An Optimal Interpolation Set for Model-Based Derivative-Free Optimization Methods

Tom M. Ragonneau*†   Zaikun Zhang*‡

31st January 2024

Abstract

This paper demonstrates the optimality of an interpolation set employed in derivative-free trust-region methods. This set is optimal in the sense that it minimizes the constant of well-poisedness in a ball centred at the starting point. It is chosen as the default initial interpolation set by many derivative-free trust-region methods based on underdetermined quadratic interpolation, including NEWUOA, BOBYQA, LINCOA, and COBYQA. Our analysis provides a theoretical justification for this choice.

Keywords   Derivative-free optimization · Model-based methods · Underdetermined quadratic interpolation · Derivative-free symmetric Broyden update · Well-poisedness

Mathematics Subject Classification (2010)   90C56 · 41A10 · 65K05 · 90C30

1   Introduction

Derivative-free optimization (DFO) methods solve the optimization problem

$$\min_{x \in \Omega \subseteq \mathbb{R}^n} f(x) \tag{1.1}$$

based on function evaluations without relying on derivative information. They are needed when classical or generalized derivatives of the objective function $f \colon \mathbb{R}^n \to \mathbb{R}$ are unavailable or too expensive to evaluate, which may also be the case for the constraint functions underlying the feasible set $\Omega \subseteq \mathbb{R}^n$. Examples of such problems arise from many areas, such as reinforcement learning [22], hyperparameter tuning [12], particle physics [10], and aircraft engineering [11]. The DFO literature is vast. For comprehensive overviews, we refer to the monographs [1,5], the survey papers [7,13–15,25], the recent thesis [23], and the references therein.

This paper studies a set employed as the initial interpolation set by many derivative-free trust-region methods [4,28] based on underdetermined quadratic interpolation, such as NEWUOA [19], and, when no constraints are present, BOBYQA [20], LINCOA, and COBYQA [23,24]. It is also employed by [3] in a method for constrained problems. We show that this interpolation set is optimal in the sense that it minimizes the constant of well-poisedness in a ball centred at the starting point, which is detailed in Theorem 3.2.

---

*Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China.

†Email: `tom.ragonneau@polyu.edu.hk`; ORCID: 0000-0003-2717-2876.

‡Email: `zaikun.zhang@polyu.edu.hk`; ORCID: 0000-0001-8934-8190.

In what follows, we denote by $\mathcal{Q}_n$ the space of polynomials on $\mathbb{R}^n$ of degree at most two, which will be referred to as quadratic polynomials for simplicity. The DFO methods that we consider maintain an interpolation set $\mathcal{Y}^k \subseteq \mathbb{R}^n$ at iteration k, and build a model $\hat{f}^k \in \mathcal{Q}_n$ for f according to the interpolation conditions

$$\hat{f}^k(y) = f(y), \quad y \in \mathcal{Y}^k. \tag{1.2}$$

Clearly, if the interpolation conditions (1.2) do not contradict each other, the model $\hat{f}^k$ is uniquely defined by (1.2) only if

$$\mathrm{card}(\mathcal{Y}^k) = \dim(\mathcal{Q}_n) = \frac{1}{2}(n+1)(n+2).$$

Such a scheme, referred to as the fully-determined quadratic interpolation, is used, for example, by UOBYQA [16]. However, the first iteration of this method requires $\mathcal{O}(n^2)$ function evaluations to construct the initial model $\hat{f}^0$. This is impracticable unless n is small. Therefore, modern interpolation-based DFO methods work with fewer interpolation points, giving rise to the underdetermined quadratic interpolation. To uniquely define $\hat{f}^k$ in this case, it is common to solve an interpolation problem of the form

$$\begin{aligned} \min_{Q \in \mathcal{Q}_n} \quad & \mathcal{F}^k(Q) \\ \text{s.t.} \quad & Q(y) = f(y), \quad y \in \mathcal{Y}^k, \end{aligned} \tag{1.3}$$

where $\mathcal{F}^k$ is a functional that promotes certain desired regularity of $\hat{f}^k$. For example, DFO [6] and MNH [26] take

$$\mathcal{F}^k(Q) = \left\| \nabla^2 Q \right\|_{\mathsf{F}}^2, \tag{1.4}$$

where $\|\cdot\|_{\mathsf{F}}$ denotes the Frobenius norm. In contrast, NEWUOA, BOBYQA, LINCOA, and COBYQA employ

$$\mathcal{F}^k(Q) = \left\| \nabla^2 Q - \nabla^2 \hat{f}^{k-1} \right\|_{\mathsf{F}}^2 \tag{1.5}$$

