

# Random projections for trust region subproblems

KY VU<sup>1</sup>, PIERRE-LOUIS POIRION<sup>2</sup>, CLAUDIA D’AMBROSIO<sup>3</sup>, LEO LIBERTI<sup>3</sup>

1. *Chinese University of Hong Kong*
2. *Huawei Research Center, Paris*
3. *CNRS LIX, École Polytechnique, F-91128 Palaiseau, France*  
Email: {dambrosio, liberti}@lix.polytechnique.fr

June 8, 2017

## Abstract

The trust region method is an algorithm traditionally used in the field of derivative free optimization. The method works by iteratively constructing surrogate models (often linear or quadratic functions) to approximate the true objective function inside some neighborhood of a current iterate. The neighborhood is called “trust region” in the sense that the model is trusted to be good enough inside the neighborhood. Updated points are found by solving the corresponding trust region subproblems. In this paper, we describe an application of random projections to solving trust region subproblems approximately.

## 1 Introduction

Derivative free optimization (DFO) (see [6, 11]) is a field of optimization including techniques for solving optimization problems in the absence of derivatives. This often occurs in the presence of *black-box functions*, i.e. functions for which no compact or computable representation is available. It may also occur for computable functions having unwieldy representations (e.g. a worst-case exponential-time or simply inefficient evaluation algorithm). A DFO problem in general form is defined as follows:

$$\min \{f(x) \mid x \in \mathcal{D}\},$$

where  $\mathcal{D}$  is a subset of  $\mathbb{R}^n$  and  $f(\cdot)$  is a continuous function such that no derivative information about it is available [5].

As an example, consider a simulator  $f(p, y)$  where  $p$  are parameters and  $y$  are state variables indexed by timesteps  $t$  (so  $y = (y^t \mid t \leq h)$ ), including some given boundary conditions  $y^0$ . The output of the simulator is a vector at each timestep  $t \leq h$ . We denote this output in function of parameters and state variables as  $f^t(p, y)$ . We also suppose that we collect some actual measurements from the process being simulated, say  $\hat{f}^t$  for some  $t \in H$ , where  $H$  is a proper index subset of  $\{1, \dots, h\}$ . We would like to choose  $p$  such that the behaviour of the simulator is as close as possible to the observed points  $\hat{f}^t$  for  $t \in H$ . The resulting optimization problem is:

$$\min \left\{ \bigoplus_{t \in H} \|\hat{f}^t - f^t(p, y)\| \mid p \in \mathcal{P} \wedge y \in \mathcal{Y} \right\},$$

where  $\mathcal{P}$  and  $\mathcal{Y}$  are appropriate domains for  $p, y$  respectively,  $\|\cdot\|$  is a given norm, and  $\oplus$  is either  $\sum$  or  $\max$ .

This is known to be a black-box optimization problem whenever  $f(\cdot)$  is given as an *oracle*, i.e. the value of  $f$  can be obtained given inputs  $p, y$ , but no other estimate can be obtained directly. In particular, the derivatives of  $f(\cdot)$  w.r.t.  $p$  and  $y$  are assumed to be impossible to obtain. Note that the lack of derivatives essentially implies that it is hard to define and therefore compute local optima. This obviously makes it even harder to compute global optima. This is why most methods in DFO focuses on finding local optima.

Trust Region (TR) methods are considered among the most suitable methods for solving DFO problems [4, 6, 14]. TR methods involve the construction of surrogate models to approximate the true function (locally) in “small” subsets  $D \subset \mathcal{D}$ , and rely on those models to search for optimal solutions within  $D$ . Such subsets are called *trust regions* because the surrogate models are “trusted” to be “good enough” limited to  $D$ . TRs are often chosen to be closed balls  $B(c, r)$  (with  $c$  center and  $r$  radius) with respect to some norms. There are several ways to obtain new data points, but the most common way is to find them as minima of the current model over the current trust region. Formally, one solves a sequence of so-called *TR subproblems*:

$$\min \{m(x) \mid x \in B(c, r) \cap \mathcal{D}\},$$

where  $m(\cdot)$  is a surrogate model of the true objective function  $f(\cdot)$ , which will then be evaluated at the solution of the TR subproblems. Depending on the discrepancy between the model and the true objective, the balls  $B(c, r)$  and the models  $m(\cdot)$  are updated: the TRs can change radius and/or center. The iterative application of this idea yields a TR method (we leave the explanation of TR methods somewhat vague since it is not our main focus; see [6] for more information).

In this paper we assume that the TR subproblem is defined by a linear or quadratic model  $m(\cdot)$  and that  $\mathcal{D}$  is a full-dimensional polyhedron defined by a set of linear inequality constraints. After appropriate scaling, the TR subproblem can be written as:

$$\min \{x^\top Qx + c^\top x \mid Ax \leq b, \|x\| \leq 1\}, \quad (1)$$

where  $Q \in \mathbb{R}^{n \times n}$  ( $Q$  is not assumed to be positive semidefinite),  $A \in \mathbb{R}^{m \times n}$ ,  $b \in \mathbb{R}^m$ . Here,  $\|\cdot\|$  refers to Euclidean norm and such notation will be used consistently throughout the paper. This problem has been studied extensively [1, 8, 17, 2]. Specifically, its complexity status varies between **P** and **NP-hard**, meaning it is **NP-hard** in general [15], but many polynomially solvable sub-cases have been discovered. For example, without linear constraints  $Ax \leq b$  the TR subproblem can be solved in polynomial time [21]; in particular, it can be rewritten as a Semidefinite Programming (SDP) problem. When adding one linear constraint or two parallel constraints, it still remains polynomial-time solvable [22, 18]; see the introduction to [1] for a more comprehensive picture.

