TRFD: A derivative-free trust-region method based on finite differences for composite nonsmooth optimization

Dânâ Davar* Geovani N. Grapiglia†

October 11, 2024

Abstract

In this work we present TRFD, a derivative-free trust-region method based on finite differences for minimizing composite functions of the form $f(x) = h(F(x))$, where F is a black-box function assumed to have a Lipschitz continuous Jacobian, and h is a known convex Lipschitz function, possibly nonsmooth. The method approximates the Jacobian of F via forward finite differences. We establish an upper bound for the number of evaluations of F that TRFD requires to find an ϵ-approximate stationary point. For L1 and Minimax problems, we show that our complexity bound reduces to $\mathcal{O}(n\epsilon^{-2})$ for specific instances of TRFD, where n is the number of variables of the problem. Assuming that h is monotone and that the components of F are convex, we also establish a worst-case complexity bound, which reduces to $\mathcal{O}(n\epsilon^{-1})$ for Minimax problems. Numerical results are provided to illustrate the relative efficiency of TRFD in comparison with existing derivative-free solvers for composite nonsmooth optimization.

1 Introduction

1.1 Problem and Contributions

We are interested in composite optimization problems of the form

$$\text{Minimize } f(x) \equiv h(F(x)) \text{ subject to } x \in \Omega, \tag{1}$$

where $F : \mathbb{R}^n \to \mathbb{R}^m$ is assumed to be continuously differentiable with Lipschitz continuous Jacobian, $h : \mathbb{R}^m \to \mathbb{R}$ is a Lipschitz convex function (possibly nonsmooth), and $\Omega \subset \mathbb{R}^n$ is a nonempty closed convex set. Specifically, we assume that $F(\cdot)$ is only accessible through an exact zeroth-order oracle, meaning that the analytical form of $F(\cdot)$ is unknown, and that for any x, all we can compute is the exact function value $F(x)$. This situation occurs when $F(x)$ is obtained as the output of a black-box software or as the result of some simulation. Examples include aerodynamic shape optimization [12], optimization of cardiovascular geometries [17, 19] or tuning of algorithmic parameters [2], just to mention a few. Standard first-order methods for solving composite optimization problems (see, e.g.,

*Department of Mathematical Engineering, ICTEAM Institute, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium (dana.davar@uclouvain.be). This author was supported by the French Community of Belgium (FSR program).

†Department of Mathematical Engineering, ICTEAM Institute, Université catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium (geovani.grapiglia@uclouvain.be). This author was partially supported by FRS-FNRS, Belgium (Grant CDR J.0081.23).

[7, 20, 3]) require the computation of the Jacobian matrices of $F(\cdot)$ at the iterates, which are not readily available when $F(\cdot)$ is accessible via a zeroth-order oracle. Therefore, in this setting one needs to rely on derivative-free methods [5, 1, 14].

One of the main classes of derivative-free methods is the class of model-based trust-region methods (see, e.g., [4]). At each iteration, this type of method builds linear or quadratic interpolation models for the components of $F(\cdot)$, considering carefully chosen points around the current iterate. Then, the corresponding model of the objective $f(\cdot)$ is approximately minimized subject to a trust-region constraint. If the resulting trial point provides a sufficient decrease in the objective function, the point is accepted as the next iterate and the radius of the trust-region may increase. Otherwise, the trial point is rejected, and the process is repeated with a reduced trust-region radius. For the class of unconstrained composite optimization problems with h being convex and Lipschitz continuous, Grapiglia, Yuan and Yuan [11] proposed a model-based version of the trust-region method of Fletcher [7]. They proved that their derivative-free method takes at most $\mathcal{O}\left(n^2|\log(\epsilon^{-1})|\epsilon^{-2}\right)$ evaluations of $F(\cdot)$ to find an ϵ-approximate stationary point of $f(\cdot)$. Considering a wider class of model-based trust-region methods, Garmanjani, Júdice and Vicente [8] established an improved evaluation complexity bound of $\mathcal{O}\left(n^2\epsilon^{-2}\right)$, also assuming that $h(\cdot)$ is convex and Lipschitz continuous. For the case in which $h(\cdot)$ is not necessarily convex, *Manifold Sampling* trust-region methods have been proposed by Larson and Menickelly [15, 13] under the general assumption that

$$h(z) \in \{h_j(z) \,:\, j \in \{1, \ldots, q\}\}, \quad \forall z \in \mathbb{R}^m,$$

where $h_j : \mathbb{R}^m \to \mathbb{R}$ is a Lipschitz continuous differentiable function, with $\nabla h_j(\cdot)$ also Lipschitz continuous. In particular, the variants of Manifold Sampling recently proposed in [13] are currently the state-of-the-art solvers for derivative-free composite optimization problems.

In recent years, improved evaluation complexity bounds have been obtained for derivative-free methods designed to smooth unconstrained optimization[1]. Specifically, Grapiglia [9, 10] proved evaluation complexity bounds of $\mathcal{O}\left(n\epsilon^{-2}\right)$ for quadratic regularization methods with finite-difference gradient approximations. Moreover, for convex problems, a bound of $\mathcal{O}\left(n\epsilon^{-1}\right)$ was also established in [10]. Motivated by these advances, in the present work we propose a derivative-free trust-region method based on finite differences for composite problems of the form (1). At its *k-th* iteration, our new method (called TRFD) approximates the Jacobian matrix of F at x_k, $J_F(x_k)$, with a matrix A_k obtained via forward finite differences. The stepsize τ_k used in the finite differences and the trust-region radius Δ_k are jointly updated in a way that ensures an error bound

$$\|J_F(x_k) - A_k\|_2 \leq \mathcal{O}\left(\Delta_k\right), \quad \forall k.$$

Assuming that $f(\cdot)$ is bounded below by f_{low}, and denoting by $L_{h,p}$ the Lipschitz constant of $h(\cdot)$ with respect to a p-norm, and by L_J the Lipschitz constant of $J_F(\cdot)$ with respect to the Euclidean norm, we show that TRFD requires no more than

$$\mathcal{O}\left(nc_{2,p}(n)^2 c_{p,2}(m) L_{h,p} L_J (f(x_0) - f_{low})\epsilon^{-2}\right) \tag{2}$$

evaluations of $F(\cdot)$ to find an ϵ-approximate stationary point of $f(\cdot)$ on Ω, where $c_{2,p}(n)$ and $c_{p,2}(m)$ are positive constants such that

$$\|x\|_2 \leq c_{2,p}(n)\|x\|_p, \quad \forall x \in \mathbb{R}^n, \quad \text{and} \quad \|z\|_p \leq c_{p,2}(m)\|z\|_2, \quad \forall z \in \mathbb{R}^m.$$

[1] Problem (1) with $\Omega = \mathbb{R}^n$, $m = 1$ and $h(z) = z$.

For L1 and Minimax problems, which are composite problems defined respectively by $h(z) = \|z\|_1$ and $h(z) = \min_{i=1,\ldots,m} \{z_i\}$, we show that the complexity bound (2) reduces to $\mathcal{O}\left(n\epsilon^{-2}\right)$ for specific instances of TRFD. This represents an improvement with respect to the complexity bound of $\mathcal{O}\left(n^2\epsilon^{-2}\right)$ proved in [8] in the context of composite nonsmooth optimization. For the case where $h(\cdot)$ is monotone and the components of $F(\cdot)$ are convex functions, we also show that TRFD requires no more than

$$\mathcal{O}\left(nc_{2,p}(n)^2 c_{p,2}(m) L_{h,p} L_J \Delta_*^2 \epsilon^{-1}\right) \tag{3}$$

evaluations of $F(\cdot)$ to find an ϵ-approximate minimizer of $f(\cdot)$ on Ω, where Δ_* is a sufficiently large upper bound on the trust-region radii. For Minimax problems, we show that the bound (3) reduces to $\mathcal{O}\left(n\epsilon^{-1}\right)$ for specific instances of TRFD. To the best of our knowledge, this is the first time that evaluation complexity bounds with linear dependence on n are obtained for a deterministic DFO method in the context of composite nonsmooth optimization problems of the form (1). Finally, we present numerical results that illustrate the relative efficiency of TRFD on L1 and Minimax problems.

1.2 Contents

The paper is organized as follows. In Section 2, we prove the relevant auxiliary results. In Section 3, we analyze the evaluation complexity of the new method for nonconvex and convex problems. Finally, in Section 4, we report numerical results for L1 and Minimax problems.

1.3 Notations

Throughout the paper, given $p \in \mathbb{N}_\infty := \{1, 2, 3, \ldots\} \cup \{\infty\}$, $\|\cdot\|_p$ denotes the p-norm of vectors or matrices (depending on the context); and $\|\cdot\|_F$ denotes the Frobenius norm. Given $x \in \Omega$, $y \in \mathbb{R}^n$ and $r > 0$, we consider the sets

$$\Omega - \{x\} := \{s \in \mathbb{R}^n : x + s \in \Omega\} \quad \text{and} \quad B_p[y; r] = \{s \in \mathbb{R}^n \, : \, \|s - y\|_p \leq r\}.$$

In addition, $[A]_j$ denotes the *j-th* column of the matrix $A \in \mathbb{R}^{m \times n}$, while $[Ad]_i$ denotes the *i-th* coordinate of the vector $Ad \in \mathbb{R}^m$.

2 Assumptions and Auxiliary results

Through the paper, we will consider the following assumptions:

A1. $\Omega \subset \mathbb{R}^n$ is a nonempty closed convex set.
A2. $F : \mathbb{R}^n \to \mathbb{R}^m$ is differentiable and its Jacobian J_F is L_J-Lipschitz on $\mathbb{R}^n$ with respect to the Euclidean norm, that is,

$$\|J_F(x) - J_F(y)\|_2 \leq L_J \|x - y\|_2, \quad \forall x, y \in \mathbb{R}^n.$$

A3. $h : \mathbb{R}^m \to \mathbb{R}$ is convex and $L_{h,p}$-Lipschitz continuous on $\mathbb{R}^m$ with respect to the p-norm, that is,

$$|h(z) - h(w)| \leq L_{h,p} \|z - w\|_p, \quad \forall z, w \in \mathbb{R}^m.$$

Remark 2.1. *By A2, given $x, y \in \mathbb{R}^n$ we have*

$$\|F(y) - F(x) - J_F(x)(y - x)\|_2 \leq \frac{L_J}{2}\|y - x\|_2^2.$$

Remark 2.2. *Since all norms are equivalent on Euclidean spaces, given $p \in \{1, 2, \ldots, +\infty\}$, there exist positive constants $c_{2,p}(n), c_{p,2}(m) \geq 1$ such that*

$$\|x\|_2 \leq c_{2,p}(n)\|x\|_p, \quad \forall x \in \mathbb{R}^n, \quad \text{and} \quad \|z\|_p \leq c_{p,2}(m)\|z\|_2, \quad \forall z \in \mathbb{R}^m. \tag{4}$$

The lemma below provides a necessary condition for a solution of (1).

