OPT2024: 16th Annual Workshop on Optimization for Machine Learning

On the Hypomonotone Class of Variational Inequalities

Khaled Alomar KHALED.ALOMAR@CISPA.DE
CISPA Helmholtz Center for Information Security

Tatjana Chavdarova TATJANA.CHAVDAROVA@BERKELEY.EDU
University of California, Berkeley & CISPA Helmholtz Center for Information Security

Abstract

This paper studies the behavior of the extragradient algorithm when applied to hypomonotone operators, a class of problems that extends beyond the classical monotone setting. While the extragradient method is widely known for its efficacy in solving variational inequalities with monotone and Lipschitz continuous operators, we demonstrate that its convergence is not guaranteed in the hypomonotone setting. We construct an example such that by choosing the starting point for all every size the extragradient diverges. Our results highlight the necessity of stronger assumptions to guarantee convergence of extragradient and to further develop the existing VI methods for broader problems.

1. Introduction

We are interested in numerically solving the following *Variational Inequality* [VI, 5, 14] problem:

$$\textit{find } x^\star \quad \text{s.t.} \quad \langle x - x^\star, F(x^\star)\rangle \geq 0, \quad \forall x \in \mathbb{R}^n \,. \tag{VI}$$

Special instances of (VI) include standard minimization with $F \equiv \nabla f$, $f : \mathbb{R}^n \to \mathbb{R}$, zero-sum min-max, and general sum multi-player games. This problem has gained recent popularity in machine learning, due to several instances that cannot be modeled by minimization only, such as Generative Adversarial Networks [6], robust versions of classification [e.g., 3, 11, 12, 15], actor-critic methods [8], and multi-agent reinforcement learning [e.g., 1, 10].

The extragradient algorithm, introduced by Korpelevich [9] in 1976, is a fundamental iterative method used to solve variational inequality problems involving monotone operators. For a detailed description of the algorithm, refer to § 3. Unlike gradient descent, the latest output (last iterate) of the extragradient method converges when applied to monotone and Lipschitz continuous operators; definitions can be found in § 3. Its popularity is due to its simplicity and efficiency for this class of problems.

However, in real-world applications, operators may not always exhibit strict monotonicity or co-coercivity. Instead, they may demonstrate hypomonotonicity, which is a weaker form of the monotonicity condition. Hypomonotone operators allow for the exploration of more general scenarios, where finding solutions is more challenging, yet still feasible for analysis.

Hypomonotonicity appears in various settings, including equilibrium problems, optimization involving non-convex structures, and certain game-theoretic models. The following inequality characterizes this problem class:

$$\langle F(x) - F(y), x - y\rangle \geq -\mu \|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n, \tag{HM}$$

where $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is the operator in question and $\mu \geq 0$ is a constant quantifying the degree of hypomonotonicity.

2. Related Works

Monotonicity and sub-classes problems. The VI problem has been widely studied. The fundamental work by Korpelevich [9] assume monotonicity. According to Chavdarova [2] methods such EG converges when the operator is monotone but not nessesarly for general operator. Beyond monotone class, extragradient type of algorithm are studied [13] in setting where monotonicity is not assumed. Two main classes are in our interest mentioned below.

Cohypomonotonicity. Cohypomontone operators on Hilbert spaces are considered a more wider class of operator or extension of monotone operators. Combettes and Pennanen [4] studied the convergence of proximal point algorithm in the cohypomontone setting. As mentioned in the appendix we can see that an operator F is cohypomontone when it's inverse of F^{-1} is hypomontone. This relation lead us to study the class of hypomotone operators and see how can we characterize it with the aim to understand the class more. In particular, we focus on establishing why the convergence of the extragradient method is not guaranteed in the hypomonotone setting.

Hypomonotonicity. Hypomonotone class of operators are mentioned in this work [7], two variant classes are metnioned, ρ–$hypomonotone$ and $maximal$ ρ–$hypomonotone$. However, to our knowledge, the convergence of the VI methods in this class has not been studied.

