

Random projections for quadratic programs over a Euclidean ball^{*}

Ky Vu¹, Pierre-Louis Poirion², Claudia D’Ambrosio³, and Leo Liberti³

¹ Dept. of Mathematics, FPT University, Hanoi, Vietnam, vukhacky@gmail.com

² RIKEN Center for Advanced Intelligence Project, Tokyo, Japan,
pierre-louis.poirion@riken.jp

³ CNRS LIX Ecole Polytechnique 91128 Palaiseau, France,
{dambrosio,liberti}@lix.polytechnique.fr

Abstract. Random projections are used as dimensional reduction techniques in many situations. They project a set of points in a high dimensional space to a lower dimensional one while approximately preserving all pairwise Euclidean distances. Usually, random projections are applied to numerical data. In this paper, however, we present a successful application of random projections to quadratic programming problems subject to polyhedral and a Euclidean ball constraint. We derive approximate feasibility and optimality results for the lower dimensional problem. We then show the practical usefulness of this idea on many random instances, as well as on two portfolio optimization instances with over 25M nonzeros in the (quadratic) risk term.

1 Introduction

In this paper we show that Random Projections (RP) can be applied to Quadratic Programming (QP) problems subject to linear inequality constraints and a single Euclidean ball constraint. We consider the following pair of QP formulations:

$$\left. \begin{array}{l} \max_y y^\top \tilde{Q}y + \tilde{c}^\top y \\ \tilde{A}y \leq \tilde{b} \\ \|y\|_2 \leq R, \end{array} \right\} \quad (1) \qquad \left. \begin{array}{l} \max_x x^\top Qx + c^\top x \\ Ax \leq b \\ \|x\|_2 \leq 1. \end{array} \right\} \quad (2)$$

In Eq. (1), y is a vector of n decision variables, \tilde{Q} is a symmetric $n \times n$ matrix, $\tilde{c} \in \mathbb{R}^n$, \tilde{A} is $m \times n$, $\tilde{b} \in \mathbb{R}^m$, and R is a positive scalar. We also assume that $\tilde{A}x \leq \tilde{b}$ defines a full dimensional polyhedron and that $\tilde{b} \geq 0$ (this can be relaxed by translation if a feasible point for $\tilde{A}x \leq \tilde{b}$ is known). No assumption is made on \tilde{Q} . Eq. (2) is a scaled version of Eq. (1), where $Q = R^2\tilde{Q}$, $c = R\tilde{c}$, $A = \tilde{A}/\mu$ (where $\mu = \max_j \|\tilde{A}_j\|_2$ and \tilde{A}_j is the j -th column of \tilde{A}), and $b = \tilde{b}/(R\mu)$. Given

^{*} This paper has received funding from the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement n. 764759 “MINOA”.

a solution x^* of Eq. (2), then $y^* = Rx^*$ is a solution of Eq. (1). Note that all the columns of A are vectors of norm ≤ 1 .

QP is now a ripe field with many applications (e.g., portfolio optimization, constrained linear regression, stable set problem, maximum cut and many more). The significance of the ball constraint is technical, but it could simply be interpreted to mean “bounded”, since for all bounded QPs we can find a large enough R (in Eq. (1)) so that all solutions fall within a ball of radius R . In practice, however, if R is too large it might lead to ill scaling of Eq. (2). Note, however, that Eq. (1) is interesting in its own right as it is the formulation of the well-known *trust region subproblem*.

If we assume that all the data are rational, then the decision version of Eq. (2) without the ball constraint is **NP**-complete [9]. Moreover, by [10, 11], the decision version of Eq. (2) without the polyhedral constraints is in **P** (and hence also in **NP**). For Eq. (2), one of the following applies: (i) some of the linear inequalities are active at the optimum; (ii) the ball inequality is active at the optimum; (iii) a combination of (i) and (ii); (iv) the optimum is unconstrained. In the first two cases the results in [9–11] apply, and the problem is in **NP**. Case (iii) falls in both of the first two categories, and the problem is still in **NP**. For case (iv) we can tell apart optimality vs. unboundedness by testing whether Q has negative eigenvalues or not [10]. Hence, the decision versions of Eq. (2) and Eq. (1) are in **NP**.

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. (2). Specifically, we will define a projected version of Eq. (2) and prove that it is likely to have approximately the same optima as the original QP. We also perform a computational verification of our claim and show that the theoretical results, which are asymptotic in nature, also apply in practice.

