Target Mirror Descent: A Unifying Framework for Solving Monotone Variational Inequalities

Yu-Wen Chen, Can Kizilkale, Murat Arcak

Department of Electrical Engineering and Computer Sciences, UC Berkeley

April 20, 2026

What is Variational Inequality (VI)?

Definition (Variational Inequality)

Given a set $\mathcal{X} \subseteq \mathbb{R}^n$ and a map $F : \mathcal{X} \to \mathbb{R}^n$, we say that $x^* \in \mathcal{X}$ solves the variational inequality, denoted as VI$(\mathcal{X}, F)$, if

$$(x - x^*)^\top F(x^*) \geq 0, \quad \forall x \in \mathcal{X}.$$

The set of all solutions is denoted as $\mathrm{SOL}(\mathcal{X}, F)$.

Example:

- objective optimization: $\min_x f(x)$ s.t. $g(x) \leq 0$. (by $F = \nabla f$)

What is Variational Inequality (VI)?

Definition (Variational Inequality)

Given a set $\mathcal{X} \subseteq \mathbb{R}^n$ and a map $F : \mathcal{X} \to \mathbb{R}^n$, we say that $x^* \in \mathcal{X}$ solves the variational inequality, denoted as VI$(\mathcal{X}, F)$, if

$$(x - x^*)^\top F(x^*) \geq 0, \quad \forall x \in \mathcal{X}.$$

The set of all solutions is denoted as $\mathrm{SOL}(\mathcal{X}, F)$.

Example: VI is much broader than (objective) optimization problem

- objective optimization: $\min_x f(x)$ s.t. $g(x) \leq 0$. (by $F = \nabla f$)
- minimax problem: $\min_\theta \max_\phi f(\theta, \phi)$.
- primal-dual dynamics
- fixed point problem: find x^* s.t. $F(x^*) = x^*$.
- multi-agent RL: multiple coupling objectives, Nash equilibria

Many well-known algorithms fail for (monotone) VIs

Definition (Monotonicity)

A map $F : \mathcal{X} \to \mathbb{R}^n$ is σ-strongly monotone if

$$(F(x) - F(y))^\top (x - y) \geq \sigma \|x - y\|^2, \quad \forall x, y \in \mathcal{X}.$$

- For a convex f, we have $F = \nabla f$ monotone.
- Naively, view F as gradients in (objective) optimization algorithms?

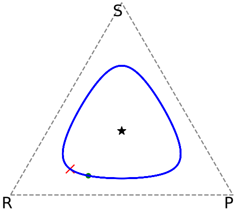
Rock-Paper-Scissor example:

$$F(x) = \begin{bmatrix} 0 & -1 & 1 \\ 1 & 0 & -1 \\ -1 & 1 & 0 \end{bmatrix} x$$

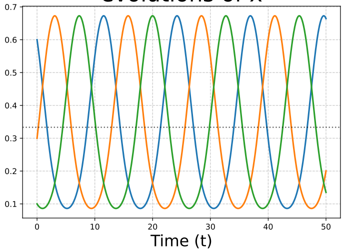
$\mathcal{X} = \Delta^3$ the simplex in $\mathbb{R}^3$

