Generalized Smooth Variational Inequalities: Methods with Adaptive Stepsizes

Daniil Vankov [1] **Angelia Nedić** [1] **Lalitha Sankar** [1]

Abstract

Variational Inequality (VI) problems have attracted great interest in the machine learning (ML) community due to their application in adversarial and multi-agent training. Despite its relevance in ML, the oft-used strong-monotonicity and Lipschitz continuity assumptions on VI problems are restrictive and do not hold in many machine learning problems. To address this, we relax smoothness and monotonicity assumptions and study structured non-monotone generalized smoothness. The key idea of our results is in adaptive stepsizes. We prove the first-known convergence results for solving generalized smooth VIs for the three popular methods, namely, projection, Korpelevich, and Popov methods. Our convergence rate results for generalized smooth VIs match or improve existing results on smooth VIs. We present numerical experiments that support our theoretical guarantees and highlight the efficiency of proposed adaptive stepsizes.

1. Introduction

We consider a constrained variational inequality (VI) problem on a set $U \subseteq \mathbb{R}^m$. Given an operator $F : U \to \mathbb{R}^m$, the variational inequality problem, denoted as VI(U, F), is defined as

$$\text{find } u^* \in U \text{ such that } \langle F(u^*), u - u^* \rangle \geq 0, \ \forall u \in U. \tag{1}$$

Many numerical problems, such as constrained optimization, saddle-point problems, and multi-agent games are important and pertinent practical examples of VI problems. The recent interest in VIs is due to their application in machine learning applications such as generative adversarial networks (GANs) (Gemp and Mahadevan, 2018; Gidel et al., 2019) and reinforcement learning (Daskalakis et al., 2020; Kotsalis et al., 2022).

The VI problem has been extensively studied (Facchinei and Pang, 2003; Tseng, 1995; Beznosikov et al., 2022) under (strong)-monotonicity and Lipschitz continuity assumptions on the operator. Two optimal solution approaches for solving VIs under these assumptions are the well-known extra-gradient (which we interchangeably refer to as the Korpelevich method) (Korpelevich, 1976), and optimistic gradient (which we refer to as the Popov method here) (Popov, 1980) methods; both require taking an extra-gradient step in every iteration though Popov requires only one oracle call to do so. Despite this progress, machine learning (ML) applications do not satisfy such assumptions (Diakonikolas et al., 2021; Zhang et al., 2020). To this end, recent works (Wei et al., 2021; Diakonikolas et al., 2021; Vankov et al., 2023) have proposed and investigated new classes of structured non-monotone operators to address non-monotonicity in VIs. Furthermore, while the bulk of research on VIs assume Lipschitz continuity or boundedness of the operator, more recently, guarantees for non-smooth VIs with only a linear growth assumption have been considered using Korpelevich (Vlatakis-Gkaragkounis et al., 2023) and Popov (Vankov et al., 2023) methods.

For the class of smooth VI operators, (Malitsky, 2020) propose adaptive golden ratios stepsizes to adapt to a local Lipschitz constant. (Böhm, 2023) utilize this adaptive schedule for the Extragradient method and prove sublinear convergence under a weak Minty assumption. (Gorbunov et al., 2022b) use clipping techniques to show convergence of stochastic Extragradient method under heavy tail noise. They use the fact that the operator is Lipschitz continuous to show that iterates of the *clipped* Korpelevich method stay within the bounded set. However, the resulting Lipschitz constants can be very large, thus leading to extremely small stepsizes and slow convergence rates. More generally, for non-smooth VIs, despite the lack of theoretical results, adaptive and normalized first-order methods are popular and efficient in practice for training GANs (Brock et al., 2019), (Karras et al., 2020). Recently, (Jelassi et al., 2022) argues that the success of the ADAM optimizer (Kingma and Ba, 2015) for GANs is due to normalization.

Non-smoothness assumptions on the operator are of increas-

[1]Department of Electrical and Computer Engineering, Arizona State University Tempe, Arizona, USA. Correspondence to: Daniil Vankov <dvankov@asu.edu>.

Proceedings of the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 235, 2024. Copyright 2024 by the author(s).

Table 1. Summary of results on rates at which the quantities R_k^2 and D_k^2 decrease with the number k of iterations. We use "Asym" as an abbreviation for asymptotic results. Performance measure for Popov method is $R_k^2 = \text{dist}^2(u_k, U^*) + \|u_k - h_{k-1}\|^2$, and for Korpelevich method is $D_k^2 = \text{dist}^2(h_k, U^*)$. Constant $\underline{\gamma}$ is a lower bound of adaptive stepsizes, i.e. $\underline{\gamma} \leq \gamma_k$ for all $k \geq 0$. Constant $C = 2^{-p/2} \min\{\frac{1}{2}(2R_1^2)^{-(p-2)/2}, 2^{-p+2}\mu\underline{\gamma}\}$. Constants B_u, B_h are upper bounds of $\text{dist}(u_k, U^*)$ and $\text{dist}(h_k, U^*)$ respectively.

Methods	$p \leq 2$	$p > 2$
Projection	Asym (Thm 4.1)	Asym (Thm 4.1)
Korpelevich	$D_0^2 \exp\left(-\frac{\mu}{B_u^{2-p}}\underline{\gamma}K\right)$ (Thm 5.3)	$D_0^2(1+\mu\underline{\gamma}(D_0^2)^{(p-2)/2}k)^{-2/(p-2)}$ (Thm 5.3)
Popov	$R_1^2 \exp\left(-\frac{\mu}{B_h^{2-p}}\underline{\gamma}K\right)$ (Thm 6.3)	$R_1^2(1+pC(R_1^2)^{(p-2)/2}k)^{-2/(p-2)}$ (Thm 6.3)

ing interest across the broader optimization literature in the context of ML. In the context of deep learning optimization (Zhang et al., 2020) observed a linear dependence between the norm of the operator and its Jacobian. Based on this observation, they introduced the (L_0, L_1)-smoothness assumption, under which the sublinear convergence rate for the clipped gradient method is derived. Following this, (Chen et al., 2023) proposed an α-symmetric class of *generalized smooth operators* that allows extending (L_0, L_1)-smoothness to non-differentiable operators. For this class, the sublinear convergence rate for normalized gradient descent was proved.

We introduce the class of structured non-monotone *generalized smooth VIs* with α-symmetric operators. We address the following two natural questions in this context and answer them both in the affirmative: (i) Can convergence guarantees be assured for VIs under the assumption of generalized smoothness of the operator?; and (ii) If so, is the convergence for generalized smooth VIs the same as that for smooth VIs?

We focus on the class of structured non-monotone VIs with p-quasi sharp operators under the weakest known assumption on generalized smoothness called α-symmetricity. The class of p-quasi sharp operators includes weak sharp and quasi-strongly monotone operators and coincides with the Saddle-Point Metric Subregularity (SP-MS) assumption for $p \geq 2$. Analogously, the class of α-symmetric operators is the weakest class of continuous operators and includes Lipschitz continuous and (L_0, L_1)-smooth operators. Our key contributions are summarized below (see also Table 1)

- We provide the first known analysis of first-order methods for solving structured non-monotone VIs under generalized smooth assumption. In particular, we focus on a class of p-quasi sharp and α-symmetric operators. The key feature of our analysis is the use of cleverly chosen adaptive stepsizes for three well-known methods, namely, projection, Korpelevich, and Popov.

- For the projection method with adaptive stepsizes, we prove asymptotic convergence to a solution. Moreover, as a consequence, we show that the method converges linearly to the 1-neighborhood of the solution.

- For Korpelevich and Popov methods, we prove asymptotic convergence to a solution. Using the α-symmetricity assumption, we show that the sequence of provided adaptive stepsizes are lower bounded for both methods. Moreover, we show a linear convergence rate for both methods for 2-quasi sharp operators. For $p > 2$, we show a rate of $O(k^{-2/(p-2)})$, which is faster than the rate $O(k^{-1/(p-1)})$ provided in (Wei et al., 2021) under more restrictive assumptions on monotonicity and Lipschitz continuity of the operator.

- When the parameters of the problem such as p, α, L_0, L_1 are unknown or overestimated, for the Korpelevich method, we introduce an easier method to fine-tune stepsize parameters. We do so by separating the clipping part from the stepsize parameter sequence. For these easy-to-tune stepsizes, we prove asymptotic convergence. In particular, we introduce a *descent inequality*, which is crucial to obtain convergence rates under different step-size schedules. Additionally, using backtracking, we obtain convergence rates of the clipped Korpelevich method without information on parameters of the problem p, α, L_0, L_1.

- Finally, we present numerical experiments to compare the performance of the three considered methods with proposed adaptive stepsizes and Korpelevich with golden ratio stepsizes (Böhm, 2023) for different α and p. We observe that our proposed stepsizes outperform golden ratio stepsizes, and as predicted by our theory, the convergence of all methods slows down as α and p increase. We also highlight an experiment with decreasing adaptive stepsizes that are independent of the problem parameters. The key takeaway from this experiment is the observation that projection appears to significantly slow down relative to the other two methods for various parameter settings.

2. Assumptions on VI

The operator F is said to be Lipschitz continuous if there exists $L > 0$ such that:

$$\|F(u) - F(v)\| \leq L\|u - v\| \quad \text{for all } u, v \in U. \tag{2}$$

While being the most common assumption used in the literature, it is quite restrictive. Recently, a weaker assumption has been considered in (Zhang et al., 2020), given as follows:

$$\|\nabla F(u)\| \leq L_0 + L_1\|F(u)\| \quad \text{for all } u \in U. \tag{3}$$

When the operator $F(\cdot)$ is differentiable and L-Lipshitz continuous, it satisfies (3) with $L_0 = L$ and $L_1 = 0$. In the experiments reported in (Zhang et al., 2020), it was observed that the neural network models tend to satisfy the condition in (3). Later in (Chen et al., 2023), a weaker smoothness condition than (3) was introduced leading to a class of operators termed *α-symmetric*, which includes operators $F(\cdot)$ satisfying the following relation for some constants $L_0, L_1 \geq 0$ and for all $u, v \in U$:

$$\|F(u) - F(v)\| \leq \left(L_0 + L_1 \max_{\theta\in[0,1]} \|F(w_\theta)\|^\alpha\right)\|u - v\|, \tag{4}$$

where $w_\theta = \theta u + (1-\theta)v$ and $\alpha \in (0, 1]$. If a differentiable operator satisfies the condition in (3), then it satisfies the condition in (4). Moreover, it has been shown in (Chen et al., 2023) that the class of differentiable and α-symmetric operators with $\alpha = 1$ is equivalent to the class of operators with linear growth of the Jacobian, as in (3). Since the class of α-symmetric operators is a rather wide class, we focus on such operators.

Assumption 2.1. Given a set $U \subseteq \mathbb{R}^m$, the operator $F(\cdot) : U \to \mathbb{R}^m$ is α-symmetric over U, i.e., for some $\alpha \in (0, 1]$ and $L_0, L_1 \geq 0$, we have for all $u, v \in U$,

$$\|F(u) - F(v)\| \leq \left(L_0 + L_1 \max_{\theta\in[0,1]} \|F(w_\theta)\|^\alpha\right)\|u - v\|, \tag{5}$$

where $w_\theta = \theta u + (1 - \theta)v$.

An alternative characterization of an α-symmetric operators has been proved in (Chen et al., 2023), as given in the following proposition.

Proposition 2.2 ((Chen et al., 2023), Proposition 1)**.** *Let $U \subseteq \mathbb{R}^m$ be a nonempty set and let $F(\cdot) : U \to \mathbb{R}^m$ be an operator. Then, we have*

(a) $F(\cdot)$ is α-symmetric with $\alpha \in (0, 1)$ and constants $L_0, L_1 \geq 0$ if and only if the following relation holds for all $y, y' \in U$,

$$\|F(y) - F(y')\| \leq \|y - y'\|(K_0 + K_1\|F(y')\|^\alpha + K_2\|y - y'\|^{\alpha/(1-\alpha)}), \tag{6}$$

where $K_0 = L_0(2^{\alpha^2/(1-\alpha)} + 1)$, $K_1 = L_1 2^{\alpha^2/(1-\alpha)}3^\alpha$, and $K_2 = L_1^{1/(1-\alpha)}2^{\alpha^2/(1-\alpha)}3^\alpha(1-\alpha)^{\alpha/(1-\alpha)}$.

(b) $F(\cdot)$ is α-symmetric with $\alpha = 1$ and constants $L_0, L_1 \geq 0$ if and only if the following relation holds for all $y, y' \in U$,

$$\|F(y) - F(y')\| \leq \|y - y'\|(L_0 + L_1\|F(y')\|)\exp(L_1\|y - y'\|). \tag{7}$$

Proposition 2.2 is useful for our analysis, since it describes α-symmetric operator by using two points $y, y' \in U$, and bypasses the evaluation of $\max_{\theta\in(0,1)} \|F(w_\theta)\|^\alpha$.

A number of structured non-monotonicity assumptions were proposed to make the assumptions on the operator closer to real-life problems. Such assumptions include strong coherency (Song et al., 2020), weak Minty (Choudhury et al., 2023), Saddle-Point Metric Subregularity (SP-MS)(Wei et al., 2021), and p-quasi sharpness (Vankov et al., 2023). Each of these assumptions imposes a special structured lower bound for the quantity $\langle F(u), u - u^*\rangle$, where u^* is a solution to the underlying VI problem. In the recent work (Vankov et al., 2023), it has been shown that the class of p-quasi sharp operators includes strongly-monotone and strongly coherent operators, and coincides with the class of operators satisfying SP-MS condition for $p > 2$ which need not be monotone. Due to the generality of the p-quasi sharpness property, we consider the class of p-quasi sharp operators. To formally introduce this class, we define the solution set for the VI(U, F), which is denoted by U^*, as follows:

$$U^* = \{u^* \in U \mid \langle F(u^*), u - u^*\rangle \geq 0 \ \text{ for all } u \in U\}.$$

Throughout this paper, we make the following assumption on the constraint set and the solution set.

Assumption 2.3. The set $U \subseteq \mathbb{R}^m$ is a nonempty closed convex set, and the solution set $U^* \subseteq \mathbb{R}^m$ is nonempty and closed.

Assumption 2.4. The operator $F(\cdot) : U \to \mathbb{R}^m$ has a p-quasi sharpness property over U relative to the solution set U^*, i.e., for some $p > 0$, $\mu > 0$, and for all $u \in U$ and all $u^* \in U^*$,

$$\langle F(u), u - u^*\rangle \geq \mu \operatorname{dist}^p(u, U^*). \tag{8}$$

Next, motivated by the example provided in (Vankov et al., 2023), we present an operator that is p-quasi sharp and α-symmetric.

Proposition 2.5. *Let* $F(u) = \begin{bmatrix} sign(u_1)|u_1|^{p-1} + u_2 \\ sign(u_2)|u_2|^{p-1} - u_1 \end{bmatrix}$ *for $u \in \mathbb{R}^2$. Then, the operator $F(\cdot)$ is p-quasi sharp*

for $p \geq 1$ with $\mu = 2^{1-p}$. Moreover, when $p > 2$, the operator is α-symmetric with $\alpha = \frac{p-2}{p-1}$, and $L_0 = 1 + (p-1)2^{1/2}4^{1/(p-1)}, L_1 = 2(p-1)2^{1/2(p-1)}$.

Note that for this example, the parameter p of quasi-sharpness depends on parameter α of α-symmetricity and vice versa. To illustrate the importance of α-symmetric assumption consider the operator $F(u)$ from Proposition 2.5 with $p = 9.0$ and $\alpha = 0.875$ on a ball $B(0, 10.0)$. Operator $F(u)$ is Lipschitz continuous in the set $B(0, 10.0)$, but the Lipschitz constant is larger than $L = 10^7 - 1$, while $L_0 = 1 + 8 \times 2^{3/4}$ and $L_1 = 16 \times 2^{1/16}$ are relatively small.

3. Methods for Solving VI

Our goal is to develop methods and provide convergence guarantees for solving non-monotone VI problems with α-symmetric operators. In this section, we consider three popular methods for solving VIs. The first method of our interest is the projection method, given as follows:

$$u_{k+1} = P_U(u_k - \gamma_k F(u_k)) \qquad \text{for all } k \geq 0, \tag{9}$$

where $\gamma_k > 0$ is the step-size and $u_0 \in U$ is an arbitrary initial point. Despite its simplicity and efficiency in optimization, the projection method may diverge for monotone VIs (Daskalakis et al., 2018). The projection method converges to the solution for strongly monotone VIs but is not optimal. In fact, two methods proposed in (Korpelevich, 1976; Popov, 1980), named extragradient and optimistic methods are optimal for strongly monotone VIs. In the recent works (Diakonikolas et al., 2021; Wei et al., 2021; Loizou et al., 2021; Vankov et al., 2023) authors considered structured non-monotone VIs. The convergence rate of Popov and Korpelevich methods for quasi-strong monotone and 2-quasi sharp operators matches optimal convergence rates (Beznosikov et al., 2022; Gorbunov et al., 2022a; Choudhury et al., 2023; Vankov et al., 2023). Motivated by these results, we consider these two methods. The Korpelevich method, is given as follows. For all $k \geq 0$,

$$\begin{aligned} u_k &= P_U(h_k - \gamma_k F(h_k)), \\ h_{k+1} &= P_U(h_k - \gamma_k F(u_k)), \end{aligned} \tag{10}$$

where γ_k is step-size and $h_0 \in U$ is an arbitrary initial point. The Popov method, is presented below. For all $k \geq 0$,

$$\begin{aligned} u_{k+1} &= P_U(u_k - \gamma_k F(h_k)), \\ h_{k+1} &= P_U(u_{k+1} - \gamma_{k+1} F(h_k)), \end{aligned} \tag{11}$$

where $\gamma_k > 0$ is the step-size and $u_0, h_0 \in U$ are arbitrary initial points.

4. Convergence Analysis for Projection Method

In this section, we provide convergence results for projection method (9). To show convergence of projection method for α-symmetric and p-quasi sharp operators we define adaptive[1] stepsizes as:

$$\gamma_k = \beta_k \min\left\{1, \frac{1}{\|F(u_k)\|}\right\}. \tag{12}$$

Theorem 4.1. *Let Assumptions 2.1, 2.3, 2.4 hold. Also, let stepsizes be given by (12) and sequence β_k be such that $\sum_{k=0}^{\infty} \beta_k = \infty, \sum_{k=0}^{\infty} \beta_k^2 < \infty$. Then, the following statements hold for the iterate sequences $\{u_k\}$ generated by the projection method (9):*

(a) The following inequality holds for any $k \geq 0$,

$$\|u_{k+1} - u^*\|^2 \leq \|u_k - u^*\|^2 - 2\gamma_k \mathrm{dist}^p(u_k, U^*) + \beta_k^2 \tag{13}$$

Moreover, for any solution $u^* \in U^*$ the sequence $\{\|u_k - u^*\|\}_{k=0}^{\infty}$ is bounded by a constant $B(u^*)$.

(b) The stepsizes sequence $\{\gamma_k\}$ is bounded by

$$\gamma_k \geq \beta_k \min\left\{1, \frac{1}{C_1}, \frac{1}{\overline{C}_1}\right\} \qquad \textit{for all } k \geq 0,$$

where C_1 and $\bar{C}_1$ are given by

$$\begin{aligned} C_1 &= B(u_0^*)(K_0 + K_1\|F(u_0^*)\|^{\alpha} \\ &\quad + K_2 B(u_0^*)^{\alpha/(1-\alpha)}) + \|F(u_0^*)\| \end{aligned}$$

$$\begin{aligned} \overline{C}_1 &= B(u_0^*)(L_0 + L_1\|F(u_0^*)\|)\exp(L_1 B(u_0^*)) \\ &\quad + \|F(u_0^*)\|, \end{aligned}$$

and $u_0^* = P_{U^*}(u_0)$.

(c) The sequence $\{u_k\}$ converges to some solution $\bar{u} \in U^*$.