RPs are usually applied to numerical data in view of speeding up algorithms which are essentially based on Euclidean distances, such as k-means or k-nearest neighbours. Since RPs ensure approximations of Euclidean distances by definition, it is perhaps not so surprising that they should work well in those settings. The focus of the present work is the much more counter-intuitive statement that a Mathematical Programming formulation might be approximately invariant (as regards feasibility and optimality) w.r.t. randomly projecting the input parameters. Similarly in spirit to our previous work on Linear Programming [13], but using a different projection and proof techniques, the results of this paper are independent of any solution algorithm, and largely independent of Euclidean distances (barring the ℓ_2 ball bounding the feasible region, which is applied to decision variables rather than data). 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 [15, 8]: two assumptions we do not make.

The rest of this paper is organized as follows. In Sect. 2 we define RPs and the projected QP. In Sect. 3 we introduce some theoretical results about random projections. In Sect. 4 we prove the main theorems about RPs applied to QP. In Sect. 5 we discuss computational results.

2 Definitions

RPs are simple but powerful tools for dimension reduction [15, 8, 12, 13, 4]. 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 Gaussian $\mathcal{N}(0, 1)$, 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. In particular, let $P \in \mathbb{R}^{d \times n}$ be a RP, e.g. sample every component of P from $\mathcal{N}(0, 1/\sqrt{d})$. Then, for any $x \in \mathbb{R}^n$ and $\varepsilon \in (0, 1)$, we have

$$\text{Prob} \left[(1 - \varepsilon) \|x\|_2^2 \leq \|Px\|_2^2 \leq (1 + \varepsilon) \|x\|_2^2 \right] \geq 1 - 2e^{-\mathcal{C}\varepsilon^2 d}, \quad (3)$$

where \mathcal{C} is a *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* [2]. It states that, for any $\varepsilon \in (0, 1)$ and for any finite set $X \subseteq \mathbb{R}^n$, there is a mapping $F : \mathbb{R}^n \rightarrow \mathbb{R}^d$, in which $d = O(\frac{\ln |X|}{\varepsilon^2})$, such that

$$\forall x, y \in X \quad (1 - \varepsilon) \|x - y\|_2^2 \leq \|F(x) - F(y)\|_2^2 \leq (1 + \varepsilon) \|x - y\|_2^2.$$

Such a mapping F can be realized as 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 the following, all norm symbols $\|\cdot\|$ will be assumed to refer to the ℓ_2 norm $\|\cdot\|_2$. We sample our RPs from Gaussian ensembles (in practice, we also specify their density, see Sect. 5).

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*:

$$\max \{x^\top (P^\top P Q P^\top P) x + c^\top P^\top P x \mid AP^\top P x \leq b, \|Px\| \leq 1\}.$$

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

$$\max_{u \in \text{Im}(P)} \{u^\top \bar{Q} u + \bar{c}^\top u \mid \bar{A} u \leq b, \|u\| \leq 1\}, \quad (4)$$

where $\text{Im}(P)$ is the image space generated by 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 $u \in \text{Im}(P)$ and study the smaller dimensional problem:

$$\max_{u \in \mathbb{R}^d} \{u^\top \bar{Q}u + \bar{c}^\top u \mid \bar{A}u \leq b, \|u\| \leq 1\}, \quad (5)$$

where u ranges in \mathbb{R}^d . As we will show later, Eq. (5) yields a good approximate solution of Eq. (2) with high probability.

3 Some properties of random projections

It is known that singular values of random matrices often concentrate around their means. In the case when the RP is sampled from Gaussian ensembles, this phenomenon is well-understood due to many current research efforts. The following lemma, which is proved in [16], uses this phenomenon to show that, when $P \in \mathbb{R}^{d \times n}$ is a Gaussian random matrix (with the number of row significantly smaller than the number of columns), then PP^\top is very close to an identity matrix. This gives an intuitive explanation as to why Eq. (5) has desirable approximate properties w.r.t. Eq. (2).

Lemma 3.1 ([16]). *Let $P \in \mathbb{R}^{d \times n}$ be a RP. Then for any $\delta > 0$ and $0 < \varepsilon < \frac{1}{2}$, with probability at least $1 - \delta$, we have $\|PP^\top - I\|_2 \leq \varepsilon$ provided that*

$$n \geq (d + 1) \ln(2d/\delta) / (\mathcal{C}_1 \varepsilon^2), \quad (6)$$

where $\|\cdot\|_2$ is the spectral norm of the matrix and $\mathcal{C}_1 > \frac{1}{4}$ is some universal constant.