3. Preliminaries

This section describes the necessary definitions. Further used background and lemmas are given in Appendix 7.1.
Notation. We denote *(i)* real-valued functions with small letters *(ii)* operators with capital letters, *(iii)* matrices with curly capital letters. $\mathcal{A}^\dagger$ denotes the complex conjugate of the matrix $\mathcal{A}$.
Extragradient [9]. We study the extragradient algorithm defined by the following update at iteration k:

$$y_{k+1} = x_k - \gamma F(x_k),$$
$$x_{k+1} = x_k - \gamma F(y_{k+1})\,.$$

Monotonicity is defined as follows.

Definition 1 (Monotonicity) *An operator $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is said to be monotone iff:*

$$\langle F(x) - F(y), x - y\rangle \geq 0\,, \qquad \forall x, y \in \mathbb{R}^n. \tag{Mnt}$$

The following definition introduces a relaxed form of monotonicity where an operator's inner product with the difference of its arguments is bounded below by a negative value.

Definition 2 (Hypomonotonicity) *An operator $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is hypomonotone with modulus $\mu \geq 0$ iff:*

$$\langle F(x) - F(y), x - y\rangle \geq -\mu\|x - y\|^2, \quad \forall x, y \in \mathbb{R}^n. \tag{HM}$$

The following property ensures bounded changes in F with respect to changes in its input, which is essential for proving convergence results in iterative algorithms.

Definition 3 (Lipschitz operator) *An operator $F\colon \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with constant $L > 0$ if:*

$$\|F(x) - F(y)\| \leq L\|x - y\|, \quad \forall x, y \in \mathbb{R}^n. \tag{Lip}$$

4. Examples of non-monotone and Hypomonotone Operators

Example 1: concave problem. Consider the following operator $T : \mathbb{R}^2 - > \mathbb{R}^2, T(x) = -\mu x, \mu > 0$. Consider $x = (1,0)^t, y = (0,1)^t \in \mathbb{R}^2, \mu = 1$. It follows that $\langle T(x) - T(y), x - y\rangle = -2 < 0$. Hence not monotone. Now to show that T is hypomontone consider $\langle -\mu x + \mu y, x - y\rangle = -\mu\langle x - y, x - y\rangle \geq -\mu\|x - y\|^2$, since $\mu > 0$ then it is $\mu-$hypomonotone. Furthermore $\|T(x) - T(y)\| = |-\mu|\|x - y\| \leq \mu\|x - y\|$, hence $\mu-$Lipchitz.

Example 2: non-monotone problem. Consider the operator $F(x, y) = (2x+y-1, -x-1.5y+1)$ to show that F is not $montone$ we use the vectors $x_1 = (0,0)^T, x_2 = (0,1)^T$ we get $\langle F(x_1) - F(x_2), x_1 - x_2\rangle = -\frac{3}{2} < 0$.

5. Main Results

While monotone problems have been extensively studied, the class of hypomonotone problems remains less understood despite its practical importance. Furthermore, we study the convergence behavior in the hypomonotone setting to uncover potential challenges and deviations from standard monotone behavior.

5.1. Theorem: Hypomonotonicity of Operators with Negative Eigenvalues

Theorem 4 *Let $\mathcal{A} \in \mathbb{R}^{n\times n}$ be a normal matrix with at least one eigenvalue whose real part is negative. Then, the operator related to matrix $\mathcal{A}$ is hypomonotone but not monotone.*

Step 1: Non-Monotonicity Since $\mathcal{A}$ has at least one eigenvalue with a negative real part, it cannot be monotone. This follows directly from the definition of monotonicity.

Step 2: Diagonalization and Eigenbasis Representation Since $\mathcal{A}$ is normal, it can be diagonalized as $\mathcal{A} = Q\Lambda Q^\top$, where Λ is the diagonal matrix of eigenvalues, and Q is an orthogonal matrix. Any vector $x \in \mathbb{C}^n$ can be written as a linear combination of eigenvectors:

$$x = \sum_{i=1}^{n} c_i v_i,$$

where v_i are the eigenvectors of $\mathcal{A}$, and c_i are the complex coefficients.