with $\hat{f}^{-1} = 0$. There are other functionals based on the ℓ_1-norm [2] or Sobolev seminorms of quadratic polynomials [27, 30]. The variational problem described by (1.3) and (1.5) defines the derivative-free symmetric Broyden update proposed by Powell (see [17, 21], [29, § 3.6], and [23, § 2.4.2]), with the functional (1.5) being inspired by the least-change property of quasi-Newton updates [8]. To ensure that (1.3) admits a solution that includes second-order information, one normally requires that

$$n + 2 \le \mathrm{card}(\mathcal{Y}^k) \le \frac{1}{2}(n+1)(n+2).$$

Finally, if the constraint functions underlying the feasible set Ω in (1.1) are nonlinear, we can approximate them using similar techniques, which is the case for COBYQA.

Most interpolation-based DFO methods do not build the interpolation point $\mathcal{Y}^k$ *ab initio* at each iteration. Instead, they choose an initial interpolation set $\mathcal{Y}^0$ and then update $\mathcal{Y}^k$ to obtain $\mathcal{Y}^{k+1}$. For example, only one point differs between $\mathcal{Y}^k$ and $\mathcal{Y}^{k+1}$ in Powell's trust-region DFO methods and COBYQA during usual iterations. This mechanism tries to make the best of the function evaluations, which are normally considered the principal cost of DFO methods. In such a case, it is clear that the initial interpolation set $\mathcal{Y}^0$ must be carefully chosen, as it will impact the optimization method for subsequent iterations. Powell devised an initial interpolation set for his methods, described in Subsection 3.1. Although the design of this set is natural, no

theoretical analysis has been provided to justify its configuration. This paper contributes such an analysis, showing that no other interpolation set is better in a well-poisedness sense [5, § 3.3].

This paper is organized as follows. Section 2 introduces the Λ-poisedness of interpolation sets in the minimum Frobenius norm sense. Section 3 then employs this notion to study the initial interpolation set used by Powell. We provide lower and upper bounds for the constant of well-poisedness of the set, evaluate this constant in a few special cases, and then demonstrate the optimality of this set under the default setting recommended by Powell. In addition, we raise some open questions motivated by our results. Section 4 concludes this paper.

2 Well-poisedness of interpolation sets

In this section, we consider an interpolation set

$$\mathcal{Y} = \{y^1, y^2, \ldots, y^m\} \subseteq \mathbb{R}^n$$

and the interpolation problem

$$
\begin{aligned}
\min_{Q \in \mathcal{Q}_n} \quad & \|\nabla^2 Q\|_\mathsf{F} \\
\text{s.t.} \quad & Q(y) = f(y), \quad y \in \mathcal{Y}.
\end{aligned}
\tag{2.1}
$$

This problem covers (1.3)–(1.4), and also covers (1.3) with (1.5) by a simple change of variable.

Definition 2.1 (Poisedness)**.** The set $\mathcal{Y}$ is *poised* in the minimum Frobenius norm sense if problem (2.1) has a unique solution for any real-valued function f.

Observe that the variational problem (2.1) is an equality-constrained quadratic programming problem with respect to the coefficients of Q. Thus, its KKT system is linear. The exact formulation of this system can be found in [17, 18]. The interpolation set $\mathcal{Y}$ is poised if this KKT system is uniquely solvable (see [17, § 2] and [5, § 5.3]). Intuitively, $\mathcal{Y}$ can be said well-poised if this system is well-conditioned. This intuition will be formalized in the sequel.

2.1 Minimum Frobenius norm Lagrange polynomials

To formally define a measure of well-poisedness of $\mathcal{Y}$, we first need to extend the classical definition of the Lagrange polynomials as follows.

Definition 2.2 (Minimum Frobenius norm Lagrange polynomials [5, Definition 5.1])**.** Assume that the interpolation set $\mathcal{Y}$ is poised. The *i*th *minimum Frobenius norm Lagrange polynomial* L_i for the interpolation problem (2.1), with $i \in \{1, 2, \ldots, m\}$, is the unique quadratic polynomial that solves

$$
\begin{aligned}
\min_{Q \in \mathcal{Q}_n} \quad & \|\nabla^2 Q\|_\mathsf{F} \\
\text{s.t.} \quad & Q(y^j) = \delta_{i,j}, \quad j \in \{1, 2, \ldots, m\},
\end{aligned}
$$

where $\delta_{i,j}$ denotes the Kronecker delta.