This lemma also tells us that, when we go from low to high dimensions, with high probability we can ensure that the norms of all the points endure small distortions. Indeed, for any vector $u \in \mathbb{R}^d$, then

$$\|P^\top u\|^2 - \|u\|^2 = \langle P^\top u, P^\top u \rangle - \langle u, u \rangle = \langle (PP^\top - I)u, u \rangle \in [-\varepsilon\|u\|^2, \varepsilon\|u\|^2],$$

due to the Cauchy-Schwarz inequality. Moreover, it implies that $\|P^\top\|_2 \leq (1 + \varepsilon)$ with probability at least $1 - \delta$.

Condition (6) is not difficult to satisfy in practice, since d is often very small compared to n . On the other hand, n should be large enough to dominate the effect of $\frac{1}{\varepsilon^2}$.

Lemma 3.2. *Let $P \in \mathbb{R}^{d \times n}$ be a RP satisfying Eq. (3) and let $0 < \varepsilon < 1$. Then there is a universal constant \mathcal{C}_0 such that the following statements hold.*

- (i) *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 $1 - 4e^{-\mathcal{C}_0 \varepsilon^2 d}$.*

- (ii) Let $\mathbf{1}$ be the all-one vector. 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 $1 - 4m e^{-C_0 \varepsilon^2 d}$.
- (iii) For any two vectors $x, y \in \mathbb{R}^n$ and a square matrix $Q \in \mathbb{R}^{n \times n}$, then with probability at least $1 - 8k e^{-C_0 \varepsilon^2 d}$, we have:

$$x^\top Qy - 3\varepsilon \|x\| \|y\| \|Q\|_* \leq x^\top P^\top P Q P^\top P y \leq x^\top Qy + 3\varepsilon \|x\| \|y\| \|Q\|_*,$$

in which $\|Q\|_*$ is the nuclear norm of Q and k is the rank of Q .

4 Approximate optimality

We now prove that the objective of the quadratic problem in Eq. (2) is approximately preserved under RPs. To do so, we study the relations between this and two other problems:

$$\begin{aligned} (\text{QP}_\varepsilon^-) \quad & \max\{u^\top P Q P^\top u + (Pc)^\top u \mid AP^\top u \leq b, \quad \|u\| \leq 1 - \varepsilon, u \in \mathbb{R}^d\} \\ (\text{QP}_\varepsilon^+) \quad & \max\{u^\top P Q P^\top u + (Pc)^\top u \mid AP^\top u \leq b + \varepsilon, \|u\| \leq 1 + \varepsilon, u \in \mathbb{R}^d\}. \end{aligned}$$

We first state the following feasibility result.

Theorem 4.1. *Let $P \in \mathbb{R}^{d \times n}$ be a RP. Let $\delta \in (0, 1)$. Assume further that Eq. (6) holds for some universal constant $C_1 > \frac{1}{4}$. Then with probability at least $1 - \delta$, for any feasible solution u of the projected problem $(\text{QP}_\varepsilon^-)$, $P^\top u$ is also feasible for the original problem in Eq. (2).*

We remark the following universal property of Theorem 4.1: with a fixed probability, feasibility holds for all vectors u (instead of a given vector).

Proof. Let C_1 be as in Lemma 3.1. Let u be any feasible solution for the projected problem $(\text{QP}_\varepsilon^-)$ and take $\hat{x} = P^\top u$. Then we have $A\hat{x} = AP^\top u \leq b$ and

$$\|P^\top u\|^2 = \langle P^\top u, P^\top u \rangle = \langle u, u \rangle + \langle (PP^\top - I)u, u \rangle \leq (1 + \varepsilon) \|u\|^2$$

with probability at least $1 - \delta$ (by Lemma 3.1). This implies that $\|\hat{x}\| \leq (1 + \varepsilon/2) \|u\|$; and since $\|u\| \leq 1 - \varepsilon$, we have $\|\hat{x}\| \leq (1 + \varepsilon/2)(1 - \varepsilon) < 1$ with probability at least $1 - \delta$, which proves the theorem. \square

Let u_ε^- and u_ε^+ be optimal solutions for these two problems, respectively. Denote by $x_\varepsilon^- = P^\top u_\varepsilon^-$ and $x_\varepsilon^+ = P^\top u_\varepsilon^+$. Let x^* be an optimal solution for the original problem in Eq. (2). We will bound $x^{*\top} Q x^* + c^\top x^*$ between $x_\varepsilon^{-\top} Q x_\varepsilon^- + c^\top x_\varepsilon^-$ and $x_\varepsilon^{+\top} Q x_\varepsilon^+ + c^\top x_\varepsilon^+$, the two values that are expected to be approximately close to each other.

