$), which yields:

$$\sigma^2 \mathsf{v}(\mathsf{RP}_\varepsilon) \leq \sigma^2 (w')^\top \bar{Q} w' + \sigma \bar{c}^\top w' + u_0^\top \bar{Q} u_0 + \bar{c}^\top u_0 = \tilde{w}^\top \bar{Q} \tilde{w} + \bar{c}^\top \tilde{w} \leq \mathsf{v}(\mathsf{RP}),$$

which concludes the proof. $\square$

We remark that the assumption $\sigma > 0$ holds whenever $0 < \varepsilon < \max\left(\frac{r}{4R}, \frac{1}{2}\right)$.

Corollary 1 *For any $\delta > 0$, let $\varepsilon \in (0, \frac{r}{R})$, σ be as above, and $d, \mathcal{C}$ be as in Eq. (19). Then, with probability at least $1 - 2\delta$, we have*

$$\mathsf{v}(\mathsf{RP}) \leq \mathsf{v}(\mathsf{P}) \leq \frac{1}{\sigma^2} \mathsf{v}(\mathsf{RP}) + 3R^2 \varepsilon \|Q\|_F + R\varepsilon \|c\|.$$

Proof By Theorem 1, with probability at least $1 - \delta$:

$$\mathsf{v}(\mathsf{RP}) \leq \mathsf{v}(\mathsf{P}) \leq \mathsf{v}(\mathsf{RP}_\varepsilon) + 3R^2 \varepsilon \|Q\|_F + R\varepsilon \|c\|.$$

On the other hand, by Theorem 2, with probability at least $1 - \delta$:

$$\mathsf{v}(\mathsf{RP}_\varepsilon) \leq \frac{1}{\sigma^2}\mathsf{v}(\mathsf{RP}).$$

The claim then follows by the union bound. $\square$

4.1 Convex QPs

In this section we assume that Q is negative semidefinite, and hence that the problem P in Eq. (1) is convex, since it maximizes a concave function subject to linear constraints.

We first recall a duality result in QP theory [5]. The Dorn dual of the QP P (Eq. (1)) is as follows:

$$\mathsf{D} \equiv \left. \begin{array}{c} \min\limits_{y,v} -y^\top Q y + b^\top v \\ A^\top v - Q y = c \\ v \geq 0, \end{array} \right\} \tag{26}$$

where $y, v \in \mathbb{R}^n$. Similarly, we define the Dorn dual of the projected problem RP (Eq. (5)):

$$\mathsf{RD} \equiv \left. \begin{array}{c} \min\limits_{z,v} -z^\top \bar{Q} z + b^\top v \\ \bar{A}^\top v - \bar{Q} z = \bar{c} \\ v \geq 0, \end{array} \right\} \tag{27}$$

where $z \in \mathbb{R}^d$ and $v \in \mathbb{R}^n$.

Theorem 3 *Let $\varepsilon \in (0, 1)$, d be the projected dimension, $\mathcal{C}$ a universal constant, Q, c, A, k, R be as above, x^* be a solution of P and $(\hat{z}, \hat{v})$ be a solution of RD. Then, with probability at least $1 - 4(m + 2k + 1)\,e^{-\mathcal{C}\varepsilon^2 d}$, we have*

$$\mathsf{v}(\mathsf{RP}) \leq \mathsf{v}(\mathsf{P}) \leq \mathsf{v}(\mathsf{RP}) + E,$$

where $E = 3\varepsilon R^2 \|Q\|_F + \varepsilon R \|c\|_2 + \varepsilon \|x^*\|_2 \min(\|A^\top \hat{v}\|_1, \|\hat{v}\|_1)$.

Proof First, by Proposition 2 we have $\mathsf{v}(\mathsf{RP}) \leq \mathsf{v}(\mathsf{P})$. Now let x^* be an optimum of P. Let $u' = Px^*$: we are going to show that u' is approximately optimal for RP. We define the auxiliary problem

$$\mathsf{RP}' \equiv \max_{u \in \mathbb{R}^d} \left\{ u^\top \bar{Q} u + \bar{c}^\top u \mid \bar{A} u \leq b + A(P^\top u' - x^*) \right\}, \tag{28}$$

and its Dorn dual

$$\mathsf{RD}' \equiv \min_{v \geq 0, z \in \mathbb{R}^d} \left\{ -z^\top \bar{Q} z + (b^\top + A(P^\top u' - x^*))v \mid \bar{A}^\top v - \bar{Q} z = \bar{c} \right\}. \tag{29}$$

By construction, u' is feasible for RP′, which implies

$$(u')^\top \bar{Q} u' + \bar{c}^\top u' \leq \mathsf{v}(\mathsf{RP}'). \tag{30}$$

We now bound the left hand side of Eq. (30) from below using Lemmata 1 and 3 and the right hand side from above using Dorn's QP duality, with probability exceeding $1 - 4(2k+1)e^{-\mathcal{C}\varepsilon^2 d}$ due to the application of the union bound to the two lemmata. For the lower bound, we have:

$$\begin{aligned}
(u')^\top \bar{Q} u' + \bar{c}^\top u' &= (x^*)^\top P^\top P Q P^\top P x^* + (Pc)^\top (P x^*) \\
&\geq (x^*)^\top Q x^* + c^\top x^* - 3\varepsilon R^2 \|Q\|_F - \varepsilon R \|c\| \\
&= \mathsf{v}(\mathsf{P}) - 3\varepsilon R^2 \|Q\|_F - \varepsilon R \|c\|.
\end{aligned}$$

For the upper bound, by weak duality we have $\mathsf{v}(\mathsf{RP}') \leq \mathsf{v}(\mathsf{RD}')$. Let $(\hat{z}, \hat{v})$ be an optimum of RD. Since RD′ has the same feasible region as RD, $(\hat{z}, \hat{v})$ is feasible for RD′. Hence

$$\mathsf{v}(\mathsf{RD}') \leq -(\hat{z})^\top \bar{Q} \hat{z} + b^\top \hat{v} + A(P^\top u' - x^*)\hat{v} = \mathsf{v}(\mathsf{RD}) + A(P^\top u' - x^*)\hat{v}.$$

By strong duality, $\mathsf{v}(\mathsf{RD}) = \mathsf{v}(\mathsf{RP})$, hence

$$\mathsf{v}(\mathsf{RD}') \leq \mathsf{v}(\mathsf{RP}) + A(P^\top u' - x^*)\hat{v}.$$