## 1.1 The issue...

In practice, the theoretical complexity of the TR subproblem Eq. (1) is a moot point, since the time taken by each function evaluation  $f(p, x)$  is normally expected to exceed the time taken to solve Eq. (1). Since the solution sought is not necessarily global, any local Nonlinear Programming (NLP) solver [7, 3] can be deployed on Eq. (1) to identify a locally optimal solution. The issue we

address in this paper arises in cases when the sheer number of variables in Eq. (1) prevents even local NLP solvers from converging to some local optima in acceptable solution times. The usefulness of TR methods is severely hampered if the TR subproblem solution phase represents a non-negligible fraction of the total solution time. As a method for addressing the issue we propose to solve the TR subproblem *approximately* in exchange for speed. We justify our choice because the general lack of optimality guarantees in DFO and black-box function optimization makes it impossible to evaluate the loss one would incur in trading off local optimality guarantees in the TR subproblem for an approximate solution. As an approximation method, we consider *random projections*.

Random projections are simple but powerful tools for dimension reduction [20, 12, 16, 13]. They are often constructed as random matrices sampled from some given distribution classes. The simplest examples are matrices sampled componentwise from independently identically distributed (i.i.d.) random variables with Gaussian  $\mathcal{N}(0, 1)$ , uniform  $[-1, 1]$  or Rademacher  $\pm 1$  distributions. Despite their simplicity, random projections are competitive with more popular dimensional reduction methods such as Principal Component Analysis (PCA) / Multi-Dimensional Scaling (MDS) [10], Isomap [19] and more. One of the most important features of a random projection 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 random projection (such as those introduced above), then for any  $x \in \mathbb{R}^n$  and  $\varepsilon \in (0, 1)$ , we have

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

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 random projections is the so-called *Johnson-Lindenstrauss lemma* [9]. 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  $d = O(\frac{\log |X|}{\varepsilon^2})$ , such that

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

Such a mapping  $F$  can be found as a realization of the random projection  $P$  above; and the existence of the correct mapping is shown (by the probabilistic method) using the union bound. Moreover, the probability of sampling a correct mapping is also very high, i.e. in practice there is often no need to re-sample  $P$ .

## 1.2 ... and how we address it

The object of this paper is the applicability of random projections to TR subproblems Eq. (1). Let  $P \in \mathbb{R}^{d \times n}$  be a random projection with i.i.d.  $\mathcal{N}(0, 1)$  entries. We want to “project” each vector  $x \in \mathbb{R}^n$  to a lower-dimension vector  $Px \in \mathbb{R}^d$  and study the following *projected problem*

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

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

$$\min_{u \in \text{Im}(P)} \{u^\top (PQP^\top) u + \bar{c}^\top u \mid \bar{A}u \leq b, \|u\| \leq 1\}, \quad (3)$$

where  $\text{Im}(P)$  is the image space generated by  $P$ . Intuitively, since  $P$  is a projection from a (supposedly very high dimensional) space to a lower dimensional space, 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:

$$\min_{u \in \mathbb{R}^d} \{u^\top (PQP^\top)u + \bar{c}^\top u \mid \bar{A}u \leq b, \|u\| \leq 1\}, \quad (4)$$

where  $u$  ranges in  $\mathbb{R}^d$ . As we will show later, Eq. (4) yields a good approximate solution of the TR subproblem with very high probability.

## 2 Random projections for linear and quadratic models

In this section, we will explain the motivations for the study of the projected problem (4).

We start with the following simple lemma, which says that linear and quadratic models can be approximated well using random projections.

### 2.1 Approximation results

#### 2.1 Lemma

Let  $P : \mathbb{R}^n \rightarrow \mathbb{R}^d$  be a random projection satisfying Eq. (2) and let  $0 < \varepsilon < 1$ . Then there is a universal constant  $C_0$  such that

(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  $1 - 4e^{-C_0\varepsilon^2d}$ .

(ii) For any  $x \in \mathbb{R}^n$  and  $A \in \mathbb{R}^{m \times n}$  whose rows are unit vectors:

$$Ax - \varepsilon \|x\| \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix} \leq AP^\top Px \leq Ax + \varepsilon \|x\| \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix}$$

with probability at least  $1 - 4me^{-C_0\varepsilon^2d}$ .

(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  $1 - 8ke^{-C_0\varepsilon^2d}$ , we have:

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

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

*Proof.*

(i) Let  $C_0$  be the same universal constant (denoted by  $\mathcal{C}$ ) in Eq. (2). By the property in Eq. (2), for

any two vectors  $u + v$ ,  $u - v$  and using the union bound, we have

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

with probability at least  $1 - 4e^{-C_0\varepsilon^2d}$ . Apply this result for  $u = \frac{x}{\|x\|}$  and  $v = \frac{y}{\|y\|}$ , we obtain the desired inequality.

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

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

The claim follows by applying Part (i) and the union bound.

(iii) Let  $Q = U\Sigma V^\top$  be the Singular Value Decomposition (SVD) of  $Q$ . Here  $U, V$  are  $(n \times k)$ -real matrices with orthogonal unit column vectors  $u_1, \dots, u_k$  and  $v_1, \dots, v_k$ , respectively and  $\Sigma = \text{diag}(\sigma_1, \dots, \sigma_k)$  is a diagonal real matrix with positive entries. Denote by  $\mathbf{1}_k = (1, \dots, 1)^\top$  the  $k$ -dimensional column vector of all 1 entries. Since

$$\begin{aligned} x^\top P^\top P Q P^\top P y &= (U^\top P^\top P x)^\top \Sigma (V^\top P^\top P y) \\ &= [U^\top x + U^\top (P^\top P - \mathbb{I}_n)x]^\top \Sigma [V^\top y + V^\top (P^\top P - \mathbb{I}_n)y] \end{aligned}$$

then the two inequalities that

$$(U^\top x - \varepsilon\|x\|\mathbf{1}_k)^\top \Sigma (V^\top y - \varepsilon\|y\|\mathbf{1}_k) \leq x^\top P^\top P Q P^\top P y \leq (U^\top x + \varepsilon\|x\|\mathbf{1}_k)^\top \Sigma (V^\top y + \varepsilon\|y\|\mathbf{1}_k)$$

occurs with probability at least  $1 - 8ke^{-C\varepsilon^2d}$  (by applying part (ii) and the union bound). Moreover