Theorem 4.2. *Let $P \in \mathbb{R}^{d \times n}$ be a RP, and let $\delta \in (0, 1)$. Let x^* be an optimal solution for the original problem Eq. (2). Then there are universal constants $C_0 > 1$ and $C_1 > \frac{1}{4}$ such that, if $d \geq \ln(m/\delta)/(C_0 \varepsilon^2)$ and Eq. (6) are satisfied, we will have the following two statements. (i) With probability at least $1 - \delta$, the solution x_ε^- is feasible for the original problem Eq. (2). (ii) With probability at least $1 - \delta$,*

$$x_\varepsilon^{-\top} Q x_\varepsilon^- + c^\top x_\varepsilon^- \leq x^{*\top} Q x^* + c^\top x^* \leq x_\varepsilon^{+\top} Q x_\varepsilon^+ + c^\top x_\varepsilon^+ + 3\varepsilon \|Q\|_* + \varepsilon \|c\|.$$

Proof. The constants \mathcal{C}_0 and \mathcal{C}_1 are chosen in the same way as before. (i) By Thm. 4.1, with probability at least $1 - \delta$, for any feasible point u of the projected problem $(\text{QP}_\varepsilon^-)$, $P^\top u$ is also feasible for the original problem Eq. (2). Therefore, it must hold also for x_ε^- .

(ii) By Part (i) above, with probability at least $1 - \delta$, x_ε^- is feasible for the original problem Eq. (2). Therefore, we have $x_\varepsilon^{-\top} Q x_\varepsilon^- + c^\top x_\varepsilon^- \leq x^{*\top} Q x^* + c^\top x^*$ with probability at least $1 - \delta$. Moreover, due to Lemma 3.2, with probability at least $1 - 8(k+1)e^{-C_0\varepsilon^2 d}$, where k is the rank of Q , we have

$$\begin{aligned} x^{*\top} Q x^* &\leq x^{*\top} P^\top P Q P^\top P x^* + 3\varepsilon \|x^*\|^2 \|Q\|_* \leq x^{*\top} P^\top P Q P^\top P x^* + 3\varepsilon \|Q\|_* \\ \text{and } c^\top x^* &\leq c^\top P^\top P x^* + \varepsilon \|c\| \|x^*\| \leq c^\top P^\top P x^* + \varepsilon \|c\|, \end{aligned}$$

since $\|x^*\| \leq 1$. Hence $x^{*\top} Q x^* + c^\top x^* \leq x^{*\top} P^\top P Q P^\top P x^* + c^\top P^\top P x^* + \varepsilon \|c\| + 3\varepsilon \|Q\|_*$. On the other hand, let $\hat{u} = P x^*$; due to Lemma 3.2, we have

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

with probability at least $1 - 4m e^{-C_0\varepsilon^2 d}$ (the last inequality holds by the assumption $b \geq 0$), and $\|\hat{u}\| = \|P x^*\| \leq (1 + \varepsilon) \|x^*\| \leq (1 + \varepsilon)$ with probability at least $1 - 2e^{-C_0\varepsilon^2 d}$ (by Lemma 3.2). Therefore, \hat{u} is a feasible solution for the problem $(\text{QP}_\varepsilon^+)$ with probability at least $1 - (4m+2)e^{-C_0\varepsilon^2 d}$. Due to the optimality of u_ε^+ for the problem $(\text{QP}_\varepsilon^+)$, it follows that

$$\begin{aligned} x^{*\top} Q x^* + c^\top x^* &\leq x^{*\top} P^\top P Q P^\top P x^* + c^\top P^\top P x^* + \varepsilon \|c\| + 3\varepsilon \|Q\|_* \\ &= \hat{u}^\top P Q P^\top \hat{u} + c^\top P^\top \hat{u} + \varepsilon \|c\| + 3\varepsilon \|Q\|_* \\ &\leq u_\varepsilon^{+\top} P Q P^\top u_\varepsilon^+ + (Pc)^\top u_\varepsilon^+ + \varepsilon \|c\| + 3\varepsilon \|Q\|_* \\ &= x_\varepsilon^{+\top} Q x_\varepsilon^+ + c^\top x_\varepsilon^+ + \varepsilon \|c\| + 3\varepsilon \|Q\|_*, \end{aligned}$$

with probability at least $1 - (4m+6)e^{-C_0\varepsilon^2 d}$, which is at least $1 - \delta$ for the chosen universal constant \mathcal{C}_0 . Hence $x^{*\top} Q x^* + c^\top x^* \leq x_\varepsilon^{+\top} Q x_\varepsilon^+ + c^\top x_\varepsilon^+ + 3\varepsilon \|Q\|_* + \varepsilon \|c\|$, which concludes the proof. \square

The above result implies that the value of $x^{*\top} Q x^* + c^\top x^*$ lies between $x_\varepsilon^{-\top} Q x_\varepsilon^- + c^\top x_\varepsilon^-$ and $x_\varepsilon^{+\top} Q x_\varepsilon^+ + c^\top x_\varepsilon^+$. It remains to prove that these two values are not so far from each other. Let $S^* = \{x \in \mathbb{R}^n \mid Ax \leq b, \|x\| \leq 1\}$. Since, by assumption, the feasible set of (2) is full dimensional, S^* is also full dimensional.