Lemma 2.3. *Suppose that A1-A3 hold. If x^* is a solution of (1), then*

$$f(x^*) \leq h(F(x^*) + J_F(x^*)s), \quad \forall s \in \Omega - \{x^*\}. \tag{5}$$

Proof. Suppose by contradiction that $f(x^*) > h(F(x^*) + J_F(x^*)\hat{s})$ for some $\hat{s} \in \Omega - \{x^*\}$. Then $\hat{s} \neq 0$ and there exists $\delta > 0$ such that

$$h(F(x^*) + J_F(x^*)\hat{s}) < f(x^*) - \delta \tag{6}$$

and

$$\delta \leq L_{h,p}c_{p,2}(m)L_J\|\hat{s}\|_2^2. \tag{7}$$

Given $\alpha \in [0, 1]$, let $\hat{x}(\alpha) = x^* + \alpha\hat{s}$. Then, using assumptions A3 and A2, and Remarks 2.1 and 2.2, we have

$$\begin{aligned}
f(\hat{x}(\alpha)) = h(F(\hat{x}(\alpha))) &= [h(F(x^* + \alpha\hat{s})) - h(F(x^*) + J_F(x^*)\alpha\hat{s})] \\
&\quad + h(F(x^*) + J_F(x^*)\alpha\hat{s}) \\
&\leq |h(F(x^* + \alpha\hat{s})) - h(F(x^*) + J_F(x^*)\alpha\hat{s})| + h(F(x^*) + J_F(x^*)\alpha\hat{s}) \\
&\leq L_{h,p}\|F(x^* + \alpha\hat{s}) - F(x^*) - J_F(x^*)\alpha\hat{s}\|_p \\
&\quad + h((1 - \alpha)F(x^*) + \alpha(F(x^*) + J_F(x^*)\hat{s})) \\
&\leq L_{h,p}c_{p,2}(m)\|F(x^* + \alpha\hat{s}) - F(x^*) - J_F(x^*)\alpha\hat{s}\|_2 \\
&\quad + (1 - \alpha)h(F(x^*)) + \alpha h(F(x^*) + J_F(x^*)\hat{s}) \\
&\leq \frac{L_{h,p}c_{p,2}(m)L_J\|\hat{s}\|_2^2}{2}\alpha^2 + (1 - \alpha)f(x^*) + \alpha h(F(x^*) + J_F(x^*)\hat{s}).
\end{aligned}$$

Then, by (6), it follows that

$$\begin{aligned}
f(\hat{x}(\alpha)) &< \frac{L_{h,p}c_{p,2}(m)L_J\|\hat{s}\|_2^2}{2}\alpha^2 + (1 - \alpha)f(x^*) + \alpha\left(f(x^*) - \delta\right) \\
&= \frac{L_{h,p}c_{p,2}(m)L_J\|\hat{s}\|_2^2}{2}\alpha^2 + f(x^*) - \delta\alpha.
\end{aligned} \tag{8}$$

Minimizing the right-hand side of the inequality above with respect to α, we obtain

$$\alpha^* = \frac{\delta}{L_{h,p}c_{p,2}(m)L_J\|\hat{s}\|_2^2}.$$

It follows from (7) that $\alpha^* \in [0,1]$. Thus, by (8) we would have

$$f(\hat{x}(\alpha^*)) < f(x^*) - \frac{\delta^2}{2L_{h,p}c_{p,2}(m)L_J\|\hat{s}\|_2^2} < f(x^*),$$

which contradicts the assumption that x^* is a solution of (1). Therefore, we conclude that (5) is true. $\square$

Lemma 2.3 motivates the following definition of stationarity, which in the unconstrained case corresponds to the definition considered by Yuan [21].

Definition 2.4. *We say that $x^* \in \Omega$ is a stationary point of f on Ω when x^* satisfies condition* (5).

Given $p \in \mathbb{N}_\infty$ and $r > 0$, let us define $\psi_{p,r} : \Omega \to \mathbb{R}$ by

$$\psi_{p,r}(x) = \frac{1}{r}\left(h(F(x)) - \min_{\substack{s \in \Omega - \{x\} \\ \|s\|_p \leq r}} h(F(x) + J_F(x)s)\right). \tag{9}$$

From the definition of $\psi_{p,r}(\,\cdot\,)$, we have the following result.

Lemma 2.5. *Suppose that A1-A3 hold and let $\psi_{p,r}(\,\cdot\,)$ be defined by* (9)*. Then,*

(a) $\psi_{p,r}(x) \geq 0 \quad \forall x \in \Omega$;
(b) $\psi_{p,r}(x^*) = 0$ if, and only if, x^* is a stationary point of f on Ω.

Proof. Given $x \in \Omega$, we have

$$\min_{\substack{s \in \Omega - \{x\} \\ \|s\|_p \leq r}} h(F(x) + J_F(x)s) \leq h(F(x)).$$

Then, by (9), we have $\psi_{p,r}(x) \geq 0$, and so (a) is true.

To prove (b), let us first assume that $x^* \in \Omega$ is a stationary point of f. By Definition 2.4,

$$h(F(x^*)) \leq h(F(x^*) + J_F(x^*)s) \quad \forall s \in \Omega - \{x^*\},$$

and so,

$$h(F(x^*)) \leq \min_{\substack{s \in \Omega - \{x^*\} \\ \|s\|_p \leq r}} h(F(x^*) + J_F(x^*)s).$$

Therefore, in view of (9), we have $\psi_{p,r}(x^*) \leq 0$. Combining this fact with (a), we conclude that $\psi_{p,r}(x^*) = 0$.

Now, suppose that $\psi_{p,r}(x^*) = 0$. Then, if $\boxed{\tilde{s} \in (\Omega - \{x^*\}) \cap B_p[0;r]}$, we have

$$h(F(x^*)) = \min_{\substack{s \in \Omega - \{x^*\} \\ \|s\|_p \leq r}} h(F(x^*) + J_F(x^*)s) \leq h(F(x^*) + J_F(x^*)\tilde{s}). \tag{10}$$

On the other hand, suppose that $\boxed{\tilde{s} \in (\Omega - \{x^*\}) \setminus B_p[0; r]}$, and define $\gamma = r/\|\tilde{s}\|_p$. Notice that $\gamma \in (0, 1)$. Then, by (10) and A3 we have

$$\begin{aligned} h(F(x^*)) \leq h(F(x^*) + J_F(x^*)\gamma\tilde{s}) &= h((1-\gamma)F(x^*) + \gamma(F(x^*) + J_F(x^*)\tilde{s})) \\ &\leq (1-\gamma)h(F(x^*)) + \gamma h(F(x^*) + J_F(x^*)\tilde{s}), \end{aligned}$$

which implies that

$$h(F(x^*)) \leq h(F(x^*) + J_F(x^*)\tilde{s}). \tag{11}$$

As a result, combining (10) and (11), we conclude that x^* is a stationary point of f on Ω. Therefore, (b) is also true. □

In view of Lemma 2.5, we will use $\psi_{p,r}(x)$ as a stationarity measure for problem (1).

Remark 2.6. *When $\Omega = \mathbb{R}^n$, $m = 1$ and $h(z) = z\ \forall z \in \mathbb{R}$, then problem* (1) *reduces to the smooth unconstrained problem* $\min_{x \in \mathbb{R}^n} F(x)$*, for which* $\psi_{2,r}(x) = \|\nabla F(x)\|_2$.

In the context of derivative-free optimization, $J_F(x)$ is unknown. Consequently, $\psi_{p,r}(x)$ is not computable. For a given $x \in \Omega$, our new method will compute a matrix $A \approx J_F(x)$ leading to the following approximate stationarity measure

$$\eta_{p,r}(x; A) := \frac{1}{r}\left(h(F(x)) - \min_{\substack{s \in \Omega - \{x\} \\ \|s\|_p \leq r}} h(F(x) + As) \right). \tag{12}$$

The error $|\psi_{p,r}(x) - \eta_{p,r}(x; A)|$ depends on how well A approximates $J_F(x)$. The next lemma provides an error bound when A is computed by forward finite differences.

Lemma 2.7. *Suppose that A2 holds. Given $x \in \mathbb{R}^n$ and $\tau > 0$, let $A \in \mathbb{R}^{m \times n}$ be defined by*

$$[A]_j = \frac{F(x + \tau e_j) - F(x)}{\tau}, \quad j = 1, ..., n. \tag{13}$$

Then

$$\|A - J_F(x)\|_2 \leq \frac{L_J\sqrt{n}}{2}\tau. \tag{14}$$

Proof. Given $j \in \{1, ..., n\}$, it follows from A2 that

$$\|F(x + \tau e_j) - F(x) - J_F(x)\tau e_j\|_2 \leq \frac{L_J}{2}\|\tau e_j\|_2^2 = \frac{L_J}{2}\tau^2.$$

Thus, by (13) we have

$$\|[A]_j - [J_F(x)]_j\|_2 \leq \frac{L_J}{2}\tau.$$

Consequently,

$$\|A - J_F(x)\|_F^2 = \sum_{j=1}^{n} \|[A]_j - [J_F(x)]_j\|_2^2 \leq n\left(\frac{L_J}{2}\tau\right)^2.$$

Therefore

$$\|A - J_F(x)\|_2 \leq \|A - J_F(x)\|_F \leq \frac{L_J\sqrt{n}}{2}\tau,$$

and so (14) is true. □

Remark 2.8. *Definition (13) implies that the i-th row of the corresponding matrix A is a forward finite-difference approximation to $\nabla F_i(x)$.*

Using Lemma 2.7, we can establish an upper bound for the error $|\psi_{p,r}(x) - \eta_{p,r}(x;A)|$ when A is constructed by forward finite differences.

Lemma 2.9. *Suppose that A1-A3 hold. Given $x \in \Omega$ and $\tau > 0$, let $A \in \mathbb{R}^{m\times n}$ be defined by (13). Then,*

$$|\psi_{p,r}(x) - \eta_{p,r}(x;A)| \leq \frac{L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)\sqrt{n}}{2}\tau, \quad \forall x \in \Omega. \tag{15}$$

Proof. By A3, it follows that $s \mapsto h(F(x) + J_F(x)s)$ defines a continuous function. Then, by A1 and the Weierstrass Theorem, there exists $\tilde{s} \in (\Omega - \{x\}) \cap B_p[0;r]$ such that

$$\min_{\substack{s\in\Omega-\{x\}\\ \|s\|_p\leq r}} h(F(x) + J_F(x)s) = h(F(x) + J_F(x)\tilde{s}).$$

Then, by A3 and (14),

$$\begin{aligned}
\psi_{p,r}(x) - \eta_{p,r}(x;A) &= \frac{1}{r}\left[h(F(x)) - h(F(x) + J_F(x)\tilde{s}) - \left(h(F(x)) - \min_{\substack{s\in\Omega-\{x\}\\ \|s\|_p\leq r}} h(F(x) + As)\right)\right] \\
&= \frac{1}{r}\left[\min_{\substack{s\in\Omega-\{x\}\\ \|s\|_p\leq r}} h(F(x) + As) - h(F(x) + J_F(x)\tilde{s})\right] \\
&\leq \frac{1}{r}\left[h(F(x) + A\tilde{s}) - h(F(x) + J_F(x)\tilde{s})\right] \\
&\leq \frac{L_{h,p}}{r}\|(A - J_F(x))\tilde{s}\|_p \\
&\leq \frac{L_{h,p}c_{p,2}(m)}{r}\|(A - J_F(x))\tilde{s}\|_2, \\
&\leq \frac{L_{h,p}c_{p,2}(m)}{r}\|A - J_F(x)\|_2\|\tilde{s}\|_2, \\
&\leq \frac{L_{h,p}c_{p,2}(m)}{r}\|A - J_F(x)\|_2 c_{2,p}(n)\|\tilde{s}\|_p, \\
&\leq \frac{L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)\sqrt{n}}{2}\tau.
\end{aligned} \tag{16}$$

In a similar way, considering $\hat{s} \in (\Omega - \{x\}) \cap B_p[0;r]$ such that

$$\min_{\substack{s\in\Omega-\{x\}\\ \|s\|_p\leq r}} h(F(x) + As) = h(F(x) + A\hat{s}),$$

we can show that

$$\eta_{p,r}(x;A) - \psi_{p,r}(x) \leq \frac{L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)\sqrt{n}}{2}\tau. \tag{17}$$

Combining (16) and (17), we see that (15) is true. □

Now, using Lemma 2.9, we can bound $\eta_{p,r}(x; A)$ from below when $\psi_{p,r}(x) > \epsilon$ and A is defined by (13) with τ being sufficiently small.

Lemma 2.10. *Suppose that A1-A3 hold. Given $x \in \Omega$ and $\tau > 0$, let A be defined by* (13)*. Given $r, \epsilon > 0$, if $\psi_{p,r}(x) > \epsilon$ and*

$$\tau \leq \frac{\max\{2\eta_{p,r}(x; A), \epsilon\}}{L_{h,p} L_J c_{p,2}(m) c_{2,p}(n)\sqrt{n}}, \tag{18}$$

then

$$\eta_{p,r}(x; A) > \frac{\epsilon}{2}.$$

Proof. From Lemma 2.9 and (18), we have

$$\psi_{p,r}(x) \leq |\psi_{p,r}(x) - \eta_{p,r}(x; A)| + \eta_{p,r}(x; A) \leq \max\left\{\eta_{p,r}(x; A), \frac{\epsilon}{2}\right\} + \eta_{p,r}(x; A). \tag{19}$$

Suppose that $\eta_{p,r}(x; A) \leq \frac{\epsilon}{2}$. Then from (19) we would have $\psi_{p,r}(x) \leq \epsilon$, contradicting the assumption that $\psi_{p,r}(x) > \epsilon$. Therefore, we must have $\eta_{p,r}(x; A) > \epsilon/2$. □

Given $0 < r_1 \leq r_2$, the next lemma establishes the relation between $\eta_{p,r_1}(x; A)$ and $\eta_{p,r_2}(x; A)$ for any given $x \in \Omega$ and $A \in \mathbb{R}^{m \times n}$.

