Journal of Scientific Computing (2024) 101:45
https://doi.org/10.1007/s10915-024-02679-y

A Quasi-Newton Subspace Trust Region Algorithm for Nonmonotone Variational Inequalities in Adversarial Learning over Box Constraints

Zicheng Qiu[1] · Jie Jiang[2] · Xiaojun Chen[1]

Received: 11 September 2023 / Revised: 17 April 2024 / Accepted: 9 September 2024 /
Published online: 5 October 2024

Abstract
The first-order optimality condition of convexly constrained nonconvex nonconcave min-max optimization problems with box constraints formulates a nonmonotone variational inequality (VI), which is equivalent to a system of nonsmooth equations. In this paper, we propose a quasi-Newton subspace trust region (QNSTR) algorithm for the least squares problems defined by the smoothing approximation of nonsmooth equations. Based on the structure of the nonmonotone VI, we use an adaptive quasi-Newton formula to approximate the Hessian matrix and solve a low-dimensional strongly convex quadratic program with ellipse constraints in a subspace at each step of the QNSTR algorithm efficiently. We prove the global convergence of the QNSTR algorithm to an ϵ-first-order stationary point of the min-max optimization problem. Moreover, we present numerical results based on the QNSTR algorithm with different subspaces for a mixed generative adversarial networks in eye image segmentation using real data to show the efficiency and effectiveness of the QNSTR algorithm for solving large-scale min-max optimization problems.

Keywords Nonmonotone variational inequality · Min-max optimization · Quasi-Newton method · Least squares problem · Generative adversarial networks

Mathematics Subject Classification 90C47 · 90C15 · 90C33 · 65K15

1 Introduction

Min-max optimization problems have wild applications in games [29], distributional robustness optimization [27], robust machine learning [26], generative adversarial networks

✉ Xiaojun Chen
xiaojun.chen@polyu.edu.hk

Zicheng Qiu
zi-cheng.qiu@connect.polyu.hk

Jie Jiang
jiangjiecq@163.com

[1] Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China

[2] College of Mathematics and Statistics, Chongqing University, Chongqing, China

(GANs) [14], reinforcement learning [7], distributed optimization [32], etc. Mathematically, a convexly constrained min-max optimization problem can be written as

$$\min_{x\in X}\max_{y\in Y} f(x,y) := \mathbb{E}_P\left[\ell(x,y,\xi)\right], \tag{1}$$

where $X \subseteq \mathbb{R}^n$ and $Y \subseteq \mathbb{R}^m$ are nonempty, closed and convex sets, ξ is an s-dimensional random vector obeying a probability distribution P with support set Ξ, $\ell : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^s \to \mathbb{R}$ is nonconvex-nonconcave for a fixed realization of ξ, i.e., $\ell(x, y, \xi)$ is neither convex in x nor concave in y. Hence the objective function $f(x, y)$ is also nonconvex-nonconcave in general.

In this paper, we are mainly interested in problem (1) arising from GANs [14], which reads

$$\min_{x\in X}\max_{y\in Y}\Big(\mathbb{E}_{P_1}\big[\log(D(y,\xi_1))\big] + \mathbb{E}_{P_2}\big[\log(1-D(y,G(x,\xi_2)))\big]\Big), \tag{2}$$

where ξ_i is an $\mathbb{R}^{s_i}$-valued random vector with probability distribution P_i for $i = 1, 2$, $D : \mathbb{R}^m \times \mathbb{R}^{s_1} \to (0, 1)$ is a discriminator, $G : \mathbb{R}^n \times \mathbb{R}^{s_2} \to \mathbb{R}^{s_1}$ is a generator.

Generally, the generator G and the discriminator D are two feedforward neural networks. For example, G can be a p-layer neural network and D can be a q-layer neural network, that is

$$\begin{aligned} G(x,\xi_2) &= \sigma_G^p(W_G^p\sigma_G^{p-1}(\cdots\sigma_G^1(W_G^1\xi_2 + b_G^1) + \cdots) + b_G^p),\\ D(y,\xi_1) &= \sigma_D^q(W_D^q\sigma_D^{q-1}(\cdots\sigma_D^1(W_D^1\xi_1 + b_D^1) + \cdots) + b_D^q), \end{aligned}$$

where $W_G^1, \cdots, W_G^p, b_G^1, \cdots, b_G^p$ and $W_D^1, \cdots, W_D^q, b_D^1, \cdots, b_D^q$ are the weight matrices, biases vectors of G and D with suitable dimensions, $\sigma_G^1, \cdots, \sigma_G^p$ and $\sigma_D^1, \cdots, \sigma_D^q$ are proper activation functions, such as ReLU, GELU, Sigmoid, etc. Denote

$$\begin{aligned} x &:= (\mathrm{vec}(W_G^1)^\top, \cdots, \mathrm{vec}(W_G^p)^\top, (b_G^1)^\top, \cdots, (b_G^p)^\top)^\top,\\ y &:= (\mathrm{vec}(W_D^1)^\top, \cdots, \mathrm{vec}(W_D^q)^\top, (b_D^1)^\top, \cdots, (b_D^q)^\top)^\top, \end{aligned}$$

where $\mathrm{vec}(\cdot)$ denotes the vectorization operator. In this case, problem (2) reduces to problem (1) if let $\xi := (\xi_1, \xi_2) \in \Xi$ and

$$\ell(x, y, \xi) := \log(D(y, \xi_1)) + \log(1 - D(y, G(x, \xi_2))).$$

Due to the nonconvexity-nonconcavity of the objective function f, problem (1) may not have a saddle point. Hence the concepts of global and local saddle points are untimely to characterize the optimum of problem (1). Recently, motivated by practical applications, the so-called global and local minimax points are proposed to describe the global and local optima of nonconvex-nonconcave min-max optimization problems in [19] from the viewpoint of sequential games. Moreover, the optimality necessary conditions for a local minimax point are studied in [19] for unconstrained min-max optimization problems. In [8, 18], the optimality conditions for a local minimax point are studied for constrained min-max optimization problems.

Numerical methods for min-max optimization problems have been extensively studied. These algorithms can be divided into four classes based on the convexity or concavity of problems: the convex-concave cases (see, e.g., [28, 30, 31, 39]), the nonconvex-concave cases (see, e.g., [22, 23, 33]), the convex-nonconcave cases (see, e.g., [22, 23, 33]) and the nonconvex-nonconcave cases (see, e.g., [11, 42]).

To solve problem (1) numerically, we first apply the sample average approximation (SAA) approach to obtain a discrete form. We collect N independent identically distributed (i.i.d.) samples of ξ (e.g., generated by the Monte Carlo method), denoted by $\xi^1, \cdots, \xi^N$, and obtain a discrete counterpart of problem (1) as below:

$$\min_{x\in X}\max_{y\in Y}\ \hat{f}_N(x,y) := \frac{1}{N}\sum_{i=1}^{N}\ell(x,y,\xi^i). \tag{3}$$

To ease the discussion, we assume that $\ell(\cdot,\cdot,\xi)$ is twice continuously differentiable with respect to (x, y) for an arbitrary $\xi \in \Xi$ in what follows.

Let $z := (x^\top, y^\top)^\top \in \mathbb{R}^{n+m}$, $Z := X \times Y \subseteq \mathbb{R}^{n+m}$ and

$$H_N(z) := \begin{pmatrix} \nabla_x \hat{f}_N(x,y) \\ -\nabla_y \hat{f}_N(x,y) \end{pmatrix}.$$

Then the first-order optimality condition of a local minimax point for problem (3) can be presented as the following nonmonotone variational inequality (VI):

$$0 \in H_N(z) + \mathcal{N}_Z(z), \tag{4}$$

where $\mathcal{N}_Z(z)$ is the normal cone to the convex set Z at z, which is defined by

$$\mathcal{N}_Z(z) := \{v : \langle v, u - z\rangle \le 0, \forall u \in Z\}.$$

We call z^* a first-order stationary point of problem (3) if it satisfies (4). Due to the nonconvexity-nonconcavity, seldom algorithms can ensure the convergence to a global or local optimal point of problem (3). In the study of the convergence of algorithms for nonconvex-nonconcave min-max problem (3), some strong assumptions, such as the Polyak-Łojasiewicz (PL) condition [42], the existence of solutions for the corresponding Minty VI of problem (4) [11] etc, are required. In fact, such assumptions are difficult to check in practice. On the other hand, some algorithms need to estimate the Lipschitz constant of $H_N(z)$, but the computation of the Lipschitz constant of $H_N(z)$ may be too costly or even intractable. In this paper, without estimating the Lipschitz constant of $H_N(z)$ and assuming the PL condition or the existence of solutions for Minty VI, we use a so-called quasi-Newton subspace trust region (QNSTR, for short) algorithm for solving problem (4).

The VI (4) can be equivalently reformulated as the following system of nonsmooth equations (see [12])

$$F_N(z) := z - \mathrm{Proj}_Z(z - H_N(z)) = 0, \tag{5}$$

where $\mathrm{Proj}_Z(\cdot)$ denotes the projection operator onto Z.

Obviously, z^* is a first-order stationary point of (3) (i.e., a solution of (4) or (5)) if it is an optimal solution of the following least squares problem:

$$\min_{z\in\mathbb{R}^{n+m}} r_N(z) := \frac{1}{2}\|F_N(z)\|^2 \tag{6}$$

and $r_N(z^*) = 0$, where $\|\cdot\|$ denotes the Euclidean norm.

The main contributions of this paper are summarized as follows. (i) We develop the QNSTR algorithm for solving the least squares problem (6) when X and Y are boxes. Based on the structure of the VI (4), we use a smoothing function to approximate the nonsmooth function F_N, adopt an adaptive quasi-Newton formula to approximate the Hessian matrix and solve a quadratic program with ellipse constraints in a subspace at each step of the QNSTR

algorithm with a relatively low computational cost. (ii) We prove the global convergence of the QNSTR algorithm to a stationary point of a smoothing least squares problem of (6), which is an ϵ-first-order stationary point of the min-max optimization problem (3) if every element of the generalized Jacobian matrix of F_N is nonsingular at it. (iii) We apply the QNSTR algorithm to GANs in eye image segmentation with real data, which validates the effectiveness and efficiency of our approach for large-scale min-max optimization problems.

This paper is organized as follows. In Sect. 2, we introduce concepts of local minimax points and first-order optimality conditions. Moreover, we investigate the asymptotic convergence between problems (1) and (3) to build the numerical foundation for the subsequent discussion. In Sect. 3, we present the QNSTR algorithm and establish its global convergence. In Sect. 4, we apply the QNSTR algorithm to solve problem (3) with examples from eye image segmentation problems and digital handwriting image generation problems based on two different real data sets. We compare the QNSTR algorithm with some existing methods including alternating Adam. Finally, we give some concluding remarks in Sect. 5.

Notations $\mathbb{N}$ denotes the set of natural numbers. $\|\cdot\|$ denotes the Euclidean norm of a vector or the norm of a matrix induced by the Euclidean norm. $\mathrm{d}(x, Y) := \inf_{y\in Y} \|x - y\|$ and $\mathrm{d}(X, Y) := \sup_{x\in X} \inf_{y\in Y} \|x - y\|$, $X, Y \subseteq \mathbb{R}^n$.

2 First-Order Stationarity and Asymptotic Convergence

In this section, we focus on the asymptotic convergence between problems (3) and (1) regarding to the global minimax point and the first-order stationary point. To this end, we first give some preliminaries on how to describe the optima of a min-max optimization problem.

Definition 1 (global and local minimax points, [19, Definitions 9 & 14])

(i) $(\hat{x}, \hat{y}) \in X \times Y$ is called a *global minimax point* of problem (1), if

$$f(\hat{x}, y) \le f(\hat{x}, \hat{y}) \le \max_{y'\in Y} f(x, y'),\ \forall (x, y) \in X \times Y.$$

(ii) $(\hat{x}, \hat{y}) \in X \times Y$ is called a *local minimax point* of problem (1), if there exist a $\delta_0 > 0$ and a function $\tau : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\tau(\delta) \to 0$ as $\delta \to 0$, such that for any $\delta \in (0, \delta_0]$ and any $(x, y) \in X \times Y$ satisfying $\|x - \hat{x}\| \le \delta$ and $\|y - \hat{y}\| \le \delta$, we have

$$f(\hat{x}, y) \le f(\hat{x}, \hat{y}) \le \max_{y'\in\{y\in Y:\|y-\hat{y}\|\le\tau(\delta)\}} f(x, y').$$

Remark 1 The concept of saddle points has been commonly used to characterize the optima of min-max problems. A point $(\hat{x}, \hat{y}) \in X \times Y$ is called a *saddle point* of problem (1), if

$$f(\hat{x}, y) \le f(\hat{x}, \hat{y}) \le f(x, \hat{y}),\ \forall (x, y) \in X \times Y, \tag{7}$$

and $(\hat{x}, \hat{y}) \in X \times Y$ is called a *local saddle point* of problem (1) if (7) holds in a neighborhood of $(\hat{x}, \hat{y})$. However, as pointed out in [19], saddle points and local saddle points may not exist in many applications of machine learning, especially in the nonconvex-nonconcave case. Also, (local) saddle points are solutions from the viewpoint of simultaneous game, where the minimization operator and the maximization operator act simultaneously. However, many applications, such as GANs and adversarial training, seek for solutions in the sense of sequential game, where the minimization operator acts first and the maximization operator acts latter. The global and local minimax points exist under some mild conditions (see [19, Proposition 11 and Lemma 16]) and also describe the optima in the sense of sequential game.

The following lemma gives the first-order necessary optimality conditions of local minimax points for problem (1).

