Perseus: A Simple and Optimal High-Order Method for Variational Inequalities

Tianyi Lin DARREN_LIN@BERKELEY.EDU
Michael. I. Jordan JORDAN@CS.BERKELEY.EDU
University of California, Berkeley

Abstract

This paper settles an open and challenging question pertaining to the design of simple and optimal high-order methods for solving smooth and monotone variational inequalities (VIs). A VI involves finding $x^\star \in \mathcal{X}$ such that $\langle F(x), x - x^\star\rangle \geq 0$ for all $x \in \mathcal{X}$ and we consider the setting in which $F : \mathbb{R}^d \mapsto \mathbb{R}^d$ is smooth with up to $(p-1)^{\text{th}}$-order derivatives. For $p = 2$, the cubic regularized Newton's method was extended to VIs with a global rate of $O(\epsilon^{-1})$ [37] while another second-order method was proposed and achieved an improved rate of $O(\epsilon^{-2/3}\log\log(1/\epsilon))$, but this method required a nontrivial line search procedure as an inner loop. High-order methods based on similar search procedures have been further developed and shown to achieve a rate of $O(\epsilon^{-2/(p+1)}\log\log(1/\epsilon))$ [5, 21, 27]. However, such procedures require fine tuning parameters and might be computationally prohibitive in practice [40], leaving the problem of developing a simple and optimal high-order VI method indeed open in the optimization theory. In this paper, we propose a p^{th}-order method that does *not* require any search procedure and provably converges to a weak solution at a rate of $O(\epsilon^{-2/(p+1)})$. We prove that our p^{th}-order method is optimal in the monotone setting by establishing a lower bound of $\Omega(\epsilon^{-2/(p+1)})$ under a standard linear span assumption. A version with restarting attains a global linear and local superlinear convergence rate for smooth and strongly monotone VIs. Furthermore, our method achieves a global rate of $O(\epsilon^{-2/p})$ for solving smooth and non-monotone VIs satisfying the Minty condition. The restarted version again attains a global linear and local superlinear convergence rate if the strong Minty condition is satisfied.

1. Introduction

Let $\mathbb{R}^d$ be a finite-dimensional Euclidean space and let $\mathcal{X} \subseteq \mathbb{R}^d$ be a closed, convex and bounded set with a diameter $D > 0$. Given that $F : \mathbb{R}^d \mapsto \mathbb{R}^d$ is a continuous operator, a fundamental assumption in optimization theory, generalizing convexity, is that F is *monotone*:

$$\langle F(x) - F(x'), x - x'\rangle \geq 0, \quad \text{for all } x, x' \in \mathbb{R}^d.$$

Another useful assumption in this context is that F is $(p-1)^{\text{th}}$-order L-smooth; in particular, that it has Lipschitz-continuous $(p-1)^{\text{th}}$-order derivative ($p \geq 1$) in the sense that there exists a constant $L > 0$ such that

$$\|\nabla^{(p-1)}F(x) - \nabla^{(p-1)}F(x')\|_{\text{op}} \leq L\|x - x'\|, \quad \text{for all } x, x' \in \mathbb{R}^d. \tag{1}$$

With these assumptions, we can formulate the main problem of interest in this paper—the *Minty variational inequality* problem [30]. This consists in finding a point $x^\star \in \mathcal{X}$ such that

$$\langle F(x), x - x^\star\rangle \geq 0, \quad \text{for all } x \in \mathcal{X}. \tag{2}$$

The solution to Eq. (2) is referred to as a *weak* solution to the variational inequality (VI) corresponding to F and $\mathcal{X}$ [13]. By way of comparison, the *Stampacchia variational inequality* problem [20] consists in finding a point $x^\star \in \mathcal{X}$ such that

$$\langle F(x^\star), x - x^\star \rangle \geq 0, \quad \text{for all } x \in \mathcal{X}, \tag{3}$$

and the solution to Eq. (3) is called a *strong* solution to the VI corresponding to F and $\mathcal{X}$. In the setting where F is continuous and monotone, the solution sets of Eq. (2) and Eq. (3) are equivalent. However, these two solution sets are different in general and a weak solution needs not exist when a strong solution exists. In addition, computing an approximate strong solution involves a higher computational burden than finding an approximate weak solution [7, 32, 33].