Remark that if $m = (n+1)(n+2)/2$, then Definition 2.2 reduces to the classical definition of the Lagrange polynomials. Moreover, it can be shown that the solution $\hat{f}$ to problem (2.1) is given by

$$\hat{f}(x) = \sum_{i=1}^m f(y^i) L_i(x)$$

3

for any $x \in \mathbb{R}^n$ [5, Lemma 5.2].

2.2 Well-poisedness in the minimum Frobenius norm sense

We are now equipped to define the notion of Λ-poisedness in the minimum Frobenius norm sense.

Definition 2.3 (Λ-poisedness in the minimum Frobenius norm sense [5, Definition 5.6]). Let Λ be a positive constant and $\mathcal{C} \subseteq \mathbb{R}^n$ be a compact set. The set $\mathcal{Y}$ is said to be Λ-*poised in the minimum Frobenius norm sense* in $\mathcal{C}$ if it is poised and

$$\max_{1 \leq i \leq m} \max_{x \in \mathcal{C}} |L_i(x)| \leq \Lambda.$$

One can show that $\mathcal{Y}$ is Λ-poised in the unit ball if and only if the condition number of the KKT system of (1.3) is bounded by some polynomials of Λ [5, Theorem 5.8]. Therefore, the notion of Λ-poisedness formalizes the intuition about well-poisedness mentioned above.

If $\mathcal{Y}$ is a poised interpolation set and $\mathcal{C} \subseteq \mathbb{R}^n$ is compact, then we will refer to

$$\Lambda_{\mathcal{C}}(\mathcal{Y}) = \max_{1 \leq i \leq m} \max_{x \in \mathcal{C}} |L_i(x)|$$

as the *constant of well-poisedness* of $\mathcal{Y}$ in $\mathcal{C}$.

3 Optimality of Powell's initial interpolation set

This section first presents the initial interpolation set devised by Powell for his trust-region DFO methods, and then analyzes its well-poisedness. We establish bounds on its constant of well-poisedness and evaluate this constant in some special cases. Finally, we point out that the default setting in Powell's methods renders an optimal interpolation set in terms of the well-poisedness in a ball centred at the starting point.

3.1 Description of the initial interpolation set

The initial interpolation that Powell designed for his methods is as follows. Without loss of generality, we assume that the starting point is at the origin. Suppose that $\Delta > 0$ is the initial trust-region radius. For $i \in \{1, 2, \ldots, 2n+1\}$, define

$$y^i = \begin{cases} 0, & \text{if } i = 1, \\ \Delta e_{i-1}, & \text{if } 2 \leq i \leq n+1, \\ -\Delta e_{i-n-1}, & \text{otherwise,} \end{cases} \tag{3.1}$$

where $e_i \in \mathbb{R}^n$ denotes the ith canonical coordinate vector in $\mathbb{R}^n$. Let m be the number of interpolation points. We focus on the case with $n+2 \leq m \leq 2n+1$ following Powell's suggestion.[1] The initial interpolation set is then chosen to be

$$\mathcal{Y}_m^0 = \{y^1, y^2, \ldots, y^m\}, \tag{3.2}$$

which can be found in [19, Equation (3.2)]. The default value for m proposed by Powell is $2n+1$. In other words, the default initial interpolation set is $\mathcal{Y}^0 = \mathcal{Y}_{2n+1}^0$. This is a natural choice,

---

[1]Powell's BOBYQA code contains a comment that "choices that exceed $2n+1$ are not recommended."

as the corresponding interpolation set is geometrically appealing. It consists of the origin and equidistant points to the origin in the positive and negative coordinate directions. In what follows, we provide a theory that justifies this natural choice, showing that $\mathcal{Y}_{2n+1}^0$ is indeed optimal in terms of well-poisedness.

3.2 Well-poisedness of the interpolation set

We now investigate the Λ-poisedness of $\mathcal{Y}_m^0$. We will do this in the closed ℓ_p-norm ball of radius Δ centred at the origin, namely

$$\mathcal{B}_p(\Delta) = \{x \in \mathbb{R}^n : \|x\|_p \leq \Delta\},$$

with $p \in [1, \infty]$. Note that we allow $p = \infty$, and that $\mathcal{B}_p(\Delta)$ is also the smallest ℓ_p-norm ball enclosing $\mathcal{Y}_m^0$. Trust-region DFO methods usually define the trust region by the Euclidean norm, so the case with $p = 2$ is the most interesting. However, it can be beneficial to define the trust region by polyhedral norms when bound or linear constraints are present, and $p = 1$ or $p = \infty$ will become more relevant.