Lemma 1 ([18, Theorem 3.2 & Corollary 3.1]) *If $(\hat{x}, \hat{y}) \in X \times Y$ is a local minimax point of problem (1), then we have*

$$\begin{cases} 0 \in \nabla_x f(\hat{x}, \hat{y}) + \mathcal{N}_X(\hat{x}), \\ 0 \in -\nabla_y f(\hat{x}, \hat{y}) + \mathcal{N}_Y(\hat{y}). \end{cases} \tag{8}$$

Definition 2 $(\hat{x}, \hat{y}) \in X \times Y$ is called a *first-order stationary point* of problem (1) if (8) holds. $(\hat{x}, \hat{y}) \in X \times Y$ is called a *first-order stationary point* of problem (3) if (8) holds with replacing f by $\hat{f}_N$.

Hereafter, we will focus on finding a first-order stationary point of (3).

As for the exponential rate of convergence of the first-order and second-order stationary points of SAA for a specific GAN, one can refer to [18, Proposition 4.3]. In what follows, we mainly focus on the almost surely convergence analysis between problems (1) and (3) as N tends to infinity. If the problem is well-behaved and the global minimax points are achievable, we consider the convergence of global minimax points between problems (1) and (3). Otherwise, the first-order stationary points (Definition 2) are getatable. Thus, we need also consider the convergence of first-order stationary points between problems (1) and (3) as N tends to infinity.

Denote the optimal value, the set of global minimax points and the set of first-order stationary points of problem (1) by ϑ_g, $\mathcal{S}_g$ and $\mathcal{S}_{1\text{st}}$, respectively. Let $\widehat{\vartheta}_g^N$, $\widehat{\mathcal{S}}_g^N$ and $\widehat{\mathcal{S}}_{1\text{st}}^N$ denote the optimal value, the set of global minimax points and the set of first-order stationary points of problem (3), respectively.

Lemma 2 *Suppose that: (a) X and Y are compact sets; (b) $\ell(x, y, \xi)$ is dominated by an integrable function for every $(x, y) \in X \times Y$. Then*

$$\sup_{(x,y)\in X\times Y} \left| \hat{f}_N(x, y) - f(x, y) \right| \to 0$$

w.p.1 as $N \to \infty$.

If, further, (c) $\|\nabla_x \ell(x, y, \xi)\|$ and $\|\nabla_y \ell(x, y, \xi)\|$ are dominated by an integrable function for every $(x, y) \in X \times Y$, then

$$\sup_{(x,y)\in X\times Y} \left\| \nabla \hat{f}_N(x, y) - \nabla f(x, y) \right\| \to 0$$

w.p.1 as $N \to \infty$.

Proof Since the samples are i.i.d. and X and Y are compact, it is known from [36, Theorem 7.53] that the above uniform convergence results hold. □

The following proposition gives the nonemptiness of $\widehat{\mathcal{S}}_g^N$, $\mathcal{S}_g$, $\widehat{\mathcal{S}}_{1\text{st}}^N$ and $\mathcal{S}_{1\text{st}}$.

Proposition 1 *If conditions (a)–(c) in Lemma 2 hold, then $\mathcal{S}_g$ and $\mathcal{S}_{1\text{st}}$ are nonempty and $\widehat{\mathcal{S}}_g^N$ and $\widehat{\mathcal{S}}_{1\text{st}}^N$ are nonempty for any $N \in \mathbb{N}$.*

Proof Since the continuity of $f(x, y)$ and $\hat{f}_N(x, y)$ w.r.t. (x, y) and the boundedness of X and Y, we know from [19, Proposition 11] the nonemptiness of $\mathcal{S}_g$ and $\widehat{\mathcal{S}}_g^N$. Note that both $\mathcal{S}_{1\text{st}}$ and $\widehat{\mathcal{S}}_{1\text{st}}^N$ are solutions of variational inequalities. Then we have from [12, Corollary 2.2.5] that $\mathcal{S}_{1\text{st}}$ and $\widehat{\mathcal{S}}_{1\text{st}}^N$ are nonempty. □

Based on the uniform laws of large numbers in Lemma 2, we have the following convergence results.

Theorem 1 *Let conditions (a)–(c) in Lemma 2 hold. Then*

$$\mathrm{d}\left(\widehat{\mathcal{S}}_g^N, \mathcal{S}_g\right) \to 0, \tag{9}$$

$$\mathrm{d}\left(\widehat{\mathcal{S}}_{1\mathrm{st}}^N, \mathcal{S}_{1\mathrm{st}}\right) \to 0, \tag{10}$$

w.p.1 as $N \to \infty$.

Proof (10) follows from [35, Proposition 19] directly. Thus, in what follows, we only consider (9).

From Proposition 1, we know that $\widehat{\mathcal{S}}_g^N$ and $\mathcal{S}_g$ are nonempty for any $N \in \mathbb{N}$. Let $z^N = (x^N, y^N) \in \widehat{\mathcal{S}}_g^N$ and $z^N \to \bar{z} = (\bar{x}, \bar{y})$ w.p.1 as $N \to \infty$. Then we just verify that $\bar{z} \in \mathcal{S}_g$ w.p.1. If $\{z^N\}$ is not a convergent sequence, due to the boundedness of X and Y, we can choose a convergent subsequence. Denote $\varphi(x) := \max_{y\in Y} f(x, y)$ and $\hat{\varphi}_N(x) := \max_{y\in Y} \hat{f}_N(x, y)$. Note that

$$\begin{aligned} \max_{x\in X} \left|\hat{\varphi}_N(x) - \varphi(x)\right| &= \max_{x\in X} \left|\max_{y\in Y} \hat{f}_N(x, y) - \max_{y\in Y} f(x, y)\right| \\ &\le \max_{(x,y)\in X\times Y} \left|\hat{f}_N(x, y) - f(x, y)\right| \\ &\to 0 \end{aligned} \tag{11}$$

w.p.1 as $N \to \infty$, where the last convergence assertion follows from Lemma 2.

Next, we show

$$\mathrm{Proj}_x \widehat{\mathcal{S}}_g^N = \arg\min_{x\in X} \hat{\varphi}_N(x) \text{ and } \mathrm{Proj}_x \mathcal{S}_g = \arg\min_{x\in X} \varphi(x), \tag{12}$$

where Proj_x denotes the projection onto the x's space. Based on the definition of global minimax points, we have, for any $(\hat{x}, \hat{y}) \in \mathcal{S}_g$, that

$$f(\hat{x}, y) \le f(\hat{x}, \hat{y}) \le \max_{y'\in Y} f(x, y'), \ \forall (x, y) \in X \times Y,$$

which implies

$$\varphi(\hat{x}) = \max_{y\in Y} f(\hat{x}, y) \le \max_{y'\in Y} f(x, y') = \varphi(x), \ \forall x \in X.$$

This means $\mathrm{Proj}_x \mathcal{S}_g \subseteq \arg\min_{x\in X} \varphi(x)$. On the other hand, for any $\hat{x} \in \arg\min_{x\in X} \varphi(x)$, we let $\hat{y} \in \arg\max_{y\in Y} f(\hat{x}, y)$. Then it is not difficult to examine that $(\hat{x}, \hat{y})$ is a global minimax point, i.e., $\arg\min_{x\in X} \varphi(x) \subseteq \mathrm{Proj}_x \mathcal{S}_g$. The $\mathrm{Proj}_x \widehat{\mathcal{S}}_g^N = \arg\min_{x\in X} \hat{\varphi}_N(x)$ can be similarly verified. Hence (12) holds.

Then (11) and (12) indicate, according to [41, Lemma 4.1], that

$$\mathrm{d}\left(\mathrm{Proj}_x \widehat{\mathcal{S}}_g^N, \mathrm{Proj}_x \mathcal{S}_g\right) \to 0 \tag{13}$$

w.p.1 as $N \to \infty$. We know from (13) that $\bar{x} \in \mathrm{Proj}_x \mathcal{S}_g$.

Moreover, we know that

$$\left|\widehat{\vartheta}_g^N - \vartheta_g\right| = \left|\min_{x\in X} \hat{\varphi}_N(x) - \min_{x\in X} \varphi(x)\right|$$

$$\begin{aligned} &\leq \max_{x\in X} \left|\hat{\varphi}_N(x) - \varphi(x)\right| \\ &\to 0 \end{aligned}$$

w.p.1 as $N \to \infty$, where ϑ_g and $\widehat{\vartheta}_g^N$ are optimal values of problems (1) and (3), respectively. Due to Lemma 2 and the continuity of f, we know that

$$\begin{aligned} \left|\hat{f}_N(x^N, y^N) - f(\bar{x}, \bar{y})\right| &\leq \left|\hat{f}_N(x^N, y^N) - f(x^N, y^N)\right| + \left|f(x^N, y^N) - f(\bar{x}, \bar{y})\right| \\ &\to 0. \end{aligned}$$

Since $\widehat{\vartheta}_g^N = \hat{f}_N(x^N, y^N)$, we know that $\vartheta_g = f(\bar{x}, \bar{y})$, which, together with $\bar{x} \in \mathrm{Proj}_x \mathcal{S}_g$, implies that $(\bar{x}, \bar{y}) \in \mathcal{S}_g$. □

Based on Theorem 1, it is well-founded for us to employ problem (3) to approximately solve problem (1). In the sequel, we will focus on how to compute an ϵ-first-order stationary point of problem (3).

3 The QNSTR Algorithm and its Convergence Analysis

In this section, we propose the QNSTR algorithm to compute an ϵ-first-order stationary point of problem (3) with a fixed sample size N. In the remainder of this paper, let $X = [a, b]$ and $Y = [c, d]$, where $a, b \in \mathbb{R}^n$, $c, d \in \mathbb{R}^m$ with $a < b$ and $c < d$ in the componentwise sense. In this case, the projection in (5) has a closed form and the function F_N can be written as

$$F_N(z) = z - \mathrm{mid}(l, u, z - H_N(z)), \tag{14}$$

where $l, u \in \mathbb{R}^{n+m}$ with $l = (a^\top, c^\top)^\top$ and $u = (b^\top, d^\top)^\top$, "mid" is the middle operator in the componentwise sense, that is

$$\mathrm{mid}(l, u, z - H_N(z))_i = \begin{cases} l_i, & \text{if} \quad (z - H_N(z))_i < l_i, \\ u_i, & \text{if} \quad (z - H_N(z))_i > u_i, \\ (z - H_N(z))_i, & \text{otherwise.} \end{cases} \quad i = 1, \cdots, n+m,$$

Since X and Y are boxes, (14) can be divided into two parts separably and rewritten as

$$F_N(z) = \begin{pmatrix} F_N^{\mathbb{1}}(z) \\ F_N^{\mathbb{2}}(z) \end{pmatrix} = \begin{pmatrix} x - \mathrm{mid}(a, b, x - \nabla_x \hat{f}_N(x, y)) \\ y - \mathrm{mid}(c, d, y + \nabla_y \hat{f}_N(x, y)) \end{pmatrix}. \tag{15}$$

3.1 Smoothing Approximation

Let $q(z) = z - H_N(z)$. The function F_N is not differentiable at z when $q_i(z) = l_i$ or $q_i(z) = u_i$ for some $1 \leq i \leq n+m$. To overcome the difficulty in computation of the generalized Hessian of the nonsmooth function $F_N(z)$, we consider its smoothing approximation

$$(F_{N,\mu})_i(z) = \begin{cases} \frac{1}{2}((H_N)_i(z) + z_i) + \frac{1}{2\mu}(u_i - q_i(z))^2 + \frac{\mu}{8} - \frac{u_i}{2}, & \text{if } |u_i - q_i(z)| \le \frac{\mu}{2}, \\ \frac{1}{2}((H_N)_i(z) + z_i) - \frac{1}{2\mu}(l_i - q_i(z))^2 - \frac{\mu}{8} - \frac{l_i}{2}, & \text{if } |l_i - q_i(z)| \le \frac{\mu}{2}, \\ (F_N)_i(z), & \text{otherwise,} \end{cases} \tag{16}$$

where $0 < \mu \le \hat{\mu} := \min_{1 \le i \le n+m}(u_i - l_i)$.

From (15), the smoothing function $F_{N,\mu}(z)$ can also be represented as

$$F_{N,\mu}(z) = \begin{pmatrix} F^{\mathbb{1}}_{N,\mu}(z) \\ F^{\mathbb{2}}_{N,\mu}(z) \end{pmatrix}, \tag{17}$$

where $F^{\mathbb{1}}_{N,\mu}(z)$ and $F^{\mathbb{2}}_{N,\mu}(z)$ are the smoothing approximations of $F^{\mathbb{1}}_N(z)$ and $F^{\mathbb{2}}_N(z)$, respectively.

We summarize some useful properties of the smoothing function $F_{N,\mu}$, which can be found in [4] and [3, Section 6].

Lemma 3 *Let $F_{N,\mu}$ be a smoothing function of F_N defined in (16). Then for any $\mu \in (0, \hat{\mu})$, $F_{N,\mu}$ is continuously differentiable and has the following properties.*

(i) There is a $\kappa > 0$ such that for any $z \in \mathbb{R}^{m+n}$ and $\mu > 0$,

$$\|F_{N,\mu}(z) - F_N(z)\| \le \kappa\mu.$$

(ii) For any $z \in \mathbb{R}^{m+n}$, we have

$$\lim_{\mu \downarrow 0} \mathrm{d}(\nabla_z F_{N,\mu}(z), \partial_C F_N(z)) = 0,$$

where $\partial_C F_N(z) = \partial(F_N(z))_1 \times \partial(F_N(z))_2 \times \cdots \times \partial(F_N(z))_{n+m}$, and $\partial(F_N(z))_i$ is the Clarke generalized gradient of $(F_N(\cdot))_i$ at z for $i = 1, \ldots, n+m$. Moreover, there exists a $\bar{\mu} > 0$ such that for any $\mu \in (0, \bar{\mu})$, we have $\nabla F_{N,\mu}(z) \in \partial_C F_N(z)$.