Step 3: Inner Product Expansion Using Eigenbasis To prove hypomonotonicity, we analyze the expression:

$$\langle x_1^* - x_2^*, x_1 - x_2\rangle = Re(\langle \mathcal{A}(x_1 - x_2), x_1 - x_2\rangle),$$

where $x_1 = \sum a_i v_i$ and $x_2 = \sum b_i v_i$. Expanding the inner product in terms of the eigenvalues λ_i, we get:

$$Re\left(\sum_{i=1}^{n} \lambda_i(|a_i|^2 + |b_i|^2 - 2Re(a_i b_i))\right).$$

Step 4: Bounding the Inner Product We now bound the inner product using the real parts of the eigenvalues:

$$Re\left(\sum_{i=1}^{n}\lambda_i(|a_i|^2+|b_i|^2-2Re(a_ib_i))\right)\geq\sum_{i=1}^{n}Re(\lambda_i)|a_i-b_i|^2.$$

Since $\lambda_{\min}$ denotes the eigenvalue of $\mathcal{A}$ with the smallest real part, we have:

$$\sum_{i=1}^{n}Re(\lambda_i)|a_i-b_i|^2\geq\mathrm{Re}(\lambda_{\min})\|x_1-x_2\|^2.$$

Step 5: Hypomonotonicity Conclusion Because the real part of the smallest eigenvalue $\mathrm{Re}(\lambda_{\min})$ is negative, we have:

$$\langle\mathcal{A}(x_1-x_2),x_1-x_2\rangle\geq-\mu\|x_1-x_2\|^2$$

for some $\mu\geq 0$, proving that the operator related to $\mathcal{A}$ is hypomonotone. The operator is not monotone because of the negative real part of $\lambda_{\min}$.

Theorem 5 *Let $\gamma>0$, $x_0=(1,0)^\top$ the the initial point and $F(x)=Ax$ to be an Operator such that*

$$A=\begin{pmatrix}-2&0\\-1&-2\end{pmatrix}$$

Then the sequence $\|x_n\|$ diverges as $n\to\infty$.

Proof By definition of extragradient, $y_{k+1}=x_k-\gamma Ax_k$ and $x_{k+1}=x_k-\gamma Ay_{k+1}$ hence it follows that $x_{k+1}=x_k(I-\gamma A+\gamma^2A^2)$.
Define $M(\gamma)=I-\gamma A+\gamma^2A^2$. By definition of A we can see that

$$M(\gamma)=\begin{pmatrix}1+2\gamma+4\gamma^2&0\\\gamma+4\gamma^2&1+2\gamma+4\gamma^2\end{pmatrix}.$$

Now consider x_0 as defined above and $x_1=M(\gamma)x_0=(1+2\gamma+4\gamma^2,\gamma+4\gamma^2)$.
let $\|.\|$ to be the euclidean norm and $\forall 0<\gamma<1$

$$\|x_1\|^2=(1+2\gamma+4\gamma^2)^2+(\gamma+4\gamma^2)^2>1,$$

now calculating x_2 we see

$$x_2=M(\gamma)x_1=\begin{pmatrix}(1+2\gamma+4\gamma^2)^2\\2(\gamma+4\gamma^2)(1+2\gamma+4\gamma^2)\end{pmatrix},$$

and $\|x_2\|>\|x_1\|$, furthermore iteratively

$$x_n=M(\gamma)^nx_1=\begin{pmatrix}(1+2\gamma+4\gamma^2)^n\\n(\gamma+4\gamma^2)(1+2\gamma+4\gamma^2)^{n-1}\end{pmatrix},$$

thus $\|x_n\|>\|x_{n-1}\|$, hence $\|x_n\|\to\infty$. ∎

6. Conclusion

In this paper, we characterized hypomonotone problems by analyzing their properties through the eigenvalues of the associated operators. Additionally, We constructed a counterexample to demonstrate divergence, showing that no matter how the step size is chosen, convergence cannot be guaranteed. This highlights the critical need for a deeper understanding of hypomonotone operators and their unique challenges.

References

[1] D. Bertsekas. *Rollout, Policy Iteration, and Distributed Reinforcement Learning*. Athena scientific optimization and computation series. Athena Scientific, 2021. ISBN 9781886529076.