We associate with each set S a positive number $\text{FULL}(S) > 0$, which is considered as a fullness measure of S and is defined as the maximum radius of any closed ball contained in S . Now, from our assumption, we have $\text{FULL}(S^*) = r^* > 0$, where r^* is the radius of the greatest ball inscribed in S^* (see Fig. 1, left).

The following lemma characterizes the fullness of S_ε^+ with respect to r^* , where $S_\varepsilon^+ = \{u \in \mathbb{R}^d \mid A P^\top u \leq b + \varepsilon, \|u\| \leq 1 + \varepsilon\}$ is the feasible set of the problem $(\text{QP}_\varepsilon^+)$.

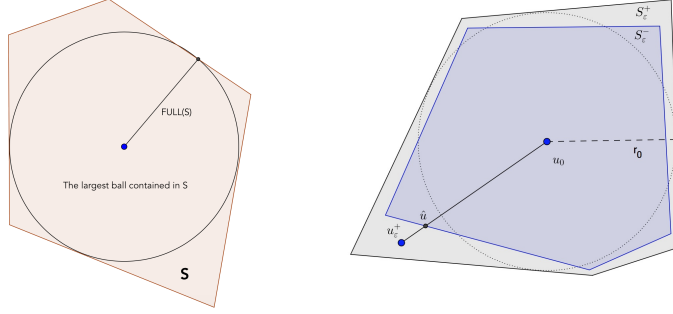


Fig. 1. Left: fullness of a set. Right: idea of the proof of Thm. 4.4.

Lemma 4.3. *Let S^* be full-dimensional with $\text{FULL}(S^*) = r^*$. Then with probability at least $1 - 3\delta$, S_ε^+ is also full-dimensional with the fullness measure $\text{FULL}(S_\varepsilon^+) \geq (1 - \varepsilon)r^*$.*

The proof of this lemma extensively uses the fact that, for any row vector $a \in \mathbb{R}^n$ $\sup_{\|u\| \leq r} a^\top u = r\|a\|$, which is actually the equality condition in the Cauchy-Schwartz inequality.

Now we will estimate the gap between the two objective functions of the problems $(\text{QP}_\varepsilon^+)$ and $(\text{QP}_\varepsilon^-)$ using the fullness measure. The theorem states that, as long as the fullness of the original polyhedron is large enough, the gap between them is small. Fig. 1, right, gives the proof idea.

Theorem 4.4. *Let $0 < \varepsilon < 0.1$. Then with probability at least $1 - 4\delta$, we have $x_\varepsilon^{-\top} Q x_\varepsilon^- + c^\top x_\varepsilon^- \leq x_\varepsilon^{+\top} Q x_\varepsilon^+ + c^\top x_\varepsilon^+ < \frac{(1+\varepsilon)^2}{(1-\varepsilon)^2} (x_\varepsilon^{-\top} Q x_\varepsilon^- + c^\top x_\varepsilon^-) + \frac{\varepsilon}{\text{FULL}(S^*)} (36 + 18\|c\|)$.*

Proof. Let $B(u_0, r_0)$ be a closed ball with maximum radius that is contained in S_ε^+ . In order to establish the relation between u_ε^+ and u_ε^- , our idea is to move u_ε^+ closer to u_0 , so that the new point is contained in S_ε^- . Therefore, its objective value will be less than the value of u_ε^- , but quite close to the objective value of u_ε^+ . We define $\hat{u} = (1 - \lambda)u_\varepsilon^+ + \lambda u_0$ for some $\lambda \in (0, 1)$: we want to find λ such that \hat{u} is feasible for $(\text{QP}_\varepsilon^-)$ while its corresponding objective value is not so different from $x_\varepsilon^{+\top} Q x_\varepsilon^+ + c^\top x_\varepsilon^+$. Since for all $\|u\| \leq r_0$, $AP^\top(u_0 + u) = AP^\top u_0 + AP^\top u \leq b + \varepsilon$. Then $AP^\top u_0 \leq b + \varepsilon - r_0(\|A_1 P^\top\|, \dots, \|A_m P^\top\|)^\top$. Therefore, we have, with probability at least $1 - \delta$,

$$AP^\top u_0 \leq b + \varepsilon - r_0(1 - \varepsilon)(\|A_1\|, \dots, \|A_m\|)^\top = b + \varepsilon - r_0(1 - \varepsilon).$$

