Stochastic Optimization in Game Theory

Shivam Patel
200070077

A unified stochastic approximation framework for learning in games
Panayotis Mertikopoulos, Ya-Ping Hsieh, Volkan Cevher ([1])

1 Introduction

Game Theoretic representations in the continuous action spaces are vital for solving real life optimization problems. Three conditions completely describe the framework required for solving any game:

- The exogenous conditions and participants in any game
- The unique set of action profiles available to each player
- The payoff vectors generated by action profiles

Given any simultaneous or extended time horizon game, it is imperative to ask whether any particular policy is a 'Nash Equilibrium' for all players. If the game reaches such an equilibrium, there is no incentive for unilateral deviation available to any player.

This question has been at the forefront of research, and concomitantly the backbone of all multi-agent selfish optimization schemes. There have been positive results which ascertain the optimality of solutions in 'nice' formulations, and at the same time it is impossible to guarantee convergence to Nash equilibrium in all classes of games.

Conventional research in the field was focused on finite action space games which are efficient in representing economic models of rationality, modelling climatic conditions and extremities, characterizing international peace treaties etc. But with the advent of computing efficiency and applications of game theory to contemporary fields like machine learning, research is being directed towards continuous space games where players can choose from a countable infinite set of actions.

Contributions Presented - The paper bridges the gap between conventional game theory and games in continuous spaces by proposing the *Mirrored Robbins-Monro (MRM)* Algorithm, which utilizes discrete updates for converging the action profiles to obtain a Nash Equilibrium. The technical results presented can be projected in three different directions -

- The notion of *subcoercivity* is introduced, which gives mathematical constraints for convergence of MRM updates to Internally Chain Transitive (ICT) sets of the domain.
- *Attractors* of any iterative algorithm are classified using appropriate Lyapunov functions, which ensure almost sure convergence in monotonous payoff gradient domains.

- *Coherent* sets are defined by observing the behaviour of non-optimal iterate perturbations and their effect on convergence of MRM.

2 Preliminaries

2.1 Notation

In the following discourse, $\mathcal{V}$ represents a d-dimensional space on real numbers, with the norm $||\cdot||$. The dual space of $\mathcal{V}$ is represented as $\mathcal{Y} = \mathcal{V}^*$, and $\langle y, x\rangle$ represents the cannonical pairing between $x \in \mathcal{V}$ and $y \in \mathcal{Y}$. Furthermore, $||y||_* := max\{\langle y, x\rangle : ||x|| \leq 1\}$ is the induced dual norm on $\mathcal{Y}$.

2.2 Normal Form Representation of Games

During the discourse, we focus on finite player games, with player i, $s.t.$ $i \in \mathcal{N} = \{1, 2, \cdots, N\}$. Each player chooses an *action* x_i from a subset $\mathcal{X}_i$ of a d-dimensional normed space $\mathcal{V}_i$. We utilize $(x_i; x_{-i})$ to represent the action chosen by player i given other players' actions. We denote the dimension of ambient space $\mathcal{V} = \Pi_i \mathcal{V}_i$ by $d = \Sigma_i d_i$. Each player's rewards are subject to his action, which is represented by the payoff function $u_i : \mathcal{X} \to \mathbb{R}$. We assume the payoff functions to be Lipschitz continuous and smooth. The gradients for these payoff functions are aptly represented through $v_i(x)$

$$v_i(x) = \nabla_{x_i} u_i(x_i; x_{-i}) \quad \text{and} \quad v(x) = (v_i(x))_{i \in \mathcal{N}}$$

2.3 Solutions to Nash Equilibrium Conditions

The defining characteristic of Nash Equilibrium is that players cannot make any strictly profitable unilateral deviations from the proposed equilibrium. Mathematically, any action $x^* \in \mathcal{X}$ is a Nash Equilibrium of the game $\mathcal{G} \equiv (\mathcal{N}, \mathcal{X}, u)$ if

$$u_i(x^*) \geq u_i(x_i; x^*_{-i}) \qquad \forall x_i \in \mathcal{X}_i, i \in \mathcal{N}$$

Nash Equilibria always exist when the individual payoff functions are strictly concave in x_i. But in other conditions, we assume the following looser conditions:

1. *Local Nash Equilibria* (LNE) are profiles $x^* \in \mathcal{X}$ for which Nash Eqm. holds in a local neighbourhood $\mathcal{U}$ of $x^* \in \mathcal{X}$.
2. *Critical Points* are action profiles $x^* \in \mathcal{X}$ which satisfy the first order derivative conditions for stationarity
$$\frac{d}{dt}|_{t=0^+} u_i(x^*_i + t(x_i - x^*_i); x^*_{-i}) \leq 0 \qquad \forall x_i \in \mathcal{X}_i, i \in \mathcal{N}$$

Now we define various conditions on the payoff profiles and gradients for players. These inequalites are represented in the given table for the sake of brevity.

