ANDERSON ACCELERATION FOR NONSMOOTH FIXED POINT PROBLEMS*

WEI BIAN† AND XIAOJUN CHEN‡

Abstract. We give new convergence results of Anderson acceleration for the composite max fixed point problem. We prove that Anderson(1) and EDIIS(1) are q-linear convergent with a smaller q-factor than existing q-factors. Moreover, we propose a smoothing approximation of the composite max function in the contractive fixed point problem. We show that the smoothing approximation is a contraction mapping with the same fixed point as the composite max fixed point problem. Our results rigorously confirm that the nonsmoothness does not affect the convergence rate of Anderson acceleration method when we use the proposed smoothing approximation for the composite max fixed point problem. Numerical results for constrained minimax problems, complementarity problems and nonsmooth differential equations are presented to show the efficiency and good performance of the proposed Anderson acceleration method with smoothing approximation.

Key words. Anderson acceleration, smoothing approximation, composite max function, minimax problem, complementarity problem

AMS subject classifications. 65H10, 68W25

1. Introduction.

In this paper, we focus on the convergence analysis of Anderson(m) and EDIIS(m) for the following composite max fixed point problem

$$u = G(u) := H(P_\Omega(Q(u))), \tag{1.1}$$

where $H : \mathbb{R}^l \to \mathbb{R}^n$ and $Q : \mathbb{R}^n \to \mathbb{R}^l$ are Lipschitz continuously differentiable functions, Ω is a box subset of $\mathbb{R}^l$, and P_Ω is the projection on Ω. Problem (1.1) arises from many applications in engineering and finance including minimax problems, complementarity problems, nonsmooth integral equations and nonsmooth differential equations.

Anderson acceleration was originally introduced in the context of integral equations by Anderson in 1965 [2]. It is a class of methods for solving the fixed point problem $u = G(u)$, where G is a continuous function from $D \subseteq \mathbb{R}^n$ to D, and uses a history of search directions to improve the convergence rate of the fixed point method

$$u_{k+1} = G(u_k). \tag{1.2}$$

Anderson acceleration method has been widely used in electronic structure computation [2, 6, 11, 22, 24, 25], chemistry and physics [1, 23], and specific optimization problems [13, 25]. In particular, Anderson acceleration is designed to solve the fixed point problem when computing the Jacobian of G is impossible or too costly. Anderson acceleration is also known as the Pilay mixing [20], DIIS (direct inversion on iterative subspace) [14, 15, 23], nonlinear GMRES method [4, 16, 26], and interface quasi-Newton [10, 12]. A formal description of Anderson acceleration is presented in Algorithm 1.1 and often called Anderson(m).

Anderson(m) maintains a history of function values of $G(\cdot)$ at u_{k-m_k+j}, $j = 0, \ldots, m_k$, where m_k is an algorithmic parameter that indicates the depth of the accelerated Anderson iterations.

*Submitted to the editors April 30, 2022.

†Department of Mathematics, Harbin Institute of Technology, Harbin, China (bianweilvse520@163.com). The work of this author was partially supported by the NSF foundation (11871178,61773136) of China .

‡Department of Applied Mathematics, The Hong Kong Polytechnic University, Hong Kong, China (maxjchen@polyu.edu.hk). The work of this author was partially supported by Hong Kong Research Grant Council grant (15300210)

Algorithm 1.1 Anderson(m)

Choose $u_0 \in D$ and a positive integer m. Set $u_1 = G(u_0)$ and $F_0 = G(u_0) - u_0$.
for $k = 1, 2, \ldots$ **do**
  set $F_k = G(u_k) - u_k$;
  choose $m_k = \min\{m, k\}$;
  solve

$$\min \left\| \sum_{j=0}^{m_k} \alpha_j F_{k-m_k+j} \right\| \qquad \text{s.t.} \sum_{j=0}^{m_k} \alpha_j = 1 \tag{1.3}$$

  to find a solution $\{\alpha_j^k : j = 0, \ldots, m_k\}$, and set

$$u_{k+1} = \sum_{j=0}^{m_k} \alpha_j^k G(u_{k-m_k+j}). \tag{1.4}$$

end for

Using these function values, Anderson(m) defines a new iterate by a linear combination of the last $m_k + 1$ iterates, where the coefficients of the linear combination are computed at each iteration by the convex optimization problem in (1.3). When $m = 0$, Anderson acceleration is the fixed point method in (1.2), which is also known as the Picard method. In practice, each m_k may be different to maintain the acceptable conditioning of $(F_{k-m_k+j})_{j=0}^{m_k}$ [25] and can be dynamically updated to improve the performance [18]. The optimization problem (1.3) in Anderson(m) does not specify the norm in its general form and using different norms will not affect the convergence [24]. Throughout this paper, we consider problem (1.3) in the sense of Euclidean norm. Notice that the description of Anderson(m) in Algorithm 1.1 is convenient for analysis, but the readers may refer to [24, 25] and references therein for its efficient implementation.

The EDIIS(m) [14] differs from Anderson(m) by adding nonnegativity constraints in (1.3), that is, replacing (1.3) by the following minimization problem

$$\min \left\| \sum_{j=0}^{m_k} \alpha_j F_{k-m_k+j} \right\| \quad \text{s.t.} \sum_{j=0}^{m_k} \alpha_j = 1, \quad \alpha_j \geq 0,\ j = 0, \ldots, m_k.$$

Suppose $G : D \to D$ is a contraction mapping with factor $c \in (0, 1)$ in the Euclidean norm $\|\cdot\|$ on a closed set $D \subset \mathbb{R}^n$, that is,

$$\|G(u) - G(v)\| \leq c\|u - v\|,\ \forall\, u, v \in D.$$

By the contraction mapping theorem [17], G has a unique fixed point $u^* \in D$, which is the unique solution of the system of nonlinear equations

$$F(u) := G(u) - u = 0.$$

Without loss of generality, we assume that there is $\mathcal{B}(\delta, u^*) := \{u \in \mathbb{R}^n : \|u - u^*\| \leq \delta\} \subset D$ with $\delta > 0$. For a contraction mapping G, it is known that the fixed-point method in (1.2) has q-linear convergence rate, that is $\|u_{k+1} - u^*\| \leq c\|u_k - u^*\|$ holds in $\mathcal{B}(\delta, u^*)$. However, the theoretical convergence analysis of Anderson(m) had not been proved for a long time after it being brought forward and widely used. The first mathematical convergence result for Anderson(m) was given by

Toth and Kelley in 2015 [24]. Under the assumption that G is Lipschitz continuously differentiable in D, Toth and Kelley [24] showed the r-linear convergence of Anderson(m) with r-factor $\hat{c} \in (c, 1)$ as follows,

$$\|F(u_k)\| \le \hat{c}^k \|F(u_0)\| \quad \text{and} \quad \|u_k - u^*\| \le \left(\frac{1+c}{1-c}\right) \hat{c}^k \|u_0 - u^*\|.$$

Without the differentiability of G, Chen and Kelley [6] showed the r-linear convergence of EDIIS(m) with r-factor $\hat{c} = c^{1/(m+1)}$ as follows,

$$\|u_k - u^*\| \le \hat{c}^k \|u_0 - u^*\|.$$

Moreover, Bian, Chen and Kelley [3] showed the q-linear convergence of Anderson(1) and EDIIS(1) with q-factor $(3c - c^2)/(1 - c)$ for general nonsmooth fixed point problems in a Hilbert space, and r-linear convergence of Anderson(m) and EDIIS(m) with r-factor $\hat{c} \in (c, 1)$ for a class of integral equations in which the operator can be written as the sum of a smooth term and a nonsmooth term having a sufficiently small Lipschitz constant. Zhang et al [27] proposed a globally convergent variant of Anderson acceleration for nonsmooth fixed point problems, but did not provide a rate of convergence. The first mathematical view to show the superiority of local convergence of Anderson method for the discretizations of the steady Navier-Stokes equations was proved by Pollock, Rebholz and Xiao in [19]. And the similar idea was extended to a more general fixed-point iterations by Evans, Pollock, Rebholz, and Xiao [9]. Most recently, Pollock and Rebholz [18] showed a novel one-step bound of Anderson method with a more general acceleration iteration, which not only sharpens the convergence results for contractive mapping in [9], but also explains some mechanism of Anderson acceleration for noncontractive cases. Overall, Anderson acceleration can significantly improve the computational performance of the fixed point method in practice. We refer the readers to [9, 11, 18, 25] and references therein for detailed discussions on its research history and practical applications.

Throughout this paper, we suppose Ω is defined by

$$\Omega = \{w \in \mathbb{R}^l \,|\, \underline{w} \le w \le \overline{w}\} \tag{1.5}$$

with $\underline{w} \in \{\{-\infty\} \cup \mathbb{R}\}^l$, $\overline{w} \in \{\{\infty\} \cup \mathbb{R}\}^l$ and $\underline{w} < \overline{w}$. Then, P_Ω can be expressed by the following composite max form

$$P_\Omega(w) = \mathrm{argmin}_{v \in \Omega} \|v - w\|^2 = \max\{\underline{w} - w, 0\} + w - \max\{w - \overline{w}, 0\}, \tag{1.6}$$

where "max" means componentwise. The formulation of P_Ω in (1.6) will play a key role in the analysis of this paper. Here we declare that $(-\infty) - a = -\infty$ and $a - (\infty) = -\infty$ for any $a \in \mathbb{R}$. When $\Omega = \mathbb{R}^l_+ := \{w \in \mathbb{R}^l \,|\, w \ge 0\}$, the expression of P_Ω in (1.6) is reduced to

$$P_\Omega(w) = \max\{-w, 0\} + w.$$

In particular, if $\underline{w}_i = -\infty$ and $\overline{w}_i = \infty$ for all $i \in \{1, \ldots, l\}$, then $G = H(Q(u))$ is Lipschitz continuously differentiable on D, which is the case considered in [24]. Thus, we focus on the case that there is at least an $i \in \{1, \ldots, l\}$ such that $-\infty < \underline{w}_i$ or $\overline{w}_i < \infty$, which means that G is nonsmooth on D in general.

The contributions of this paper are new convergence results of Anderson acceleration method for composite max fixed point problem (1.1). In section 2, we prove that Anderson(1) and EDIIS(1) are q-linear convergent for problem (1.1) with q-factor $\hat{c} \in (\frac{2c-c^2}{1-c}, 1)$, which can be strictly smaller than the existing q-factor $(3c-c^2)/(1-c)$ proved in [3, 24]. In section 3, we give the contraction consistent properties between G and its smoothing approximations. Then, we propose a new smoothing approximation $\mathcal{G}(\cdot,\mu)$ of G. We show that there is $\bar{\mu}>0$, such that $\mathcal{G}(\cdot,\mu)$ is continuously differentiable, contractive on D, and $u^*=\mathcal{G}(u^*,\mu)=G(u^*)$, for any fixed $\mu\in(0,\bar{\mu}]$. To improve the ability and performance of Anderson acceleration method for solving problem (1.1), we propose a smoothing Anderson acceleration (s-Anderson(m)) in Algorithm 3.1 with the proposed smoothing function of G and updating scheme for smoothing parameters. We prove that s-Anderson(m) for (1.1) owns the same r-linear convergence rate as Anderson(m) for continuously differentiable problems. In section 4, we use numerical examples from constrained minimax problems, pricing American options and nonsmooth Dirichlet problem to illustrate our theoretical results. Preliminary numerical results show that s-Anderson(m) can efficiently solve the nonsmooth fixed point problem (1.1) and outperform Anderson(m) in most cases.

2. q-linear convergence of Anderson(1) and EDIIS(1).

For $m_k=1$, the optimal solution of problem (1.3) owns the closed form $(1-\alpha_k,\alpha_k)^{\mathrm{T}}$ with

$$\alpha_k=\frac{F_k^{\mathrm{T}}(F_k-F_{k-1})}{\|F_k-F_{k-1}\|^2} \tag{2.1}$$

and the iterate can be expressed as

$$u_{k+1}=(1-\alpha_k)G(u_k)+\alpha_k G(u_{k-1}). \tag{2.2}$$

In the remainder of this paper, we need the following assumption.

ASSUMPTION 2.1. *Functions Q and H in (1.1) satisfy the following conditions.*

(i) Q is Lipschitz continuously differentiable on D with Lipschitz constant c_Q.

(ii) H is Lipschitz continuously differentiable on an open set D_H containing Ω as a subset with Lipschitz constant c_H.

(iii) $c:=c_Hc_Q<1$.

Note that the Lipschitz continuous differentiability of Q and H cannot imply the differentiability of G on D due to the existence of projection operator P_Ω in its formulation. Since P_Ω is Lipschitz continuous with Lipschitz constant 1, from

$$\begin{aligned}\|H(P_\Omega(Q(u)))-H(P_\Omega(Q(v)))\| \le & c_H\|P_\Omega(Q(u))-P_\Omega(Q(v))\| \\ \le & c_H\|Q(u)-Q(v)\|\le c_Hc_Q\|u-v\|,\end{aligned}$$

we find that G in (1.1) is a contraction mapping on D with factor $c=c_Hc_Q$ under Assumption 2.1. Then, it gives

$$(1-c)\|u-u^*\|\le\|F(u)\|\le(1+c)\|u-u^*\|,\quad \forall u\in D. \tag{2.3}$$

The following theorem shows that the local q-linear convergence factor of Anderson(1) and EDIIS(1) can be improved to any $\hat{c}\in(\frac{2c-c^2}{1-c},1)$ for (1.1), which can be strictly smaller than the factor $\frac{3c-c^2}{1-c}$ given in [3, 24].