We obtain

$$\mathsf{v}(\mathsf{P}) - E_1 \leq (u')^\top \bar{Q} u' + \bar{c}^\top u' \leq \mathsf{v}(\mathsf{RP}') \leq \mathsf{v}(\mathsf{RD}') \leq \mathsf{v}(\mathsf{RP}) + E_2,$$

where $E_1 = 3\varepsilon R^2 \|Q\|_F + \varepsilon R \|c\|$ and $E_2 = A(P^\top u' - x^*)\hat{v}$, whence

$$\mathsf{v}(\mathsf{RP}) \leq \mathsf{v}(\mathsf{P}) \leq \mathsf{v}(\mathsf{RP}) + E_1 + E_2.$$

We remark that $E_2 = \langle A^\top \hat{v}, P^\top u' - x^* \rangle$. By Hölder's inequality,

$$E_2 \leq \|A^\top \hat{v}\|_1 \|P^\top u' - x^*\|_\infty.$$

By Lemma 2 applied with I_n replacing the symbol A in the lemma, every component of $P^\top u' - x^*$ is bounded above by $\varepsilon \|x^*\|_2$ with probability at least $1 - 4m e^{-\mathcal{C}\varepsilon^2 d}$, hence $E_2 \leq \varepsilon \|x^*\|_2 \|A^\top \hat{v}\|_1$ (∗). We can now re-write E_2 as $\langle \hat{v}, A(P^\top u' - x^*) \rangle$. Again by Hölder's inequality, we have

$$E_2 \leq \|\hat{v}\|_1 \|A(P^\top u' - x^*)\|_\infty.$$

Note that $\|A(P^\top u' - x^*)\|_\infty = \max_i |\langle A_i, P^\top u' - x^* \rangle| \leq \max_i \|A_i\| \|P^\top u' - x^*\|$, where the last inequality is Cauchy–Schwartz. Note also that all the rows of A have unit norm; again by Lemma 2 applied with I_n, every component of $P^\top u' - x^*$ is bounded

above by $\varepsilon \|x^*\|_2$ with probability at least $1 - 4me^{-C\varepsilon^2 d}$, hence $E_2 \leq \varepsilon \|x^*\|_2 \|\hat{v}\|_1$. So we can take the minimum of the two upper bounds for E_2, as claimed. The probability cited in the statement again holds by the union bound. □

Although the bound in Theorem 3 only applies to convex QPs (convexity is key in Dorn's proof [5]), the error appears additively in the bound expression, rather than multiplicatively as in the case of Theorem 2. This is a more satisfactory situation in view of Theorem 1.

4.2 Critique of the error bounds

The bounds derived in Theorem 1, Corollary 1, and Theorem 3 all depend on the error term

$$\hat{E} = R\varepsilon(3R\|Q\|_F + \|c\|_2).$$

Let $\lambda_{\max}$ be the largest eigenvalue of Q and $c_{\max} = \|c\|_\infty$. Then $\|Q\|_F \leq \lambda_{\max}\sqrt{n}$ and $\|c\|_2 \leq c_{\max}\sqrt{n}$. If we suppose that R, $\lambda_{\max}$, and $c_{\max}$ are fixed, then $\hat{E} = O(\varepsilon\sqrt{n})$. If we want $\hat{E}$ to be $O(1)$, we need $\varepsilon = O(\frac{1}{\sqrt{n}})$, which, by the lower bound on d given in the statement of Theorem 1, implies that $d = \Omega(n \ln(m/\delta))$, i.e. the projected dimension should actually be greater than the original dimension for the bounds to be tight. Our proposed projection framework should therefore perform better on instances for which $\|Q\|_F \ll \sqrt{n}$ and $\|c\|_2 \ll \sqrt{n}$.

We now write $\hat{E}$ as a function of $v(P)$. First, we remark that the only purpose of R is to bound the value of $\|x^*\|_2$, where x^* is an optimal solution of P. We can therefore rewrite $\hat{E}$ as follows:

$$\hat{E} = 3\varepsilon \|x^*\|_2^2 \|Q\|_F + \varepsilon \|c\|_2 \|x^*\|_2.$$

Letting θ^* be the angle between x^* and c, we can rewrite the second term as $\varepsilon \|c\|_2 \|x^*\|_2 = \frac{\varepsilon}{\cos\theta^*} c^\top x^*$, as long as $\theta^* \neq \frac{\pi}{2}$. For a general QP, however, we can assume that this holds almost surely. Similarly, let α^* be the angle between $x^*(x^*)^\top$ and Q in the Euclidean space $\mathbb{R}^{n \times n}$ with the scalar product $\langle A, B \rangle = \mathsf{trace}(A^\top B)$. This yields:

$$\cos \alpha^* \|x^*(x^*)^\top\|_F \|Q\|_F = \langle x^*(x^*)^\top, Q \rangle.$$

We remark that $\|\cdot\|_F$ is the norm associated with $\langle \cdot, \cdot \rangle$ in $\mathbb{R}^{n \times n}$. Again, we can assume that, for a general QP, $\alpha^* \neq \frac{\pi}{2}$ almost surely. We also have that $\langle x^*(x^*)^\top, Q \rangle = (x^*)^\top Q x^*$ and $\|x^*(x^*)^\top\|_F = \|x^*\|_2^2$, which yields:

$$\|x^*\|_2^2 \|Q\|_F = \frac{(x^*)^\top Q x^*}{\cos \alpha^*}.$$

Consequently,

$$\hat{E} = \frac{3\varepsilon}{\cos \alpha^*} (x^*)^\top Q x^* + \frac{\varepsilon}{\cos \theta^*} c^\top x^*.$$

For instances where $\cos \alpha^*$ and $\cos \theta^*$ have the same sign, we conclude that

$$\hat{E} \leq 3\varepsilon \max \left(\frac{1}{|\cos \alpha^*|}, \frac{1}{|\cos \theta^*|} \right) \mathsf{v}(P).$$

We observe that, in this case, $\hat{E}$ is now written multiplicatively w.r.t. $\mathsf{v}(P)$, and that the error coefficient $3\varepsilon \max \left(\frac{1}{\cos \alpha^*}, \frac{1}{\cos \theta^*} \right)$ is independent of the dimension.