[2] Tatjana Chavdarova, Ya-Ping Hsieh, and Michael I. Jordan. Continuous-time analysis for variational inequalities: An overview and desiderata, 2022.

[3] Rune Christiansen, Niklas Pfister, Martin Emil Jakobsen, Nicola Gnecco, and Jonas Peters. A causal framework for distribution generalization. *arXiv:2006.07433*, 2020.

[4] Patrick L. Combettes and Teemu Pennanen. Proximal methods for cohypomonotone operators. *SIAM Journal on Control and Optimization*, 43(2):731–742, 2004.

[5] Francisco Facchinei and Jong-Shi Pang. *Finite-Dimensional Variational Inequalities and Complementarity Problems Vol I*. Springer-Verlag, 2003.

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *Advances in Neural Information Processing Systems*, volume 27, 2014.

[7] A. N. Iusem, T. Pennanen., and B. F. Svaiter. Inexact variants of the proximal point algorithm without monotonicity. *SIAM Journal on Optimization*, 13(4):1080–1097, 2003. doi: 10.1137/S1052623401399587.

[8] Vijay Konda and John Tsitsiklis. Actor-critic algorithms. In S. Solla, T. Leen, and K. Müller, editors, *Advances in Neural Information Processing Systems*, volume 12. MIT Press, 1999.

[9] G. M. Korpelevich. The extragradient method for finding saddle points and other problems. *Ekonomika i Matematicheskie Metody*, 12(4):747–756, 1976.

[10] Ryan Lowe, Yi Wu, Aviv Tamar, Jean Harb, Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. *Neural Information Processing Systems (NIPS)*, 2017.

[11] Aleksander Madry, Aleksandar Makelov, Ludwig Schmidt, Dimitris Tsipras, and Adrian Vladu. Towards deep learning models resistant to adversarial attacks. In *ICLR*, 2018.

[12] Santiago Mazuelas, Andrea Zanoni, and Aritz Pérez. Minimax classification with 0-1 loss and performance guarantees. In *NeurIPS*, volume 33, pages 302–312. Curran Associates, Inc., 2020.

[13] Thomas Pethick, Olivier Fercoq, Puya Latafat, Panagiotis Patrinos, and Volkan Cevher. Solving stochastic weak minty variational inequalities without increasing batch size, 2023.

[14] Guido Stampacchia. Formes Bilineaires Coercitives Sur Les Ensembles Convexes. *Académie des Sciences de Paris*, 258:4413–4416, 1964.

[15] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. *arXiv:1312.6199*, 2014.

7. Appendix

A similar structure to hypomonotonicity is *cohypomonotonicity*, which establishes a hypomonotonicity relationship with the inverse operators.

Definition 6 (Cohypomonotonicity) *An operator F is cohypomonotone if its inverse F^{-1} is hypomonotone with respect to a constant $\rho \geq 0$ on the set $F(C)$. For all $x, y \in F(C)$:*

$$\langle F^{-1}(x) - F^{-1}(y), x - y \rangle \geq -\rho \|x - y\|^2. \tag{cHM}$$

7.1. Useful Lemmas

This section lists the necessary lemmas that we rely on.

Normal matrices play a crucial role in diagonalization and spectral analysis.

Definition 7 (Normal Matrix) *A matrix $\mathcal{A} \in \mathbb{C}^{n \times n}$ is normal if it commutes with its conjugate transpose: $\mathcal{A}\mathcal{A}^\dagger = \mathcal{A}^\dagger \mathcal{A}$.*

The following is essential for spectral decomposition and understanding operator behavior via eigenvalues.

Theorem 8 (Spectral Theorem) *Any normal matrix $\mathcal{A}$—as per Def. 7—can be diagonalized by a unitary matrix S, yielding: $\mathcal{A} = SDS^\dagger$, where D is a diagonal matrix of eigenvalues.*

Theorem 9 *Let $\mathcal{A} \in \mathbb{R}^{n \times n}$ be a normal matrix with at least one eigenvalue whose real part is negative. Then, the operator related to matrix $\mathcal{A}$ is hypomonotone and not monotone.*

Proof Since $\mathcal{A}$ has an eigenvalue with a negative real part, it is not monotone; which follows straightforwardly from the spectral viewpoint of the monotonicity definition. The following shows that $\mathcal{A}$ is a hypomonotone operator.