In Fig. 11 in Appendix A, we show the approximation error $\|F_{N,\mu}(z) - F_N(z)\|$ over $X \times Y$ as $\mu \downarrow 0$ with different N.

Definition 3 (ϵ-first-order stationary points) For given $\epsilon > 0$, a point z is called an *ϵ-first-order stationary point* of problem (3), if $\|F_N(z)\| \le \epsilon$.

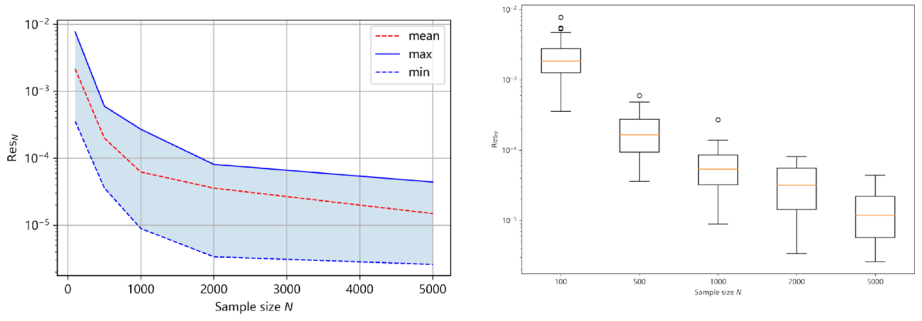


Fig. 1 The convergence of res_N as N grows (left: The range of res_N with different N, right: The boxplot of res_N with different N)

From (i) of Lemma 3, if z^* is an ϵ-first-order stationary point of problem (3), i.e., $\|F_N(z^*)\| \leq \epsilon$, then for any $\mu \in (0, \frac{\epsilon}{\kappa})$, we have

$$\|F_{N,\mu}(z^*)\| - \|F_N(z^*)\| \leq \|F_{N,\mu}(z^*) - F_N(z^*)\| \leq \kappa\mu \leq \epsilon, \tag{18}$$

which implies $\|F_{N,\mu}(z^*)\| \leq \|F_N(z^*)\| + \epsilon \leq 2\epsilon$. On the other hand, if z^* satisfies $\|F_{N,\mu}(z^*)\| \leq \frac{\epsilon}{2}$ for some $\epsilon > 0$ and $\mu \in (0, \frac{\epsilon}{2\kappa})$, then we have

$$\|F_N(z^*)\| - \|F_{N,\mu}(z^*)\| \leq \|F_{N,\mu}(z^*) - F_N(z^*)\| \leq \kappa\mu \leq \frac{\epsilon}{2},$$

which implies $\|F_N(z^*)\| \leq \|F_{N,\mu}(z^*)\| + \frac{\epsilon}{2} \leq \epsilon$, that is, z^* is an ϵ-first-order stationary point of problem (3).

Now we consider the smoothing least squares problem with a fixed small smoothing parameter $\mu > 0$:

$$\min_{z\in\mathbb{R}^{n+m}} r_{N,\mu}(z) := \frac{1}{2}\|F_{N,\mu}(z)\|^2. \tag{19}$$

Let $J_{N,\mu}(z)$ be the Jacobian matrix of $F_{N,\mu}(z)$. The gradient of the function $r_{N,\mu}$ is

$$\nabla r_{N,\mu}(z) = J_{N,\mu}(z)^\top F_{N,\mu}(z).$$

A vector z^* is called a first-order stationary point of problem (19) if $\nabla r_{N,\mu}(z^*) = 0$. If $J_{N,\mu}(z^*)$ is nonsingular, then $F_{N,\mu}(z^*) = 0$. From (i) of Lemma 3, $\|F_N(z^*)\| = \|F_N(z^*) - F_{N,\mu}(z^*)\| \leq \kappa\mu \leq \epsilon$ when $\mu \in (0, \epsilon/\kappa)$. This means that a first-order stationary point z^* of problem (19) is an ϵ-first-order stationary point of problem (3) if $\mu \in (0, \epsilon/\kappa)$ and $J_{N,\mu}(z^*)$ is nonsingular. Note that $\partial_C F_N(z)$ is a compact set for any $z \in X \times Y$. From (ii) of Lemma 3, if all matrices in $\partial_C F_N(z^*)$ are nonsingular, then there is $\mu_0 > 0$ such that for any $\mu \in (0, \mu_0)$, $J_{N,\mu}(z^*)$ is nonsingular.

If $J_{N,\mu}(z^*)$ is singular, the assumptions of local convergence theorems in [10, 15] for Gauss-Newton methods to solve the least squares problem (19) fail. In the next subsection, we prove the convergence of the QNSTR algorithm for the least squares problem (19) to a stationary point of (19) without assuming the nonsigularity of $J_{N,\mu}(z^*)$.

3.2 The QNSTR Algorithm

In this subsection, we present the QNSTR algorithm with a fixed sample size N and a fixed smoothing parameter μ. For simplicity, in this subsection, we use $F(z)$, $J(z)$ and $r(z)$ to denote $F_{N,\mu}(z)$, $J_{N,\mu}(z)$ and $r_{N,\mu}(z)$, respectively. Moreover, we use $F^{\mathbb{1}}(z)$, $F^{\mathbb{2}}(z)$ to represent $F^{\mathbb{1}}_{N,\mu}(z)$ and $F^{\mathbb{2}}_{N,\mu}(z)$, respectively.

For an arbitrary point $z_0 \in X \times Y$ and a positive number R_0, we denote the level set $\bar{S} := \{z \in \mathbb{R}^{n+m} \mid r(z) \leq r(z_0)\}$ and define a set $S(R_0) := \{z \in \mathbb{R}^{n+m} \mid \|z - z'\| \leq R_0, \forall z' \in \bar{S}\}$. By the definition of F and the boundedness of X and Y, we have $r(z) = \frac{1}{2}\|F(z)\|^2 \to \infty$ if $\|z\| \to \infty$. Hence both $\bar{S}$ and $S(R_0)$ are bounded.

Let $J_1(z) = \nabla F^{\mathbb{1}}(z)$ and $J_2(z) = \nabla F^{\mathbb{2}}(z)$. Then from

$$r(z) := \frac{1}{2}\|F^{\mathbb{1}}(z)\|^2 + \frac{1}{2}\|F^{\mathbb{2}}(z)\|^2,$$

if $F^{\mathbb{1}}$ and $F^{\mathbb{2}}$ are twice differentiable at z, the Hessian matrix

$$\nabla^2 r(z) = \frac{1}{2}\nabla^2\|F^{\mathbb{1}}(z)\|^2 + \frac{1}{2}\nabla^2\|F^{\mathbb{2}}(z)\|^2,$$

can be written as

$$\nabla^2\|F^{\mathbb{1}}(z)\|^2 = J_1(z)^\top J_1(z) + \sum_{i=1}^{n}(F^{\mathbb{1}})_i(z)\nabla^2(F^{\mathbb{1}})_i(z), \tag{20}$$

$$\nabla^2\|F^{\mathbb{2}}(z)\|^2 = J_2(z)^\top J_2(z) + \sum_{i=1}^{m}(F^{\mathbb{2}})_i(z)\nabla^2(F^{\mathbb{2}})_i(z). \tag{21}$$

If $F^{\mathbb{1}}$ and $F^{\mathbb{2}}$ are not twice differentiable at z, the generalized Hessian of r at z, denote $\partial(\nabla r(z))$, is the convex hull of all $(m+n)\times(m+n)$ matrices obtained as the limit of a sequence of the form $\nabla^2 r(z^k)$, where $z^k \to z$ and $F^{\mathbb{1}}$ and $F^{\mathbb{2}}$ are twice differentiable at z^k [5]. Hence from (20)–(21) and the twice continuous differentiability of $\hat{f}_N$, we know that there is a positive number M_1 such that $\|H\| \le M_1$ for any $H \in \partial(\nabla r(z))$, $z \in S(R_0)$. Moreover from [5, Proposition 2.6.5], there is a positive number M_2 such that

$$\left\|\nabla r(z) - \nabla r(z')\right\| \le M_2\left\|z - z'\right\|, \ \forall z, z' \in S(R_0). \tag{22}$$

To give a globally convergent algorithm for problem (19) without using the second derivatives, we keep the term $J_1(z)^\top J_1(z)$ and $J_2(z)^\top J_2(z)$ in (20)–(21), and approximate

$$\sum_{i=1}^{n}(F^{\mathbb{1}})_i(z)\nabla^2(F^{\mathbb{1}})_i(z) \ \text{ and } \ \sum_{i=1}^{m}(F^{\mathbb{2}})_i(z)\nabla^2(F^{\mathbb{2}})_i(z).$$

Specifically, the Hessian matrix at the k-th iteration point z_k is approximated by H_k with

$$H_k = J_1(z_k)^\top J_1(z_k) + J_2(z_k)^\top J_2(z_k) + A_k, \tag{23}$$

where

$$A_k = \begin{pmatrix} B_k & O \\ O & C_k \end{pmatrix}.$$

Here the matrices B_k and C_k are computed by the truncated BFGS quasi-Newton formula as follows.

$$B_{k+1} = \begin{cases} \bar{B}_{k+1} & \text{if } \|\bar{B}_{k+1}\| \le \gamma \ \&\ \frac{(s_k^{\mathbb{1}})^\top v_k^{\mathbb{1}}}{(s_k^{\mathbb{1}})^\top s_k^{\mathbb{1}}} \ge \bar{\epsilon} \\ \|F^{\mathbb{1}}(z_{k+1})\| I_n & \text{otherwise,} \end{cases} \tag{24}$$

$$C_{k+1} = \begin{cases} \bar{C}_{k+1} & \text{if } \|\bar{B}_{k+1}\| \le \gamma \ \&\ \frac{(s_k^{\mathbb{2}})^\top v_k^{\mathbb{2}}}{(s_k^{\mathbb{2}})^\top s_k^{\mathbb{2}}} \ge \bar{\epsilon} \\ \|F^{\mathbb{2}}(z_{k+1})\| I_m & \text{otherwise,} \end{cases} \tag{25}$$

where

$$\bar{B}_{k+1} = B_k - \frac{B_k s_k^{\mathbb{1}}(s_k^{\mathbb{1}})^\top B_k^\top}{(s_k^{\mathbb{1}})^\top B_k s_k^{\mathbb{1}}} + \frac{v_k^{\mathbb{1}}(v_k^{\mathbb{1}})^\top}{(v_k^{\mathbb{1}})^\top s_k^{\mathbb{1}}} \tag{26}$$

and

$$\bar{C}_{k+1} = C_k - \frac{C_k s_k^{\mathbb{2}}(s_k^{\mathbb{2}})^\top C_k^\top}{(s_k^{\mathbb{2}})^\top C_k s_k^{\mathbb{2}}} + \frac{v_k^{\mathbb{2}}(v_k^{\mathbb{2}})^\top}{(v_k^{\mathbb{2}})^\top s_k^{\mathbb{2}}}. \tag{27}$$

Here $\bar{\epsilon}$ and γ are given positive parameters, $s_k^{\mathbb{1}} = x_{k+1} - x_k$, $s_k^{\mathbb{2}} = y_{k+1} - y_k$,

$$v_k^{\mathbb{1}} = (\nabla_x F^{\mathbb{1}}(z_{k+1}) - \nabla_x F^{\mathbb{1}}(z_k))^\top F^{\mathbb{1}}(z_{k+1}) \|F^{\mathbb{1}}(z_{k+1})\| / \|F^{\mathbb{1}}(z_k)\|,$$

$$v_k^{\mathbb{2}} = (\nabla_y F^{\mathbb{2}}(z_{k+1}) - \nabla_y F^{\mathbb{2}}(z_k))^\top F^{\mathbb{2}}(z_{k+1}) \|F^{\mathbb{2}}(z_{k+1})\| / \|F^{\mathbb{2}}(z_k)\|.$$

Notice that the approximation form (23) is proposed in [47]. However, the matrix A_k in [47] is defined by using $F(z)$ and $\nabla F(z)$ at z_{k+1}, z_k. In this paper, we use $F^{\mathbb{1}}(z)$, $F^{\mathbb{2}}(z)$, $\nabla F^{\mathbb{1}}(z)$ and $\nabla F^{\mathbb{2}}(z)$ at z_{k+1}, z_k to define a two-block diagonal matrix A_k in (23), based on the structure of VI in (17).

In [47], a back tracking line search is used to obtain a stationary point of the least squares problem. In this paper, we use a subspace trust-region method to solve problem (19) with global convergence guarantees. Comparing with the quasi-Newton method with back tracking line search in [47], the QNSTR algorithm solves a strongly convex quadratic subproblem in a low dimension at each step, which is efficient to solve large-scale min-max optimization problems with real data. See Sect. 4 for more details.

In what follows, for simplification, we use J_k and F_k to denote $J(z_k)$, $F(z_k)$, respectively. Let $g_k = \nabla r(z_k)$. Choose $\{d_k^1, \cdots, d_k^{L-1}\}$ such that $V_k := \left[-g_k \; d_k^1 \cdots d_k^{L-1}\right] \in \mathbb{R}^{(n+m)\times L}$ has L linearly independent column vectors. Let

$$c_k := V_k^\top g_k, \quad G_k := V_k^\top V_k, \quad Q_k := V_k^\top H_k V_k.$$

Then, to obtain the stepsize α_k at the iteration point z_k, we solve the following strongly convex quadratic program in an L-dimensional space:

$$\begin{aligned} \alpha_k = \underset{\alpha\in\mathbb{R}^L}{\arg\min}\; & m_k(\alpha) := r(z_k) + c_k^\top \alpha + \tfrac{1}{2}\alpha^\top Q_k \alpha \\ \text{s.t.} \quad & \|V_k \alpha\| \le \Delta_k, \end{aligned} \tag{28}$$

where $\Delta_k > 0$ is the trust-region radius.