Although this analysis is only valid if $\operatorname{sgn} \cos \alpha^* = \operatorname{sgn} \cos \theta^*$, and does not take the multiplicative error factor $1/\sigma^2$ into account, it shows that bounds can be slack if the solution (or its Gramian) defines angles close to $\pi/2$ with c (or Q). It also shows that, for families of instances of increasing sizes where these cosines are always close to 1, the error $\hat{E}$ should not increase with size. We note that the algorithms used to randomly generate the instances tested in Sect. 5 do not take α^*, θ^* into account, which could contribute to explain the results trend.

5 Computational experiments

Most papers about RPs are entirely theoretical, although several exceptions exist (e.g. [4,21,24]). RPs are also used in practice in a variety of application settings [3,15, 17,27]. In this section, we discuss computational experiments which showcase the applicability of the techniques presented above.

All tests were carried out on a 4-CPU machine with 64 GB RAM, each CPU of which has 8 cores (Intel Xeon CPU E5-2620 v4@2.10 GHz). We used Python 3.7 as an interface to the QP barrier algorithm from IBM-ILOG CPLEX 12.8 [7] with default configuration. The QP barrier algorithm in CPLEX was always given the origin as a starting point (aside from the experiments in Sect. 5.7.2).

We fixed $\varepsilon = 0.1$ in all experiments, and used the formula $d = \frac{1}{\varepsilon^2} \ln(n)$ to compute the projected dimension. While our theoretical results require d to be bounded below by a function of $\ln(m)$, the fact that m, n are linked by a polynomial relationship in most cases of interest (including those studied in this paper) makes this difference negligible. This choice was made because, after some preliminary tests, it seemed to strike the better compromise between CPU advantage and solution quality.

5.1 Sparse RPs

Although we developed our theory for dense Gaussian RPs, in practice one can decrease computational costs considerably by exploiting sparsity [1,11].

All of the results of this paper actually hold (unchanged) also for *sub-gaussian* RPs [22, Sect. 9.3.1]. After some preliminary testing with various types of dense and sparse sub-gaussian RPs, we elected to use $d \times n$ matrices where each component is

sampled from N(0, $\bar{\sigma}$), where the standard deviation is $\bar{\sigma} = \frac{1}{\sqrt{d\gamma}}$, with some given probability $\gamma \in (0, 1)$, which defines the density of the RP matrix. Since, to the best of our knowledge, there is no proof in the literature that these matrices are valid RPs, we provide one here.

Proposition 3 *Given $n \in \mathbb{N}$ and $d = O(\frac{1}{\varepsilon^2} \ln n)$, the set of $d \times n$ matrices $P = (P_{ij})$ where each P_{ij} is sampled from N(0, $\bar{\sigma}$) with probability γ, and is equal to zero with probability $1 - \gamma$, provides a valid random projection.*

Proof We show that $\sqrt{d}P$ has subgaussian isotropic rows with zero mean, i.e. for each row ρ of $\sqrt{d}P$ we have $\mathbb{E}(\rho^\top \rho) = I_n$. The first property occurs because each component of ρ is either zero or is sampled from N(0, $1/\sqrt{\gamma}$), which has zero mean and has tail bounded by a negative exponential function [22, Ex. 2.5.8]. We now prove isotropy: if $i \neq j$ then $\mathbb{E}(\rho_i \rho_j) = 0$ because the two components are sampled independently. Now observe that any component of $\sqrt{d\gamma}P$ is distributed as B(γ)N(0, 1), where $B(\gamma)$ is a Bernoulli distribution with parameter γ. Finally, if $i = j$ we have

$$\mathbb{E}(\rho_i^2) = \mathbb{E}(\rho_i^2) - \mathbb{E}^2(\rho_i) \quad \text{because } \mathbb{E}(\rho_i) = 0$$
$$= \mathsf{Var}(\rho_i) = \frac{\mathsf{Var}(\mathsf{B}(\gamma)\mathsf{N}(0, 1))}{\gamma}.$$

By independence we have that the variance of the product above is equal to:

$$\mathsf{Var}(\rho_i) = \mathbb{E}[\mathsf{B}(\gamma)]^2 \mathsf{Var}(\mathsf{N}(0, 1)) + \mathbb{E}[\mathsf{N}(0, 1)]^2 \mathsf{Var}(\mathsf{B}(\gamma)) + \mathsf{Var}(\mathsf{B}(\gamma))\mathsf{Var}(\mathsf{N}(0, 1))$$
$$= \gamma^2 + \gamma(1 - \gamma) = \gamma.$$

Hence $\mathbb{E}(\rho_i^2) = 1$ as claimed. Now the result follows by [22, Sect. 9.3.1]. $\qquad\square$

The density of the RP was always fixed at 0.2 in all experiments.

5.2 Random instances

Our first computational test is on randomly generated feasible instances of Eq. (1) with Q negative semidefinite. While our projection technique is independent of the convexity of the objective function, we make this assumption in order to compute guaranteed global maxima in acceptable CPU times for comparison purposes. Our theoretical analysis is also mostly independent of convexity, aside from those bound results for which convexity is explicitly assumed (e.g. Theorem 3).

We consider three sets of random instances: `random`, `pairs`, `cuberot`. In all of them, the $n \times n$ matrix Q defining the quadratic form is generated starting with a negative identity matrix $-I_n$. With probability given by a prescribed density parameter `dens`, we populate the off-diagonal upper-triangular entries with samples from a uniform distribution on $\left[-\frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{n}}\right]$ (the lower-triangular part is copied from the upper one to make Q symmetric). This yields the negative of a random sparse diagonally dominant matrix, which is negative definite. Each component of the vector c defining

Table 1 Number of instances with given number of variables and constraints in each test set

#	n	d	m		
			Random	Pairs	Cuberot
6	1000	691	100	200	2100
6	1000	691	1000	2000	3000
6	2000	760	100	200	4100
6	2000	760	1000	2000	5000
6	3000	801	100	200	6100
6	3000	801	1000	2000	7000
6	4000	829	100	200	8100
6	4000	829	1000	2000	9000

the linear part of the objective function is sampled from a uniform distribution on $[0, 1]$ and then scaled by $\|c\|_2$ to make it a unit vector. The three instance sets differ in the random generation of A, b defining the feasible set $\{x \mid Ax \leq b\}$.

Here are the details about the three generation methods.

