Random projections for quadratic programs[1]

Claudia D'Ambrosio[1], Leo Liberti[1], Pierre-Louis Poirion[2], Ky Vu[3]

1. *CNRS LIX, École Polytechnique, F-91128 Palaiseau, France*
   Email:`{dambrosio,liberti}@lix.polytechnique.fr`
2. *RIKEN Center for Advanced Intelligence Project, Tokyo, Japan*
3. *Dept. of Mathematics, FPT University, Hanoi, Vietnam*

June 10, 2019

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. While random projections are usually applied to numerical data, we show 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. We also show how to retrieve a feasible solution of the original problem from the lower-dimensional solution of the projected problem. We then show that the retrieved solution can be used to bound the optimal objective function value of the original problem from below and above, and show that the lower and upper bounds are not too far apart. We then discuss a set of computational results on randomly generated instances, as well as a variant of Markowitz' portfolio problem.

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 can be used to construct an approximate solution for the original QP. We consider the following QP formulation:

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

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 on Eq. (1).

[1] This paper has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement n. 764759 "MINOA". This paper generalizes [24].

1. We assume that $Ax \leq b$ defines a full dimensional polytope.
2. We assume we are given the radius R of a ball containing the polytope $Ax \leq b$.
3. We also 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 $\forall i \leq n\ \frac{A_i{}^\top x}{\mu_i} \leq \frac{b_i}{\mu_i}$, which satisfies the assumption.

No assumption is made on Q, which may be positive/negative semidefinite or indefinite. In the following, all norm symbols $\|\cdot\|$ will be assumed to refer to the ℓ_2 norm $\|\cdot\|$, 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 [7, 19]). If we assume that all the data are rational, then the decision version of Eq. (1) is **NP**-complete [20].

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 approximately the same optima as the original QP. We also perform a computational verification of our claim, and discuss the extent to which the theoretical results, which are asymptotic in nature, can also be applied 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 [3] or k-nearest neighbours [9, 10, 2]. Since, according to the Johnson-Lindenstrauss lemma [11] RPs ensure 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 invarian (as regards feasibility and optimality) w.r.t. randomly projecting the input parameters. Similarly in spirit to our previous work on Linear Programming (LP) [25], but using a 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 [27, 17]: two assumptions we do not make.

This paper is a considerable extension of [24]: we remove the need for a ball constraint, thus providing a generalization to *any* QP satisfying the assumptions 1-3 above. While the general structure of this article is similar to that of [24], the removal of the ball constraint allowed us to streamline certain results (e.g. Prop. 4.1), but forced us to completely change some others (e.g. Thm. 4.4). In some cases we were able to improve some bounds (e.g. Lemma 3.3). Consequently, almost everything in this paper is an original contribution, including the proof for our sparse random projections in Sect. 5.1 (different, as far as we know, from those in the literature) as well as the computational experiments.

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 taking 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 [25]. Thus, the restriction to the LP case of the results in this paper yields a new analysis for the projected formulation of [25], 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 [27, 17, 26, 25, 13]. 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 [23]: e.g. $\mathsf{N}(0,1)$, 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 [21]. In particular, let $P \in \mathbb{R}^{d\times n}$ be a RP every component of which is sampled from $\mathsf{N}(0, 1/\sqrt{d})$ (where $1/d$ is the variance). Then, for any $x \in \mathbb{R}^n$ and $\varepsilon \in (0,1)$, we have

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

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 [11]. 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

$$\forall x, y \in X \qquad (1-\varepsilon)\|x-y\|^2 \le \|F(x)-F(y)\|^2 \le (1+\varepsilon)\|x-y\|^2. \tag{3}$$

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 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 application.

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*:

$$\max\{x^\top(P^\top PQP^\top P)x + c^\top P^\top Px \mid AP^\top Px \le b\}.$$

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

$$\max_{u\in\,\mathsf{Im}(P)} \{u^\top \bar{Q} u + \bar{c}^\top u \mid \bar{A}u \le b\}, \tag{4}$$

where $\mathsf{Im}(P)$ 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 $u \in \mathsf{Im}(P)$ and study the smaller dimensional problem:

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

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.

Let $u^* \in \mathbb{R}^d$ be a solution of RP. We look for a solution $x' \in \mathbb{R}^n$ which is feasible in P. Consider $x' = P^\top u^*$. We have $Ax' = AP^\top u^* \leq b$ by feasibility of u^*, which establishes the feasibility of x' in P. The approximate optimality of x' will be established in Sect. 4.

2.1 Proposition
We have $\mathsf{v}(\mathsf{RP}) \leq \mathsf{v}(\mathsf{P})$.

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

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

Since P, RP are maximization problems, the result follows. □

Since P is in maximization form, we remark that Prop. 2.1 does not prove that RP is a relaxation of P.

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 3.1 shows that RPs approximately preserve scalar products. Lemma 3.2 shows that RPs approximately preserve double-sided inequalities. Lemma 3.3 shows that RPs approximately preserve quadratic forms.

3.1 Lemma
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 $1 - 4e^{-\mathcal{C}\varepsilon^2 d}$.

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

$$\begin{aligned}|\langle Pu, Pv\rangle - \langle u,v\rangle| &= \frac{1}{4}\big|\|P(u+v)\|^2 - \|P(u-v)\|^2 - \|u+v\|^2 + \|u-v\|^2\big| \\ &\le \frac{1}{4}\big|\|P(u+v)\|^2 - \|u+v\|^2\big| + \frac{1}{4}\big|\|P(u-v)\|^2 - \|u-v\|^2\big| \\ &\le \frac{\varepsilon}{4}(\|u+v\|^2 + \|u-v\|^2) = \frac{\varepsilon}{2}(\|u\|^2 + \|v\|^2),\end{aligned}$$