Hence $AP^\top \hat{u} = (1 - \lambda)AP^\top u_\varepsilon^+ + \lambda AP^\top u_0 \leq b + \varepsilon - \lambda r_0(1 - \varepsilon) \leq b + \varepsilon - \frac{1}{2}\lambda r_0$, since we can assume w.l.o.g. that $\varepsilon \leq \frac{1}{2}$. Hence, $AP^\top \hat{u} \leq b$ if we choose $\varepsilon \leq \lambda \frac{r_0}{2}$. Furthermore $\|\hat{u}\| \leq 1 + \varepsilon$. Hence, when we choose $\lambda = 2\frac{\varepsilon}{r_0}$, then $\frac{1 - \varepsilon}{1 + \varepsilon} \hat{u}$ is feasible for the problem $(\text{QP}_\varepsilon^-)$ with probability at least $1 - \delta$. Therefore,

$$\begin{aligned}
& \frac{1+\varepsilon}{1-\varepsilon} u_\varepsilon^- \top P Q P^\top u_\varepsilon^- + (Pc)^\top u_\varepsilon^- \text{ is greater than or equal to } \hat{u}^\top P Q P^\top \hat{u} + (Pc)^\top \hat{u} = \\
& = (u_\varepsilon^+ + \lambda(u_0 - u_\varepsilon^+))^\top P Q P^\top (u_\varepsilon^+ + \lambda(u_0 - u_\varepsilon^+)) + (Pc)^\top \hat{u} \\
& = u_\varepsilon^+ \top P Q P^\top u_\varepsilon^+ + \lambda u_\varepsilon^+ \top P Q P^\top (u_0 - u_\varepsilon^+) + \lambda(u_0 - u_\varepsilon^+)^\top P Q P^\top u_\varepsilon^+ \\
& \quad + \lambda^2 (u_0 - u_\varepsilon^+)^\top P Q P^\top (u_0 - u_\varepsilon^+) + (Pc)^\top \hat{u}.
\end{aligned}$$

However, from Lemma 3.2 and the Cauchy-Schwartz inequality, we have

$$\begin{aligned}
|u_\varepsilon^+ \top P Q P^\top (u_0 - u_\varepsilon^+)| & \leq \|P^\top u_\varepsilon^+\| \|Q\|_2 \|P^\top (u_0 - u_\varepsilon^+)\| \\
& \leq (1 + \varepsilon)^2 \|u_\varepsilon^+\| \|Q\|_2 \|(u_0 - u_\varepsilon^+)\| \leq 2(1 + \varepsilon)^4 \|Q\|_2
\end{aligned}$$

(since $\|u_\varepsilon^+\|$ and $\|u_\varepsilon^-\| \leq 1 + \varepsilon$), and similarly for other terms. We then have

$$\hat{u}^\top P Q P^\top \hat{u} \geq u_\varepsilon^+ \top P Q P^\top u_\varepsilon^+ - (4\lambda + 4\lambda^2)(1 + \varepsilon)^4 \|Q\|_2.$$

Since $\varepsilon < 0.1$, we have $(1 + \varepsilon)^4 < 2$ and we can assume that $\lambda < 1$. Then we have

$$\begin{aligned}
\hat{u}^\top P Q P^\top \hat{u} & > u_\varepsilon^+ \top P Q P^\top u_\varepsilon^+ - 16\lambda \|Q\|_2 \\
& = u_\varepsilon^+ \top P Q P^\top u_\varepsilon^+ - 32\varepsilon/r_0 \quad (\text{since } \|Q\|_2 = 1) \\
& \geq u_\varepsilon^+ \top P Q P^\top u_\varepsilon^+ - 32\varepsilon/((1 - \varepsilon)\text{FULL}(S^*)) \quad (\text{due to Lemma 4.3}) \\
& > u_\varepsilon^+ \top P Q P^\top u_\varepsilon^+ - 36\varepsilon/\text{FULL}(S^*) \quad (\text{since } \varepsilon \leq 0.1),
\end{aligned}$$

with probability at least $1 - 2\delta$. Furthermore, we have

$$c^\top P^\top \hat{u} = c^\top P^\top u_\varepsilon^+ + \lambda c^\top P^\top (u_0 - u_\varepsilon^+) \geq c^\top P^\top u_\varepsilon^+ - \frac{4(1 + \varepsilon)\varepsilon}{r_0} \|Pc\|.$$