THEOREM 2.1. *Let $\{u_k\}$ be the sequence generated by Anderson(1) for (1.1). Suppose Assumption 2.1 holds and $\bar{c}=\frac{2c-c^2}{1-c}<1$. For any $\hat{c}\in(\bar{c},1)$, if u_0 is sufficiently close to u^*, then $\{u_k\}$ converges to u^* q-linearly with factor $\hat{c}$, i.e.*

$$\|F(u_{k+1})\|\le\hat{c}\|F(u_k)\|,\quad k=0,1,\dots. \tag{2.4}$$

Proof. Give $\epsilon > 0$. Reduce $\delta > 0$ if necessary such that $\delta \le \epsilon$ and $\mathcal{B}(\delta, u^*) \subseteq D$. Since $c \le \bar{c} < \hat{c}$, (2.4) is trivially true for $k = 0$. Then, we prove (2.4) by induction and assume it holds for $0 \le k \le K$. Let

$$0 < \varrho \le \min\{1, (\overline{w}_i - \underline{w}_i)/3 : i = 1, 2, \ldots, l\}.$$

Here we declare that $\overline{w}_i - \underline{w}_i = \infty$ if $\overline{w}_i = \infty$ or/and $\underline{w}_i = -\infty$.

By (2.1), we have

$$|\alpha_k| \le \frac{\|F(u_k)\|}{\|F(u_k) - F(u_{k-1})\|} \quad \text{and} \quad |1 - \alpha_k| \le \frac{\|F(u_{k-1})\|}{\|F(u_k) - F(u_{k-1})\|}, \ \forall k. \tag{2.5}$$

Similar to the analysis in [3, Theorem 1.3] and by the hypothesis in (2.4) for $0 \le k \le K$, we have that

$$|\alpha_k| \le \frac{\hat{c}}{1 - \hat{c}} \quad \text{and} \quad \|u_k - u^*\| \le \frac{\|F(u_k)\|}{1 - c} \le \frac{\hat{c}^k (1 + c)\|u_0 - u^*\|}{1 - c}. \tag{2.6}$$

Then,

$$\begin{aligned}\|u_{K+1} - u^*\| =& \|(1 - \alpha_K)G(u_K) + \alpha_K G(u_{K-1}) - G(u^*)\| \\ \le& c|1 - \alpha_K|\|u_K - u^*\| + c|\alpha_K|\|u_{K-1} - u^*\| \le \frac{2c(1 + c)\hat{c}^K}{(1 - c)(1 - \hat{c})}\|u_0 - u^*\|.\end{aligned}$$

Similarly,

$$\|(1 - \alpha_K)u_K + \alpha_K u_{K-1} - u^*\| \le \frac{2(1 + c)\hat{c}^K}{(1 - c)(1 - \hat{c})}\|u_0 - u^*\|.$$

Thus, there exists $\delta_0 > 0$ such that if $u_0 \in \mathcal{B}(\delta_0, u^*)$, then

$$u_k \in \mathcal{B}(\delta, u^*),\ k = 0, \ldots, K + 1 \quad \text{and} \quad (1 - \alpha_K)u_K + \alpha_K u_{K-1} \in \mathcal{B}(\delta, u^*).$$

Now we estimate $\|F(u_{K+1})\|$ by using

$$\|F(u_{K+1})\| = \|G(u_{K+1}) - u_{K+1}\| \le \|A_K\| + \|B_K\|, \tag{2.7}$$

where

$$A_K = G(u_{K+1}) - G((1 - \alpha_K)u_K + \alpha_K u_{K-1}), \quad B_K = G((1 - \alpha_K)u_K + \alpha_K u_{K-1}) - u_{K+1}.$$

The estimate of A_k is straightforward as it is in [3, 6, 24], which gives

$$\|A_K\| \le c\|(1 - \alpha_K)(G(u_K) - u_K) + \alpha_K(G(u_{K-1}) - u_{K-1})\| \le c\|F(u_K)\|. \tag{2.8}$$

Now, we estimate $\|B_K\|$. First, we note that $\psi(t) = \begin{cases} \max\{0, t\} & \text{if } |t| > \varrho \\ \dfrac{(t + \varrho)^2}{4\varrho} & \text{if } |t| \le \varrho \end{cases}$ is a smoothing approximation of $\max\{t, 0\}$. Then, by (1.6),

$$\Phi(w) = \Psi(\underline{w} - w) + w - \Psi(w - \overline{w})$$

is a smoothing approximation of $P_\Omega(w)$, where

$$\Psi(v) = (\psi(v_1), \dots, \psi(v_l))^{\mathrm{T}}.$$

By virtue of the value of ϱ, for any $i \in \{1, \dots, l\}$ and $w_i \in \mathbb{R}$, at most one of $|\underline{w}_i - w_i| \le \varrho$ and $|w_i - \overline{w}_i| \le \varrho$ holds. Then, since $|\psi'(t)| \le 1$, $\forall t \in \mathbb{R}$, for any $w, \tilde{w} \in \mathbb{R}^l$, we obtain

$$\|\Phi(w) - \Phi(\tilde{w})\| \le 2\|w - \tilde{w}\|. \tag{2.9}$$

Next, recalling the definition of ψ, we have

$$\max\{t, 0\} - \psi(t) = \begin{cases} 0 & \text{if } |t| > \varrho \\ -(\varrho - |t|)^2/4\varrho & \text{if } |t| \le \varrho, \end{cases}$$

which implies the absolute value and Lipschitz constant of $\max\{t, 0\} - \psi(t)$ on $\mathbb{R}$ are upper bounded by $\varrho/4$ and $1/2$, respectively. Then, for any $w, \tilde{w} \in \mathbb{R}^l$, we have

$$\|P_\Omega(w) - \Phi(w)\| \le \sqrt{l}\varrho/4, \tag{2.10}$$

$$\|P_\Omega(w) - \Phi(w) - (P_\Omega(\tilde{w}) - \Phi(\tilde{w}))\| \le \tfrac{1}{2}\|w - \tilde{w}\|. \tag{2.11}$$

Denote

$$G_S(u) = H(\Phi(Q(u))) \quad \text{and} \quad G_N(u) = G(u) - G_S(u).$$

Then from the definition of u_{K+1} in (2.2), we have

$$\|B_K\| \le \|M_K\| + \|N_K\|, \tag{2.12}$$

with

$$M_K = G_S((1 - \alpha_K)u_K + \alpha_K u_{K-1}) - (1 - \alpha_K)G_S(u_K) - \alpha_K G_S(u_{K-1})$$

and

$$N_K = G_N((1 - \alpha_K)u_K + \alpha_K u_{K-1}) - (1 - \alpha_K)G_N(u_K) - \alpha_K G_N(u_{K-1}).$$

Notice that ψ is Lipschitz continuously differentiable on $\mathbb{R}$. By the Lipschitz continuous differentiability of Q and H, G_S is Lipschitz continuously differentiable on $\mathcal{B}(\delta, u^*)$, which inspires us to estimate M_k exactly by the same way as in [24, Corollary 2.5] to get

$$\|M_k\| \le \frac{\gamma|\alpha_K||1 - \alpha_K|\|u_K - u_{K-1}\|^2}{2} \le \frac{\gamma\|F(u_{K-1})\|}{2(1-c)^2}\|F(u_K)\| \le \frac{\gamma(1+c)\epsilon}{2(1-c)^2}\|F(u_K)\|, \tag{2.13}$$

where γ is the Lipschitz constant of G_S' on $\mathcal{B}(\delta, u^*)$ and we use $\|F(u_{K-1})\| \le (1+c)\|u_{K-1} - u^*\| \le (1+c)\delta \le (1+c)\epsilon$ in the last inequality.

The final stage of this proof is to evaluate $\|N_K\|$, which is the main part in this proof.

To do this, the first thing is to evaluate the Lispchitz constant of G_N around u^*. For any $u, v \in \mathcal{B}(\delta, u^*)$, by the Lipschitz continuous differentiability of H and the mean value theorem for a vector-valued function, we have

$$\begin{aligned} &\|G_N(u) - G_N(v)\| = \|H(P_\Omega(Q(u))) - H(\Phi(Q(u))) - H(P_\Omega(Q(v))) + H(\Phi(Q(v)))\| \\ &= \left\| \left(\int_0^1 H'(\hat{\xi}(t))\mathrm{d}t \right) (P_\Omega(Q(u)) - P_\Omega(Q(v))) - \left(\int_0^1 H'(\bar{\xi}(t))\mathrm{d}t \right) (\Phi(Q(u)) - \Phi(Q(v))) \right\|, \end{aligned}$$

where $\hat{\xi}(t) = tP_\Omega(Q(v)) + (1-t)P_\Omega(Q(u))$ and $\bar{\xi}(t) = t\Phi(Q(v)) + (1-t)\Phi(Q(u))$. Denote

$$G_N^1 = \left(\int_0^1 H'(\hat{\xi}(t))\mathrm{d}t\right)(P_\Omega(Q(u)) - P_\Omega(Q(v))) - \left(\int_0^1 H'(\bar{\xi}(t))\mathrm{d}t\right)(P_\Omega(Q(u)) - P_\Omega(Q(v))),$$

$$G_N^2 = \left(\int_0^1 H'(\bar{\xi}(t))\mathrm{d}t\right)(P_\Omega(Q(u)) - P_\Omega(Q(v))) - \left(\int_0^1 H'(\bar{\xi}(t))\mathrm{d}t\right)(\Phi(Q(u)) - \Phi(Q(v))),$$

then

$$\|G_N(u) - G_N(v)\| \le \|G_N^1\| + \|G_N^2\|.$$

By (2.9), (2.10), and the definitions of $\hat{\xi}(t)$ and $\bar{\xi}(t)$, for any $t \in [0,1]$, it holds

$$\|\hat{\xi}(t) - \bar{\xi}(t)\| \le \|P_\Omega(Q(v)) - \Phi(Q(v))\| + \|P_\Omega(Q(u)) - \Phi(Q(u))\| \le \sqrt{l}\varrho \,/\, 2\,. \tag{2.14}$$

Due to the convexity of Ω, $\hat{\xi}(t) \in \Omega$ for all $t \in [0,1]$. Then, by (2.14), we can suppose $\bar{\xi}(t) \in D_H$ for all $t \in [0,1]$ by reducing ϱ if necessary. Moreover, since $u, v \in \mathcal{B}(\delta, u^*)$, $\bar{\xi}(t)$ and $\hat{\xi}(t)$ are bounded for all $t \in [0,1]$. Then, using the Lipschitz continuous differentiability of H, there exists $\theta > 0$ such that it holds

$$\int_0^1 \left\| H'(\hat{\xi}(t)) - H'(\bar{\xi}(t)) \right\| \mathrm{d}t \le \theta \max_{0\le t\le 1} \|\hat{\xi}(t) - \bar{\xi}(t)\|,$$

combining which with (2.14) gives

$$\|G_N^1\| \le \left(\int_0^1 \left\| H'(\hat{\xi}(t)) - H'(\bar{\xi}(t)) \right\| \mathrm{d}t\right) \|P_\Omega(Q(u)) - P_\Omega(Q(v))\| \le \left(\frac{\sqrt{l}\varrho\theta c_Q}{2}\right)\|u - v\|.$$

Thus, by reducing ϱ if necessary, we obtain

$$\|G_N^1\| \le \epsilon \|u - v\|. \tag{2.15}$$

To evaluate G_N^2, by (2.11) and $\bar{\xi}(t) \in D_H$ for all $t \in [0,1]$, we have

$$\begin{aligned} \|G_N^2\| \le& \left(\int_0^1 \|H'(\bar{\xi}(t))\|\mathrm{d}t\right) \|P_\Omega(Q(u)) - \Phi(Q(u)) - (P_\Omega(Q(v)) - \Phi(Q(v)))\| \\ \le& \frac{1}{2} c_H c_Q \|u - v\| = \frac{1}{2} c \|u - v\|. \end{aligned} \tag{2.16}$$

Hence, (2.16) together with (2.15) gives that the Lipschitz constant of G_N around u^* can be bounded by $\frac{1}{2}c + \epsilon$. Using it to N_K, we have

$$\begin{aligned} \|N_K\| =& \|G_N(u_K - \alpha_K(u_K - u_{K-1})) - G_N(u_K) + \alpha_K(G_N(u_K) - G_N(u_{K-1}))\| \\ \le& (\frac{1}{2}c + \epsilon) 2|\alpha_K| \|u_K - u_{K-1}\|. \end{aligned} \tag{2.17}$$

Then, (2.5) and (2.17) imply

$$\|N_K\| \le \frac{c + 2\epsilon}{1 - c} \|F(u_K)\|. \tag{2.18}$$

We obtain from (2.7), (2.8), (2.12), (2.13) and (2.18) that

$$\|F(u_{K+1})\| \le (\bar{c} + \iota\epsilon)\,\|F(u_K)\|$$

with $\iota = \frac{\gamma(1+c)}{2(1-c)^2} + \frac{2}{1-c}$. Due to the arbitrariness of $\epsilon \in (0,1)$, the estimation in (2.4) holds for $k = K+1$ by reducing ϵ if necessary so that $\iota\epsilon \le \hat{c} - \bar{c}$. This completes the proof. ∎

The important technique in the proof of Theorem 2.1 is the decomposition method of $F(u_{k+1})$, especially the structure and analysis of N_k, which reduces the Lipschitz constant of the nonsmooth part of B_k by half.