VIs capture a wide range of problems in optimization theory and beyond, including saddle-point problems and models of equilibria in the game-theoretic settings [9, 22, 45]. Moreover, the challenge of designing solution methods for VIs with provable worst-case bounds has been a central topic during several past decades; see [13, 19]. These foundations have been inspirational to machine learning researchers in recent years, where general saddle-point problems have found applications, including generative adversarial networks (GANs) [16] and multi-agent learning in games [6, 29]. Some of these applications in ML induce a non-monotone structure, with representative examples including the training of robust neural networks [28] or robust classifiers [44].

Building on seminal work in the context of high-order optimization [2, 3], we tackle the challenge of developing p^{th}-order methods for VIs via an inexact solution of regularized subproblems based on a $(p-1)^{\text{th}}$-order Taylor expansion of F. Accordingly, we make the following assumptions throughout this paper.

A1. $F : \mathbb{R}^d \mapsto \mathbb{R}^d$ is $(p-1)^{\text{th}}$-order L-smooth.

A2. The subproblem based on a $(p-1)^{\text{th}}$-order Taylor expansion of F and a convex and bounded set $\mathcal{X}$ can be computed approximately in an efficient manner (see Section **??** for details).

For the first-order VI methods ($p = 1$), [35] has proved that the extragradient (EG) method [1, 23] converges to a weak solution with a global rate of $O(\epsilon^{-1})$ if F is monotone and Eq. (1) holds. There are other methods that achieve the same global rate, including forward-backward splitting method [46], optimistic gradient (OG) method [24, 31, 43] and dual extrapolation method [38]. All these above methods match the lower bound of [41] and are thus optimal.

Comparing to first-order counterparts, the investigations of second-order and high-order methods ($p \geq 2$) are rare, as exploiting high-order derivative information is much more involved for VIs [34, 37]. Aiming to fill this gap, some recent works studied high-order extension of first-order VI methods [5, 21, 27]. These extensions could attain a rate of $O(\epsilon^{-2/(p+1)} \log\log(1/\epsilon))$ but require nontrivial line search procedures at each iteration. These line search procedures require fine tuning parameters and can be computationally prohibitive from a practical viewpoint. Thus, the problem of designing a **simple** and **optimal** high-order method remains open. Indeed, [40, Page 305] noted the difficulty of removing the line search procedure without sacrificing the rate of convergence and highlighted this as an open and challenging question. We summarize the problem as follows:

Can we design a simple and optimal p^{th}-order VI method without line search?

In this paper, we present an affirmative answer to this problem by identifying a p^{th}-order method that achieves a global rate of $O(\epsilon^{-2/(p+1)})$ while dispensing entirely with the line search inner loop.

The core idea of the proposed method is to incorporate a simple and elegant adaptive strategy into a straightforward high-order generalization of dual extrapolation method.

There are three main reasons why we choose the dual extrapolation method as a base algorithm for our high-order methods. First, the dual extrapolation method has its own merit as summarized in [38] and the first second-order VI method with a global convergence rate of $O(\epsilon^{-1})$ [37] was developed based on the dual extrapolation step. In this context, our method can be interpreted as a simplification and generalization of this method. Second, the dual extrapolation step is known as an important conceptual algorithmic component in the optimization literature and there exists a close relationship between extrapolation and acceleration in the context of first-order methods for smooth convex optimization [25, 26]. This is different from the EG method which is in fact an approximation proximal point method [31]. Thus, it would deepen our understanding of the dual extrapolation step if we can design the simple and optimal high-order VI method based on this scheme. Finally, the multi-agent/game-theoretical online learning is a natural application of the VI methods. In this context, the dual extrapolation method outperforms the EG method in some certain aspects; indeed, one of the common performance guarantee for measuring the algorithm is *no-regret*. It is well known that the dual extrapolation method is a no-regret learning algorithm [29]. In contrast, the EG method was recently shown to be not a no-regret learning algorithm [15]. Although it remains unclear how to study high-order methods using the online learning perspective, we believe that it is worth developing the high-order dual extrapolation methods for equilibrium computation.