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. □

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

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

The claim follows by applying Lemma 3.1 and the union bound. □

3.3 Lemma
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 $Q \in \mathbb{R}^{n\times n}$, then with probability at least $1 - 8\,k\,e^{-\mathcal{C}\varepsilon^2 d}$, we have:

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

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:

$$\begin{aligned}x^\top P^\top PQP^\top Py &= (U^\top P^\top Px)^\top \Sigma (V^\top P^\top Py) \\ &= \big[U^\top x + U^\top(P^\top P - I_n)x\big]^\top \Sigma \big[V^\top y + V^\top(P^\top P - I_n)y\big].\end{aligned}$$

Therefore,

$$\begin{aligned}|x^\top P^\top PQP^\top Py - x^\top Qy| \;&\le\; |(U^\top x)^\top\,\Sigma\,V^\top(P^\top P - I_n)y| + |(U^\top(P^\top P - I_n)x)^\top\,\Sigma\,V^\top y| \\ &+\; |(U^\top(P^\top P - I_n)x)^\top\,\Sigma\,V^\top(P^\top P - I_n)y|, \qquad (6)\end{aligned}$$

since $U^\top x \Sigma V^\top x = x^\top Q y$. By Lemma 3.2, we have

$$\forall i \le n \quad |(\Sigma V^\top (P^\top P - I_n)y)_i| \le \varepsilon \sigma_i \|y\|. \tag{7}$$

Moreover, since Σ is symmetric, we also have

$$\forall i \le n \quad |((U^\top (P^\top P - I_n)x)^\top \Sigma)_i| \le \varepsilon \sigma_i \|x\|. \tag{8}$$

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

$$\langle U^\top x \,, \Sigma V^\top (P^\top P - I_n)y\rangle \le \|U^\top x\| \, \|V^\top (P^\top P - I_n)y\| \tag{9}$$

$$\langle (U^\top (P^\top P - I_n)x)^\top \Sigma \,, V^\top y\rangle \le \|(U^\top (P^\top P - I_n)x)^\top \Sigma\| \, \|V^\top y\| \tag{10}$$

$$\langle (U^\top (P^\top P - I_n)x)^\top \Sigma \,, V^\top (P^\top P - I_n)y\rangle \le \|(U^\top (P^\top P - I_n)x)^\top \Sigma\| \, \|V^\top (P^\top P - I_n)y\|. \tag{11}$$

We claim that

$$\|U^\top x\| \le \|x\| \tag{12}$$

$$\|V^\top y\| \le \|y\|. \tag{13}$$

First, we observe that

$$\|U^\top x\|^2 = \langle U^\top x, U^\top x\rangle = \langle UU^\top x, x\rangle. \tag{14}$$

We consider the left vector $UU^\top x$ as an inner product of $UU^\top$ and x. By sub-multiplicativity of the matrix norm induced by the ℓ_2 norm we obtain $\|UU^\top x\| \le \|UU^\top\|_2 \, \|x\|$. Note that $\|UU^\top\| = \|U^\top U\|$ because matrix products commute under norms. Now, $U^\top U = I_k$ since $U^\top$ has orthonormal rows. Hence the largest eigenvalue of $U^\top U$ is 1. By definition of the ℓ_2 matrix norm, this implies that $\|U^\top U\|_2 = 1$. Hence we have $\|UU^\top\|_2 \, \|x\| = \|x\|$. By Eq. (14) and Cauchy-Schwartz,

$$\langle UU^\top x, x\rangle \le \|UU^\top x\| \, \|x\| \le \|x\| \, \|x\| = \|x\|^2,$$

which immediately implies $\|U^\top x\| \le \|x\|$. The argument for V, y is the same. Next, we use Eq. (12)-(13) and Eq. (7)-(8) to bound Eq. (9)-(11). We obtain:

$$\langle U^\top x \,, \Sigma V^\top (P^\top P - I_n)y\rangle \le \varepsilon \|x\| \, \|y\| \, \|\sigma\| \tag{15}$$

$$\langle (U^\top (P^\top P - I_n)x)^\top \Sigma \,, V^\top y\rangle \le \varepsilon \|y\| \, \|x\| \, \|\sigma\| \tag{16}$$

$$\langle (U^\top (P^\top P - I_n)x)^\top \Sigma \,, V^\top (P^\top P - I_n)y\rangle \le \varepsilon^2 \|x\| \, \|y\| \, \|\sigma\|, \tag{17}$$

where the bound for Eq. (17) was obtained by setting $\Sigma = I_n$ in Eq. (7). Finally, we obtain

$$|x^\top P^\top P Q P^\top P y - x^\top Q y| \le 3\varepsilon \|x\| \, \|y\| \, \|\sigma\| \tag{18}$$

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

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

$$\mathsf{RP}_\varepsilon \equiv \max\{u^\top \bar{Q} u + \bar{c}^\top u \mid \bar{A} u \le b + R\varepsilon \;\wedge\; u \in \mathbb{R}^d\}. \tag{19}$$

For a MP formulation P, we will denote its optimal objective function value by $\mathsf{v}(P)$.

We first state the following (obvious) feasibility result.

4.1 Proposition
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 $A\hat{x} = AP^\top u \leq b$. □

Let u^- be an optimal solution for RP (Eq. (5)) and u^+ be an optimal solution for RP_ε (Eq. (19)). 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 $\mathsf{v}(\mathsf{P}) = x^{*\top}Qx^* + c^\top x^*$ between $\mathsf{v}(\mathsf{RP}) = x^{-\top}Qx^- + c^\top x^-$ and $\mathsf{v}(\mathsf{RP}_\varepsilon) = x^{+\top}Qx^+ + c^\top x^+$, the two values that are expected to be approximately close to each other.