According to Definition 2.3, the set $\mathcal{Y}_m^0$ is Λ_p-poised in $\mathcal{B}_p(\Delta)$ in the minimum Frobenius norm sense with

$$\Lambda_p = \max_{1 \leq i \leq m} \max_{\|x\|_p \leq \Delta} |L_i(x)|, \tag{3.3}$$

where L_i, for $i \in \{1, 2, \ldots, m\}$, is the ith minimum Frobenius norm Lagrange polynomial associated with $\mathcal{Y}_m^0$. Indeed, Λ_p is the constant of well-poisedness of $\mathcal{Y}_m^0$ in $\mathcal{B}_p(\Delta)$. We will focus on this constant in what follows.

3.2.1 Formulation of the Lagrange polynomials

To study Λ_p, we first present explicit formulae for L_i for all $i \in \{1, 2, \ldots, m\}$. These formulae are given in [19, § 3] without proof.

Lemma 3.1. *For all $x \in \mathbb{R}^n$ and $m \in \{n + 2, n + 3, \ldots, 2n + 1\}$, the expression of L_i, for $i \in \{1, 2, \ldots, m\}$, is given by*

$$L_i(x) = \begin{cases} 1 - \dfrac{1}{\Delta^2} \displaystyle\sum_{j=1}^{m-n-1} x_j^2 - \dfrac{1}{\Delta} \sum_{j=m-n}^{n} x_j, & \text{if } i = 1, \\[2ex] \dfrac{x_{i-1}^2}{2\Delta^2} + \dfrac{x_{i-1}}{2\Delta}, & \text{if } 2 \leq i \leq m - n, \\[2ex] \dfrac{x_{i-1}}{\Delta}, & \text{if } m - n + 1 \leq i \leq n + 1, \\[2ex] \dfrac{x_{i-n-1}^2}{2\Delta^2} - \dfrac{x_{i-n-1}}{2\Delta}, & \text{if } n + 2 \leq i \leq m. \end{cases} \tag{3.4}$$

Here, x_j denotes the jth entry of x for each $j \in \{1, 2, \ldots, n\}$, and we define $\sum_{j=m-n}^{n} x_j = 0$ in the formulation of L_1 if $m = 2n + 1$.

Proof. Let $i \in \{1, 2, \ldots, m\}$ be fixed and let L be a quadratic polynomial satisfying

$$L(y^j) = \delta_{i,j}, \quad j \in \{1, 2, \ldots, m\}. \tag{3.5}$$

First, it is straightforward to verify that L_i satisfies the interpolation conditions (3.5). Hence, it suffices to show that $\|\nabla^2 L_i\|_\mathsf{F} \leq \|\nabla^2 L\|_\mathsf{F}$.

Consider any $j \in \{1, 2, \ldots, m - n - 1\}$. Denote the jth diagonal entries of $\nabla^2 L$ and $\nabla^2 L_i$ by $(\nabla^2 L)_{j,j}$ and $(\nabla^2 L_i)_{j,j}$, respectively. According to equation (3.1), we have

$$y^1 = 0, \quad y^{j+1} = \Delta e_j, \quad \text{and} \quad y^{n+j+1} = -\Delta e_j.$$

Since L and L_i are quadratic polynomials sharing the same values on $\{y^1, y^{j+1}, y^{n+j+1}\}$, Taylor expansions of the quadratic polynomial L around y^1 yield

$$\left(\nabla^2 L\right)_{j,j} = \frac{L(y^{j+1}) + L(y^{n+j+1}) - 2L(y^1)}{\Delta^2} = \left(\nabla^2 L_i\right)_{j,j}.$$

On the other hand, it is easy to check according to (3.4) that all the entries of $\nabla^2 L_i$ are zero except for the first $m - n - 1$ diagonal entries. Therefore,

$$\left\|\nabla^2 L_i\right\|_\mathsf{F}^2 \leq \left\|\nabla^2 L\right\|_\mathsf{F}^2,$$

which completes the proof. $\qquad\square$

3.2.2 Bounds for the constant of well-poisedness

The next lemma simplifies the expression of Λ_p defined in (3.3) for further computations.