$$\begin{aligned} (U^\top x - \varepsilon\|x\|\mathbf{1}_k)^\top \Sigma (V^\top y - \varepsilon\|y\|\mathbf{1}_k) &= x^\top Q y - \varepsilon\|x\|(\mathbf{1}_k^\top \Sigma V^\top y) - \varepsilon\|y\|(x^\top U \Sigma \mathbf{1}_k) + \varepsilon^2\|x\|\|y\| \sum_{i=1}^k \sigma_i \\ &= x^\top Q y - \varepsilon(\sigma_1, \dots, \sigma_k)(\|x\|V^\top y + \|y\|U^\top x) + \varepsilon^2\|x\|\|y\| \sum_{i=1}^k \sigma_i, \end{aligned}$$

and

$$(U^\top x + \varepsilon\|x\|\mathbf{1}_k)^\top \Sigma (V^\top y + \varepsilon\|y\|\mathbf{1}_k) = x^\top Q y + \varepsilon(\sigma_1, \dots, \sigma_k)(\|x\|V^\top y + \|y\|U^\top x) + \varepsilon^2\|x\|\|y\| \sum_{i=1}^k \sigma_i.$$

Therefore,

$$\begin{aligned} |x^\top P^\top P Q P^\top P y - x^\top Q y| &\leq \|x\|\|y\| \left( 2\varepsilon \sqrt{\sum_{i=1}^k \sigma_i^2} + \varepsilon^2 \sum_{i=1}^k \sigma_i \right) \\ &\leq 3\varepsilon\|x\|\|y\| \sum_{i=1}^k \sigma_i = 3\varepsilon\|x\|\|y\| \|Q\|_* \end{aligned}$$

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

It is known that singular values of random matrices often concentrate around their expectations. In the case when the random matrix is sampled from Gaussian ensembles, this phenomenon is well-understood due to many current research efforts. The following lemma, which is proved in [23], 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 the identity matrix.

### 2.2 Lemma ([23])

Let  $P \in \mathbb{R}^{d \times n}$  be a random matrix in which each entry is an i.i.d  $\mathcal{N}(0, \frac{1}{\sqrt{n}})$  random variable. 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

$$n \geq \frac{(d+1) \log(\frac{2d}{\delta})}{C_1 \varepsilon^2}, \quad (5)$$

where  $\|\cdot\|_2$  is the spectral norm of the matrix and  $C_1 > \frac{1}{4}$  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

$$\|P^\top u\|^2 - \|u\|^2 = \langle P^\top u, P^\top u \rangle - \langle u, u \rangle = \langle (PP^\top - I)u, u \rangle \in \text{range}(-\varepsilon\|u\|^2, \varepsilon\|u\|^2),$$

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

Condition (5) is not difficult to satisfy, since  $d$  is often very small compared to  $n$ . It suffices that  $n$  should be large enough to dominate the effect of  $\frac{1}{\varepsilon^2}$ .

## 2.2 Trust region subproblems with linear models

We will first work with a simple case, i.e. when the surrogate models used in TR methods are linear:

$$\min \{c^\top x \mid Ax \leq b, \|x\| \leq 1, x \in \mathbb{R}^n\}. \quad (6)$$

We will establish a relationship between problem (6) and its corresponding projected problem:

$$\min \{(Pc)^\top u \mid AP^\top u \leq b, \|u\| \leq 1 - \varepsilon, u \in \mathbb{R}^d\}. \quad (P_\varepsilon^-)$$

Note that, in the above problem, we shrink the unit ball by  $\varepsilon$ . We first obtain the following feasibility result:

### 2.3 Theorem

Let  $P \in \mathbb{R}^{d \times n}$  be a random matrix in which each entry is an i.i.d  $\mathcal{N}(0, \frac{1}{\sqrt{n}})$  random variable. Let  $\delta \in (0, 1)$ . Assume further that

$$n \geq \frac{(d+1) \log(\frac{2d}{\delta})}{C_1 \varepsilon^2},$$

for some universal constant  $\mathcal{C}_1 > \frac{1}{4}$ . Then with probability at least  $1 - \delta$ , for any feasible solution  $u$  of the projected problem  $(P_\varepsilon^-)$ ,  $P^\top u$  is also feasible for the original problem (6).

We remark the following universal property of Theorem 2.3: with a fixed probability, feasibility holds for all (instead of a specific) vectors  $u$ .

*Proof.* Let  $\mathcal{C}_1$  be as in Lemma 2.2. Let  $u$  be any feasible solution for the projected problem  $(P_\varepsilon^-)$  and take  $\hat{x} = P^\top u$ . Then we have  $A\hat{x} = AP^\top u \leq b$  and

$$\|\hat{x}\|^2 = \|P^\top u\|^2 = u^\top P^\top P u = u^\top u + u^\top (P^\top P - I)u \leq (1 + \varepsilon)\|u\|^2,$$

with probability at least  $1 - \delta$  (by Lemma 2.2). This implies that  $\|\hat{x}\| \leq (1 + \varepsilon/2)\|u\|$ ; and since  $\|u\| \leq 1 - \varepsilon$ , we have

$$\|\hat{x}\| \leq (1 + \frac{\varepsilon}{2})(1 - \varepsilon) < 1,$$

with probability at least  $1 - \delta$ , which proves the theorem.  $\square$

In order to estimate the quality of the objective function value, we define another projected problem, which can be considered as a relaxation of  $(P_\varepsilon^-)$  (we enlarge the feasible set by  $\varepsilon$ ):

$$\min \{(Pc)^\top u \mid AP^\top u \leq b + \varepsilon, \|u\| \leq 1 + \varepsilon, u \in \mathbb{R}^d\}. \quad (P_\varepsilon^+)$$

Intuitively, these two projected problems  $(P_\varepsilon^-)$  and  $(P_\varepsilon^+)$  are very close to each other when  $\varepsilon$  is small enough (under some additional assumptions, such as the “fullness” of the original polyhedron). Moreover, to make our discussions meaningful, we need to assume that they are feasible (in fact, it is enough to assume the feasibility of  $(P_\varepsilon^-)$  only).

Let  $u_\varepsilon^-$  and  $u_\varepsilon^+$  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 (1). We will bound  $c^\top x^*$  between  $c^\top x_\varepsilon^-$  and  $c^\top x_\varepsilon^+$ , the two values that are expected to be approximately close to each other.