Lemma 2.11. *Suppose that A1 and A3 hold. Given $x \in \Omega$, $A \in \mathbb{R}^{m \times n}$ and $0 < r_1 \leq r_2$, we have*

$$\eta_{p,r_1}(x; A) \geq \eta_{p,r_2}(x; A).$$

Proof. Denote $\boxed{\alpha = \frac{r_1}{r_2}}$ and let $s^* \in (\Omega - \{x\}) \cap B_p[0; r_2]$ be such that

$$\min_{\substack{s \in \Omega - \{x\} \\ \|s\|_p \leq r_2}} h(F(x) + As) = h(F(x) + As^*).$$

Then, $\|\alpha s^*\| \leq r_1$ and so

$$\min_{\substack{s \in \Omega - \{x\} \\ \|s\|_p \leq r_1}} h(F(x) + As) \leq h(F(x) + \alpha A s^*),$$

which implies that

$$\eta_{p,r_1}(x; A) = \frac{1}{r_1}\left(h(F(x)) - \min_{\substack{s \in \Omega - \{x\} \\ \|s\|_p \leq r_1}} h(F(x) + As)\right) \geq \frac{1}{r_1}\left(h(F(x)) - h(F(x) + \alpha A s^*)\right). \tag{20}$$

On the other hand, using the convexity of h (from A3) and the fact that $\alpha \in (0, 1]$, we also have

$$\begin{aligned} h(F(x) + \alpha A s^*) &= h((1 - \alpha)F(x) + \alpha(F(x) + As^*)) \\ &\leq (1 - \alpha)h(F(x)) + \alpha h(F(x) + As^*) \\ &= h(F(x)) + \alpha[h(F(x) + As^*) - h(F(x))]. \end{aligned} \tag{21}$$

Finally, combining (20), (21) and the definition of α, we obtain

$$\eta_{p,r_1}(x; A) \geq \frac{\alpha}{r_1}[h(F(x)) - h(F(x) + As^*)] = \frac{1}{r_2}[h(F(x)) - h(F(x) + As^*)] = \eta_{p,r_2}(x; A),$$

and the proof is complete. □

3 Trust-region method based on finite differences

Leveraging the auxiliary results established in the previous section, we propose a derivative-free **T**rust-**R**egion Method based on **F**inite-**D**ifference Jacobian approximations, which we refer to as TRFD.

Algorithm 1: TRFD

Step 0. Given $x_0 \in \Omega$, $\epsilon > 0$, $\sigma > 0$, $\alpha \in (0,1)$, $\theta \in (0,1]$ and the Lipschitz constant $L_{h,p}$ of $h(\,\cdot\,)$, define

$$\tau_0 = \frac{\epsilon}{L_{h,p}\sigma c_{p,2}(m)c_{2,p}(n)\sqrt{n}},$$

where $c_{p,2}(m)$ and $c_{2,p}(n)$ satisfy (4). Choose Δ_0 and Δ_* such that $\tau_0\sqrt{n} \leq \Delta_0 \leq \Delta_*$ and set $k := 0$.

Step 1. Construct $A_k \in \mathbb{R}^{m\times n}$ with

$$[A_k]_j = \frac{F(x_k + \tau_k e_j) - F(x_k)}{\tau_k}, \quad j = 1, ..., n.$$

and compute $\eta_{p,\Delta_*}(x_k; A_k)$ defined in (12).

Step 2. If $\eta_{p,\Delta_*}(x_k; A_k) \geq \epsilon/2$, go to Step 3. Otherwise, define $x_{k+1} = x_k$, $\Delta_{k+1} = \Delta_k$, $\tau_{k+1} = \frac{1}{2}\tau_k$, set $k := k+1$ and go to Step 1.

Step 3 Let d_k^* be a solution of the trust-region subproblem

$$\begin{aligned} \min_{d\in\mathbb{R}^n} \quad & h(F(x_k) + A_k d) \\ \text{s.t.} \quad & \|d\|_p \leq \Delta_k \\ & x_k + d \in \Omega. \end{aligned} \tag{22}$$

Compute $d_k \in B_p[0; \Delta_k] \cap (\Omega - \{x_k\})$ such that

$$h(F(x_k)) - h(F(x_k) + A_k d_k) \geq \theta\left[h(F(x_k)) - h(F(x_k) + A_k d_k^*)\right]. \tag{23}$$

Step 4. Compute

$$\rho_k = \frac{h(F(x_k)) - h(F(x_k + d_k))}{h(F(x_k)) - h(F(x_k) + A_k d_k)}. \tag{24}$$

If $\rho_k \geq \alpha$, define $x_{k+1} = x_k + d_k$, $\Delta_{k+1} = \min\{2\Delta_k, \Delta_*\}$, $\tau_{k+1} = \tau_k$, set $k := k+1$ and go to Step 1.

Step 5 Set $x_{k+1} = x_k$, $\Delta_{k+1} = \frac{1}{2}\Delta_k$. If $\tau_k\sqrt{n} \leq \Delta_{k+1}$, define $\tau_{k+1} = \tau_k$, $A_{k+1} = A_k$, set $k := k+1$ and go to Step 3. Otherwise, define $\tau_{k+1} = \frac{1}{2}\tau_k$, set $k := k+1$ and go to Step 1.

In TRFD, we have four types of iterations:

1. **Unsuccessful iterations of type I** ($\mathcal{U}^{(1)}$): those where $\eta_{p,\Delta_*}(x_k; A_k) < \epsilon/2$.

2. **Successful iterations** ($\mathcal{S}$): those where $\eta_{p,\Delta_*}(x_k; A_k) \geq \epsilon/2$ and $\rho_k \geq \alpha$.

3. **Unsuccessful iterations of type II** ($\mathcal{U}^{(2)}$): those where $\eta_{p,\Delta_*}(x_k; A_k) \geq \epsilon/2$, $\rho_k < \alpha$, and $\tau_k\sqrt{n} \leq \Delta_{k+1}$.
4. **Unsuccessful iterations of type III** ($\mathcal{U}^{(3)}$): those where $\eta_{p,\Delta_*}(x_k; A_k) \geq \epsilon/2$, $\rho_k < \alpha$, and $\tau_k\sqrt{n} > \Delta_{k+1}$.

The lemma below shows that the finite-difference stepsize τ_k is always bounded from above by $\Delta_k/\sqrt{n}$.

Lemma 3.1. *Given $T \geq 1$, let $\{\tau_k\}_{k=0}^T$ and $\{\Delta_k\}_{k=0}^T$ be generated by TRFD. Then*

$$\tau_k\sqrt{n} \leq \Delta_k, \tag{25}$$

for $k = 0, \ldots, T$.

Proof. We will prove this result by induction over k. In view of the choice of Δ_0 at Step 0 of TRFD, we see that (25) is true for $k = 0$. Assuming that (25) is true for some $k \in \{0, \ldots, T-1\}$, we will show that it is also true for $k+1$. Considering our classification of iterations, we have four possible cases.

Case I: $k \in \mathcal{U}^{(1)}$.

In this case, by Step 2 of TRFD, we have $\tau_{k+1} = \frac{1}{2}\tau_k$ and $\Delta_{k+1} = \Delta_k$. Thus, by the induction assumption,

$$\tau_{k+1}\sqrt{n} = \frac{1}{2}\tau_k\sqrt{n} < \tau_k\sqrt{n} \leq \Delta_k = \Delta_{k+1},$$

that is, (25) holds for $k+1$.

Case II: $k \in \mathcal{S}$.

In this case, by Step 4 of TRFD, we have $\tau_{k+1} = \tau_k$ and $\Delta_{k+1} \geq \Delta_k$. Thus, by the induction assumption,

$$\tau_{k+1}\sqrt{n} = \tau_k\sqrt{n} \leq \Delta_k \leq \Delta_{k+1},$$

which means that (25) is true for $k+1$.

Case III: $k \in \mathcal{U}^{(2)}$

In this case, by Step 5 of TRFD, we have $\tau_k\sqrt{n} \leq \Delta_{k+1}$, and $\tau_{k+1} = \tau_k$. Thus

$$\tau_{k+1}\sqrt{n} = \tau_k\sqrt{n} \leq \Delta_{k+1},$$

that is, (25) is true for $k+1$.

Case IV: $k \in \mathcal{U}^{(3)}$.

In this case, by Step 5 of TRFD we have $\tau_{k+1} = \frac{1}{2}\tau_k$ and $\Delta_{k+1} = \frac{1}{2}\Delta_k$. Thus, by the induction assumption,

$$\tau_{k+1}\sqrt{n} = \frac{1}{2}\tau_k\sqrt{n} \leq \frac{1}{2}\Delta_k = \Delta_{k+1},$$

and so (25) is true for $k+1$, which concludes the proof. ☐

In view of Lemmas 2.7 and 3.1, the matrices A_k in TRFD satisfy

$$\|J_F(x_k) - A_k\|_2 \leq \frac{L_J}{2}\Delta_k \quad \forall k.$$

Thanks to this error bound, we can derive the following sufficient condition for an iteration to be successful.

Lemma 3.2. *Suppose that A1-A3 hold. If $\psi_{p,\Delta_*}(x_k) > \epsilon$ and*

$$\Delta_k \leq \frac{(1-\alpha)\theta\eta_{p,\Delta_*}(x_k; A_k)}{L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)^2}, \tag{26}$$

then $k \in \mathcal{S}$.

Proof. From Step 0 of TRFD, we have $\alpha \in (0,1)$ and $\theta \in (0,1]$. Then, it follows from Lemma 3.1, (26) and $c_{2,p}(n) \geq 1$ that

$$\tau_k \leq \frac{\Delta_k}{\sqrt{n}} \leq \frac{2\eta_{p,\Delta_*}(x_k; A_k)}{L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)\sqrt{n}}.$$

Since $\psi_{p,\Delta_*}(x_k) > \epsilon$, by Lemma 2.10 we get

$$\eta_{p,\Delta_*}(x_k; A_k) \geq \epsilon/2.$$

Therefore, to conclude that $k \in S$, it remains to show that $\rho_k \geq \alpha$. On the one hand, by A3, (4) and A2 we have

$$\begin{aligned}
& h(F(x_k + d_k)) - h(F(x_k) + A_k d_k) \\
= \ & h(F(x_k + d_k)) - h(F(x_k) + J_F(x_k)d_k) + h(F(x_k) + J_F(x_k)d_k) - h(F(x_k) + A_k d_k) \\
\leq \ & |h(F(x_k + d_k)) - h(F(x_k) + J_F(x_k)d_k)| + |h(F(x_k) + J_F(x_k)d_k) - h(F(x_k) + A_k d_k)| \\
\leq \ & L_{h,p}\|F(x_k + d_k) - F(x_k) - J_F(x_k)d_k\|_p + L_{h,p}\|(J_F(x_k) - A_k)d_k\|_p \\
\leq \ & L_{h,p}c_{p,2}(m)\|F(x_k + d_k) - F(x_k) - J_F(x_k)d_k\|_2 + L_{h,p}c_{p,2}(m)\|J_F(x_k) - A_k\|_2\|d_k\|_2 \\
\leq \ & L_{h,p}c_{p,2}(m)\frac{L_J}{2}\|d_k\|_2^2 + L_{h,p}c_{p,2}(m)\frac{L_J\sqrt{n}}{2}\tau_k\|d_k\|_2 \\
\leq \ & (0.5)\, L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)^2\|d_k\|_p^2 + (0.5)\, L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)\tau_k\sqrt{n}\|d_k\|_p \\
\leq \ & (0.5)\, L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)^2\Delta_k^2 + (0.5)\, L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)\tau_k\sqrt{n}\Delta_k.
\end{aligned}$$

Then, by Lemma 3.1 and $c_{2,p}(n) \geq 1$, we have

$$\begin{aligned}
h(F(x_k + d_k)) - h(F(x_k) + A_k d_k) \ & \leq \ (0.5)\, L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)\left(c_{2,p}(n) + 1\right)\Delta_k^2 \\
& \leq \ L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)^2\Delta_k^2.
\end{aligned} \tag{27}$$

On the other hand, by (23) and (12) we have

$$\begin{aligned}
h(F(x_k)) - h(F(x_k) + A_k d_k) \ & \geq \ \theta\left[h(F(x_k)) - h(F(x_k) + A_k d_k^*)\right] \\
& = \ \theta\Delta_k\left[\frac{1}{\Delta_k}\left(h(F(x_k)) - h(F(x_k) + A_k d_k^*)\right)\right] \\
& = \ \theta\Delta_k\eta_{p,\Delta_k}(x_k; A_k).
\end{aligned}$$

Since $\Delta_k \leq \Delta_*$, it follows from Lemma 2.11 that

$$h(F(x_k)) - h(F(x_k) + A_k d_k) \geq \theta \Delta_k \eta_{p,\Delta_*}(x_k; A_k). \tag{28}$$

Now, combining (24), (27), (28) and (26), we obtain

$$\begin{aligned} 1 - \rho_k &= \frac{h(F(x_k)) - h(F(x_k) + A_k d_k) - [h(F(x_k)) - h(F(x_k + d_k))]}{h(F(x_k)) - h(F(x_k) + A_k d_k)} \\ &= \frac{h(F(x_k + d_k)) - h(F(x_k) + A_k d_k)}{h(F(x_k)) - h(F(x_k) + A_k d_k)} \leq \frac{L_{h,p} L_J c_{p,2}(m) c_{2,p}(n)^2 \Delta_k^2}{\theta \Delta_k \eta_{p,\Delta_*}(x_k; A_k)} \\ &= \frac{L_{h,p} L_J c_{p,2}(m) c_{2,p}(n)^2 \Delta_k}{\theta \eta_{p,\Delta_*}(x_k; A_k)} \leq 1 - \alpha. \end{aligned}$$

Therefore, $\rho_k \geq \alpha$, and we conclude that $k \in \mathcal{S}$. ☐

Now we can obtain a lower bound on the trust-region radii.