Lemma 3.2. *For any $m \in \{n + 2, n + 3, \ldots, 2n + 1\}$ and any $p \in [1, \infty]$, we have*

$$\Lambda_p = \max_{\|x\|_p \leq \Delta} \left|L_1(x)\right|. \tag{3.6}$$

Proof. According to Lemma 3.1, for each $i \in \{2, 3, \ldots, n + 1\}$, $L_i(x)$ only depends on x_{i-1} for all $x \in \mathbb{R}^n$, and hence

$$\max_{\|x\|_p \leq \Delta} \left|L_i(x)\right| = \max_{t \in [-\Delta, \Delta]} \left|L_i(te_{i-1})\right| = 1.$$

Similarly, for each $i \in \{n + 2, n + 3, \ldots, m\}$, since $L_i(x)$ only depends on x_{i-n-1} for all $x \in \mathbb{R}^n$, we have

$$\max_{\|x\|_p \leq \Delta} \left|L_i(x)\right| = \max_{t \in [-\Delta, \Delta]} \left|L_i(te_{i-n-1})\right| = 1.$$

Meanwhile, since $L_1(y^1) = 1$ and $y^1 \in \mathcal{B}_p(\Delta)$, we have

$$\max_{\|x\|_p \leq \Delta} \left|L_1(x)\right| \geq L_1(y^1) = 1.$$

Hence, (3.6) holds according to the definition of Λ_p in (3.3). $\qquad\square$

We are now equipped to develop lower and upper bounds for Λ_p. For convenience, we define henceforth

$$0^0 = 0 \quad \text{and} \quad \frac{\infty}{\infty} = 1.$$

Theorem 3.1. *For any $m \in \{n + 2, n + 3, \ldots, 2n + 1\}$ and any $p \in [1, \infty]$, we have*

$$1 + (2n + 1 - m)^{\frac{p-1}{p}} \leq \Lambda_p \leq n. \tag{3.7}$$

6

In particular, we have $\Lambda_\infty = \max\{n-1, 2n-m+2\}$.

Proof. We will establish the bounds in (3.7) using the formulation of Λ_p in Lemma 3.2. For the lower bound, by considering only the points in $\mathbb{R}^n$ whose leading $m - n - 1$ entries are zeros and whose remaining $2n + 1 - m$ entries are equal, we have

$$\Lambda_p = \max_{\|x\|_p \leq \Delta} |L_1(x)| \geq \max_{t \in \mathbb{R}} \left\{ 1 - \frac{1}{\Delta}(2n+1-m)t : (2n+1-m)^{\frac{1}{p}}|t| \leq \Delta \right\} = 1 + (2n+1-m)^{\frac{p-1}{p}}.$$

We now establish the upper bound. For any $p \geq 1$, we have $\mathcal{B}_p(\Delta) \subseteq \mathcal{B}_\infty(\Delta)$, so that $\Lambda_p \leq \Lambda_\infty$. Therefore, we only need to show that $\Lambda_\infty \leq \max\{n-1, 2n-m+2\} \leq n$. For any $x \in \mathcal{B}_\infty(\Delta)$,

$$-L_1(x) = -1 + \frac{1}{\Delta^2} \sum_{j=1}^{m-n-1} x_j^2 + \frac{1}{\Delta} \sum_{j=m-n}^{n} x_j \leq n - 1, \tag{3.8}$$

and

$$L_1(x) = 1 - \frac{1}{\Delta^2} \sum_{j=1}^{m-n-1} x_j^2 - \frac{1}{\Delta} \sum_{j=m-n}^{n} x_j \leq 1 - \frac{1}{\Delta} \sum_{j=m-n}^{n} x_j \leq 2n - m + 2. \tag{3.9}$$

Thus

$$\Lambda_\infty = \max_{\|x\|_\infty \leq \Delta} |L_1(x)| \leq \max\{n-1, 2n-m+2\},$$

which completes the proof of (3.7).

Finally, remark that the right-hand side in (3.8) is attained by the vector whose entries are all Δ. Moreover, the right-hand side in (3.9) is attained by the vector whose first $m-n-1$ entries are zero and whose remaining entries are all $-\Delta$. Hence, $\Lambda_\infty = \max\{n-1, 2n-m+2\}$. $\qquad\square$

3.2.3 Some special cases

Theorem 3.1 provides the value of Λ_∞, which is either $n-1$ or n since $m \geq n+2$. There are more cases where we can evaluate Λ_p, as we will detail in the following. First, when $1 \leq p \leq 2$, Λ_p actually equals the lower bound in (3.7).

Proposition 3.1. *For any $m \in \{n+2, n+3, \ldots, 2n+1\}$ and any $p \in [1, 2]$, we have*

$$\Lambda_p = 1 + (2n+1-m)^{\frac{p-1}{p}}. \tag{3.10}$$

Proof. According to Theorem 3.1, it suffices to show that the right-hand side of (3.10) is an upper bound for Λ_p. We will prove this using Lemma 3.2.