## 2.4 Theorem

Let  $P \in \mathbb{R}^{d \times n}$  be a random matrix in which each entry is an i.i.d  $\mathcal{N}(0, \frac{1}{\sqrt{n}})$  random variable. Let  $\delta \in (0, 1)$ . Then there are universal constants  $\mathcal{C}_0 > 1$  and  $\mathcal{C}_1 > \frac{1}{4}$  such that if the two following conditions

$$d \geq \frac{\log(m/\delta)}{\mathcal{C}_0 \varepsilon^2}, \quad \text{and} \quad n \geq \frac{(d+1) \log(\frac{2d}{\delta})}{\mathcal{C}_1 \varepsilon^2},$$

are satisfied, we will have:

- (i) With probability at least  $1 - \delta$ , the solution  $x_\varepsilon^-$  is feasible for the original problem (1).
- (ii) With probability at least  $1 - \delta$ , we have:

$$c^\top x_\varepsilon^- \geq c^\top x^* \geq c^\top x_\varepsilon^+ - \varepsilon \|c\|.$$

*Proof.* Select  $\mathcal{C}_0$  and  $\mathcal{C}_1$  as in Lemma 2.1 and Lemma 2.2. Note that the second condition is the requirement for the Lemma 2.2 to hold, and the first condition is equivalent to  $m e^{-\mathcal{C}_0 \varepsilon^2 d} \leq \delta$ . Therefore, we can choose the universal constant  $\mathcal{C}_0$  such that  $1 - (4m + 6) e^{-\mathcal{C}_0 \varepsilon^2 d} \geq 1 - \delta$ .

(i) From the previous theorem, with probability at least  $1 - \delta$ , for any feasible point  $u$  of the projected problem  $(P_\varepsilon^-)$ ,  $P^\top u$  is also feasible for the original problem (6). Therefore, it must hold also for  $x_\varepsilon^-$ .

(ii) From part (i), with probability at least  $1 - \delta$ ,  $x_\varepsilon^-$  is feasible for the original problem (1). Therefore, we have

$$c^\top x_\varepsilon^- \geq c^\top x^*,$$

with probability at least  $1 - \delta$ .

Moreover, due to Lemma 2.1, with probability at least  $1 - 4e^{-\mathcal{C}_0\varepsilon^2d}$ , we have

$$c^\top x^* \geq c^\top P^\top P x^* - \varepsilon \|c\| \|x^*\| \geq c^\top P^\top P x^* - \varepsilon \|c\|,$$

(the last inequality follows from  $\|x^*\| \leq 1$ ). On the other hand, let  $\hat{u} = P x^*$ , due to Lemma 2.1, we have

$$AP^\top \hat{u} = AP^\top P x^* \leq Ax^* + \varepsilon \|x^*\| \begin{bmatrix} 1 \\ \cdots \\ 1 \end{bmatrix} \leq Ax^* + \varepsilon \begin{bmatrix} 1 \\ \cdots \\ 1 \end{bmatrix} \leq b + \varepsilon,$$

with probability at least  $1 - 4me^{-\mathcal{C}_0\varepsilon^2d}$ , and

$$\|\hat{u}\| = \|P x^*\| \leq (1 + \varepsilon) \|x^*\| \leq (1 + \varepsilon),$$

with probability at least  $1 - 2e^{-\mathcal{C}_0\varepsilon^2d}$ , by Property (2). Therefore,  $\hat{u}$  is a feasible solution for the problem  $(P_\varepsilon^+)$  with probability at least  $1 - (4m + 2)e^{-\mathcal{C}_0\varepsilon^2d}$ .

Since  $u_\varepsilon^+$  is the optimal solution for the problem  $(P_\varepsilon^+)$ , it follows that

$$c^\top x^* \geq c^\top P^\top P x^* - \varepsilon \|c\| = c^\top P^\top \hat{u} - \varepsilon \|c\| \geq c^\top P^\top u_\varepsilon^+ - \varepsilon \|c\| = c^\top x_\varepsilon^+ - \varepsilon \|c\|,$$

with probability at least  $1 - (4m + 6)e^{-\mathcal{C}_0\varepsilon^2d}$ , which is at least  $1 - \delta$  for our chosen universal constant  $\mathcal{C}_0$ .  $\square$

We have showed that  $c^\top x^*$  is bounded between  $c^\top x_\varepsilon^-$  and  $c^\top x_\varepsilon^+$ . Now we will estimate the gap between these two bounds. First of all, as stated at the outset, we assume that the feasible set

$$S^* = \{x \in \mathbb{R}^n \mid Ax \leq b, \|x\| \leq 1\}$$

is full-dimensional. This is a natural assumption since otherwise the polyhedron  $\{x \in \mathbb{R}^n \mid Ax \leq b\}$  is almost surely not full-dimensional either.

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 the our assumption, we have  $\text{FULL}(S^*) = r^* > 0$ , where  $r^*$  is the radius of the greatest ball inscribed in  $S^*$  (see Fig. 1).

The following lemma characterizes the fullness of  $S_\varepsilon^+$  with respect to  $r^*$ , where

$$S_\varepsilon^+ = \{u \in \mathbb{R}^d \mid AP^\top u \leq b + \varepsilon, \|u\| \leq 1 + \varepsilon\},$$

that is, the feasible set of the problem  $(P_\varepsilon^+)$ .