A key question for solving problem (28) is how to compute Q_k efficiently when H_k is huge. In fact, Q_k can be calculated efficiently without computing and storing the full information H_k. From (23), we can write Q_k as

$$Q_k = V_k^\top J_k^\top J_k V_k + V_k^\top A_k V_k. \tag{29}$$

For the term $V_k^\top J_k^\top J_k V_k$ in (29), we compute $J_k V_k$ in a columnwise way: For a sufficiently small $\epsilon > 0$,

$$J_k g_k \approx \frac{F(z_k + \epsilon g_k) - F(z_k)}{\epsilon}, \quad J_k d_k^i \approx \frac{F(z_k + \epsilon d_k^i) - F(z_k)}{\epsilon}, \; i = 1, \cdots, L-1.$$

On the other hand, $A_k V_k$ in term $V_k^\top A_k V_k$ can also be computed columnwisely by a series of vector-vector products.

We give the QNSTR algorithm in Algorithm 1.

Algorithm 1 The QNSTR Algorithm

Input: $\bar{\Delta} > 0$, $\Delta_0 \in (0, \bar{\Delta})$, $\beta_1 < 1 < \beta_2$, $0 \leq \eta < \zeta_1 < \zeta_2 \leq 1$, tolerance parameter $\delta > 0$, $\epsilon > 0$, $z_0 \in \mathbb{R}^{n+m}$.

1: If $\|g_k\| \leq \delta$ or $\|F(z_k)\| \leq \epsilon$, terminate. Otherwise solve (28) for α_k.
2: Compute the reduction ratio at iterate k:

$$\rho_k = \frac{r(z_k) - r(z_k + V_k\alpha_k)}{m_k(0) - m_k(\alpha_k)} \tag{30}$$

3: **if** $\rho_k < \zeta_1$ **then**
4: $\quad \Delta_{k+1} = \beta_1 \Delta_k$
5: **else**
6: $\quad$ **if** $\rho_k \geq \zeta_2$ and $\|V_k\alpha_k\| = \Delta_k$, **then**
7: $\quad\quad \Delta_{k+1} = \min\{\beta_2\Delta_k, \bar{\Delta}\}$
8: $\quad$ **else**
9: $\quad\quad \Delta_{k+1} = \Delta_k$
10: $\quad$ **end if**
11: **end if**
12: **if** $\rho_k > \eta$ **then**
13: $\quad z_{k+1} = z_k + V_k\alpha_k$
14: **else**
15: $\quad z_{k+1} = z_k$
16: **end if**

Trust-region algorithms are a class of popular numerical methods for optimization problems [6, 43]. Our QNSTR algorithm uses the special structure of the VI in (15) and (17) to construct the subproblem (28). The global convergence of the QNSTR algorithm is given in the following theorem.

Theorem 2 *Suppose that X and Y are nonempty and bounded boxes and $\hat{f}_N$ is twice continuously differentiable. Let $\{z_k\}_{k=0}^{\infty}$ be generated by Algorithm 1. Then there exists an $M > 0$ such that $\left\|\nabla r(z) - \nabla r(z')\right\| \leq M \left\|z - z'\right\|$ for any $z, z' \in S(R_0)$ and $\|H_k\| \leq M$ for $k \in \mathbb{N}$. Moreover, we have $\lim_{k\to\infty} \|g_k\| = 0$.*

The proof is given in Appendix B.

To end this section, we give some remarks about Theorem 2.

Remark 2 From Corollary 2.2.5 in [12], the continuity of F and boundness of X and Y imply that the solution set $\mathcal{Z}^*$ of $F(z) = 0$ is nonempty and bounded. Hence the set of minimizers of the least squares problem $\min r(z)$ with the optimal value zero is coincident with the solution set of $F(z) = 0$. If all matrices in $\partial_C F_N(z)$ for $z \in \mathcal{Z}^*$ are nonsingular, we know from the compactness of $\partial_C F_N(z)$ and (ii) of Lemma 3 that there is a $\mu_0 > 0$ such that for any $\mu \in (0, \mu_0)$, $J(z)$ is nonsingular and $\sup_{\mu\in(0,\mu_0)} \left\|J(z)^{-1}\right\| \leq C$ for some $C > 0$. Thus,

$$\|F(z)\| = \left\|(J(z)^\top)^{-1}\nabla r(z)\right\| \leq \left\|(J(z)^\top)^{-1}\right\| \|\nabla r(z)\| \leq C \|\nabla r(z)\|.$$

We have from (i) of Lemma 3, i.e., $\|F_N(z) - F(z)\| \leq \kappa\mu$ that

$$\|F_N(z)\| \leq C \|\nabla r(z)\| + \kappa\mu.$$

Thus, for any $\epsilon > 0$, we can properly select parameters δ and μ such that $\|\nabla r(z)\| \leq \delta$ and $C\delta + \kappa\mu \leq \epsilon$.

According to Theorem 2, we can find a point z_k such that

$$\|g_k\| = \|\nabla r(z_k)\| \leq \delta,$$

with an given stopping criterion parameter δ. The numerical experiments in the next section show that the QNSTR algorithm can result in an ϵ-first-order stationary point of problem (3).

4 Numerical Experiments

In this section, we report some numerical results via the QNSTR algorithm for finding an ϵ-first-order stationary point of problem (3). Also, we compare the QNSTR algorithm with some state-of-the-art algorithms for minimax problems. All of the numerical experiments in this paper are implemented on TensorFlow 1.13.1, Python 3.6.9 and Cuda 10.0 on a server with 1 Tesla P100-PCIE GPU with 16 GB memory at 1.3285 GHz and an operating system of 64 bits in the University Research Facility in Big Data Analytics (UBDA) of The Hong Kong Polytechnic University. (UBDA website: https://www.polyu.edu.hk/ubda/.)

We test our algorithm with two practical problems. One is a GAN based image generation problem for MNIST hand-writing data. The other one is a mix model for image segmentation on Digital Retinal Images for Vessel Extraction (DRIVE) data. In the first experiment, we use notation QNSTR(L) to denote that the dimension of the subspace spanned by the columns of V_k is L in the QNSTR algorithm and test the efficiency of QNSTR(L) under different choices of L and directions $\{d_k^i\}_{i=1}^{L-1}$. In the second experiment, we apply the QNSTR algorithm to a medical image segmentation problem.

To ensure that H_N is continuously differentiable, we choose Gaussian error Linear Units (GELU) [16]

$$\sigma(x) = x \int_{-\infty}^{x} \frac{e^{\frac{-t^2}{2\Sigma^2}}}{\sqrt{2\pi}\,\Sigma} dt$$

with $\Sigma = 10^{-4}$ as the activation function in each hidden layer in D and G, and Sigmoid

$$\sigma(x) = \frac{e^{-x}}{1 + e^{-x}}$$

as the activation function of the output layers in D and G.

We test different choices of the subspace spanned by the columns of V_k. Specifically, denote

$$\begin{aligned} V_k^z &= [-g_k, (z_k - z_{k-1}), \cdots, (z_{k-L+2} - z_{k-L+1})], \\ V_k^F &= [-g_k, F(z_k), \cdots, F(z_{k-L+2})], \\ V_k^g &= -[g_k, g_{k-1}, \cdots, g_{k-L+1}] \end{aligned}$$

for $L \geq 2$. Among all the experiments in the sequel, the initial point z_0 is the result for running 10,000 steps of the alternating Adam with step size 0.0005. The parameters in Algorithm 1 are set as $\bar{\Delta} = 100$, $\Delta_0 = 1$, $\beta_1 = 0.5$, $\beta_2 = 2$, $\eta = 0.01$, $\zeta_1 = 0.02$, $\zeta_2 = 0.05$. The parameters $\bar{\epsilon}$ and γ in (24) and (25) are chosen as $\bar{\epsilon} = 10^{-4}$ and $\gamma = 10^3$.

Table 1 Means, variances and 95% CIs of res_N with different N

N	500	1000	2000	5000
Mean	2.01×10^{-4}	6.22×10^{-5}	3.58×10^{-5}	1.49×10^{-5}
std	1.65×10^{-4}	4.95×10^{-5}	2.93×10^{-5}	1.18×10^{-5}
95% CI	$[1.65, 2.36]_{\times 10^{-4}}$	$[4.95, 7.49]_{\times 10^{-5}}$	$[2.93, 4.23]_{\times 10^{-5}}$	$[1.18, 1.79]_{\times 10^{-5}}$

4.1 Numerical Performance on MNIST Data

In this subsection, we report some preliminary numerical results of the QNSTR algorithm for solving the following SAA counterpart

$$\min_{x \in X} \max_{y \in Y} \frac{1}{N} \sum_{i=1}^{N} \Big(\log(D(y, \xi_1^i)) + \log(1 - D(y, G(x, \xi_2^i))) \Big) \tag{30}$$

of problem (2) by using MNIST handwritting data. We consider a two-layer GAN, where

$$\begin{aligned} &\xi_1^i \in \mathbb{R}^{784},\ \xi_2^i \in \mathbb{R}^{100}, \\ &W_G^1 \in \mathbb{R}^{N_G^1 \times 100},\ W_G^2 \in \mathbb{R}^{784 \times N_G^1},\ W_D^1 \in \mathbb{R}^{N_D^1 \times 784},\ W_D^2 \in \mathbb{R}^{1 \times N_D^1} \end{aligned}$$

with different choices of dimensions N_G^1 and N_D^1 for hidden outputs. Here $\{\xi_2^i\}$ are generated by an uniform distribution $\mathcal{U}(-1, 1)^{100}$. All initial weight matrices W_G^i and W_D^i for $i = 1, 2$ are randomly generated by using the Gaussian distribution with mean 0 and standard deviation 0.1, and all initial bias vectors b_G^i and b_D^i for $i = 1, 2$ are set to be zero. X and Y are set as $[-1, 1]^n$ and $[-1, 1]^m$, respectively.

4.1.1 Numerical Results for SAA Problems

We first study the performance of SAA problems under different sample sizes on a GAN model with a two-layer generator with $N_G^1 = 64$ and a two-layer discriminator with $N_D^1 = 64$, respectively. We set $\hat{N} = 10000$ as the benchmark to approximate the original problem (2) and let $N = 100, 500, 1000, 2000, 5000$. For each N, we solve $F_N(z) = 0$, 50 times with different samples by using the QNSTR algorithm. We stop the iteration either

$$\|F_N(z_k)\| \leq 10^{-5} \tag{31}$$

or the number of iterations exceeds 5000.

In these experiments, we set $\mu = 10^{-8}$. We use z_N^* to denote the first point that satisfies (31) in the iteration for $N = 500, 1000, 2000, 5000$, and we measure its optimality by the residual

$$\text{res}_N := \|z_N^* - \text{mid}\left(l, u, z_N^* - H_{\hat{N}}(z_N^*)\right)\|.$$

Figure 1 shows the convergence of res_N to zero as N grows. Table 1 presents the average of mean, standard deviation (std) and the width of 95% CI of res_N. It shows that all values decrease as the sample size N increases. Both Fig. 1 and Table 1 validate the convergence results in Sect. 2.

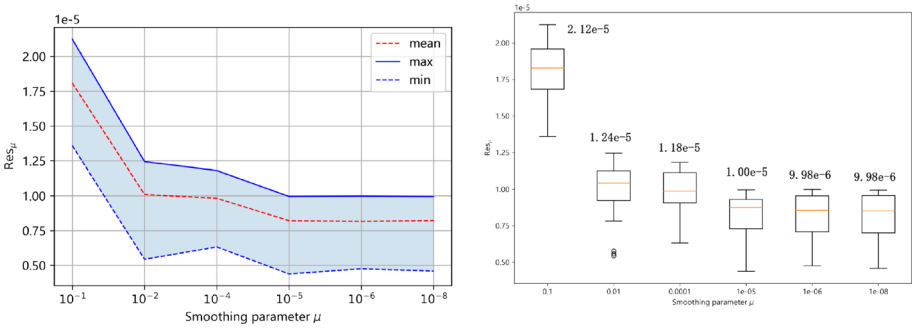


Fig. 2 The convergence of res_μ as μ decreases (left: The range of res_μ with different μ; right: The boxplot of res_μ with different μ)

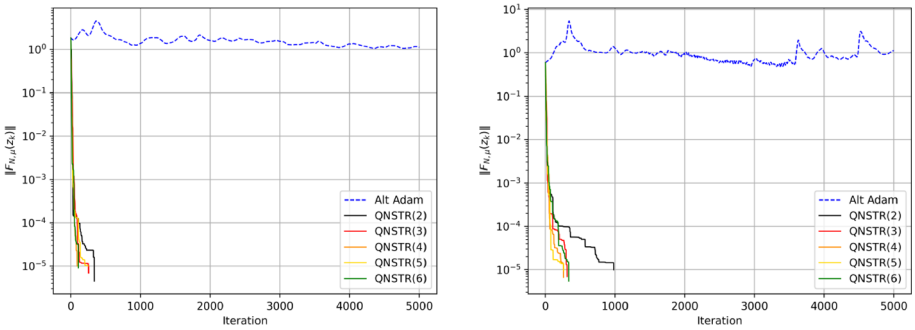


Fig. 3 The residual $\|F_{N,\mu}(z_k)\|$ with V_k^z (left: $N_G^1 = N_D^1 = 64$; right: $N_G^1 = N_D^1 = 128$)

4.1.2 Numerical Results for Smoothing Approximation

In this subsection, for a fixed sample size $N = 2000$, we study how the smoothing parameter μ affects the residual $\|F_{N,\mu}(z)\|$. All numerical results in this part are based on a GAN model which is constituted of a two-layer generator with $N_G^1 = 64$ and a two-layer discriminator with $N_D^1 = 64$. Specifically, for $\mu = 10^{-t}$, $t = 1, 2, 4, 5, 6, 8$, we generate 50 test problems, respectively. For each μ, we solve problem $F_{N,\mu}(z) = 0$ by the QNSTR algorithm. We stop the iteration either condition $\|F_{N,\mu}(z_k)\| \leq 10^{-5}$ holds or the number of iterations exceeds 5000.