Interestingly, to achieve the descent inequality (13) we don't use Assumption 2.1. But this assumption plays a crucial role in showing the upper bound on a sequence $\{\|F(u_k)\|\}$, and as a result a lower bound on the stepsizes. In order to get converges from equation (13), it is required to show that stepsize sequence is square summable, i.e. $\sum_{k=0}^{\infty} \gamma_k^2 \leq \infty$.

Corollary 4.2. *The projection method with adaptive stepsizes as given in (12) and Theorem 4.1 enjoys a linear convergence rate to a 1-neighborhood of the solution set for $p \geq 2$.*

Corollary 4.2 follows straight from equation (13) in Theorem 4.1 (a) and stepsizes from (Stich, 2019). In Section 8,

[1] By adaptive stepsizes we mean stepsizes that depend on the operator value at the current point and previous points.

we show that this rate holds in our experiments. Moreover, results of Theorem 4.1 hold for normalized projection method (4), that can be viewed as deterministic version of normalized SGD (Jelassi et al., 2022):

$$u_{k+1} = P_U\left(u_k - \beta_k \frac{1}{\|F(u_k)\|} F(u_k)\right), \ \forall\, k \geq 0.$$

5. Convergence Analysis for Korpelevich Method

In this section, we consider the Korpelevich method with adaptive stepsizes. We show the asymptotic convergence of the method and provide its convergence rate. Firstly, we derive a basic inequality for Korpelevich method (10) with arbitrary stepsizes without any assumptions on the operator. This lemma is the basis for all the subsequent results.

Lemma 5.1. *Let U be a nonempty closed convex set. Then, for the iterate sequences $\{u_k\}$ and $\{h_k\}$ generated by the Korpelevich method (10) we have for all $y \in U$ and $k \geq 0$,*

$$\begin{aligned}\|h_{k+1} - y\|^2 \leq &\|h_k - y\|^2 - \|h_k - u_k\|^2 \\ &- 2\gamma_k \langle F(u_k), u_k - y\rangle \\ &+ \gamma_k^2 \|F(h_k) - F(u_k)\|^2.\end{aligned} \tag{14}$$

The proof of Lemma 5.1 can be found in Appendix D. Next we provide adaptive step-sizes for Korpelevich method employed with an α-symmetric operator with $\alpha \in (0, 1]$. When $\alpha \in (0, 1)$, the stepsizes are defined as follows: for all $k \geq 0$,

$$\begin{aligned}\gamma_k = \min\Bigg\{&\frac{1}{4\mu}, \frac{1}{3\sqrt{2}K_0}, \frac{1}{\|F(h_k)\|}, \\ &\frac{1}{3\sqrt{2}K_1\|F(h_k)\|^\alpha}, \frac{1}{3\sqrt{2}K_2}\Bigg\},\end{aligned} \tag{15}$$

while for $\alpha = 1$, the stepsizes are given by: for all $k \geq 0$,

$$\gamma_k = \min\left\{\frac{1}{4\mu}, \frac{1}{2\sqrt{2}eL_0}, \frac{1}{2\sqrt{2}eL_1\|F(h_k)\|}\right\}. \tag{16}$$

The constants $L_0 > 0$ and $L_1 > 0$ are related to the α-smoothness of the operator, while $K_0, K_1, K_2 > 0$ are defined in terms of L_0 and L_1 as given in Proposition 2.2.

We present a theorem on the convergence of the Korpelevich method (10) with adaptive stepsizes (15), (16), to the solution set for p-quasi sharp and α-symmetric operators with $\alpha \in (0, 1]$. There are two key ideas behind the proof.

Firstly, one need to upper bound $\gamma_k^2\|F(h_k) - F(u_k)\|^2$ from Lemma 5.1 to get descent inequality. To do that, we use Proposition 2.2 of α-symmetric operator and carefully chosen adaptive stepsizes $\{\gamma_k\}$ (15), (16). The stepsizes utilize the idea of normalization of operator or operator clipping. Notice that based on Proposition 2.2, $\|F(h_k) - F(h_{k-1})\|$ depends on $\|F(h_k)\|$ and $\|u_k - h_k\|$, while based on the structure of the method quantity $\|u_k - h_k\|$ can be upper bounded by $\gamma_k\|F(h_k)\|$. This makes the stepsizes γ_k depended only on operator value at point h_k and parameters from Assumption 2.1.

Secondly, to obtain convergence guarantees it is crucial to find a lower bound for the stepsizes sequence. Such a lower bound is possible to find when an operator is p-quasi sharp, or more generally when $\langle F(h_k), h_k - u^*\rangle$ is positive. When this property holds, it is easy to observe that quantity $\|h_k - u^*\|^2$ is non-increasing for all $u^* \in U^*$ and $k \geq 0$. Using this fact and assumption on α-symmetric operator we are able to present a lower bound on the adaptive stepsizes (15), (16). Finally, we apply deterministic version of Robbins-Siegmund lemma (Robbins and Siegmund, 1971)(see Lemma B.3 in Appendix) and get the desired convergence results. Formally, the theorem summarizing the result on convergence and lower bound on stepsizes is given below.

Theorem 5.2. *Let Assumptions 2.1, 2.3, 2.4 hold. Also, let the stepsizes be given by (15) if $\alpha \in (0, 1)$ and by (16) if $\alpha = 1$. Then, the following statements hold for the iterate sequences $\{u_k\}$ and $\{h_k\}$ generated by the Korpelevich method (10):*

(a) The following descent inequality holds for any solution $u^* \in U^*$ and any $k \geq 0$,

$$\begin{aligned}\|h_{k+1} - u^*\|^2 \leq \|h_k - u^*\|^2 - \frac{1}{2}\|u_k - h_k\|^2 \\ - 2\gamma_k\mu \mathrm{dist}^p(u_k, U^*).\end{aligned} \tag{17}$$

(b) The stepsizes sequence $\{\gamma_k\}$ is bounded below, i.e.,

$$\gamma_k \geq \underline{\gamma} \qquad \text{for all } k \geq 0.$$

For $\alpha \in (0, 1)$, the constant $\underline{\gamma}$ is given by

$$\underline{\gamma} = \min\left\{\frac{1}{4\mu}, \frac{1}{3\sqrt{2}K_0}, \frac{1}{C_1}, \frac{1}{3\sqrt{2}K_1C_1^\alpha}, \frac{1}{3\sqrt{2}K_2}\right\},$$

and for $\alpha = 1$, it is given by

$$\underline{\gamma} = \min\left\{\frac{1}{4\mu}, \frac{1}{2\sqrt{2}eL_0}, \frac{1}{2\sqrt{2}eL_1\overline{C}_1}\right\},$$

where the constants $K_0, K_1, K_2 > 0$ are as given in Proposition 2.2, while C_1 and $\bar{C}_1$ are given by

$$\begin{aligned}C_1 =& D_0(K_0 + K_1\|F(P_{U^*}(h_0))\|^\alpha + K_2(D_0)^{\frac{\alpha}{(1-\alpha)}}) \\ &+ \|F(P_{U^*}(h_0))\|, \\ \overline{C}_1 =& D_0(L_0 + L_1\|F(P_{U^*}(h_0))\|)\exp(L_1D_0) \\ &+ \|F(P_{U^*}(h_0))\|,\end{aligned}$$

and $D_0^2 = \mathrm{dist}^2(h_0, U^*)$.

(c) The iterates u_k and h_k converge to a solution $\bar{u} \in U^*$.

The proof of these results can be found in Appendix D. The conditions on a stepsize sequence $\{\gamma_k\}$ for Korpelevich method in Theorem 5.2 are less restrictive than conditions on $\{\gamma_k\}$ for projection method from Theorem 4.1. In Theorem 4.1, stepsizes should be diminishing, while for Korpelevich method, it is possible to have stepsizes with a global nonzero lower bound $\underline{\gamma}$. Next, we provide convergence rate for Korpelevich method.

Theorem 5.3. *Let Assumptions 2.1, 2.3, 2.4 hold. Consider the Korpelevich method (10) with the stepsizes γ_k given by (15) for $\alpha \in (0,1)$, and by (16) for $\alpha = 1$, and let*

$$D_k^2 = \mathrm{dist}^2(h_k, U^*) \qquad \text{for all } k \geq 0.$$

Then, the following results hold for the quantity D_k^2:
Case (a) *If the operator $F(\cdot)$ is p-quasi-sharp with $p \leq 2$, then*

$$D_{k+1}^2 \leq (1 - \frac{\mu}{(B_u)^{2-p}}\underline{\gamma})D_k^2 \qquad \text{for all } k \geq 0, \tag{18}$$

where $B_u \in \mathbb{R}$ is such that $\mathrm{dist}(u_k, U^*) \leq B_u$, $\forall k \geq 0$.

Case (b) *If the operator $F(\cdot)$ is p-quasi-sharp with $p > 2$, then*

$$D_{k+1}^2 \leq \frac{D_0^2}{(1 + \mu\underline{\gamma}k(D_0^2)^{(p-2)/2})^{2/(p-2)}} \qquad \text{for all } k \geq 1. \tag{19}$$

The constant $\underline{\gamma}$ in (18) and (19) is the lower-bound on the stepsizes provided in Theorem 5.2.

The proof of these results can be found in Appendix D. Notice that when F is L_0-Lipschitz continuous, the constant $L_1 = 0$, and $\alpha = 1$, then the stepsizes will be simplified to:

$$\gamma_k = \min\left\{\frac{1}{4\mu}, \frac{1}{2\sqrt{2}eL_0}\right\}. \tag{20}$$

In this case, the rate provided Theorem 5.3(a) recovers the optimal rate of $\mathcal{O}(\frac{L_0}{\mu}\exp(-\frac{\mu}{L_0}k))$ for 2-quasi-sharp operators. Notice that for $p > 2$, from equation (17) in Theorem 5.2 (a), it follows that convergence to 1-neighborhood of a solution is linear.

6. Convergence Analysis for Popov Method

We provide convergence results for Popov method (11) for solving structured non-monotone non-Lipschitz VIs. We start with a lemma on the iterates of Popov method which serve as a base for further convergence results. We collate all our proofs in Appendix E.

Lemma 6.1. *Let Assumption 2.3 hold. Then, for the iterate sequences $\{u_k\}$ and $\{h_k\}$ generated by the Popov method (11) we have for all $y \in U$ and $k \geq 1$,*

$$\begin{aligned}\|u_{k+1} - y\|^2 \leq \|u_k - y\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 \\ - 2\gamma_k\langle F(h_k), h_k - y\rangle \\ + 2\gamma_k^2\|F(h_k) - F(h_{k-1})\|^2.\end{aligned}$$

The stepsizes for Popov method for α-symmetric operators with different $\alpha \in (0,1]$ are given below. When $\alpha \in (0,1)$ the stepsizes for all $k \geq 1$ are:

$$\begin{aligned}\gamma_k = \min\Bigg\{&\frac{1}{\|F(h_{k-1})\|}, \frac{1}{6\sqrt{2}K_0}, \frac{1}{6\sqrt{2}K_1\|F(h_{k-1})\|^\alpha}, \\ &\frac{1}{6\sqrt{2}K_2(\|u_k - h_{k-1}\| + 1)^{\alpha/(1-\alpha)}}, \frac{1}{4\mu}\Bigg\}.\end{aligned} \tag{21}$$

For $\alpha = 1$, the stepsizes for Popov method for all $k \geq 1$ are

$$\begin{aligned}\gamma_k = \min\Bigg\{&\frac{1}{4\mu}, \frac{1}{\|F(h_{k-1})\|}, \\ &\frac{1}{2\sqrt{2}(L_0 + L_1\|F(h_{k-1})\|)\exp(L_1\|u_k - h_{k-1}\| + 1)}\Bigg\}\end{aligned} \tag{22}$$

To prove the convergence of Popov method we use the same two ideas as in the proof of Korpelevich method: (i) provide an upper bound on $\gamma_k^2\|F(h_k) - F(h_{k-1})\|^2$ utilizing the adaptive stepsizes (21), and (22); (ii) provide a lower bound on the stepsizes sequence. Based on Proposition 2.2, the term $2\gamma_k^2\|F(h_k) - F(h_{k-1})\|^2$ from Lemma 6.1 depends on $\|F(h_{k-1})\|$ and $\|h_k - h_{k-1}\|$. Unlike the bound in the analysis of Korpelevich method, in Popov method, the quantity $\|h_k - h_{k-1}\|$ can't be upper bounded by $\|F(h_{k-1})\|$. To address this issue, using the projection inequality and the method update rule, we provide a bound of $\|h_k - h_{k-1}\|$ using $\gamma_k\|F(h_{k-1})\|$ and $\|u_k - h_{k-1}\|$. Notice that, based on the method updates, the stepsize γ_k can't depend on h_k. Thus, normalization by $\|F(h_k)\|$, $\|F(h_k) - F(h_{k-1})\|$ or $\|h_k - h_{k-1}\|$ is not possible. Thus, γ_k uses normalization by $\|u_k - h_{k-1}\|$ and $\|F(h_{k-1})\|$. We next prove asymptotic convergence of Popov method with adaptive stepsizes.

Theorem 6.2. *Let Assumptions 2.1, 2.3, 2.4 hold. Also, let the stepsizes be given by (21) if $\alpha \in (0,1)$ and by (22) if $\alpha = 1$. Then, the following statements hold for the iterate sequences $\{u_k\}$ and $\{h_k\}$ generated by the Popov method:*

(a) The descent inequality holds for arbitrary solution $u^* \in U^*$ and any $k \geq 1$

$$\begin{aligned}\|u_{k+1} - u^*\|^2 + \|u_{k+1} - h_k\|^2 \leq \|u_k - u^*\|^2 \\ + \frac{1}{2}\|u_k - h_{k-1}\|^2 - \frac{1}{2}\|u_k - h_k\|^2 \\ - 2\gamma_k\langle F(h_k), h_k - u^*\rangle.\end{aligned} \tag{23}$$

(b) The stepsizes sequence $\{\gamma_k\}$ is bounded below, i.e.,

$$\gamma_k \geq \underline{\gamma} \qquad \text{for all } k \geq 1.$$

For $\alpha \in (0,1)$, the constant $\underline{\gamma}$ is given by

$$\begin{aligned}\underline{\gamma} = \min\Bigg\{&\frac{1}{4\mu}, \frac{1}{C_1}, \frac{1}{6\sqrt{2}K_0}, \frac{1}{6\sqrt{2}K_1(C_1)^{\alpha/2}}, \\ &\frac{1}{6\sqrt{2}K_2(\sqrt{2}R_1 + 1)^{\alpha/(1-\alpha)}}\Bigg\},\end{aligned} \tag{24}$$

and for $\alpha = 1$, it is given by

$$\underline{\gamma} = \min\left\{\frac{1}{4\mu}, \frac{1}{C_1}, \frac{1}{2\sqrt{2}(L_0 + L_1 C_1)\exp(\sqrt{2}L_1 R_1 + 1)}\right\}. \tag{25}$$

The constants K_0, K_1, K_2 are as given in Proposition 2.2, while C_1 and $\bar{C}_1$ are given by

$$\begin{aligned} C_1 =& \sqrt{2}R_1(K_0 + K_1\|F(P_{U^*}(u_1))\|^\alpha \\ &+ K_2(\sqrt{2}R_1)^{1/(1-\alpha)}) + \|F(P_{U^*}(u_1))\|, \end{aligned}$$

$$\begin{aligned} \bar{C}_1 =& \sqrt{2}R_1(L_0 + L_1\|F(P_{U^*}(u_1))\|)\exp(\sqrt{2}L_1 R_1) \\ &+ \|F(P_{U^*}(u_1))\|, \end{aligned}$$

and $R_1^2 = \|u_1 - P_{U^*}(u_1)\|^2 + \|u_1 - h_0\|^2$.

(c) The iterates u_k and h_k converge to a solution $\bar{u} \in U^*$.

Based on results of Theorem 6.2, we provide convergence rates of Popov method iterates to the solution set. Formally, in the next theorem we provide convergence rates for Popov method (11) with adaptive stepsizes (21) for α-symmpetric operators with $\alpha \in (0, 1]$.

Theorem 6.3. *Let Assumptions 2.1, 2.3, 2.4 hold. Consider the Popov method (11) with srepsizes γ_k given by (21) for $\alpha \in (0, 1)$ and by (22) for $\alpha = 1$, and let*

$$R_k^2 = \mathrm{dist}^2(u_k, U^*) + \|u_k - h_{k-1}\|^2 \quad \textit{for all } k \geq 0.$$

Then, the following results for the quantity R_k^2:

Case (a) *If operator $F(\cdot)$ is p-quasi-sharp with $p \leq 2$, then the following inequality holds for all $k \geq 1$,*

$$R_{k+1}^2 \leq \left(1 - \frac{\mu}{(B_h)^{2-p}}\underline{\gamma}\right) R_k^2, \tag{26}$$

where $B_h \in \mathbb{R}$ is such that $\mathrm{dist}(h_k, U^*) \leq B_h$, $\forall k \geq 0$.
Case (b) *If operator $F(\cdot)$ is p-quasi-sharp with $p > 2$, then the following inequality holds for all $k \geq 1$,*

$$R_{k+1}^2 \leq \frac{R_1^2}{(1 + pC2^{-p/2}(R_1^2)^{(p-2)/2}k)^{2/(p-2)}}, \tag{27}$$

where $C = \min\{\frac{1}{2}(2R_1^2)^{-(p-2)/2}, 2^{-p+2}\mu\underline{\gamma}\}$. The constant $\underline{\gamma}$ in (24) and (25) is the lower-bound on the stepsizes provided in Theorem 6.2.

7. Korpelevich Method with Clipping

In practice, constants L_0, L_1 and corresponding K_0, K_1, K_2 are unknown and have to be estimated. The estimation of such constants can be a difficult problem and lead to large values of L_0, L_1. Instead of estimating the parameters of the problem one can fine-tune stepsizes. Moreover, in the recent papers on generalized smoothness (Zhang et al., 2020; Chen et al., 2023), the theoretical stepsizes were not applied in numerical results but were fine-tuned. To address the issue of fine-tuning, in this section, we provide analysis for Korpelevich method with clipping. The clipped stepsizes are given by:

$$\gamma_k = \beta_k \min\left\{1, \frac{1}{\|F(h_k)\|}\right\}. \tag{28}$$

The parameter β_k can be easily fine-tuned, while normalization in the stepsizes remains independent from α, L_0, L_1 and corresponding K_0, K_1, K_2. Moreover, in real-life problems, it is also possible to use different schedule techniques (Defazio et al., 2023) for β_k. Next, we show that Korpelevich method with a simple choice of decreasing sequence β_k converges to the solution and has a better practical performance.

Theorem 7.1. *Let Assumptions 2.1, 2.3, 2.4 hold. Also, let the stepsizes be given by (28) and parameter $\{\beta_k\}$ be such that there exists $N > 0$ such that $\beta_k(K_0 + K_1 + K_2) < 1$, $(L_0 + L_1)\beta_k \exp(L_1\beta_k) < 1$ and $\beta_k \leq \frac{1}{4\mu}$ for all $k \geq N$. Then, the following statements hold for the iterate sequences $\{u_k\}$ and $\{h_k\}$ generated by the Korpelevich method (10):*

(a) The following descent inequality holds for any solution $u^* \in U^*$ and any $k \geq 0$,

$$\begin{aligned} \|h_{k+1} - u^*\|^2 \leq& \|h_k - u^*\|^2 - 2\gamma_k\mu\mathrm{dist}^p(u_k, U^*) \\ &- (1 - \beta_k^2 C_2(\beta_k))\|u_k - h_k\|^2, \end{aligned} \tag{29}$$

where

$$C_2(\beta_k) = \max\{(K_0 + K_1 + K_2)^2, (L_0 + L_1)\exp(L_1\beta_k)\}.$$

(b) The stepsizes sequence $\{\gamma_k\}$ is bounded below, i.e.,

$$\gamma_k \geq \beta_k \min\left\{1, \frac{1}{C_1}\right\} \qquad \textit{for all } k \geq 0,$$

where the constant C_1 is given by

$$\begin{aligned} C_1 = \bar{D}_N \max\{&(K_0 + K_1\|F(P_{U^*}(h_{\bar{N}}))\|^\alpha \\ &+ K_2(\bar{D}_N)^{\frac{\alpha}{(1-\alpha)}}), \\ &(L_0 + L_1\|F(P_{U^*}(h_{\bar{N}}))\|)\exp(L_1\bar{D}_N)\} \\ &+ \|F(P_{U^*}(h_{\bar{N}}))\|. \end{aligned}$$

where $(\bar{D}_N)^2 = \max_{k\in[0,N]}\{\mathrm{dist}^2(h_k, U^*)\}$, and $\bar{N} = \arg\max_{k\in[0,N]}\{\mathrm{dist}^2(h_k, U^*)\}$

(c) If $\sum_{k=0}^\infty \beta_k = \infty$, then the iterates u_k and h_k converge to a solution $\bar{u} \in U^*$.