## 2.5 Lemma

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

$$\text{FULL}(S_\varepsilon^+) \geq (1 - \varepsilon)r^*.$$

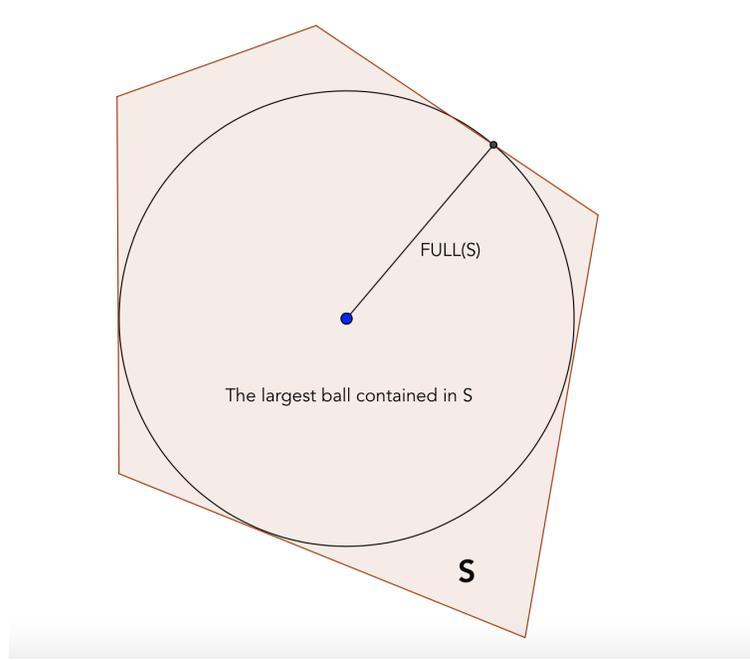


Figure 1: Fullness of a set.

In the proof of this lemma, we will extensively use 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.

*Proof.* For any  $i \in \{1, \dots, n\}$ , let  $A_i$  denotes the  $i$ th row of  $A$ . Let  $B(x_0, r^*)$  be a closed ball contained in  $S^*$ . Then for any  $x \in \mathbb{R}^n$  with  $\|x\| \leq r^*$ , we have  $A(x_0 + x) = Ax_0 + Ax \leq b$ , which implies that for any  $i \in \{1, \dots, n\}$ ,

$$b_i \geq (Ax_0)_i + \sup_{\|x\| \leq r^*} A_i x = (Ax_0)_i + r^* \|A_i\| = (Ax_0)_i + r^*, \quad (7)$$

hence

$$b \geq Ax_0 + r^* \quad \text{or equivalently,} \quad Ax_0 \leq b - r^*.$$

By Lemma 2.1, with probability at least  $1 - \delta$ , we have

$$AP^\top Px_0 \leq Ax_0 + \varepsilon \leq b - r^* + \varepsilon.$$

Let  $u \in \mathbb{R}^n$  with  $\|u\| \leq (1 - \varepsilon)r^*$ , then for any  $i \in \{1, \dots, n\}$ , by the above inequality we have:

$$\begin{aligned} (AP^\top(Px_0 + u))_i &= (AP^\top Px_0)_i + (AP^\top u)_i \\ &\leq b_i + \varepsilon - r^* + (AP^\top)_i u \\ &= b_i + \varepsilon - r^* + A_i P^\top u \end{aligned}$$

where  $(AP^\top)_i$  denotes the  $i$ th row of  $AP^\top$ . Since it holds for all such vector  $u$ , hence, by Cauchy-Schwartz inequality,

$$(AP^\top(Px_0 + u))_i \leq b_i + \varepsilon - r^* + (1 - \varepsilon)r^*\|A_iP^\top\|.$$

Using Eq. (2) and the union bound, we can see that with probability at least  $1 - 2me^{-c\varepsilon^2d} \geq 1 - \delta$ , we have: for all  $i \in \{1, \dots, m\}$ ,  $\|A_iP^\top\| \leq (1 + \varepsilon)\|A_i\| = (1 + \varepsilon)$ . Hence

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

with probability at least  $1 - 2\delta$ .

In other words, with probability at least  $1 - 2\delta$ , the closed ball  $B^*$  centered at  $Px_0$  with radius  $(1 - \varepsilon)r^*$  is contained in  $\{u \mid AP^\top u \leq b + \varepsilon\}$ .

Moreover, since  $B(x_0, r^*)$  is contained in  $S^*$ , which is the subset of the unit ball, then  $\|x_0\| \leq 1 - r^*$ .

With probability at least  $1 - \delta$ , for all vectors  $u$  in  $B(Px_0, (1 - \varepsilon)r^*)$ , we have

$$\|u\| \leq \|Px_0\| + (1 - \varepsilon)r^* \leq (1 + \varepsilon)\|x_0\| + (1 - \varepsilon)r^* \leq (1 + \varepsilon)(1 - r^*) + (1 - \varepsilon)r^* \leq 1 + \varepsilon.$$

Therefore, by the definition of  $S_\varepsilon^+$  we have

$$B(Px_0, (1 - \varepsilon)r^*) \subseteq S_\varepsilon^+,$$

which implies that the fullness of  $S_\varepsilon^+$  is at least  $(1 - \varepsilon)r^*$ , with probability at least  $1 - 3\delta$ .  $\square$

Now we will estimate the gap between the two objective functions of the problems  $(P_\varepsilon^+)$  and  $(P_\varepsilon^-)$  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.

## 2.6 Theorem

With probability at least  $1 - 2\delta$ , we have

$$c^\top x_\varepsilon^+ \leq c^\top x_\varepsilon^- \leq c^\top x_\varepsilon^+ + \frac{18\varepsilon}{\text{FULL}(S^*)}\|c\|.$$

*Proof.*

Let  $B(u_0, r_0)$  a closed ball with maximum radius that is contained in  $S_\varepsilon^+$ .

In order to establish the relation between  $u_\varepsilon^+$  and  $u_\varepsilon^-$ , our idea is to move  $u_\varepsilon^+$  closer to  $u_0$  (defined in the above lemma), so that the new point is contained in  $S_\varepsilon^-$ . Therefore, its objective value will be at least that of  $u_\varepsilon^-$ , but quite close to the objective value of  $u_\varepsilon^+$ .

We define  $\hat{u} := (1 - \lambda)u_\varepsilon^+ + \lambda u_0$  for some  $\lambda \in (0, 1)$  specified later. We want to find  $\lambda$  such that  $\hat{u}$  is feasible for  $P_\varepsilon^-$  while its corresponding objective value is not so different from  $c^\top x_\varepsilon^+$ .