3 Online Learning Algorithms

Depending on the type of information available to players, we classify continuous action space games into *Oracle Based Methods* and *Payoff Based Methods*.

Inequality	Shorthand	Condition	Domain of Validity
Stampacchia Variational Inequality	SVI	$\langle v(x^*), x - x^* \rangle \leq 0$	$\forall x \in \mathcal{X}$
Minty Variational Inequality	MVI	$\langle v(x), x - x^* \rangle \leq 0$	$\forall x \in \mathcal{X}$
Variational Stability	VS	$\langle v(x), x - x^* \rangle < 0$	$\forall x \in \mathcal{U} \backslash \{x^*\}$ of $x \in \mathcal{X}$
Neutral Stability	NS	$\langle v(x), x - x^* \rangle \leq 0$	$\forall x \in \mathcal{U}$ of $x \in \mathcal{X}$
Global Variational Stability	GVS	$\langle v(x), x - x^* \rangle < 0$	$\forall x \in \mathcal{X} \backslash \{x^*\}$
Global Neutral Stability	GNS($\equiv$ MVI)	$\langle v(x), x - x^* \rangle \leq 0$	$\forall x \in \mathcal{X}$

Table 1: Some variational inequalities on payoff functions and their gradients on action space $\mathcal{X}$

3.1 Oracle Based Methods

Oracle based black box models return only noisy estimates of payoff gradients to players. The Stochastic First Order Condition (SFO) can be described as

$$V(x, \theta) = v(x) + Err(x; \theta)$$

Some standard algorithms for Oracle Based Methods are described below.

1. *Stochastic Gradient Ascent* (SGA)- The simplest iterative update method, where all players simultaneously update their action policies.

$$X_{i,n+1} = X_{i,n} + \gamma_n V_i(X_n; \theta_n)$$

2. *Sequential Gradient Ascent* (SeqGA) - Players update their actions in a round-robin manner, instead of making simultaneous updates.

$$X_{i,n+1} = X_{i,n} + \gamma_n V_i(\cdots, X_{i-1,n+1}, X_{i,n}, X_{i+1,n}, \cdots; \theta_n)$$

This algorithm is extensively used in training GANs [2].

3. *Extra Gradient* (EG) - The iterative updates are made by using the explore-exploit paradigm, where the player takes a gradient step to an intermediate location in the landscape, and then updates his state by using the gradient obtained from that intermediate state.

$$X_{i,n+1/2} = X_{i,n} + \gamma_n V_i(X_n; \theta_n),$$
$$X_{i,n+1} = X_{i,n} + \gamma_n V_i(X_{n+1/2}; \theta_{n+1/2})$$

To explore its applications in robust RL and GAN training, refer [3] and [4].

4. *Optimistic Gradient* (OG) - The expensive EG requires two oracle queries per iteration. To overcome this, we reuse intermediate step gradient as obtained from last step.

$$X_{i,n+1/2} = X_{i,n} + \gamma_n V_i(X_{n-1/2}; \theta_{n-1}),$$
$$X_{i,n+1} = X_{i,n} + \gamma_n V_i(X_{n+1/2}; \theta_n)$$

5. *Exponential Weights* (EW) - In games where players select randomized strategies $\alpha_{i,n} \in \mathcal{A}_i$ and observe payoff vectors either in terms of *pure payoffs* or *mixed payoffs*. In such contexts, most widely used algorithm is the EW updates

$$Y_{i,n+1} = Y_{i,n} + \gamma_n V_i(X_n; \alpha_n)$$
$$X_{i,n+1} = \Lambda_i(Y_{i,n+1})$$