We use z_μ^* to denote the first point that satisfies (31) in the iteration, and we measure the residual of z_μ^* by

$$\text{res}_\mu := \|z_\mu^* - \text{mid}\left(l, u, z_\mu^* - H_N(z_\mu^*)\right)\|.$$

The numerical results are presented in Fig. 2, which shows that res_μ decreases as smoothing parameter μ decreases. In fact, the residual res_μ becomes stable when $\mu \leq 10^{-5}$. Note that it is not difficult to obtain $\kappa = \frac{\sqrt{n+m}}{8}$ in (18). Also, when $N_D^1 = N_G^1 = 64$, we have $n + m = 107729$. If $\|F_{N,\mu}(z)\| \leq 10^{-5}$ and $\mu = 10^{-8}$, then we have $\|F_N(z)\| \leq \kappa\mu + \epsilon \leq \frac{\sqrt{107729}}{8} \times 10^{-8} + 10^{-5} \approx 1.041 \times 10^{-5}$. This is consistent with the theoretical results in Sect. 3.

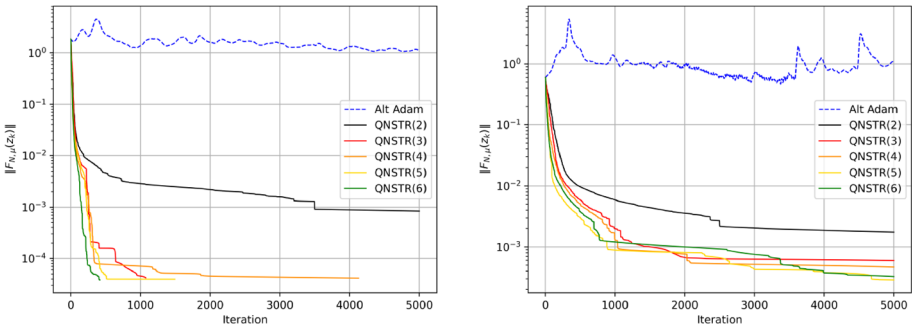


Fig. 4 The residual $\|F_{N,\mu}(z_k)\|$ with V_k^F (left: $N_G^1 = N_D^1 = 64$; right: $N_G^1 = N_D^1 = 128$)

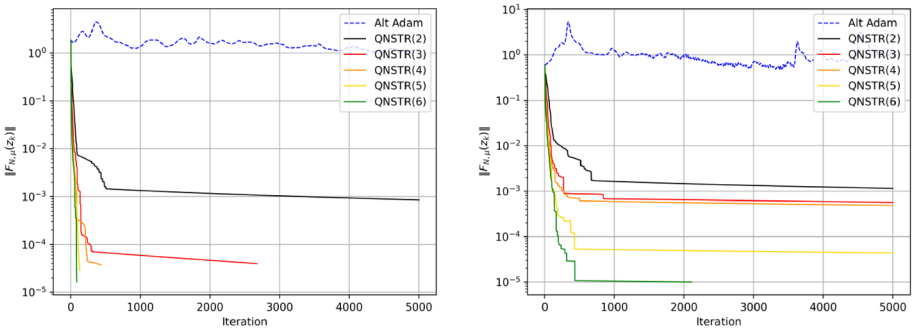


Fig. 5 The residual $\|F_{N,\mu}(z_k)\|$ with V_k^g (left: $N_G^1 = N_D^1 = 64$; right: $N_G^1 = N_D^1 = 128$)

4.1.3 Comparison Experiments

In this subsection, we report some numerical results to compare the QNSTR algorithm with some commonly-used methods. To this end, we set $N = 2000$ and $\mu = 10^{-8}$. We use $\|F_{N,\mu}(z_k)\|$ to measure performance of these algorithms and apply Frechet Inception Distance (FID) score to measure the quality of image generated by the generator G trained by different algorithms. We terminate all these algorithms when one of the three cases holds: $\|F_{N,\mu}(z_k)\| \leq 10^{-5}$, $\|\nabla F_{N,\mu}(z_k)^\top F_{N,\mu}(z_k)\| \leq 10^{-8}$, the number of iterations exceeds 5000.

We present the numerical results in Figs. 3, 4, 5. It is not difficult to observe from these figures that the QNSTR algorithm outperforms.

We also compare the QNSTR algorithm with some commonly-used algorithms in training GANs including simultaneous and alternate gradient descent-ascent (GDA) [9, 45], simultaneous and alternate optimistic gradient descent-ascent (OGDA) [9, 45], γ-alternate Adam, projected point algorithm (PPA) [11, 24]. The initial point z_0 of all these methods are given by alternating Adam with stepsize 0.0005 and 10000 iterations. We use the grid search for the selection of hyper-parameters in these methods. For simultaneous and alternate GDA and simultaneous and alternate OGDA, the stepsize is $\alpha = 0.5, 0.05, 0.005, 0.001, 0.0005$. For γ-alternate Adam, the ratio is $\gamma = 1, 2, 3, 5, 10$. For PPA, the proximal parameter is $L = 10, 100, 500, 1000, 5000$ and stepsize is $\frac{0.1}{2L}$. Besides, the stopping criteria of the k-th inner loop is $\frac{0.01}{k^2}$ or the number of iterations exceeds 100. The comparison results between the QNSTR algorithm and the γ-alternate Adam, the QNSTR algorithm and PPA are given

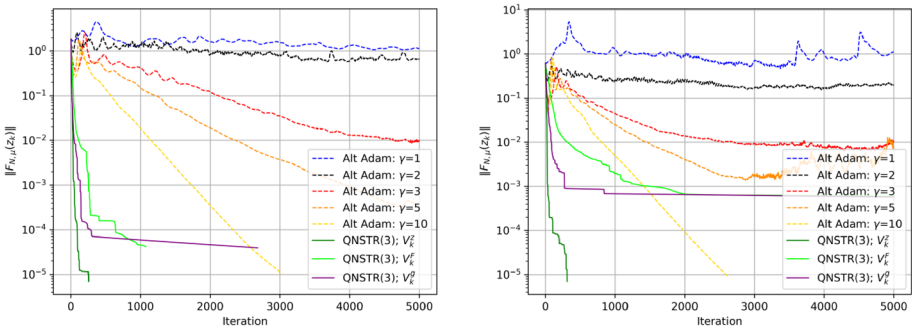


Fig. 6 Comparison results between the QNSTR algorithm for V_k^z, V_k^F, V_k^g and γ-alternate Adam (left: $N_G^1 = N_D^1 = 64$; right $N_G^1 = N_D^1 = 128$)

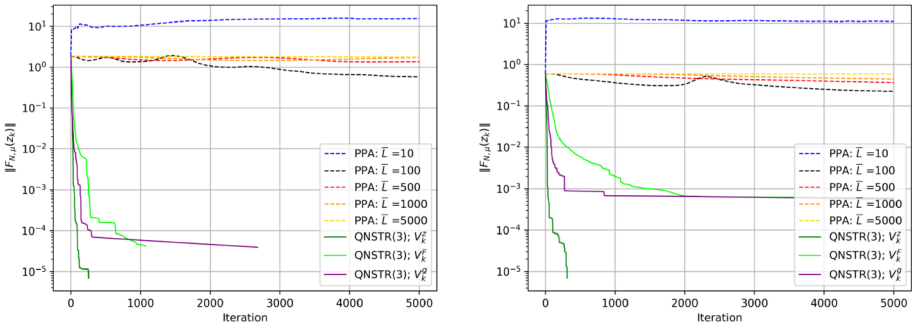


Fig. 7 Comparison results between the QNSTR algorithm for V_k^z, V_k^F, V_k^g and PPA (left: $N_G^1 = N_D^1 = 64$; right $N_G^1 = N_D^1 = 128$)

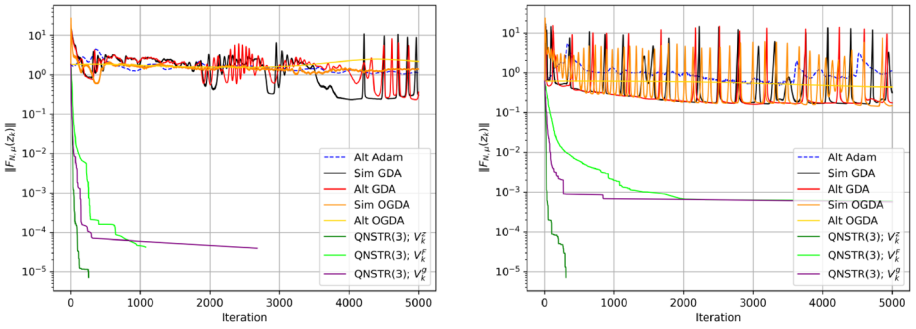


Fig. 8 Comparison results between the QNSTR algorithm for V_k^z, V_k^F, V_k^g and simultaneous GDA, alternate GDA, simultaneous OGDA, alternate OGDA (left: $N_G^1 = N_D^1 = 64$; right: $N_G^1 = N_D^1 = 128$)

in Figs. 6 and 7, respectively. According to the results, we can see the γ-alternate Adam also shows a outstanding convergence tendency under a suitable hyper-parameter and the QNSTR algorithm can even better than the γ-alternate Adam if a suitable searching space V_k is selected. The PPA performs poorly in these comparisons.

The comparison results between the QNSTR algorithm and simultaneous and alternate GDA, simultaneous and alternate OGDA are given in Fig. 8. To show these comparison results

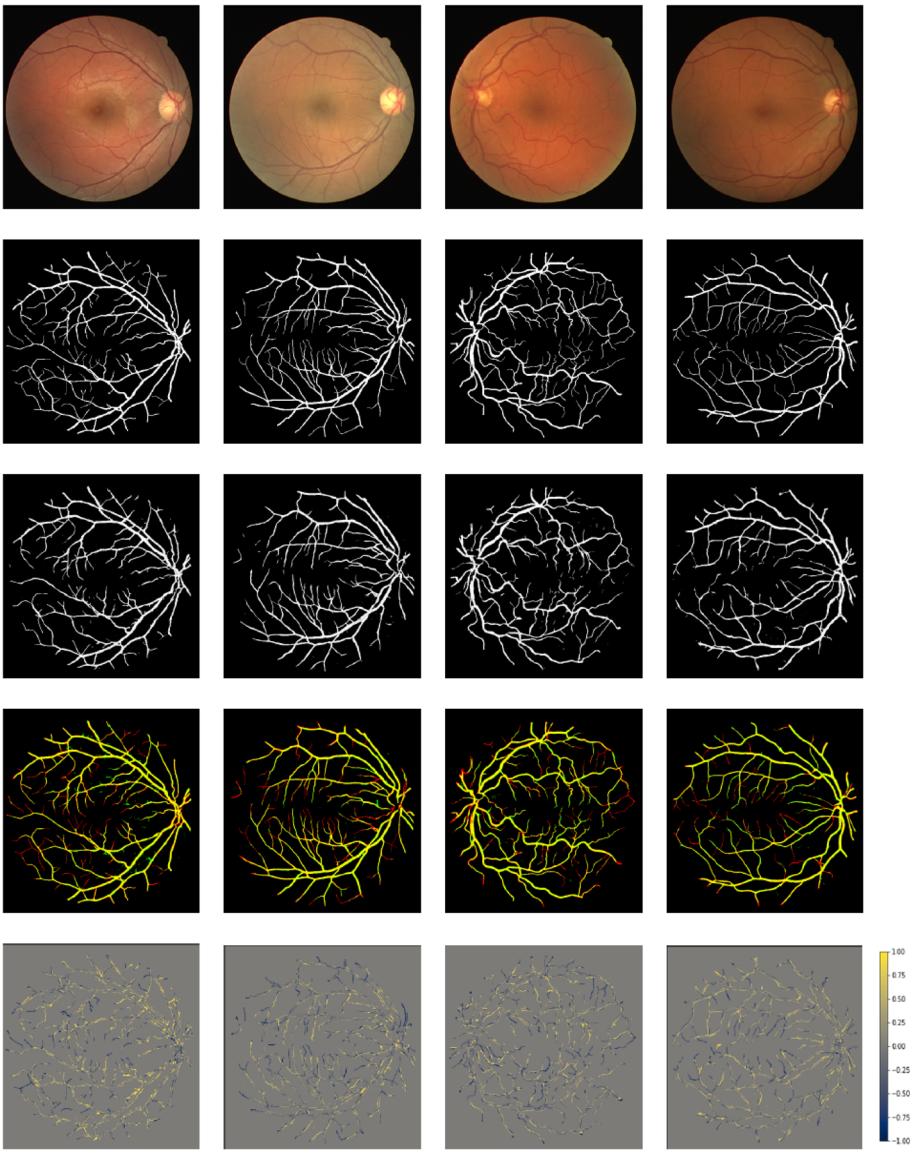


Fig. 9 Row 1. fundus image, Row 2. manual segmentation, Row 3. vessel map generated by GANs with QNSTR algorithm, Row 4. yellow(correct); red(wrong); green(missing), Row 5. Error (Color figure online)

more clearly, we only give the results of each method with the optimal stepsize α in the search range.

We can observe from these figures that the QNSTR algorithm outperforms, which validates that the QNSTR algorithm is more efficient in finding an ϵ-first-order stationary point of problem (30).