In EDIIS(1), α_k is chosen as the minimizer of the optimization problem

$$\min \frac{1}{2}\left\|(1-\alpha)F_k + \alpha F_{k-1}\right\|^2, \quad \text{s.t.} \quad 0 \le \alpha \le 1.$$

This is a convex optimization problem and its solution α_k can be expressed by the formulation with the middle operator as

$$\alpha_k = \text{mid}\left\{0,\ \frac{F_k^{\mathrm{T}}(F_k - F_{k-1})}{\|F_{k-1} - F_k\|^2},\ 1\right\}^{1}.$$

Following the proof of Theorem 2.1, it is clear that (2.13) and (2.18) hold for $\alpha_k = 0$ and $\alpha_k = 1$, which are the points that we only need to check for the EDIIS(1) with respect to Anderson(1). Thus, we have the following statement.

Corollary 2.2. *Suppose that the assumptions of Theorem 2.1 hold. Then the sequence $\{u_k\}$ generated by EDIIS(1) satisfies (2.4).*

Since the results in Theorem 2.1 and Corollary 2.2 are local convergence results of Anderson(1) and EDIIS(1), the Lipschitz continuous differentiability of Q and H around u^* and $P_\Omega(Q(u^*))$ is enough to guarantee these statements.

3. Anderson acceleration method with smoothing approximation.

3.1. Smoothing approximation. In this subsection, we introduce some smoothing approximations of the nonsmooth contraction mapping G for finding its fixed point. For a function $\omega : \mathbb{R}^n \times (0,1] \to \mathbb{R}^n$, $\omega'(y,\mu)$ always denotes the derivative of ω with respect to y for fixed $\mu \in (0,1]$ in what follows. We define a smoothing function of $\max\{t,0\}$ at first.

Definition 3.1. *[5] We call $\psi : \mathbb{R} \times (0,1] \to \mathbb{R}$ a smoothing function of $\max\{t,0\}$ in $\mathbb{R}$, if $\psi(\cdot,\mu)$ is continuously differentiable in $\mathbb{R}$ for any fixed $\mu > 0$, and the following conditions hold.*

(i) *There is a $\kappa_\psi > 0$ such that for any $t \in \mathbb{R}$ and $\mu \in (0,1]$, $|\psi(t,\mu) - \max\{t,0\}| \le \kappa_\psi \mu$.*

(ii) *For any $t \in \mathbb{R}$, it holds $\{\lim_{s\to t\,\mu\downarrow 0} \psi'(s,\mu)\} \subseteq \partial(\max\{t,0\})$, where ∂ indicates the Clarke subdifferential [8].*

Definition 3.1-(i) implies that $\lim_{s\to t\,\mu\downarrow 0} \psi(s,\mu) = \max\{t,0\}$ and Definition 3.1-(ii) implies the gradient consistency. Smoothing functions for the max function have been studied in numerical methods for optimization and differential equations [5]. Four widely used smoothing functions of

---

[1] $\text{mid}(0,a,1) = \begin{cases} 0, & a < 0 \\ a, & a \in [0,1] \\ 1, & a > 1. \end{cases}$

$\max\{t,0\}$ are as follows:

$$\psi_1(t,\mu) = t + \mu\ln(1+e^{-\frac{t}{\mu}}), \qquad \psi_2(t,\mu) = \frac{1}{2}(t+\sqrt{t^2+4\mu^2}), \tag{3.1}$$
$$\psi_3(t,\mu) = \begin{cases} \max\{0,t\} & \text{if } |t| > \mu \\ \dfrac{(t+\mu)^2}{4\mu} & \text{if } |t| \le \mu, \end{cases} \qquad \psi_4(t,\mu) = \begin{cases} t + \dfrac{\mu}{2}e^{-\frac{t}{\mu}} & \text{if } t > 0 \\ \dfrac{\mu}{2}e^{\frac{t}{\mu}} & \text{if } t \le 0. \end{cases}$$

Let ψ be a smoothing function of $\max\{t,0\}$. For $v \in \mathbb{R}^l$, set

$$\Phi(v,\mu) = (\phi_1(v_1,\mu), \phi_2(v_2,\mu), \ldots, \phi_l(v_l,\mu))^T, \tag{3.2}$$

where

$$\phi_i(t,\mu) = \psi(\underline{w}_i - t,\mu) + t - \psi(t - \overline{w}_i,\mu), \ \ i = 1,2,\ldots,l. \tag{3.3}$$

It is clear that $\Phi(\cdot,\mu)$ is continuously differentiable on $\mathbb{R}^l$ for any fixed $\mu \in (0,1]$, and by (1.6), we have

$$\lim_{s\to t\,\mu\downarrow 0} \phi_i(s,\mu) = P_{[\underline{w}_i,\overline{w}_i]}(t) \quad \text{and} \quad |\phi_i(t,\mu) - P_{[\underline{w}_i,\overline{w}_i]}(t)| \le 2\kappa_\psi\mu, \quad \forall t \in \mathbb{R},\ \mu \in (0,1]. \tag{3.4}$$

Then, since $\underline{w}_i < \overline{w}_i$ for all $i = 1,2,\ldots,l$, we obtain

$$\begin{cases} \lim\limits_{s\to t,\mu\downarrow 0} \phi_i'(s,\mu) = \{0\} & \text{if } t < \underline{w}_i \text{ or } t > \overline{w}_i \\ \lim\limits_{s\to t,\mu\downarrow 0} \phi_i'(s,\mu) = \{1\} & \text{if } \underline{w}_i < t < \overline{w}_i \\ \left\{\lim\limits_{s\to t,\mu\downarrow 0} \phi_i'(s,\mu)\right\} \subseteq [0,1] & \text{if } t = \underline{w}_i \text{ or } t = \overline{w}_i. \end{cases} \tag{3.5}$$

PROPOSITION 3.2. *Let ψ be a smoothing function of $\max\{t,0\}$ with parameter κ_ψ in Definition 3.1-(i). Suppose Assumption 2.1 holds and $\Omega + \mathcal{B}(2\kappa_\psi\sqrt{l}, \boldsymbol{0}) \subseteq D_H$, then the function*

$$\mathcal{G}(u,\mu) = H(\Phi(Q(u),\mu)) \tag{3.6}$$

owns the following properties.

(i) *$\mathcal{G}(\cdot,\mu)$ is continuously differentiable on D for any fixed $\mu \in (0,1]$.*
(ii) *There is a $\kappa_G > 0$ such that for any $u \in D$ and $\mu \in (0,1]$, $\|\mathcal{G}(u,\mu) - G(u)\| \le \kappa_G\mu$.*
(iii) *For any $u \in D$, $\limsup_{z\to u,\mu\downarrow 0} \|\mathcal{G}'(z,\mu)\| \le c$.*
(iv) *For any $c_S \in (c,1)$, there exists $\hat{\mu} \in (0,1]$ such that for any fixed $\mu \in (0,\hat{\mu}]$, $\|\mathcal{G}'(u,\mu)\| \le c_S$, $\forall u \in D$, which implies that $\mathcal{G}(\cdot,\mu)$ is a contraction mapping on D with factor c_S, i.e.*

$$\|\mathcal{G}(u,\mu) - \mathcal{G}(v,\mu)\| \le c_S\|u - v\|, \quad \textit{for all } u,v \in D,\ \mu \in (0,\hat{\mu}]. \tag{3.7}$$

(v) *Let u_μ be a fixed point of $\mathcal{G}(\cdot,\mu)$, then $\|u_\mu - u^*\| \le \left(\frac{\kappa_G}{1-c}\right)\mu$, which further implies $\lim_{\mu\to 0} u_\mu = u^*$.*

Proof. From (3.4), we can claim that

$$\|\Phi(Q(u),\mu) - P_\Omega(Q(u))\| \le 2\kappa_\psi\sqrt{l}\mu. \tag{3.8}$$

Since $\Omega + \mathcal{B}(2\kappa_\psi\sqrt{l}, \mathbf{0}) \subseteq D_H$, by the continuous differentiability of Q, H and $\Phi(\cdot,\mu)$, (i) and (ii) hold with $\kappa_G = 2c_H\kappa_\psi\sqrt{l}$.

Note that

$$\mathcal{G}'(z,\mu) = H'(w)_{w=\Phi(Q(z),\mu)}\Phi'(v,\mu)_{v=Q(z)}Q'(z). \tag{3.9}$$

Recalling (3.5) and the definition of Φ in (3.2), we get

$$\|\Phi'(v,\mu)_{v=Q(z)}\| = \|\mathrm{diag}(\phi_i'(v_i,\mu)_{v_i=Q_i(z)})\| \le 1. \tag{3.10}$$

Then, the continuous differentiability of H and Q combining with the estimations in (3.8), (3.9) and (3.10) gives that

$$\limsup_{z\to u, \mu\downarrow 0} \|\mathcal{G}'(z,\mu)\| \le \|H'(w)_{w=P_\Omega(Q(u))}\|\|Q'(u)\| \le c_H c_Q = c,$$

which guarantees items (iii) and (iv).

Since u_μ and u^* are the fixed points of $\mathcal{G}(\cdot,\mu)$ and G on D, respectively, by (ii), we have

$$\|u_\mu - u^*\| = \|\mathcal{G}(u_\mu,\mu) - G(u^*)\| \le \|\mathcal{G}(u_\mu,\mu) - G(u_\mu)\| + \|G(u_\mu) - G(u^*)\| \le \kappa_G\mu + c\|u_\mu - u^*\|,$$

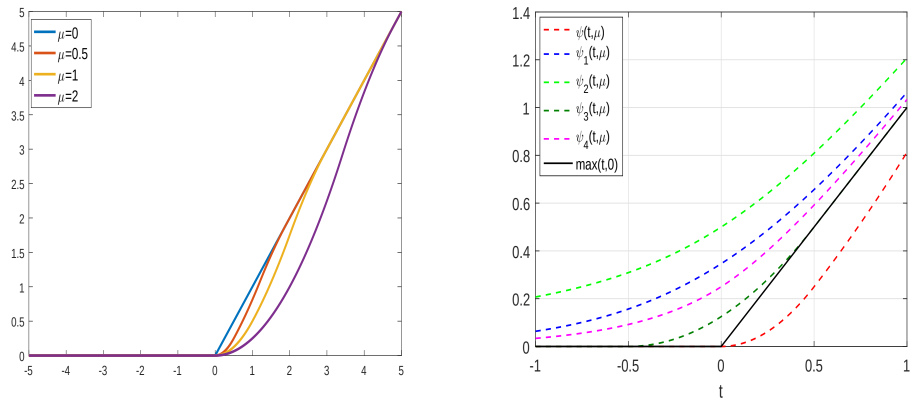
which gives the results in (v) by simple deduction. We complete the proof. ∎

If G satisfies Assumption 2.1, Proposition 3.2-(iv) says that its smoothing approximations in (3.6) also own the contractive property when μ is sufficiently small. Inspired by the proof of Proposition 3.2, if u^k is an approximate fixed point of $\mathcal{G}(u,\mu_k)$ with accuracy tolerance ϵ_k, i.e. $\|\mathcal{G}(u^k,\mu_k) - u^k\| \le \epsilon_k$, then we also have $\lim_{k\to\infty} u^k = u^*$, if $\lim_{k\to\infty}\mu_k = 0$ and $\lim_{k\to\infty}\epsilon_k = 0$. Moreover, the error estimation in Proposition 3.2-(v) holds always no matter $\mathcal{G}(\cdot,\mu)$ is contractive or not. Proposition 3.2-(v) also gives an upper bound of the error on the fixed point of G and its smoothing approximation, which is defined by the parameter κ_G coming from the structure of the smoothing approximation function and the contraction factor of G.

Remark 3.1. *Following the proof of Proposition 3.2, condition $\Omega + \mathcal{B}(2\kappa_\psi\sqrt{l}, \mathbf{0}) \subseteq D_H$ is only used to guarantee $\Phi(Q(u),\mu) \in D_H$ for all $u \in D$ and $\mu \in (0,1]$. So, the statements (i) and (ii) in Proposition 3.2 hold for any $\mu \in (0,\tilde{\mu}]$ with parameter $\tilde{\mu} \in (0,1]$ satisfying $\Omega + \mathcal{B}(2\kappa_\psi\sqrt{l}\tilde{\mu}, \mathbf{0}) \subseteq D_H$.*

3.2. A modified Anderson(m) algorithm. In this subsection, we will propose an Anderson acceleration algorithm for the nonsmooth fixed point problem (1.1) based on the smoothing approximation method. At first, we study the new smoothing function of $\max\{t,0\}$ as follows, which has more desirable properties for solving (1.1):

$$\psi(t,\mu) = \begin{cases} 0 & \text{if } t \le 0 \\ \dfrac{t^2}{2\mu} & \text{if } 0 < t \le \mu \\ \dfrac{1}{4}(t-\mu)^2 + t - \dfrac{1}{2}\mu & \text{if } \mu < t \le \mu + \sqrt{\mu} \\ -\dfrac{1}{4}(t-\mu-2\sqrt{\mu})^2 + t & \text{if } \mu + \sqrt{\mu} < t \le \mu + 2\sqrt{\mu} \\ t & \text{if } t > \mu + 2\sqrt{\mu}. \end{cases} \tag{3.11}$$



(a) (b)

Fig. 3.1: Smoothing functions of $\max\{t,0\}$: (a) $\psi(t,\mu)$ in (3.11) with different values of μ; (b) $\max\{t,0\}$, $\psi(t,\mu)$ in (3.11) and the four smoothing functions $\psi_i(t,\mu)$ in (3.1) with $\mu = 0.3$.

Fig. 3.1(a) shows the smoothing function $\psi(\cdot,\mu)$ in (3.11) with different values of μ, while Fig. 3.1(b) shows the relationships of $\max\{t,0\}$ and its smoothing functions defined in (3.1) and (3.11). Since ψ in (3.11) is a smoothing function of $\max\{t,0\}$ with Definition 3.1, the results in Proposition 3.2 also holds for $\mathcal{G}(u,\mu)$ defined in (3.6) with ψ in (3.11). In what follows, we will present some more desirable properties of ψ in (3.11).