4.2 Theorem
Let $P \in \mathbb{R}^{d\times n}$ be a RP, and let $\delta \in (0,1)$. There is a universal constant $\mathcal{C} > 1$ such that, if $d \geq \ln(m/\delta)/(\mathcal{C}\varepsilon^2)$, 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 Prop. 4.1 x^- is feasible in Eq. (1). By Prop. 2.1 we have $\mathsf{v}(\mathsf{RP}) \leq \mathsf{v}(\mathsf{P})$. Moreover, by Lemma 3.3, with probability at least $1 - 8(k+1)e^{-\mathcal{C}\varepsilon^2 d}$, where k is the rank of Q, we have

$$\begin{aligned} x^{*\top}Qx^* \leq\ & x^{*\top}P^\top PQP^\top Px^* + 3\varepsilon\|x^*\|^2\,\|Q\|_F && \leq x^{*\top}P^\top PQP^\top Px^* + 3R^2\varepsilon\|Q\|_F \\ \text{and} \qquad c^\top x^* \leq\ & c^\top P^\top Px^* + \varepsilon\|c\|\,\|x^*\| && \leq c^\top P^\top Px^* + R\varepsilon\|c\|, \end{aligned}$$

since $\|x^*\| \leq 1$. Hence

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

On the other hand, let $\hat{u} = Px^*$; by Lemma 3.2, we have

$$AP^\top\hat{u} = AP^\top Px^* \leq Ax^* + \varepsilon\|x^*\|\mathbf{1} \leq Ax^* + R\varepsilon\mathbf{1} \leq b + R\varepsilon$$

with probability at least $1 - 4\,m\,e^{-\mathcal{C}\varepsilon^2 d}$ (the last inequality holds by the assumption $b \geq 0$). Therefore, $\hat{u}$ is a feasible solution for the problem (19) with probability at least $1 - 4me^{-\mathcal{C}\varepsilon^2 d}$. Due to the optimality of u^+ for problem (19), it follows that

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

with probability at least $1 - (4m+6)e^{-\mathcal{C}\varepsilon^2 d}$, which is at least $1-\delta$ for the chosen universal constant $\mathcal{C}$. Hence $\mathsf{v}(\mathsf{P}) \leq \mathsf{v}(\mathsf{RP}_\varepsilon) + 3R^2\varepsilon\|Q\|_F + R\varepsilon\|c\|$, which concludes the proof. □

The above result implies that the value of $\mathsf{v}(\mathsf{P})$ lies between $\mathsf{v}(\mathsf{RP})$ and $\mathsf{v}(\mathsf{RP}_\varepsilon)$ (plus some error term). We will now prove that these two values are not so far from each other. In order to achieve this goal, we associate with each set S a "fullness measure" $\text{FULL}(S) > 0$, 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 $\text{FULL}(S) = r > 0$ (see Fig. 1, left). We define S^+ as the feasible

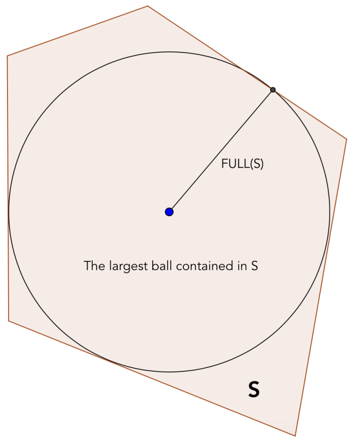

Figure 1: Fullness of a set.

region of Eq. (19). The following lemma characterizes the fullness of S^+ with respect to r. Its proof extensively uses the fact that, for any row vector $a \in \mathbb{R}^n$, we have $\max_{\|u\| \leq r} a^\top u = r\|a\|$ (the equality condition in the Cauchy-Schwartz inequality).

4.3 Lemma
With probability at least $1 - 3\delta$, S^+ is full-dimensional with $\text{FULL}(S^+) \geq (1 - \varepsilon)r$.

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 $\|x\| \leq r$, we have $x_0 + x \in S$, and hence $A(x_0 + x) = Ax_0 + Ax \leq b,$. This implies that

$$\forall i \leq n \quad b_i \geq (Ax_0)_i + \max_{\|x\| \leq r} A_i x = (Ax_0)_i + r\|A_i\| = (Ax_0)_i + r. \tag{20}$$

Hence $Ax_0 + r \leq b$ or, equivalently, $Ax_0 \leq b - r$. By Lemma 3.2, with probability at least $1 - \delta$, we have $AP^\top Px_0 \leq Ax_0 + R\varepsilon \leq b - r + R\varepsilon$. Let $u \in \mathbb{R}^n$ with $\|u\| \leq (1 - \varepsilon)r$, then, by the above inequality, we have:

$$\begin{aligned} \forall i \leq n \quad (AP^\top(Px_0 + u))_i &= (AP^\top Px_0)_i + (AP^\top u)_i \leq b_i + R\varepsilon - r + (AP^\top)_i u \\ &= b_i + R\varepsilon - r + A_i P^\top u, \end{aligned}$$

where $(AP^\top)_i$ denotes the ith row of $AP^\top$. Since this holds for all such vectors u, by the Cauchy-Schwartz inequality we have

$$(AP^\top(Px_0 + u))_i \leq b_i + R\varepsilon - r + (1 - \varepsilon)r\|A_i P^\top\|.$$

Using Eq. (2) and the union bound, we can see that, with probability at least $1 - 2me^{-\mathcal{C}\varepsilon^2 d} \geq 1 - \delta$, we have

$$\forall i \leq n \quad \|A_i P^\top\| \leq (1 + \varepsilon)\|A_i\| = (1 + \varepsilon).$$

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

$$AP^\top(Px_0 + u) \leq b + R\varepsilon - r + (1-\varepsilon)r(1+\varepsilon) \leq b + R\varepsilon$$

with probability at least $1 - 2\delta$. In other words, with probability at least $1 - 2\delta$, the closed ball $\bar{B}$ centered at Px_0 with radius $(1-\varepsilon)r$ is contained in $\{u \mid AP^\top u \leq b + R\varepsilon\}$. Therefore, by definition of S^+ we have $B\big(Px_0, (1-\varepsilon)r\big) \subseteq S^+$, which implies that the fullness of S^+ is at least $(1-\varepsilon)r$, with probability at least $1 - 3\delta$. □

Now we will estimate the gap between the two objective functions of the problems Eq. (19) and Eq. (5) 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 Thm. 4.2 are not themselves unbounded above and below.