1. `Random`: A given number q of vectors A_i are sampled (componentwise) uniformly from $[0, 1]^n$, sparsified so that each vector has density `dens`, and scaled so that their ℓ_2 norm is in $[0.5, 0.6]$; each vector defines a half-space $A_i x \leq \|A_i\|_2^2$ containing the origin [14, Sect. 2], and having a defining hyperplane which is at distance $\|A_i\|_2$ from the origin. Polyhedra sampled this way are not guaranteed to be bounded, nor is their circumscribing sphere radius easy to compute. The number of constraints is $m = q$.
2. `Pairs`: Similar to `random`, but each vector A_i gives rise to two parallel half-spaces: $A_i x \leq \|A_i\|_2^2$ and $A_i x \geq -\|A_i\|_2^2$, which, together, define a split containing the origin. Polyhedra sampled this way are not guaranteed to be bounded, nor is their circumscribing sphere radius easy to compute. The number of constraints is $m = 2q$.
3. `Cuberot`: Similar to `random`, but intersected with a randomly rotated hypercube centered at the origin with half-side length $R/\sqrt{n}$, where R is a given scalar. These polyhedra are guaranteed to be bounded, and have a circumscribing sphere radius equal to R. The number of constraints is $m = n + q$.

All our randomly sampled polyhedra turned out to be bounded in the direction of the objective function.

We generated 48 instances for each of the above types, for:

– Number of variables $n \in \{1000, 2000, 3000, 4000\}$;
– Constraint generation parameter $q \in \{100, 1000\}$;
– Generation density `dens` $\in \{0.1, 0.9\}$;
– Circumscribing sphere radius $R \in \{\frac{1}{2}, 1, 10\}$.

To help give an overall idea of how the number of constraints m changes with the number of variables n, Table 1 reports the number of instances having n variables and m constraints in each instance set.

Table 2 Number of variables and constraints on the portfolio instance set

ID	1	2	3	4	5	6	7	8	9	10	11	12
n	1000	1000	1000	1000	1000	1000	2000	2000	2000	2000	2000	2000
d	691	691	691	691	691	691	760	760	760	760	760	760
m	2102	2302	2502	2602	2702	2902	4102	4302	4502	4602	4702	4902
ID	13	14	15	16	17	18	19	20	21	22	23	24
n	3000	3000	3000	3000	3000	3000	4000	4000	4000	4000	4000	4000
d	801	801	801	801	801	801	829	829	829	829	829	829
m	6102	6302	6502	6602	6702	6902	8102	8302	8502	8602	8702	8902

5.3 Portfolio optimization instances

We consider a realistic variant of the classic portfolio optimization problem [13]. Its objective function is a scalarized version of risk minimization and return maximization. Short-selling is allowed by considering decision variables in $[-1, 1]$ (instead of the more standard $[0, 1]$). Furthermore, we limit the investment to q given "investment areas" (e.g. start-ups, tech companies, emerging countries, micro-credit, and so on), defined as sets of shares $T_1, \ldots, T_q$, each of which may not be allocated more than $\tau_1, \ldots, \tau_q$ fraction of budget, yielding constraints:

$$\forall p \leq q \quad \sum_{j \in T_p} x_j \leq \tau_p. \tag{31}$$

The unbiased estimator for a covariance matrix is $\frac{1}{n-1} \sum_{j \leq n} (x_j - \bar{x})(x_j - \bar{x})^\top$ (where $\bar{x}$ is the sample mean of the x_j), i.e. of a random variable $Z = X - \mathbb{E}(X)$ with zero mean and samples $z_j = x_j - \bar{x}$. This implies that $\sum_j z_j = 0$, which means that the rank of the estimator is $n - 1$. We generate random $n \times (n-1)$ matrices Y with components sampled uniformly from $[0, 1]$, and corresponding covariance matrices as $Q = YY^\top$ with rank $n - 1$. The mean returns are random vectors in $[0, 1]^n$. The constraints are: $-1 \leq x \leq 1$, $-1 \leq \sum_j x_j \leq 1$, and Eq. (31).

We generated 24 portfolio instances, for:

– Number of variables $n \in \{1000, 2000, 3000, 4000\}$;
– Investment areas $q \in \{100, 300, 500, 600, 700, 900\}$.

Table 2 below gives the number of variables and constraints for our instance set.

5.4 Results

We discuss results for QPs over random polytopes and portfolio instances separately. Our fundamental performance measures are ratios.

The objective function ratio measures the error between the optimal objective function value $f^* = \mathsf{v}(\mathsf{P})$ of the original problem P and the optimal objective function

Table 3 Aggregate results for QPs over random polytopes

	All			Random		
	r	c	CPU$_P$	r	c	CPU$_P$
avg	0.625	0.44	176.8	0.659	0.47	42.58
stdev	0.197	0.39	310.26	0.144	0.49	43.11
min	0.251	0.07	2.35	0.402	0.07	2.35
max	1.593	1.93	2082.39	0.813	1.93	186.15
	Pairs			Cuberot		
	r	c	CPU$_P$	r	c	CPU$_P$
avg	0.634	0.43	60.42	0.582	0.43	427.39
stdev	0.174	0.36	59.49	0.249	0.31	434.78
min	0.251	0.07	3.05	0.266	0.09	18.82
max	0.944	1.32	259.07	1.593	1.33	2082.39

value $\bar{f} = \mathsf{v}(\mathsf{RP})$ of the projected problem. This error is scaled so it is always greater than or equal to zero:

$$\mathsf{r} = \frac{|f^* - \bar{f}|}{|f^*|}.$$

Clearly, better performances yield r values closer to zero.

The CPU ratio $\mathsf{c} = \mathsf{CPU_{RP}}/\mathsf{CPU_P}$ measures the ratio between the time taken to solve the projected instance RP, and the time taken to solve the original one P. Clearly, better performances yield c values closer to zero.

By "solve", we mean the following:

– In the case of the original instances, the time taken to pass it to CPLEX, and solve it;
– In the case of the projected instances, the time taken to project the instance data, pass it to CPLEX, solve it, and compute the retrieved solution (see Proposition 1).

5.4.1 Random QPs

We first report aggregate results for r, c, CPU$_P$ in Table 3.

We immediately note that the cuberot instances globally take more time than the other classes: this is due to their large number of constraints.

A finer analysis shows that $f^* < \bar{f}$ for two of the cuberot instances with $R = 0.5$, due to a solver failure in computing f^* (this is further discussed in Sect. 5.7.1 below). If we require $R \geq 1$ we obtain the results (for r and c only) in Table 4.