Proposition 3.3. *Function $\psi(t,\mu)$ in (3.11) is continuously differentiable with respect to t for any fixed $\mu \in (0,1]$ and satisfies the following properties.*

(i) *$|\psi(t,\mu) - \max\{t,0\}| \le \frac{1}{2}\mu$, for any $t \in \mathbb{R}$ and $\mu \in (0,1]$.*

(ii) *$\psi(t,\mu) = \max\{t,0\}$ if $t \le 0$ or $t \ge \mu + 2\sqrt{\mu}$.*

(iii) *For any $\mu \in (0,1]$, $\psi'(t,\mu) = 0$ if $t \le 0$, $0 \le \psi'(t,\mu) \le 1 + \frac{1}{2}\sqrt{\mu}$ if $0 < t < \mu + 2\sqrt{\mu}$, and $\psi'(t,\mu) = 1$ if $t \ge \mu + 2\sqrt{\mu}$.*

Proof. By the definition of ψ in (3.11), we obtain

$$|\psi(t,\mu) - \max\{t,0\}| = \begin{cases} 0 & \text{if } t \le 0 \\ |t^2/2\mu - t| \le \mu/2 & \text{if } 0 < t \le \mu \\ |(t-\mu)^2/4 - \mu/2| \le \mu/2 & \text{if } \mu < t \le \mu + \sqrt{\mu} \\ (t - \mu - 2\sqrt{\mu})^2/4 \le \mu/4 & \text{if } \mu + \sqrt{\mu} < t \le \mu + 2\sqrt{\mu} \\ 0 & \text{if } t > \mu + 2\sqrt{\mu}, \end{cases}$$

which implies the statements in (i) and (ii).

By straightforward calculation, we can verify that $\psi(t,\mu)$ is continuously differentiable with respect to t for any fixed $\mu \in (0,1]$ and the estimation in (iii) holds. ∎

By Proposition 3.3-(ii), it holds that for any fixed $t \in \mathbb{R}$, there exists $\bar{\mu} > 0$ such that $\psi(t,\mu) = \max\{t,0\}$, $\forall \mu \in (0,\bar{\mu}]$, which is the main advantage of ψ in (3.11) compared with the other four smoothing functions of $\max\{t,0\}$ in (3.1). Following the proof of Proposition 3.3, we can further obtain the following properties of ϕ_i in (3.3) with ψ in (3.11).

PROPOSITION 3.4. *For any fixed* $\mu \in (0,1]$, *functions* $\phi_i(\cdot,\mu)$ *in* (3.3) *with* ψ *in* (3.11), $i = 1,2,\dots,l$, *are continuously differentiable and satisfy the following properties:*

(i) $|\phi_i(t,\mu) - P_{[\underline{w}_i,\overline{w}_i]}(t)| \le \frac{1}{2}\mu$, *for any* $t \in \mathbb{R}$;

(ii) $\phi_i(t,\mu) = P_{[\underline{w}_i,\overline{w}_i]}(t)$ *if* $t \le \underline{w}_i - \mu - 2\sqrt{\mu}$ *or* $\underline{w}_i \le t \le \overline{w}_i$ *or* $t \ge \overline{w}_i + \mu + 2\sqrt{\mu}$;

(iii) $|\phi_i'(t,\mu)| \le 1$, *for any* $t \in \mathbb{R}$.

Proof. By Proposition 3.3-(ii), we have

$$\begin{aligned}\psi(\underline{w}_i - t,\mu) &= \max\{\underline{w}_i - t, 0\} \quad \text{if } t \ge \underline{w}_i \text{ or } t \le \underline{w}_i - \mu - 2\sqrt{\mu},\\ \psi(t - \overline{w}_i,\mu) &= \max\{t - \overline{w}_i, 0\} \quad \text{if } t \le \overline{w}_i \text{ or } t \ge \overline{w}_i + \mu + 2\sqrt{\mu}.\end{aligned}$$

Then, for any $\mu \in (0,1]$ and $t \in \mathbb{R}$, at most one of $\psi(\underline{w}_i - t,\mu) = \max\{\underline{w}_i - t, 0\}$ and $\psi(t - \overline{w}_i,\mu) = \max\{t - \overline{w}_i, 0\}$ holds. Then, the results (i) and (ii) in Proposition 3.3 imply items (i) and (ii) in this proposition.

In what follows, we consider the estimation in item (iii). From (3.11), we have

$$|\phi_i'(t,\mu)| = |\psi'(\underline{w}_i - t,\mu) + 1 - \psi'(t - \overline{w}_i,\mu)|$$

$$= \begin{cases} 0 & \text{if } t \le \underline{w}_i - \mu - 2\sqrt{\mu}\\ |\underline{w}_i - t - \mu - 2\sqrt{\mu}|/2 \le \sqrt{\mu}/2 & \text{if } \underline{w}_i - \mu - 2\sqrt{\mu} \le t < \underline{w}_i - \mu - \sqrt{\mu}\\ |t + \mu - \underline{w}_i|/2 \le \sqrt{\mu}/2 & \text{if } \underline{w}_i - \mu - \sqrt{\mu} < t < \underline{w}_i - \mu\\ |t - \underline{w}_i + \mu|/\mu \le 1 & \text{if } \underline{w}_i - \mu \le t < \underline{w}_i\\ 1 & \text{if } \underline{w}_i \le t < \overline{w}_i\\ |\mu + \overline{w}_i - t|/\mu \le 1 & \text{if } \overline{w}_i \le t < \overline{w}_i + \mu\\ |t - \overline{w}_i - \mu|/2 \le \sqrt{\mu}/2 & \text{if } \overline{w}_i + \mu \le t < \overline{w}_i + \mu + \sqrt{\mu}\\ |t - \overline{w}_i - \mu - 2\sqrt{\mu}|/2 \le \sqrt{\mu}/2 & \text{if } \overline{w}_i + \mu + \sqrt{\mu} \le t < \overline{w}_i + \mu + 2\sqrt{\mu}\\ 0 & \text{if } t \ge \overline{w}_i + \mu + 2\sqrt{\mu}. \end{cases} \tag{3.12}$$

Thus, (iii) holds. ∎

In what follows, we will use the smoothing function of $\max\{t,0\}$ in (3.11) to construct a smoothing approximation of $P_\Omega(v)$ on $\mathbb{R}^l$, which is also with the formulation in (3.2). Then, we can give a smoothing approximation of G in (1.1) by the formulation of (3.6) with (3.11).

Set $\varpi_1 = \min\{3, \underline{w}_i - Q_i(u^*) : i \in \{i : Q_i(u^*) < \underline{w}_i\}\}$, $\varpi_2 = \min\{3, Q_i(u^*) - \overline{w}_i : i \in \{i : Q_i(u^*) > \overline{w}_i\}\}$, and by Assumption 2.1-(ii), denote $\eta \in (0,1]$ the parameter such that

$$\Omega + \mathcal{B}(\sqrt{l}\eta/2, \mathbf{0}) \subseteq D_H. \tag{3.13}$$

Then, we define parameter $\bar{\mu}$ by

$$\bar{\mu} = \min\{\eta, (\varpi_1/3)^2, (\varpi_2/3)^2\}. \tag{3.14}$$

THEOREM 3.5. *Suppose Assumption 2.1 holds. Besides the properties in Proposition 3.2, function* $\mathcal{G}(u,\mu)$ *in* (3.6) *with* ψ *defined in* (3.11) *owns the following properties.*

(i) *For any fixed* $\mu \in (0,\eta]$, $\mathcal{G}(\cdot,\mu)$ *is a contractive mapping on* D *with contraction factor no larger than* c *in Assumption 2.1.*

(ii) $\|\mathcal{G}(u,\mu) - G(u)\| \le \kappa\mu$ *for all* $u \in D$ *and* $\mu \in (0,\eta]$ *with* $\kappa = c_H\sqrt{l}/2$*;*
(iii) $\mathcal{G}(u^*,\mu) = G(u^*) = u^*$*,* $\forall \mu \in (0,\bar{\mu}]$*, where* $\bar{\mu}$ *is defined by* (3.14)*.*

Proof. By Proposition 3.4-(i), it holds

$$\|\Phi(Q(u),\mu) - P_\Omega(Q(u))\| \le \sqrt{l}\mu/2. \tag{3.15}$$

Then, $\Phi(Q(u),\mu) \in D_H$ for all $u \in D$ and $\mu \in (0,\eta]$ can be guaranteed by the condition $\Omega + \mathcal{B}(\sqrt{l}\eta/2, \mathbf{0}) \subseteq D_H$.

(i) Using the Lipschitz property of H and Q again, for any $u, v \in D$ and $\mu \in (0,\eta]$, we obtain

$$\begin{aligned}\|\mathcal{G}(u,\mu) - \mathcal{G}(v,\mu)\| \le& c_H\|\Phi(Q(u),\mu) - \Phi(Q(v),\mu)\| \\ \le& c_H\|Q(u) - Q(v)\| \le c_H c_Q\|u - v\| = c\|u - v\|,\end{aligned}$$

where the second inequality follows from Proposition 3.4-(iii). Thus, for any $\mu \in (0,\eta]$, $\mathcal{G}(u,\mu)$ is a contractive mapping on D with factor no larger than c.

(ii) By the Lipschitz property of H on D_H and $\Phi(Q(u),\mu) \in D_H$ for all $u \in D$ and $\mu \in (0,\eta]$, it holds

$$\begin{aligned}\|\mathcal{G}(u,\mu) - G(u)\| =& \|H(\Phi(Q(u),\mu)) - H(\max\{Q(u),0\})\| \\ \le& c_H\|\Phi(Q(u),\mu) - \max\{Q(u),0\}\| \le \kappa\mu,\end{aligned}$$

where the last inequality follows from (3.15) with $\kappa = \sqrt{l}c_H/2$.

(iii) Denote $I_1 = \{i : Q_i(u^*) < \underline{w}_i\}$, $I_2 = \{i : \underline{w}_i \le Q_i(u^*) \le \overline{w}_i\}$ and $I_3 = \{i : Q_i(u^*) > \overline{w}_i\}$. First, we can easily find that

$$\phi_i(Q_i(u^*),\mu) = Q_i(u^*) = P_{[\underline{w}_i,\overline{w}_i]}(Q_i(u^*)), \quad \forall i \in I_2.$$

Next, for $i \in I_1$, by the definition of ϖ_1 and $\bar{\mu} \le (\varpi_1/3)^2 \le 1$, we have

$$Q_i(u^*) \le \underline{w}_i - \varpi_1 = \underline{w}_i - 3\sqrt{\bar{\mu}} \le \underline{w}_i - \mu - 2\sqrt{\mu}, \quad \forall i \in I_1,\ 0 < \mu \le \bar{\mu},$$

by Proposition 3.4-(ii), which implies

$$\phi_i(Q_i(u^*),\mu) = P_{[\underline{w}_i,\overline{w}_i]}(Q_i(u^*)), \quad \forall i \in I_1,\ 0 < \mu \le \bar{\mu}. \tag{3.16}$$

Similarly, for $i \in I_3$, we obtain

$$Q_i(u^*) \ge \overline{w}_i + \mu + 2\sqrt{\mu}, \quad \forall i \in I_3,\ 0 < \mu \le \bar{\mu},$$

which gives (3.16) for $i \in I_3$. Thus, for any $\mu \in (0,\bar{\mu}]$, we have $\Phi(Q(u^*),\mu) = P_\Omega(Q(u^*))$ and thus $\mathcal{G}(u^*,\mu) = G(u^*) = u^*$. We complete the proof. ∎

Inspired by Theorem 3.5-(iii), when $\mu \le \bar{\mu}$ with $\bar{\mu}$ defined in (3.14), u^* is also the fixed point of $\mathcal{G}(u,\mu)$, and from Theorem 3.5-(i), we further have

$$(1-c)\|u - u^*\| \le \|\mathcal{F}(u,\mu)\| \le (1+c)\|u - u^*\|, \quad \forall u \in D,\ \mu \in (0,\bar{\mu}], \tag{3.17}$$

where $\mathcal{F}(u,\mu) = \mathcal{G}(u,\mu) - u$.