Since for all  $\|u\| \leq r_0$ :

$$AP^\top(u_0 + u) = AP^\top u_0 + AP^\top u \leq b + \varepsilon,$$

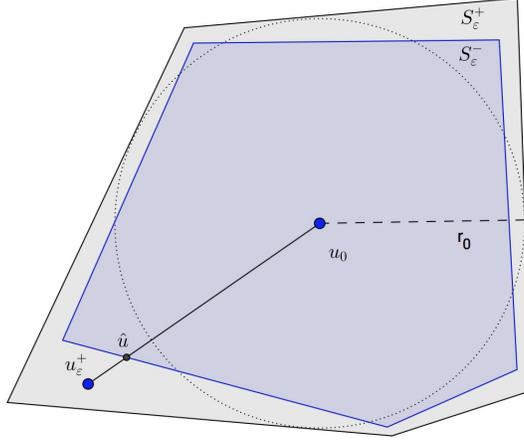


Figure 2: Idea of the proof

then, similarly to Eq. (7),

$$AP^\top u_0 \leq b + \varepsilon - r_0 \begin{pmatrix} \|A_1 P^\top\| \\ \vdots \\ \|A_m P^\top\| \end{pmatrix}.$$

Therefore, we have, with probability at least  $1 - \delta$ ,

$$AP^\top u_0 \leq b + \varepsilon - r_0(1 - \varepsilon) \begin{pmatrix} \|A_1\| \\ \vdots \\ \|A_m\| \end{pmatrix} = b + \varepsilon - r_0(1 - \varepsilon).$$

Hence

$$AP^\top \hat{u} = (1 - \lambda)AP^\top u_\varepsilon^+ + \lambda AP^\top u_0 \leq b + \varepsilon - \lambda r_0(1 - \varepsilon) \leq b + \varepsilon - \frac{1}{2}\lambda r_0,$$

as we can assume w.l.o.g. that  $\varepsilon \leq \frac{1}{2}$ . Hence,  $AP^\top \hat{u} \leq b$  if we choose  $\varepsilon \leq \frac{\lambda r_0}{2}$ . Moreover, let  $u \in B(u_0, r_0)$  such that  $u$  is collinear with  $0$  and  $u_0$ , and such that with  $\|u\| = r_0$ ; we then have

$$\|u_0\| \leq \|u_0 + u\| - r_0 \leq 1 + \varepsilon - r_0,$$

so that

$$\|\hat{u}\| \leq (1 - \lambda)\|u_\varepsilon^+\| + \lambda\|u_0\| \leq (1 - \lambda)(1 + \varepsilon) + \lambda(1 + \varepsilon - r_0) = 1 + \varepsilon - \lambda r_0,$$

which is less than or equal to  $1 - \varepsilon$  for

$$\varepsilon \leq \frac{1}{2}\lambda r_0.$$

We now can choose  $\lambda = 2\frac{\varepsilon}{r_0}$ . With this choice,  $\hat{u}$  is a feasible point for the problem  $P_\varepsilon^-$ . Therefore, we have

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

By the above Lemma, we know that  $r_0 \geq (1 - \varepsilon)r^*$ , therefore, we have:

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

with probability at least  $1 - \delta$ .

The claim of the theorem now follows, by noticing that, when  $0 \leq \varepsilon \leq \frac{1}{2}$ , we can simplify:

$$\frac{2(1 + \varepsilon)^2}{(1 - \varepsilon)} \leq \frac{2(1 + \frac{1}{2})^2}{(1 - \frac{1}{2})} = 9.$$

□

### 2.3 Trust region subproblems with quadratic models

In this subsection, we consider the case when the surrogate models using in TR methods are quadratic and defined as follows:

$$\min \{x^\top Qx + c^\top x \mid Ax \leq b, \|x\| \leq 1, x \in \mathbb{R}^n\}. \quad (8)$$

Similar to the previous section, we also study the relations between this and two other problems:

$$\min \{u^\top PQP^\top u + (Pc)^\top u \mid AP^\top u \leq b, \|u\| \leq 1 - \varepsilon, u \in \mathbb{R}^d\} \quad (Q_\varepsilon^-)$$

and

$$\min \{u^\top PQP^\top u + (Pc)^\top u \mid AP^\top u \leq b + \varepsilon, \|u\| \leq 1 + \varepsilon, u \in \mathbb{R}^d\}. \quad (Q_\varepsilon^+)$$

We will just state the following feasibility result, as the proof is very similar to that of Thm. 2.3.

#### 2.7 Theorem

Let  $P \in \mathbb{R}^{d \times n}$  be a random matrix in which each entry is an i.i.d  $\mathcal{N}(0, \frac{1}{\sqrt{n}})$  random variable. Let  $\delta \in (0, 1)$ . Assume further that

$$n \geq \frac{(d + 1) \log(\frac{2d}{\delta})}{\mathcal{C}_1 \varepsilon^2},$$

for some universal constant  $\mathcal{C}_1 > \frac{1}{4}$ . Then with probability at least  $1 - \delta$ , for any feasible solution  $u$  of the projected problem  $(Q_\varepsilon^-)$ ,  $P^\top u$  is also feasible for the original problem (8).

Let  $u_\varepsilon^-$  and  $u_\varepsilon^+$  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 (8). We will bound  $x^{*\top} Qx^* + c^\top x^*$  between  $x_\varepsilon^{-\top} Qx_\varepsilon^- + c^\top x_\varepsilon^-$  and  $x_\varepsilon^{+\top} Qx_\varepsilon^+ + c^\top x_\varepsilon^+$ , the two values that are expected to be approximately close to each other.

**2.8 Theorem**

Let  $P \in \mathbb{R}^{d \times n}$  be a random matrix in which each entry is an i.i.d  $\mathcal{N}(0, \frac{1}{\sqrt{n}})$  random variable. Let  $\delta \in (0, 1)$ . Let  $x^*$  be an optimal solution for the original problems (8). Then there are universal constants  $\mathcal{C}_0 > 1$  and  $\mathcal{C}_1 > \frac{1}{4}$  such that if the two following conditions

$$d \geq \frac{\log(m/\delta)}{\mathcal{C}_0 \varepsilon^2}, \quad \text{and} \quad n \geq \frac{(d+1) \log(\frac{2d}{\delta})}{\mathcal{C}_1 \varepsilon^2},$$

are satisfied, we will have:

- (i) With probability at least  $1 - \delta$ , the solution  $x_\varepsilon^-$  is feasible for the original problem (8).
- (ii) With probability at least  $1 - \delta$ , we have:

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

*Proof.* The constants  $\mathcal{C}_0$  and  $\mathcal{C}_1$  are chosen in the same way as before.