We also observed that sparsity does not significantly impact the r statistic, but lowers c by around 20% on average.

The results in Tables 3 and 4 are very good w.r.t. the CPU time (see Fig. 2). The maximum values of $\mathsf{c} > 1$ (which occur when the time taken to solve the original problem is smaller than the time taken to solve the projected one) are limited to the

**Table 4** Aggregate results for r, c for QPs over random polytopes with $R \geq 1$

	All		Random		Pairs		Cuberot	
	r	c	r	c	r	c	r	c
avg	0.606	0.44	0.660	0.47	0.604	0.43	0.555	0.42
stdev	0.161	0.39	0.144	0.48	0.169	0.36	0.150	0.29
min	0.251	0.07	0.402	0.07	0.251	0.07	0.269	0.09
max	0.902	1.80	0.813	1.80	0.902	1.32	0.726	1.23
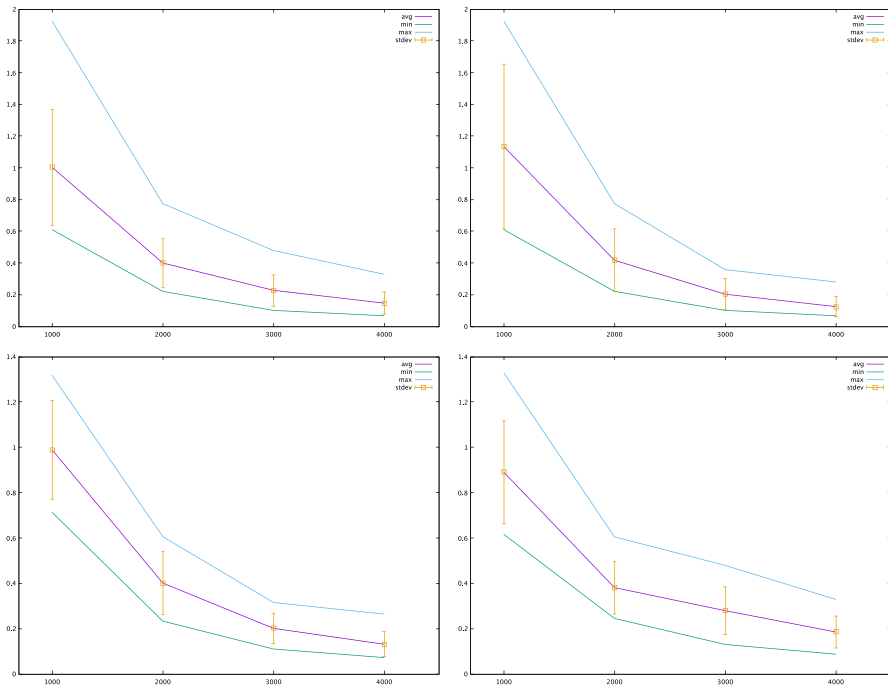
Fig. 2 Mean, standard deviation, minimum, maximum values of c versus n, for all instances (top left), random (top right), pairs (bottom left), cuberot (bottom right)

smallest instances, i.e. some instances with $n = 1000$, where the projection does not significantly decrease the number of variables. If restricted to instances with $n \geq 2000$, the maximum over c is 0.77. The objective function values of the projected problems, however, are not very close to those of the original problem.

We now refine the analysis of r. We plot its mean, standard deviation, minimum and maximum versus n in Fig. 3.

The mean of r appears to increase logarithmically with n. Since the projected problem RP has d variables, and $d = O(\ln(n))$, this seems to imply that the objective function error for this random projection increases with the number d of variables of the projected problem. For random instances the standard deviation is small. For pairs instance it is large and regular. For cuberot instances it is large and irregular. We
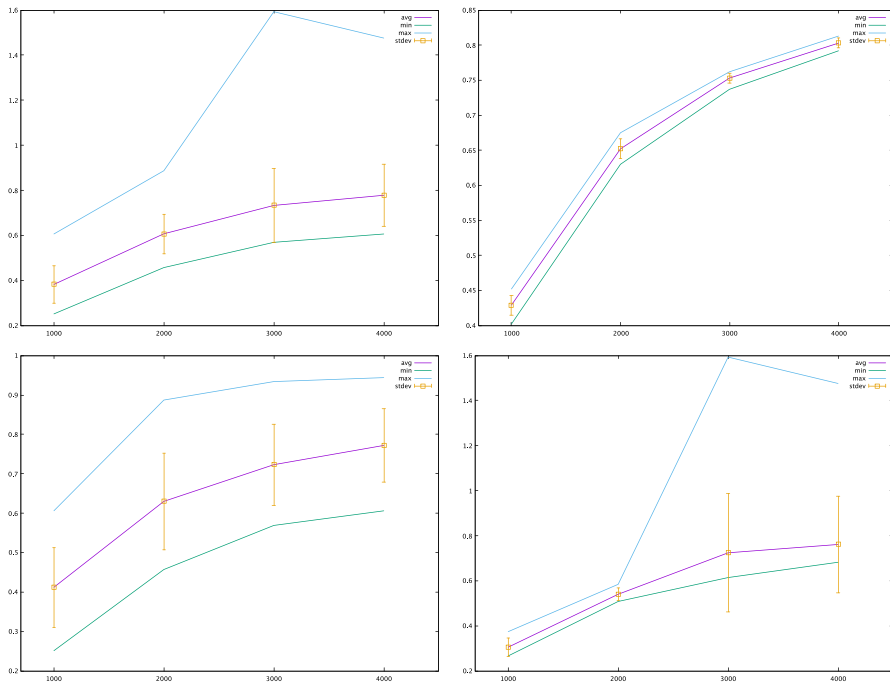
Fig. 3 Mean, standard deviation, minimum, maximum values of r versus n, for all instances (top left), `random` (top right), `pairs` (bottom left), `cuberot` (bottom right)

Table 5 Aggregate results for portfolio instances

|	r	c	CPU$_P$
avg	0.478	0.83	66.78
stdev	0.242	0.42	54.29
min	0.065	0.30	7.36
max	0.744	1.66	157.74

remark that this logarithmic behaviour is not due to the scaling of r: a similar behaviour can be noted also for plots of $|f^* - \bar{f}|$. In general, the type of concentration of measure phenomena that are more helpful are those where the error decreases as the number of variables increases (e.g. [24]). Unfortunately, it seems that the application of RPs to QPs does not exhibit this property; some possible theoretical reasons are provided in Sect. 4.2.