Where Λ_i is an appropriate choice of the *Logit Function*

$$\Lambda_i = \frac{(exp(y_{i\alpha_i}))_{\alpha_i \in \mathcal{A}_i}}{\Sigma_{\alpha_i \in \mathcal{A}_i} exp(y_{i\alpha_i})}$$

There arise two different settings of information feedback, depending on the oracle model:

(a) Full Information Feedback: The mixed payof vectors are observed by the player

$$V_i(X_n; \alpha_n) = v_i(X_{i,n}; X_{-i,n})$$

(b) Realization-based Feedback: Only the pure payoff vectors of the players are observed

$$V_i(X_n; \alpha_n) = v_i(\alpha_{i,n}; \alpha_{-i,n})$$

EW Algorithm is long used in online learning and game theory, see Vovk [5].

6. *Mirror-Prox Algorithm* (MP) - Accumulating the gradient updates in a different domain $\mathcal{Y}$ and then maping back to the feasible set of action profiles $\mathcal{X}$ given us the MP algorithm. We use a *mirror-map* $Q : \mathcal{Y} \to \mathcal{X}$ for bijective mapping.

$$Y_{n+1/2} = Y_n + \gamma_n V(X_n; \theta_n) \qquad Y_{n+1} = Y_n + \gamma_n V(X_{n+1/2}; \theta_{n+1/2})$$
$$X_{n+1/2} = Q(Y_{n+1/2}) \qquad X_{n+1} = Q(Y_{n+1})$$

3.2 Payoff Based Methods

In games like auctions, networks, online advertisement recommendations etc. it is difficult to estimate payoff gradients. We only observe the payoffs through queries, and need to estimate the gradients through such black-box payoff queries. Some algorithms for payoff based approximations are -

7. *Single Point Stochastic Approximation* (SPSA) - A simple method for obtaining zeroth-order feedback reconstruction of gradients is using the SPSA algorithm. In any query state $\hat{X}_{i,n}$ we add a perturbation to the base state $X_{i,n}$ along a randomly chosen direction $W_{i,n}$, and scale it accordingly with a weighing parameter δ_n.

$$\hat{X}_{i,n} = X_{i,n} + \delta_n W_{i,n}$$
$$X_{i,n+1} = X_{i,n} + \gamma_n (u_i(\hat{X}_n)/\delta_n) W_{i,n}$$

For a constrained setting implementation of SPSA, refer Bravo et al[6]. Note that they show the validity of our analysis when the constrained set $\mathcal{X}$ is compact.

8. *Dampened Gradient Approximation* (DGA) - An 'Explore-Exploit' version of the SPSA is provided by the two stage DGA algorithm.

$$X_{i,n+1/2} = X_{i,n} + (1/n) W_{i,n}$$
$$X_{i,n+1} = X_{i,n}[1 + (u_i(X_{n+1/2}) - u_i(X_n)) W_{i,n}]$$

Please refer Bervoets et al [7] for the original paper on DGA updates for games with $\mathcal{X}_i = [0, \infty)$.

9. *EXP3 Algorithm* - This algorithm resembles the bandit case where each player observes only the payoffs of his own actions that were played (note no distinction between pure and randomized strategies). We employ the importance weighted estimator (IWE)

$$V_{i\alpha_i}(\hat{X}_n; \hat{\alpha}_n) = \frac{1(\hat{\alpha}_{i,n}=\alpha_i)}{\hat{X}_{i\alpha_i,n}} u_i(\hat{\alpha}_{i,n}; \hat{\alpha}_{-i,n}) \qquad \forall \alpha_i \in \mathcal{A}_i, i \in \mathcal{N}$$

The above obtained IWE is plugged into EW to obtain EXP3 algorithm.

As a closing remark, we note that multiple algorithms of different mathematical flavours can be generated by mixing the above techniques.

4 Stochastic Approximation Framework

4.1 Framework of Template

The generic structure that we use to generalize above mentioned algorithms is the *Mirrored Robbins-Monro* Algorithm.

$$Y_{n+1} = Y_n + \gamma_n \hat{v}_n \qquad X_{n+1} = Q(Y_{n+1})$$

Where:

1. $X_n = (X_{i,n})_{i \in \mathcal{N}} \in \mathcal{X}$ represents the state of the algorithm at each stage $n = 1, 2, \cdots$
2. $Y_n = (Y_{i,n})_{i \in \mathcal{N}} \in \mathcal{Y}$ represents the accumulating state of the gradients of iterates.
3. $\hat{v}_n = (\hat{v}_{i,n})_{i \in \mathcal{N}} \in \mathcal{Y}$ is the payoff gradient (noisy) related to individual's payoffs.
4. γ_n represents the step size, and $\Sigma_n \gamma_n = \infty$ (typically $\gamma_n \propto 1/n^p$ for some $p > 0$).
5. $Q : \mathcal{Y} \to \mathcal{X}$ is the mirror map.