4.4 Theorem
Let $\varepsilon \in (0, \frac{r}{R})$. Then with probability at least $1 - 4\delta$, we have

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

where $\sigma = 1 - \frac{R}{r}\frac{\varepsilon}{(1-\varepsilon)^2} > 0$, $r = \textsc{full}(S)$, and S is the feasible region of P.

Proof. The fact that $\mathsf{v}(\mathsf{RP}) \leq \mathsf{v}(\mathsf{RP}_\varepsilon)$ 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. (19)),

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

Assuming $b > 0$, this is equivalent to

$$\forall i \leq m \quad \frac{1}{b_i}\bar{A}_i w \leq 1 + \frac{R\varepsilon}{b_i}. \tag{21}$$

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. (21) 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

$$\rho^2 \mathsf{v}(\mathsf{RP}_\varepsilon) \leq \rho^2 w^\top \bar{Q} w + \rho \bar{c}^\top w = \bar{w}^\top \bar{Q}\bar{w} + \bar{c}^\top \bar{w} \leq \mathsf{v}(\mathsf{RP}) \tag{22}$$

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 4.3, there is a ball $B(u_0, (1-\varepsilon)r)$ contained in S^+, the feasible region of RP_ε. We define new variables $u' = u - u_0$. Replacing u by $u' + u_0$ in Eq. (19) 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^+, we have:

$$\forall i \le m \quad \max\{\bar{A}_i u' \mid \|u'\| \le (1-\varepsilon)r\} \le \hat{b}_i + R\varepsilon. \tag{23}$$

We now observe that $\max\{\bar{A}_i u' \mid \|u'\| \le (1-\varepsilon)r\} = \|\bar{A}_i\|(1-\varepsilon)r$ for every $i \le m$, 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), $\|\bar{A}_i\| = \|A_i P^\top\| = \|PA_i\| \ge (1-\varepsilon)\|A_i\|$. Since $\|A_i\| = 1$ by Assumption 3, we conclude that $\|\bar{A}_i\| \ge 1-\varepsilon$. By Eq. (23) we have $\forall i \le m \quad (1-\varepsilon)^2 r - R\varepsilon \le \hat{b}_i$. This yields

$$\min_i \hat{b}_i \ge (1-\varepsilon)^2 r - R\varepsilon. \tag{24}$$

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. (24) 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 $i \le m$. 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. (22). We let w' be the optimal solution of RP_ε after the variable change, so that $w = w' + u_0$. We define $\tilde{w} = \sigma w' + u_0$ and note that it is feasible since $\bar{A}_i \tilde{w} = \sigma \bar{A}_i w' + \bar{A}_i u_0 \le \hat{b}_i + \bar{A}_i u_0 = b_i$. We therefore replace w in Eq. (22) by $w' + u_0$ and ρ by σ (since $\sigma \le \rho$), which yields:

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

which concludes the proof. □

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 dual of the QP P (Eq. (1)) is as follows:

$$\left.\begin{array}{rl} \mathsf{D} \equiv \min\limits_{y,v} & -y^\top Q y + b^\top v \\ & A^\top v - Qy = c \\ & v \ge 0, \end{array}\right\} \tag{25}$$

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

$$\left.\begin{array}{rl} \mathsf{RD} \equiv \min\limits_{z,v} & -z^\top \bar{Q} z + b^\top v \\ & \bar{A}^\top v - \bar{Q} z = \bar{c} \\ & v \ge 0, \end{array}\right\} \tag{26}$$

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

4.5 Theorem
Let 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

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

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

Proof. First, by Prop. 2.1 we have $\mathsf{v}(\mathsf{RP}) \le \mathsf{v}(\mathsf{P})$. 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

$$\mathsf{RP}' \equiv \max\{u^\top \bar{Q} u + \bar{c}^\top u \mid \bar{A}u \le b + A(P^\top u' - x^*) \ \wedge\ u \in \mathbb{R}^d\}, \tag{27}$$

and its dual

$$\mathsf{RD}' \equiv \min\{-z^\top \bar{Q} z + (b^\top + A(P^\top u' - x^*))v \mid \bar{A}^\top v - \bar{Q}z = \bar{c} \wedge v \ge 0 \wedge z \in \mathbb{R}^d\}. \tag{28}$$

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

$$(u')^\top \bar{Q} u' + \bar{c}^\top u' \le \mathsf{v}(\mathsf{RP}'). \tag{29}$$

We now bound the left hand side of Eq. (29) from below using Lemmata 3.1 and 3.3 and the right hand side from above using QP duality. For the lower bound, we have:

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

For the upper bound, by weak duality we have $\mathsf{v}(\mathsf{RP}') \le \mathsf{v}(\mathsf{RD}')$. 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

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

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

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

We obtain

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

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

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

We remark that $E_2 = \langle A^\top \hat{v}\,, P^\top u' - x^*\rangle$. By Cauchy-Schwartz, $E_2 \le \|A^\top \hat{v}\|_1 \, \|P^\top u' - x^*\|_\infty$. By Lemma 3.2, every component of $P^\top u' - x^*$ is bounded above by $\varepsilon \|x^*\|_2$, hence

$$E_1 + E_2 \le 3\varepsilon R^2 \|Q\|_F + \varepsilon R\|c\| + \varepsilon \|x^*\|_2 \|A^\top \hat{v}\|_1.$$

By writing $E_2 = \langle \hat{v}\,, A(P^\top u' - x^*)\rangle$ we derive $E_2 \le \varepsilon \|x^*\|_2 \|\hat{v}\|_1$, since all the rows of A have unit norm. □

Although the bound in Thm. 4.5 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 Thm. 4.4. This is a more satisfactory situation in view of Thm. 4.2.