5.5 Portfolio optimization with short sells

We first report aggregate results for r, c in Table 5. Again, the results in Table 5 are good w.r.t. c but not r. Again, we refine the analysis of c and r: we plot their mean, standard deviation, minimum and maximum versus n in Figs. 4 and 5.
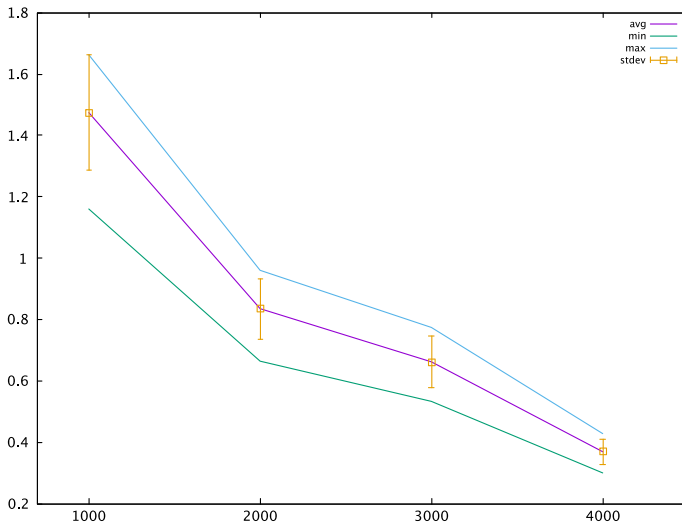
Fig. 4 Mean, standard deviation, minimum, maximum values of c versus n for portfolio optimization instances
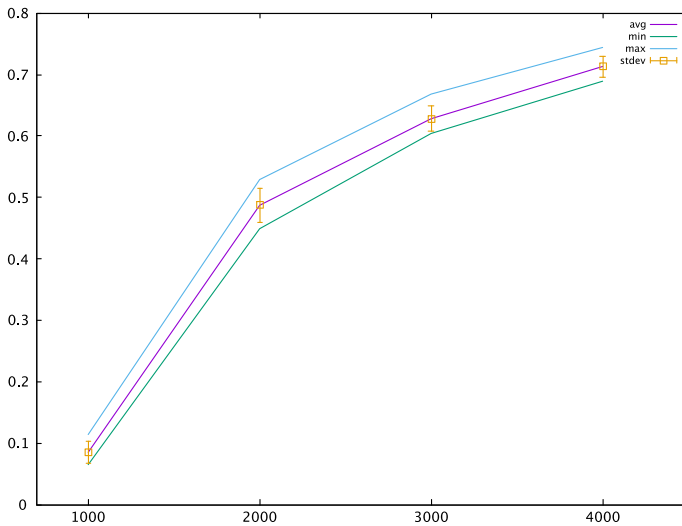
Fig. 5 Mean, standard deviation, minimum, maximum values of r versus n for portfolio optimization instances

We observe the same behaviour as with random QPs: r appears to be proportional to the number of variables of the projected problem (see Sect. 4.2).

5.6 Comparison with the results from [23]

In terms of objective function value, the results in Tables 3 and 4 are worse than those obtained in [23]. Although in [23] we do not solve the same problem, perform the same analysis, use the same random generation techniques, or report the same metric, a short comparative discussion is pertinent insofar as in [23] we also discussed results for randomly projected QPs over polyhedra (and a single unit ball constraint), and the reported results were better.

The main difference is that the results from [23] displayed an infeasibility error w.r.t. the norm constraint. When the retrieved optima were scaled back to be feasible w.r.t. the norm constraint, the objective function values corresponding to the scaled retrieved solutions had errors of similar magnitudes to those reported here.

5.7 Practical usefulness

Based on our computational experiments, our techniques are shown to provide an effective way to compute a good feasible point. We see at least two scenarios where our proposed techniques might be helpful: (i) when the QP at hand is so large that no off-the-shelf solver will be able to provide any output at all; (ii) in order to provide a good starting point for further improvement. Since the time difference between solving the projected and the original problem increases with size (in favor of the projected problem), there is increasingly more time, w.r.t. solving the original problem, to improve the retrieved solution from the projected problem.

5.7.1 Finding a feasible solution to the original QP

The CPLEX barrier solver failed to solve two among the random instances we generated. The two instances are:

1. Cuberot, $n = 3000$, $m = 7000$, dens $= 0.9$, $R = 0.5$;
2. Cuberot, $n = 4000$, $m = 9000$, dens $= 0.9$, $R = 0.5$.

These are challenging instances because of their large size: the first has around 3.38M nonzeroes in the lower triangle of Q and around 50M nonzeroes in the lower triangle of $AA^\top$. The second instance has around 5.38M nonzeroes in the lower triangle of Q and around 84M nonzeroes in the lower triangle of $AA^\top$. In both of these cases, the solver reported primal–dual infeasibility. On the other hand, the projected instances were solved to optimality and produced a feasible point for the original instance (by Proposition 2).

5.7.2 Speeding up the solver by a good feasible starting point

Some improvement results (the scenario (ii) above) were obtained by running CPLEX on the original problem using the retrieved solution of the projected problem as starting point, and a slightly relaxed convergence criterion: for the largest and densest instances we were able to derive the optimal solution this way (projection then improvement)

Table 6 Performance of projection, solution and improvement

Instance set	n	m	R	c	$c_{0.01}$
Cuberot	4000	8100	0.5	0.87	0.63
Cuberot	4000	8100	1.0	0.92	0.78
Cuberot	4000	8100	10.0	0.89	0.78
Cuberot	4000	9000	0.5	0.93	0.80
Cuberot	4000	9000	1.0	0.92	0.77
Cuberot	4000	9000	10.0	0.80	0.56
Pairs	4000	2000	0.5	0.76	0.68
Pairs	4000	2000	1.0	0.72	0.63
Pairs	4000	2000	10.0	0.79	0.71
Random	4000	1000	0.5	0.77	0.70
Random	4000	1000	1.0	0.66	0.55
Random	4000	1000	10.0	0.78	0.72

in globally less time than with a straight CPLEX solve of the original problem. The results, in terms of $c = \text{CPU}_{\text{RP}}/\text{CPU}_{\text{P}}$ (where CPU_{RP} includes the construction of the projected formulation, its solution, the retrieval of the corresponding feasible point in the original formulation, and the solution of the original formulation from this point) are given in Table 6.