$$x_{t+1} = x_t - \eta F(x_t)$$

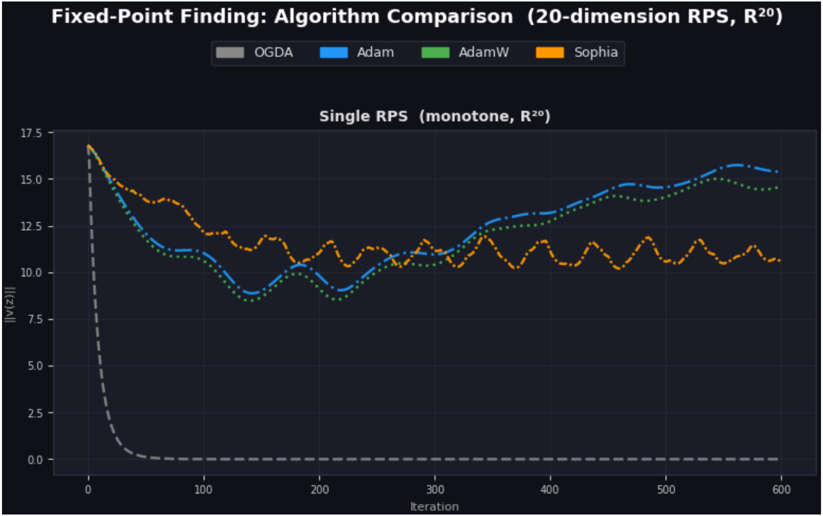
trajectory on simplex



evolutions of x



Adam / AdamW / Sophia also fail for (monotone) VIs



Some successful algorithms for (monotone) VIs

- Proximal Point Method:

$$x_{t+1} = \arg\min_{x} \left\{ f(x) + \frac{1}{2\eta}\|x - x_t\|^2 \right\} \implies x_{t+1} = (I + \eta F)^{-1}(x_t)$$

- Extragradient:

$$\omega_t = \arg\min_{y \in \mathcal{X}} \ \{\langle \eta F(x_t), y\rangle + D_h(y, x_t)\}$$
$$x_{t+1} = \arg\min_{y \in \mathcal{X}} \ \{\langle \eta F(\omega_t), y\rangle + D_h(y, x_t)\}$$

- Splitting method:
  decompose F into $A + B$ with $(I + A)^{-1} \,\&\, (I + B)^{-1}$ easy to get

- Brown-von Neumann-Nash:
  game theory dynamics

- Discounted Mirror Descent:
  Nesterov dual averaging

Overview of our paper: target mirror descent (TMD)

- We propose the target mirror descent (TMD) framework, which
  - stabilizes monotone VIs.
  - corrects the update direction via a target point feedback mechanism.
  - performs a unified convergence analysis.

- We demonstrate that, by appropriate design choices, TMD subsumes
  - proximal point algorithms
  - extragradient methods
  - splitting methods
  - BNN dynamics
  - forward-backward-forward dynamics
  - discounted mirror descent and its higher-order extension

- We study an ensemble version of TMD:
  - parallel TMDs with a shared update direction
  - owing to the decoupling structure of TMD

Overview of our paper: target mirror descent (TMD)

- We propose the target mirror descent (TMD) framework, which
  - stabilizes monotone VIs.
  - corrects the update direction via a target point feedback mechanism.
  - performs a unified convergence analysis.

- We demonstrate that, by appropriate design choices, TMD subsumes
  - proximal point algorithms
  - extragradient methods
  - splitting methods
  - BNN dynamics
  - forward-backward-forward dynamics
  - discounted mirror descent and its higher-order extension

- We study an ensemble version of TMD:
  - parallel TMDs with a shared update direction
  - owing to the decoupling structure of TMD

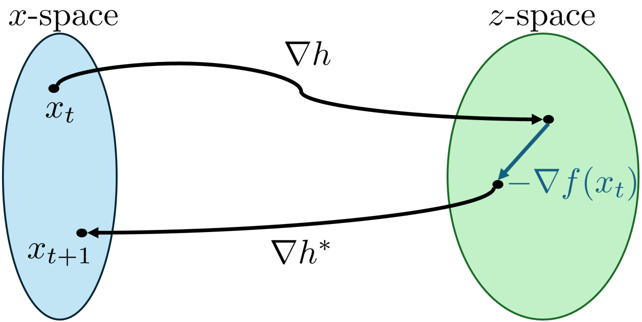
Overview of our paper: target mirror descent (TMD)

- We propose the target mirror descent (TMD) framework, which
  - stabilizes monotone VIs.
  - corrects the update direction via a target point feedback mechanism.
  - performs a unified convergence analysis.