(i) From the previous theorem, with probability at least  $1 - \delta$ , for any feasible point  $u$  of the projected problem  $(P_\varepsilon^-)$ ,  $P^\top u$  is also feasible for the original problem (8). Therefore, it must hold also for  $x_\varepsilon^-$ .

(ii) From part (i), with probability at least  $1 - \delta$ ,  $x_\varepsilon^-$  is feasible for the original problem (8). Therefore, we have

$$x_\varepsilon^{-\top} Q x_\varepsilon^- + c^\top x_\varepsilon^- \geq x^{*\top} Q x^* + c^\top x^*,$$

with probability at least  $1 - \delta$ .

Moreover, due to Lemma 2.1, with probability at least  $1 - 8(k+1)e^{-\mathcal{C}_0 \varepsilon^2 d}$ , where  $k$  is the rank of  $Q$ , we have

$$x^{*\top} Q x^* \geq x^{*\top} P^\top P Q P^\top P x^* - 3\varepsilon \|x^*\|^2 \|Q\|_* \geq x^{*\top} P^\top P Q P^\top P x^* - 3\varepsilon \|Q\|_*,$$

and

$$c^\top x^* \geq c^\top P^\top P x^* - \varepsilon \|c\| \|x^*\| \geq c^\top P^\top P x^* - \varepsilon \|c\|,$$

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

$$x^{*\top} Q x^* + c^\top x^* \geq x^{*\top} P^\top P Q P^\top P x^* + c^\top P^\top P x^* - \varepsilon \|c\| - 3\varepsilon \|Q\|_*,$$

On the other hand, let  $\hat{u} = P x^*$ , due to Lemma 2.1, we have

$$A P^\top \hat{u} = A P^\top P x^* \leq A x^* + \varepsilon \|x^*\| \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix} \leq A x^* + \varepsilon \begin{bmatrix} 1 \\ \dots \\ 1 \end{bmatrix} \leq b + \varepsilon,$$

with probability at least  $1 - 4m e^{-\mathcal{C}_0 \varepsilon^2 d}$ , and

$$\|\hat{u}\| = \|P x^*\| \leq (1 + \varepsilon) \|x^*\| \leq (1 + \varepsilon),$$

with probability at least  $1 - 2e^{-\mathcal{C}_0 \varepsilon^2 d}$  (by Lemma 2.1). Therefore,  $\hat{u}$  is a feasible solution for the problem  $(P_\varepsilon^+)$  with probability at least  $1 - (4m + 2)e^{-\mathcal{C}_0 \varepsilon^2 d}$ . Due to the optimality of  $u_\varepsilon^+$  for the

problem  $(P_\varepsilon^+)$ , it follows that

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

with probability at least  $1 - (4m + 6)e^{-\mathcal{C}_0 \varepsilon^{2d}}$ , which is at least  $1 - \delta$  for the chosen universal constant  $\mathcal{C}_0$ . Moreover,

$$c^\top x^* \geq c^\top P^\top P x^* - \varepsilon \|c\| = c^\top P^\top \hat{u} - \varepsilon \|c\| \geq c^\top P^\top u_\varepsilon^+ - \varepsilon \|c\| = c^\top x_\varepsilon^{+\top} - \varepsilon \|c\|,$$

Hence

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

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. For that, we also use the definition of fullness measure. We have the following result:

### 2.9 Theorem

Let  $0 < \varepsilon < 0.1$ . Then with probability at least  $1 - 4\delta$ , we have

$$x_\varepsilon^{+\top} Q x_\varepsilon^+ + c^\top x_\varepsilon^+ \leq x_\varepsilon^{-\top} Q x_\varepsilon^- + c^\top x_\varepsilon^- < x_\varepsilon^{+\top} Q x_\varepsilon^+ + c^\top x_\varepsilon^+ + \frac{\varepsilon}{\text{FULL}(S^*)} (36 + 18\|c\|).$$

*Proof.* Let  $B(u_0, r_0)$  be a closed ball with maximum radius that is contained in  $S_\varepsilon^+$ .

We define  $\hat{u} := (1 - \lambda)u_\varepsilon^+ + \lambda u_0$  for some  $\lambda \in (0, 1)$  specified later. We want to find “small”  $\lambda$  such that  $\hat{u}$  is feasible for  $Q_\varepsilon^-$  while its corresponding objective value is still close to  $x_\varepsilon^{+\top} Q x_\varepsilon^+ + c^\top x_\varepsilon^+$ .

Similar to the proof of Th. 2.6, when we choose  $\lambda = \frac{2\varepsilon}{r_0}$ , then  $\hat{u}$  is feasible for the problem  $Q_\varepsilon^-$  with probability at least  $1 - \delta$ .

Therefore,  $u_\varepsilon^{-\top} P Q P^\top u_\varepsilon^- + (Pc)^\top u_\varepsilon^-$  is smaller than or equal to

$$\begin{aligned} &\hat{u}^\top P Q P^\top \hat{u} + (Pc)^\top \hat{u} \\ &= (u_\varepsilon^+ + \lambda(u_0 - u_\varepsilon^+))^\top P Q P^\top (u_\varepsilon^+ + \lambda(u_0 - u_\varepsilon^+)) + (Pc)^\top \hat{u} \\ &= u_\varepsilon^{+\top} P Q P^\top u_\varepsilon^+ + \lambda u_\varepsilon^{+\top} P Q P^\top (u_0 - u_\varepsilon^+) + \lambda(u_0 - u_\varepsilon^+)^\top P Q P^\top u_\varepsilon^+ \\ &\quad + \lambda^2 (u_0 - u_\varepsilon^+)^\top P Q P^\top (u_0 - u_\varepsilon^+) + (Pc)^\top \hat{u}. \end{aligned}$$