Consider any $x \in \mathcal{B}_p(\Delta)$. Note that $\|x\|_2 \leq \|x\|_p \leq \Delta$ for $p \in [1, 2]$. According to Lemma 3.1 and the Hölder inequality, we have

$$-L_1(x) = -1 + \frac{1}{\Delta^2} \sum_{j=1}^{m-n-1} x_j^2 + \frac{1}{\Delta} \sum_{j=m-n}^{n} x_j$$

$$\leq -1 + \frac{1}{\Delta^2} \|x\|_2^2 + \frac{1}{\Delta}(2n+1-m)^{\frac{p-1}{p}} \|x\|_p$$

$$\leq (2n+1-m)^{\frac{p-1}{p}}.$$

Similarly,

$$L_1(x) \le 1 + \frac{1}{\Delta} \sum_{j=m-n}^{n} |x_j| \le 1 + \frac{1}{\Delta}(2n+1-m)^{\frac{p-1}{p}} \|x\|_p \le 1 + (2n+1-m)^{\frac{p-1}{p}}.$$

Therefore, invoking Lemma 3.2, we have

$$\Lambda_p = \max_{\|x\|_p \le \Delta} |L_1(x)| \le 1 + (2n+1-m)^{\frac{p-1}{p}},$$

which concludes the proof. $\qquad\square$

We can also evaluate Λ_p under the default and natural setting $m = 2n+1$ recommended by Powell. To this end, recall the following elementary fact.

Lemma 3.3. *For any $p \in [1,\infty]$ and $q \in [1,\infty]$, we have*

$$\max_{\|x\|_p \le 1} \|x\|_q = \max\{1, n^{\frac{1}{q}-\frac{1}{p}}\}. \tag{3.11}$$

Proof. If p or q is infinity, then (3.11) is straightforward to verify. So, we assume that both of them are finite.

Consider the case where $p \le q$. For $x \in \mathcal{B}_p(1)$, we have $\|x\|_q^q \le \|x\|_p^p \le 1$, and this bound is attained at the first coordinate vector $e_1 \in \mathcal{B}_p(1)$, so that (3.11) holds in this case.

We now suppose that $p > q$. Let $\mathbb{1} \in \mathbb{R}^n$ denote the vector with all entries being one, $r = p/q$, and $s = r/(r-1) = p/(p-q)$. For $x \in \mathcal{B}_p(1)$, define $z = (|x_1|^q, |x_2|^q, \dots, |x_n|^q)$. According to the Hölder inequality, we have

$$\|x\|_q = \left(\mathbb{1}^{\mathsf{T}} z\right)^{\frac{1}{q}} \le \left(\|\mathbb{1}\|_s \|z\|_r\right)^{\frac{1}{q}} = n^{\frac{1}{sq}} \|x\|_p \le n^{\frac{p-q}{qp}}.$$

Moreover, this bound is attained at $x^* = n^{-\frac{1}{p}}\mathbb{1}$, which proves (3.11) for $p > q$. $\qquad\square$

Proposition 3.2. *For any $p \in [1,\infty]$, if $m = 2n+1$, then*

$$\Lambda_p = \max\left\{1, n^{\frac{p-2}{p}} - 1\right\}.$$

Proof. Since $m = 2n+1$, we have $L_1(x) = 1 - \Delta^{-2}\|x\|_2^2$ according to Lemma 3.1. Then, it is clear that

$$\max_{\|x\|_p \le \Delta} L_1(x) = \max_{\|x\|_p \le \Delta} \left(1 - \Delta^{-2}\|x\|_2^2\right) = 1. \tag{3.12}$$

Moreover, according to Lemma 3.3, we have

$$\max_{\|x\|_p \le \Delta} -L_1(x) = \max_{\|x\|_p \le \Delta} \left(\Delta^{-2}\|x\|_2^2 - 1\right) = \max_{\|z\|_p \le 1} \|z\|_2^2 - 1 = \max\{0, n^{\frac{p-2}{p}} - 1\}. \tag{3.13}$$

The desired result is obtained by combining (3.12) and (3.13) with Lemma 3.2. $\qquad\square$

Now we can show that $\mathcal{Y}_{2n+1}^0$ attains the minimal constant of well-poisedness in $\mathcal{B}_p(\Delta)$ among all interpolation sets that contain the origin, provided that $1 \le p \le 2$. In this sense, $\mathcal{Y}_{2n+1}^0$ is an optimal interpolation set in $\mathcal{B}_p(\Delta)$ for $p \in [1,2]$. Indeed, the upper bound for p can be slightly

larger than 2, as is detailed in Theorem 3.2. Recall that $p = 2$ is of interest for most trust-region DFO methods, including those by Powell.