Table 2 FID scores of different algorithms with $N_D^1 = N_G^1 = 64$

QNSTR					
L	2	3	4	5	6
V_k^z	30.33	30.38	38.86	37.64	38.57
V_k^F	30.98	31.86	32.26	32.47	32.00
V_k^g	32.09	33.17	33.21	33.46	33.06
PPA					
$\overline{L}$	10	100	500	1000	5000
	32.77	31.93	36.15	37.26	37.73
γ-alt adam					
γ	1	2	3	5	10
	28.30	32.02	32.37	32.31	32.22
sGDA, aGDA, sOGDA, aOGDA					
α	0.5	0.05	0.01	0.005	0.0005
sGDA	34.11	30.87	31.56	32.97	34.57
aGDA	32.21	30.86	31.53	31.96	33.87
aOGDA	30.73	30.14	31.47	32.02	34.06
aOGDA	30.97	31.91	31.87	32.46	33.82

Table 3 FID scores of different algorithms with $N_D^1 = N_G^1 = 128$

QNSTR					
L	2	3	4	5	6
V_k^z	30.33	30.08	37.48	33.22	34.15
V_k^F	33.20	31.33	30.27	32.43	31.83
V_k^g	28.30	27.73	37.64	27.85	26.82
PPA					
$\overline{L}$	10	100	500	1000	5000
	29.85	27.32	27.06	38.76	30.48
γ-alt adam					
γ	1	2	3	5	10
	26.13	27.73	29.64	30.98	30.70
sGDA, aGDA, sOGDA, aOGDA					
α	0.5	0.05	0.01	0.005	0.0005
sGDA	31.43	30.18	30.85	32.64	34.57
aGDA	31.21	29.98	30.46	32.33	34.87
sOGDA	30.91	28.91	30.24	31.85	34.06
aOGDA	32.97	30.17	29.87	30.46	33.52

We also record the final FID score of each algorithm's output. All results are given in Tables 2 and 3. They show the generator of GANs trained by the QNSTR algorithm can generate high quality images.

4.2 DRIVE Data

Image segmentation is an important component in many visual understanding systems, which is the process of partitioning a digital image into multiple image segments [38]. Image segmentation plays a central role in a broad range of applications [13], including medical image analysis, autonomous vehicles (e.g., navigable surface and pedestrian detection), video surveillance and augmented reality. One of the well-known paradigm for image segmentation is based on some kinds of manual designed loss functions. However, they usually lead to the blurry segmentation boundary [17].

In 2016, Phillip et. al introduced a generative adversarial network framework into their objective function to implement an image-to-image translation problem [17], and found the blurry output given CNN under l_1 norm can be reduced. At the same year, Pauline et. al introduced a mix objective function combining by GAN and cross-entropy loss on semantic image segmentation problem [25], also implemented a better performance. The similar idea of mix GAN and traditional loss can also be found in [37, 46].

The mix model has the following form

$$\min_{x\in X}\max_{y\in Y}\left\{\hat{f}_N(x,y) := \frac{1}{N}\sum_{i=1}^{N}\lambda\cdot\psi\left(\xi_1^i, G(x,\xi_2^i)\right) + \left(\frac{1}{N}\sum_{i=1}^{N}\left(\log(D(y,\xi_1^i)) + \log(1-D(y,G(x,\xi_2^i)))\right)\right)\right\}, \tag{32}$$

where X, Y are two bounded boxes, $\{(\xi_1^i, \xi_2^i)\}_{i=1}^N$ is the finite collected data, ξ_2^i is the original data while the ξ_1^i is the corresponding label. Problem (32) is a special case of problem (3), which can be viewed as a discrete generative adversarial problem (30) with an extra classical supervision term. The classical supervision part, i.e.,

$$\min_{x\in X}\frac{1}{N}\sum_{i=1}^{N}\left[\psi\left(\xi_1^i, G(x,\xi_2^i)\right)\right] \tag{33}$$

is to minimize the difference between the output of given ξ_2 on G and its corresponding label ξ_1. The model can be regarded as a combination of a classical supervised learning problem and a generative adversarial problem with a trade-off parameter $\lambda \in [0, \infty)$. When $\lambda = 0$, problem (32) reduces to a classical supervised learning problem (33). When $\lambda \to \infty$, problem (32) tends to a vanilla GAN.

The fundoscopic exam is an important procedure to provide information to diagnose different retinal degenerative diseases such as Diabetic Retinopathy, Macular Edema and Cytomegalovirus Retinitis. A high accurate system to sketch out the blood vessel and find abnormalities on fundoscopic images is necessary. Although the supervision deep learning frameworks such as Unet are able to segment macro vessel accurately, they failed for segmenting microvessels with high certainty. In this part, we will train a Unet as a generator in our framework on DRIVE data. We download the data includes 20 eye blood vessel images with manual segmentation label from the open source website (https://drive.grand-challenge.org/). We applied 16 images as training data while the other 4 as testing data. In this experiment, the structure of segmentation model G is U-net [34] which includes 18 layers with $n = 121435$ parameters, and the structure of discrimination model D is a deep convolutional neural network which contains 5 convolutional layers and 1 fully connected layer with

Table 4 The performance of QNSTR algorithm and alternating Adam for model (32), and other methods in [1, 20, 21] for model (33)

Molds	F1 score	Sensitivity	Specificity	Accuracy	AUC-ROC	SSIM
Residual Unet [1]	0.8149	0.7726	0.9820	0.9553	0.9779	–
RecurrentUnet [1]	0.8155	0.7751	0.9816	0.9556	0.9782	–
R2Unet [1]	0.8171	0.7792	0.9813	0.9556	0.9784	–
DFUNet [20]	0.8190	0.7863	0.9805	0.9558	0.9779	0.8789
IterNet [21]	0.8205	0.7735	0.9838	0.9573	0.9816	0.9008
Alt Adam	0.7856	0.7830	0.9807	0.9551	0.9747	0.8908
QNSTR	0.7990	0.8327	0.9808	0.9648	0.9791	0.8936

$m = 142625$ parameters. The feasible sets X and Y are set as $[-5, 5]^n$ and $[-5, 5]^m$, respectively. We use activation function GELU except Sigmoid at the output layer of D and G. We compare our results based on problem (32) with some existing models. In our experiment, we use $\lambda = 10$ and V_k^z with $L = 4$.

We compute traditional metrics such as F1-score, Sensitivity, Specificity and Accuracy. The form of these metrics are given as follows:

$$\begin{aligned}
\text{Sensivity} &= \frac{1}{N}\sum_{i=1}^{N}\frac{|\text{GT}_i \cap \text{SR}_i|}{|\text{GT}_i \cap \text{SR}_i| + |\text{GT}_i \cap \text{SR}_i^c|},\\
\text{Specificity} &= \frac{1}{N}\sum_{i=1}^{N}\frac{|\text{GT}_i^c \cap \text{SR}_i^c|}{|\text{GT}_i^c \cap \text{SR}_i^c| + |\text{GT}_i^c \cap \text{SR}_i|},\\
\text{Accuracy} &= \frac{1}{N}\sum_{i=1}^{N}\frac{|\text{GT}_i \cap \text{SR}_i| + |\text{GT}_i^c \cap \text{SR}_i^c|}{\Omega},\\
\text{Precision} &= \frac{1}{N}\sum_{i=1}^{N}\frac{|\text{GT}_i \cap \text{SR}_i|}{|\text{GT}_i \cap \text{SR}_i| + |\text{GT}_i^c \cap \text{SR}_i|},\\
\text{F1} &= \frac{2\text{Precision} \times \text{Sensitivity}}{\text{Precision} + \text{Sensitivity}},
\end{aligned}$$

where Ω is the Universe set for all index of pixels in image, GT_i is the grouth truth vessel index for i-th image, SR is the index of pixel labelled as vessel in i-th image's segmentation result. Furthermore, we compute Area Under Curve-Receiver Operating Characteristic (AUC-ROC) [2] and Structural Similarity Index Measure (SSIM) [40].

Table 4 shows that the QNSTR algorithm for solving problem (32) is more promising for blood vessel segmentation. In Fig. 9, we visualize the error between vessel map generated by problem (32) with the QNSTR algorithm and the manual segmentation. In Fig. 10, we compare the segmentation results of problem (32) based on the QNSTR algorithm and the alternating Adam approach, respectively.

5 Conclusion

In this paper, we propose a new QNSTR algorithm for solving the large-scale min-max optimization problem (3) via the nonmonotone VI (4). Based on the structure of the problem, we use a smoothing function $F(\cdot, \mu)$ to approximate the nonsmooth function F_N, and consider

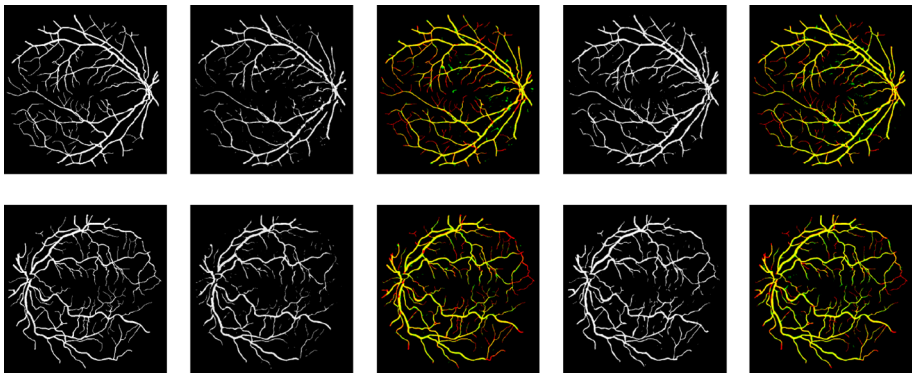
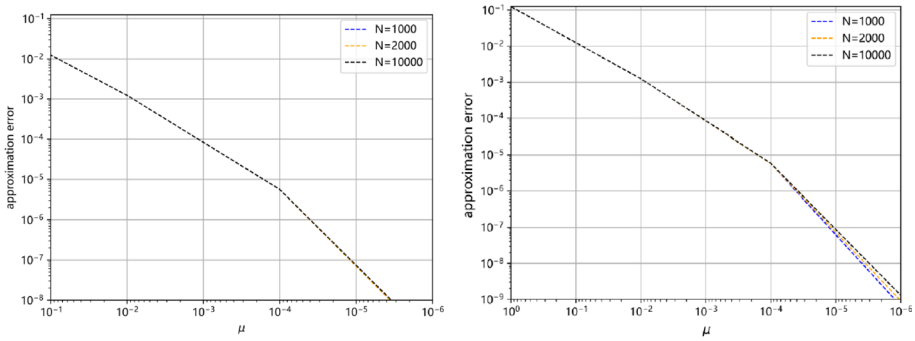


Fig. 10 Comparison of Alternating Adam and QNSTR Algorithm. Columns from left to right are: 1. manual segmentation, 2. vessel map generated by GANs with Alternating Adam, 3. yellow(correct); red(wrong); green(missing) of Alternating Adam, 4. vessel map generated by GANs with proposed Algorithm, 5. yellow(correct); red(wrong); green(missing) of QNSTR Algorithm (Color figure online)



(a) $N_G^1 = N_D^1 = 64$ (b) $N_G^1 = N_D^1 = 128$

Fig. 11 Smoothing approximation error of $F_{N,\mu}(z)$ to $F_N(z)$ for $N = 1000, 2000, 10000$

the smoothing least squares problem (19). We adopt an adaptive quasi-Newton formula in [47] to approximate the Hessian matrix and solve a strongly convex quadratic program with ellipse constraints in a low-dimensional subspace at each step of the QNSTR algorithm. We prove the global convergence of the QNSTR algorithm to a stationary point of the least squares problem, which is an ϵ-first-order stationary point of the min-max optimization problem if every element of the generalized Jacobian of F_N is nonsingular at the point. In our numerical experiments, we test the QNSTR algorithm by using two real data sets: MNIST data and DRIVE data. Preliminary numerical results validate that the QNSTR algorithm outperforms some existing algorithms.

Appendix A: Smoothing Approximation of $F_N(z)$

We consider problem (30) with a two-layers discriminator and a two-layers generator using MNIST handwritting data. All notations are the same as those in Sect. 4.1. Set $X = [-5, 5]$, $Y = [-5, 5]$, $N_G^1 = N_D^1 = 64$, $N_G^1 = N_D^1 = 128$, $\mu = 10^{-t}, t = 0, 1, \ldots, 6$ and $N =$

1000, 2000, 10, 000, respectively. Based on the uniform distribution over $[-10, 10]^{n+m}$, we generate 1000 points $z^i \in [-10, 10]^{n+m}$, $i = 1, \cdots, 1000$. Denote the "approximation error" by

$$\text{approximation error} := \frac{1}{1000}\sum_{i=1}^{1000} \|F_{N,\mu}(z^i) - F_N(z^i)\|_\infty.$$

In Fig. 11, we plot the average of "approximation error" under different choices of μ and N with 20 sets of 1000 points in $[-10, 10]^{n+m}$. From the figure, we can observe that for each $N = 1000, 2000, 10000$, the "approximation error" converges to zero as μ tends to zero.

Appendix B: Proof of Theorem 2

As we discussed in subsection 3.2, there is a positive number M_1 such that $\|H\| \le M_1$ for any $H \in \partial(\nabla r(z))$, $z \in S(R_0)$. Moreover there is a positive number M_2 such that $\left\|\nabla r(z) - \nabla r(z')\right\| \le M_2 \left\|z - z'\right\|$, $\forall z, z' \in S(R_0)$.

Since z_k is updated only when $\rho_k > \eta$ in Algorithm 1, we have $z_k \in S(R_0)$. From the continuous differentiability of r over $S(R_0)$, there is a positive number M_3 such that $\|J_k\| \le M_3$ for $k \in \mathbb{N}$. Moreover, from (23), $\|A_k\| = \max\{\|B_k\|, \|C_k\|\} \le \max\{\gamma, \|F(z_0)\|\}$. Hence $\|H_k\| \le M_3 + \max\{\gamma, \|F(z_0)\|\}$. Let $M \ge \max\{M_2, M_3 + \gamma, M_3 + \|F(z_0)\|\}$. Then we have $\left\|\nabla r(z) - \nabla r(z')\right\| \le M \left\|z - z'\right\|$ for any $z, z' \in S(R_0)$ and $\|H_k\| \le M$ for $k \in \mathbb{N}$.