However, from Lemma 2.1 and the Cauchy-Schwartz inequality, we have

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

Similar for other terms, then we have

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

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

$$\begin{aligned} \hat{u}^\top P Q P^\top \hat{u} &< u_\varepsilon^{+\top} P Q P^\top u_\varepsilon^+ + 16\lambda \|Q\|_2 \\ &= u_\varepsilon^{+\top} P Q P^\top u_\varepsilon^+ + \frac{32\varepsilon}{r_0} \quad (\text{since } \|Q\|_2 = 1) \\ &\leq u_\varepsilon^{+\top} P Q P^\top u_\varepsilon^+ + \frac{32\varepsilon}{(1 - \varepsilon)\text{FULL}(S^*)} \quad (\text{due to Lemma 2.5}) \\ &< u_\varepsilon^{+\top} P Q P^\top u_\varepsilon^+ + \frac{36\varepsilon}{\text{FULL}(S^*)} \quad (\text{since } \varepsilon \leq 0.1), \end{aligned}$$

with probability at least  $1 - 2\delta$ . Furthermore, similarly to the proof of Th. 2.6, we have

$$c^\top P^\top \hat{u} = (1 - \lambda)(Pc)^\top u_\varepsilon^+ + \lambda(Pc)^\top u_0 \leq c^\top P^\top u_\varepsilon^+ + \frac{4(1 + \varepsilon)^2 \varepsilon}{(1 - \varepsilon)r^*} \|c\| \leq c^\top P^\top u_\varepsilon^+ + \frac{18\varepsilon}{\text{FULL}(S^*)} \|c\|,$$

with probability at least  $1 - 2\delta$ . □

## Conclusion

In this paper, we have shown theoretically that random projection can be used in combination with trust region method to study high-dimensional derivative free optimization. We have proved that the solutions provided by solving low-dimensional projected version of trust region subproblems are good approximations of the true ones. We hope this provides a useful insight in the solution of very high dimensional derivative free problems.

As a last remark, we note in passing that the results of this paper extend to the case of arbitrary (non-unit and non-zero) trust region radii, through scaling. This observation also suggests that the LP and QP trust region problem formulations analyzed in this paper can inner- and outer-approximate arbitrary bounded LPs and QPs.

## References

- [1] D. Bienstock and A. Michalka. Polynomial solvability of variants of the trust-region subproblem. In *Proceedings of the 25th annual ACM Symposium on Discrete Algorithms*, volume 25 of *SODA*, pages 380–390, Philadelphia, 2014. ACM.
- [2] Samuel Burer and Boshi Yang. The trust region subproblem with non-intersecting linear constraints. *Mathematical Programming*, 149(1):253–264, 2015.
- [3] COIN-OR. *Introduction to IPOPT: A tutorial for downloading, installing, and using IPOPT*, 2006.

- [4] A. R. Conn, N. Gould, A. Sartenaer, and Ph. L. Toint. Convergence properties of minimization algorithms for convex constraints using a structured trust region. *SIAM Journal on Optimization*, 6(4):1059–1086, 1996.
- [5] A. R. Conn, K. Scheinberg, and Ph. L. Toint. Recent progress in unconstrained nonlinear optimization without derivatives. *Mathematical Programming*, 79(1):397, 1997.
- [6] A. R. Conn, K. Scheinberg, and L.N. Vicente. *Introduction to derivative-free optimization*. Society for Industrial and Applied Mathematics, Philadelphia, 2009.
- [7] P.E. Gill. *User's Guide for SNOPT 5.3*. Systems Optimization Laboratory, Department of EESOR, Stanford University, California, February 1999.
- [8] V. Jeyakumar and G. Y. Li. Trust-region problems with linear inequality constraints: Exact sdp relaxation, global optimality and robust optimization. *Math. Program.*, 147(1-2):171–206, October 2014.
- [9] 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. AMS.
- [10] I. Jolliffe. *Principal Component Analysis*. Springer, Berlin, 2nd edition, 2010.
- [11] Oliver Kramer, David Echeverría Ciaurri, and Slawomir Koziel. Derivative-free optimization. In Slawomir Koziel and Xin-She Yang, editors, *Computational Optimization, Methods and Algorithms*, pages 61–83. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.
- [12] V.K. Ky, P.L. Poirion, and L. Liberti. Using the johnson-lindenstrauss lemma in linear and integer programming. Technical Report 1507.00990, arXiv:, 2015.
- [13] L. Liberti, P.L. Poirion, and Vu Khac Ky. Solving LP using random projections. *Electronic Notes in Discrete Mathematics*, 55:53–56, 2016.
- [14] M. Marazzi and J. Nocedal. Wedge trust region methods for derivative free optimization. *Mathematical Programming*, 91(2):289–305, 2002.
- [15] K. Murty and S. Kabadi. Some NP-complete problems in quadratic and nonlinear programming. *Mathematical Programming*, 39:117–129, 1987.
- [16] Mert Pilanci and Martin J Wainwright. Randomized sketches of convex programs with sharp guarantees. In *Information Theory (ISIT), 2014 IEEE International Symposium on*, pages 921–925. IEEE, 2014.
- [17] M. Salahi and S. Fallahi. Trust region subproblem with an additional linear inequality constraint. *Optimization Letters*, 10(4):821–832, 2016.
- [18] Jos F Sturm and Shuzhong Zhang. On cones of nonnegative quadratic functions. *Mathematics of Operations Research*, 28(2):246–267, 2003.
- [19] J. Tenenbaum, V. de Silva, and J. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290:2319–2322, 2000.

- [20] David P. Woodruff. Sketching as a tool for numerical linear algebra. *Found. Trends Theor. Comput. Sci.*, 10:1–157, October 2014.
- [21] Yinyu Ye. A new complexity result on minimization of a quadratic function with a sphere constraint. In Christodoulos A. Floudas and Panos M. Pardalos, editors, *Recent Advances in Global Optimization*, pages 19–31. Princeton University Press, 1992.
- [22] Yinyu Ye and Shuzhong Zhang. New results on quadratic minimization. *SIAM Journal on Optimization*, 14(1):245–267, 2003.
- [23] 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*, volume 30 of *Journal of Machine Learning Research: Workshop and Conference Proceeding*, pages 135–157. JMLR, 2013.