5 Computational experiments

Most papers about RPs are entirely theoretical, although several exceptions exist (e.g. [22, 4]). RPs are also used in practice in a variety of application settings [3, 16, 28, 18]. 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 64GB RAM, each CPU of which has 8 cores (Intel Xeon CPU E5-2620 v4@2.10GHz). We used Python 3.7 as an interface to the QP barrier algorithm from IBM-ILOG CPLEX 12.8 [8] with default configuration. In all cases, the QP barrier algorithm was given the origin as a starting point.

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, 12].

All of the results of this paper actually hold (unchanged) also for *sub-gaussian* RPs [23, §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 $\mathsf{N}(0, \bar{\sigma})$, where the standard deviation is $\bar{\sigma} = \frac{1}{\sqrt{d\delta}}$, with some given probability $\delta \in (0, 1)$, 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.

5.1 Proposition
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 $\mathsf{N}(0, \bar{\sigma})$ with probability δ, and is equal to zero with probability $1 - \delta$, 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 $\mathsf{N}(0, 1/\sqrt{\delta})$, which has zero mean and has tail bounded by a negative exponential function [23, 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 $\sqrt{d(\delta)}P$ is distributed as $\mathsf{B}(\delta)\mathsf{N}(0, 1)$, where $B(\delta)$ is a Bernoulli distribution with parameter δ. Finally, if $i = j$ we have

$$\begin{aligned} \mathbb{E}(\rho_i^2) &= \mathbb{E}(\rho_i^2) - \mathbb{E}^2(\rho_i) & \text{because } \mathbb{E}(\rho_i) = 0 \\ &= \mathsf{Var}(\rho_i) = \frac{\mathsf{Var}(\mathsf{B}(\delta)\mathsf{N}(0,1))}{\delta}, \end{aligned}$$

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

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

Hence $\mathbb{E}(\rho_i^2) = 1$ as claimed. Now the result follows by [23, §9.3.1]. □

5.2 Random instances

Our first computational test is on randomly generated feasible instances of Eq. (1) with Q negative semidefinite. We make this assumption in order to compute guaranteed global maxima in acceptable CPU times for comparison purposes: our projection technique is independent of the convexity of the objective function. Our theoretical analysis is also mostly independent of convexity, aside from those bound results for which convexity is explicitly assumed (e.g. Thm. 4.5).

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 $[-\frac{1}{n\sqrt{n}}, \frac{1}{n\sqrt{n}}]$ (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 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 [15, §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.

			m	
#	n	`random`	`pairs`	`cuberot`
6	1000	100	200	2100
6	1000	1000	2000	3000
6	2000	100	200	4100
6	2000	1000	2000	5000
6	3000	100	200	6100
6	3000	1000	2000	7000
6	4000	100	200	8100
6	4000	1000	2000	9000

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

5.3 Portfolio optimization instances

We consider a realistic variant of the classic portfolio optimization problem [14]. 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, ...), 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:

$$\forall p \le q \quad \sum_{j \in T_p} x_j \le \tau_p. \tag{30}$$

The unbiased estimator for a covariance matrix is $\frac{1}{n-1} \sum_{j \le n} (x_j - \bar{x})(x_j - \bar{x})^\top$ (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: $-1 \le x \le 1$, $-1 \le \sum_j x_j \le 1$, and Eq. (30).

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.

instance	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
m	2102	2302	2502	2602	2702	2902	4102	4302	4502	4602	4702	4902
instance	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
m	6102	6302	6502	6602	6702	6902	8102	8302	8502	8602	8702	8902

Table 2: Number of variables and constraints on the portfolio 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 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, and can only exceed one if the signs of $f^*, \bar{f}$ differ:

$$\mathsf{r} = \frac{|f^* - \bar{f}|}{\max(|f^*|, |\bar{f}|)}.$$

Clearly, better performances yield r values closer to zero.

The CPU ratio $\mathsf{c} = \mathsf{CPU}_{\mathsf{RP}}/\mathsf{CPU}_{\mathsf{P}}$ 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 write the instance to disk in AMPL [6] format, have AMPL read it, pass it to CPLEX, and solve it;
- in the case of the projected instances, the time taken to project the instance data, write it to disk as a projected instance in AMPL format, have AMPL read it, pass it to CPLEX, solve it, and compute the retrieved solution (see Prop. 4.1).

We set up timings this way because, with large-scale instances, the time taken to read and format instances may also be considerable w.r.t. the time taken to solve it.

5.4.1 Random QPs

We first report aggregate results for $\mathsf{r}, \mathsf{c}, \mathsf{CPU}_{\mathsf{P}}$ in Table 3. We immediately note that the `cuberot`

	all			`random`			`pairs`			`cuberot`		
	r	c	$\mathsf{CPU}_{\mathsf{P}}$	r	c	$\mathsf{CPU}_{\mathsf{P}}$	r	c	$\mathsf{CPU}_{\mathsf{P}}$	r	c	$\mathsf{CPU}_{\mathsf{P}}$
mean	0.625	0.44	176.80	0.659	0.47	42.58	0.634	0.43	60.42	0.582	0.43	427.39
stdev	0.197	0.39	310.26	0.144	0.49	43.11	0.174	0.36	59.49	0.249	0.31	434.78
min	0.251	0.07	2.35	0.402	0.07	2.35	0.251	0.07	3.05	0.266	0.09	18.82
max	1.593	1.93	2082.39	0.813	1.93	186.15	0.944	1.32	259.07	1.593	1.33	2082.39

Table 3: Aggregate results for QPs over random polytopes.

instances globally take way more time than the other classes: this is due to the large number of their constraints.

A finer analysis shows that $\mathsf{r} > 1$ (i.e. the signs of $f^*, \bar{f}$) only change for some of the `cuberot` instances with $R = 0.5$. 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

	all		random		pairs		cuberot	
	r	c	r	c	r	c	r	c
mean	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

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

around 20% on average.