Contributions. Our contribution can be summarized as follows:

1. We present a new p^{th}-order method for solving smooth and monotone VIs where F has a Lipschitz continuous $(p-1)^{\text{th}}$-order derivative and $\mathcal{X}$ is convex and bounded. We prove that the number of calls of subproblem solvers required by our method to find an ϵ-weak solution is bounded by
$$O\left(\left(\frac{LD^{p+1}}{\epsilon}\right)^{\frac{2}{p+1}}\right).$$
We prove that our method is optimal by establishing the matching lower bound under a linear span assumption. Moreover, we present a restarted version of our method for solving smooth and strongly monotone VIs, i.e., there exists a constant $\mu > 0$ such that
$$\langle F(x) - F(x'), x - x' \rangle \geq \mu \|x - x'\|^2, \quad \text{for all } x, x' \in \mathbb{R}^d.$$
We show that the number of calls of subproblem solvers required to find $\hat{x} \in \mathcal{X}$ satisfying $\|\hat{x} - x^\star\| \leq \epsilon$ is bounded by
$$O\left((\kappa D^{p-1})^{\frac{2}{p+1}} \log_2\left(\frac{D}{\epsilon}\right)\right),$$
where $\kappa = L/\mu$ refers to the condition number of F. The restarted version also achieves local superlinear convergence for the case of $p \geq 2$.

2. We show how to modify our framework such that it can be used for solving smooth and non-monotone VIs satisfying the so-called Minty condition (see Definition 5). We also note that

a line search procedure is not required. We prove that the number of calls of subproblem solvers to find an ϵ-strong solution is bounded by

$$O\left(\left(\frac{LD^{p+1}}{\epsilon}\right)^{\frac{2}{p}}\right).$$

The restarted version is developed for solving smooth and non-monotone VIs satisfying the strong Minty condition (see Definition 7). We show that the number of calling subproblem solvers required to find $\hat{x} \in \mathcal{X}$ satisfying $\|\hat{x} - x^\star\| \leq \epsilon$ is bounded by

$$O\left(\max\{(\kappa_{\text{Minty}} D^{p-1})^{\frac{2}{p}}, (\kappa_{\text{Minty}} D^{p-1})^{\frac{2}{p+1}}\} \log_2\left(\frac{D}{\epsilon}\right)\right),$$

where $\kappa_{\text{Minty}} = L/\mu_{\text{Minty}}$ refers to the Minty condition number of F. In addition, the restarted version of our method achieves local superlinear convergence for the case of $p \geq 2$.

2. Preliminaries and Technical Background

The regularity conditions that we consider for an operator $F : \mathbb{R}^d \mapsto \mathbb{R}^d$ are as follows.

Definition 1 *F is k^{th}-order L-smooth if $\|\nabla^{(k)} F(x) - \nabla^{(k)} F(x')\| \leq L\|x - x'\|$ for all x, x'.*

Definition 2 *F is μ-strongly-monotone if $\langle F(x) - F(x'), x - x'\rangle \geq \mu\|x - x'\|^2$ for all x, x'. If $\mu = 0$, we recover the definition of monotonicity for a continuous operator.*

Assumption 3 *The following statements hold true: (i) $F : \mathbb{R}^d \mapsto \mathbb{R}^d$ is $(p-1)^{\text{th}}$-order L-smooth; (ii) $\mathcal{X}$ is closed, convex and bounded with a diameter $D > 0$.*

As for the boundedness condition for $\mathcal{X}$, it is standard in the VI literature [13]. This condition not only guarantees the validity of the most natural optimality criterion in the monotone setting—the gap function [35, 38]—but additionally it is satisfied in a wide range of real-world applications [13]. On the other hand, there is a line of works focusing on relaxing the boundedness condition via appeal to other notions of approximate solutions [7, 32–34]. For simplicity, we retain the boundedness condition and leave the analysis for the cases with unbounded constraint sets to future work.

Under monotonicity, it is well known that any ϵ-approximate strong solution is an ϵ-approximate weak solution but the reverse does not hold in general. These definitions motivate the use of a gap function, $\text{GAP}(\cdot) : \mathcal{X} \mapsto \mathbb{R}_+$, defined by $\text{GAP}(\hat{x}) = \sup_{x \in \mathcal{X}} \langle F(x), \hat{x} - x\rangle$ with which we measure the optimality of a point $\hat{x} \in \mathcal{X}$ output by various iterative solution methods.