We know that $r_0 \geq (1 - \varepsilon)r^*$, hence

$$c^\top P^\top \hat{u} \geq c^\top P^\top u_\varepsilon^+ - \frac{4(1 + \varepsilon)\varepsilon}{r_0} \|Pc\| \geq c^\top P^\top u_\varepsilon^+ - \frac{4(1 + \varepsilon)^2 \varepsilon}{(1 - \varepsilon)r^*} \|c\|,$$

with probability at least $1 - \delta$. The results holds by $\varepsilon < 0.1$. \square

5 Computational results

Although we developed our theory for dense Gaussian RPs, in practice one can decrease computational costs considerably by using sparsity [1, 3]. All of the results of this paper, aside from Lemma 3.1, actually hold (unchanged) also for *sub-gaussian* RPs. Amongst sub-gaussian RPs, we elect to use $d \times n$ matrices where each component is sampled from $\mathbf{N}(0, \frac{1}{\sqrt{d}})$ with some given probability $\text{dens} \in (0, 1)$. Although we are not going to include the generalization of Lemma 3.1 to sub-gaussian RPs here for lack of space, the proof exploits the fact that the largest and smallest singular values of sub-gaussian RPs are approximately the same.

All tests were carried out on a single core of a 4-CPU machine with 64GB RAM, each CPU of which has 8 cores (Intel Xeon CPU E5-2620 v4@2.10GHz).

5.1 Random instances

Our first computational test is carried out of randomly generated feasible instances of Eq. (2) with Q negative semidefinite (we make this assumption in order to compute guaranteed global maxima in acceptable CPU times for comparison purposes: the projection technique is independent of the convexity of the objective function). We generate all instances varying the following parameters: number of constraints $m \in \{10, 100, 1000\}$, of variables $n \in \{2000, 3000\}$, random number generation distribution $\text{distr} \in \{0, 1\}$ (choosing between uniform distributions $U(0, 1)$ and $U(-1, 1)$) and density $\text{dens} \in \{0.1, 0.6\}$ for matrices A and Q . As mentioned above, our RPs are sparse random Gaussian matrices P with $\varepsilon \in \{0.10, 0.15, 0.20\}$ and density $\text{dens}_P \in \{0.2, 0.5, 1.0\}$. This yields a total of 216 solution logs (all instances with all RP generation methods) obtained using the IPOPT solver [14]. We benchmark means, standard deviations, maximum and minimum values for: (i) CPU time (solution of original vs. projected problem to optimality, where the projected CPU time also includes RP sampling, matrix multiplication and solution retrieval time); (ii) objective function ratio $\rho = \frac{|f_{\text{org}}^* - f_{\text{retr}}^*|}{\max(|f_{\text{org}}^*|, |f_{\text{retr}}^*|)}$, where f_{org}^* is the optimal objective function value of Eq. (2) and f_{retr}^* is the value of the objective function of Eq. (2) evaluated at the solution retrieved from the projected problem Eq. (5); (iii) average feasibility errors denoted **ace**, **are**, **be** for, respectively, $Ax \leq b$, ranges $-1 \leq x \leq 1$ and $\|x\| \leq 1$. All our coding was carried out in Julia+JuMP [5].

	CPU _{org}	CPU _{proj}	ρ	ace	are	be
mean	37.691	14.590	0.103	0.0	0.0	2.237
stdev	49.984	15.057	0.070	0.0	0.0	0.916
min	8.750	2.170	0.000	0.0	0.0	0.921
max	198.350	61.340	0.485	0.0	0.0	3.886

Table 1. Computational results for random instances.

Table 1 shows a consistent behaviour of our projection technique: projected formulations take considerably less time to solve (despite pre- and post-processing steps), and yield solutions having objective function values within around 10% of the optimum, with no feasibility error w.r.t. linear and range constraints. There is a large ball error, however, which we are unable to explain at this time — we are looking into it. Scaling the retrieved solution back to norm 1 yields a feasible point but increases ρ considerably (to around 0.4 on average).

Table 2 shows the trade-off between approximation quality and efficiency in function of the parameters ε and dens_P of the RP (blackier is better). The best compromise appears to be achieved for $\varepsilon = 0.15$ and $\text{dens}_P = 0.2$.

dens_P	0.2			0.5			1.0		
	CPU_{org}	CPU_{proj}	ρ	CPU_{org}	CPU_{proj}	ρ	CPU_{org}	CPU_{proj}	ρ
0.10	37.57	19.91	0.07	37.90	21.02	0.06	37.76	24.36	0.07
0.15	37.56	11.26	0.10	37.81	11.95	0.09	37.78	12.16	0.11
0.20	37.50	10.04	0.14	37.71	10.30	0.15	37.62	10.32	0.14

Table 2. Trade-off between approximation and efficiency for ε and dens_P .