The proof of these results can be found in Appendix F. While the quantity $\bar{D}_N$ can be large, it does not affect the asymptotic convergence of the method. In case, when β_k is such

Algorithm 1 Korpelevich Method with Backtracking

Input: $h_0 \in \mathbb{R}^n, \beta_0 = \beta, 0 < q < 1, k \geq 1$
for $t = 0, \ldots, k$ **do**
  while True **do**
    $\gamma_t = \beta_t \min\{1, \frac{1}{\|F(h_t)\|}\}$
    $\bar{u}_t = P_U(h_t - \gamma_k F(h_t))$
    if $\gamma_t^2 \|F(h_t) - F(\bar{u}_t)\|^2 \leq \frac{1}{2}\|h_t - \bar{u}_t\|^2$ **then**
      break
    else
      $\beta_t = q \times \beta_t$
    end if
  end while,
  $u_t \leftarrow \bar{u}_t$
  $h_{t+1} = P_U(h_t - \gamma_k F(u_t))$
  $\beta_{t+1} \leftarrow \beta_t$
end for

that $\beta_k(K_0 + K_1 + K_2) < 1$, $(L_0 + L_1)\beta_k \exp(L_1\beta_k) < 1$ and $\beta_k \leq \frac{1}{4\mu}$ for all $k \geq 0$, for all $N \geq 0$, the quantity $\bar{D}_N$ is equal to $\mathrm{dist}(h_0, U^*)$.

Corollary 7.2. *Consider Korpelevich method (10) with adaptive stepsizes given in (28). Let $\beta_k(K_0+K_1+K_2) < 1$, $(L_0 + L_1)\beta_k \exp(L_1\beta_k) < 1$ and $\beta_k \leq \frac{1}{4\mu}$ for all $k \geq 0$. Then if $p = 2$ and $\beta_k = \beta$, the method has a linear convergence rate. If $p > 2$ then the method has convergence rate of $O((\sum_{i=0}^k \beta_i)^{-2/(p-2)})$.*

These results are direct consequence from equation (29) in Theorem 7.1 and Lemma 6 (Section 2 in (Polyak, 1987)).

To avoid parameter estimation and diminishing stepsizes, we combine the results in Theorem 7.1 with a backtracking technique for parameter β_k. The generality of Theorem 7.1 allows us to use backtracking technique for β_k and obtain convergence rates $O(\exp(-k))$ for $p \leq 2$ and $O(k^{-2/(p-2)})$ for $p > 2$ without any additional information on parameters of the problem p, α, L_0, L_1. We define Korpelevich method with clipping and backtracing in Algorithm 1.

Theorem 7.3. *Let Assumptions 2.1, 2.3, 2.4 hold. Let the iterate sequences $\{u_k\}$ and $\{h_k\}$ be generated by the Algorithm 1, then the sequences $\{u_k\}$ and $\{h_k\}$ converge to a solution $\bar{u} \in U^*$. Moreover, there exists a lower bound $\underline{c} \in \mathbb{R}$ such that for all $k \geq 0$,*

$$\min\left\{1, \frac{1}{\|F(h_k)\|}\right\} \geq \underline{c}.$$

Additionally, let $\beta \leq \frac{1}{4\mu}$ and let

$$D_k^2 = \mathrm{dist}^2(h_k, U^*) \qquad \text{for all } k \geq 0.$$

Then, the following results hold for the quantity D_k^2:
Case (a) *If the operator $F(\cdot)$ is p-quasi-sharp with $p \leq 2$, then*

$$D_{k+1}^2 \leq (1 - \frac{\mu}{(B_u)^{2-p}}\,\underline{c}\,\beta_k)D_k^2 \qquad \text{for all } k \geq 0, \tag{30}$$

where $B_u \in \mathbb{R}$ is such that $\mathrm{dist}(u_k, U^*) \leq B_u$, $\forall k \geq 0$.

Case (b) *If the operator $F(\cdot)$ is p-quasi-sharp with $p > 2$, then for all $k \geq 0$,*

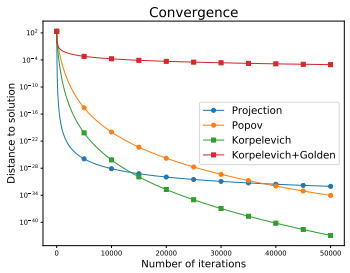
$$D_{k+1}^2 \leq \frac{D_0^2}{(1 + \mu\,\underline{c}\sum_{i=0}^{k-1}\beta_i(D_0^2)^{(p-2)/2})^{2/(p-2)}}. \tag{31}$$

8. Numerical Results

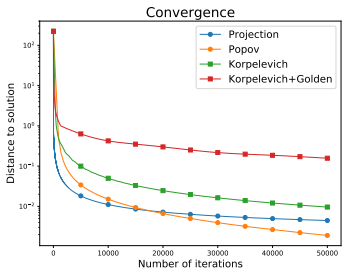
We study the performance of the projection, Korpelevich and Popov methods, for different values of α and p and under Assumptions 2.1 and 2.4. We consider an unconstrained VI($\mathbb{R}^2, F$), where $F(\cdot)$ is an α-symmetric and p-quasi sharp operator from Proposition 2.5. We set parameters of the problem to be $\{(\alpha \approx 0.090., p = 2.1), (\alpha \approx 0.66, p = 4.0), (\alpha \approx 0.86, p = 8.0)\}$. We also compare our results to the Korpelevich method with golden ratio stepsizes (Böhm, 2023).

In Figure 1, we plot the distance to the solution set as a function of the number of iterations. In particular, the stepsizes for the Korpelevich and Popov methods are chosen according to Theorems 5.3 and 6.3, respectively, while that for the projection method is chosen according to Theorem 4.1 with $\beta_k = \frac{100}{100+k}$. We observe that all three methods converge faster than the Korpelevich method with golden-ratio stepsizes for all three choices for α, p. We also notice that when p and α increase, all methods slow down as predicted by the theory; moreover, projection outperforms other methods. This is due to the fact that stepsizes in Theorems 5.3 and 6.3 depend on the parameters of the problem and decrease significantly when the values K_0, K_1, K_2 increase; on the other hand, the stepsizes in Theorem 4.1 for the projection method are independent of parameters.

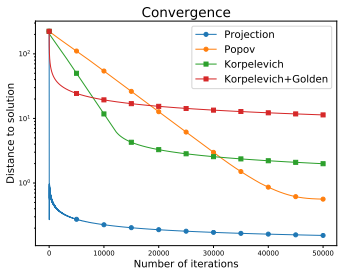
In Figure 2, we compare Korpelevich method with clipping and backtracking (Algorithm 1) with $\beta = 1.0$ and $q = 0.75$ and methods with fully adaptive stepsizes for both Korpelevich (from Theorem 7.1) and Popov; for the latter, we use the following schedule: $\gamma_k = \beta_k \min\{1, \frac{1}{\|F(h_k)\|}, \frac{1}{(\|u_k - h_{k-1}\|+1)^{\alpha/(1-\alpha)}}\}$, with the same $\beta_k = \frac{a}{b+k}$ for all methods, where $a = b = 100$. We choose the parameter $\beta_k = \frac{a}{b+k}$ as it is the simplest diminishing stepsizes one can use in practice. We observe that backtracking and diminishing stepsizes accelerate both methods, and increasing values of the parameters α, p do not affect performance. Additionally, we conducted an experiment on GAN training for a 2D ring dataset. The results of the experiments, showing the advantage of the first-order method with clipping, are included in Appendix G.



(a) $(\alpha \approx 0.09, p = 2.1)$

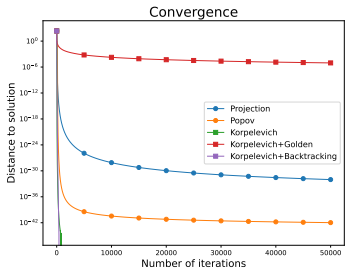


(b) $(\alpha \approx 0.66, p = 4.0)$

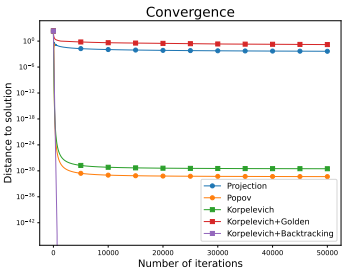


(c) $(\alpha \approx 0.86, p = 8.0)$

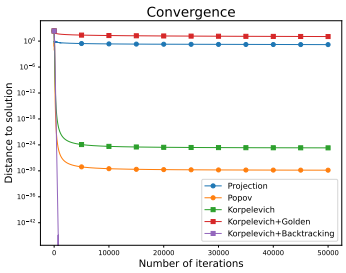
Figure 1. Comparison of projection, Korpelevich and Popov methods with adaptive stepsizes and Korpelevich method with golden ratio stepsizes for different (α, p).



(a) $(\alpha \approx 0.09, p = 2.1)$



(b) $(\alpha \approx 0.66, p = 4.0)$



(c) $(\alpha \approx 0.86, p = 8.0)$

Figure 2. Comparison of projection, Korpelevich and Popov methods with clipping and decreasing stepsizes, Korpelevich method with golden ratio stepsizes and Korpelevich method with clipping and backtracking for different (α, p).

9. Concluding Remarks

For a class of structured non-monotone VI problems with α-symmetric and p-quasi sharp operators, we have proved the first-known convergence results under the weakest assumption on generalized smoothness. Our convergence rates of $O(k^{-2/(p-2)})$ for Korpelevich and Popov methods with adaptive stepsizes improve the existing rate of $O(k^{-1/(p-1)})$ provided in (Wei et al., 2021) under more restrictive assumption on monotonicity and Lipschitz continuity of the operator. Our numerical experiments suggests that for α-symmetric operators, our adaptive stepsizes show more promise than golden ratio stepsizes (Böhm, 2023).

There are a few potential directions for future research: (i) extending our analysis for weak Minty conditions; (ii) analyzing adaptive methods for solving generalized smooth stochastic VIs.

Acknowledgments

We thank the anonymous reviewers for their valuable comments. This work is supported in part by NSF grants CIF-2134256, SCH-2205080 and CIF-2007688.

Impact Statement

This paper presents work whose goal is to advance the field of Machine Learning. There are many potential societal consequences of our work, none of which we feel must be specifically highlighted here.

References

A. Beznosikov, B. Polyak, E. Gorbunov, D. Kovalev, and A. Gasnikov. Smooth monotone stochastic variational inequalities and saddle point problems–survey. *arXiv preprint arXiv:2208.13592*, 2022.

A. Böhm. Solving nonconvex-nonconcave min-max problems exhibiting weak minty solutions. *Trans. Mach. Learn. Res.*, 2023.

A. Brock, J. Donahue, and K. Simonyan. Large scale GAN training for high fidelity natural image synthesis. In *International Conference on Learning Representations*, 2019.

Z. Chen, Y. Zhou, Y. Liang, and Z. Lu. Generalized-smooth nonconvex optimization is as efficient as smooth nonconvex optimization. In *International Conference on Machine Learning*, 2023.

S. Choudhury, E. Gorbunov, and N. Loizou. Single-call stochastic extragradient methods for structured non-monotone variational inequalities: Improved analysis under weaker conditions. *preprint arXiv:2302.14043*, 2023.

C. Daskalakis, A. Ilyas, V. Syrgkanis, and H. Zeng. Training gans with optimism. In *International Conference on Learning Representations*, 2018.

C. Daskalakis, D. J. Foster, and N. Golowich. Independent policy gradient methods for competitive reinforcement learning. *Advances in neural information processing systems*, 33:5527–5540, 2020.

A. Defazio, A. Cutkosky, H. Mehta, and K. Mishchenko. When, why and how much? adaptive learning rate scheduling by refinement. *arXiv preprint arXiv:2310.07831*, 2023.

J. Diakonikolas, C. Daskalakis, and M. I. Jordan. Efficient methods for structured nonconvex-nonconcave min-max optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 2746–2754. PMLR, 2021.

F. Facchinei and J.-S. Pang. *Finite-dimensional variational inequalities and complementarity problems*. Springer, 2003.

I. Gemp and S. Mahadevan. Global convergence to the equilibrium of gans using variational inequalities. *arXiv preprint arXiv:1808.01531*, 2018.

G. Gidel, H. Berard, G. Vignoud, P. Vincent, and S. Lacoste-Julien. A variational inequality perspective on generative adversarial networks. In *International Conference on Learning Representations*, 2019.

E. Gorbunov, H. Berard, G. Gidel, and N. Loizou. Stochastic extragradient: General analysis and improved rates. In *International Conference on Artificial Intelligence and Statistics*, pages 7865–7901. PMLR, 2022a.

E. Gorbunov, M. Danilova, D. Dobre, P. Dvurechenskii, A. Gasnikov, and G. Gidel. Clipped stochastic methods for variational inequalities with heavy-tailed noise. *Advances in Neural Information Processing Systems*, 35: 31319–31332, 2022b.

S. Jelassi, D. Dobre, A. Mensch, Y. Li, and G. Gidel. Dissecting adaptive methods in gans. *arXiv preprint arXiv:2210.04319*, 2022.

T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen, and T. Aila. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 8110–8119, 2020.

D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

G. M. Korpelevich. The extragradient method for finding saddle points and other problems. *Matecon*, 12:747–756, 1976.

G. Kotsalis, G. Lan, and T. Li. Simple and optimal methods for stochastic variational inequalities, i: operator extrapolation. *SIAM Journal on Optimization*, 32(3):2041–2073, 2022.

N. Loizou, H. Berard, G. Gidel, I. Mitliagkas, and S. Lacoste-Julien. Stochastic gradient descent-ascent and consensus optimization for smooth games: Convergence analysis under expected co-coercivity. *Advances in Neural Information Processing Systems*, 34:19095–19108, 2021.

Y. Malitsky. Golden ratio algorithms for variational inequalities. *Mathematical Programming*, 184(1-2):383–410, 2020.

B. T. Polyak. Introduction to optimization. optimization software. *Inc., Publications Division, New York*, 1:32, 1987.

L. D. Popov. A modification of the arrow-hurwicz method for search of saddle points. *Mathematical notes of the Academy of Sciences of the USSR*, 28(5):845–848, 1980.

H. Robbins and D. Siegmund. A convergence theorem for non negative almost supermartingales and some applications. In *Optimizing methods in statistics*, pages 233–257. Elsevier, 1971.

C. Song, Z. Zhou, Y. Zhou, Y. Jiang, and Y. Ma. Optimistic dual extrapolation for coherent non-monotone variational inequalities. *Advances in Neural Information Processing Systems*, 33:14303–14314, 2020.

S. U. Stich. Unified optimal analysis of the (stochastic) gradient method. *preprint arXiv:1907.04232*, 2019.

P. Tseng. On linear convergence of iterative methods for the variational inequality problem. *Journal of Computational and Applied Mathematics*, 60(1-2):237–252, 1995.

D. Vankov, A. Nedić, and L. Sankar. Last iterate convergence of Popov method for non-monotone stochastic variational inequalities. *arXiv preprint arXiv:2310.16910*, 2023.

E.-V. Vlatakis-Gkaragkounis, A. Giannou, Y. Chen, and Q. Xie. Stochastic methods in variational inequalities: Ergodicity, bias and refinements. *arXiv preprint arXiv:2306.16502*, 2023.

C. Wei, C. Lee, M. Zhang, and H. Luo. Linear last-iterate convergence in constrained saddle-point optimization. In *International Conference on Learning Representations*, 2021.

J. Zhang, T. He, S. Sra, and A. Jadbabaie. Why gradient clipping accelerates training: A theoretical justification for adaptivity. In *International Conference on Learning Representations*, 2020.

A. Proof of Proposition 2.5

Proof. Notice that the variational inequality VI($\mathbb{R}^2, F$) has a unique solution $u^* = (0, 0)$. To see this that the operator is p-quasi sharp take an arbitrary $u \in \mathbb{R}^2$. Then, we have:

$$\begin{aligned}\langle F(u), u - u^* \rangle &= \left\langle \begin{bmatrix} \text{sign}(u_1)|u_1|^{p-1} + u_2 \\ \text{sign}(u_2)|u_2|^{p-1} - u_1 \end{bmatrix}, \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\rangle \\ &= |u_1|^p + |u_2|^p \\ &\geq 2^{1-p}|u_1 + u_2|^p \\ &\geq 2^{1-p}\left(\sqrt{u_1^2 + u_2^2}\right)^p = 2^{1-p}\text{dist}^p(u, U^*),\end{aligned}$$

where the first equality holds due to the Jensen inequality for a convex function $|\cdot|^p$, since $p \geq 1$. The second inequality holds because due to $\|\cdot\|_1 \geq \|\cdot\|_2$ and monotonicity of $|\cdot|^p$. Thus, for $p \geq 1$, the operator $F(\cdot)$ is p-quasi sharp with $\mu = 2^{1-p}$.