The results in Tables 3-4 appear to be very good w.r.t. the CPU time. The maximum values of $c > 1$ (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 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. 2. The mean of r appears to increase logarithmically with n. Since the

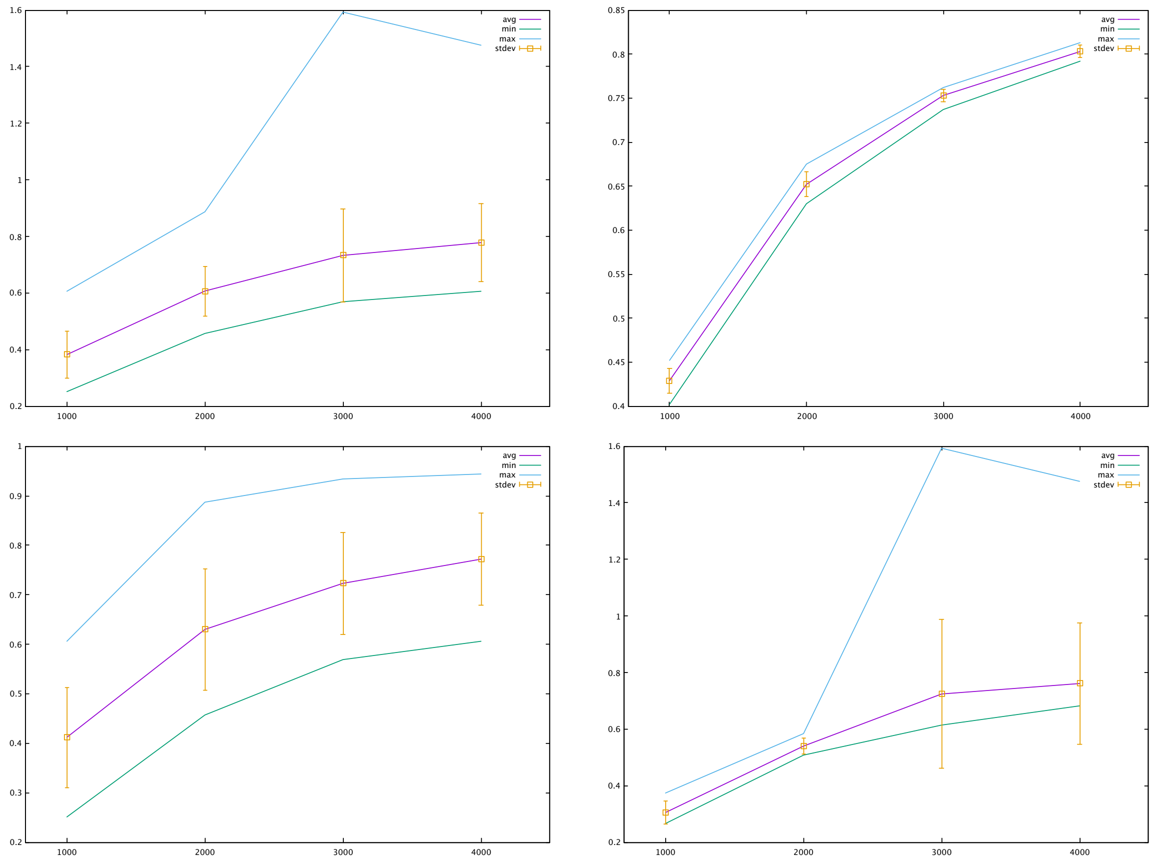

Figure 2: Mean, standard deviation, minimum, maximum values of r versus n, for all instances (top left), `random` (top right), `pairs` (bottom left), `cuberot` (bottom right).

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 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. [25]). Unfortunately, it seems that the application of RPs to QPs does not exhibit this property.

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

	r	c	$\mathsf{CPU_P}$
mean	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

Table 5: Aggregate results for portfolio instances.

but not r. Again, we refine the analysis of r: we plot its mean, standard deviation, minimum and maximum versus n in Fig. 3. We observe the same behaviour as with random QPs: r appears to be proportional to the number of variables of the projected problem.

5.6 Comparison with the results from [24]

In terms of objective function value, the results in Tables 3-4 are worse than those obtained in [24].

Although in [24] 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 [24] we also discussed results for randomly projected QPs over polyhedra (and a single unit ball constraint), and the reported results were better.

On the other hand, the results from [24] displayed an infeasibility error w.r.t. the norm constraint. Moreover, 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, it appears that our techniques provide a good 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

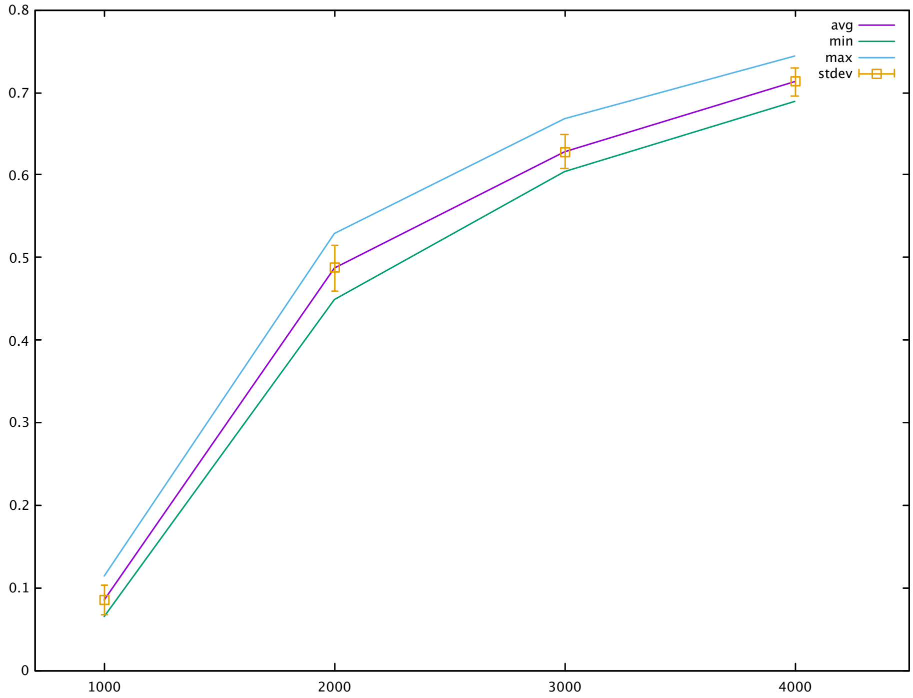

Figure 3: Mean, standard deviation, minimum, maximum values of r versus n for portfolio optimization instances.