Theorem 3.2. *Assume that $m = 2n + 1$, and that either $n \leq 2$ or*

$$1 \leq p \leq \frac{2 \log n}{\log(n/2)}. \tag{3.14}$$

If an interpolation set containing 0 is Λ-poised in $\mathcal{B}_p(\Delta)$, then $\Lambda \geq \Lambda_p$.

Proof. Under the assumptions, we have $\Lambda_p = 1$ according to Proposition 3.2. If an interpolation set containing 0 is Λ-poised in $\mathcal{B}_p(\Delta)$, then we have $\Lambda \geq 1$, because the Lagrange polynomial corresponding to 0 takes the value 1 at 0. Thus, the theorem holds. $\square$

It is worth mentioning that Proposition 3.1 also implies that $\mathcal{Y}_{2n+1}^0$ renders $\Lambda_p = 1$ for $p \in [1, 2]$. Theorem 3.2 shows that this equality holds for a larger range of p, because the right-hand side in (3.14) is greater than 2 when $n > 2$.

3.2.4 Remarks and open questions

Note that Λ_p can be regarded as a function of m. Theorem 3.2 implies that $m^* = 2n+1$ minimizes this function if p satisfies inequality (3.14). Thus, it is natural to ask whether $2n+1$ minimizes Λ_p for any $p \geq 1$, to which we do not have an answer yet.

The definition of $\mathcal{Y}_m^0$ in (3.1) and (3.2) assumes that $m \leq 2n + 1$. Even though larger values of m are not recommended in practice, Powell [19] proposed an extension of $\mathcal{Y}_m^0$ for $m > 2n + 1$. With such an extension, we can define Λ_p by (3.3) for any $m \in \{n+2, n+3, \ldots, (n+1)(n+2)/2\}$. It is interesting to ask whether $2n + 1$ still minimizes Λ_p for any $p \geq 1$ after the extension. We expect that the analysis will be more challenging than what we have done. One of the challenges is that Lemma 3.2 does not hold when $m > 2n + 1$, and hence, the estimations of Λ_p will become more involved.

Another interesting open question is whether an interpolation set containing the origin can be Λ-poised in $\mathcal{B}_p(\Delta)$ with $\Lambda < \Lambda_p$. Theorem 3.2 provides a negative answer when p satisfies (3.14). If the answer is negative for all $p \geq 1$, then the optimality of $2n + 1$ mentioned in the above two paragraphs is true, and $\mathcal{Y}_{2n+1}^0$ is an optimal interpolation set in $\mathcal{B}_p(\Delta)$ for all $p \geq 1$.

4 Conclusion

We have analyzed the well-poisedness (in the minimum Frobenius norm sense) of an interpolation set that appears in many trust-region DFO methods, particularly those by Powell [19, 20]. It is proved to be an optimal interpolation set under the default setting, because it minimizes the constant of well-poisedness in a ball centred at the starting point. Our analysis justifies the natural configuration of this set from the viewpoint of interpolation theory.

The spirit of our analysis is similar to that of [9], which proves that a widely used direction set is indeed optimal for directional direct search methods based on sufficient decrease. While proving nothing surprising, this kind of investigation deepens our understanding of certain algorithmic strategies that we often employ but rarely ask why.

References

[1] C. Audet and W. Hare. *Derivative-Free and Blackbox Optimization.* Springer Ser. Oper. Res. Financ. Eng. Springer, Cham, Switzerland, 2017.

[2] A. S. Bandeira, K. Scheinberg, and L. N. Vicente. Computation of sparse low degree interpolating polynomials and their application to derivative-free optimization. *Math. Program.*, 134:223–257, 2012.

[3] P. D. Conejo, E. W. Karas, and L. G. Pedroso. A trust-region derivative-free algorithm for constrained optimization. *Optim. Methods Softw.*, 30:1126–1145, 2015.

[4] A. R. Conn, N. I. M. Gould, and Ph. L. Toint. *Trust-Region Methods.* MPS-SIAM Ser. Optim. SIAM, Philadelphia, PA, USA, 2000.

[5] A. R. Conn, K. Scheinberg, and L. N. Vicente. *Introduction to Derivative-Free Optimization.* MPS-SIAM Ser. Optim. SIAM, Philadelphia, PA, USA, 2009.

[6] A. R. Conn, K. Scheinberg, and Ph. L. Toint. A derivative free optimization algorithm in practice. In *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, pages 129–139, St. Louis, MO, USA, 1998. AIAA.

[7] A. L. Custódio, K. Scheinberg, and L. N. Vicente. Methodologies and software for derivative-free optimization. In T. Terlaky, M. F. Anjos, and S. Ahmed, editors, *Advances and Trends in Optimization with Engineering Applications*, MOS-SIAM Ser. Optim., pages 495–506. SIAM, Philadelphia, PA, USA, 2017.