Definition 4 *A point* $\hat{\mathrm{x}} \in \mathcal{X}$ *is an* ϵ-weak solution to the monotone VI corresponding to $F : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\mathcal{X} \subseteq \mathbb{R}^d$ *if we have* $\text{GAP}(\hat{x}) \leq \epsilon$. *If* $\epsilon = 0$, *then* $\hat{x} \in \mathcal{X}$ *is a* weak solution.

In the strongly monotone setting, we let $\mu > 0$ denote the modulus of strong monotonicity for F. Under Assumption 3, we define $\kappa := L/\mu$ as the *generalized condition number* of F.

We also study the case in which F is non-monotone but satisfies the (strong) Minty condition.

Definition 5 *The VI corresponding to $F : \mathbb{R}^d \mapsto \mathbb{R}^d$ and $\mathcal{X} \subseteq \mathbb{R}^d$ satisfies the Minty condition if there exists $x^\star \in \mathcal{X}$ such that $\langle F(x), x - x^\star\rangle \geq 0$ for all $x \in \mathcal{X}$.*

Algorithm 1 Perseus(p, x_0, L, T, opt)

Input: order p, initial point $x_0 \in \mathcal{X}$, parameter L, iteration number T and $\mathsf{opt} \in \{0, 1, 2\}$.
Initialization: set $s_0 = 0_d \in \mathbb{R}^d$.
for $k = 0, 1, 2, \ldots, T$ **do**
  STEP 1: If $x_k \in \mathcal{X}$ is a solution of the VI, then **stop**.
  STEP 2: Compute $v_{k+1} = \text{argmax}_{v \in \mathcal{X}}\{\langle s_k, v - x_0\rangle - \frac{1}{2}\|v - x_0\|^2\}$.
  STEP 3: Compute $x_{k+1} \in \mathcal{X}$ such that Eq. (6) holds true.
  STEP 4: Compute $\lambda_{k+1} > 0$ such that $\frac{1}{20p-8} \leq \frac{\lambda_{k+1} L\|x_{k+1} - v_{k+1}\|^{p-1}}{p!} \leq \frac{1}{10p+2}$.
  STEP 5: Compute $s_{k+1} = s_k - \lambda_{k+1} F(x_{k+1})$.
end for
Output: $\hat{x} = \begin{cases} \tilde{x}_T = \frac{1}{\sum_{k=1}^T \lambda_k} \sum_{k=1}^T \lambda_k x_k, & \text{if } \mathsf{opt} = 0, \\ x_{k_T} \text{ for } k_T = \text{argmin}_{1 \leq k \leq T} \|x_k - v_k\|, & \text{else if } \mathsf{opt} = 1, \\ x_T, & \text{otherwise.} \end{cases}$

Accordingly, we define the residue function $\text{RES}(\cdot) : \mathcal{X} \mapsto \mathbb{R}_+$ given by

$$\text{RES}(\hat{x}) = \sup_{x \in \mathcal{X}} \langle F(\hat{x}), \hat{x} - x\rangle, \tag{4}$$

which measures the optimality of a point $\hat{x} \in \mathcal{X}$ achieved by iterative solution methods.

Definition 6 *A point* $\hat{\mathrm{x}} \in \mathcal{X}$ *is an* ϵ*-strong solution to the non-monotone VI corresponding to* $F : \mathbb{R}^d \mapsto \mathbb{R}^d$ *and* $\mathcal{X} \subseteq \mathbb{R}^d$ *if we have* $\text{RES}(\hat{x}) \leq \epsilon$*. If* $\epsilon = 0$*, then* $\hat{x} \in \mathcal{X}$ *is a strong solution.*

Proceeding a step further, we define the strong Minty condition and define $\kappa_{\text{Minty}} := L/\mu_{\text{Minty}}$ to be the *Minty condition number* of F if the VI satisfies the μ_{Minty}-strong Minty condition.

Definition 7 *The VI corresponding to* $F : \mathbb{R}^d \mapsto \mathbb{R}^d$ *and* $\mathcal{X} \subseteq \mathbb{R}^d$ *satisfies the* μ_{Minty}*-strong Minty condition if there exists* $x^\star \in \mathcal{X}$ *such that* $\langle F(x), x - x^\star\rangle \geq \mu_{\text{Minty}}\|x - x^\star\|^2$ *for all* $x \in \mathcal{X}$*.*

There are many application problems that can be formulated as non-monotone VIs satisfying (strong) Minty condition, including product pricing [8, 12, 14] and competitive exchange economies [4].