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.

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 instances we were able to derive the optimal solution this way (projection then improvement) in globally less time than with a straight CPLEX solve of the original problem.

We would also like to warn readers on 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 dense, projected matrices $\bar{Q}$. 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. 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.

5.8 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 [29, 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 failed. We, however, went ahead and computed aggregated results for random QPs (Sect. 5.2) and portfolio instances (Sect. 5.3), given in Table 6. The plot of mean, standard deviation, minimum and maximum versus n for

	all		random		pairs		cuberot	
	r	c	r	c	r	c	r	c
mean	0.302	0.44	0.094	0.48	0.268	0.41	0.543	0.42
stdev	0.282	0.39	0.059	0.49	0.256	0.35	0.258	0.31
min	0.006	0.06	0.006	0.06	0.052	0.07	0.237	0.06
max	1.596	1.96	0.244	1.96	0.934	1.34	1.596	1.31

	Portfolio	
	r	c
mean	0.273	0.86
stdev	0.056	0.41
min	0.121	0.33
max	0.349	1.69

Table 6: Aggregate results for r, c for QPs over random polytopes (left) and portfolio optimization problems (right) with different scaling for the quadratic term.

QPs over random polytopes is given in Fig. 4. The corresponding plot for portfolio optimization is given in Fig. 5. 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. The scaled objective value difference between original and projected problem appears to increases logarithmically (empirically) in function of the number of variables of the original problem.

References

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

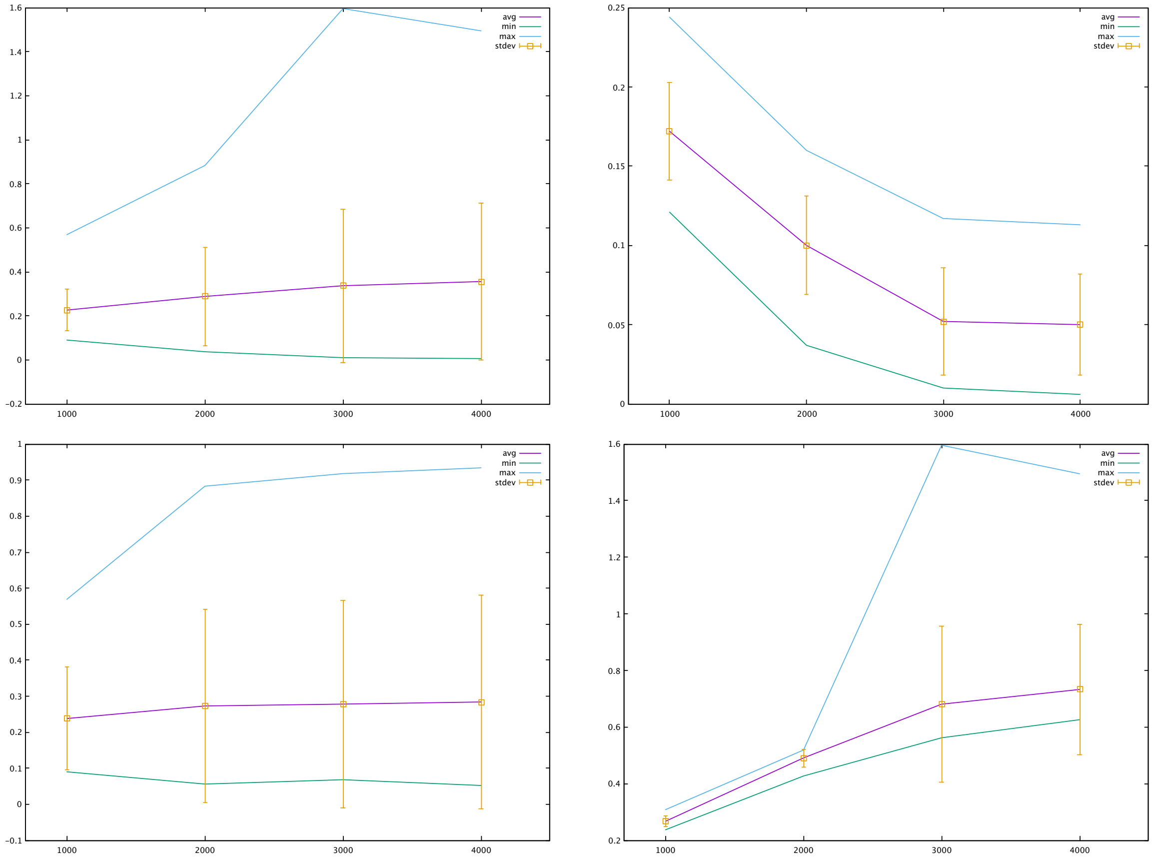

Figure 4: Mean, standard deviation, minimum, maximum values of `r` versus n, for all instances (top left), `random` (top right), `pairs` (bottom left), `cuberot` (bottom right), using a different scaling for the quadratic term.

[2] N. Ailon and B. Chazelle. Approximate nearest neighbors and fast Johnson-Lindenstrauss lemma. In *Proceedings of the Symposium on the Theory Of Computing*, volume 06 of *STOC*, Seattle, 2006. ACM.