5.2 Two large portfolio instances

We consider two large-scale Markowitz portfolio [6, 7] instances where the objective is a scalarized version of risk minimization (using correlation rather than covariance for better scaling) and return maximization. The system $Ax \leq b$ encodes the portfolio constraints $0 \leq x \leq 1$ (with $-x \leq 0$ being part of the inequality constraints) and $\sum_j x_j \leq 1$, which imply $\|x\|_2 \leq 1$. The stock price data were obtained from Kaggle ([goo.gl/XHfhi2](https://www.kaggle.com/goo.gl/XHfhi2)), and yielded fully dense Q matrices. We have used the $\varepsilon = 0.15$ and $\text{dens}_P = 0.2$ RP settings obtained from Table 2. Our results are presented in Table 3. The computational savings

Instance	n	$\text{nnz}(Q)$	CPU_{org}	CPU_{proj}	ρ	ace	are	be
etfs	1344	902,496	534.32	11.38	0.270	0.026	0.001	1.570
stocks	7163	25,650,703	266,713.40	132.78	0.007	0.023	0.001	3.927

Table 3. Results for two large instances of Markowitz’ portfolio problem.

are remarkable, the optimal objective function values are within a reasonable approximation ratio, but the retrieved solutions are slightly infeasible w.r.t. linear and range constraints. Specifically, some of the components of the retrieved solutions are very slightly negative (0.001 for **etfs** and 0.0005 for **stocks** on average), which is an issue we had also observed in applying RPs to Linear Programs [13]. The ball errors are again high.

6 Conclusion

We prove that random projections can be used to generate lower dimensional QPs, bounded by a Euclidean ball constraint, which have approximately the same global optimum with arbitrarily high probability as their original counterparts. Computational results are exhibited to substantiate our claim and show the applicability of our techniques.

As a corollary, we remark that our results are also applicable to reduce the number of variables of inequality constrained LPs (with a Euclidean ball constraint), since there is no assumption on Q (so $Q = 0$ is a possibility). We are in

the process of deriving theorems which ensure better bounds given this specific structure.

References

1. Achlioptas, D.: Database-friendly random projections: Johnson-Lindenstrauss with binary coins. *Journal of Computer and System Sciences* 66, 671–687 (2003)
2. 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. American Mathematical Society, Providence (1984)
3. Kane, D., Nelson, J.: Sparser Johnson-Lindenstrauss transforms. *Journal of the ACM* 61(1), 4 (2014)
4. Liberti, L., Vu, K.: Barvinok’s naive algorithm in distance geometry. *Operations Research Letters* 46, 476–481 (2018)
5. Lubin, M., Dunning, I.: Computing in operations research using Julia. *INFORMS Journal on Computing* 27(2), 238–248 (2015), <http://dx.doi.org/10.1287/ijoc.2014.0623>
6. Markowitz, H.: Portfolio selection. *The Journal of Finance* 7(1), 77–91 (1952)
7. Mencarelli, L., D’Ambrosio, C.: Complex portfolio selection via convex mixed-integer quadratic programming: a survey. *International Transactions in Operational Research* (to appear)
8. 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)
9. Vavasis, S.: Quadratic programming is in NP. *Information Processing Letters* 36, 73–77 (1990)
10. Vavasis, S., Zippel, R.: Proving polynomial-time for sphere-constrained quadratic programming. Tech. Rep. 90-1182, Dept. of Comp. Sci., Cornell University (1990)
11. Vavasis, S.: *Nonlinear Optimization: Complexity Issues*. Oxford University Press, Oxford (1991)
12. Vu, K., Poirion, P.L., Liberti, L.: Gaussian random projections for euclidean membership problems. *Discrete Applied Mathematics* doi:10.1016/j.dam.2018.08.025 (2018)
13. Vu, K., Poirion, P.L., Liberti, L.: Random projections for linear programming. *Mathematics of Operations Research* doi:10.1287/moor.2017.0894 (to appear)
14. Wächter, A., Biegler, L.: On the implementation of an interior-point filter line-search algorithm for large-scale nonlinear programming. *Mathematical Programming* 106(1), 25–57 (2006)
15. Woodruff, D.: Sketching as a tool for linear algebra. *Foundations and Trends in Theoretical Computer Science* 10(1-2), 1–157 (2014)
16. 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)*. *Proceedings of Machine Learning Research*, vol. 30, pp. 135–157. jmlr.org (2013)