In the rest of Appendix B, we prove $\lim_{k\to\infty} \|g_k\| = 0$.

If $g_k = 0$ for some $k > 0$, then Algorithm 1 terminates and Theorem 2 holds. In the remainder, we only consider the case that $g_k \neq 0$.

We next consider the following one-dimensional problem:

$$\min_{\tau} m_k(\tau\alpha_k^s) \quad \text{s.t.} \quad \|\tau V_k \alpha_k^s\| \le \Delta_k, \quad \tau > 0, \tag{34}$$

where α_k^s is an optimal solution of

$$\min_{\alpha} c_k^\top \alpha \quad \text{s.t.} \ \|V_k\alpha\| \le \Delta_k. \tag{35}$$

Let τ_k denote an arbitrary optimal solution of problem (34). Then $\alpha_k^C := \tau_k\alpha_k^s$ is a feasible solution of problem (28).

In what follows, we give the closed form of α_k^C step by step. For this purpose, we consider the KKT condition of problem (35) as follows:

$$\lambda G_k\alpha + c_k = 0, \qquad 0 \le \lambda \perp \Delta_k^2 - \alpha^\top G_k\alpha \ge 0,$$

where λ is a multiplier. Since $g_k \neq 0$ and V_k is of full column rank, we have $c_k \neq 0$. Thus, $\lambda > 0$, and the KKT system gives

$$\alpha = -\frac{1}{\lambda} G_k^{-1} c_k, \qquad \Delta_k^2 = \alpha^\top G_k \alpha.$$

Then we obtain $\frac{1}{\lambda} = \sqrt{\frac{\Delta_k^2}{(G_k^{-1}c_k)^\top G_k (G_k^{-1}c_k)}}$, and the solution of (35) can written as

$$\begin{aligned}
\alpha_k^s &= -\frac{\Delta_k}{\sqrt{(G_k^{-1}c_k)^\top G_k (G_k^{-1}c_k)}} G_k^{-1}c_k \\
&= -\frac{\Delta_k}{\sqrt{c_k^\top G_k^{-1}c_k}} G_k^{-1}c_k \\
&= -\frac{\Delta_k}{\sqrt{g_k^\top V_k (V_k^\top V_k)^{-1} V_k^\top g_k}} G_k^{-1}c_k \\
&= -\frac{\Delta_k}{\|g_k\|} G_k^{-1}c_k.
\end{aligned}$$

Hence, the objective function of (34) has the following form

$$\begin{aligned}
m_k(\tau\alpha_k^s) &= r(z_k) + \tau c_k^\top \alpha_k^s + \frac{\tau^2}{2}(\alpha_k^s)^\top Q_k \alpha_k^s \\
&= r(z_k) - \tau \frac{\Delta_k}{\|g_k\|} c_k^\top G_k^{-1}c_k + \frac{\tau^2}{2}\left(\frac{\Delta_k}{\|g_k\|}\right)^2 (G_k^{-1}c_k)^\top Q_k G_k^{-1}c_k \\
&= r(z_k) - \Delta_k\|g_k\|\tau + \frac{\tau^2}{2}\left(\frac{\Delta_k}{\|g_k\|}\right)^2 g_k^\top H_k g_k
\end{aligned}$$

and the constraint of (34) satisfies

$$\|\tau\alpha_k^s\|_{G_k} = \tau\frac{\Delta_k}{\|g_k\|}\sqrt{c_k^\top G_k^{-1} G_k G_k^{-1} c_k} = \tau\Delta_k \le \Delta_k,$$

which is equivalent to $0 < \tau \le 1$.

Therefore, problem (34) can be equivalently rewritten as

$$\min_\tau \; -\Delta_k\|g_k\|\tau + \tfrac{\tau^2}{2}\left(\tfrac{\Delta_k}{\|g_k\|}\right)^2 g_k^\top H_k g_k, \quad \text{s.t.} \quad 0 < \tau \le 1. \tag{36}$$

Since H_k is positive definite (see (23), (26) and (27)), problem (36) has the unique solution

$$\tau_k = \min\left(\|g_k\|^3/(\Delta_k g_k^\top H_k g_k), 1\right).$$

Finally, we obtain

$$\alpha_k^C = -\min\left(\|g_k\|^3/(\Delta_k g_k^\top H_k g_k), 1\right)\frac{\Delta_k}{\|g_k\|} G_k^{-1}c_k. \tag{37}$$

Lemma 4 *Let α_k be the unique optimal solution of subproblem (28) in the k-th step. Then*

$$m_k(0) - m_k(\alpha_k) \ge \frac{1}{2}\|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\|H_k\|}\right).$$

Proof Since α_k^C is a feasible solution of problem (28), we have

$$m_k(0) - m_k(\alpha_k) \ge m_k(0) - m_k(\alpha_k^C).$$

In what follows, we verify

$$m_k(0) - m_k(\alpha_k^C) \ge \frac{1}{2}\|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\|H_k\|}\right).$$

If $\|g_k\|^3/(\Delta_k g_k^\top H_k g_k) < 1$, substituting α_k^C (see (37)) into (28), we have

$$\begin{aligned} m_k(0) - m_k(\alpha_k^C) &= -c_k^\top \alpha_k^C - \frac{1}{2}(\alpha_k^C)^\top Q_k \alpha_k^C \\ &= \frac{\|g_k\|^2}{g_k^\top H_k g_k} c_k^\top G_k^{-1} c_k - \frac{1}{2}\frac{\|g_k\|^4}{(g_k^\top H_k g_k)^2} c_k^\top G_k^{-1} Q_k G_k^{-1} c_k \\ &= \frac{\|g_k\|^4}{g_k^\top H_k g_k} - \frac{1}{2}\frac{\|g_k\|^4}{(g_k^\top H_k g_k)^2} g_k^\top H_k g_k \\ &= \frac{1}{2}\frac{\|g_k\|^4}{g_k^\top H_k g_k} \geq \frac{1}{2}\frac{\|g_k\|^4}{\|H_k\|\|g_k\|^2} = \frac{1}{2}\frac{\|g_k\|^2}{\|H_k\|}. \end{aligned} \tag{38}$$

If $\|g_k\|^3/(\Delta_k g_k^\top H_k g_k) \geq 1$ (i.e., $g_k^\top H_k g_k \leq \frac{\|g_k\|^3}{\Delta_k}$), we have

$$\begin{aligned} m_k(0) - m_k(\alpha_k^C) &= -c_k^\top \alpha_k^C - \frac{1}{2}(\alpha_k^C)^\top Q_k \alpha_k^C \\ &= \frac{\Delta_k}{\|g_k\|}\|c_k^\top G_k^{-1} c_k\| - \frac{1}{2}\frac{\Delta_k^2}{\|g_k\|^2} c_k^\top G_k^{-1} Q_k G_k^{-1} c_k \\ &= \Delta_k\|g_k\| - \frac{1}{2}\frac{\Delta_k^2}{\|g_k\|^2} g_k^\top H_k g_k \\ &\geq \Delta_k\|g_k\| - \frac{1}{2}\frac{\Delta_k^2}{\|g_k\|^2}\frac{\|g_k\|^3}{\Delta_k} \\ &= \frac{1}{2}\|g_k\|\Delta_k. \end{aligned} \tag{39}$$

Combining (38) and (39), we complete the proof. □

Lemma 5 *Under assumptions of Theorem 2, for any index k, there exists a $\bar{k} > k$ such that $\|g_{\bar{k}}\| < \|g_k\|/2$.*

Proof We give the proof by contradiction. Suppose that there exists a $\hat{k}$ with $\|g_{\hat{k}}\| = 2\epsilon$ for some $\epsilon > 0$, and $\|g_k\| \geq \epsilon$, $\forall k \geq \hat{k}$. Then we know from Lemma 4 that

$$m_k(0) - m_k(\alpha_k) \geq \frac{1}{2}\|g_k\| \min\left(\Delta_k, \frac{\|g_k\|}{\|H_k\|}\right) \geq \frac{1}{2}\epsilon \min\left(\Delta_k, \frac{\epsilon}{M}\right). \tag{40}$$

According to the definition of ρ_k in (30), we have

$$\begin{aligned} |\rho_k - 1| &= \left|\frac{r(z_k) - r(z_k + V_k\alpha_k) - (m_k(0) - m_k(\alpha_k))}{m_k(0) - m_k(\alpha_k)}\right| \\ &= \left|\frac{m_k(\alpha_k) - r(z_k + V_k\alpha_k)}{m_k(0) - m_k(\alpha_k)}\right|. \end{aligned} \tag{41}$$

By Taylor expansion, we have

$$r(z_k + V_k\alpha_k) = r(z_k) + g_k^\top V_k\alpha_k + \int_0^1 (\nabla r(z_k + tV_k\alpha_k) - \nabla r(z_k))^\top V_k\alpha_k \mathrm{d}t.$$

Then

$$\begin{aligned} |m_k(\alpha_k) - r(z_k + V_k\alpha_k)| &= \left|\frac{1}{2}\alpha_k^\top Q_k\alpha_k - \int_0^1 (\nabla r(z_k + tV_k\alpha_k) - \nabla r(z_k))^\top V_k\alpha_k \mathrm{d}t\right| \\ &\leq (M/2)\|V_k\alpha_k\|^2 + M\|V_k\alpha_k\|^2 \leq 3\Delta_k^2 M/2, \end{aligned} \tag{42}$$

where the first inequality follows from $Q_k = V_k^\top H_k V_k$ and the mean-value theorem, and the second inequality follows from $\|V_k \alpha_k\| \le \Delta_k$ due to the constraint in problem (28).

Then, by (40), (41) and (42), we get

$$|\rho_k - 1| \le \frac{3\Delta_k^2 M/2}{(\epsilon/2)\min(\Delta_k, \epsilon/M)}.$$

Denote

$$\tilde{\Delta} := \min\left(\frac{(1-\zeta_1)\epsilon}{3M}, R_0\right).$$

For any $\Delta_k \le \tilde{\Delta}$, we have

$$|\rho_k - 1| \le \frac{3\Delta_k^2 M/2}{(\epsilon/2)\min(\Delta_k, \epsilon/M)} = \frac{3M\Delta_k^2}{\epsilon\Delta_k} = \frac{3M\Delta_k}{\epsilon} \le \frac{3M\tilde{\Delta}}{\epsilon} \le 1 - \zeta_1,$$

which implies $\rho_k \ge \zeta_1$, where the first equality follows from the fact that

$$\Delta_k \le \tilde{\Delta} = \min\left(\frac{(1-\zeta_1)\epsilon}{3M}, R_0\right) \le \frac{\epsilon}{3M} < \frac{\epsilon}{M}.$$

The above observation together with update rules in Algorithm 1 indicates that $\Delta_{k+1} \ge \Delta_k$ when $\Delta_k \le \tilde{\Delta}$ (and thus $\rho_k \ge \zeta_1$). In other words, if $\Delta_k > \tilde{\Delta}$, $\rho_k < \zeta_1$ holds. In this case, $\Delta_{k+1} = \beta_1 \Delta_k > \beta_1 \tilde{\Delta}$. To summarize the two cases, we then have

$$\Delta_k \ge \min(\Delta_{k-1}, \beta_1\tilde{\Delta}) \ge \cdots \ge \min(\Delta_{\hat{k}}, \beta_1\tilde{\Delta}), \quad \forall k \ge \hat{k}. \tag{43}$$

Then we proved sequence $\{\Delta_k\}_{k\ge\hat{k}}^{\infty}$ is bounded from below. Note that there exists an infinite subsequence, denoted by $\mathcal{K}$, of $\{\hat{k}, \hat{k}+1, \cdots\}$ such that, for any $k \in \mathcal{K}$, one of the following two cases holds.

Case 1: $\Delta_{k+1} = \beta_1 \Delta_k$. It is easy to obtain $\Delta_k \to 0$ as $k \overset{\mathcal{K}}{\to} \infty$ since $\beta_1 < 1$, which is contradicted to the fact that Δ_k is bounded from below (see (43)).

Case 2: $\rho_k \ge \zeta_1$. We have from the definition of ρ_k (see (30)) and $\rho_k \ge \zeta_1$ that

$$r(z_k) - r(z_{k+1}) \ge \zeta_1(m_k(0) - m_k(\alpha_k)) \ge \zeta_1 \frac{1}{2}\epsilon \min(\Delta_k, \epsilon/M) > 0,$$

where the second inequality follows from Lemma 4 and $\|g_k\| \ge \epsilon$ for $k \in \mathcal{K}$.

Therefore, $\{r(z_k)\}_{k\in\mathcal{K}}$ is strictly decreasing. Since $\{r(z_k)\}_{k\in\mathcal{K}}$ is bounded from below (note that $r(z) \ge 0$ for any z), we know that the sequence $\{r(z_k)\}_{k\in\mathcal{K}}$ is convergent and $r(z_k) - r(z_{k+1}) \downarrow 0$ as $k \overset{\mathcal{K}}{\to} \infty$. Thus, $\Delta_k \to 0$ as $k \overset{\mathcal{K}}{\to} \infty$, which is also contradicted with (43). □

Now we are ready to prove $\lim_{k\to\infty} \|g_k\| = 0$.