- We demonstrate that, by appropriate design choices, TMD subsumes
  - proximal point algorithms
  - extragradient methods
  - splitting methods
  - BNN dynamics
  - forward-backward-forward dynamics
  - discounted mirror descent and its higher-order extension

- We study an ensemble version of TMD:
  - parallel TMDs with a shared update direction
  - owing to the decoupling structure of TMD

Background for mirror descent



- $z_t = \nabla h(x_t)$
- $z_{t+1} = z_t - \eta \nabla f(x_t)$
- $x_{t+1} = \nabla h^*(z_{t+1})$

- ∇h: mirror map
- h^* is the convex conjugate of h
- $\nabla h^* = (\nabla h)^{-1}$
- Gradients live in the dual space!

Proposed target mirror descent (TMD)

- The proposed TMD:

$$\begin{aligned}\dot{z}(t) &= \alpha\left(S(T(x(t))) - S(x(t))\right) - \beta\Phi(x(t)),\\ x(t) &= \nabla h^*(z(t)).\end{aligned}$$

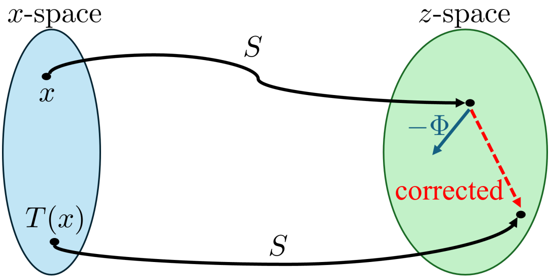
  - Φ: a surrogate function of F
  - ∇h^*: mirror map
  - S: mapping to dual space
  - T: a 'target' mechanism that determines a target point $T(x(t))$
  - $S \circ T - S$: feedback correction

Proposed target mirror descent (TMD)

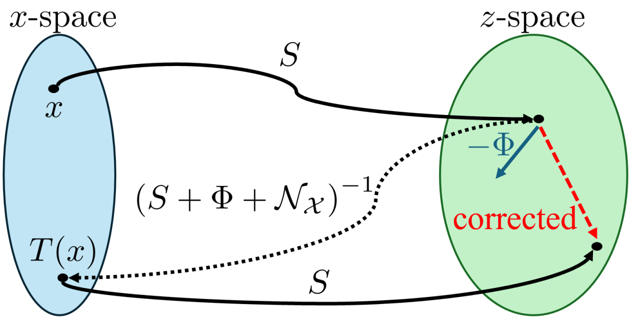
- The proposed TMD:

$$\dot{z}(t) = \alpha\left(S(T(x(t))) - S(x(t))\right) - \beta\Phi(x(t)),$$
$$x(t) = \nabla h^*(z(t)).$$

  - Φ: a surrogate function of F
  - ∇h^*: mirror map
  - S: mapping to dual space
  - T: a 'target' mechanism that determines a target point $T(x(t))$
  - $S \circ T - S$: feedback correction



Proposed TMD — graphical interpretation



Interpretation

- Φ: a surrogate function of F
- S: mapping to dual space
- T: a 'target' mechanism

Conditions on S, T, and Φ:

- $\mathrm{SOL}(\mathcal{X}, \Phi) \subseteq \mathrm{SOL}(\mathcal{X}, F)$
- S is strongly monotone
- $T = (S + \Phi + \mathcal{N}_{\mathcal{X}})^{-1} \circ S$
- $\exists \bar{x} \in \mathcal{X}$ variationally stable (VS) w.r.t. Φ

TMD as a unifying framework

	S	Φ	α	β
Proximal Point Method	∇h	ηF	1	0
Extragradient (EG)	$\nabla h - \eta F$	ηF	1	0
EG+	$\nabla h - \eta_1 F$	$\eta_1 F$	$\frac{\eta_2}{\eta_1}$	0
Douglas-Rachford Splitting[1]	I	$T_{\mathrm{DR}}^{-1} - I$	1	0
Forward-Backward Splitting	I	$T_{\mathrm{FB}}^{-1} - I$	1	0
Brown-von Neumann-Nash	∇h	excess payoff[2]	$\frac{1}{\eta}$	0
Forward-Backward-Forward	$I - \eta F$	ηF	1	0
Discounted Mirror Descent w/	∇h	ηF	γ	0
calibration improvement	$\nabla h - \eta F$	ηF	γ	γ

[1] Splitting method: $x_{k+1} = T_{\mathrm{op}}(x_k)$, where $T_{\mathrm{op}} \in \{T_{\mathrm{DR}}, T_{\mathrm{FB}}\}$. Thus, Φ is the resolvent operator.