Remark 3.2. *Proposition 3.4-(ii) shows that* $\mathcal{G}(u,\mu) = G(u)$*, for any* $\mu \in (0,1]$ *and* $u \in D$ *satisfying* $Q(u) \in \Omega$*. Thus, if* u^* *is the fixed point of* $\mathcal{G}(\cdot,\mu)$ *for a given* $\mu \in (0,1]$ *and* $Q(u^*) \in \Omega$*, then we can justify that* u^* *is also the fixed point of* G*.*

Algorithm 3.1 s-Anderson(m)

Choose $u_0 \in D$ and a positive integer m.
Set parameters $\sigma_1, \sigma_2 \in (0,1)$, $\gamma > 0$ and a sufficiently small positive parameter $\epsilon < \gamma \|F(u_0)\|^2$.
Let $F_0 = G(u_0) - u_0$, $\mu_0 = \gamma\|F_0\|^2$, $\mathcal{F}_0 = \mathcal{G}(u_0, \mu_0) - u_0$ and $u_1 = \mathcal{G}(u_0, \mu_0)$.
for $k = 1, 2, ...$ **do**
set $F_k = G(u_k) - u_k$,
if $\|F_k\| \le \sigma_1 \|F_{k-1}\|$, then let

$$\mu_k = \mu_{k-1},$$

otherwise, let

$$\mu_k = \max\{\epsilon, \sigma_2 \mu_{k-1}\};$$

set $\mathcal{F}_k = \mathcal{G}(u_k, \mu_k) - u_k$;
choose $m_k = \min\{m, k\}$;
solve

$$\min \left\| \sum_{j=0}^{m_k} \alpha_j \mathcal{F}_{k-m_k+j} \right\| \qquad \text{s.t. } \sum_{j=0}^{m_k} \alpha_j = 1 \tag{3.18}$$

to find a solution $\{\alpha_j^k : j = 0, \dots, m_k\}$, and set

$$u_{k+1} = \sum_{j=0}^{m_k} \alpha_j^k \mathcal{G}(u_{k-m_k+j}, \mu_{k-m_k+j}); \tag{3.19}$$

end for

By Theorem 3.5, when we use (3.6) with (3.11) as the smoothing approximation of G, $\mathcal{G}(\cdot, \mu)$ is contractive and $u_\mu = u^*$ for $\mu \in (0, \bar{\mu}]$, where u_μ is the fixed point of $\mathcal{G}(\cdot, \mu)$. Then, we can apply Anderson(m) or EDIIS(m) to find a fixed point of G by using $\mathcal{G}(\cdot, \mu)$ in the algorithms. If u_0 is sufficiently close to u^*, then $\mu_0 := \gamma\|F(u_0)\|^2 < \bar{\mu}$. In such case, we can let $\mu_k := \mu_0$ for all k. However, u^* is unknown, and the value of $\bar{\mu}$ in (3.14) is often difficult to be evaluated in practice. Thus, we use an updating scheme on μ_k in Algorithm 3.1 to improve the ability and performance of the Anderson acceleration methods for nonsmooth fixed point problems. In s-Anderson(m), we replace $G(u)$ in Anderson(m) by $\mathcal{G}(u, \mu)$ and update μ step by step. The strategy for updating μ_k in Algorithm 3.1 is based on the reduction of the norms of the residual function at u_k and u_{k-1}. If $\|F_k\| \le \sigma_1\|F_{k-1}\|$, then it means that using μ_{k-1} can reduce the norm of the residual function at u_k sufficiently. Hence we let $\mu_k = \mu_{k-1}$ for the next iteration. Otherwise, we set $\mu_k = \max\{\epsilon, \sigma_2\mu_{k-1}\}$.

Same as the condition on the coefficients $\{\alpha_j^k : j = 1, \dots, m_k\}$ used in [6, 24], we need the following assumption on them in (3.18).

ASSUMPTION 3.1. *There exists an* $M_\alpha \ge 1$ *such that* $\sum_{j=0}^{m_k} |\alpha_j^k| \le M_\alpha$ *holds for all* $k \ge 1$.

Before proving the local r-linear convergence of s-Anderson(m), we need predefine some necessary parameters used in the forthcoming proof and give some preliminary analysis.

- a: Combining (3.9), (3.10) with the Lipschitz property of $Q'(u)$, $\text{diag}(\phi'(Q_i(u), \mu))$ and $H'(\Phi(Q(u), \mu))$ on D, there exists a constant $a > 0$ such that $\mathcal{G}'(u, \mu)$ is Lipschitz continuous on $\mathcal{B}(\delta, u^*)$ with constant a. This means,

$$\mathcal{G}(u,\mu) = \mathcal{G}(u^*,\mu) + \mathcal{G}'(u^*,\mu)(u-u^*) + \Delta_u, \quad \forall u \in \mathcal{B}(\delta, u^*),\ \mu \in [\epsilon, \eta], \tag{3.20}$$

where $\|\Delta_u\| \le \frac{1}{2}a\|u-u^*\|^2$.

- δ_1: Since $\mathcal{G}(u,\mu)$ is Lipschitz continuous, from Theorem 2.2 in [6], there exists $\delta_1 \in (0,\delta]$ such that if $\|u_0 - u^*\| \le \delta_1$, we have the r-linear convergence of Anderson(m) on solving $\mathcal{F}(u,\hat{\mu}) := \mathcal{G}(u,\hat{\mu}) - u = 0$ with any $\hat{\mu} \in [\epsilon, \bar{\mu}]$, that is

$$\limsup_{k\to\infty} \left(\frac{\|\mathcal{F}(u_k,\hat{\mu})\|}{\|\mathcal{F}(u_0,\hat{\mu})\|} \right)^{1/k} \le c, \tag{3.21}$$

where c is a contraction factor of $\mathcal{G}(u,\hat{\mu})$ on D by Theorem 3.5-(i).

- δ_0: Let

$$\delta_0 := \min\{\delta_1, \frac{\sqrt{\bar{\mu}}}{\sqrt{\gamma}(1+c)}, \frac{\sqrt{\eta}}{\sqrt{\gamma}(1+c)}, \frac{(1-c)\delta_1}{M_\alpha(1+c)}, \frac{1-c}{\varpi}\}, \tag{3.22}$$

where $\bar{\mu}$ is defined in (3.14) and $\varpi = \frac{a(M_\alpha^2+M_\alpha)(1+c)+2M_\alpha\sqrt{l}c_H\gamma(1+c)^2(1-c)}{2(1-c)^2}$.

Lemma 3.6. *If $\|u_0 - u^*\| \le \delta_0$, then for the sequences $\{\mu_k\}$, $\{u^k\}$ and $\{\mathcal{F}_k\}$ generated by s-Anderson(m) in Algorithm 3.1, it holds that*

$$\mu_k \le \bar{\mu}, \quad \Omega + \mathcal{B}(\sqrt{l}\mu_k/2, \boldsymbol{0}) \subseteq D_H, \quad u_k \in \mathcal{B}(\delta_1, u^*) \quad \text{and} \quad \|\mathcal{F}_k\| \le \|\mathcal{F}_0\|. \tag{3.23}$$

Proof. Since

$$\gamma\|F(u_0)\|^2 \le \gamma(1+c)^2\|u_0-u^*\|^2 \le \min\{\bar{\mu},\eta\},$$

then $\mu_k \le \min\{\bar{\mu},\eta\}$ by the updating method of μ_k in s-Anderson(m) for $k \ge 0$. From (3.13), we find that the first two relations in (3.23) hold.

Then, by Theorem 3.5-(i) and (iii), we have

$$\mathcal{G}(u^*,\mu_k) = G(u^*) = u^* \text{ and } \|\mathcal{G}(u,\mu_k) - \mathcal{G}(v,\mu_k)\| \le c\|u-v\|, \quad \forall k \ge 0,\ u,v \in \mathcal{B}(\delta_1,u^*). \tag{3.24}$$

We next prove the last two statements of (3.23) by induction, where we see that they are true for $k=0$ and we suppose both of them hold for $0 \le k \le K$.

Owning to (3.24), we have

$$\begin{aligned}
\|u_{K+1} - u^*\| &= \left\| \sum_{j=0}^{m_K} \alpha_j^K \mathcal{G}(u_{K-m_K+j}, \mu_{K-m_K+j}) - \sum_{j=0}^{m_K} \alpha_j^K \mathcal{G}(u^*, \mu_{K-m_K+j}) \right\| \\
&\le M_\alpha c \max_j \|u_{K-m_K+j} - u^*\| \le \frac{M_\alpha c}{1-c} \max_j \|\mathcal{F}_{K-m_K+j}\| \\
&\le \frac{M_\alpha c}{1-c}\|\mathcal{F}_0\| \le \frac{M_\alpha c(1+c)}{1-c}\|u_0-u^*\|,
\end{aligned}$$

which gives $u_{K+1} \in \mathcal{B}(\delta_1, u^*)$ by the condition of δ_0. Then, the third result in (3.23) holds for $k = K+1$.

Similarly, $\sum_{j=0}^{m_K} \alpha_j^K u_{K-m_K+j} \in \mathcal{B}(\delta_1, u^*) \subseteq D$. Formulas (3.19) and (3.24) imply

$$\|\mathcal{F}_{K+1}\| = \|\mathcal{G}(u_{K+1}, \mu_{K+1}) - u_{K+1}\|$$
$$\leq c\|u_{K+1} - \sum_{j=0}^{m_K} \alpha_j^K u_{K-m_K+j}\| + \|\mathcal{G}(\sum_{j=0}^{m_K} \alpha_j^K u_{K-m_K+j}, \mu_{K+1}) - \sum_{j=0}^{m_K} \alpha_j^K \mathcal{G}(u_{K-m_K+j}, \mu_{K-m_K+j})\|$$
$$\leq c\|\mathcal{F}_K\| + A_K + B_K, \tag{3.25}$$

where

$$A_K = \left\| \mathcal{G}(\sum_{j=0}^{m_K} \alpha_j^K u_{K-m_K+j}, \mu_{K+1}) - \sum_{j=0}^{m_K} \alpha_j^K \mathcal{G}(u_{K-m_K+j}, \mu_{K+1}) \right\|,$$

$$B_K = \left\| \sum_{j=0}^{m_K} \alpha_j^K \mathcal{G}(u_{K-m_K+j}, \mu_{K+1}) - \sum_{j=0}^{m_K} \alpha_j^K \mathcal{G}(u_{K-m_K+j}, \mu_{K-m_K+j}) \right\|.$$

Then, by (3.20), we estimate $\|A_K\|$ by the same way as in [6, 24] to get

$$\begin{aligned} \|A_K\| = & \left\| \Delta_{\sum_{j=0}^{m_K} \alpha_j^K u_{K-m_K+j}} - \sum_{j=0}^{m_K} \alpha_j^K \Delta_{u_{K-m_K+j}} \right\| \\ \leq & \frac{a(M_\alpha^2 + M_\alpha)}{2} \max_j \|u_{K-m_K+j} - u^*\|^2 \\ \leq & \frac{a(M_\alpha^2 + M_\alpha)}{2(1-c)^2} \max_j \|\mathcal{F}_{K-m_K+j}\|^2 \\ \leq & \frac{a(M_\alpha^2 + M_\alpha)(1+c)\|u_0 - u^*\|}{2(1-c)^2} \|\mathcal{F}_0\|. \end{aligned} \tag{3.26}$$

To evaluate $\|B_K\|$, by Theorem 3.5-(ii), (2.3), (3.17) and (3.24), we have

$$\|B_K\| \leq M_\alpha \kappa(\mu_{K-m_K} + \mu_{K+1}) \leq 2M_\alpha \kappa \mu_0 = 2M_\alpha \kappa \gamma \|F(u_0)\|^2 \leq \frac{2M_\alpha \kappa \gamma (1+c)^2 \|u_0 - u^*\|}{1-c} \|\mathcal{F}_0\|. \tag{3.27}$$

Together (3.25), (3.26), (3.27) with the assumption of (3.23) for $k = K$, gives

$$\|\mathcal{F}_{K+1}\| \leq (c + \varpi \|u_0 - u^*\|) \|\mathcal{F}_0\|.$$

Then the fourth relation in (3.23) holds for $k = K+1$ by δ_0 satisfying $c + \varpi \delta_0 \leq 1$. We complete the proof for (3.23). ∎

THEOREM 3.7. *Suppose Assumption 2.1 and Assumption 3.1 hold. If u_0 is sufficiently close to u^*, then the sequence $\{u^k\}$ generated by s-Anderson(m) in Algorithm 3.1 converges to the solution of (1.1) with the r-linear convergence rates of*

$$\limsup_{k\to\infty} \left(\frac{\|u_k - u^*\|}{\|u_0 - u^*\|} \right)^{1/k} \leq c \quad \text{and} \quad \limsup_{k\to\infty} \left(\frac{\|F(u_k)\|}{\|F(u_0)\|} \right)^{1/k} \leq c. \tag{3.28}$$

Proof. Let $\|u_0 - u^*\| \le \delta_0$ with δ_0 in (3.22). Then, $\mu_0 \le \bar{\mu}$. By the updating method of μ_k, there exist K and $\hat{\mu} \in [\epsilon, \bar{\mu}]$ such that $\mu_k = \hat{\mu}$, for all $k \ge K$.

By Lemma 3.6, as $\|u_0 - u^*\| \le \delta_0$, we have $\|u_K - u^*\| \le \delta_1$. Then, by (3.21), we have

$$\limsup_{k\to\infty} \left(\frac{\|\mathcal{F}(u_k, \mu_k)\|}{\|\mathcal{F}(u_K, \mu_K)\|} \right)^{1/(k-K)} \le c,$$

which implies

$$\limsup_{k\to\infty} \left(\frac{\|\mathcal{F}(u_k, \mu_k)\|}{\|\mathcal{F}(u_0, \mu_0)\|} \right)^{1/k} \le c. \tag{3.29}$$

From (3.17), we obtain

$$\frac{\|\mathcal{F}(u_k, \mu_k)\|}{\|\mathcal{F}(u_0, \mu_0)\|} \ge \left(\frac{1-c}{1+c} \right) \frac{\|u_k - u^*\|}{\|u_0 - u^*\|},$$

which combines with (3.29) and $\limsup_{k\to\infty} \left(\frac{1-c}{1+c} \right)^{1/k} = 1$ gives the first estimation in (3.28). In light of (2.3) and the first relation in (3.28), we further obtain the second result in (3.28). ∎

From the updating method of $\{\mu_k\}$ in s-Anderson(m), it is a case that $\lim_{k\to\infty} \mu_k > \epsilon$, which means that there exists K such that $\|F_k\| \le \sigma_1 \|F_{k-1}\|$, $\forall k \ge K$. Combining this with Theorem 3.7, we note that if $\lim_{k\to\infty} \mu_k > \epsilon$, then s-Anderson(m) not only owns the r-linear convergence in (3.29), but also has the q-linear convergence on residual $\|F(u_k)\|$ with factor σ_1. Moreover, following the statements in Theorem 3.7, even if we have no knowledge on $\bar{\mu}$ and η, the local convergence properties of s-Anderson(m) in Theorem 3.7 are always valid with any $\sigma_1, \sigma_2 \in (0, 1)$ by setting ϵ sufficiently small. In particular, if μ_0 is sufficiently small such that μ_k is unchanged in s-Anderson(m), then s-Anderson(m) is just Anderson(m) on $\mathcal{G}(u, \mu_0)$. A simple consideration is that the results in Theorem 3.7 also hold if we let $\mu_k := \epsilon$ with ϵ being sufficiently small. Similar results in Theorem 3.7 also hold for the EDIIS(m) with the same smoothing approach.