3. Our Method

We summarize our p^{th}-order method, which we denote as Perseus(p, x_0, L, T, opt), in Algorithm 1. Its major novelty lies in an adaptive strategy used for updating λ_{k+1} (see **Step 4**). This modification is simple yet important. It serves as the key for obtaining a global rate of $O(\epsilon^{-2/(p+1)})$ (monotone) and $O(\epsilon^{-2/p})$ (non-monotone with the Minty condition) under Assumption 3. Our methods also allow the subproblem to be solved inexactly and we give options for choosing the type of outputs.

We remark that **Step 3** resorts to the computation of an approximate strong solution to the VI in which we define $F_{v_{k+1}}(x)$ as the sum of a high-order polynominal and a regularization term:

$$F_{v_{k+1}}(x) = F(v_{k+1}) + \ldots + \tfrac{1}{(p-1)!}\nabla^{(p-1)} F(v_{k+1})[x - v_{k+1}]^{p-1} + \tfrac{5L}{(p-1)!}\|x - v_{k+1}\|^{p-1}(x - v_{k+1}),$$

where we write the VI of interest in the subproblem as follows:

$$\text{Find } x_{k+1} \in \mathcal{X} \text{ such that } \langle F_{v_{k+1}}(x_{k+1}), x - x_{k+1}\rangle \geq 0 \text{ for all } x \in \mathcal{X}. \tag{5}$$

Algorithm 2 Perseus-restart(p, x_0, L, σ, D, T, opt)

Input: order p, initial point $x_0 \in \mathcal{X}$, parameters L, σ, D, iteration number T and $\textsf{opt} \in \{0, 1, 2\}$.

Initialization: set $T_{\text{inner}} = \begin{cases} \lceil (\frac{2^{p+1}(5p-2)}{p!} \frac{LD^{p-1}}{\sigma})^{\frac{2}{p+1}} \rceil, & \text{if } \textsf{opt} = 0, \\ \lceil (\frac{2^{p+2}(5p+1)}{p!} \frac{LD^{p-1}}{\sigma})^{\frac{2}{p}} + \left(\frac{2^{p+5}}{p!} \frac{LD^{p-1}}{\sigma}\right)^{\frac{2}{p+1}} \rceil, & \text{elseif } \textsf{opt} = 1, \\ 1, & \text{otherwise.} \end{cases}$

for $k = 0, 1, 2, \ldots, T$ **do**

  STEP 1: If $x_k \in \mathcal{X}$ is a solution of the VI, then **stop**.

  STEP 2: Compute $x_{k+1} = \textsf{Perseus}(p, x_k, L, T_{\text{inner}}, \textsf{opt})$.

end for

Output: x_{T+1}.

Since $F_{v_{k+1}}$ is continuous and $\mathcal{X}$ is closed, convex and bounded, Harker and Pang [19, Theorem 3.1] guarantees that a strong solution to the VI in Eq. (5) exists.

In the monotone setting, the VI in Eq. (5) is monotone and thus computationally tractable [10], with the following approximation condition:

$$\sup_{x \in \mathcal{X}} \langle F_{v_{k+1}}(x_{k+1}), x_{k+1} - x \rangle \leq \tfrac{L}{p!} \|x_{k+1} - v_{k+1}\|^{p+1}. \tag{6}$$

Indeed, if $p = 1$, we have that $\nabla F_{v_{k+1}}(x) = L \cdot I_{d\times d}$ that is positive semidefinite for all $x \in \mathbb{R}^d$ where $I_{d\times d} \in \mathbb{R}^{d\times d}$ is an identity matrix. Otherwise, we obtain from Assumption 3 that

$$\nabla F_{v_{k+1}}(x) \succeq \nabla F(x) + \tfrac{4L}{(p-1)!}\|x - v_{k+1}\|^{p-1} I_{d\times d} + \tfrac{L}{(p-2)!}\|x - v_{k+1}\|^{p-2}(x - v_{k+1})(x - v_{k+1})^\top.$$

In the non-monotone setting, the VI in Eq. (5) is not necessarily monotone and the computation of an approximate strong solution is intractable in general [11]. However, we note that $F_{v_{k+1}}$ is the sum of a high-order polynomial and a regularization term; this special structure might lend itself to efficient numerical methods. In optimization setting, this problem can be solved efficiently [17, 18].