[8] J. E. Dennis, Jr. and R. B. Schnabel. Least change secant updates for quasi-Newton methods. *SIAM Rev.*, 21:443–459, 1979.

[9] M. Dodangeh, L. N. Vicente, and Z. Zhang. On the optimal order of worst case complexity of direct search. *Optim. Lett.*, 10:699–708, 2016.

[10] J. S. Eldred, J. Larson, M. Padidar, E. Stern, and S. M. Wild. Derivative-free optimization of a rapid-cycling synchrotron. *Optim. Eng.*, 24:1289–1319, 2023.

[11] A. Gazaix, F. Gallard, V. Ambert, D. Guénot, M. Hamadi, S. Grihon, P. Sarouille, T. Druot, J. Brezillon, V. Gachelin, J. Plakoo, N. Desfachelles, N. Bartoli, T. Lefebvre, S. Gürol,

B. Pauwels, C. Vanaret, and R. Lafage. Industrial application of an advanced bi-level MDO formulation to aircraft engine pylon optimization. In *AIAA Aviation 2019 Forum*, Dallas, TX, USA, 2019. AIAA.

[12] H. Ghanbari and K. Scheinberg. Black-box optimization in machine learning with trust region based derivative free algorithm. Technical Report 17T-005, COR@L, Bethlehem, PA, USA, 2017.

[13] J. Larson, M. Menickelly, and S. M. Wild. Derivative-free optimization methods. *Acta Numer.*, 28:287–404, 2019.

[14] M. J. D. Powell. A view of unconstrained minimization algorithms that do not require derivatives. *ACM Trans. Math. Software*, 1:97–107, 1975.

[15] M. J. D. Powell. Direct search algorithms for optimization calculations. *Acta Numer.*, 7:287–336, 1998.

[16] M. J. D. Powell. UOBYQA: unconstrained optimization by quadratic approximation. *Math. Program.*, 92:555–582, 2002.

[17] M. J. D. Powell. Least Frobenius norm updating of quadratic models that satisfy interpolation conditions. *Math. Program.*, 100:183–215, 2004.

[18] M. J. D. Powell. On updating the inverse of a KKT matrix. In Y. Yuan, editor, *Numerical Linear Algebra and Optimization*, pages 56–78. Science Press, Beijing, China, 2004.

[19] M. J. D. Powell. The NEWUOA software for unconstrained optimization without derivatives. In G. Di Pillo and M. Roma, editors, *Large-Scale Nonlinear Optimization*, volume 83 of *Nonconvex Optimization and Its Applications*, pages 255–297. Springer, Boston, MA, USA, 2006.

[20] M. J. D. Powell. The BOBYQA algorithm for bound constrained optimization without derivatives. Technical Report DAMTP 2009/NA06, Department of Applied Mathematics and Theoretical Physics, University of Cambridge, Cambridge, UK, 2009.

[21] M. J. D. Powell. Beyond symmetric Broyden for updating quadratic models in minimization without derivatives. *Math. Program.*, 138:475–500, 2013.

[22] H. Qian and Y. Yu. Derivative-free reinforcement learning: a review. *Front. Comput. Sci.*, 15:156336, 2021.

[23] T. M. Ragonneau. *Model-Based Derivative-Free Optimization Methods and Software*. PhD thesis, Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China, 2022.

[24] T. M. Ragonneau and Z. Zhang. COBYQA: Constrained Optimization BY Quadratic Approximations. https://www.cobyqa.com, 2023.

[25] L. M. Rios and N. V. Sahinidis. Derivative-free optimization: a review of algorithms and comparison of software implementations. *J. Global Optim.*, 56:1247–1293, 2013.

[26] S. M. Wild. MNH: a derivative-free optimization algorithm using minimal norm Hessians. In *The Tenth Copper Mountain Conference on Iterative Methods*, Copper Mountain, CO, USA, 2008.

[27] P. Xie and Y. Yuan. Least H^2 norm updating quadratic interpolation model function for derivative-free trust-region algorithms. arXiv:2302.12017, 2023.

[28] Y. Yuan. Recent advances in trust region algorithms. *Math. Program.*, 151:249–281, 2015.

[29] Z. Zhang. *On Derivative-Free Optimization Methods* (in Chinese). PhD thesis, Chinese Academy of Sciences, Beijing, China, 2012.

[30] Z. Zhang. Sobolev seminorm of quadratic functions with applications to derivative-free optimization. *Math. Program.*, 146:77–96, 2014.