Next, we show that $F(\cdot)$ is α-symmetric. Consider arbitrary $u, v \in \mathbb{R}^2$. Then, using the triangle inequality we obtain

$$\begin{aligned}\|F(u) - F(v)\| &= \left\| \begin{bmatrix} \text{sign}(u_1)|u_1|^{p-1} + u_2 \\ \text{sign}(u_2)|u_2|^{p-1} - u_1 \end{bmatrix} - \begin{bmatrix} \text{sign}(v_1)|v_1|^{p-1} + v_2 \\ \text{sign}(v_2)|v_2|^{p-1} - v_1 \end{bmatrix} \right\| \\ &\leq \|\phi(u) - \phi(v)\| + \|u - v\|, \end{aligned} \tag{32}$$

where $\phi(u) = \begin{bmatrix} \text{sign}(u_1)|u_1|^{p-1} \\ \text{sign}(u_2)|u_2|^{p-1} \end{bmatrix}$. Notice that, for $p > 2$, the operator $\phi(u)$ is differentiable, and its Jacobian is given by

$$\nabla\phi(u) = (p-1)\begin{bmatrix} |u_1|^{p-2} & 0 \\ 0 & |u_2|^{p-2} \end{bmatrix}.$$

Then, by the mean value theorem:

$$\|\phi(u) - \phi(v)\| \leq \|u - v\| \max_{\theta \in [0,1]} \|\nabla\phi(w_\theta)\|, \tag{33}$$

where $w_\theta = \theta u + (1 - \theta)v$. For the norm of $\nabla\phi(z)$, we have

$$\begin{aligned}\|\nabla\phi(z)\| &= (p-1)\max\{|z_1|^{p-2}, |z_2|^{p-2}\} \\ &= (p-1)\|(|z_1|^{p-2}, |z_2|^{p-2})'\|_\infty \\ &\leq (p-1)\|(|z_1|^{p-2}, |z_2|^{p-2})'\|_2, \end{aligned} \tag{34}$$

where we used the fact that $\|z\|_\infty = \max\{z_1, z_2\}$ and $\|z\|_\infty \leq \|z\|_2$ for an arbitrary z. Expanding further we obtain:

$$\begin{aligned}\|\nabla\phi(z)\| &\leq (p-1)\sqrt{|z_1|^{2(p-2)} + |z_2|^{2(p-2)}} \\ &= (p-1)\sqrt{\left(|z_1|^{2(p-1)}\right)^{(p-2)/(p-1)} + \left(|z_2|^{2(p-1)}\right)^{(p-2)/(p-1)}}. \end{aligned} \tag{35}$$

Since $|\cdot|^{(p-2)/(p-1)}$ is a concave function for $p > 1$, it follows that

$$\begin{aligned}\left(|z_1|^{2(p-1)}\right)^{(p-2)/(p-1)} + \left(|z_2|^{2(p-1)}\right)^{(p-2)/(p-1)} &\leq 2\left(\frac{|z_1|^{2(p-1)} + |z_2|^{2(p-1)}}{2}\right)^{(p-2)/(p-1)} \\ &= 2^{1/(p-1)}\left(|z_1|^{2(p-1)} + |z_2|^{2(p-1)}\right)^{(p-2)/(p-1)}. \end{aligned} \tag{36}$$

Notice that

$$|z_1|^{2(p-1)} + |z_2|^{2(p-1)} = \|(|z_1|^{p-1}, |z_2|^{p-1})'\|^2 = \|\phi(z)\|^2.$$

Combining these facts we obtain for all $z \in \mathbb{R}^2$,

$$\|\nabla\phi(z)\| \leq (p-1)2^{1/2(p-1)}\|\phi(z)\|^{(p-2)/(p-1)}. \tag{37}$$

Now, we show that

$$\|\phi(z)\| \leq 2\|F(z)\| + \sqrt{2}\, 4^{1/(p-2)}. \tag{38}$$

By triangle inequality and the definitions of $F(\cdot)$ and $\phi(\cdot)$, it follows that

$$\|\phi(z)\| \leq \|F(z)\| + \|z\|. \tag{39}$$

For $p = 2$,

$$\begin{aligned} \|z\| &= \sqrt{z_1^2 + z_2^2} \\ &= \frac{1}{\sqrt{2}}\sqrt{(z_1+z_2)^2 + (z_2-z_1)^2} \\ &= \frac{1}{\sqrt{2}}\|F(z)\|. \end{aligned} \tag{40}$$

Combining this equality with the preceding inequality we conclude, that equation (38) holds for $p = 2$. Let $p > 2$, consider two cases, (1) $\|z\| \leq \sqrt{2}\, 4^{1/(p-2)}$, (2) $\|z\| > \sqrt{2}\, 4^{1/(p-2)}$. For case (1), by inequality (39), we see that equation (38) holds. Now, let $\|z\| > \sqrt{2}\, 4^{1/(p-2)}$. Without loss of generality, assume that $|z_1| \geq |z_2|$, then:

$$\begin{aligned} |z_1| &= \|z\|_\infty \\ &\geq \frac{1}{\sqrt{2}}\|z\|_2 \\ &> 4^{1/(p-2)}. \end{aligned} \tag{41}$$

Now, we estimate $\|z\|$, since $|z_1| \geq |z_2|$:

$$\begin{aligned} \|z\|^2 &= |z_1|^2 + |z_2|^2 \\ &= |z_1|^2 + |z_2|^2 + |z_1|^{2(p-1)} - |z_1|^p|z_1|^{p-2} \\ &\leq |z_1|^2 + |z_2|^2 + |z_1|^{2(p-1)} - |z_1|^p|z_1|^{p-2} + |z_2|^{2(p-1)}, \end{aligned} \tag{42}$$

where in equality we add and subtract $|z_1|^{2(p-1)}$, and in the inequality we add $|z_2|^{2(p-1)}$. Since $|z_1| > 4^{1/(p-2)}$, then $|z_1|^p \geq 4|z_1|^2$ and using this fact we obtain:

$$\begin{aligned} \|z\|^2 &\leq |z_1|^2 + |z_2|^2 + |z_1|^{2(p-1)} - |z_1|^p|z_1|^{p-2} + |z_2|^{2(p-1)} \\ &\leq |z_1|^2 + |z_2|^2 + |z_1|^{2(p-1)} - 4|z_1|^2|z_1|^{p-2} + |z_2|^{2(p-1)} \\ &\leq |z_1|^2 + |z_2|^2 + |z_1|^{2(p-1)} - 2|z_1|^2(|z_1|^{p-2} + |z_2|^{p-2}) + |z_2|^{2(p-1)} \\ &\leq |z_1|^2 + |z_2|^2 + |z_1|^{2(p-1)} - 2|z_1||z_2|(|z_1|^{p-2} + |z_2|^{p-2}) + |z_2|^{2(p-1)}, \end{aligned} \tag{43}$$

where in the last two inequalities we used the fact that $|z_1| \geq |z_2|$. Then:

$$\begin{aligned} \|z\|^2 &\leq |z_1|^2 + |z_2|^2 + |z_1|^{2(p-1)} - 2|z_1|^{p-1}|z_2| - 2|z_1||z_2|^{p-1} + |z_2|^{2(p-1)} \\ &\leq |z_1|^2 + |z_2|^2 + |z_1|^{2(p-1)} + 2\text{sign}(z_1)|z_1|^{p-1}|z_2| - 2\text{sign}(z_2)|z_1||z_2|^{p-1} + |z_2|^{2(p-1)} \\ &= \|F(z)\|^2. \end{aligned} \tag{44}$$

Therefore, for case (2), $\|z\| \leq \|F(z)\|$. Then it holds that

$$\begin{aligned} \|\phi(z)\| &\leq \|F(z)\| + \|z\| \\ &\leq 2\|F(z)\| + \sqrt{2}\, 4^{1/(p-2)}. \end{aligned} \tag{45}$$

Using equations (37), (38), and the fact that $\frac{(p-2)}{(p-1)} < 1$ we obtain:

$$\begin{aligned}\|\nabla\phi(w_\theta)\| &\leq (p-1)2^{1/2(p-1)}(2\|F(w_\theta)\| + \sqrt{2}\,4^{1/(p-2)})^{(p-2)/(p-1)} \\ &\leq 2(p-1)2^{1/2(p-1)}\|F(w_\theta)\|^{(p-2)/(p-1)} + (p-1)2^{1/2}4^{1/(p-1)}. \end{aligned} \tag{46}$$

Then by taking maximum over $\theta \in [0,1]$ on both sides we obtain

$$\max_{\theta\in[0,1]} \|\nabla\phi(w_\theta)\| \leq (p-1)2^{1/2}4^{1/(p-1)} + 2(p-1)2^{1/2(p-1)} \max_{\theta\in[0,1]} \|F(w_\theta)\|^{(p-2)/(p-1)}. \tag{47}$$

Combining equations (32), (33), and (47) we obtain

$$\|F(u) - F(v)\| \leq (1 + (p-1)2^{1/2}4^{1/(p-1)} + 2(p-1)2^{1/2(p-1)} \max_{\theta\in[0,1]} \|F(w_\theta)\|^{(p-2)/(p-1)})\|u-v\|. \tag{48}$$

Thus, the operator is α-symmetric with $L_0 = 1 + (p-1)2^{1/2}4^{1/(p-1)}$, $L_1 = 2(p-1)2^{1/2(p-1)}$ and $\alpha = \frac{p-2}{p-1}$. ■

B. Technical Lemmas

In our analysis, we use the properties of the projection operator $P_U(\cdot)$ given in the following lemma.

Lemma B.1. *(Theorem 1.5.5 and Lemma 12.1.13 in (Facchinei and Pang, 2003)) Given a nonempty convex closed set $U \subset \mathbb{R}^m$, the projection operator $P_U(\cdot)$ has the following properties:*

$$\langle v - P_U(v), u - P_U(v)\rangle \leq 0 \quad \textit{for all } u \in U, v \in \mathbb{R}^m, \tag{49}$$

$$\|u - P_U(v)\|^2 \leq \|u - v\|^2 - \|v - P_U(v)\|^2 \quad \textit{for all } u \in U, v \in \mathbb{R}^m, \tag{50}$$

$$\|P_U(u) - P_U(v)\| \leq \|u - v\| \quad \textit{for all } u, v \in \mathbb{R}^m. \tag{51}$$

In the forthcoming analysis, we use Lemma 11 (Polyak, 1987), which is stated below.

Lemma B.2. *[Chapter 2, Lemma 11 (Polyak, 1987)] Let $\{v_k\}, \{z_k\}, \{a_k\}$, and $\{b_k\}$ be nonnegative random scalar sequences such that almost surely for all $k \geq 0$,*

$$\mathbb{E}[v_{k+1} \mid \mathcal{F}_k] \leq (1 + a_k)v_k - z_k + b_k, \tag{52}$$

where $\mathcal{F}_k = \{v_0, \ldots, v_k, z_0, \ldots, z_k, a_0, \ldots, a_k, b_0, \ldots, b_k\}$, and a.s. $\sum_{k=0}^{\infty} a_k < \infty$, $\sum_{k=0}^{\infty} b_k < \infty$. *Then, almost surely, $\lim_{k\to\infty} v_k = v$ for some nonnegative random variable v and $\sum_{k=0}^{\infty} z_k < \infty$.*

As a direct consequence of Lemma B.2, when the sequences $\{v_k\}, \{z_k\}, \{a_k\}, \{b_k\}$ are deterministic, we obtain the following result.

Lemma B.3. *Let $\{\bar{v}_k\}, \{\bar{z}_k\}, \{\bar{a}_k\}, \{\bar{b}_k\}$ be nonnegative scalar sequences such that for all $k \geq 0$,*

$$\bar{v}_{k+1} \leq (1 + \bar{a}_k)\bar{v}_k - \bar{z}_k + \bar{b}_k, \tag{53}$$

where $\sum_{k=0}^{\infty} \bar{a}_k < \infty$ and $\sum_{k=0}^{\infty} \bar{b}_k < \infty$. Then, $\lim_{k\to\infty} \bar{v}_k = \bar{v}$ for some scalar $\bar{v} \geq 0$ and $\sum_{k=0}^{\infty} \bar{z}_k < \infty$.

Lemma B.4. *Let a_1, a_2 be nonnegative scalar and $p > 0$. Then the following inequality holds:*

$$(a_1 + a_2)^p \leq 2^{p-1}(a_1^p + a_2^p).$$

Proof. Let $a = (a_1, a_2), b = (1, 1)$, then by Hölder's inequality:

$$\begin{aligned} a_1 + a_2 &= |\langle a, b\rangle| \\ &\le \|a\|_p \|b\|_{p/(p-1)} \\ &\le (a_1^p + a_2^p)^{1/p}(1+1)^{(p-1)/p}. \end{aligned}$$

Raising the inequality in the power p we get the desired relation. ■

Lemma B.5. *[Chapter 2, Lemma 6 (Polyak, 1987)] Let $\{x_k\}$, $\{a_k\}$ be nonnegative scalar sequences, and $q > 0$ such that for all $k \ge 0$,*

$$x_{k+1} \le x_k - a_k x_k^{1+q}. \tag{54}$$

Then

$$x_k \le \frac{x_0}{(1 + q x_0^q \sum_{i=0}^{k-1} a_i)^{1/q}}. \tag{55}$$

In particular, when $a_k = a$, then

$$x_k \le \frac{x_0}{(1 + a q x_0^q k)^{1/q}}. \tag{56}$$

C. Projection Method Analysis

Proof of Theorem 4.1.

Proof. Let $k \ge 0$ be arbitrary but fixed. From the definition of u_{k+1} in (9), we have $\|u_{k+1} - y\|^2 = \|P_U(u_k - \gamma_k F(u_k)) - y\|^2$ for all $y \in U$. Using the non-expansiveness property of projection operator (51) we obtain for all $y \in U$ and $k \ge 0$,

$$\begin{aligned} \|u_{k+1} - y\|^2 &\le \|u_k - \gamma_k F(u_k) - y\|^2 \\ &= \|u_k - y\|^2 - 2\gamma_k \langle F(u_k), u_k - y\rangle + \gamma_k^2 \|F(u_k)\|^2. \end{aligned} \tag{57}$$

By the definition of the stepsizes (12), $\gamma_k = \beta_k \min\{1, \frac{1}{\|F(u_k)\|}\}$. Thus, we obtain:

$$\begin{aligned} \|u_{k+1} - y\|^2 &\le \|u_k - \gamma_k F(u_k) - y\|^2 \\ &= \|u_k - y\|^2 - 2\gamma_k \langle F(u_k), u_k - y\rangle + \beta_k^2 \min\{1, \frac{1}{\|F(u_k)\|^2}\}\|F(u_k)\|^2 \\ &\le \|u_k - y\|^2 - 2\gamma_k \langle F(u_k), u_k - y\rangle + \beta_k^2. \end{aligned} \tag{58}$$

Plugging in $y = u^* \in U^*$, where u^* is an arbitrary solution, and using p-quasi sharpness we get:

$$\begin{aligned} \|u_{k+1} - u^*\|^2 &\le \|u_k - u^*\|^2 - 2\gamma_k \langle F(u_k), u_k - u^*\rangle + \beta_k^2 \\ &\le \|u_k - u^*\|^2 - 2\gamma_k \mathrm{dist}^p(u_k, U^*) + \beta_k^2. \end{aligned} \tag{59}$$

The equation (59) satisfies the condition of Lemma B.3 with

$$\begin{aligned} \bar{v}_k &= \|u_k - u^*\|^2, \quad \bar{a}_k = 0, \\ \bar{z}_k &= 2\mu\gamma_k \,\mathrm{dist}^p(u_k, U^*), \quad \bar{b}_k = \beta_k^2. \end{aligned} \tag{60}$$

By Lemma B.2, it follows that the sequence $\{\bar{v}_k\}$ converges to a non-negative scalar for any $u^* \in U^*$. Therefore, for all solutions $u^* \in U^*$ the sequence $\{\|u_k - u^*\|\}$ is bounded by $B(u^*)$.

(b) Now, we want to estimate $\|F(u_k)\|$, since this term is present in the denominator of the stepsize. We add and subtract $F(u_0^*)$, where $u_0^* = P_{U^*}(u_0)$ is a projection of u_0 onto the solution set U^*, and get $\|F(u_k)\| = \|F(u_k) - F(u_0^*) + F(u_0^*)\| \le \|F(u_k) - F(u_0^*)\| + \|F(u_0^*)\|$. We can estimate the first term using the α-symmetric assumption on the operator.

Case(I) $\alpha \in (0,1)$

$$\|F(u_k) - F(u^*)\| \leq \|u_k - u_0^*\|(K_0 + K_1\|F(u_0^*)\|^\alpha + K_2\|u_k - u_0^*\|^{\alpha/(1-\alpha)}). \tag{61}$$

Earlier, in part **(a)** we proved that for the solution u^* the following bound hold for any $k \geq 0$

$$\|u_k - u_0^*\| \leq B(u_0^*).$$

Using this fact and equation (61) for we obtain that for all $k \geq 0$:

$$\|F(u_k)\| \leq B(u_0^*)(K_0 + K_1\|F(u_0^*)\|^\alpha + K_2 B(u_0^*)^{\alpha/(1-\alpha)}) + \|F(u_0^*)\|. \tag{62}$$

We showed that the sequence $\{\|F(u_k)\|\}$ is upper bounded by some constant C_1. Where $C_1 = B(u_0^*)(K_0 + K_1\|F(u_0^*)\|^\alpha + K_2 B(u_0^*)^{\alpha/(1-\alpha)}) + \|F(u_0^*)\|$. Using these facts, we conclude that for all $k \geq 0$:

$$\begin{aligned}\gamma_k &= \beta_k \min\{1, \frac{1}{\|F(u_k)\|}\} \\ &\geq \min\{1, \frac{1}{C_1}\}.\end{aligned} \tag{63}$$

Case(II) $\alpha = 1$

For $\alpha = 1$ by Proposition 2.2 we have

$$\|F(u_k) - F(u_0^*)\| \leq \|u_k - u_0^*\|(L_0 + L_1\|F(u_0^*)\|)\exp(L_1\|u_k - u_0^*\|). \tag{64}$$

We can get the following bound for $\|F(u_k)\|$, for all $k \geq 0$:

$$\begin{aligned}\|F(u_k)\| &\leq \|F(u_k) - F(u_0^*)\| + \|F(u_0^*)\| \\ &\leq \|u_k - u_0^*\|(L_0 + L_1\|F(u_0^*)\|)\exp(L_1\|u_k - u_0^*\|) + \|F(u_0^*)\|.\end{aligned} \tag{65}$$

Earlier, in part (a) we proved that for the solution u^* the following bound holds for any $k \geq 0$

$$\|u_k - u^*\| \leq B(u_0^*).$$

Using these facts and equation (65) for all $k \geq 0$:

$$\begin{aligned}\|F(u_k)\| &\leq \|u_k - u_0^*\|(L_0 + L_1\|F(u_0^*)\|)\exp(L_1\|u_k - u_0^*\|) + \|F(u^*)\| \\ &\leq \|u_k - u^*\|(L_0 + L_1\|F(u_0^*)\|)\exp(L_1 B(u_0^*)) + \|F(u_0^*)\|.\end{aligned} \tag{66}$$

We showed that the sequence $\{\|F(u_k)\|\}$ is upper bounded by some constant $\overline{C}_1$, where $\overline{C}_1 = B(u_0^*)(L_0 + L_1\|F(u_0^*)\|)\exp(L_1 B(u_0^*)) + \|F(u_0^*)\|$. Then, we conclude that

$$\begin{aligned}\gamma_k &= \beta_k \min\{1, \frac{1}{\|F(u_k)\|}\} \\ &\geq \beta_k \min\{1, \frac{1}{\overline{C}_1}\}\end{aligned} \tag{67}$$

For both cases $\alpha \in (0,1)$ and $\alpha = 1$ in equations (63), (67) we showed that stepsizes sequence $\{\gamma_k\}$ is lower bounded by $\beta_k \min\{1, \frac{1}{C_1}, \frac{1}{\overline{C}_1}\}$.

(c) Now we show that $\{u_k\}$ converges to a solution of $VI(U, F)$. Using the provided bound on the stepsize sequence γ_k from part (b) and equation (59) we obtain

$$\begin{aligned}\|u_{k+1}-u^*\|^2 &\le \|u_k-u^*\|^2 - 2\gamma_k \mathrm{dist}^p(u_k,U^*) + \beta_k^2 \\ &\le \|u_k-u^*\|^2 - 2\beta_k \min\{1,\frac{1}{C_1},\frac{1}{\overline{C}_1}\}\,\mathrm{dist}^p(u_k,U^*) + \beta_k^2 \end{aligned} \tag{68}$$

The equation (68) satisfies the condition of Lemma B.3 with

$$\begin{aligned} \bar{v}_k &= \|u_k-u^*\|^2, \quad \bar{a}_k = 0, \\ \bar{z}_k &= 2\mu\beta_k \min\{1,\frac{1}{C_1},\frac{1}{\overline{C}_1}\}\,\mathrm{dist}^p(u_k,U^*), \quad \bar{b}_k = \beta_k^2. \end{aligned} \tag{69}$$

Then

$$\sum_{k=0}^{\infty} \beta_k \mathrm{dist}^p(u_k,U^*) < \infty.$$

Since, $\sum_{k=0}^{\infty}\beta_k = \infty$ then it follows that

$$\liminf_{k\to\infty} \mathrm{dist}^p(u_k,U^*) = 0. \tag{70}$$

Since $\bar{v}_k$ converges for any given $u^* \in U^*$ we can conclude that $\|u_k - u^*\|$ converges for all $u^* \in U^*$. Therefore, the sequence $\{u_k\}$ is bounded and has accumulation points. Let $\{k_i\}$ be an index sequence, such that

$$\lim_{i\to\infty} \mathrm{dist}^p(u_{k_i},U^*) = \liminf_{k\to\infty} \mathrm{dist}^p(u_k,U^*) = 0. \tag{71}$$

We assume, that the sequence $\{u_{k_i}\}$ is convergent with a limit point $\bar{u}$, otherwise, we choose a convergent subsequence. Therefore,

$$\lim_{i\to\infty} \|u_{k_i} - \bar{u}\| = 0. \tag{72}$$

Then by (71), $\mathrm{dist}(\bar{u},U^*) = 0$, thus $\bar{u} \in U^*$ since U^* is closed. And since the sequence $\{\|u-u^*\|\}$ converges for all $u^* \in U^*$ and by (72) we have

$$\lim_{k\to\infty} \|u_k - \bar{u}\| = 0. \tag{73}$$

■

D. Korpelevich Method Analysis

Proof of Lemma 5.1.