Lemma 3.3. *Suppose that A1-A3 hold and, given $T \geq 1$, let $\{\Delta_k\}_{k=0}^{T}$ be generated by TRFD. If*

$$\psi_{p,\Delta_*}(x_k) > \epsilon, \quad \textit{for} \quad k = 0, \ldots, T-1,$$

then

$$\Delta_k \geq \frac{(1-\alpha)\theta\epsilon}{4L_{h,p} \max\{\sigma, L_J\} c_{p,2}(m) c_{2,p}(n)^2} \equiv \Delta_{\min}(\epsilon), \quad \textit{for} \quad k = 0, \ldots, T. \tag{29}$$

Proof. First, let us prove by induction that

$$\Delta_k \geq \min\left\{\Delta_0, \frac{(1-\alpha)\theta\epsilon}{4L_{h,p} L_J c_{p,2}(m) c_{2,p}(n)^2}\right\} \equiv \tilde{\Delta}_{\min}(\epsilon), \quad \text{for} \quad k = 0, \ldots, T. \tag{30}$$

Clearly, the inequality in (30) is true for $k = 0$. Suppose that the inequality in (30) is true for some $k \in \{0, \ldots, T-1\}$. If $k \in \mathcal{U}^{(1)}$, then $\Delta_{k+1} = \Delta_k \geq \tilde{\Delta}_{\min}(\epsilon)$ and so (30) holds for $k+1$. Now, suppose that $k \notin \mathcal{U}^{(1)}$. In this case, we have $\eta_{p,\Delta_*}(x_k; A_k) \geq \epsilon/2$. Thus, if

$$\Delta_k \leq \frac{(1-\alpha)\theta\epsilon}{2L_{h,p} L_J c_{p,2}(m) c_{2,p}(n)^2}, \tag{31}$$

then by Lemma 3.2, $k \in \mathcal{S}$. Consequently, Step 4 of TRFD and the induction assumption imply that

$$\Delta_{k+1} = \min\{2\Delta_k, \Delta_*\} \geq \min\{\Delta_k, \Delta_*\} = \Delta_k \geq \tilde{\Delta}_{\min}(\epsilon).$$

Now, suppose that (31) is not true. Since in any case we have $\Delta_{k+1} \geq \frac{1}{2}\Delta_k$, we will have

$$\Delta_{k+1} \geq \frac{1}{2}\Delta_k > \frac{(1-\alpha)\theta\epsilon}{4L_{h,p} L_J c_{p,2}(m) c_{2,p}(n)^2} \geq \tilde{\Delta}_{\min}(\epsilon).$$

This shows that (30) is true. Finally, it follows from Step 0 of TRFD that

$$\Delta_0 \geq \tau_0 \sqrt{n} = \frac{\epsilon}{L_{h,p} \sigma c_{p,2}(m) c_{2,p}(n)} \geq \frac{(1-\alpha)\theta\epsilon}{4L_{h,p} \max\{\sigma, L_J\} c_{p,2}(m) c_{2,p}(n)^2}.$$

Thus, from the definition of $\tilde{\Delta}_{\min}(\epsilon)$ in (30), we see that

$$\tilde{\Delta}_{\min}(\epsilon) \geq \frac{(1-\alpha)\theta\epsilon}{4L_{h,p} \max\{\sigma, L_J\} c_{p,2}(m) c_{2,p}(n)^2} = \Delta_{\min}(\epsilon). \tag{32}$$

Then, combining (30) and (32), we conclude that (29) is true. ☐

3.1 Worst-Case Complexity Bound for Nonconvex Problems

Given $j \in \{0, 1, 2, \ldots, \}$, let

$$\begin{aligned} \mathcal{S}_j &= \{0, 1, \ldots, j\} \cap \mathcal{S}, \\ \mathcal{U}_j^{(i)} &= \{0, 1, \ldots, j\} \cap \mathcal{U}^{(i)}, \quad i \in \{1, 2, 3\}. \end{aligned}$$

Also, let

$$T_g(\epsilon) = \inf \{k \in \mathbb{N} \,:\, \psi_{p,\Delta_*}(x_k) \leq \epsilon\} \tag{33}$$

be the index of the first iteration in which $\{x_k\}_{k \geq 0}$ reaches an ϵ-approximate stationary point, if it exists. Our goal is to obtain a finite upper bound for $T_g(\epsilon)$. Assuming that $T_g(\epsilon) \geq 1$, it follows from the notation above that

$$\begin{aligned} T_g(\epsilon) &= \left| \mathcal{S}_{T_g(\epsilon)-1} \cup \left(\mathcal{U}^{(1)}_{T_g(\epsilon)-1} \cup \mathcal{U}^{(2)}_{T_g(\epsilon)-1} \cup \mathcal{U}^{(3)}_{T_g(\epsilon)-1} \right) \right| \\ &\leq \left| \mathcal{S}_{T_g(\epsilon)-1} \right| + \left| \mathcal{U}^{(1)}_{T_g(\epsilon)-1} \cup \mathcal{U}^{(3)}_{T_g(\epsilon)-1} \right| + \left| \mathcal{U}^{(2)}_{T_g(\epsilon)-1} \cup \mathcal{U}^{(3)}_{T_g(\epsilon)-1} \right|. \end{aligned} \tag{34}$$

In the next three lemmas, we will provide upper bounds for each of the three terms in (34). To that end, let us consider the following additional assumption:

A4. There exists $f_{low} \in \mathbb{R}$ such that $f(x) \geq f_{low}$ for all $x \in \mathbb{R}^n$.

The next lemma provides an upper bound on $\left|\mathcal{S}_{T_g(\epsilon)-1}\right|$.

Lemma 3.4. *Suppose that A1-A4 hold and that $T_g(\epsilon) \geq 1$. Then*

$$|\mathcal{S}_{T_g(\epsilon)-1}| \leq \frac{8L_{h,p} \max\{\sigma, L_J\}\, c_{p,2}(m) c_{2,p}(n)^2 (f(x_0) - f_{low})}{\alpha(1-\alpha)\theta^2} \epsilon^{-2}. \tag{35}$$

Proof. Let $k \in S_{T_g(\epsilon)-1}$, that is, $\eta_{p,\Delta_*}(x_k; A_k) \geq \epsilon/2$ and $\rho_k \geq \alpha$. Then, by (23), (12), $\Delta_k \leq \Delta_*$ and Lemma 2.11, we have

$$\begin{aligned} f(x_k) - f(x_{k+1}) &= h(F(x_k)) - h(F(x_k + d_k)) \\ &\geq \alpha \left[h(F(x_k)) - h(F(x_k) + A_k d_k) \right] \\ &\geq \alpha\theta \left[h(F(x_k)) - h(F(x_k) + A_k d_k^*) \right] \\ &= \alpha\theta\Delta_k \left[\frac{1}{\Delta_k} \left(h(F(x_k)) - h(F(x_k) + A_k d_k^*) \right) \right] \\ &= \alpha\theta\Delta_k \eta_{p,\Delta_k}(x_k; A_k) \\ &\geq \alpha\theta\Delta_k \eta_{p,\Delta_*}(x_k; A_k) \\ &\geq \frac{\alpha\theta\epsilon}{2} \Delta_k. \end{aligned}$$

Consequently, it follows from Lemma 3.3 that

$$f(x_k) - f(x_{k+1}) \geq \frac{\alpha(1-\alpha)\theta^2}{8L_{h,p} \max\{\sigma, L_J\}\, c_{p,2}(m) c_{2,p}(n)^2} \epsilon^2 \quad \text{when } k \in \mathcal{S}_{T_g(\epsilon)-1}. \tag{36}$$

Let $\mathcal{S}^c_{T_g(\epsilon)-1} = \{0, 1, \ldots, T_g(\epsilon) - 1\} \setminus \mathcal{S}_{T_g(\epsilon)-1}$. Notice that, when $k \in \mathcal{S}^c_{T_g(\epsilon)-1}$, then $f(x_{k+1}) = f(x_k)$. Thus, it follows from A4 and (36) that

$$\begin{aligned} f(x_0) - f_{low} &\geq f(x_0) - f(x_{T_g(\epsilon)}) = \sum_{k=0}^{T_g(\epsilon)-1} f(x_k) - f(x_{k+1}) \\ &= \sum_{k \in \mathcal{S}_{T_g(\epsilon)-1}} f(x_k) - f(x_{k+1}) + \sum_{k \in \mathcal{S}^c_{T_g(\epsilon)-1}} f(x_k) - f(x_{k+1}) \\ &= \sum_{k \in \mathcal{S}_{T_g(\epsilon)-1}} f(x_k) - f(x_{k+1}) \\ &\geq |\mathcal{S}_{T_g(\epsilon)-1}| \frac{\alpha(1-\alpha)\theta^2}{8L_{h,p} \max\{\sigma, L_J\} c_{p,2}(m) c_{2,p}(n)^2} \epsilon^2, \end{aligned}$$

which implies that (35) is true. □

The next lemma provides an upper bound on $\left|\mathcal{U}^{(1)}_{T_g(\epsilon)-1} \cup \mathcal{U}^{(3)}_{T_g(\epsilon)-1}\right|$.

Lemma 3.5. *Suppose that A1-A3 hold and that $T_g(\epsilon) \geq 2$. If $T \in \{2, \ldots, T_g(\epsilon)\}$, then*

$$\left|\mathcal{U}^{(1)}_{T-1} \cup \mathcal{U}^{(3)}_{T-1}\right| \leq \left\lceil \left| \log_2 \left(\frac{\tau_0 \sqrt{n}}{\Delta_{\min}(\epsilon)} \right) \right| \right\rceil, \tag{37}$$

where $\Delta_{\min}(\epsilon)$ is defined in (29).