Remark 3.3. *According to Rademacher's theorem, a locally Lipschitz continuous function G is differentiable almost everywhere. If ψ is a smoothing function of $\max\{t, 0\}$, Proposition 3.2 says that the contraction factor of $\mathcal{G}(\cdot, \mu)$ on D can be sufficiently close to the contraction factor of G as μ is sufficiently small. Theorem 3.5 gives an upper bound of the contraction factor of $\mathcal{G}(\cdot, \mu)$ on D with ψ defined in (3.11). By the structure of ψ in (3.11), if G is not continuously differentiable at u^*, which means that there is $i \in \{1, \dots, l\}$ such that $Q_i(u^*) = \underline{w}_i$ or $\overline{w}_i$, then the contraction factor of $\mathcal{G}(\cdot, \mu)$ with (3.11) can be strictly smaller than the contraction factor of G around u^* as μ is smaller than a threshold.*

For example, if $G(u) = (\max\{u_1/2, 0\}, 1 - u_2/4)^{\mathrm{T}}$, the exact contraction factor of G around its fixed point $u^* = (0, 4/5)^{\mathrm{T}}$ is $1/2$. Let $\mathcal{G}(u, \mu) = (\psi(u_1/2, \mu), 1 - u_2/4)^{\mathrm{T}}$ with the definition of ψ in (3.11). For any given $\mu \in [\epsilon, 1]$, we note that

$$\|\mathcal{G}'(u, \mu)\| \le \max\{|u_1|/(2\mu), 1/4\}, \quad \forall u \in \mathcal{B}(\delta, u^*)$$

with $\delta \le \epsilon/2$, which implies that the contraction factor of $\mathcal{G}(\cdot, \mu)$ is no larger than $1/4$ when $\mu \in [\epsilon, 1]$. These results combining the analysis in Theorem 3.7 show that as u_0 is sufficiently close to u^*, s-Anderson(m) is r-linearly convergent to the fixed point of G with factor no larger than $1/4$, which is strictly smaller than the contraction factor of G around u^*. And the contraction factor of $\mathcal{G}(u, \mu)$ on $\mathcal{B}(\delta, u^*)$ is decreasing as μ is increasing in $[\epsilon, 1]$.

4. Numerical applications and examples. In this section, we illustrate our new convergence results of Anderson acceleration for nonsmooth fixed point problem (1.1) by three applications. All the numerical experiments are performed in MATLAB 2016a on a Lenovo PC547 (3.00GHz, 2.00GB of RAM). When $m \geq 2$, proceeding as in [11, 25], we write the problem in (1.3) by the following equivalent form

$$\theta^k \in \arg\min_{\theta \in \mathbb{R}^{m_k}} \left\| F_k - \sum\nolimits_{j=0}^{m_k-1} \theta_j (F_{k-m_k+j+1} - F_{k-m_k+j}) \right\| \tag{4.1}$$

and then

$$u_{k+1} = G(u_k) - \sum\nolimits_{j=0}^{m_k-1} \theta_j^k (G(u_{k-m_k+j+1}) - G(u_{k-m_k+j}))),$$

in terms of the original iterations α_j^k in (1.3), where $\alpha_0^k = \theta_0^k$, $\alpha_j^k = \theta_j^k - \theta_{j-1}^k$ for $1 \leq j \leq m_k - 1$ and $\alpha_{m_k}^k = 1 - \theta_{m_k-1}^k$. To solve (4.1), we consult the method based on the pseudoinverse introduced in [11], and it has been shown that the deteriorating condition of the least-squares matrix does not necessarily interfere with convergence [24]. This method is also used to find the α_j^k in s-Anderson(m) in Algorithm 3.1. For s-Anderson(m), we always set $\epsilon = 10^{-10}$, $\gamma = 1/n$ and $\sigma_1 = \sigma_2 = 0.6$ for comparison. And we stop Anderson(m) in Algorithm 1.1 and s-Anderson(m) in Algorithm 3.1 when

$$\frac{\|F(u_k)\|}{\|F(u_0)\|} \leq 10^{-14} \quad \text{or} \quad k \geq 7000. \tag{4.2}$$

It should be noticed that the stopped criterion for s-Anderson(m) also uses the value of $\|F(u_k)\|/\|F(u_0)\|$ not the residual on smoothing approximation $\mathcal{F}(u, \mu)$. From these numerical results in Examples 4.1-4.3, we have the following observations.

(i) Both Anderson(m) and s-Anderson(m) can be used to solve the considered problems, in which the contraction mappings G are nonsmooth at the fixed points. Though the theoretical results of them are built up for local convergence, it is satisfactory that all the numerical experiments in this section are convergent with random initial points.

(ii) For both Anderson(m) and s-Anderson(m), as presented in the experiments, the best choice of m is problem dependent.

(iii) s-Anderson(m) performs better than Anderson(m) for most cases, and the local convergence of $\|F(u_k)\|/\|F(u_0)\|$ by s-Anderson(m) is also faster. Since the mapping $\mathcal{G}(u, \mu)$ used in s-Anderson(m) only has small difference with $G(u)$ in Anderson(m), the generated u_k in the former iterations cannot bring obvious differences on $\|F_k\|/\|F_0\|$ when $\|F_k\|$ is relatively large. However, after certain iterations, $\|F_k\|$ is reduced significantly and the advantages of s-Anderson(m) appears clearly. So it is reasonable that s-Anderson(m) outperforms Anderson(m) when the accuracy is high.

(iv) The superiorities of s-Anderson(m) over Anderson(m) become more and more obvious as the number of elements in $\{i : Q_i(u^*) = \underline{w}_i \text{ or } \overline{w}_i\}$ increases.

4.1. Minimax optimization problem. Constrained minimax optimization problem is often modeled by

$$\min_{x \in \mathcal{X}} \max_{y \in \mathcal{Y}} f(x, y), \tag{4.3}$$

where $f : \mathcal{X} \times \mathcal{Y} \to \mathbb{R}$ is a convex-concave function over closed, convex sets $\mathcal{X} \subseteq \mathbb{R}^{n_1}$ and $\mathcal{Y} \subseteq \mathbb{R}^{n_2}$. Such models are widely used in game theory, machine learning and parallel computing. Due to the convexity and concavity of f with respect to x and y, respectively, $((x^*)^{\mathrm{T}}, (y^*)^{\mathrm{T}})^{\mathrm{T}}$ is a saddle point of (4.3), if and only if it satisfies

$$\begin{cases} x^* = P_{\mathcal{X}}(x^* - \alpha \nabla_x f(x^*, y^*)) \\ y^* = P_{\mathcal{Y}}(y^* + \beta \nabla_y f(x^*, y^*)) \end{cases} \tag{4.4}$$

with $\alpha, \beta > 0$. Denote

$$u = \begin{pmatrix} x \\ y \end{pmatrix}, \ \Lambda = \begin{pmatrix} \alpha I_{n_1} & 0 \\ 0 & \beta I_{n_2} \end{pmatrix}, \ L(u) := L(x,y) = \begin{pmatrix} \nabla_x f(x,y) \\ -\nabla_y f(x,y) \end{pmatrix}, \ \Omega = \mathcal{X} \times \mathcal{Y}.$$

Then, (4.4) is expressed by $u^* = P_\Omega(u^* - \Lambda L(u^*))$, which is reduced to a fixed point problem of G with

$$G(u) := P_\Omega(u - \Lambda L(u)). \tag{4.5}$$

The mapping in (4.5) can be formulated by (1.1) with $Q(u) = u - \Lambda L(u)$ and $H(v) = v$.

ASSUMPTION 4.1. *The mapping L is strongly monotone and Lipschitz continuous, i.e. there exist positive parameters τ_L and c_L such that for all $u, \tilde{u} \in \Omega$, it holds*

$$(L(u) - L(\tilde{u}))^{\mathrm{T}}(u - \tilde{u}) \geq \tau_L \|u - \tilde{u}\|^2,$$
$$\|L(u) - L(\tilde{u})\| \leq c_L \|u - \tilde{u}\|.$$

For $u, \tilde{u} \in \Omega$, by the Lipschitz property of P_Ω and Assumption 4.1, when $\alpha = \beta$, we obtain

$$\begin{aligned} & \|P_\Omega(u - \alpha L(u)) - P_\Omega(\tilde{u} - \alpha L(\tilde{u}))\|^2 \\ \leq & \|u - \alpha L(u) - \tilde{u} + \alpha L(\tilde{u})\|^2 \\ = & \|u - \tilde{u}\|^2 + \alpha^2 \|L(u) - L(\tilde{u})\|^2 - 2\alpha (u - \tilde{u})^T (L(u) - L(\tilde{u})) \\ \leq & \left(1 + \alpha^2 c_L^2 - 2\alpha \tau_L\right) \|u - \tilde{u}\|^2. \end{aligned}$$

It is easy to verify that $1 + \alpha^2 c_L^2 - 2\alpha\tau_L \in (0,1)$, if $\alpha \in \left(0, 2\tau_L/c_L^2\right)$. Hence under Assumption 4.1, if $\alpha = \beta \in \left(0, 2\tau_L/c_L^2\right)$, then G in (4.5) is a contractive mapping with factor $c = \sqrt{1 + \alpha^2 c_L^2 - 2\alpha\tau_L}$ and the conclusions in Theorem 3.7 hold for G in (4.5), which prompts us to find the fixed point of G by using s-Anderson(m) with the smoothing approximation of G defined in (3.6). To show the effectiveness of the corresponding theoretical results and the effect of s-Anderson(m) on solving problem (4.3), we conduct the numerical experiment on a special case of (4.3), which comes from the two-payers Nash game problems.

Example 4.1. *Consider*

$$\min_{x \in \mathbb{R}_+^{n_1}} \max_{y \in \mathbb{R}_+^{n_2}} f(x,y) := \frac{1}{2} x^{\mathrm{T}} A x + x^{\mathrm{T}} B y - \frac{1}{2} y^{\mathrm{T}} C y + a^{\mathrm{T}} x - b^{\mathrm{T}} y, \tag{4.6}$$

where $A \in \mathbb{R}^{n_1 \times n_1}$ and $C \in \mathbb{R}^{n_2 \times n_2}$ are symmetric positive definite matrices, $B \in \mathbb{R}^{n_1 \times n_2}$, $a \in \mathbb{R}^{n_1}$ and $b \in \mathbb{R}^{n_2}$ are random matrix and vectors. Denote $\lambda_{\min}(A)$ and $\lambda_{\min}(C)$ the minimal eigenvalues of A and C, respectively. Let $\Omega = \mathbb{R}_+^{n_1+n_2}$ and $L(u) = Mu + d$ with $u = (x^{\mathrm{T}}, y^{\mathrm{T}})^{\mathrm{T}}$, $M = \begin{pmatrix} A & B \\ -B^{\mathrm{T}} & C \end{pmatrix}$ and $d = \begin{pmatrix} a \\ b \end{pmatrix}$, which satisfies Assumption 4.1 with

$$\tau_L = \min\{\lambda_{\min}(A), \lambda_{\min}(C)\} \quad \text{and} \quad c_L = \|M\|. \tag{4.7}$$

Based on the above analysis, the solution of (4.6) can be transformed to the fixed point of (4.5), and when we choose

$$\alpha = \beta = \tau_L / c_L^2, \tag{4.8}$$

G in (4.5) is a contractive mapping with factor $c = \|I_{n_1+n_2} - \alpha M\|$. For given positive integers $n_1 = 1000$, $n_2 = 500$ and $s_1 = 0.3$, we generate matrices A, C and B as follows:

$$\mathtt{A1} = 1 + \mathtt{a} * \mathtt{rand}(\mathtt{n_1}, 1); \mathtt{U1} = \mathtt{orth}(\mathtt{rand}(\mathtt{n_1}, \mathtt{n_1})); \mathtt{A} = \mathtt{U1}' * \mathtt{diag}(\mathtt{A1}) * \mathtt{U1};$$
$$\mathtt{C1} = 2 + \mathtt{b} * \mathtt{rand}(\mathtt{n_2}, 1); \mathtt{U2} = \mathtt{orth}(\mathtt{rand}(\mathtt{n_2}, \mathtt{n_2})); \mathtt{C} = \mathtt{U2}' * \mathtt{diag}(\mathtt{C1}) * \mathtt{U2};$$
$$\mathtt{B} = \mathtt{sprand}(\mathtt{n_1}, \mathtt{n_2}, \mathtt{s_1}); \mathtt{B} = \mathtt{full}(\mathtt{B})/\mathtt{norm}(\mathtt{B});$$