Proof. Let $k \ge 0$ be arbitrary but fixed. By the definition of u_{k+1} in (10), we have $\|h_{k+1}-y\|^2 = \|P_U(h_k-\gamma_k F(u_k))-y\|$ for any $y \in U$. Using the projection inequality (50) of Lemma B.1 we obtain for any $y \in U$,

$$\begin{aligned}\|h_{k+1}-y\|^2 &\le \|h_k-\gamma_k F(u_k)-y\|^2 - \|h_{k+1}-h_k+\gamma_k F(u_k)\|^2 \\ &\le \|h_k-y\|^2 - \|h_{k+1}-h_k\|^2 + 2\gamma_k\langle F(u_k), y-h_{k+1}\rangle. \end{aligned} \tag{74}$$

Next, we consider the term $\|h_{k+1}-h_k\|^2$, where we add and subtract u_k, and thus get

$$\|h_{k+1}-h_k\|^2 = \|h_{k+1}-u_k\| + \|h_k-u_k\|^2 - 2\langle h_{k+1}-u_k, h_k-u_k\rangle. \tag{75}$$

Adding and subtracting $2\gamma_k\langle F(h_k), u_k-h_{k+1}\rangle$, and combining (74) and (75) we obtain

$$\begin{aligned}\|h_{k+1}-y\|^2 &\le \|h_k-y\|^2 - \|h_{k+1}-u_k\|^2 - \|h_k-u_k\|^2 + 2\langle h_{k+1}-u_k, h_k-u_k\rangle \\ &\quad + 2\gamma_k\langle F(u_k), y-u_k+u_k-h_{k+1}\rangle + 2\gamma_k\langle F(h_k), u_k-h_{k+1}\rangle - 2\gamma_k\langle F(h_k), u_k-h_{k+1}\rangle \\ &\le \|h_k-y\|^2 - \|h_{k+1}-u_k\|^2 - \|h_k-u_k\|^2 + 2\langle h_{k+1}-u_k, h_k-\gamma_k F(h_k)-u_k\rangle \\ &\quad + 2\gamma_k\langle F(u_k), y-u_k\rangle + 2\gamma_k\langle F(h_k)-F(u_k), h_{k+1}-u_k\rangle. \end{aligned} \tag{76}$$

By the projection inequality (49) where $u = h_{k+1}$, and $v = h_k - \gamma_k F(h_k)$ we obtain that

$$2\langle h_{k+1} - u_k, h_k - \gamma_k F(h_k) - u_k\rangle \leq 0.$$

Now using Cauchy-Schwartz inequality and well know relation $2ab \leq a^2 + b^2$ for $a, b \in \mathbb{R}$, we obtain:

$$\begin{aligned} 2\gamma_k\langle F(h_k) - F(u_k), h_{k+1} - u_k\rangle &\leq 2\gamma_k\|F(h_k) - F(u_k)\|\|h_{k+1} - u_k\| \\ &\leq \gamma_k^2\|F(h_k) - F(u_k)\|^2 + \|h_{k+1} - u_k\|^2. \end{aligned} \tag{77}$$

Combining the preceding two estimates with (76), we obtain the desired result:

$$\|h_{k+1} - y\|^2 \leq \|h_k - y\|^2 - \|h_k - u_k\|^2 - 2\gamma_k\langle F(u_k), u_k - y\rangle + \gamma_k^2\|F(h_k) - F(u_k)\|^2. \tag{78}$$

■

Proof of Theorem 5.2

Proof. **(a)** By Lemma 5.1, the following inequality holds for any $y \in U$ and all $k \geq 0$:

$$\|h_{k+1} - y\|^2 \leq \|h_k - y\|^2 - \|h_k - u_k\|^2 - 2\gamma_k\langle F(u_k), u_k - y\rangle + \gamma_k^2\|F(h_k) - F(u_k)\|^2. \tag{79}$$

We want to estimate the last term on the LHS of the inequality using the fact that operator $F(\cdot)$ is an α-symmetric operator.

Case (I) $\alpha \in (0, 1)$.

Based on the alternative characterization of α-symmetric operators from Proposition 2.2(a) (as given in (6)), when $\alpha \in (0, 1)$, the next inequality holds for any $k \geq 1$:

$$\|F(h_k) - F(u_k)\| \leq \|h_k - u_k\|(K_0 + K_1\|F(h_k)\|^\alpha + K_2\|h_k - u_k\|^{\alpha/(1-\alpha)}). \tag{80}$$

By the projection property (50) and the stepsize choice (15) :

$$\|h_k - u_k\| \leq \gamma_k\|F(h_k)\| \leq 1.$$

Then by the well know relation $(\sum_{i=1}^m a_i)^2 \leq m\sum_{i=1}^m a_i^2$ we obtain:

$$\begin{aligned} \|F(h_k) - F(u_k)\|^2 &\leq \|h_k - u_k\|^2(K_0 + K_1\|F(h_k)\|^\alpha + K_2)^2 \\ &\leq 3K_0^2\|h_k - u_k\|^2 + 3K_1^2\|F(h_k)\|^{2\alpha}\|h_k - u_k\|^2 + 3K_2^2\|h_k - u_k\|^2. \end{aligned} \tag{81}$$

Based on the step-size choice: $\gamma_k \leq \frac{1}{3\sqrt{2}K_0}$, $\gamma_k \leq \frac{1}{3\sqrt{2}K_1\|F(h_k)\|^\alpha}$, $\gamma_k \leq \frac{1}{3\sqrt{2}K_2}$ and then:

$$\begin{aligned} \gamma_k^2\|F(h_k) - F(u_k)\|^2 &\leq \frac{1}{6}\|h_k - u_k\|^2 + \frac{1}{6}\|h_k - u_k\|^2 + \frac{1}{6}\|h_k - u_k\|^2 \\ &= \frac{1}{2}\|h_k - u_k\|^2. \end{aligned} \tag{82}$$

Case (II) $\alpha = 1$.

Based on the alternative characterization of α-symmetric operators from Proposition 2.2(b) (as given in (7)), when $\alpha = 1$, the next inequality holds for any $k \geq 0$:

$$\|F(h_k) - F(u_k)\| \leq \|h_k - u_k\|(L_0 + L_1\|F(h_k)\|)\exp(L_1\|h_k - u_k\|). \tag{83}$$

By the projection property, triangle inequality and step size choice (16):

$$\|h_k - u_k\| \leq \gamma_k\|F(h_k)\| \leq \frac{1}{2\sqrt{2}eL_1} < \frac{1}{L_1}.$$

Using the well know relation $(\sum_{i=1}^m a_i)^2 \leq m \sum_{i=1}^m a_i^2$ we obtain:

$$\begin{aligned}\|F(h_k) - F(u_k)\|^2 &\leq \|h_k - u_k\|^2 (L_0 + L_1\|F(h_k)\|)^2 e^2 \\ &\leq 2e^2 L_0^2 \|h_k - u_k\|^2 + 2L_1^2 e^2 \|F(h_k)\|^2 \|h_k - u_k\|^2.\end{aligned} \tag{84}$$

Using the step-size choice: $\gamma_k \leq \frac{1}{2\sqrt{2}eL_0}$, $\gamma_k \leq \frac{1}{2\sqrt{2}L_1 e\|F(h_k)\|}$, it follows that:

$$\gamma_k^2 \|F(h_k) - F(u_k)\|^2 \leq \frac{1}{2}\|h_k - u_k\|^2. \tag{85}$$

For both cases $\gamma_k^2\|F(h_k) - F(u_k)\|^2 \leq \frac{1}{2}\|h_k - u_k\|^2$, then combining this fact with (79) we obtain that for any $k \geq 0$:

$$\|h_{k+1} - y\|^2 \leq \|h_k - y\|^2 - \frac{1}{2}\|u_k - h_k\|^2 - 2\gamma_k \langle F(u_k), u_k - y\rangle. \tag{86}$$

Now we plug $y = u^*$ into equation (86), where $u^* \in U^*$ is an arbitrary solution. Thus, by using p-quasi sharpness of the operator, we obtain the following recursive inequality:

$$\|h_{k+1} - u^*\|^2 \leq \|h_k - u^*\|^2 - \frac{1}{2}\|u_k - h_k\|^2 - 2\gamma_k \mu \text{dist}^p(u_k, U^*). \tag{87}$$

From this inequality, we conclude that for any $u^* \in U^*$, the sequence $\{\|h_k - u^*\|\}$ is bounded.

(b) Now, we want to estimate $\|F(h_k)\|$, since this term is present in the denominator of the stepsize. Let $h_0^* = P_{U^*}(h_0)$ be a projection of h_0 onto the solution set U^*. We add and subtract $F(h_0^*)$, and get $\|F(h_k)\| = \|F(h_k) - F(h_0^*) + F(h_0^*)\| \leq \|F(h_k) - F(h_0^*)\| + \|F(h_0^*)\|$. We say that $\|F(h_0^*)\| = D_2$ where $D_2 > 0$ is some constant. We can estimate the first term using the α-symmetric assumption on the operator class.

Case(I) $\alpha \in (0, 1)$

Based on the alternative characterization of α-symmetric operators from Proposition 2.2(a) (as given in (6)), when $\alpha \in (0, 1)$, the next inequality holds for any $k \geq 0$:

$$\|F(h_k) - F(h_0^*)\| \leq \|h_k - h_0^*\|(K_0 + K_1\|F(h_0^*)\|^\alpha + K_2\|h_k - h_0^*\|^{\alpha/(1-\alpha)}). \tag{88}$$

Earlier, in equation (87) we proved that for arbitrary solution u^* the following bound hold for any $k \geq 0$

$$\|h_{k+1} - u^*\|^2 \leq \|h_0 - u^*\|^2.$$

Let $R_0 = \|h_0 - h_0^*\|^2$, then

$$\|h_k - h_0^*\| \leq \sqrt{R_0}.$$

Using this fact and equation (88) for we obtain that for all $k \geq 0$:

$$\|F(h_k)\| \leq \sqrt{R_0}(K_0 + K_1 D_2^\alpha + K_2 \sqrt{R_0}^{\alpha/(1-\alpha)}) + D_2. \tag{89}$$

We showed that the sequence $\{\|F(h_k)\|\}$ is upper bounded by some constant C_1. Where $C_1 = \sqrt{R_0}(K_0 + K_1 D_2^\alpha + K_2\sqrt{R_0}^{\alpha/(1-\alpha)}) + D_2$. Since function x^a is monotone for any $x > 0$, and $a > 0$ then $\|F(h_k)\|^\alpha \leq C_1^\alpha$. Using these facts, we conclude that

$$\begin{aligned}\gamma_k &= \min\{\frac{1}{4\mu}, \frac{1}{3\sqrt{2}K_0}, \frac{1}{\|F(h_k)\|}, \frac{1}{3\sqrt{2}K_1\|F(h_k)\|^\alpha}, \frac{1}{3\sqrt{2}K_2}\} \\ &\geq \min\{\frac{1}{4\mu}, \frac{1}{3\sqrt{2}K_0}, \frac{1}{C_1}, \frac{1}{3\sqrt{2}K_1 C_1^\alpha}, \frac{1}{3\sqrt{2}K_2}\}\end{aligned} \tag{90}$$

Case(II) $\alpha = 1$ Based on the alternative characterization of α-symmetric operators from Proposition 2.2(b) (as given in (7)), when $\alpha = 1$, the next inequality holds for any $k \geq 0$:

$$\|F(h_k) - F(h_0^*)\| \leq \|h_k - h_0^*\|(L_0 + L_1\|F(h_0^*)\|)\exp(L_1\|h_k - h_0^*\|). \tag{91}$$

We can get the following bound for $\|F(h_{k-1})\|$, for all $k \geq 1$:

$$\begin{aligned}\|F(h_k)\| &\leq \|F(h_k) - F(h_0^*)\| + \|F(h_0^*)\| \\ &\leq \|h_k - h_0^*\|(L_0 + L_1\|F(h_0^*)\|)\exp(L_1\|h_k - h_0^*\|) + \|F(h_0^*)\|\end{aligned} \tag{92}$$

Earlier, in equation (87) we proved that for arbitrary solution u^* the following bound hold for any $k \geq 0$

$$\|h_k - u^*\|^2 \leq \|h_0 - u^*\|^2.$$

Let $R_0 = \|h_0 - h_0^*\|^2$, Then, it follows that for all $k \geq 1$:

$$\|h_k - h_0^*\| \leq \sqrt{R_0}.$$

Using these facts and equation (91) for all $k \geq 1$:

$$\begin{aligned}\|F(h_k)\| &\leq \|h_k - h_0^*\|(L_0 + L_1\|F(h_0^*)\|)\exp(L_1\|h_k - h_0^*\|) + \|F(h_0^*)\| \\ &\leq \sqrt{R_0}(L_0 + L_1 D_2)\exp(L_1\sqrt{R_0}) + D_2.\end{aligned} \tag{93}$$

We showed that the sequence $\{\|F(h_k)\|\}$ is upper bounded by some constant $\overline{C}_1$, where $\overline{C}_1 = \sqrt{R_1}(L_0 + L_1 D_2)\exp(L_1\sqrt{R_0}) + D_2$. Then, we conclude that

$$\begin{aligned}\gamma_k &= \min\{\frac{1}{4\mu}, \frac{1}{2\sqrt{2}eL_0}, \frac{1}{2\sqrt{2}eL_1\|F(h_k)\|}\} \\ &\geq \min\{\frac{1}{4\mu}, \frac{1}{2\sqrt{2}eL_0}, \frac{1}{2\sqrt{2}eL_1\overline{C}_1}\}\end{aligned} \tag{94}$$

For both cases $\alpha \in (0, 1)$ and $\alpha = 1$ in equations (89), (94) we showed that stepsizes sequence $\{\gamma_k\}$ is lower bounded by some constant $\underline{\gamma}$. Where

$$\underline{\gamma} = \min\{\frac{1}{4\mu}, \frac{1}{3\sqrt{2}K_0}, \frac{1}{C_1}, \frac{1}{3\sqrt{2}K_1 C_1^{\alpha}}, \frac{1}{3\sqrt{2}K_2}\} \text{ for } \alpha \in (0, 1),$$

$$\underline{\gamma} = \min\{\frac{1}{4\mu}, \frac{1}{\overline{C}_1}, \frac{1}{2\sqrt{2}eL_0}, \frac{1}{2\sqrt{2}eL_1\overline{C}_1}\} \text{ for } \alpha = 1,$$

i.e. $\gamma_k \geq \underline{\gamma}$ for all $k \geq 0$.

(c) Based on the results from part (b), for both cases ($\alpha \in (0, 1)$ and $\alpha = 1$) there exists some positive constant $\underline{\gamma}$ such that $\underline{\gamma} \leq \gamma_k$ for all $k \geq 0$. Using this fact and equation (87), we obtain for any $k \geq 1$:

$$\|h_{k+1} - u^*\|^2 \leq \|h_k - u^*\|^2 - \frac{1}{2}\|u_k - h_k\|^2 - 2\underline{\gamma}\mu\text{dist}^p(u_k, U^*). \tag{95}$$

The equation (95) satisfies the conditions of Lemma B.3 with

$$\begin{aligned}\bar{v}_k &= \|h_k - h^*\|^2, \quad \bar{a}_k = 0, \\ \bar{z}_k &= \frac{1}{2}\|u_k - h_k\|^2 + 2\mu\underline{\gamma}\,\text{dist}^p(u_k, U^*), \quad \bar{b}_k = 0.\end{aligned} \tag{96}$$

By Lemma B.2, it follows that the sequence $\{\bar{v}_k\}$ converges to a non-negative scalar for any $u^* \in U^*$, and we have

$$\sum_{k=0}^{\infty} \operatorname{dist}^p(u_k, U^*) < \infty, \quad \sum_{k=0}^{\infty} \|u_k - h_k\|^2 < \infty.$$

Thus, it follows that

$$\lim_{k\to\infty} \operatorname{dist}^p(u_k, U^*) = 0. \tag{97}$$

$$\lim_{k\to\infty} \|u_k - h_k\| = 0. \tag{98}$$

Since $\bar{v}_k$ converges for any given $u^* \in U^*$ we can conclude that $\|h_k - u^*\|$ converges for all $u^* \in U^*$. Therefore, the sequence $\{h_k\}$ is bounded and has accumulation points. In view of relation (98), the sequences $\{u_k\}$ and $\{h_k\}$ have the same accumulation points. Then, there exists a convergent subsequnce $\{h_{k_i}\}$, such that $\bar{u} \in U$ is its limit point, i.e.,

$$\lim_{i\to\infty} \|h_{k_i} - \bar{u}\| = 0. \tag{99}$$

By relation (98), it follows that

$$\lim_{i\to\infty} \|u_{k_i} - \bar{u}\| = 0.$$

By the continuity of the distance function $\operatorname{dist}(\cdot, U^*)$, from relation (97) we conclude that $\operatorname{dist}(\bar{u}, U^*) = 0$, which implies that $\bar{u} \in U^*$ since the set U^* is closed. Since the sequence $\{\|h_k - u^*\|^2\}$ converges for any $u^* \in U^*$, it follows that $\{\|h_k - \bar{u}\|^2\}$ converges , and by relation (99) we conclude that $\lim_{k\to\infty} \|h_k - \bar{u}\|^2 = 0$.