Given that $\mathcal{A}$ is diagonalizable, there exist eigenvalues $\lambda_1, \dots, \lambda_n$ and a corresponding set of linearly independent eigenvectors $v_1, \dots, v_n$. Any vector $x \in \mathbb{C}^n$ can be expressed as $x = \sum_{i=1}^n c_i v_i$, where c_i are complex coefficients.

To prove the hypomonotonicity of $\mathcal{A}$, consider $x_1, x_2 \in \mathbb{C}^n$ with $x_1^* = \mathcal{A}x_1$ and $x_2^* = \mathcal{A}x_2$. We write $x_1 = \sum_{i=1}^n a_i v_i$ and $x_2 = \sum_{i=1}^n b_i v_i$, where v_i are eigenvectors and a_i, b_i are coordinates in this basis.

Using the sesquilinearity of the inner product (which accommodates complex vectors), we have: First we notice that the coordinates of x_1 and x_2 are real:

$$\begin{aligned}
&\langle x_1^* - x_2^*, x_1 - x_2\rangle = Re(\langle x_1^* - x_2^*, x_1 - x_2\rangle)\\
&= Re(\langle \mathcal{A}x_1 - \mathcal{A}x_2, x_1 - x_2\rangle)\\
&= Re(\langle \mathcal{A}x_1, x_1\rangle - \langle \mathcal{A}x_1, x_2\rangle - \langle \mathcal{A}x_2, x_1\rangle + \langle \mathcal{A}x_2, x_2\rangle)\\
&= Re(\sum_{i=1}^{n} \lambda_i |a_i|^2 - \sum_{i=1}^{n} \lambda_i a_i \overline{b_i} - \sum_{i=1}^{n} \lambda_i \overline{a_i} b_i + \sum_{i=1}^{n} \lambda_i |b_i|^2)\\
&= Re(\sum_{i=1}^{n} \lambda_i |a_i|^2 - \sum_{i=1}^{n} \lambda_i a_i \overline{b_i} - \sum_{i=1}^{n} \lambda_i \overline{a_i} b_i + \sum_{i=1}^{n} \lambda_i |b_i|^2)\\
&= Re(\sum_{i=1}^{n} \lambda_i (|a_i|^2 + |b_i|^2 - (a_i \overline{b_i} + \overline{a_i} b_i))\\
&= Re(\sum_{i=1}^{n} \lambda_i (|a_i|^2 + |b_i|^2)) - Re(\sum_{i=1}^{n} \lambda_i 2Re(a_i b_i))\\
&= Re(\sum_{i=1}^{n} \lambda_i (|a_i|^2 + |b_i|^2)) - \sum_{i=1}^{n} Re(\lambda_i) 2Re(a_i b_i))\\
&\geq Re(\sum_{i=1}^{n} \lambda_i (|a_i|^2 + |b_i|^2)) - \sum_{i=1}^{n} Re(\lambda_i) 2|a_i b_i|)\\
&= Re(\sum_{i=1}^{n} \lambda_i (|a_i|^2 + |b_i|^2 - 2|a_i||b_i|))\\
&= \sum_{i=1}^{n} Re(\lambda_i) |a_i - b_i|^2\\
&\geq \mathrm{Re}(\lambda_{\min}) \sum_{i=1}^{n} |a_i - b_i|^2\\
&= \mathrm{Re}(\lambda_{\min}) \|x_1 - x_2\|^2
\end{aligned}$$

where $\lambda_{\min}$ denotes the eigenvalue of $\mathcal{A}$ with the smallest real part. Note that we use $\mathrm{Re}(\lambda_{\min})$ because the hypomonotonicity depends on the real part of the eigenvalues.

Therefore, since $\mathrm{Re}(\lambda_{\min})$ is negative, the operator related to $\mathcal{A}$ is hypomonotone but not monotone. ■