Proof. Suppose by contradiction that

$$\left|\mathcal{U}^{(1)}_{T-1} \cup \mathcal{U}^{(3)}_{T-1}\right| > \left\lceil \left| \log_2 \left(\frac{\tau_0 \sqrt{n}}{\Delta_{\min}(\epsilon)} \right) \right| \right\rceil. \tag{38}$$

Notice that

$$\left|\mathcal{U}^{(1)}_0 \cup \mathcal{U}^{(3)}_0\right| \leq 1 \quad \text{and} \quad \left|\mathcal{U}^{(1)}_{k+1} \cup \mathcal{U}^{(3)}_{k+1}\right| \leq \left|\mathcal{U}^{(1)}_k \cup \mathcal{U}^{(3)}_k\right| + 1, \quad \forall k. \tag{39}$$

It follows from (38) and (39) that there exists $k_* \in \{0, \ldots, T-2\}$ such that

$$\left|\mathcal{U}^{(1)}_{k_*} \cup \mathcal{U}^{(3)}_{k_*}\right| = \left\lceil \left| \log_2 \left(\frac{\tau_0 \sqrt{n}}{\Delta_{\min}(\epsilon)} \right) \right| \right\rceil.$$

In view of (39), $\boxed{\text{for any } k \in \mathbb{N} \cap [k_*, T-1]}$ we have

$$\left|\mathcal{U}^{(1)}_k \cup \mathcal{U}^{(3)}_k\right| \geq \left|\mathcal{U}^{(1)}_{k_*} \cup \mathcal{U}^{(3)}_{k_*}\right| \geq \left| \log_2 \left(\frac{\tau_0 \sqrt{n}}{\Delta_{\min}(\epsilon)} \right) \right| \geq -\log_2 \left(\frac{\Delta_{\min}(\epsilon)}{\tau_0 \sqrt{n}} \right).$$

Thus

$$-\left|\mathcal{U}^{(1)}_k \cup \mathcal{U}^{(3)}_k\right| \leq \log_2 \left(\frac{\Delta_{\min}(\epsilon)}{\tau_0 \sqrt{n}} \right),$$

and so

$$\tau_k = (0.5)^{\left|\mathcal{U}^{(1)}_k \cup \mathcal{U}^{(3)}_k\right|} \tau_0 = 2^{-\left|\mathcal{U}^{(1)}_k \cup \mathcal{U}^{(3)}_k\right|} \tau_0 \leq \frac{\Delta_{\min}(\epsilon)}{\sqrt{n}}. \tag{40}$$

By (40), the definition of $\Delta_{\min}(\epsilon)$ in (29), and $c_{2,p}(n) \geq 1$, we have

$$\tau_k \leq \frac{\epsilon}{L_{h,p} L_J c_{p,2}(m) c_{2,p}(n)\sqrt{n}}.$$

Since we also have $\psi_{p,\Delta_*}(x_k) > \epsilon$, it follows from Lemma 2.10 that $\eta_{p,\Delta_*}(x_k) \geq \epsilon/2$. Therefore, the *k-th* iteration is not an unsuccessful iteration of type I, i.e., $\boxed{k \notin \mathcal{U}^{(1)}}$. In addition, (40) and Lemma 3.3 imply that

$$\tau_k\sqrt{n} \leq \Delta_{\min}(\epsilon) \leq \Delta_{k+1},$$

which means that the *k-th* iteration is not an unsuccessful iteration of type III, i.e., $\boxed{k \notin \mathcal{U}^{(3)}}$. In summary, $\boxed{k \notin \mathcal{U}^{(1)} \cup \mathcal{U}^{(3)}}$ and so

$$\left|\mathcal{U}_k^{(1)} \cup \mathcal{U}_k^{(3)}\right| = \left|\mathcal{U}_{k-1}^{(1)} \cup \mathcal{U}_{k-1}^{(3)}\right|.$$

Thus, for any $k_* < k \leq T-1$,

$$\left|\mathcal{U}_k^{(1)} \cup \mathcal{U}_k^{(3)}\right| = \left|\mathcal{U}_{k-1}^{(1)} \cup \mathcal{U}_{k-1}^{(3)}\right| = \ldots = \left|\mathcal{U}_{k_*}^{(1)} \cup \mathcal{U}_{k_*}^{(3)}\right|$$

In particular,

$$\left|\mathcal{U}_{T-1}^{(1)} \cup \mathcal{U}_{T-1}^{(3)}\right| = \left|\mathcal{U}_{k_*}^{(1)} \cup \mathcal{U}_{k_*}^{(3)}\right| = \left\lceil \left|\log_2\left(\frac{\tau_0\sqrt{n}}{\Delta_{\min}(\epsilon)}\right)\right| \right\rceil,$$

contradicting (38). □

Remark 3.6. *By the definition of* τ_0 *(at Step 0 of TRFD) and the definition of* $\Delta_{\min}(\epsilon)$ *in* (29)*, we have*

$$\frac{\tau_0\sqrt{n}}{\Delta_{\min}(\epsilon)} = \frac{4\max\{\sigma, L_J\}\, c_{2,p}(n)}{\sigma(1-\alpha)\theta}. \tag{41}$$

The lemma below provides an upper bound on $\left|\mathcal{U}_{T_g(\epsilon)-1}^{(2)} \cup \mathcal{U}_{T_g(\epsilon)-1}^{(3)}\right|$.

Lemma 3.7. *Suppose that A1-A3 hold and that* $T_g(\epsilon) \geq 1$*. If* $T \in \{1, \ldots, T_g(\epsilon)\}$*, then*

$$\left|\mathcal{U}_{T-1}^{(2)} \cup \mathcal{U}_{T-1}^{(3)}\right| \leq \log_2\left(\frac{4L_{h,p}\max\{\sigma, L_J\}\, c_{p,2}(m) c_{2,p}(n)^2 \Delta_0}{(1-\alpha)\theta}\epsilon^{-1}\right) + |\mathcal{S}_{T-1}|. \tag{42}$$

Proof. By the update rules for Δ_k in TRFD, we have

$$\begin{aligned}
\Delta_{k+1} &= \frac{1}{2}\Delta_k, \text{ if } k \in \mathcal{U}_{T-1}^{(2)} \cup \mathcal{U}_{T-1}^{(3)},\\
\Delta_{k+1} &= \Delta_k, \quad \text{if } k \in \mathcal{U}_{T-1}^{(1)},\\
\Delta_{k+1} &\leq 2\Delta_k, \ \text{ if } k \in \mathcal{S}_{T-1}.
\end{aligned}$$

In addition, by Lemma 3.3 we have

$$\Delta_k \geq \Delta_{\min}(\epsilon) \quad \text{for} \quad k = 0, \ldots, T,$$

where $\Delta_{\min}(\epsilon)$ is defined in (29). Thus, considering $\nu_k = 1/\Delta_k$, it follows that

$$
\begin{aligned}
2\nu_k &= \nu_{k+1}, \quad \text{if } k \in \mathcal{U}^{(2)}_{T-1} \cup \mathcal{U}^{(3)}_{T-1}, && (43)\\
\nu_k &= \nu_{k+1}, \quad \text{if } k \in \mathcal{U}^{(1)}_{T-1}, && (44)\\
\frac{1}{2}\nu_k &\leq \nu_{k+1}, \quad \text{if } k \in \mathcal{S}_{T-1}, && (45)
\end{aligned}
$$

and

$$\nu_k \leq \Delta_{\min}(\epsilon)^{-1} \quad \text{for} \quad k = 0, \ldots, T. \tag{46}$$

In view of (43)-(46), we have

$$2^{\left|\mathcal{U}^{(2)}_{T-1} \cup \mathcal{U}^{(3)}_{T-1}\right|} (0.5)^{|\mathcal{S}_{T-1}|} \nu_0 \leq \nu_T \leq \Delta_{\min}(\epsilon)^{-1}.$$

Then, taking the logarithm in both sides we get

$$\left|\mathcal{U}^{(2)}_{T-1} \cup \mathcal{U}^{(3)}_{T-1}\right| - |\mathcal{S}_{T-1}| \leq \log_2\left(\frac{\Delta_{\min}(\epsilon)^{-1}}{\nu_0}\right) = \log_2\left(\frac{\Delta_0}{\Delta_{\min}(\epsilon)}\right),$$

which together with (29) implies that (42) is true. □

Now, combining the previous results, we obtain the following worst-case complexity bound on the number of iterations required by TRFD to find an ϵ-approximate stationary point.

Theorem 3.8. *Suppose that A1-A4 hold and let $T_g(\epsilon)$ be defined by (33). Then*

$$
\begin{aligned}
T_g(\epsilon) \leq\ & \frac{16L_{h,p}\max\{\sigma, L_J\}\, c_{p,2}(m)c_{2,p}(n)^2(f(x_0) - f_{low})}{\alpha(1-\alpha)\theta^2}\epsilon^{-2} + \left\lceil\left|\log_2\left(\frac{4\max\{\sigma, L_J\}\, c_{2,p}(n)}{\sigma(1-\alpha)\theta}\right)\right|\right\rceil \\
& + \log_2\left(\frac{4L_{h,p}\max\{\sigma, L_J\}\, c_{p,2}(m)c_{2,p}(n)^2\Delta_0}{(1-\alpha)\theta}\epsilon^{-1}\right) + 1. && (47)
\end{aligned}
$$

Proof. If $T_g(\epsilon) \leq 1$, then (47) is clearly true. Let us assume that $T_g(\epsilon) \geq 2$. By (34),

$$T_g(\epsilon) \leq \left|\mathcal{S}_{T_g(\epsilon)-1}\right| + \left|\mathcal{U}^{(1)}_{T_g(\epsilon)-1} \cup \mathcal{U}^{(3)}_{T_g(\epsilon)-1}\right| + \left|\mathcal{U}^{(2)}_{T_g(\epsilon)-1} \cup \mathcal{U}^{(3)}_{T_g(\epsilon)-1}\right|.$$

Then, (47) follows directly from Lemmas 3.4, 3.5 and 3.7, together with (41). □

Since each iteration of TRFD requires at most $(n+1)$ evaluations of $F(\,\cdot\,)$, from Theorem 3.8 we obtain the following upper bound on the total number of evaluations of $F(\,\cdot\,)$ required by TRFD to find an ϵ-approximate stationary point.

Corollary 3.9. *Suppose that A1-A4 hold and let $FE_{T_g(\epsilon)-1}$ be the total number of function evaluations executed by TRFD up to the $(T_g(\epsilon)-1)$-st iteration. Then*

$$
\begin{aligned}
FE_{T_g(\epsilon)-1} \leq\ & (n+1)\left[\frac{16L_{h,p}\max\{\sigma, L_J\}\, c_{p,2}(m)c_{2,p}(n)^2(f(x_0) - f_{low})}{\alpha(1-\alpha)\theta^2}\epsilon^{-2}\right. \\
& \left. + \left\lceil\left|\log_2\left(\frac{4\max\{\sigma, L_J\}\, c_{2,p}(n)}{\sigma(1-\alpha)\theta}\right)\right|\right\rceil + \log_2\left(\frac{4L_{h,p}\max\{\sigma, L_J\}\, c_{p,2}(m)c_{2,p}(n)^2\Delta_0}{(1-\alpha)\theta}\epsilon^{-1}\right) + 1\right].
\end{aligned}
$$

In view of Corollary 3.9, TRFD needs no more than

$$\mathcal{O}\left(n\, c_{2,p}(n)^2 c_{p,2}(m) L_{h,p} L_J (f(x_0) - f_{low})\epsilon^{-2}\right)$$

evaluations of $F(\cdot)$ to find x_k such that $\psi_{p,\Delta_*}(x_k) \leq \epsilon$. In what follows, Tables 1 and 2 specify this complexity bound for the L1 and Minimax problems with respect to different choices of the p-norm used in TRFD.

L1 Problems: case $h(z) = \|z\|_1\ \forall z \in \mathbb{R}^m$				
p-norm in TRFD	$L_{h,p}$	$c_{p,2}(m)$	$c_{2,p}(n)$	Evaluation Complexity Bound
$p = 1$	1	$\sqrt{m}$	1	$\mathcal{O}\left(n\sqrt{m} L_J (f(x_0) - f_{low})\epsilon^{-2}\right)$
$p = 2$	$\sqrt{m}$	1	1	$\mathcal{O}\left(n\sqrt{m} L_J (f(x_0) - f_{low})\epsilon^{-2}\right)$
$p = \infty$	m	1	$\sqrt{n}$	$\mathcal{O}\left(n^2 m L_J (f(x_0) - f_{low})\epsilon^{-2}\right)$

Table 1: Complexity bounds for problems with objective function of the form $f(\cdot) = \|F(\cdot)\|_1$.

Minimax Problems: case $h(z) = \max_{i=1,\ldots,m}\{z_i\}\ \forall z \in \mathbb{R}^m$				
p-norm in TRFD	$L_{h,p}$	$c_{p,2}(m)$	$c_{2,p}(n)$	Evaluation Complexity Bound
$p = 1$	1	$\sqrt{m}$	1	$\mathcal{O}\left(n\sqrt{m} L_J (f(x_0) - f_{low})\epsilon^{-2}\right)$
$p = 2$	1	1	1	$\mathcal{O}\left(n L_J (f(x_0) - f_{low})\epsilon^{-2}\right)$
$p = \infty$	1	1	$\sqrt{n}$	$\mathcal{O}\left(n^2 L_J (f(x_0) - f_{low})\epsilon^{-2}\right)$

Table 2: Complexity bounds for problems with objective function of the form $f(\cdot) = \max_{i=1,\ldots,m}\{F_i(\cdot)\}$.

Notice that in both cases, considering TRFD with $p = 1$ or $p = 2$, we obtain evaluation complexity bounds of $\mathcal{O}\left(n\epsilon^{-2}\right)$, with linear dependence on the number of variables n. This represents an improvement over the bound of $\mathcal{O}\left(n^2\epsilon^{-2}\right)$ established in [8] for a model-based derivative-free trust-region method for composite nonsmooth optimization.

3.2 Worst-Case Complexity Bound for Convex Problems

Let us consider the additional assumptions:

A5. $F_i(\cdot)$ is convex for $i = 1, \ldots, m$.

A6. $h(\cdot)$ is monotone, i.e., $h(u) \leq h(v)$ if $u_i \leq v_i$ for $i = 1, \ldots, m$.