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

[4] S. Dasgupta. Experiments with random projection. In *Proceedings of the 16th Conference on Uncertainty in Artificial Intelligence*, pages 143–151, San Francisco, 2000. Morgan Kaufman.

[5] W. Dorn. Duality in Quadratic Programming. *Quarterly of Applied Mathematics*, 18(2):155–162, 1960.

[6] R. Fourer and D. Gay. *The AMPL Book*. Duxbury Press, Pacific Grove, 2002.

[7] N. Gould and P. Toint. A Quadratic Programming bibliography. Technical Report 2000-1, RAL Numerical Analysis Group, oct 2001.

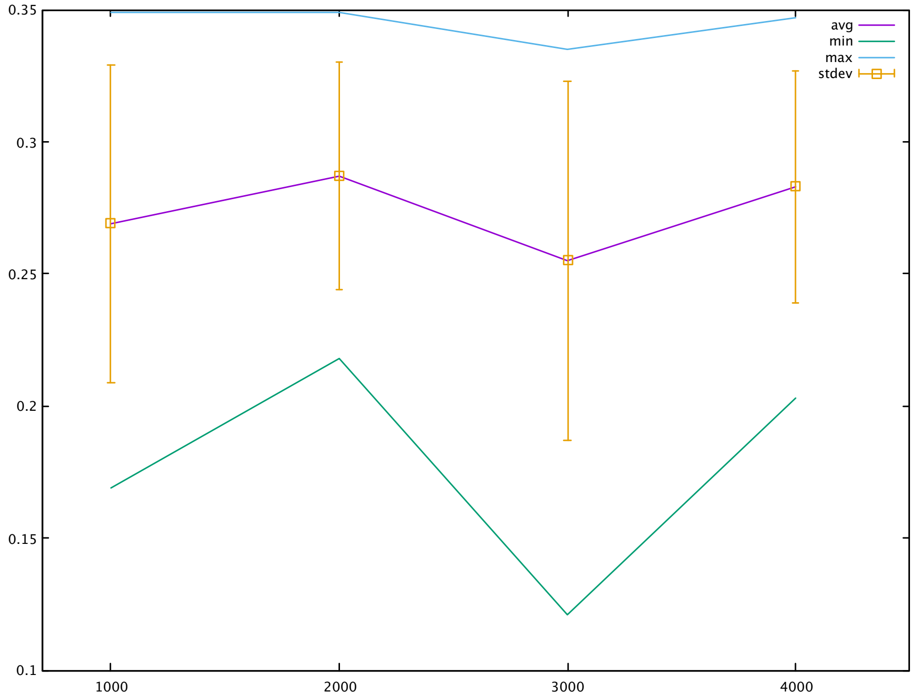

Figure 5: Mean, standard deviation, minimum, maximum values of r versus n.

[8] IBM. *ILOG CPLEX 12.8 User's Manual.* IBM, 2017.

[9] P. Indyk and R. Motwani. Approximate nearest neighbors: towards removing the curse of dimensionality. In *Proceedings of the Symposium on the Theory Of Computing*, volume 30 of *STOC*, pages 604–613, New York, 1998. ACM.

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

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

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

[13] L. Liberti and K. Vu. Barvinok's naive algorithm in distance geometry. *Operations Research Letters*, 46:476–481, 2018.

[14] H. Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952.

[15] J. May and R. Smith. Random polytopes: their definition, generation and aggregate properties. *Mathematical Programming*, 24:39–54, 1982.

[16] S. Paul, C. Boutsidis, M. Magdon-Ismail, and P. Drineas. Random projections for linear support vector machines. *ACM Transactions on Knowledge Discovery from Data*, 8(4):22:1–22:25, 2014.

[17] M. Pilanci and M. Wainwright. Randomized sketches of convex programs with sharp guarantees. In *International Symposium on Information Theory (ISIT)*, pages 921–925, Piscataway, 2014. IEEE.

[18] M. Pilanci and M. Wainwright. Newton sketch: A linear time optimization algorithm with linear-quadratic convergence. *SIAM Journal on Optimization*, 27(1):205–245, 2017.

[19] J. Shim. A survey of quadratic programming applications to business and economics. *International Journal of Systems Science*, 14(1):105–115, 1983.

[20] S. Vavasis. Quadratic programming is in NP. *Information Processing Letters*, 36:73–77, 1990.

[21] S. Vempala. *The Random Projection Method*. Number 65 in DIMACS Series in Discrete Mathematics and Theoretical Computer Science. American Mathematical Society, Providence, RI, 2004.

[22] S. Venkatasubramanian and Q. Wang. The Johnson-Lindenstrauss transform: An empirical study. In *Algorithm Engineering and Experiments*, volume 13 of *ALENEX*, pages 164–173, Providence, 2011. SIAM.

[23] R. Vershynin. *High-dimensional geometry*. Cambridge University Press, Cambridge, 2018.

[24] K. Vu, P.-L. Poirion, C. D'Ambrosio, and L. Liberti. Random projections for quadratic programs over a Euclidean ball. In A. Lodi et al., editor, *Integer Programming and Combinatorial Optimization (IPCO)*, LNCS, New York, 2019. Springer.

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

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

[27] D. Woodruff. Sketching as a tool for linear algebra. *Foundations and Trends in Theoretical Computer Science*, 10(1-2):1–157, 2014.

[28] J. Yang, X. Meng, and M. Mahoney. Quantile regression for large-scale applications. *SIAM Journal of Scientific Computing*, 36(5):S78–S110, 2014.

[29] L. Zhang, M. Mahdavi, R. Jin, T. Yang, and S. Zhu. Recovering the optimal solution by dual random projection. In S. Shalev-Shwartz and I. Steinwart, editors, *Conference on Learning Theory (COLT)*, volume 30 of *Proceedings of Machine Learning Research*, pages 135–157. `jmlr.org`, 2013.