Restarting. We summarize the restarted version of our p^{th}-order method in Algorithm 2. This method, which we refer to as Perseus-restart(p, x_0, L, σ, D, T, opt), combines Algorithm 1 with a restart scheme [36, 39, 40, 42]. Intuitively, the restart scheme stops an algorithm when a criterion is satisfied and then restarts the algorithm with an new input. At each iteration of Algorithm 2, we use $x_{k+1} = \textsf{Perseus}(p, x_k, L, t, \textsf{opt})$ as a subroutine. In other words, we restart Perseus every $t \geq 1$ iterations and take advantage of average iterates or best iterates to generate x_{k+1} from x_k. Moreover, the choice of t can be specialized to different settings and/or different type of convergence guarantees. Indeed, we set $\textsf{opt} = 0$ for the strong monotone setting, $\textsf{opt} = 1$ for the non-monotone setting satisfying the strong Minty condition and $\textsf{opt} = 2$ for a local convergence guarantee.

References

[1] A. S. Antipin. Method of convex programming using a symmetric modification of Lagrange function. *Matekon*, 14(2):23–38, 1978.

[2] M. Baes. Estimate sequence methods: extensions and approximations. *Institute for Operations Research, ETH, Zürich, Switzerland*, 2009.

[3] E. G. Birgin, J. L. Gardenghi, J. M. Martínez, S. A. Santos, and P. L. Toint. Worst-case evaluation complexity for unconstrained nonlinear optimization using high-order regularized models. *Mathematical Programming*, 163(1-2):359–368, 2017.

[4] L. Brighi and R. John. Characterizations of pseudomonotone maps and economic equilibrium. *Journal of Statistics and Management Systems*, 5(1-3):253–273, 2002.

[5] B. Bullins and K. A. Lai. Higher-order methods for convex-concave min-max optimization and monotone variational inequalities. *SIAM Journal on Optimization*, 32(3):2208–2229, 2022.

[6] N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

[7] Y. Chen, G. Lan, and Y. Ouyang. Accelerated schemes for a class of variational inequalities. *Mathematical Programming*, 165(1):113–149, 2017.

[8] S. C. Choi, W. S. DeSarbo, and P. T. Harker. Product positioning under price competition. *Management Science*, 36(2):175–199, 1990.

[9] R. Cottle, F. Giannessi, and J-L. Lions. *Variational Inequalities and Complementarity Problems: Theory and Applications*. John Wiley & Sons, 1980.

[10] C. D. Dang and G. Lan. On the convergence properties of non-Euclidean extragradient methods for variational inequalities with generalized monotone operators. *Computational Optimization and Applications*, 60(2):277–310, 2015.

[11] C. Daskalakis, S. Skoulakis, and M. Zampetakis. The complexity of constrained min-max optimization. In *STOC*, pages 1466–1478, 2021.

[12] C. Ewerhart. Cournot games with biconcave demand. *Games and Economic Behavior*, 85: 37–47, 2014.

[13] F. Facchinei and J-S. Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems*. Springer Science & Business Media, 2007.

[14] G. Gallego and M. Hu. Dynamic pricing of perishable assets under competition. *Management Science*, 60(5):1241–1259, 2014.

[15] N. Golowich, S. Pattathil, and C. Daskalakis. Tight last-iterate convergence rates for no-regret learning in multi-player games. In *NeurIPS*, pages 20766–20778, 2020.

[16] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In *NeurIPS*, pages 2672–2680, 2014.

[17] N. I. M Gould, S. Lucidi, M. Roma, and P. L. Toint. Solving the trust-region subproblem using the Lanczos method. *SIAM Journal on Optimization*, 9(2):504–525, 1999.

[18] N. I. M. Gould, D. P. Robinson, and H. S. Thorne. On solving trust-region and other regularised subproblems in optimization. *Mathematical Programming Computation*, 2(1):21–57, 2010.