■

Proof of Theorem 5.3

Proof. Since the solution set U^* is closed, there exists a projection h_k^* of the iterate h_k on the set U^*, i.e., there exists a point $h_k^* \in U^*$ such that $\|h_k - h_k^*\| = \operatorname{dist}(h_k, U^*)$. Thus, by letting in $u^* = h_k^*$ in (17) (the result of Theorem 5.2(a)), we obtain for all $k \geq 0$,

$$\|h_{k+1} - h_k^*\|^2 \leq \|h_k - h_k^*\|^2 - \frac{1}{2}\|u_k - h_k\|^2 - 2\mu\gamma_k \operatorname{dist}^p(u_k, U^*). \tag{100}$$

In view of $\|h_k - h_k^*\| = \operatorname{dist}(h_k, U^*)$ and $\operatorname{dist}(h_{k+1}, U^*) \leq \|h_{k+1} - h_k^*\|$, it follows that for all $k \geq 1$,

$$\operatorname{dist}^2(h_{k+1}, U^*) \leq \operatorname{dist}^2(h_k, U^*) - \frac{1}{2}\|u_k - h_k\|^2 - 2\mu\gamma_k \operatorname{dist}^p(u_k, U^*). \tag{101}$$

Next, we estimate the term $\operatorname{dist}^p(u_k, U^*)$ in (101). By the triangle inequality, we have

$$\|h_k - u^*\| \leq \|u_k - h_k\| + \|u_k - u^*\| \qquad \text{for all } u^* \in U^*,$$

and by taking the minimum over $u^* \in U^*$ on both sides of the preceding relation, we obtain

$$\operatorname{dist}(h_k, U^*) \leq \|u_k - h_k\| + \operatorname{dist}(u_k, U^*). \tag{102}$$

Case (a) $p \leq 2$ Squaring both sides of inequality (102), and using $(\sum_{i=1}^m a_i)^2 \leq m \sum_{i=1}^m a_i^2$, which is valid for any scalars a_i, $i = 1, \ldots, m$, and any integer $m \geq 1$, we obtain

$$-2\operatorname{dist}^2(u_k, U^*) \leq 2\|u_k - h_k\|^2 - \operatorname{dist}^2(h_k, U^*). \tag{103}$$

In Theorem 5.2 we showed that the sequence $\{u_k\}$ convergence to a solution $\bar{u} \in U^*$. Hence, there exists $B_u \in \mathbb{R}$ such that the sequence $\operatorname{dist}(u_k, U^*)$ is bounded by B_u. Then, since $p \leq 2$, it holds that for all $k \geq 0$

$$\operatorname{dist}^{2-p}(u_k, U^*) \leq (B_u)^{2-p},$$

and consequently,

$$\text{dist}^2(u_k, U^*) = \text{dist}^p(u_k, U^*)\text{dist}^{2-p}(u_k, U^*) \leq (B_u)^{2-p}\text{dist}^p(u_k, U^*).$$

Thus,

$$-\text{dist}^p(u_k, U^*) \leq -\frac{1}{(B_u)^{2-p}}\text{dist}^2(u_k, U^*)$$

Using the preceding inequality in equation (103) we obtain the following bound:

$$-\text{dist}^p(u_k, U^*) \leq -\frac{1}{(B_u)^{2-p}}\text{dist}^2(u_k, U^*) \leq \frac{1}{(B_u)^{2-p}}\|u_k - h_k\|^2 - \frac{1}{2(B_u)^{2-p}}\text{dist}^2(h_k, U^*) \tag{104}$$

Combining equation (104) with (101), we get that for any $k \geq 0$:

$$\text{dist}^2(h_{k+1}, U^*) \leq (1 - \frac{\mu}{(B_u)^{2-p}}\gamma_k)\text{dist}^2(h_k, U^*) - \frac{1}{2}\|u_k - h_k\|^2 + \frac{2\mu}{(B_u)^{2-p}}\gamma_k \|u_k - h_k\|^2. \tag{105}$$

Without loss of generality we consider case when $B_u \geq 1$ (otherwise, we can choose B_u = 1), then:

$$\text{dist}^2(h_{k+1}, U^*) \leq (1 - \frac{\mu}{(B_u)^{2-p}}\gamma_k)\text{dist}^2(h_k, U^*) - \frac{1}{2}\|u_k - h_k\|^2 + 2\mu\gamma_k \|u_k - h_k\|^2. \tag{106}$$

Based on stepsizes choice $\gamma_k\mu \leq \frac{1}{4}$ and based on the result of Theorem 5.2 (b) the stepsizes sequence $\{\gamma_k\}$ is lower bounded, i.e., $\gamma_k \geq \underline{\gamma}$ for all $k \geq 0$. Then, for all $k \geq 0$:

$$\text{dist}^2(h_{k+1}, U^*) \leq (1 - \frac{\mu}{(B_u)^{2-p}}\underline{\gamma})\text{dist}^2(h_k, U^*). \tag{107}$$

Thus, by defining $D_k^2 = \text{dist}^2(h_k, U^*)$ we obtain that for all $k \geq 0$,

$$D_{k+1}^2 \leq (1 - \frac{\mu}{(B_u)^{2-p}}\underline{\gamma})D_k^2. \tag{108}$$

Case (b) $p > 2$ By using Lemma B.4, we further obtain

$$\begin{aligned}\text{dist}^p(h_k, U^*) &\leq (\|u_k - h_k\| + \text{dist}(u_k, U^*))^p \\ &\leq 2^{p-1}\|u_k - h_k\|^p + 2^{p-1}\,\text{dist}^p(u_k, U^*).\end{aligned} \tag{109}$$

Using projection inequality (50), and stepsizes choice we obtain:

$$\|u_k - h_k\| \leq \|\gamma_k F(h_k)\| \leq 1.$$

Combining this result with equation (109) we get:

$$\begin{aligned}\text{dist}^p(h_k, U^*) &\leq 2^{p-1}\|u_k - h_k\|^p + 2^{p-1}\text{dist}^p(u_k, U^*) \\ &\leq 2^{p-1}\|u_k - h_k\|^2 + 2^{p-1}\,\text{dist}^p(u_k, U^*).\end{aligned} \tag{110}$$

And by rearranging terms, we obtain the following:

$$-\text{dist}^p(u_k, U^*) \leq \|u_k - h_k\|^2 - 2^{-p+1}\,\text{dist}^p(h_k, U^*). \tag{111}$$

Combining this inequality with (101), we obtain that for any $k \geq 1$:

$$\text{dist}^2(u_{k+1}, U^*) \leq \text{dist}^2(u_k, U^*) - (\frac{1}{2} - 2\mu\gamma_k)\|u_k - h_k\|^2 - 2^{-p+2}\gamma_k\mu\,\text{dist}^p(h_k, U^*). \tag{112}$$

By the stepsizes choice $\gamma_k \leq \frac{1}{4\mu}$ then $(\frac{1}{2} - 2\mu\gamma_k)\|u_k - h_k\|^2 \leq 0$, and $\gamma_k \geq \underline{\gamma}$ for all $k \geq 0$. Thus, by defining $D_k^2 = \text{dist}^2(h_k, U^*)$ we obtain that for all $k \geq 0$,

$$D_{k+1}^2 \leq D_k^2 - 2^{-p+2}\underline{\gamma}\mu\,(D_k^2)^{p/2}. \tag{113}$$

Using Lemma B.5 with $x_k = D_k^2$, $q = \frac{p-2}{2}$, $a_k = 2^{-p+2}\underline{\gamma}$, we obtain, for all $k \geq 0$:

$$D_{k+1}^2 \leq \frac{D_0^2}{(1 + (p-2)\underline{\gamma}2^{-p+1}(D_0^2)^{(p-2)/2}(k+1))^{2/(p-2)}}. \tag{114}$$

■

E. Popov Method Analysis

Proof of Lemma 6.1

Proof. Let $k \geq 1$ be arbitrary but fixed. From the definition of u_{k+1} in (11), we have $\|u_{k+1}-y\|^2 = \|P_U(u_k-\gamma_k F(h_k))-y\|$ for any $y \in U$. Using the projection inequality (50) of Lemma B.1 we obtain for any $y \in U$,

$$\begin{aligned}\|u_{k+1}-y\|^2 &\leq \|u_k-\gamma_k F(h_k)-y\|^2 - \|u_k-\gamma_k F(h_k)-u_{k+1}\|^2 \\ &= \|u_k-y\|^2 - \|u_{k+1}-u_k\|^2 - 2\gamma_k\langle F(h_k), u_{k+1}-y\rangle.\end{aligned} \tag{115}$$

We next consider the term $\|u_{k+1}-u_k\|^2$, where we add and subtract h_k, and thus obtain

$$\begin{aligned}\|u_{k+1}-u_k\|^2 =&\|(u_{k+1}-h_k)-(u_k-h_k)\|^2 \\ =&\|u_{k+1}-h_k\|^2 + \|u_k-h_k\|^2 - 2\langle u_k-h_k, u_{k+1}-h_k\rangle \\ =&\|u_{k+1}-h_k\|^2 + \|u_k-h_k\|^2 - 2\langle u_k-\gamma_k F(h_{k-1})-h_k, u_{k+1}-h_k\rangle \\ &- 2\gamma_k\langle F(h_{k-1}), u_{k+1}-h_k\rangle,\end{aligned} \tag{116}$$

where the last equality is obtained by adding and subtracting $2\gamma_k\langle F(h_{k-1},), u_{k+1}-h_k\rangle$. Next, we use another projection property (49) with $v = u_k-\gamma_k F(h_{k-1}), u = u_{k+1}$ and obtain:

$$\langle u_k-\gamma_k F(h_{k-1})-h_k, u_{k+1}-h_k\rangle \leq 0. \tag{117}$$

Therefore,

$$\|u_{k+1}-u_k\|^2 \geq \|u_{k+1}-h_k\|^2 + \|u_k-h_k\|^2 - 2\gamma_k\langle F(h_{k-1}), u_{k+1}-h_k\rangle. \tag{118}$$

Combining (115) and (118) we can see that for any $y \in U$,

$$\begin{aligned}\|u_{k+1}-y\|^2 \leq&\|u_k-y\|^2 - \|u_{k+1}-h_k\|^2 - \|u_k-h_k\|^2 - 2\gamma_k\langle F(h_k), u_{k+1}-h_k\rangle \\ &- 2\gamma_k\langle F(h_k), h_k-y\rangle + 2\gamma_k\langle F(h_{k-1}), u_{k+1}-h_k\rangle \\ =&\|u_k-y\|^2 - \|u_{k+1}-h_k\|^2 - \|u_k-h_k\|^2 - 2\gamma_k\langle F(h_k), h_k-y\rangle \\ &+ 2\gamma_k\langle F(h_{k-1})-F(h_k), u_{k+1}-h_k\rangle.\end{aligned} \tag{119}$$

To estimate the last inner product in (119), we write

$$\langle F(h_{k-1})-F(h_k), u_{k+1}-h_k\rangle \leq \|F(h_{k-1})-F(h_k)\|\,\|u_{k+1}-h_k\|.$$

From the definitions of u_{k+1} and h_{k+1} in (11), we have $u_{k+1} = P_U(u_k-\gamma_k F(h_k))$ and $h_k = P_U(u_k-\gamma_k F(h_{k-1}))$. Thus, by using the Lipschitz continuity of the projection operator (see equation (51) in Lemma 1), we obtain $\|u_{k+1}-h_k\| \leq \gamma_k\|F(h_{k-1})-F(h_k)\|$, implying that

$$\langle F(h_{k-1})-F(h_k), u_{k+1}-h_k\rangle \leq \gamma_k\|F(h_{k-1})-F(h_k)\|^2.$$

Upon substituting the preceding estimate back in relation (119), we obtain the desired relation:

$$\begin{aligned}\|u_{k+1}-y\|^2 \leq&\|u_k-y\|^2 - \|u_{k+1}-h_k\|^2 - \|u_k-h_k\|^2 - 2\gamma_k\langle F(h_k), h_k-y\rangle \\ &+ 2\gamma_k^2\|F(h_k)-F(h_{k-1})\|^2.\end{aligned} \tag{120}$$

■

Proof of Theorem 6.2

Proof. By Lemma 6.1 we have that the following inequality holds for all $k \geq 1$ and all $y \in U$,

$$\begin{aligned}\|u_{k+1}-y\|^2 \leq&\|u_k-y\|^2 - \|u_{k+1}-h_k\|^2 - \|u_k-h_k\|^2 - 2\gamma_k\langle F(h_k), h_k-y\rangle \\ &+ 2\gamma_k^2\|F(h_k)-F(h_{k-1})\|^2.\end{aligned} \tag{121}$$

We want to estimate the last term on the LHS of the inequality (121) using the fact that the operator $F(\cdot)$ is an α-symmetric operator.

Case (I) $\alpha \in (0,1)$

Using the definition of α-symmetric operator (as given in (6)), we obtain that for all $k \geq 1$,

$$\begin{aligned}\|F(h_k) - F(h_{k-1})\|^2 &\leq \|h_k - h_{k-1}\|^2 (K_0 + K_1\|F(h_{k-1})\|^\alpha + K_2\|h_k - h_{k-1}\|^{\alpha/(1-\alpha)})^2 \\ &\leq 3K_0^2\|h_k - h_{k-1}\|^2 + 3K_1^2\|F(h_{k-1})\|^{2\alpha}\|h_k - h_{k-1}\|^2 \\ &\quad + 3K_2^2\|h_k - h_{k-1}\|^{2\alpha/(1-\alpha)}\|h_k - h_{k-1}\|^2,\end{aligned} \tag{122}$$

where the last inequality is obtained by using relation $(\sum_{i=1}^m a_i)^2 \leq m\sum_{i=1}^m a_i^2$.

Using the definition of the iterate updates (11), the projection property (50) for $v = h_{k-1} - \gamma_k F(h_{k-1}), u = h_{k-1}$ and triangle inequality we get:

$$\|h_k - h_{k-1}\| \leq \|u_k - h_{k-1} - \gamma_k F(h_{k-1})\| \leq \|u_k - h_{k-1}\| + \gamma_k\|F(h_{k-1})\|.$$

By the definition of the stepsize γ_k, we have that $\gamma_k \leq \frac{1}{\|F(h_{k-1})\|}$. Therefore, it follows that

$$\|h_k - h_{k-1}\| \leq \|u_k - h_{k-1}\| + \gamma_k\|F(h_{k-1})\| \leq \|u_k - h_{k-1}\| + 1. \tag{123}$$

Combining this inequality with (122) we obtain:

$$\begin{aligned}\|F(h_k) - F(h_{k-1})\|^2 &\leq 3K_0^2\|h_k - h_{k-1}\|^2 + 3K_1^2\|F(h_{k-1})\|^{2\alpha}\|h_k - h_{k-1}\|^2 \\ &\quad + 3K_2^2(\|u_k - h_{k-1}\| + 1)^{2\alpha/(1-\alpha)}\|h_k - h_{k-1}\|^2.\end{aligned} \tag{124}$$

By stepsize choice, $\gamma_k \leq \frac{1}{6\sqrt{2}K_0}$, $\gamma_k \leq \frac{1}{6\sqrt{2}K_1\|F(h_{k-1})\|^\alpha}$, and $\gamma_k \leq \frac{1}{6\sqrt{2}K_2^2(\|u_k - h_{k-1}\|+1)^{\alpha/(1-\alpha)}}$, then we obtain:

$$\begin{aligned}2\gamma_k^2\|F(h_k) - F(h_{k-1})\|^2 &\leq \frac{1}{12}\|h_k - h_{k-1}\|^2 + \frac{1}{12}\|h_k - h_{k-1}\|^2 \\ &\quad + \frac{1}{12}\|h_k - h_{k-1}\|^2 \\ &= \frac{1}{4}\|h_k - h_{k-1}\|^2.\end{aligned} \tag{125}$$

Case (II) $\alpha = 1$

Using the definition of α-symmetric operator (as given in (7)), we obtain that for all $k \geq 1$,

$$\|F(h_k) - F(h_{k-1})\| \leq \|h_k - h_{k-1}\|(L_0 + L_1\|F(h_k)\|)\exp(L_1\|h_k - h_{k-1}\|). \tag{126}$$

By the projection property and triangle inequality:

$$\|h_k - h_{k-1}\| \leq \|u_k - h_{k-1} - \gamma_k F(h_{k-1})\| \leq \|u_k - h_{k-1}\| + \gamma_k\|F(h_{k-1})\|.$$

By the definition of the stepsize γ_k, we have that $\gamma_k \leq \frac{1}{\|F(h_{k-1})\|}$. Therefore, it follows that

$$\|h_k - h_{k-1}\| \leq \|u_k - h_{k-1}\| + 1.$$

Then:

$$\|F(h_k) - F(h_{k-1})\| \leq \|h_k - h_{k-1}\|(L_0 + L_1\|F(h_{k-1})\|)\exp(L_1\|u_k - h_{k-1}\| + 1). \tag{127}$$

Based on the step-size choice, $\gamma_k \leq \frac{1}{2\sqrt{2}(L_0+L_1\|F(h_{k-1})\|)\exp(L_1\|u_k - h_{k-1}\|+1)}$, we have:

$$2\gamma_k^2\|F(h_k) - F(h_{k-1})\|^2 \leq \frac{1}{4}\|h_k - h_{k-1}\|^2. \tag{128}$$

For both cases quantity $2\gamma_k^2\|F(h_k) - F(h_{k-1})\|^2 \leq \frac{1}{4}\|h_k - h_{k-1}\|^2$ is upperbounded for all $k \geq 1$. Then combining upperbounds in equations (125), and (128) with equation (121) we obtain that for any $k \geq 1$ the next inequality holds:

$$\begin{aligned}\|u_{k+1} - y\|^2 \leq \|u_k - y\|^2 - \|u_{k+1} - h_k\|^2 - \|u_k - h_k\|^2 \\ - 2\gamma_k\langle F(h_k), h_k - y\rangle + \frac{1}{4}\|h_k - h_{k-1}\|^2.\end{aligned} \tag{129}$$

Now using the triangle inequality and relation $(\sum_{i=1}^m a_i)^2 \leq m\sum_{i=1}^m a_i^2$ we obtain:

$$\|h_k - h_{k-1}\|^2 \leq 2\|u_k - h_k\|^2 + 2\|u_k - h_{k-1}\|^2.$$

Using the preceding etimate in equation (129) we get:

$$\begin{aligned}\|u_{k+1} - y\|^2 + \|u_{k+1} - h_k\|^2 \leq \|u_k - y\|^2 + \frac{1}{2}\|u_k - h_{k-1}\|^2 \\ - \frac{1}{2}\|u_k - h_k\|^2 - 2\gamma_k\langle F(h_k), h_k - y\rangle.\end{aligned} \tag{130}$$

Now we plug $y = u^*$ where u^* is an arbitrary solution and use the assumption in the theorem on quasi sharpness of the operator. Doing so we obtain the following recursive inequality for any $u^* \in U^*$ and $k \geq 1$

$$\begin{aligned}\|u_{k+1} - u^*\|^2 + \|u_{k+1} - h_k\|^2 &\leq \|u_k - u^*\|^2 + \frac{1}{2}\|u_k - h_{k-1}\|^2 \\ &- \frac{1}{2}\|u_k - h_k\|^2 - 2\gamma_k\langle F(h_k), h_k - u^*\rangle \\ &\leq \|u_k - u^*\|^2 + \|u_k - h_{k-1}\|^2 \\ &- \frac{1}{2}\|u_k - h_{k-1}\|^2 - \frac{1}{2}\|u_k - h_k\|^2.\end{aligned} \tag{131}$$

The first inequality holds because $\langle F(h_k), h_k - u^*\rangle$ is positive for all $u^* \in U^*$ when $F(\cdot)$ is p-quasi sharp. We conclude that for any $u^* \in U^*$, the norms $\|u_k - u^*\|$ and $\|u_{k+1} - h_k\|$ are bounded.

(b) Now, we want to estimate $\|F(h_{k-1})\|$, since this term is present in the denominator of the stepsizes. Let $u_1^* = P_{U^*}(u_1)$ be a projection of u_1 onto the solution set U^*. We add and subtract $F(u_1^*)$, and get $\|F(h_{k-1})\| = \|F(h_{k-1}) - F(u_1^*) + F(u_1^*)\| \leq \|F(h_{k-1}) - F(u_1^*)\| + \|F(u_1^*)\|$. Now, we estimate the first term using the assumption on the operator.