Then, we set $n = n_1 + n_2$, α and β be defined by (4.8) with the parameters in (4.7). It is clear that G in (4.5) is nonsmooth at u^* if there exists i such that $(Mu^* + d)_i = 0$ and $u_i^* = 0$. So, for given $s_2 = 0.5$, we generate the fixed point u^* (`sol` in the code) with $n \times s_2$ elements of 0 and vector $d \in \mathbb{R}^n$ such that the corresponding elements of $Mu^* + d$ are also 0 by the following codes:

$$\mathtt{index} = \mathtt{randperm}(\mathtt{n}); \mathtt{index1} = \mathtt{index}(1 : \mathtt{s_2} * \mathtt{n});$$
$$\mathtt{sol} = 0.1 + 0.9 * \mathtt{rand}(\mathtt{n}, 1); \mathtt{sol}(\mathtt{index1}) = 0; \mathtt{M} = [\mathtt{A} \quad \mathtt{B}; -\mathtt{B}' \quad \mathtt{C}]; \mathtt{d} = -\mathtt{M} * \mathtt{sol};$$

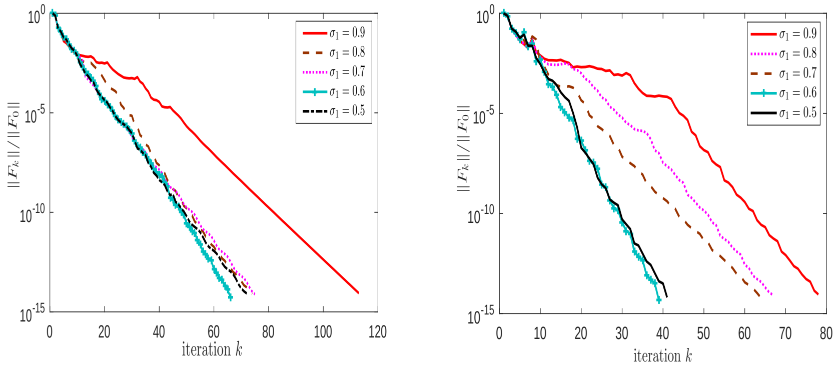
Let $u_0 = \mathrm{zeros}(n, 1)$. For different values of a and b, which influence the contractive factor of G in (4.5), the number of iterations of Anderson(m) and s-Anderson(m) to find u_k satisfying (4.2) are shown in Table 4.1, where the values are the mean values of 50 random experiments. From Table 4.1, we see that though the contractive factors of G are all very close to 1, both Anderson(m) and s-Anderson(m) work well, and s-Anderson(m) performs better for most cases. Throughout the whole table, the smallest iterations for all cases are presented by s-Anderson(m) with $m = 3$ or $m = 5$. Fig. 4.1 plots the convergence behaviors of s-Anderson(1) and s-Anderson(3) with some different values of $\sigma_1 = \sigma_2$, where the best is located at $\sigma_1 = \sigma_2 = 0.6$. This is an interesting thing that we can let the value of ϵ be sufficiently small to guarantee the efficiency of s-Anderson(m), and control the values of σ_1 and σ_2 to improve its convergence behaviours. How to choose better parameters is an interesting topic for further study.

Parameters		Anderson(m)/s-Anderson(m)					
a, b	c	$m = 0$	$m = 1$	$m = 2$	$m = 3$	$m = 5$	$m = 10$
0, 0	0.835	150/**148**	72/**65**	60/**57**	58/**48**	62/<u>**43**</u>	78/**54**
2, 1	0.893	230/**218**	**68**/85	71/71	71/**65**	80/<u>**58**</u>	80/**66**
1, 1	0.895	246/**236**	**74**/92	76/**73**	77/<u>**65**</u>	84/**71**	94/**80**
1, 2	0.943	465/**446**	147/**129**	113/**105**	117/<u>**102**</u>	125/**107**	159/**126**
3, 1	0.941	407/**379**	114/**104**	105/**96**	108/<u>**93**</u>	111/**94**	122/**102**
3, 3	0.961	609/**565**	194/194	136/**123**	144/**123**	149/<u>**122**</u>	174/**139**

Table 4.1: Numerical results of Anderson(m) and s-Anderson(m) for Example 4.1

4.2. Complementarity problem. Given a continuously differentiable function $f : \mathbb{R}^n \to \mathbb{R}^n$, the complementarity problem is to find v such that

$$v \geq 0, \quad f(v) \geq 0, \quad v^{\mathrm{T}} f(v) = 0.$$



(a) s-Anderson(1) (b) s-Anderson(3)

Fig. 4.1: Convergence of $\|F_k\|/\|F_0\|$ by s-Anderson(1) and s-Anderson(3) with different values of $\sigma_1 = \sigma_2$ for Example 4.1

This problem is denoted as CP(f), which is equivalent to $v = \max\{v - f(v), 0\}$. Let $Q(v) = v - f(v)$. If $\|I - f'(v)\| \le c < 1$, then $G(v) = \max\{Q(v), 0\}$ is a contraction mapping with factor c.

If $f(v) = Mv + q$ with $M \in \mathbb{R}^{n\times n}$ and $q \in \mathbb{R}^n$, the CP(f) is the linear complementarity problem, denoted as LCP(q, M). Suppose $M = (m_{ij})_{n\times n}$ is strictly diagonally dominate with positive diagonal elements in the following sense,

$$\sum_{i=1, i\neq j}^{n} |m_{ij}| < m_{ii} \quad \text{and} \quad \sum_{j=1, j\neq i}^{n} |m_{ij}| < m_{ii}.$$

Let Λ =diag(m_{ii}). Then LCP(q, M) is equivalent to

$$\Lambda v \ge 0, \quad M\Lambda^{-1}\Lambda v + q \ge 0, \quad (\Lambda v)^{\mathrm{T}}(M\Lambda^{-1}\Lambda v + q) = 0$$

and can be solved via LCP$(q, M\Lambda^{-1})$. Moreover, from

$$\begin{aligned} \|I - M\Lambda^{-1}\| &\le \sqrt{\|(\Lambda - M)\Lambda^{-1}\|_1 \|(\Lambda - M)\Lambda^{-1}\|_\infty} \\ &= \sqrt{\max \frac{1}{m_{ii}} \sum_{j=1, j\neq i}^{n} |m_{ij}|}\sqrt{\max \frac{1}{m_{ii}} \sum_{i=1, i\neq j}^{n} |m_{ij}|} =: c < 1, \end{aligned}$$

$G(u) = \max((I - M\Lambda^{-1})u - q, 0)$ is a contraction mapping. Let $Q(u) = (I - M\Lambda^{-1})u - q$. We define a smoothing approximation of G by (3.6), which is also a contraction mapping with factor c and satisfies the conditions in Assumption 2.1. Thus, if u^* is the fixed point of the above defined G, then $v^* = \Lambda^{-1}u^*$ is the solution of LCP(q, M).

Example 4.2. *Pricing American options in a partial differential equation framework with finite difference methods or finite element methods lead to a linear complementarity problem*

$$v - a \ge 0, \quad Mv - b \ge 0, \quad (v - a)^{\mathrm{T}}(Mv - b) = 0, \tag{4.9}$$

where v is the value of an American option, a is from a given payoff function, b is from an initial guess of the value and its changing rate, and M is from differential operators [21].

Let $u = v - a$ and $q = Ma - b$, then (4.9) is the standard form of LCP(q, M). We set

$$M = \begin{pmatrix} 2+\gamma_1 h^2 & -1+0.5h\tau_1 & & & \\ -1-0.5h\tau_2 & 2+\gamma_2 h^2 & -1+0.5h\tau_2 & & \\ & \ddots & \ddots & \ddots & \\ & & & 2+\gamma_{n-1}h^2 & -1+0.5h\tau_{n-1} \\ & & & -1-0.5h\tau_n & 2+\gamma_n h^2 \end{pmatrix}.$$

Here, M is the matrix from the centered difference formulate for

$$-\frac{\partial^2 V}{\partial x^2}(t,x) + \tau(t,x)\frac{\partial V}{\partial x}(t,x) + \gamma(t,x)V(t,x)$$

at a fixed time t, where $h = 1/(n+1)$ is the mesh size of discretization, and $\gamma(t,x) > 0$ and $\tau(t,x)$ are given functions. If $|\tau_i| = |\tau_{i+1}| < 2(n+1)$, $i = 1, 2, \ldots, n-1$, the matrix M is a strictly diagonal dominate matrix, and thus a P-matrix. Then, the LCP(q, M) has a unique solution u^* for any $q \in \mathbb{R}^n$, which is also the fixed point of the nonsmooth fixed point problem

$$u = G(u) = \max\{(I - \eta M)u - \eta q, 0\}, \tag{4.10}$$

with $\eta = \frac{1}{2+\gamma h^2}$. Here function G in (4.10) is a contraction mapping with the contraction factor $c = 2\eta$ and G_i is not differentiable at the solution u^* for

$$i \in \mathcal{N} := \{i : ((I - \eta M)u^* - \eta q)_i = 0\}. \tag{4.11}$$

Throughout this example, we choose $u_0 = 0.5 * \mathrm{ones}(n, 1)$ and set $\gamma(t,x) \equiv 10^3$, $\tau(t,x) \equiv -1$. For given n and $\Theta \in (0,1)$ (`theta`), we randomly generate the solution u^* (`sol`) and corresponding q as follows

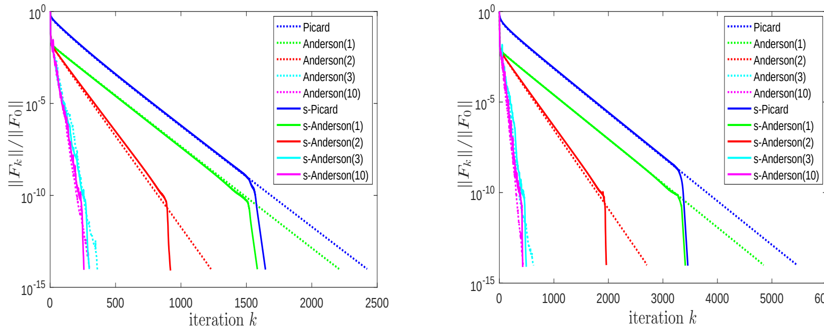
$$\texttt{sol} = \max\{\mathrm{rand}(\texttt{n}, \texttt{1}) - \texttt{theta}, \texttt{0}\}; \quad \texttt{q} = -\texttt{M} * \texttt{sol}; \tag{4.12}$$

By the setting of this problem, there are around $\Theta \times n$ components in $\mathcal{N}$ defined by (4.11).

First, we compare the performance of Anderson(m) and s-Anderson(m) with different values of m. Set $\Theta = 0.4$, and $n = 200, 300$ in (4.12). The convergence of $\|F_k\|/\|F_0\|$ for Anderson(m) and s-Anderson(m) with $m = 0, 1, 2, 3, 10$ are plotted in Fig. 4.2, from which we can see that s-Anderson(m) is faster than Anderson(m) always and s-Anderson(10) is the best. In [18], the following dynamically updating of depth m_k is introduced and used,

$$m_k = \mathrm{median}([m_1; \tilde{m}_k; m_2]) \quad \textit{with} \quad \tilde{m}_k = \mathrm{ceil}(-\log_{10}\|F_k\|), \tag{4.13}$$

where m_1 and m_2 are positive integers to control the lower and upper bounds of m. In particular, if $m_1 = m_2$, then the corresponding algorithms are just Anderson(m) and s-Anderson(m) with $m = m_1 = m_2$. Fig. 4.3 shows the number of iterations of Anderson(m) and s-Anderson(m) to satisfy the stop criterion in (4.2) using dynamic depth selection (4.13) with $m = m_1 = m_2$ and $m_1 \neq m_2$, in which the best result is located at $m = m_1 = m_2 = 8$ by s-Anderson(m). From Fig. 4.3, we find that the number of iterations is not monotone decreasing as m is increasing. Whether the dynamic depth selection approaches can improve the convergence of Anderson acceleration methods is an interesting topic for further research.