[2] Excess payoff determines the target mechanism T.

Convergence analysis of TMD

Theorem

TMD asymptotically approaches $\mathrm{SOL}(\mathcal{X}, \Phi)$.

Proof.

- Lyapunov function: Bregman divergence.
- Barbalat's lemma shows: $x(t) - T(x(t)) \to 0$.
- Any fixed point of T is in $\mathrm{SOL}(\mathcal{X}, \Phi)$.

$\blacksquare$

If with a slightly stronger condition: $\mathrm{SOL}(\mathcal{X}, \Phi)$ is VS w.r.t. Φ, then

Corollary

With the above condition, TMD converges to $x^* \in \mathrm{SOL}(\mathcal{X}, \Phi)$.

Ensemble TMD

- Multiple algorithms solve the same problem (in parallel) using distinct mirror maps, while a shared ensemble state drives their dual updates.
- Formally, the ensemble TMD is

$$
\begin{aligned}
\dot{z}^i(t) &= \alpha(S \circ T - S)(x^{\mathrm{en}}(t)) - \beta\Phi(x^{\mathrm{en}}(t)), \quad z^i(0) \in \mathbb{R}^n, \\
x^i(t) &= \nabla(h^i)^*(z^i(t)), \\
x^{\mathrm{en}}(t) &= \frac{1}{N}\sum_{k=1}^{N} x^k(t).
\end{aligned}
$$

- It is the decouple nature of (S, T, Φ) and mirror maps that enables the ensemble TMD.

Ensemble TMD performs geometric ensemble

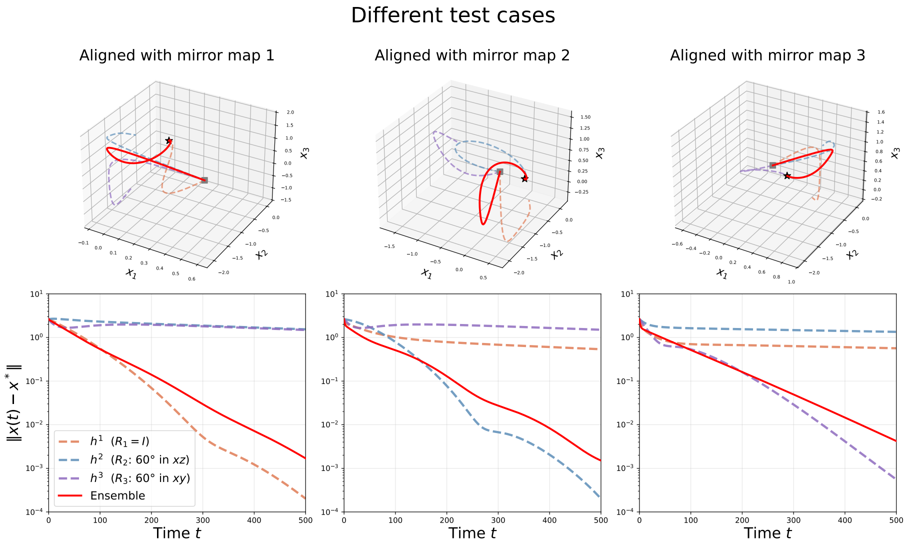
Equivalence of ensemble TMD

The ensemble state x^{en} satisfies

$$\begin{aligned} \dot{z}(t) &= \alpha(S \circ T - S)(x^{\mathrm{en}}(t)) - \beta\Phi(x^{\mathrm{en}}(t)) \\ x^{\mathrm{en}}(t) &= \nabla(h^{\mathrm{en}})^*(z(t)), \quad \text{with} \quad z(0) = 0, \end{aligned}$$

for some synthesized mirror map h^{en}.

Simulation for ensemble TMD



Conclusion