A7. $f(\cdot) = h(F(\cdot))$ has a global minimizer x^* on Ω and

$$D_0 \equiv \sup_{x \in \mathcal{L}_f(x_0)} \{\|x - x^*\|_p\} < +\infty,$$

for $\mathcal{L}_f(x_0) = \{x \in \Omega\, :\, f(x) \leq f(x_0)\}$.

The lemma below establishes the relationship between the stationarity measure and the functional residual when the reference radius r is sufficiently large.

Lemma 3.10. *Suppose that A1, A2, A5, A6 and A7 hold, and let $x_k \in \mathcal{L}_f(x_0)$. If $r \geq D_0$, then*

$$\psi_{p,r}(x_k) \geq \frac{1}{r}(f(x_k) - f(x^*)).$$

Proof. By A2 and A5, for each $i \in \{1, \ldots, m\}$ we have

$$F_i(x_k + s) \geq F(x_k) + \langle \nabla F_i(x_k), s \rangle, \quad \forall s \in \Omega - \{x_k\}.$$

Thus, it follows from A6 that

$$h(F(x_k + s)) \geq h(F(x_k) + J_F(x_k)s), \quad \forall s \in \Omega - \{x_k\},$$

and so

$$\min_{\substack{s \in \Omega - \{x_k\} \\ \|s\|_p \leq r}} h(F(x_k + s)) \geq \min_{\substack{s \in \Omega - \{x_k\} \\ \|s\|_p \leq r}} h(F(x_k) + J_F(x_k)s). \tag{48}$$

Let $s^* = x^* - x_k$. Then $s^* \in \Omega - \{x_k\}$ and, by A7 and $r \geq D_0$, we also have $\|s^*\|_p \leq D_0 \leq r$. Therefore

$$\min_{\substack{s \in \Omega - \{x_k\} \\ \|s\|_p \leq r}} h(F(x_k + s)) = h(F(x_k + s^*)) = h(F(x^*)) = f(x^*). \tag{49}$$

Combining (48) and (49), we obtain

$$f(x^*) \geq \min_{\substack{s \in \Omega - \{x_k\} \\ \|s\|_p \leq r}} h(F(x_k) + J_F(x_k)s),$$

which implies that

$$\psi_{p,r}(x_k) = \frac{1}{r}\left(h(F(x_k)) - \min_{\substack{s \in \Omega - \{x_k\} \\ \|s\|_p \leq r}} h(F(x_k) + J_F(x_k)s) \right) \geq \frac{1}{r}\left(f(x_k) - f(x^*) \right),$$

which concludes the proof. □

The lemma below provides a lower bound on the approximate stationarity measure in terms of the functional residual.

Lemma 3.11. *Suppose that A1-A3 and A5-A7 hold, and let $\{x_k\}$ be generated by TRFD. If $\Delta_* \geq D_0$, then*

$$\eta_{p,\Delta_*}(x_k; A_k) \geq \frac{f(x_k) - f(x^*)}{\left(\frac{L_J}{\sigma} + 1\right)\Delta_*}$$

whenever $k \notin \mathcal{U}^{(1)}$.

Proof. Suppose that $k \notin \mathcal{U}^{(1)}$. In this case, we have $\eta_{p,\Delta_*}(x_k; A_k) \geq \epsilon/2$. Then, it follows from Lemma 2.9 and from the definition of τ_0 at Step 0 of TRFD that

$$\begin{aligned}
\psi_{p,\Delta_*}(x_k) &\leq |\psi_{p,\Delta_*}(x_k) - \eta_{p,\Delta_*}(x_k; A_k)| + |\eta_{p,\Delta_*}(x_k; A_k)| \\
&\leq \frac{L_{h,p} L_J c_{p,2}(m) c_{2,p}(n)\sqrt{n}}{2}\tau_k + \eta_{p,\Delta_*}(x_k; A_k) \\
&\leq \frac{L_{h,p} L_J c_{p,2}(m) c_{2,p}(n)\sqrt{n}}{2}\tau_0 + \eta_{p,\Delta_*}(x_k; A_k) \\
&= \frac{L_J \epsilon}{2\sigma} + \eta_{p,\Delta_*}(x_k; A_k) \\
&\leq \left(\frac{L_J}{\sigma} + 1\right)\eta_{p,\Delta_*}(x_k; A_k).
\end{aligned}$$

Therefore, by Lemma 3.10, we obtain

$$\eta_{p,\Delta_*}(x_k;A_k) \geq \frac{\psi_{p,\Delta_*}(x_k)}{\left(\frac{L_J}{\sigma}+1\right)} \geq \frac{f(x_k)-f(x^*)}{\left(\frac{L_J}{\sigma}+1\right)\Delta_*}.$$

Therefore, the statement is proved. □

Next we establish an upper bound for $\frac{f(x_k)-f(x^*)}{\Delta_k}$ when the functional residual is sufficiently large.

Lemma 3.12. *Suppose that A1-A3 and A5-A7 hold, and let $\{x_k\}_{k=0}^{T}$ be generated by TRFD. If $\Delta_* \geq \Delta_0$ and*

$$f(x_k)-f(x^*) > \Delta_*\epsilon \quad \text{for } k=0,\ldots,T-1, \tag{50}$$

then

$$\left(\frac{1}{\Delta_k}\right)(f(x_k)-f(x^*)) \leq \max\left\{\left(\frac{1}{\Delta_0}\right)(f(x_0)-f(x^*)), \frac{2\left(\frac{L_J}{\sigma}+1\right)\Delta_* L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)^2}{(1-\alpha)\theta}\right\} \equiv \beta \tag{51}$$

for $k=0,\ldots,T$.

Proof. By the definition of β, (51) is true for $k=0$. Suppose that (51) is true for some $k \in \{0,\ldots,T-1\}$. Let us show that it is also true for $k+1$.

Case 1: $k \in \mathcal{U}^{(1)} \cup \mathcal{S}$

In this case, we have $\Delta_{k+1} \geq \Delta_k$. Since $f(x_{k+1}) \leq f(x_k)$, it follows that

$$\left(\frac{1}{\Delta_{k+1}}\right)(f(x_{k+1})-f(x^*)) \leq \left(\frac{1}{\Delta_k}\right)(f(x_k)-f(x^*)) \leq \beta,$$

where the last inequality is the induction assumption. Therefore, (51) holds for $k+1$ in this case.

Case 2: $k \in \mathcal{U}^{(2)} \cup \mathcal{U}^{(3)}$

In this case we have

$$\Delta_{k+1} = \frac{1}{2}\Delta_k. \tag{52}$$

In addition, in view of (50) and $\Delta_* \geq D_0$, it follows from Lemma 3.10 that $\psi_{p,\Delta_*}(x_k) > \epsilon$. Therefore, we must have

$$\Delta_k > \frac{(1-\alpha)\theta\eta_{p,\Delta_*}(x_k;A_k)}{L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)^2} \tag{53}$$

since otherwise, by Lemma 3.2, we would have $k \in \mathcal{S}$, contradicting our assumption that $k \in \mathcal{U}^{(2)} \cup \mathcal{U}^{(3)}$. Notice that (53) is equivalent to

$$\left(\frac{1}{\Delta_k}\right)\eta_{p,\Delta_*}(x_k;A_k) < \frac{L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)^2}{(1-\alpha)\theta}.$$

Finally, it follows from (52), Lemma 3.11 and (53) that

$$\begin{aligned}\left(\frac{1}{\Delta_{k+1}}\right)(f(x_{k+1})-f(x^*)) &= \left(\frac{2}{\Delta_k}\right)(f(x_{k+1})-f(x^*)) \leq \left(\frac{2}{\Delta_k}\right)(f(x_k)-f(x^*)) \\ &\leq \frac{2\left(\frac{L_J}{\sigma}+1\right)\Delta_*}{\Delta_k}\eta_{p,\Delta_*}(x_k;A_k) \\ &< 2\left(\frac{L_J}{\sigma}+1\right)\Delta_*\frac{L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)^2}{(1-\alpha)\theta} \\ &= \frac{2\left(\frac{L_J}{\sigma}+1\right)\Delta_* L_{h,p}L_J c_{p,2}(m)c_{2,p}(n)^2}{(1-\alpha)\theta} \\ &\leq \beta,\end{aligned}$$

that is, (51) also holds for $k+1$ in this case. □

Let

$$T_f(\epsilon)=\inf\{k\in\mathbb{N}\,:\,f(x_k)-f(x^*)\leq\Delta_*\epsilon\} \tag{54}$$

be the index of the first iteration in which $\{x_k\}_{k\geq 0}$ reaches a $\Delta_*\epsilon$-approximate solution of (1), if it exists. Our goal is to establish a finite upper bound for $T_f(\epsilon)$. In this context, the lemma below provides an upper bound on $\left|\mathcal{S}_{T_f(\epsilon)-1}\right|$.

Lemma 3.13. *Suppose that A1-A7 hold. Given $\epsilon>0$, if $T_f(\epsilon)\geq 2$ and $\Delta_*\geq D_0$, then*

$$\left|\mathcal{S}_{T_f(\epsilon)-1}\right| \leq 1+\frac{\left(\frac{L_J}{\sigma}+1\right)\beta}{\alpha\theta}\epsilon^{-1}, \tag{55}$$

where β is defined in (51).

Proof. Let $k\in\mathcal{S}_{T_f(\epsilon)-2}$. By Lemmas 2.11, 3.11 and 3.12, we have

$$\begin{aligned}f(x_k)-f(x_{k+1}) &\geq \alpha\left[h(F(x_k))-h(F(x_k)+A_kd_k)\right] \\ &\geq \alpha\theta\left[h(F(x_k))-h(F(x_k)+A_kd_k^*)\right] \\ &= \alpha\theta\Delta_k\eta_{p,\Delta_k}(x_k;A_k) \\ &\geq \alpha\theta\Delta_k\eta_{p,\Delta_*}(x_k;A_k) \\ &\geq \alpha\theta\Delta_k\frac{f(x_k)-f(x^*)}{\left(\frac{L_J}{\sigma}+1\right)\Delta_*} \\ &= \frac{\alpha\theta(f(x_k)-f(x^*))^2}{\left(\frac{L_J}{\sigma}+1\right)\Delta_*\left(\frac{1}{\Delta_k}\right)(f(x_k)-f(x^*))} \\ &\geq \frac{\alpha\theta(f(x_k)-f(x^*))^2}{\left(\frac{L_J}{\sigma}+1\right)\Delta_*\beta}.\end{aligned} \tag{56}$$

Denoting $\delta_k=f(x_k)-f(x^*)$, (56) becomes

$$\delta_k-\delta_{k+1}\geq\frac{\alpha\theta}{\left(\frac{L_J}{\sigma}+1\right)\Delta_*\beta}\delta_k^2.$$

Consequently,

$$\frac{1}{\delta_{k+1}} - \frac{1}{\delta_k} = \frac{\delta_k - \delta_{k+1}}{\delta_k \delta_{k+1}} \geq \frac{\frac{\alpha\theta}{\left(\frac{L_J}{\sigma}+1\right)\Delta_* \beta}\delta_k^2}{\delta_k^2} = \frac{\alpha\theta}{\left(\frac{L_J}{\sigma}+1\right)\Delta_* \beta}. \tag{57}$$

Since $\delta_{k+1} = \delta_k$ for any $k \in \{0, ..., T_f(\epsilon) - 2\} \setminus \mathcal{S}_{T_f(\epsilon)-2}$, it follows from (57) that

$$\begin{aligned} \frac{1}{\delta_{T_f(\epsilon)-1}} - \frac{1}{\delta_0} &= \sum_{k=0}^{T_f(\epsilon)-2} \frac{1}{\delta_{k+1}} - \frac{1}{\delta_k} = \sum_{k \in \mathcal{S}_{T_f(\epsilon)-2}} \frac{1}{\delta_{k+1}} - \frac{1}{\delta_k} \\ &\geq \left|\mathcal{S}_{T_f(\epsilon)-2}\right| \frac{\alpha\theta}{\left(\frac{L_J}{\sigma}+1\right)\Delta_* \beta}. \end{aligned}$$

Therefore

$$\Delta_* \epsilon < f(x_{T_f(\epsilon)-1}) - f(x^*) = \delta_{T_f(\epsilon)-1} \leq \frac{\left(\frac{L_J}{\sigma}+1\right)\Delta_* \beta}{\alpha\theta \left|\mathcal{S}_{T_f(\epsilon)-2}\right|},$$

which implies that

$$\left|\mathcal{S}_{T_f(\epsilon)-1}\right| \leq 1 + \left|\mathcal{S}_{T_f(\epsilon)-2}\right| \leq 1 + \frac{\left(\frac{L_J}{\sigma}+1\right)\beta}{\alpha\theta}\epsilon^{-1},$$

that is, (55) is true. □

The next lemma establishes the relationship between $T_f(\epsilon)$ and $T_g(\epsilon)$.