[19] P. T. Harker and J-S. Pang. Finite-dimensional variational inequality and nonlinear complementarity problems: A survey of theory, algorithms and applications. *Mathematical Programming*, 48(1):161–220, 1990.

[20] P. Hartman and G. Stampacchia. On some non-linear elliptic differential-functional equations. *Acta Mathematica*, 115:271–310, 1966.

[21] R. Jiang and A. Mokhtari. Generalized optimistic methods for convex-concave saddle point problems. *ArXiv Preprint: 2202.09674*, 2022.

[22] D. Kinderlehrer and G. Stampacchia. *An Introduction to Variational Inequalities and Their Applications*. SIAM, 2000.

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

[24] 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.

[25] G. Lan and Y. Zhou. An optimal randomized incremental gradient method. *Mathematical Programming*, 171(1):167–215, 2018.

[26] G. Lan and Y. Zhou. Random gradient extrapolation for distributed and stochastic optimization. *SIAM Journal on Optimization*, 28(4):2753–2782, 2018.

[27] T. Lin and M. I. Jordan. On monotone inclusions, acceleration and closed-loop control. *ArXiv Preprint: 2111.08093*, 2021.

[28] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu. Towards deep learning models resistant to adversarial attacks. In *ICLR*, 2018. URL https://openreview.net/forum?id=rJzIBfZAb.

[29] P. Mertikopoulos and Z. Zhou. Learning in games with continuous action sets and unknown payoff functions. *Mathematical Programming*, 173(1):465–507, 2019.

[30] G. J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29(3):341–346, 1962.

[31] A. Mokhtari, A. E. Ozdaglar, and S. Pattathil. Convergence rate of o(1/k) for optimistic gradient and extragradient methods in smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 30(4):3230–3251, 2020.

[32] R. D. C. Monteiro and B. F. Svaiter. On the complexity of the hybrid proximal extragradient method for the iterates and the ergodic mean. *SIAM Journal on Optimization*, 20(6):2755–2787, 2010.

[33] R. D. C. Monteiro and B. F. Svaiter. Complexity of variants of Tseng's modified FB splitting and Korpelevich's methods for hemivariational inequalities with applications to saddle-point and convex optimization problems. *SIAM Journal on Optimization*, 21(4):1688–1720, 2011.

[34] R. D. C. Monteiro and B. F. Svaiter. Iteration-complexity of a Newton proximal extragradient method for monotone variational inequalities and inclusion problems. *SIAM Journal on Optimization*, 22(3):914–935, 2012.

[35] A. Nemirovski. Prox-method with rate of convergence o(1/t) for variational inequalities with Lipschitz continuous monotone operators and smooth convex-concave saddle point problems. *SIAM Journal on Optimization*, 15(1):229–251, 2004.

[36] A. S. Nemirovskii and Y. E. Nesterov. Optimal methods of smooth convex minimization. *USSR Computational Mathematics and Mathematical Physics*, 25(2):21–30, 1985.

[37] Y. Nesterov. Cubic regularization of Newton's method for convex problems with constraints. Technical report, Université catholique de Louvain, Center for Operations Research and Econometrics (CORE), 2006.

[38] Y. Nesterov. Dual extrapolation and its applications to solving variational inequalities and related problems. *Mathematical Programming*, 109(2):319–344, 2007.

[39] Y. Nesterov. Gradient methods for minimizing composite functions. *Mathematical Programming*, 140(1):125–161, 2013.

[40] Y. Nesterov. *Lectures on Convex Optimization*, volume 137. Springer, 2018.

[41] Y. Ouyang and Y. Xu. Lower complexity bounds of first-order methods for convex-concave bilinear saddle-point problems. *Mathematical Programming*, 185(1):1–35, 2021.

[42] B. O'donoghue and E. Candes. Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics*, 15(3):715–732, 2015.

[43] 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.

[44] A. Sinha, H. Namkoong, and J. Duchi. Certifiable distributional robustness with principled adversarial training. In *ICLR*, 2018. URL https://openreview.net/forum?id=Hk6kPgZA-.

[45] R. Trémolières, J-L. Lions, and R. Glowinski. *Numerical Analysis of Variational Inequalities*. Elsevier, 2011.

[46] P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38(2):431–446, 2000.