Proof (of $\lim_{k\to\infty} \|g_k\| = 0$) Let

$$\epsilon := \frac{1}{2}\|g_k\| \quad \text{and} \quad R := \min\left(\frac{\epsilon}{M}, R_0\right). \tag{44}$$

Note that $\mathcal{B}(z_k, R) = \{z : \|z - z_k\| \le R\} \subseteq S(R_0)$, and thus $\nabla r(\cdot)$ is Lipschitz continuous on $\mathcal{B}(z_k, R)$ with Lipschitz modulus M. Thus, for $\forall z \in \mathcal{B}(z_k, R)$, we have

$$\|\nabla r(z) - \nabla r(z_k)\| \le M\|z - z_k\| \le MR = M\min\left(\frac{\epsilon}{M}, R_0\right) \le \epsilon.$$

For $\forall z \in \mathcal{B}(z_k, R)$, we have by the triangle inequality that

$$\|\nabla r(z)\| \geq \|g_k\| - \|\nabla r(z) - \nabla r(z_k)\| = 2\epsilon - \|\nabla r(z) - \nabla r(z_k)\| \geq 2\epsilon - \epsilon = \epsilon.$$

According to Lemma 5, we know that there exists an index $l \geq k$ satisfying $\|g_{l+1}\| < \epsilon$. Moreover, we assume that z_{l+1} is the first point that iterates out of the ball $\mathcal{B}(z_k, R)$ after z_k as well as satisfying $\|g_{l+1}\| < \epsilon$. Thus, $\|g_i\| \geq \epsilon$ for $i = k, k+1, \cdots, l$. Then we have

$$\begin{aligned} r(z_k) - r(z_{l+1}) &= \sum_{i=k}^{l} r(z_i) - r(z_{i+1}) = \sum_{\substack{i=k,\\ z_i \neq z_{i+1}}}^{l} \rho_i (m_i(0) - m_i(\alpha_i)) \\ &\geq \sum_{\substack{i=k,\\ z_i \neq z_{i+1}}}^{l} \eta (m_i(0) - m_i(\alpha_i)) \geq \frac{\eta}{2}\epsilon \sum_{\substack{i=k,\\ z_i \neq z_{i+1}}}^{l} \min\left(\Delta_i, \frac{\epsilon}{M}\right), \end{aligned} \tag{45}$$

where the second equality follows from (30), the first inequality follows from $\rho_i < \eta$ when $z_i \neq z_{i+1}$. Since $\|g_k\| = 2\epsilon$ and $\|g_{l+1}\| < \epsilon$, we have $z_{l+1} \neq z_k$, which implies that $\{k, \cdots, l\} \cap \{j : z_j \neq z_{j+1}\} \neq \emptyset$.

If $\Delta_i \leq \epsilon/M$ for all $i \in \{k, \cdots, l\} \cap \{j : z_j \neq z_{j+1}\}$, we continue (45) as follows:

$$\begin{aligned} r(z_k) - r(z_{l+1}) &\geq \frac{\eta}{2}\epsilon \sum_{\substack{i=k,\\ z_i \neq z_{i+1}}}^{l} \Delta_i \geq \frac{\eta}{2}\epsilon \sum_{i=k}^{l} \|z_{i+1} - z_i\| \\ &\geq \frac{\eta}{2}\epsilon \|z_k - z_{l+1}\| \geq \frac{\eta}{2}\epsilon R = \frac{\eta}{2}\epsilon \min\left(\frac{\epsilon}{M}, R_0\right), \end{aligned}$$

where the second inequality follows from $\|z_{i+1} - z_i\| \leq \Delta_i$, the third inequality follows from the triangle inequality, the last inequality follows from the fact that z_{l+1} is the first point that iterates out of the ball $\mathcal{B}(z_k, R)$ after z_k.

If $\Delta_i > \epsilon/M$ for some $i \in \{k, \cdots, l\} \cap \{j : z_j \neq z_{j+1}\}$, we continue (45) as follows:

$$r(z_k) - r(z_{l+1}) \geq \frac{\eta}{2}\epsilon \sum_{\substack{i=k,\\ z_i \neq z_{i+1}}}^{l} \frac{\epsilon}{M} \geq \frac{\eta}{2}\epsilon \frac{\epsilon}{M},$$

where the last inequality follows from $\{k, \cdots, l\} \cap \{j : z_j \neq z_{j+1}\} \neq \emptyset$. To summarize, we obtain

$$r(z_k) - r(z_{l+1}) \geq \frac{\eta}{2}\epsilon \min\left(\frac{\epsilon}{M}, R_0\right). \tag{46}$$

Since the sequence $\{r(z_i)\}_{i=0}^{\infty}$ is a decreasing and bounded sequence from below, there exists $r^* \geq 0$ such that $\lim_{i\to\infty} r(z_i) = r^*$. Hence

$$r(z_k) - r^* \geq r(z_k) - r(z_{l+1}) \geq \frac{\eta}{2}\epsilon \min\left(\frac{\epsilon}{M}, R_0\right) = \frac{\eta}{4}\|g_k\| \min\left(\frac{\|g_k\|}{2M}, R_0\right),$$

where the second inequality follows from (46), the last equality follows from (44).

Due to the arbitrariness of k, by letting $k \to \infty$, we know

$$\frac{\eta}{4}\|g_k\| \min\left(\frac{\|g_k\|}{2M}, R_0\right) \to 0,$$

which implies $\lim_{k\to\infty} \|g_k\| = 0$. Then the proof is complete. □

Acknowledgements We would like to thank two referees for their very helpful comments. We would like to thank Prof. Yinyu Ye for his suggestion to add more search directions in the Dimension Reduced Second-Order Method proposed in [44].

Funding Open access funding provided by The Hong Kong Polytechnic University This work is supported by CAS AMSS-PolyU Joint Laboratory of Applied Mathematics, University Research Facility in Big Data Analytics, PolyU and the Hong Kong Research Grant Council PolyU15300021, C5036-21E, N-PolyU507/22.

Data Availibility All used data sets in Sect. 4.1 were downloaded from the website https://www.tensorflow.org/datasets/catalog/mnist, and all used data sets in Sect. 4.2 were downloaded from the website https://drive.grand-challenge.org/.

Declarations

Conflict of interest The authors declare that they have no Conflict of interest.

References

1. Alom, M.Z., Hasan, M., Yakopcic, C., Taha, T.M., Asari, V.K.: Recurrent residual convolutional neural network based on U-net (R2U-Net) for medical image segmentation. arXiv preprint arXiv:1802.06955 (2018)
2. Bradley, A.P.: The use of the area under the ROC curve in the evaluation of machine learning algorithms. Pattern Recognit. **30**, 1145–1159 (1997)
3. Chen, X.: Superlinear convergence of smoothing quasi-Newton methods for nonsmooth equations. J. Comput. Appl. Math. **80**, 105–126 (1997)
4. Chen, X., Qi, L., Sun, D.: Global and superlinear convergence of the smoothing Newton method and its application to general box constrained variational inequalities. Math. Comput. **67**, 519–540 (1998)
5. Clarke, F.H.: Optimization and Nonsmooth Analysis. Wiley, New York (1983)
6. Conn, A.R., Gould, N.I., Toint, P.L.: Trust Region Methods. SIAM, Philadelphia (2000)
7. Dai, B., Shaw, A., Li, L., Xiao, L., He, N., Liu, Z., Chen, J., Song, L.: SBEED: convergent reinforcement learning with nonlinear function approximation. In: Proceedings of the International Conference on Machine Learning, pp. 1125–1134. PMLR (2018)
8. Dai, Y.H., Zhang, L.: Optimality conditions for constrained minimax optimization. CSIAM Trans. Appl. Math. **1**, 296–315 (2020)
9. Daskalakis, C., Ilyas, A., Syrgkanis, V., Zeng, H.: Training gans with optimism. arXiv preprint arXiv:1711.00141 (2017)
10. Dennis, J.E., Schnabel, R.B.: Numerical Methods for Unconstrained Optimization and Nonlinear Equations. Prentice-Hall, Englewood Cliffs, NJ (1983)
11. Diakonikolas, J., Daskalakis, C., Jordan, M.I.: Efficient methods for structured nonconvex-nonconcave min-max optimization. In: Proceedings of the International Conference on Artificial Intelligence and Statistics, pp. 2746–2754. PMLR (2021)
12. Facchinei, F., Pang, J.S.: Finite-dimensional Variational Inequalities and Complementarity Problems. Springer, New York (2003)
13. Forsyth, D.A., Ponce, J.: Computer Vision: A Modern Approach. Prentice Hall Professional Technical Reference, Upper Saddle River (2002)
14. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: Generative adversarial nets. Adv. Neural Inform. Process. Syst. **27** (2014)

15. Gratton, S., Lawless, A.S., Nichols, N.K.: Approximate Gauss-Newton methods for nonlinear least squares problems. SIAM J. Optim. **18**, 106–132 (2007)
16. Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus). arXiv preprint arXiv:1606.08415 (2016)
17. Isola, P., Zhu, J.Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1125–1134 (2017)
18. Jiang, J., Chen, X.: Optimality conditions for nonsmooth nonconvex-nonconcave min-max problems and generative adversarial networks. SIAM J. Math. Data Sci. **5**, 693–722 (2023)
19. Jin, C., Netrapalli, P., Jordan, M.: What is local optimality in nonconvex-nonconcave minimax optimization? In: Proceedings of International Conference on Machine Learning, pp. 4880–4889. PMLR (2020)
20. Jin, Q., Meng, Z., Pham, T.D., Chen, Q., Wei, L., Su, R.: DUNet: a deformable network for retinal vessel segmentation. Knowl.-Based Syst. **178**, 149–162 (2019)
21. Li, L., Verma, M., Nakashima, Y., Nagahara, H., Kawasaki, R.: Iternet: Retinal image segmentation utilizing structural redundancy in vessel networks. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3656–3665 (2020)
22. Lin, T., Jin, C., Jordan, M.: On gradient descent ascent for nonconvex-concave minimax problems. In: Proceedings of the International Conference on Machine Learning, pp. 6083–6093. PMLR (2020)
23. Lin, T., Jin, C., Jordan, M.I.: Near-optimal algorithms for minimax optimization. In: Proceedings of the Conference on Learning Theory, pp. 2738–2779. PMLR (2020)
24. Liu, M., Rafique, H., Lin, Q., Yang, T.: First-order convergence theory for weakly-convex-weakly-concave min-max problems. J. Mach. Learn. Res. **22**, 7651–7684 (2021)
25. Luc, P., Couprie, C., Chintala, S., Verbeek, J.: Semantic segmentation using adversarial networks. arXiv preprint arXiv:1611.08408 (2016)
26. Madry, A., Makelov, A., Schmidt, L., Tsipras, D., Vladu, A.: Towards deep learning models resistant to adversarial attacks. arXiv preprint arXiv:1706.06083 (2017)
27. Mohajerin Esfahani, P., Kuhn, D.: Data-driven distributionally robust optimization using the Wasserstein metric: performance guarantees and tractable reformulations. Math. Program. **171**, 115–166 (2018)
28. Monteiro, R.D., Svaiter, B.F.: Complexity of variants of Tseng's modified FB splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. SIAM J. Optim. **21**, 1688–1720 (2011)
29. Myerson, R.B.: Game Theory: Analysis of Conflict. Harvard University Press, Cambridge (1991)
30. Nemirovski, A.: Prox-method with rate of convergence O (1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. SIAM J. Optim. **15**, 229–251 (2004)
31. Nesterov, Y.: Dual extrapolation and its applications to solving variational inequalities and related problems. Math. Program. **109**, 319–344 (2007)
32. Rabbat, M., Nowak, R.: Distributed optimization in sensor networks. In: Proceedings of the 3rd International Symposium on Information Processing in Sensor Networks, pp. 20–27 (2004)
33. Rafique, H., Liu, M., Lin, Q., Yang, T.: Weakly-convex-concave min-max optimization: provable algorithms and applications in machine learning. Optim. Method Softw. **37**, 1087–1121 (2022)
34. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Proceedings of the Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, pp. 234–241. Springer (2015)
35. Shapiro, A.: Monte Carlo sampling methods. In: Ruszczyński, A., Shapiro, A. (eds.) Handbooks in Operations Research and Management Science, pp. 353–425. Elsevier, Amsterdam (2003)
36. Shapiro, A., Dentcheva, D., Ruszczynski, A.: Lectures on Stochastic Programming: Modeling and Theory. SIAM, Philadelphia (2021)
37. Son, J., Park, S.J., Jung, K.H.: Towards accurate segmentation of retinal vessels and the optic disc in fundoscopic images with generative adversarial networks. J. Digit. Imaging **32**, 499–512 (2019)
38. Szeliski, R.: Computer Vision: Algorithms and Applications. Springer Nature, Switzerland (2022)
39. Tseng, P.: On accelerated proximal gradient methods for convex-concave optimization. submitted to SIAM J. Optim. **2** (2008)
40. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**, 600–612 (2004)
41. Xu, H.: Uniform exponential convergence of sample average random functions under general sampling with applications in stochastic programming. J. Math. Anal. Appl. **368**, 692–710 (2010)
42. Yang, J., Kiyavash, N., He, N.: Global convergence and variance reduction for a class of nonconvex-nonconcave minimax problems. Adv. Neural. Inf. Process. Syst. **33**, 1153–1165 (2020)
43. Yuan, Y.: Recent advances in trust region algorithms. Math. Program. **151**, 249–281 (2015)

44. Zhang, C., Ge, D., Jiang, B., Ye, Y.: DRSOM: A dimension reduced second-order method and preliminary analyses. arXiv preprint arXiv:2208.00208 (2022)
45. Zhang, G., Wang, Y., Lessard, L., Grosse, R.B.: Near-optimal local convergence of alternating gradient descent-ascent for minimax optimization. In: Proceedings of the International Conference on Artificial Intelligence and Statistics, pp. 7659–7679. PMLR (2022)
46. Zhang, Y., Yang, L., Chen, J., Fredericksen, M., Hughes, D.P., Chen, D.Z.: Deep adversarial networks for biomedical image segmentation utilizing unannotated images. In: Proceedings of the Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, pp. 408–416. Springer (2017)
47. Zhou, W., Chen, X.: Global convergence of a new hybrid Gauss-Newton structured BFGS method for nonlinear least squares problems. SIAM J. Optim. **20**, 2422–2441 (2010)