We also ran tests using the formula $d = \frac{\mathcal{C}}{\varepsilon^2} \ln(n)$ for different values of $\mathcal{C}$ between 0.01 and 1.3. The trend shows that the smaller d is, the better. The results of c for $\mathcal{C} = 0.01$ are shown in the corresponding column (labelled $c_{0.01}$) of Table 6.

We do not report quality statistics because the improvement phase always found the same optimum as the original problem (with the same optimal objective function value), which is unsurprising, since the improvement phase actually solves the original problem.

5.8 Practical limitations

We would also like to warn readers about two important features of typical QPs which adversely impact our methodology. Many QPs have a quadratic form based on Q being a diagonal matrix. Typically such matrices are extremely sparse, whereas our RP techniques would yield smaller, but denser, projected matrices $\bar{Q}$, which make the projected QPs harder to solve. Secondly, many QPs have variable bounds that most solvers can deal with directly. When projected, however, simple bounds might turn into dense linear constraints, which impact most QP solvers adversely: and we have not yet found a QP projection technique that can treat variable bounds satisfactorily. Thus, many QPs might fail to benefit from our proposed techniques because their original form is amenable to solver simplifications which their projected forms are not.

Table 7 Aggregate results for r, c for QPs over random polytopes (left) and portfolio optimization problems (right) with different scaling for the quadratic ter

	All		Random		Pairs		Cuberot			Portfolio	
	r	c	r	c	r	c	r	c		r	c
avg	0.302	0.44	0.094	0.48	0.268	0.41	0.543	0.42	Mean	0.273	0.86
stdev	0.282	0.39	0.059	0.49	0.256	0.35	0.258	0.31	SD	0.056	0.41
min	0.006	0.06	0.006	0.06	0.052	0.07	0.237	0.06	Min	0.121	0.33
max	1.596	1.96	0.244	1.96	0.934	1.34	1.596	1.31	Max	0.349	1.69

5.9 Different scalings: a conjecture

Based on the intuitive idea that quadratic terms contribute larger errors than linear ones, we propose to use different scalings for the RP in the quadratic term and in the linear term. Consider an RP $P \sim \mathsf{N}(0, \frac{1}{\sqrt{d}})^{dn}$ and the projected objective function

$$\frac{d}{n} x^\top P^\top P Q P^\top P x + c^\top P^\top P x.$$

We set $u = Px$, yielding

$$\frac{d}{n} u^\top P Q P^\top u + c^\top P^\top u.$$

Motivated by [28, Cor. 7], which suggests that $\frac{1}{n} P P^\top \approx I_d$ with high probability, we tried to prove that $\frac{1}{n} u^\top P Q P^\top u$ is somehow "close" to $x^\top Q x$, and, so far, failed. We, however, went ahead and computed aggregated results for random QPs (Sect. 5.2) and portfolio instances (Sect. 5.3), given in Table 7. The plot of mean, standard deviation, minimum and maximum versus n for QPs over random polytopes is given in Fig. 6. The corresponding plot for portfolio optimization is given in Fig. 7. It is clear that, with the exception of the cuberot instance class, weighing the quadratic term by $\frac{d}{n}$ is beneficial in practice for the application of RPs to QPs. In our attempts to derive a corresponding analysis we also tried a few other different weights, but these failed to deliver results as promising as those obtained with $\frac{d}{n}$. Currently, we do not know if these improved results have occurred by chance, or because of a deeper reason. We leave this as an open question for future research.

6 Conclusion

In this paper we discussed the application of RPs to QPs. It turns out that it is possible to solve randomly projected QPs and obtain optima whose value approximates the optimal value of the original problem, in far less time. Theoretically, the error bounds for the approximation are not unconditionally good, unfortunately. Empirically, the scaled objective value difference between original and projected problem shows a
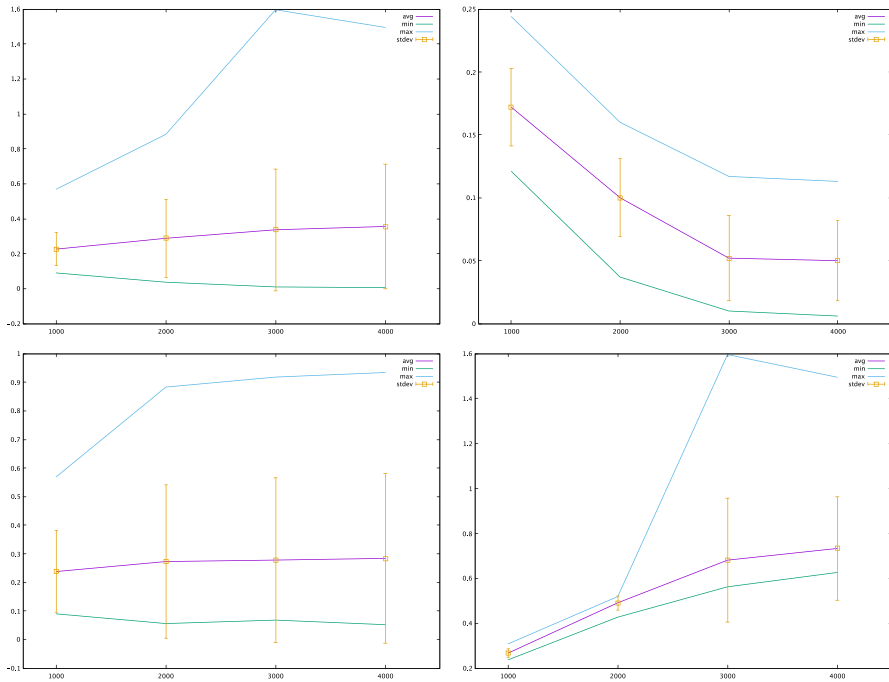
Fig. 6 Mean, standard deviation, minimum, maximum values of r versus n, for all instances (top left), `random` (top right), `pairs` (bottom left), `cuberot` (bottom right), using a different scaling for the quadratic term
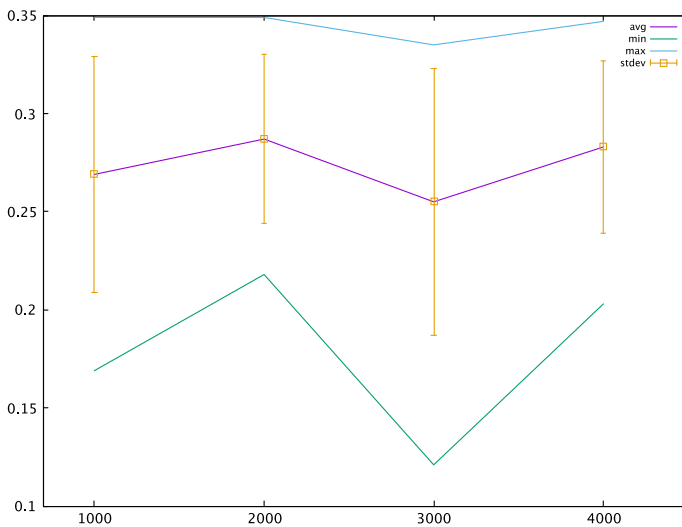
Fig. 7 Mean, standard deviation, minimum, maximum values of r versus n