Lemma 3.14. *Suppose that A2, A5, A6 and A7 hold, and let $T_f(\epsilon)$ and $T_g(\epsilon)$ be defined by* (54) *and* (33)*, respectively. If $\Delta_* \geq D_0$, then $T_f(\epsilon) \leq T_g(\epsilon)$.*

Proof. Suppose by contradiction that $T_f(\epsilon) > T_g(\epsilon)$. In this case, by $\Delta_* \geq D_0$ and Lemma 3.10, we would arrive at the contradiction

$$\epsilon < \frac{1}{\Delta_*}(f(x_{T_g(\epsilon)}) - f(x^*)) \leq \psi_{p,\Delta_*}(x_{T_g(\epsilon)}) \leq \epsilon.$$

Therefore, we must have $T_f(\epsilon) \leq T_g(\epsilon)$. □

The following theorem gives an upper bound on the number of iterations required by TRFD to reach a $\Delta_*\epsilon$-approximate solution of (1).

Theorem 3.15. *Suppose that A1-A7 and let $T_f(\epsilon)$ be defined by* (54)*. If $\Delta_* \geq D_0$, then*

$$\begin{aligned} T_f(\epsilon) \leq & 2\left[1 + \frac{(\frac{L_J}{\sigma}+1)\beta}{\alpha\theta}\epsilon^{-1}\right] + \left\lceil \left|\log_2\left(\frac{4\max\{\sigma, L_J\}\, c_{2,p}(n)}{\sigma(1-\alpha)\theta}\right)\right| \right. \\ & \left. + \log_2\left(\frac{4L_{h,p}\max\{\sigma, L_J\}\, c_{p,2}(m) c_{2,p}(n)^2 \Delta_0}{(1-\alpha)\theta}\epsilon^{-1}\right)\right\rceil, \end{aligned} \tag{58}$$

where β is defined in (51).

Proof. If $T_f(\epsilon) \leq 1$, then (58) is true. Let us assume that $T_f(\epsilon) \geq 2$. As in the proof of Theorem 3.8, we have

$$T_f(\epsilon) \leq \left|\mathcal{S}_{T_f(\epsilon)-1}\right| + \left|\mathcal{U}^{(1)}_{T_f(\epsilon)-1} \cup \mathcal{U}^{(3)}_{T_f(\epsilon)-1}\right| + \left|\mathcal{U}^{(2)}_{T_f(\epsilon)-1} \cup \mathcal{U}^{(3)}_{T_f(\epsilon)-1}\right|. \tag{59}$$

By Lemma 3.14, we have $T_f(\epsilon) \leq T_g(\epsilon)$. Thus, it follows from Lemmas 3.5 and 3.7 that

$$\left|\mathcal{U}^{(1)}_{T_f(\epsilon)-1} \cup \mathcal{U}^{(3)}_{T_f(\epsilon)-1}\right| \leq \left\lceil \left|\log_2\left(\frac{\tau_0\sqrt{n}}{\Delta_{\min}(\epsilon)}\right)\right| \right\rceil \tag{60}$$

and

$$\left|\mathcal{U}^{(2)}_{T_f(\epsilon)-1} \cup \mathcal{U}^{(3)}_{T_f(\epsilon)-1}\right| \leq \log_2\left(\frac{4L_{h,p}\max\{\sigma, L_J\}\, c_{p,2}(m)c_{2,p}(n)^2\Delta_0}{(1-\alpha)\theta}\epsilon^{-1}\right) + |\mathcal{S}_{T_f(\epsilon)-1}|, \tag{61}$$

where $\Delta_{\min}(\epsilon)$ is defined in (29). Then, combining (59), Lemma 3.13, (60), (61) and (41), we conclude that (58) is true. □

Since each iteration of TRFD requires at most $(n+1)$ evaluations of $F(\,\cdot\,)$, from Theorem 3.15 we obtain the following upper bound on the total number of evaluations of $F(\,\cdot\,)$ required by TRFD to find a $\Delta_*\epsilon$-approximate solution of (1).

Corollary 3.16. *Suppose that A1-A7 hold and let $FE_{T_f(\epsilon)-1}$ be the total number of function evaluations executed by TRFD up to the $(T_f(\epsilon)-1)$-st iteration. If $\Delta_* \geq D_0$, then*

$$\begin{aligned} FE_{T_f(\epsilon)-1} \quad &\leq \quad (n+1)\left[2\left[1 + \frac{(\frac{L_J}{\sigma}+1)\beta}{\alpha\theta}\epsilon^{-1}\right] + \left\lceil\left|\log_2\left(\frac{4\max\{\sigma, L_J\}\, c_{2,p}(n)}{\sigma(1-\alpha)\theta}\right)\right|\right\rceil \right.\\ &\qquad \left. + \log_2\left(\frac{4L_{h,p}\max\{\sigma, L_J\}\, c_{p,2}(m)c_{2,p}(n)^2\Delta_0}{(1-\alpha)\theta}\epsilon^{-1}\right)\right]. \end{aligned}$$

In view of Corollary 3.16 and the definition of β in (51), if $h(\,\cdot\,)$ is monotone and the components $F_i(\,\cdot\,)$ are convex, then TRFD, with a sufficiently large Δ_*, needs no more than

$$\mathcal{O}\left(n\, c_{2,p}(n)^2 c_{p,2}(m) L_{h,p} L_J \Delta_* \epsilon^{-1}\right)$$

function evaluations to find x_k such that

$$f(x_k) - f(x^*) \leq \Delta_*\epsilon.$$

Thus, given $\epsilon_f > 0$, if we use TRFD with $\epsilon = \epsilon_f/\Delta_*$, then it will need no more than

$$\mathcal{O}\left(n\, c_{2,p}(n)^2 c_{p,2}(m) L_{h,p} L_J \Delta_*^2 \epsilon_f^{-1}\right)$$

function evaluations to find x_k such that

$$f(x_k) - f(x^*) \leq \epsilon_f.$$

Table 3 below specifies the complexity bound for the Minimax problem, which is a composite nonsmooth problem of the form (1) whose function $h(\,\cdot\,)$ is monotone.

Minimax Problems: case $h(z) = \max_{i=1,\ldots,m}\{z_i\}\ \forall z \in \mathbb{R}^m$				
p-norm in TRFD	$L_{h,p}$	$c_{p,2}(m)$	$c_{2,p}(n)$	Evaluation Complexity Bound
$p=1$	1	$\sqrt{m}$	1	$\mathcal{O}\left(n\sqrt{m}L_J\Delta_*^2\epsilon_f^{-1}\right)$
$p=2$	1	1	1	$\mathcal{O}\left(nL_J\Delta_*^2\epsilon_f^{-1}\right)$
$p=\infty$	1	1	$\sqrt{n}$	$\mathcal{O}\left(n^2L_J\Delta_*^2\epsilon_f^{-1}\right)$

Table 3: Complexity bounds for problems with objective function of the form $f(\cdot) = \max_{i=1,\ldots,m}\{F_i(\cdot)\}$.

When Ω is a polyhedron, for $p=1$ and $p=\infty$, the computation of $\eta_{p,\Delta_*}(x_k;\Delta_k)$ and d_k in TRFD can be performed by solving linear programming problems. The complexity bounds in Table 3 suggest that one should use $p=1$ when $\sqrt{m}<n$, and $p=\infty$ otherwise. On the other hand, the best complexity bound, of $\mathcal{O}\left(n\epsilon^{-1}\right)$, is obtained with $p=2$. However, in this case, the computation of $\eta_{p,\Delta_*}(x_k;\Delta_k)$ and d_k requires solving linear problems subject to a quadratic constraint.

4 Numerical experiments

We performed numerical experiments with Matlab implementations of TRFD. Specifically, two classes of test problems were considered: unconstrained L1 problems (see subsection 4.1) and unconstrained Minimax problems (see subsection 4.2). We compared TRFD against Manifold Sampling Primal [13] and against the derivative-free trust-region method proposed in [11]. For each problem, a budget of 100 simplex gradients was allowed to each solver[2]. In addition, our implementations of TRFD were equipped with the following stopping criteria:

$$\Delta_k \leq 10^{-13} \quad \text{or} \quad \eta_{p,\Delta_*}(x_k;A_k) \leq 10^{-13}.$$

Implementations are compared using data profiles [18][3], where a code M is said to solve a given problem when it reaches x_M such that

$$\frac{f(x_0)-f(x_M)}{f(x_0)-f(x_{Best})} \geq 1 - \textit{Tolerance},$$

where $f(x_{Best})$ is the lowest function value found among all the methods. All experiments were performed with MATLAB (R2023a) on a PC with microprocessor 13-*th* Gen Intel(R) Core(TM) i5-1345U 1.60 GHz and 32 GB of RAM memory.

4.1 L1 problems

Here we considered problems of the form

$$\min_{x\in\mathbb{R}^n} \|F(x)\|_1.$$

[2]One simplex gradient corresponds to $n+1$ function evaluations, with n being the number of variables of the problem.

[3]The data profiles were generated using the code *data_profile.m*, freely available at the website `https://www.mcs.anl.gov/~more/dfo/`.

We tested 53 functions $F : \mathbb{R}^n \to \mathbb{R}^m$ defined by Moré and Wild [18], for which $2 \leq n \leq 12$ and $2 \leq m \leq 65$. The following codes were compared:

- **TRFD-L1**: Implementation of TRFD with $p = 1$ and parameters $\epsilon = 10^{-15}$, $\alpha = 0.15$, $L_{h,p} = 1$, $\Delta_0 = \max\{1, \tau_0\sqrt{n}\}$, $\Delta_* = 1000$ and

$$\sigma = \frac{\epsilon}{L_{h,p} c_{p,2}(m) c_{2,p}(n) \sqrt{n}\sqrt{eps}},$$

where eps is the machine precision, $c_{1,2}(m) = \sqrt{m}$ and $c_{2,1}(n) = 1$. The computation of $\eta_{p,\Delta_*}(x_k; \Delta_k)$ and d_k is performed using the MATLAB function *linprog.m*.

- **MS-P**: Implementation of Manifold Sampling Primal [13], freely available on GitHub[4]. The initial parameters are given in the file *check_inputs_and_initialize.m*, while the outer function $h(\cdot)$ was provided by the file *one_norm.m*.

- **DFL1S**: Implementation of the trust-region method in [11] adapted to the case $h(\cdot) = \|\cdot\|_1$.

Data profiles are shown in Figure 1. As we can see, in this particular test set, TRFD-L1 outperforms both MS-P and DFL1S, being able to solve a higher percentage of problems within the allowed budget of $100(n+1)$ evaluations of $F(\cdot)$ across all the tolerances considered.

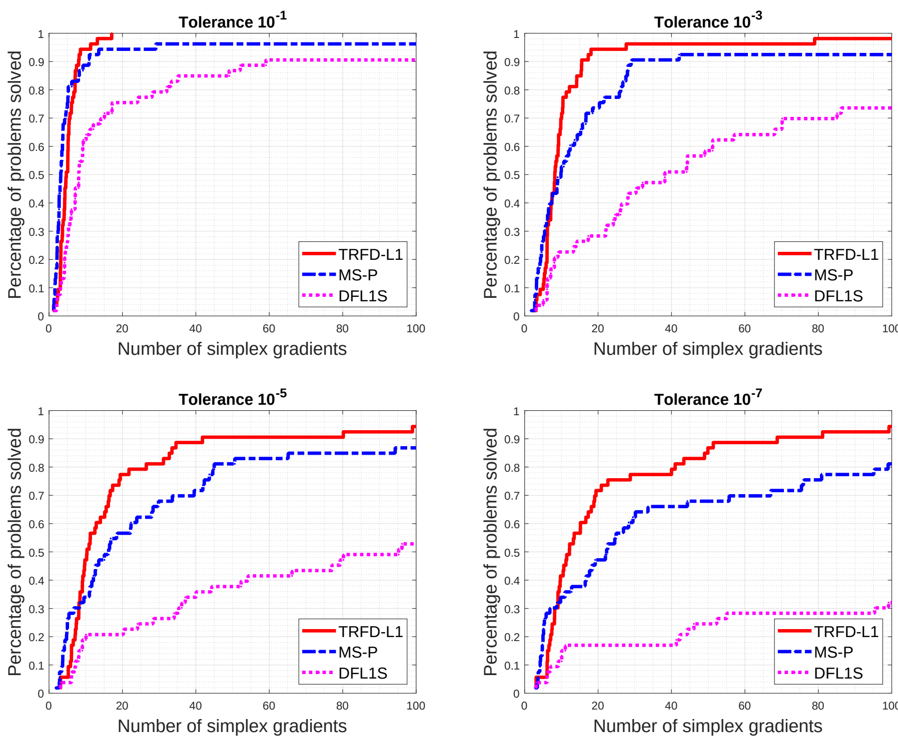

Figure 1: Data profiles of TRFD-L1, MS-P and DFL1S on L1 problems

[4]`https://github.com/POptUS/IBCDFO`.