Case (I) $\alpha \in (0, 1)$ Using the definition of α-symmetric operator (as given in (6)), we obtain that for all $k \geq 1$,

$$\|F(h_{k-1}) - F(u_1^*)\| \leq (K_0 + K_1\|F(u_1^*)\|^\alpha)\|h_{k-1} - u_1^*\| + K_2\|h_{k-1} - u_1^*\|^{1/(1-\alpha)}. \tag{132}$$

we can get the following bound for $\|F(h_{k-1})\|$, for all $k \geq 1$:

$$\begin{aligned}\|F(h_{k-1})\| &\leq \|F(h_{k-1}) - F(u_1^*)\| + \|F(u_1^*)\| \\ &\leq (K_0 + K_1\|F(u_1^*)\|^\alpha)\|h_{k-1} - u_1^*\| + K_2\|h_{k-1} - u_1^*\|^{1/(1-\alpha)} + \|F(u_1^*)\|.\end{aligned} \tag{133}$$

Earlier, in equation (131) we proved that for arbitrary solution u^* the following bounds hold for any $k \leq 0$

$$\|u_{k+1} - u^*\|^2 + \|u_{k+1} - h_k\|^2 \leq \|u_1 - u^*\|^2 + \|u_1 - h_0\|^2.$$

Let $R_1^2 = \|u_1 - u_1^*\|^2 + \|u_1 - h_0\|^2$, then

$$(\|u_k - u_1^*\| + \|u_k - h_{k-1}\|)^2 \leq 2\|u_k - u_1^*\|^2 + 2\|u_k - h_{k-1}\|^2 \leq 2R_1^2.$$

Therefore, it follows that for all $k \geq 1$

$$\|h_{k-1} - u_1^*\| \leq \|u_k - u_1^*\| + \|u_k - h_{k-1}\| \leq \sqrt{2}R_1,$$

$$\|u_k - h_{k-1}\| \leq \sqrt{2}R_1.$$

Then using the facts above and (133) for all $k \geq 1$:

$$\|F(h_{k-1})\| \leq (K_0 + K_1\|F(u_1^*)\|^\alpha)\sqrt{2}R_1 + K_2(\sqrt{2}R_1)^{1/(1-\alpha)} + \|F(u_1^*)\|. \tag{134}$$

Now, we define a new constant $C_1 = (K_0 + K_1\|F(u_1^*)\|^\alpha)\sqrt{2}R_1 + K_2(\sqrt{2}R_1)^{1/(1-\alpha)} + \|F(u_1^*)\|$, then $\|F(h_{k-1})\| \leq C_1$ for all $k \geq 1$. Since function $(x)^s$ is monotone for any $x > 0$ and any $s > 0$, then $\|F(h_{k-1})\|^{\alpha/2} \leq C_1^{\alpha/2}$ and $(\|u_k - h_{k-1}\| + 1)^{\alpha/(1-\alpha)} \leq (\sqrt{2}R_1 + 1)^{\alpha/(1-\alpha)}$. Using these facts, we conclude that

$$\begin{aligned}\gamma_k &= \min\{\frac{1}{4\mu}, \frac{1}{\|F(h_{k-1})\|}, \frac{1}{6\sqrt{2}K_0}, \frac{1}{6\sqrt{2}K_1\|F(h_{k-1})\|^{\alpha/2}}, \frac{1}{6\sqrt{2}K_2(\|u_k - h_{k-1}\| + 1)^{\alpha/(1-\alpha)}}\} \\ &\geq \min\{\frac{1}{4\mu}, \frac{1}{C_1}, \frac{1}{6\sqrt{2}K_0}, \frac{1}{6\sqrt{2}K_1C_1^{\alpha/2}}, \frac{1}{6\sqrt{2}K_2(\sqrt{2}R_1 + 1)^{\alpha/(1-\alpha)}}\}.\end{aligned} \tag{135}$$

Case (II) $\alpha = 1$ Using the definition of α-symmetric operator (as given in (7)), we obtain that for all $k \geq 1$,

$$\|F(h_k) - F(u_1^*)\| \leq \|h_k - u_1^*\|(L_0 + L_1\|F(u_1^*)\|)\exp(L_1\|h_k - u_1^*\|). \tag{136}$$

Hence, we obtain the following bound for $\|F(h_{k-1})\|$, for all $k \geq 1$:

$$\begin{aligned}\|F(h_{k-1})\| &\leq \|F(h_{k-1}) - F(u_1^*)\| + \|F(u_1^*)\| \\ &\leq \|h_k - u_1^*\|(L_0 + L_1\|F(u_1^*)\|)\exp(L_1\|h_k - u_1^*\|) + \|F(u_1^*)\|\end{aligned} \tag{137}$$

Earlier, in equation (131) we proved that for arbitrary solution u^* the following bounds hold for any $k \geq 0$

$$\|u_{k+1} - u^*\|^2 + \|u_{k+1} - h_k\|^2 \leq \|u_1 - u^*\|^2 + \|u_1 - h_0\|^2.$$

Let $R_1^2 = \|u_1 - u_1^*\|^2 + \|u_1 - h_0\|^2$, then

$$(\|u_k - u_1^*\| + \|u_k - h_{k-1}\|)^2 \leq 2\|u_k - u_1^*\|^2 + 2\|u_k - h_{k-1}\|^2 \leq 2R_1^2.$$

Therefore, it follows that for all $k \geq 1$:

$$\|h_{k-1} - u_1^*\| \leq \|u_k - u_1^*\| + \|h_{k-1} - u_k\| \leq \sqrt{2}R_1,$$

$$\|u_k - h_{k-1}\| \leq \sqrt{2}R_1.$$

Then using the preceding relation and equation (137) for all $k \geq 1$:

$$\begin{aligned}\|F(h_{k-1})\| &\leq \|h_k - u_1^*\|(L_0 + L_1\|F(u_1^*)\|)\exp(L_1\|h_k - u_1^*\|) + \|F(u_1^*)\| \\ &\leq \sqrt{2}R_1(L_0 + L_1\|F(u_1^*)\|)\exp(L_1\sqrt{2}R_1) + \|F(u_1^*)\|.\end{aligned} \tag{138}$$

Now, we define a new constant $\overline{C}_1 = \sqrt{2}R_1(L_0 + L_1\|F(u_1^*)\|)\exp(L_1\sqrt{2}R_1) + \|F(u_1^*)\|$, then $\|F(h_{k-1})\| \leq \overline{C}_1$ for all $k \geq 1$. And since function $\exp(x)$ is monotone for any $x > 0$, then $\exp(L_1\|u_k - h_{k-1}\| + 1) \leq \exp(L_1\sqrt{2}R_1 + 1)$. Using these facts, we conclude that

$$\begin{aligned}\gamma_k &= \min\{\frac{1}{4\mu}, \frac{1}{\|F(h_{k-1})\|}, \frac{1}{4(L_0 + L_1\|F(h_{k-1})\|)\exp(L_1\|u_k - h_{k-1}\| + 1)}\} \\ &\geq \min\{\frac{1}{4\mu}, \frac{1}{\overline{C}_1}, \frac{1}{4(L_0 + L_1\overline{C}_1)\exp(L_1\sqrt{2}R_1 + 1)}\}\end{aligned} \tag{139}$$

For both cases $\alpha \in (0,1)$ and $\alpha = 1$ in equations (135), (139) we showed that stepsizes sequence $\{\gamma_k\}$ is lower bounded by constant $\underline{\gamma}$, where

$$\underline{\gamma} = \min\{\frac{1}{4\mu}, \frac{1}{C_1}, \frac{1}{6\sqrt{2}K_0}, \frac{1}{6\sqrt{2}K_1(C_1)^{\alpha/2}}, \frac{1}{6\sqrt{2}K_2(\sqrt{2}R_1+1)^{\alpha/(1-\alpha)}}\} \text{ for } \alpha \in (0,1)$$
$$\underline{\gamma} = \min\{\frac{1}{4\mu}, \frac{1}{\overline{C}_1}, \frac{1}{4(L_0 + L_1\overline{C}_1)\exp(L_1\sqrt{2}R_1+1)}\} \text{ for } \alpha = 1 \tag{140}$$

(c) Based on the results from part (b) $\underline{\gamma} \leq \gamma_k$ for all $k \geq 1$. Then using p-quasi-sharpness of the operator and equation (130) with $y = u^*$, where u^* is an arbitrary solution, we obtain for any $k \geq 1$:

$$\begin{aligned}
\|u_{k+1} - u^*\|^2 + \|u_{k+1} - h_k\|^2 &\leq \|u_k - u^*\|^2 + \frac{1}{2}\|u_k - h_{k-1}\|^2. \\
&- \frac{1}{2}\|u_k - h_k\|^2 - 2\gamma_k\langle F(h_k), h_k - u^*\rangle \\
&\leq \|u_k - u^*\|^2 + \|u_k - h_{k-1}\|^2 \\
&- \frac{1}{2}\|u_k - h_k\|^2 - \frac{1}{2}\|u_k - h_{k-1}\|^2 - 2\gamma_k\mu\,\mathrm{dist}^p(h_k, U^*) \\
&\leq \|u_k - u^*\|^2 + \|u_k - h_{k-1}\|^2 \\
&- \frac{1}{2}\|u_k - h_k\|^2 - \frac{1}{2}\|u_k - h_{k-1}\|^2 - 2\underline{\gamma}\mu\,\mathrm{dist}^p(h_k, U^*).
\end{aligned} \tag{141}$$

The equation (141) satisfies the condition of Lemma B.3 with

$$\bar{v}_k = \|u_k - u^*\|^2 + \|u_k - h_{k-1}\|^2, \quad \bar{a}_k = 0,$$
$$\bar{z}_k = \frac{1}{2}\|u_k - h_{k-1}\|^2 + \frac{1}{2}\|u_k - h_k\|^2 + 2\mu\underline{\gamma}\,\mathrm{dist}^p(h_k, U^*), \quad \bar{b}_k = 0 \tag{142}$$

By Lemma B.2, it follows that the sequence $\{\bar{v}_k\}$ converges to a non-negative scalar for any $u^* \in U^*$, and surely we have

$$\sum_{k=N}^{\infty} \mathrm{dist}^p(h_k, U^*) < \infty, \quad \sum_{k=N}^{\infty} (\|u_k - h_k\|^2 + \|u_k - h_{k-1}\|^2) < \infty.$$

Thus, it follows that

$$\lim_{k\to\infty} \mathrm{dist}^p(h_k, U^*) = 0 \tag{143}$$

$$\lim_{k\to\infty} \|u_k - h_k\| = 0 \tag{144}$$

$$\lim_{k\to\infty} \|u_k - h_{k-1}\| = 0 \tag{145}$$

Since $\bar{v}_k$ converges for any given $u^* \in U^*$ and based on equation (145) we can conclude that $\|u_k - u^*\|$ converges for all $u^* \in U^*$. Therefore, the sequence $\{u_k\}$ is bounded and has accumulation points. In view of relation (144), the sequences $\{u_k\}$ and $\{h_k\}$ have the same accumulation points. Then, there exists a convergent subsequnce $\{u_{k_i}\}$, such that $\bar{u} \in U$ its limit point, i.e.,

$$\lim_{i\to\infty} \|u_{k_i} - \bar{u}\| = 0 \tag{146}$$

By relation (144), it follows that

$$\lim_{i\to\infty} \|h_{k_i} - \bar{u}\| = 0$$

By continuity of the distance function $\mathrm{dist}(\cdot, U^*)$, from relation (143) we conclude that $\mathrm{dist}(\bar{u}, U^*) = 0$, which implies that $\bar{u} \in U^*$ since the set U^* is closed. Since the sequence $\{\|u_k - u^*\|^2\}$ converges for any $u^* \in U^*$, it follows that $\{\|u_k - \bar{u}\|^2\}$ converges , and by relation (146) we conclude that $\lim_{k\to\infty} \|u_k - \bar{u}\|^2 = 0$. ■

Proof of Theorem 6.3

Proof. Since the solution set U^* is closed, there exists a projection u_k^* of the iterate u_k on the set U^*, i.e., there exists a point $u_k^* \in U^*$ such that $\|u_k - u_k^*\| = \text{dist}(u_k, U^*)$. Thus, by letting $u^* = u_k^*$ in (141), we obtain for all $k \geq 1$,

$$\begin{aligned}\|u_{k+1} - u_k^*\|^2 + \|u_{k+1} - h_k\|^2 \leq \|u_k - u_k^*\|^2 + \|u_k - h_{k-1}\|^2 \\ - \frac{1}{2}\|u_k - h_k\|^2 - \frac{1}{2}\|u_k - h_{k-1}\|^2 - 2\mu\gamma_k \,\text{dist}^p(h_k, U^*).\end{aligned} \tag{147}$$

In view of $\|u_k - u_k^*\| = \text{dist}(u_k, U^*)$ and $\text{dist}(u_{k+1}, U^*) \leq \|u_{k+1} - u_k^*\|$, it follows that for all $k \geq 1$,

$$\begin{aligned}\text{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2 \leq \text{dist}^2(u_k, U^*) + \|u_k - h_{k-1}\|^2 \\ - \frac{1}{2}\|u_k - h_k\|^2 - \frac{1}{2}\|u_k - h_{k-1}\|^2 - 2\mu\gamma_k \,\text{dist}^p(h_k, U^*).\end{aligned} \tag{148}$$

Based on the result of Theorem 6.2 (b) the stepsizes are lower bounded, i.e. $\gamma_k \geq \underline{\gamma}$, then for all $k \geq 1$ we have:

$$\begin{aligned}\text{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2 \leq \text{dist}^2(u_k, U^*) + \|u_k - h_{k-1}\|^2 \\ - \frac{1}{2}\|u_k - h_k\|^2 - \frac{1}{2}\|u_k - h_{k-1}\|^2 - 2\mu\underline{\gamma} \,\text{dist}^p(h_k, U^*).\end{aligned} \tag{149}$$

Next, we estimate the term $\text{dist}^p(h_k, U^*)$ in (149). By the triangle inequality, we have

$$\|u_k - u^*\| \leq \|u_k - h_k\| + \|h_k - u^*\| \qquad \text{for all } u^* \in U^*,$$

and by taking the minimum over $u^* \in U^*$ on both sides of the preceding relation, we obtain

$$\text{dist}(u_k, U^*) \leq \|u_k - h_k\| + \text{dist}(h_k, U^*). \tag{150}$$

Case (a) $p \leq 2$

Squaring both sides of inequality (150), and using $(\sum_{i=1}^m a_i)^2 \leq m \sum_{i=1}^m a_i^2$, which is valid for any scalars $a_i, i = 1, \ldots, m$, and any integer $m \geq 1$, we obtain

$$-2\,\text{dist}^2(h_k, U^*) \leq 2\|u_k - h_k\|^2 - \text{dist}^2(u_k, U^*). \tag{151}$$

In Theorem 6.2 we showed that the sequence $\{h_k\}$ convergence to a solution $\bar{u} \in U^*$. Hence, there exists $B_h \in \mathbb{R}$ such that the sequence $\text{dist}(h_k, U^*)$ is bounded by B_h. Then, since $p \leq 2$, it holds that for all $k \geq 0$

$$\text{dist}^{2-p}(h_k, U^*) \leq (B_h)^{2-p},$$

and consequently,

$$\text{dist}^2(h_k, U^*) = \text{dist}^p(h_k, U^*)\text{dist}^{2-p}(h_k, U^*) \leq (B_h)^{2-p}\text{dist}^p(h_k, U^*).$$

Thus,

$$-\text{dist}^p(h_k, U^*) \leq -\frac{1}{(B_h)^{2-p}}\text{dist}^2(h_k, U^*)$$

Using the preceding inequality in equation (151) we obtain the following bound:

$$-\text{dist}^p(h_k, U^*) \leq -\frac{1}{(B_h)^{2-p}}\text{dist}^2(h_k, U^*) \leq \frac{1}{(B_h)^{2-p}}\|u_k - h_k\|^2 - \frac{1}{2(B_h)^{2-p}}\text{dist}^2(u_k, U^*) \tag{152}$$

Combining equation (152) with (149), and adding and subtracting $\mu\underline{\gamma}\|u_k - h_{k-1}\|^2$ we obtain that for any $k \geq 1$:

$$\begin{aligned}\text{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2 \leq \text{dist}^2(u_k, U^*) + (1 - \frac{\mu}{(B_h)^{2-p}}\underline{\gamma})\|u_k - h_{k-1}\|^2 \\ - (\frac{1}{2} - \frac{2\mu}{(B_h)^{2-p}}\underline{\gamma})\|u_k - h_k\|^2 - (\frac{1}{2} - \frac{\mu}{(B_h)^{2-p}}\underline{\gamma})\|u_k - h_{k-1}\|^2 \\ - \frac{\mu}{(B_h)^{2-p}}\underline{\gamma}\,\text{dist}^2(u_k, U^*).\end{aligned} \tag{153}$$

Without loss of generality we consider case when $B_h \geq 1$ (otherwise, we can choose $B_h = 1$), then:

$$\begin{aligned}\operatorname{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2 \leq \left(1 - \frac{\mu}{(B_h)^{2-p}}\underline{\gamma}\right)\left(\operatorname{dist}^2(u_k, U^*) + \|u_k - h_{k-1}\|^2\right) \\ - (\frac{1}{2} - 2\mu\underline{\gamma})\|u_k - h_k\|^2 - (\frac{1}{2} - \mu\underline{\gamma})\|u_k - h_{k-1}\|^2.\end{aligned} \tag{154}$$

By the stepsize choice, $\underline{\gamma} \leq \frac{1}{4\mu}$ then $-(\frac{1}{2} - 2\mu\underline{\gamma})\|u_k - h_k\|^2 \leq 0$ and $-(\frac{1}{2} - \mu\underline{\gamma})\|u_k - h_{k-1}\|^2 \leq 0$. Thus, by defining $R_k^2 = \operatorname{dist}^2(u_k, U^*) + \|u_k - h_{k-1}\|^2$ we obtain that for all $k \geq 1$,

$$R_{k+1}^2 \leq \left(1 - \frac{\mu}{(B_h)^{2-p}}\underline{\gamma}\right) R_k^2. \tag{155}$$

Case (b) $p > 2$ By using Lemma B.4, we further obtain

$$\begin{aligned}\operatorname{dist}^p(u_k, U^*) &\leq (\|u_k - h_k\| + \operatorname{dist}(h_k, U^*))^p \\ &\leq 2^{p-1}\|u_k - h_k\|^p + 2^{p-1}\operatorname{dist}^p(h_k, U^*).\end{aligned} \tag{156}$$

Using projection inequality (50), and stepsizes choice we obtain:

$$\|u_k - h_k\| \leq \|\gamma_k F(h_{k-1})\| \leq 1,$$
$$\|u_k - h_k\|^{p-2} \leq 1.$$

Combining this result with (156)

$$\operatorname{dist}^p(u_k, U^*) \leq 2^{p-1}\|u_k - h_k\|^2 + 2^{p-1}\operatorname{dist}^p(h_k, U^*). \tag{157}$$

And by rearranging terms we obtain:

$$-\operatorname{dist}^p(h_k, U^*) \leq \|u_k - h_k\|^2 - 2^{-p+1}\operatorname{dist}^p(u_k, U^*) \tag{158}$$

Combining this inequality with (149), we obtain that for any $k \geq 1$:

$$\begin{aligned}\operatorname{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2 \leq \operatorname{dist}^2(u_k, U^*) + \|u_k - h_{k-1}\|^2 - (\frac{1}{2} - 2\mu\underline{\gamma})\|u_k - h_k\|^2 \\ - \frac{1}{2}\|u_k - h_{k-1}\|^2 - 2^{-p+2}\mu\underline{\gamma}\operatorname{dist}^p(u_k, U^*).\end{aligned} \tag{159}$$

By the stepsizes choice $\gamma_k \leq \frac{1}{4\mu}$ for all $k \geq 1$, then $(\frac{1}{2} - 2\mu\underline{\gamma})\|u_k - h_k\|^2 \leq 0$, and for all $k \geq 1$

$$\begin{aligned}\operatorname{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2 \leq \operatorname{dist}^2(u_k, U^*) + \|u_k - h_{k-1}\|^2 \\ - \frac{1}{2}\|u_k - h_{k-1}\|^2 - 2^{-p+2}\mu\underline{\gamma}\operatorname{dist}^p(u_k, U^*).\end{aligned} \tag{160}$$

To get convergence rate results we define $R_k^2 = \operatorname{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2$. Now we estimate $-\frac{1}{2}\|u_k - h_{k-1}\|^2$. From the result of Theorem 6.2 (a) we get that for all $k \geq 1$, $\|u_k - h_{k-1}\| \leq \sqrt{2}R_1$. Then for all $k \geq 1$

$$\|u_k - h_{k-1}\|^{p-2}\|u_k - h_{k-1}\|^2 \leq (2R_1^2)^{(p-2)/2}\|u_k - h_{k-1}\|^2 \tag{161}$$

Then by multiplying both sides by $-\frac{1}{2}\|u_k - h_{k-1}\|^{p-2}$ we obtain:

$$-\frac{1}{2}\|u_k - h_{k-1}\|^2 \leq -\frac{1}{2}(2R_1^2)^{-(p-2)/2}\|u_k - h_{k-1}\|^p \tag{162}$$

Using this bound and equation (160) we get:

$$\begin{aligned}\operatorname{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2 \leq \operatorname{dist}^2(u_k, U^*) + \|u_k - h_{k-1}\|^2 \\ - \frac{1}{2}(2R_1^2)^{-(p-2)/2}\|u_k - h_{k-1}\|^p - 2^{-p+2}\mu\underline{\gamma}\operatorname{dist}^p(u_k, U^*).\end{aligned} \tag{163}$$

Let $C_4 = \min\{\frac{1}{2}(2R_1^2)^{-(p-2)/2}, 2^{-p+2}\mu\underline{\gamma}\}$, note that $C_4 \leq 2^{-p+2}\mu\underline{\gamma} \leq 4\mu\underline{\gamma} \leq 1$ when $p > 2$. Then we obtain for all $k \geq 1$

$$\begin{aligned}\text{dist}^2(u_{k+1}, U^*) + \|u_{k+1} - h_k\|^2 \leq \text{dist}^2(u_k, U^*) + \|u_k - h_{k-1}\|^2 \\ - C_4(\|u_k - h_{k-1}\|^p + \text{dist}^p(u_k, U^*)).\end{aligned} \tag{164}$$

Finally, using ineqaulity (B.4) with $a_1 = \|u_k - h_{k-1}\|^p$, $a_2 = \text{dist}^2(u_k, U^*)$ we obtain

$$-((\|u_k - h_{k-1}\|^2)^{p/2} + (\text{dist}^2(u_k, U^*))^{p/2}) \leq 2^{-p/2+1}(\|u_k - h_{k-1}\|^2 + \text{dist}^2(u_k, U^*))^{p/2} \tag{165}$$

Letting $R_k^2 = \|u_k - h_{k-1}\|^2 + \text{dist}^2(u_k, U^*)$, and using the fact that $C_4 2^{-p/2+1} \leq 1$ for $p > 2$, we obtain that for all $k \geq 1$:

$$R_{k+1}^2 \leq R_k^2 - C_4 2^{-p/2+1}(R_k^2)^{p/2}. \tag{166}$$

Using Lemma B.5 with $x_k = R_k^2$, $q = \frac{p-2}{2}$, $a = C_4 2^{1-p/2}$ we obtain, for all $k \geq 1$:

$$R_{k+1}^2 \leq \frac{R_1^2}{(1 + (p-2)C_4 2^{-(p+2)/2}(R_1^2)^{(p-2)/2}k)^{2/(p-2)}}. \tag{167}$$

■

F. Analysis of Korpelevich Method with Clipping 7

Proof of Theorem 7.1

Proof. **(a)** By Lemma 5.1, the following inequality holds for any $y \in U$ and all $k \geq 0$:

$$\|h_{k+1} - y\|^2 \leq \|h_k - y\|^2 - \|h_k - u_k\|^2 - 2\gamma_k\langle F(u_k), u_k - y\rangle + \gamma_k^2\|F(h_k) - F(u_k)\|^2. \tag{168}$$

We want to estimate the last term on the LHS of the inequality using the fact that operator $F(\cdot)$ is an α-symmetric operator.