(a) $n = 200$ (b) $n = 300$

Fig. 4.2: Convergence of $\|F_k\|/\|F_0\|$ by Anderson(m) and s-Anderson(m) for Example 4.2 with $n = 200$ and $n = 300$

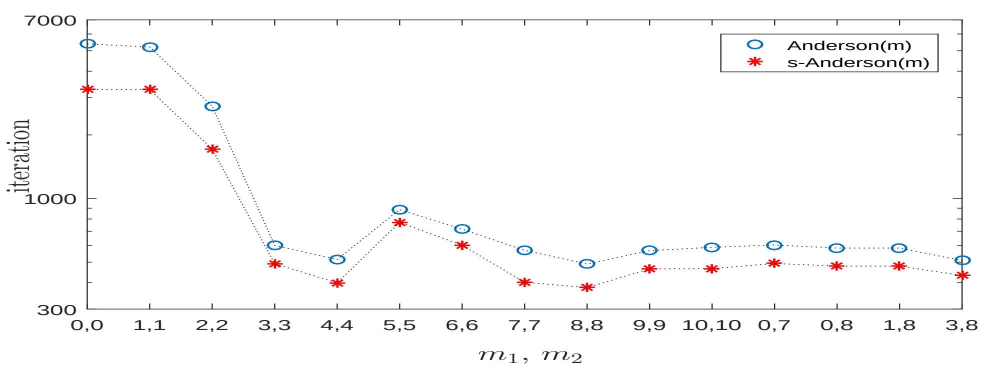
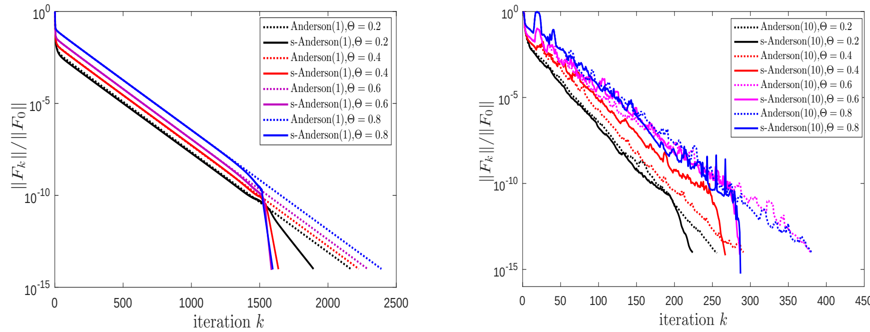


Fig. 4.3: Performance of Anderson(m) and s-Anderson(m) using dynamic depth selection (4.13) with $m = m_1 = m_2$ and $m_1 \neq m_2$ for Example 4.2

Next, we test the performance of Anderson(m) and s-Anderson(m) for different values of Θ, *since its value controls the number of dimensions, on which* G *is nonsmooth at* u^*. *Let* $n = 200$. *For* $\Theta = 0.2$, 0.4, 0.6 *and* 0.8, *we plot the convergence of* $\|F_k\|/\|F_0\|$ *by Anderson(m) and s-Anderson(m) with* $m = 1, 10$ *in Fig. 4.4. The displayed results in Fig. 4.4 show that s-Anderson(m) is faster than Anderson(m) for all these cases. In particular, as* Θ *is larger, the superiority on the local convergence rate of s-Anderson(m) compared with Anderson(m) is more obvious, which corresponds to the observation (iv) given at the beginning of this section.*

4.3. Nonsmooth Dirichlet problem. Consider the Dirichlet problem [7]

$$\left\{ \begin{array}{lll} -\Delta v + \beta v & = \lambda \max\{v - \varphi(x,y), 0\} + \psi(x,y) & \text{in} \quad \Xi \\ v & = f(x,y) \quad \text{on} \quad \Xi, & \end{array} \right. \tag{4.14}$$



(a) (b)

Fig. 4.4: Performance of Anderson(m) and s-Anderson(m) with $m = 1, 10$ for Example 4.2 with four different values of Θ

where $\Xi = (0,1) \times (0,1)$, $\bar{\Xi}$ denotes the boundary of Ξ, $\varphi, \psi \in C(\bar{\Xi}) \cap C^1(\Xi)$, $f \in C(\bar{\Xi})$, $\beta > 0$ and $\lambda \in \mathbb{R}$. Using the five point centered finite difference method for the Dirichlet problem (4.14) with a mesh size h at grid (x_i, y_j) gives

$$(4.15) \qquad -v_{i,j+1} - v_{i,j-1} + 4v_{i,j} - v_{i+1,j} - v_{i-1,j} + \beta h^2 v_{i,j} = \lambda h^2 \max\{v_{i,j} - \varphi_{i,j}, 0\} + h^2 \psi_{i,j}.$$

By transforming $(v_{i,j})$ to a vector u, (4.15) can be illustrated by the following system

$$(4.16) \qquad (-L + 4I - U + \beta h^2 I)u = \lambda h^2 \max\{u + p, 0\} + q,$$

where L and U are lower and upper diagonal matrices with nonnegative elements, $h = 1/(\sqrt{n}+1)$, $p, q \in \mathbb{R}^n$ are the corresponding vectors transformed by $\varphi_{i,j}$ and $h^2\psi_{i,j}$. Then, (4.16) is equivalent to the following fixed point problem

$$(4.17) \qquad u = G(u) := \frac{1}{4+\beta h^2}(L+U)u + \frac{h^2\lambda}{4+\beta h^2}\max\{u+p, 0\} + \frac{1}{4+\beta h^2}q.$$

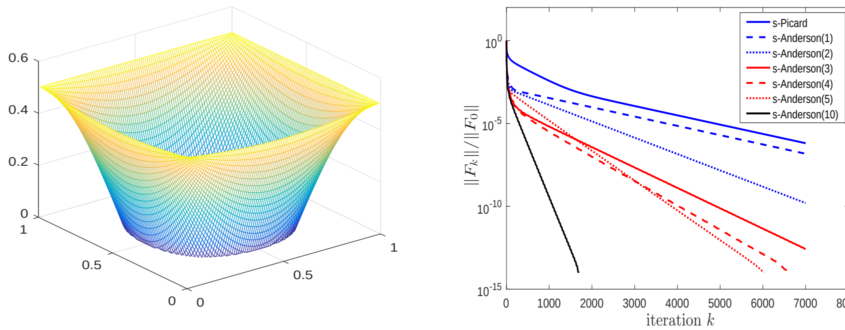
When $\beta > |\lambda|$, from

$$\|G(u) - G(v)\| \leq \frac{4 + |\lambda| h^2}{4 + \beta h^2}\|u - v\|,$$

the function G in (4.17) is a contraction mapping with factor $c := \frac{4+|\lambda|h^2}{4+\beta h^2}$.

Example 4.3. *We consider the nonsmooth fixed point problem (4.17) from the finite difference discretization of the nonsmooth Dirichlet problem (4.14). Let the solution of problem (4.14) be* $v(x,y) = \max(-\sin(x\pi)\sin(y\pi) + 0.5, 0)$*, and* u^* *present the values of* $v(x,y)$ *at the mesh points for given mesh size* $h = 1/(\sqrt{n}+1)$*. We randomly generate* $p = -0.4 * \mathrm{rand(n,1)}$ *and set* $q = (4+\beta h^2)u^* - (L+U)u^* - \lambda h^2 \max\{u^* + p, 0\}$ *with* $\lambda = 1$ *and* $\beta = 2$*. Notice that the contraction factor of* G *is very close to 1 at this situation.*

When $n = 64 \times 64$*, the original function is plotted in Fig. 4.5(a), in which we can see that it is nonsmooth. Choosing the initial point* $u_0 = 0.5 * \mathrm{rand}(n,1)$*, the convergence performance of*



(a) $v(x, y) = \max(-\sin(x\pi)\sin(y\pi) + 0.5, 0)$ (b) Convergence rate

Fig. 4.5: Solution and convergence performance of s-Anderson(m) for Example 4.3 with $n = 64 \times 64$

$\|F_k\|/\|F_0\|$ for s-Anderson(m) are plotted in Fig. 4.5(b). For different values of n, the convergence rates at the stopped point, defined by $(\|F_k\|/\|F_0\|)^{1/k}$, are listed in Table 4.2. This example shows that s-Anderson(m) can effectively solve this problem with a contraction factor very close to 1, and s-Anderson(m) is faster as m increases from 0 to 20.

$\sqrt{n}$	$1-c$	$m=0$	$m=1$	$m=2$	$m=3$	$m=5$	$m=10$	$m=20$
16	8.635e-4	9.793e-01	9.789e-01	9.632e-01	9.519e-01	9.191e-01	8.501e-01	7.750e-01
32	2.294e-4	9.944e-01	9.940e-01	9.897e-01	9.860e-01	9.784e-01	9.480e-01	9.078e-01
64	5.916e-5	9.978e-01	9.978e-01	9.968e-01	9.959e-01	9.946e-01	9.832e-01	9.758e-01
128	1.502e-5	9.991e-01	9.985e-01	9.983e-01	9.976e-01	9.976e-01	9.964e-01	9.930e-01

Table 4.2: Values of $(\|F_k\|/\|F_0\|)^{1/k}$ by s-Anderson(m) for Example 4.3

5. Conclusions.

Anderson acceleration does not use derivatives in its iterations, but it is difficult to prove its convergence without continuous differentiability. Most existing convergence results of Anderson acceleration are established under the assumption that the involved function is continuously differentiable [6, 9, 19, 24, 25]. For a special class of nonsmooth functions that is a sum of a smooth term and a nonsmooth term with a small Lipschitz constant, convergence of Anderson acceleration is proved in a recent paper [3]. In this paper, we give new convergence results of Anderson acceleration for nonsmooth fixed point problem (1.1), which has a composite max function in G. Theorem 2.1 shows that Anderson(1) is q-linear convergent with a q-factor $\hat{c} \in (\frac{2c-c^2}{1-c}, 1)$, which can be strictly smaller than $\frac{3c-c^2}{1-c}$ given in [3, 24]. Moreover, we construct a smoothing approximation $\mathcal{G}(\cdot, \mu)$ for the nonsmooth function G in (3.6), where $\mathcal{G}(\cdot, \mu)$ is also a contraction mapping and has the same fixed point as G. Then, we propose an Anderson accelerated algorithm with $\mathcal{G}(u, \mu)$ and prove its local r-linear convergence with factor c for nonsmooth fixed point problem (1.1), which is same as the convergence rate of Anderson acceleration for the continuously differentiable case.

Acknowledgment We would like to thank Prof. Tim Kelley for introducing us the research topics on Anderson acceleration for nonsmooth fixed problems. We would also like to thank Prof

Angela Kunoth and the two referees for their helpful comments.

REFERENCES

[1] H. An, X. Jia and H. Walker, *Anderson acceleration and application to the three-temperature energy equations*, J. Comput. Phys., 347(2017), pp. 1–17.
[2] D. G. Anderson, *Iterative procedures for nonlinear integral equations*, J. ACM, 12(1965), pp. 547–560.
[3] W. Bian, X. Chen and T. Kelley, *Anderson acceleration for a class of nonsmooth integral equations,* SIAM J. Sci. Comput., 43(2021), pp. S1–S20.
[4] N. N. Carlson and K. Miller, *Design and application of a gradient weighted moving finite element code I: In one dimension*, SIAM J. Sic. Comp., 19(1998), pp. 766–798.
[5] X. Chen, *Smoothing methods for nonsmooth, nonconvex minimization*, Math. Program., 134(2012), pp. 71–99.
[6] X. Chen and C. Kelley, *Convergence of the EDIIS algorithm for nonlinear equations*, SIAM J. Sci. Comput., 41(2019), pp. A365–A379.
[7] X. Chen, L. Qi and Z. Nashed, *Smoothing methods and semismooth methods for nondifferentiable operator equations*, SIAM J. Numer. Anal. 38 (2000), pp. 1200-1216.
[8] F. H. Clarke, *Optimization and Nonsmooth Analysis*, SIAM Publisher, New York, 1990.
[9] C. Evans, S. Pollock, L. G. Rebholz, and M. Xiao, *A proof that Anderson acceleration improves the convergence rate in linearly converging fixed point methods (but not in those converging quadratically)*, SIAM J. Numer. Anal., 58(2020), pp. 788–810.
[10] V. Eyert, *A comparative study on methods for convergence acceleration of iterative vector sequences*, J. Comput. Phys., 124(1996), pp. 271–285.
[11] H. Fang and Y. Saad, *Two classes of multisecant methods for nonlinear acceleration*, Numer. Linear Algebra Appl., 16(2009), pp. 167–221.
[12] R. Haelterman, J. Degroote, D. Van Heule and J. Vierendeels, *The quasi-Newton least squares method: A new and fast secant method analyzed for linear systems*, SIAM J. Numer. Anal., 47(2009), pp. 2347–2368.
[13] N. Higham and N. Strabic, *Anderson acceleration of the alternating projections method for computing the nearest correlation matrix*, Numer. Algorithms, 72(2016), pp. 1021–1042.
[14] K. N. Kudin, G. E. Scuseria and E. Cances, *A black-box self-consistent field convergence algorithm: One step closer*, J. Chem. Phys., 16(2002), pp. 8255–8261.
[15] L. Lin and C. Yang, *Elliptic preconditioner for accelerating the self-consistent field iteration in Kohn-Sham density functional theory*, SIAM J. Sic. Comp., 35(2013), pp. S277–S298.
[16] K. Miller, *Nonlinear Krylov and moving nodes in the method of lines*, J. Comp. Appl. Math., 183(2005), pp. 275–287.
[17] J.M. Ortega and W.C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, (1970).
[18] S. Pollock, and L. G. Rebholz, *Anderson acceleration for contractive and noncontractive operators*, IMA J. Numer. Anal., 41(2021), pp. 2841–2872.
[19] S. Pollock, L. G. Rebholz, and M. Xiao, *Anderson-accelerated convergence of Picard iterations for incompressible Navier-Stokes equations*, SIAM J. Numer. Anal., 57(2019), pp. 615–637.
[20] P. Pulay, *Improved SCF convergence acceleration*, J. Comp. Chem., 3(1982), pp. 556–560.
[21] C. Reisinger and J.H. Witte, *On the use of policy iteration as an easy way of pricing American options*, SIAM J. Financial Math., 3(2012), pp. 459–478.
[22] R. Schneider, T. Rohwedder, A. Neelov and J. Blauert, *Direct minimization for calculating invariant subspaces in density functional computations of the electronic structure*, J. Comp. Math., 27(2008), pp. 360–387.
[23] P. Stasiak and M. Matsen, *Efficiency of pseudo-spectral algorithms with Anderson mixing for the SCFT of periodic block-copolymer phases*, Eur. Phys. J. E, 34(2011), pp. 1–9.
[24] A. Toth and C. Kelley, *Convergence analysis for Anderson acceleration*, SIAM J. Numer. Anal., 53(2015), pp. 805–819.
[25] H. F. Walker and P. Ni, *Anderson acceleration for fixed-point iterations*, SIAM J. Numer. Anal., 49(2011), pp. 1715–1735.
[26] T. Washio and C. W. Oosterlee, *Krylov subspace acceleration for nonlinear multigrid schemes*, Elec. Trans. Num. Anal., 6(1997), pp. 271–290.
[27] J. Zhang, B. O'Donoghue, S. Boyd, Global convergent type-I Anderson acceleration for nonsmooth fixed-point iterations, SIAM J. Optim., 30(2020), pp. 3170-3197.