4.2 Minimax problems

We also considered problems of the form

$$\min_{x\in\mathbb{R}^n}\max_{i=1,\ldots,m}\{F_i(x)\}.$$

We tested 43 functions $F:\mathbb{R}^n\to\mathbb{R}^m$ defined by Lukšan and Vlcek [16] and Di Pillo et al. [6], for which $2\leq n\leq 50$ and $2\leq m\leq 130$. On these problems, the following codes were compared:

- **TRFD-M** Implementation of TRFD with $p=1$ if $\sqrt{m}<n$, and $p=\infty$ if $\sqrt{m}\geq n$. Parameters are the same used in TRFD-L1, with constants $L_{h,1}=1$, $L_{h,\infty}=1$, $c_{1,2}(m)=\sqrt{m}$, $c_{2,1}(n)=1$, $c_{\infty,2}(m)=1$ and $c_{2,\infty}(n)=\sqrt{n}$. Subproblems are solved using the MATLAB function *linprog.m*.

- **TRFD-M2** Implementation of TRFD with $p=2$. Parameters are the same used in TRFD-L1, with constants $L_{h,2}=1$, $c_{2,2}(m)=1$ and $c_{2,2}(n)=1$. Subproblems are solved using the MATLAB function *fmincon.m*.

- **MS-P**: Implementation of Manifold Sampling Primal [13], with the outer function $h(\cdot)$ provided by the file *pw_maximum*.

- **DFMS**: Implementation described in Section 7 of [11].

Figure 2 presents the data profiles comparing TRFD-M, MS-P and DFMS. As shown, TRFD-M and MS-P exhibited similar performances and both outperformed DFMS across all tolerances considered.

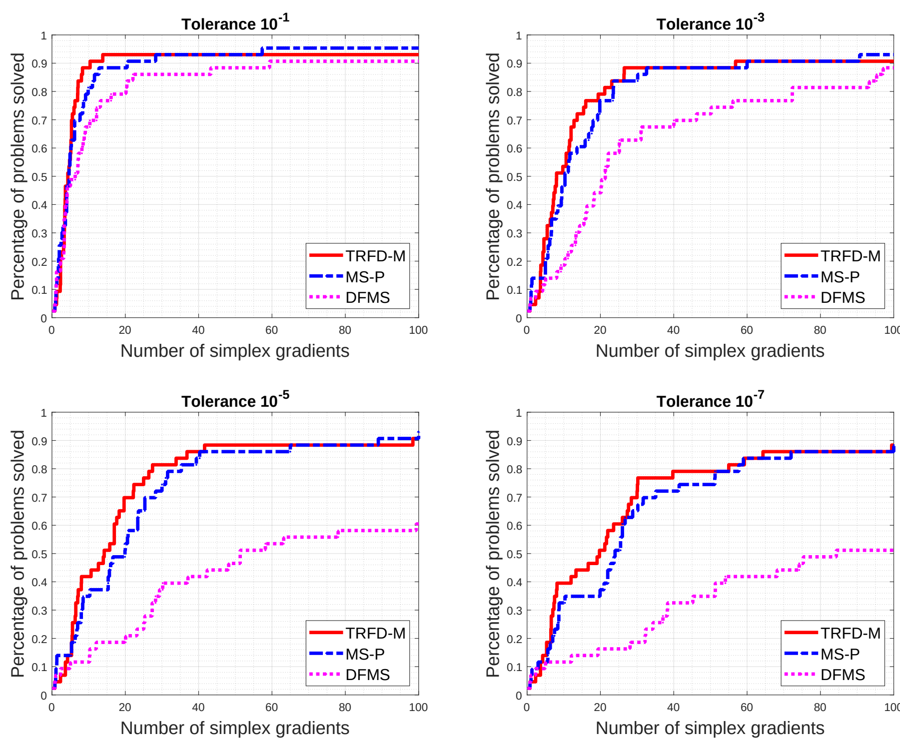

Figure 2: Data profiles of TRFD-M, MS-P and DFMS on Minimax problems

We also compared TRFD-M2 against MS-P and TRFD-M. The data profiles are shown in Figure 3. For tolerances 10^{-3} and 10^{-5}, TRFD-M2 performed slightly better than MS-P and TRFD-M. However, for tolerance 10^{-7}, both MS-P and TRFD-M outperformed TRFD-M2.

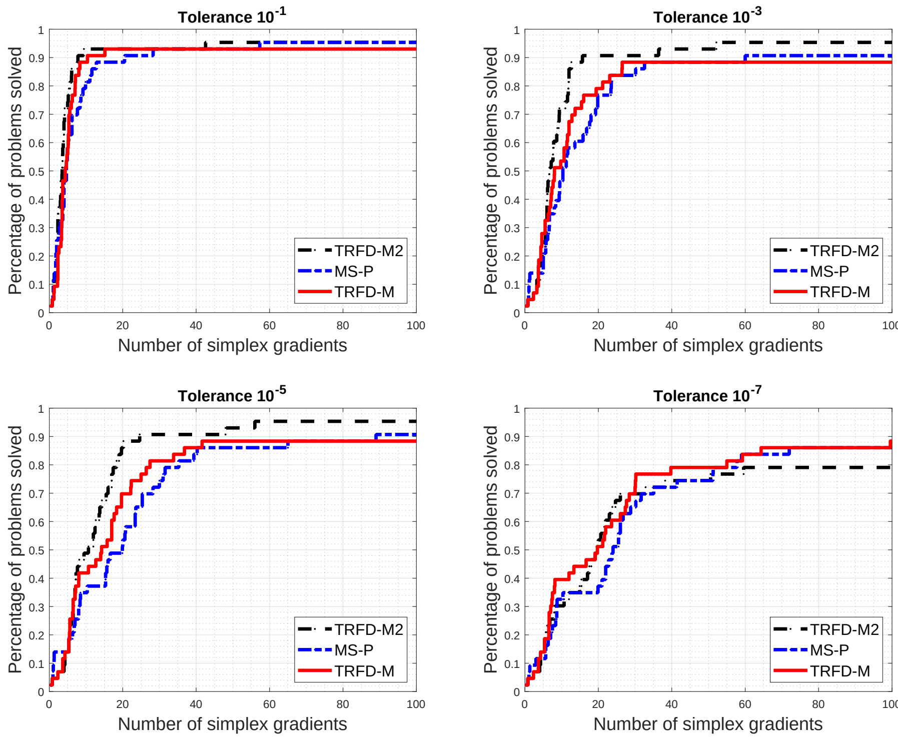

Figure 3: Data profiles of TRFD-M2, MS-P and TRFD-M on Minimax problems

5 Conclusion

In this paper, we introduced TRFD, a derivative-free trust-region method for minimizing composite functions of the form $f(x) = h(F(x))$ over a convex set Ω. In the proposed method, trial points are obtained by minimizing models of the form $h(M_k(x_k + d))$ subject to the constraints $\|d\|_p \leq \Delta_k$ and $x_k + d \in \Omega$. Unlike existing model-based derivative-free methods for composite nonsmooth optimization, in which $M_k(x_k + d)$ is built as a linear or quadratic interpolation model of F around x_k, TRFD employs $M_k(x_k + d) = F(x_k) + A_k d$, where A_k is an approximation for the Jacobian of F at x_k, constructed using finite differences defined by a stepsize τ_k. Special rules for updating τ_k and Δ_k allowed us to establish improved evaluation complexity bounds for TRFD in the nonconvex case. In particular, for L1 and Minimax problems, we proved that TRFD with $p = 1$ and $p = 2$ requires no more than $O(n\epsilon^{-2})$ evaluations of $F(\,\cdot\,)$ to find an ϵ-approximate stationary point. Moreover, under the assumptions that $h(\,\cdot\,)$ is monotone and that the components of $F(\,\cdot\,)$ are convex, we established a complexity bound for the number of evaluations of $F(\,\cdot\,)$ that TRFD requires to find an ϵ-approximate minimizer of $f(\,\cdot\,)$ on Ω. For Minimax problems, our bound reduces to $O(n\epsilon^{-1})$ when we use $p = 1$ or $p = 2$ in TRFD. We concluded by presenting numerical results comparing implementations of TRFD against two model-based derivative-free trust-region methods, namely, Manifold Sampling

[13] and the derivative-free method from [11]. For L1 problems, TRFD outperformed the other two solvers, while for Minimax problems, TRFD demonstrated a competitive performance with Manifold Sampling.

Acknowledgments

We are grateful to Jeffrey Larson for interesting discussions about composite optimization and for his assistance with the MS-P code.

References

[1] C. Audet and W. Hare, *Introduction: tools and challenges in derivative-free and blackbox optimization*, Springer, 2017.

[2] C. Audet and D. Orban, *Finding optimal algorithmic parameters using derivative-free optimization*, SIAM J. Optim., 17 (2006), pp. 642–664.

[3] C. Cartis, N. I. Gould, and P. L. Toint, *On the evaluation complexity of composite function minimization with applications to nonconvex nonlinear programming*, SIAM J. Optim., 21 (2011), pp. 1721–1739.

[4] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Global convergence of general derivative-free trust-region algorithms to first-and second-order critical points*, SIAM J. Optim., 20 (2009), pp. 387–415.

[5] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to derivative-free optimization*, SIAM, 2009.

[6] G. Di Pillo, L. Grippo, and S. Lucidi, *A smooth method for the finite minimax problem*, Math. Program., 60 (1993), pp. 187–214.

[7] R. Fletcher, *A model algorithm for composite nondifferentiable optimization problems*, in Nondifferential and Variational Techniques in Optimization, Springer, 1982, pp. 67–76.

[8] R. Garmanjani, D. Júdice, and L. N. Vicente, *Trust-region methods without using derivatives: worst case complexity and the nonsmooth case*, SIAM J. Optim., 26 (2016), pp. 1987–2011.

[9] G. N. Grapiglia, *Quadratic regularization methods with finite-difference gradient approximations*, Comput. Optim. Appl., 85 (2023), pp. 683–703.

[10] G. N. Grapiglia, *Worst-case evaluation complexity of a derivative-free quadratic regularization method*, Optim. Lett., 18 (2024), pp. 195–213.

[11] G. N. Grapiglia, J. Yuan, and Y.-x. Yuan, *A derivative-free trust-region algorithm for composite nonsmooth optimization*, Comput. Appl. Math., 35 (2016), pp. 475–499.

[12] H. R. Karbasian and B. C. Vermeire, *Gradient-free aerodynamic shape optimization using large eddy simulation*, Comput. & Fluids, 232 (2022), p. 105185.

[13] J. Larson and M. Menickelly, *Structure-aware methods for expensive derivative-free nonsmooth composite optimization*, Math. Program. Comput., 16 (2024), pp. 1–36.

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

[15] J. Larson, M. Menickelly, and B. Zhou, *Manifold sampling for optimizing nonsmooth nonconvex compositions*, SIAM J. Optim., 31 (2021), pp. 2638–2664.

[16] L. Lukšan and J. Vlcek, *Test problems for nonsmooth unconstrained and linearly constrained optimization*, Technical report 798, Institute of Computer Science, Academy of Sciences of the Czech Republic, Prague, 2000.

[17] A. L. Marsden, J. A. Feinstein, and C. A. Taylor, *A computational framework for derivative-free optimization of cardiovascular geometries*, Comput. Methods Appl. Mech. Engrg., 197 (2008), pp. 1890–1905.

[18] J. J. Moré and S. M. Wild, *Benchmarking derivative-free optimization algorithms*, SIAM J. Optim., 20 (2009), pp. 172–191.

[19] J. B. Russ, R. L. Li, A. R. Herschman, H. Waisman, V. Vedula, J. W. Kysar, and D. Kalfa, *Design optimization of a cardiovascular stent with application to a balloon expandable prosthetic heart valve*, Mater. Des., 209 (2021), p. 109977.

[20] Y.-x. Yuan, *Conditions for convergence of trust region algorithms for nonsmooth optimization*, Math. Program., 31 (1985), pp. 220–228.

[21] Y.-x. Yuan, *On the superlinear convergence of a trust region algorithm for nonsmooth optimization*, Math. Program., 31 (1985), pp. 269–285.