logarithmic increase in function of the number of variables of the original problem, which is also not good. The CPU time, however, shows a logarithmic decrease. The technique is, however, very useful in cases where the instance size is so large that the solver fails. In such cases, we show that we can successfully compute good feasible points, and use them as starting point to actually solve the instance to optimality.

References

1. Achlioptas, D.: Database-friendly random projections: Johnson–Lindenstrauss with binary coins. J. Comput. Syst. Sci. **66**, 671–687 (2003)
2. Ailon, N., Chazelle, B.: Approximate nearest neighbors and fast Johnson–Lindenstrauss lemma. In: Proceedings of the Symposium on the Theory Of Computing, Volume '06 of STOC. ACM, Seattle (2006)
3. Boutsidis, C., Zouzias, A., Drineas, P.: Random projections for k-means clustering. In: Advances in Neural Information Processing Systems, NIPS. NIPS Foundation, La Jolla, pp. 298–306 (2010)
4. Dasgupta, S.: Experiments with random projection. In: Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence, pp. 143–151. Morgan Kaufman, San Francisco (2000)
5. Dorn, W.: Duality in quadratic programming. Q. Appl. Math. **18**(2), 155–162 (1960)
6. Gould, N., Toint, P.: A quadratic programming bibliography. In: Technical Report 2000–2001. RAL Numerical Analysis Group (2001)
7. IBM: ILOG CPLEX 12.8 User's Manual. IBM (2017)
8. Indyk, P., Motwani, R.: Approximate nearest neighbors: towards removing the curse of dimensionality. In: Proceedings of the Symposium on the Theory Of Computing, Volume 30 of STOC, pp. 604–613. ACM, New York (1998)
9. Indyk, P., Naor, A.: Nearest neighbor preserving embeddings. ACM Trans. Algorithms **3**(3), 31 (2007)
10. Johnson, W., Lindenstrauss, J.: Extensions of Lipschitz mappings into a Hilbert space. In: Hedlund, G. (ed.) Conference in Modern Analysis and Probability. Contemporary Mathematics, vol. 26, pp. 189–206. AMS, Providence, RI (1984)
11. Kane, D., Nelson, J.: Sparser Johnson–Lindenstrauss transforms. J. ACM **61**(1), 4 (2014)
12. Liberti, L., Vu, K.: Barvinok's naive algorithm in distance geometry. Oper. Res. Lett. **46**, 476–481 (2018)
13. Markowitz, H.: Portfolio selection. J. Finance **7**(1), 77–91 (1952)
14. May, J., Smith, R.: Random polytopes: their definition, generation and aggregate properties. Math. Program. **24**, 39–54 (1982)
15. Paul, S., Boutsidis, C., Magdon-Ismail, M., Drineas, P.: Random projections for linear support vector machines. ACM Trans Knowl Discov Data **8**(4), 22:1–22:25 (2014)
16. Pilanci, M., Wainwright, M.: Randomized sketches of convex programs with sharp guarantees. In: International Symposium on Information Theory (ISIT), pp. 921–925. IEEE, Piscataway (2014)
17. Pilanci, M., Wainwright, M.: Newton sketch: a linear time optimization algorithm with linear-quadratic convergence. SIAM J. Optim. **27**(1), 205–245 (2017)
18. Shim, J.: A survey of quadratic programming applications to business and economics. Int. J. Syst. Sci. **14**(1), 105–115 (1983)
19. Vavasis, S.: Quadratic programming is in NP. Inf. Process. Lett. **36**, 73–77 (1990)
20. Vempala, S.: The random projection method. In: Number 65 in DIMACS Series in Discrete Mathematics and Theoretical Computer Science. AMS, Providence, RI (2004)
21. Venkatasubramanian, S., Wang, Q.: The Johnson–Lindenstrauss transform: an empirical study. In: Algorithm Engineering and Experiments, Volume 13 of ALENEX, pp. 164–173. SIAM, Providence, RI (2011)
22. Vershynin, R.: High-dimensional Probability. CUP, Cambridge (2018)

23. Vu, K., Poirion, P.-L., D'Ambrosio, C., Liberti, L.: Random projections for quadratic programs over a Euclidean ball. In: Lodi, A., et al. (eds.) Integer Programming and Combinatorial Optimization (IPCO). LNCS, vol. 11480, pp. 442–452. Springer, New York (2019)
24. Vu, K., Poirion, P.-L., Liberti, L.: Random projections for linear programming. Math. Oper. Res. **43**(4), 1051–1071 (2018)
25. Vu, K., Poirion, P.-L., Liberti, L.: Gaussian random projections for Euclidean membership problems. Discrete Appl. Math. **253**, 93–102 (2019)
26. Woodruff, D.: Sketching as a tool for linear algebra. Found. Trends Theor. Comput. Sci. **10**(1–2), 1–157 (2014)
27. Yang, J., Meng, X., Mahoney, M.: Quantile regression for large-scale applications. SIAM J. Sci. Comput. **36**(5), S78–S110 (2014)
28. Zhang, L., Mahdavi, M., Jin, R., Yang, T., Zhu, S.: Recovering the optimal solution by dual random projection. In: Shalev-Shwartz S., Steinwart I. (eds) Conference on Learning Theory (COLT), Volume 30 of Proceedings of Machine Learning Research, pp. 135–157. ⟨jmlr.org⟩ (2013)