Case (I) $\alpha \in (0, 1)$.

Using the alternative characterization of α-symmetric operators from Proposition 2.2(a) (as given in (6)), when $\alpha \in (0, 1)$, the next inequality holds for any $k \geq 1$:

$$\|F(h_k) - F(u_k)\| \leq \|h_k - u_k\|(K_0 + K_1\|F(h_k)\|^\alpha + K_2\|h_k - u_k\|^{\alpha/(1-\alpha)}). \tag{169}$$

By the projection property (50) and the stepsize choice (15) :

$$\|h_k - u_k\| \leq \gamma_k\|F(h_k)\| = \beta_k \min\{1, \frac{1}{\|F(h_k)\|}\}\|F(h_k)\| \leq \beta_k \leq 1.$$

Then by the relation $(\sum_{i=1}^m a_i)^2 \leq m\sum_{i=1}^m a_i^2$ we obtain:

$$\begin{aligned}\gamma_k\|F(h_k) - F(u_k)\| &\leq \gamma_k(K_0 + K_1\|F(h_k)\|^\alpha + K_2)\|h_k - u_k\| \\ &\leq \beta_k(K_0 \min\{1, \frac{1}{\|F(h_k)\|}\} + K_1 \min\{1, \frac{1}{\|F(h_k)\|}\}\|F(h_k)\|^\alpha + K_2 \min\{1, \frac{1}{\|F(h_k)\|}\})\|h_k - u_k\| \\ &\leq \beta_k(K_0 + K_1 + K_2)\|h_k - u_k\|.\end{aligned} \tag{170}$$

Case (II) $\alpha = 1$.

Based on the alternative characterization of α-symmetric operators from Proposition 2.2(b) (as given in (7)), when $\alpha = 1$, the next inequality holds for any $k \geq 0$:

$$\|F(h_k) - F(u_k)\| \leq \|h_k - u_k\|(L_0 + L_1\|F(h_k)\|)\exp(L_1\|h_k - u_k\|). \tag{171}$$

By the projection property, triangle inequality and step size choice (28):

$$\|h_k - u_k\| \leq \gamma_k \|F(h_k)\| = \beta_k \min\{1, \frac{1}{\|F(h_k)\|}\}\|F(h_k)\| \leq \beta_k.$$

Then, we obtain:

$$\begin{aligned}\gamma_k\|F(h_k) - F(u_k)\| &\leq \gamma_k(L_0 + L_1\|F(h_k)\|)\exp(L_1\beta_k)\|h_k - u_k\| \\ &= \exp(L_1\beta_k)(L_0\beta_k \min\{1, \frac{1}{\|F(h_k)\|}\} + L_1\beta_k \min\{1, \frac{1}{\|F(h_k)\|}\}\|F(h_k)\|)\|h_k - u_k\| \\ &\leq \exp(L_1\beta_k)\beta_k(L_0 + L_1)\|h_k - u_k\|.\end{aligned} \tag{172}$$

Now let $C_a(\beta_k) = K_0 + K_1 + K_2$ when $\alpha \in (0, 1)$, and $C_a(\beta_k) = \exp(L_1\beta_k)(L_0 + L_1)$ when $\alpha = 1$. Then for both cases$\gamma_k^2\|F(h_k) - F(u_k)\|^2 \leq \beta_k^2 C_a^2(\beta_k)\|h_k - u_k\|^2$, then combining this fact with (168) we obtain that for any $k \geq 0$:

$$\|h_{k+1} - y\|^2 \leq \|h_k - y\|^2 - (1 - C_a^2(\beta_k))\|u_k - h_k\|^2 - 2\gamma_k\langle F(u_k), u_k - y\rangle. \tag{173}$$

Now we plug $y = u^*$ into equation (173), where $u^* \in U^*$ is an arbitrary solution. Thus, by using p-quasi sharpness of the operator, we obtain the following recursive inequality:

$$\|h_{k+1} - u^*\|^2 \leq \|h_k - u^*\|^2 - (1 - C_a^2(\beta_k))\|u_k - h_k\|^2 - 2\gamma_k\mu\text{dist}^p(u_k, U^*). \tag{174}$$

Now, if there exists $N > 0$ such that $C_a^2(\beta_k) < 1$ for all $k \geq N$, then we conclude that for any $u^* \in U^*$, the sequence $\{\|h_k - u^*\|\}$ is bounded $\bar{D}_N$ by where $\bar{D}_N = \max_{k\in[0,N]}\{\text{dist}(h_k, U^*)\}$ for all $k \geq 0$.

(b) Now, we want to estimate $\|F(h_k)\|$, since this term is present in the denominator of the stepsize. Let $h_0^* = P_{U^*}(h_0)$ be a projection of h_0 onto the solution set U^*. We add and subtract $F(h_0^*)$, and get $\|F(h_k)\| = \|F(h_k) - F(h_0^*) + F(h_0^*)\| \leq \|F(h_k) - F(h_0^*)\| + \|F(h_0^*)\|$. We can estimate the first term using the α-symmetric assumption on the operator class.

Case(I) $\alpha \in (0, 1)$

Based on the alternative characterization of α-symmetric operators from Proposition 2.2(a) (as given in (6)),

$$\|F(h_k) - F(h_0^*)\| \leq \|h_k - h_0^*\|(K_0 + K_1\|F(h_0^*)\|^\alpha + K_2\|h_k - h_0^*\|^{\alpha/(1-\alpha)}). \tag{175}$$

Earlier, in part (a) we proved that for arbitrary solution u^* the following bound holds for any $k \geq 0$

$$\|h_k - u^*\| \leq \bar{D}_N.$$

Therefore it holds for $u^* = h_0^*$,

$$\|h_k - h_0^*\| \leq \bar{D}_N.$$

Using this fact and equation (175) for we obtain that for all $k \geq 0$:

$$\|F(h_k)\| \leq \bar{D}_N(K_0 + K_1\|F(h_0^*)\|^\alpha + K_2\bar{D}_N^{\alpha/(1-\alpha)}) + \|F(h_0^*)\|. \tag{176}$$

We showed that the sequence $\{\|F(h_k)\|\}$ is upper bounded by some constant C_1. Where $C_1 = \bar{D}_N(K_0 + K_1\|F(h_0^*)\|^\alpha + K_2\bar{D}_N^{\alpha/(1-\alpha)}) + \|F(h_0^*)\|$. Using this fact, we conclude that for all $k \geq 0$

$$\begin{aligned}\gamma_k &= \beta_k \min\{1, \frac{1}{\|F(h_k)\|}\} \\ &\geq \beta_k \min\{1, \frac{1}{C_1}\}\end{aligned} \tag{177}$$

Case(II) $\alpha = 1$

Based on the alternative characterization of α-symmetric operators from Proposition 2.2(b) (as given in (7)),

$$\|F(h_k) - F(h_0^*)\| \leq \|h_k - h_0^*\|(L_0 + L_1\|F(h_0^*)\|)\exp(L_1\|h_k - h_0^*\|). \tag{178}$$

Using the preceding relation we get the following bound for $\|F(h_{k-1})\|$, for all $k \geq 1$:

$$\begin{aligned}\|F(h_k)\| &\leq \|F(h_k) - F(h_0^*)\| + \|F(h_0^*)\| \\ &\leq \|h_k - h_0^*\|(L_0 + L_1\|F(h_0^*)\|)\exp(L_1\|h_k - h_0^*\|) + \|F(h_0^*)\|\end{aligned} \tag{179}$$

Earlier, in part(a) we proved that, for arbitrary solution $u^* \in U^*$, the following bound hold for any $k \geq 0$

$$\|h_k - u^*\| \leq \bar{D}_N.$$

Then, it holds for $u^* = h_0^*$, and for all $k \geq 0$ it holds that

$$\|h_k - h_0^*\| \leq \bar{D}_N.$$

Using this fact and equation (179) for all $k \geq 0$:

$$\begin{aligned}\|F(h_k)\| &\leq \|h_k - h_0^*\|(L_0 + L_1\|F(h_0^*)\|)\exp(L_1\|h_k - h_0^*\|) + \|F(h_0^*)\| \\ &\leq \bar{D}_N(L_0 + L_1\|F(h_0^*)\|)\exp(L_1\bar{D}_N) + \|F(h_0^*)\|.\end{aligned} \tag{180}$$

We showed that the sequence $\{\|F(h_k)\|\}$ is upper bounded by some constant $\overline{C}_1$. Where $\overline{C}_1 = \bar{D}_N(L_0 + L_1\|F(h_0^*)\|)\exp(L_1\bar{D}_N) + \|F(h_0^*)\|$. Then, we conclude that

$$\begin{aligned}\gamma_k &= \beta_k \min\{1, \frac{1}{\|F(h_k)\|}\} \\ &\geq \beta_k \min\{1, \frac{1}{\overline{C}_1}\}.\end{aligned} \tag{181}$$

For both cases $\alpha \in (0,1)$ and $\alpha = 1$ in equations (176), (181) we showed that stepsizes sequence $\{\gamma_k\}$ is lower bounded by some constant $\beta_k \min\{1, \frac{1}{C_1}, \frac{1}{\overline{C}_1}\}$.

(c) Using the results from part (b) and equation (174), we obtain for any $k \geq 0$:

$$\|h_{k+1} - u^*\|^2 \leq \|h_k - u^*\|^2 - \frac{1}{2}\|u_k - h_k\|^2 - 2\beta_k\mu \min\{1, \frac{1}{C_1}, \frac{1}{\overline{C}_1}\}\,\mathrm{dist}^p(u_k, U^*). \tag{182}$$

The equation (182) satisfies the condition of Lemma B.3 with

$$\begin{aligned}\bar{v}_k &= \|h_k - h^*\|^2, \quad \bar{a}_k = 0, \\ \bar{z}_k &= \frac{1}{2}\|u_k - h_k\|^2 + 2\mu\beta_k \min\{1, \frac{1}{C_1}, \frac{1}{\overline{C}_1}\}\,\mathrm{dist}^p(u_k, U^*), \quad \bar{b}_k = 0.\end{aligned} \tag{183}$$

By Lemma B.2, it follows that the sequence $\{\bar{v}_k\}$ converges to a non-negative scalar for any $u^* \in U^*$, we have

$$\sum_{k=0}^{\infty} \beta_k \mathrm{dist}^p(u_k, U^*) < \infty, \quad \sum_{k=0}^{\infty} \|u_k - h_k\|^2 < \infty.$$

Thus, it follows that

$$\lim_{k\to\infty} \beta_k \mathrm{dist}^p(u_k, U^*) = 0, \tag{184}$$

and since $\sum_{k=0}^{\infty} \beta_k = \infty$,

$$\liminf_{k\to\infty} \operatorname{dist}^p(u_k, U^*) = 0, \tag{185}$$

$$\lim_{k\to\infty} \|u_k - h_k\| = 0. \tag{186}$$

Since $\bar{v}_k$ converges for any given $u^* \in U^*$ we can conclude that $\|h_k - u^*\|$ converges for all $u^* \in U^*$. Therefore, the sequence $\{h_k\}$ is bounded and has accumulation points. In view of relation (186), the sequences $\{u_k\}$ and $\{h_k\}$ have the same accumulation points. Let $\{k_i\}$ be an index sequence, such that

$$\lim_{i\to\infty} \operatorname{dist}^p(u_{k_i}, U^*) = \liminf_{k\to\infty} \operatorname{dist}^p(u_k, U^*) = 0, \tag{187}$$

We assume that the sequence $\{u_{k_i}\}$ is convergent with a limit point $\bar{u}$, otherwise, we choose a convergent subsequence:

$$\lim_{i\to\infty} \|u_{k_i} - \bar{u}\| = 0 \tag{188}$$

By relation (186), it follows that

$$\lim_{i\to\infty} \|h_{k_i} - \bar{u}\| = 0 \tag{189}$$

By continuity of the distance function $\operatorname{dist}(\cdot, U^*)$, from relation (184) we conclude that $\operatorname{dist}(\bar{u}, U^*) = 0$, which implies that $\bar{u} \in U^*$ since the set U^* is closed. Since the sequence $\{\|h_k - u^*\|^2\}$ converges for any $u^* \in U^*$, it follows that $\{\|h_k - \bar{u}\|^2\}$ converges , and by relation (189) we conclude that $\lim_{k\to\infty} \|h_k - \bar{u}\|^2 = 0$.

■

Proof of Theorem 7.3

Proof. By Lemma 5.1, the following inequality holds for any $y \in U$ and all $k \geq 0$:

$$\|h_{k+1} - y\|^2 \leq \|h_k - y\|^2 - \|h_k - u_k\|^2 - 2\gamma_k \langle F(u_k), u_k - y\rangle + \gamma_k^2 \|F(h_k) - F(u_k)\|^2. \tag{190}$$

By Algorithm 1, $\gamma_k^2 \|F(h_k) - F(u_k)\|^2 \leq \frac{1}{2}\|u_k - h_k\|$, then:

$$\|h_{k+1} - y\|^2 \leq \|h_k - y\|^2 - \frac{1}{2}\|h_k - u_k\|^2 - 2\gamma_k \langle F(u_k), u_k - y\rangle. \tag{191}$$

Now we plug $y = u^*$ into equation (191), where $u^* \in U^*$ is an arbitrary solution. Thus, by using p-quasi sharpness of the operator, we obtain the following recursive inequality:

$$\|h_{k+1} - y\|^2 \leq \|h_k - y\|^2 - \frac{1}{2}\|h_k - u_k\|^2 - 2\mu\gamma_k \operatorname{dist}^p(u_k, U^*). \tag{192}$$

Equation (192) is similar to equation (174) in the proof of Theorem 7.1 with $C_a(\beta_k) = \frac{1}{2}$. By following the proof of Theorem 7.1, we obtain (i) there exist a lower bound $\underline{c} \in \mathbb{R}$ on $\min\{1, \frac{1}{\|F(h_k)\|}\}$, i.e. $\min\{1, \frac{1}{\|F(h_k)\|}\} \geq \underline{c}$ for all $k \geq 0$ (ii) iterates $\{h_k\}$ and $\{u_k\}$ converge to a solution $\bar{u} \in U^*$.

Moreover, equation (192) is equivalent to equation (100) of the proof of Theorem 5.3 on convergence rates. Since $\gamma_k = \beta_k \min\{1, \frac{1}{\|F(h_k)\|}\} \geq \beta_k \underline{c}$ and $\beta_k \leq \frac{1}{4\mu}$, by following the proof of Theorem 5.3 we obtain:

Case (a) $p \leq 2$

For all $k \geq 0$,

$$D_{k+1}^2 \leq (1 - \frac{\mu}{(B_u)^{2-p}} \beta_k \underline{c}) D_k^2, \tag{193}$$

where $D_k^2 = \operatorname{dist}^2(h_k, U^*)$.

Case (b) $p > 2$

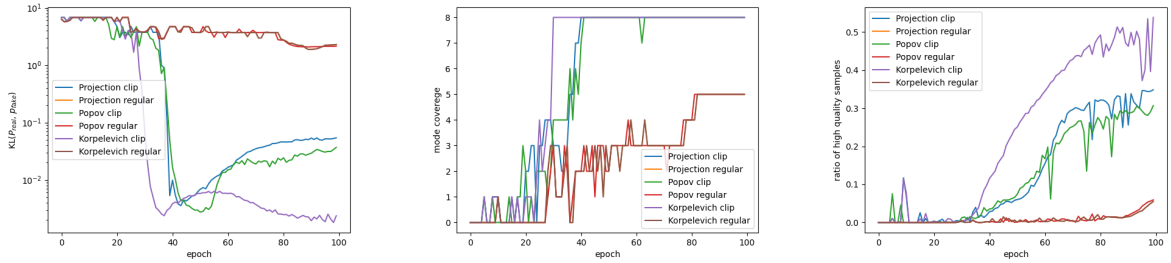
For all $k \geq 0$:

$$D_{k+1}^2 \leq \frac{D_0^2}{(1 + (D_0^2)^{(p-2)/2} \underline{c} \sum_{i=0}^{k} \beta_i)^{2/(p-2)}}. \tag{194}$$

■

G. Additional Experiments

We present the experiments on training GAN for the 2D Ring dataset, a mixture of 8 equal-prior Gaussian distributions, with mean $\cos(2\pi i/8), \sin(2\pi i/8)$ for $i \in \{1, \ldots, 8\}$ and variance 10^{-4}. Both the generator and discriminator networks have 4 fully-connected layers with hidden sizes 200 and 400, respectively. We use LSGAN objective functions and run methods for 100 epochs with a batch size of 128, with $\gamma_k = 10^{-3}$ for methods without clipping, and $\gamma_k = 10^{-3} \min\{1, \frac{1}{\|F(h_k)\|}\}$ for methods with clipping. Based on our experiments, all three considered methods with clipping outperform the same methods without clipping. The results of this experiment are shown in Figure 3 and Table 2



(a) KL Divergence between real distribution and fake

(b) Mode coverage

(c) High quality samples

Figure 3. Comparison of projection, Korpelevich and Popov methods with and without clipping.

Table 2. Comparison of projection, Korpelevich and Popov methods with and without clipping after 100 epochs.

Method	KL(p_{real}, p_{fake})	Ratio of High Quality Samples	Mode Coverage
Projection w/o clipping	2.286	0.055	5
Projection with clipping	0.053	0.348	8
Popov w/o clipping	2.142	0.0593	5
Popov with clipping	0.036	0.306	8
Korpelevich w/o clipping	2.286	0.055	5
Korpelevich with clipping	0.002	0.538	8