4.1.1 Analysis of Gradient Signal from Oracle

Let us decompose the gradient signal $\hat{v}_n$ in MRM as

$$\hat{v}_n = v(X_n) + U_n + b_n$$

Where

$$U_n = \hat{v}_n - [\hat{v}_n | \mathcal{F}_n] \qquad \text{and} \qquad b_n = [\hat{v}_n | \mathcal{F}_n] - v(X_n)$$

We assume that $U_n, b_n, \hat{v}_n$ are upper bounded by B_n, σ_n, M_n in a particular normed space $||\cdot||^q$. The three terms respectively represent the upper bound on bias, fluctuation and magnitude of the gradient estimate $\hat{v}_n$. Let us assume the asymptotic behaviour

$$\gamma_n = \gamma / n^p \qquad B_n = \mathcal{O}(1/n^b) \qquad M_n = \mathcal{O}(n^s)$$

4.1.2 Player's Mirror Map Q

The mirror map $Q = (Q_i)_{i \in \mathcal{N}} : \mathcal{Y} \to \mathcal{X}$ is the defining element for transforming the auxiliary domain to the state space of action profiles. Q is defined by the means of a regularizer on $\mathcal{X}$.

Any function $h_i : \mathcal{V} \to \mathbb{R} \cup \infty$ is said to be the regularizer on $\mathcal{X}_i$ if:

1. dom $h_i = \{x_i \in \mathcal{V}_i : h_i(x_i) < \infty\} = \mathcal{X}_i$, i.e. h_i is supported on $\mathcal{X}_i$
2. h_i is strongly convex and continuous on $\mathcal{X}_i$

$$h_i(\lambda x_i + (1-\lambda)x_i') \leq \lambda h_i(x_i) + (1-\lambda)h_i(x_i') - \tfrac{1}{2}K_i\lambda(1-\lambda)||x_i' - x_i||^2$$

The mirror map Q associated with h_i is defined for $y_i \in \mathcal{Y}_i$ as

$$Q_i(y_i) = arg\ max_{x_i \in \mathcal{X}_i}\{\langle y_i, x_i\rangle - h_i(x_i)\}$$

and the image $\mathcal{X}_{h_i} = im\ Q$ of Q_i is called the *prox-domain* of h_i.

4.1.3 Representing Online Learning Algorithms through MRM

Here, we show how existing online learning algorithms are special cases of the MRM algorithm, using appropriate $Q(\cdot)$ functions and iteration rules.

1. Stochastic Gradient Ascent: Take $Q(y) = y$ and run MRM with gradient signals as $\hat{v}_n = V(X_n; \theta_n)$.
2. Sequential Gradient Ascent: Same as in SGA, but with sequential updates for each player instead of simultaneous updates for all players.
3. Extra-Gradient: Same as in SGA, but the gradient signal is now $\hat{v}_n = V(X_{n+1/2}; \theta_{n+1/2})$.
4. Optimistic Gradient: Same as SGA, but with gradient signal $\hat{v}_n = V(X_{n+1/2}; \theta_n)$.
5. Exponential Weights: The oracle gradient signals depend on the information feedback structure used. The mirror map $Q(\cdot)$ is the logit function

$$\Lambda_i = \frac{(exp(y_{i\alpha_i}))_{\alpha_i \in \mathcal{A}_i}}{\Sigma_{\alpha_i \in \mathcal{A}_i} exp(y_{i\alpha_i})}$$

6. Mirror-Prox algorithm: The gradient signal is same as in EG, i.e. $\hat{v}_n = V(X_{n+1/2}; \theta_{n+1/2})$. The mirror map is general.
7. Single Point Stochastic Approximation: The gradient signal is $\hat{v}_{i,n} = (u_i(\hat{X}_n)/\delta_n)W_{i,n}$. Other details are same as in SGA.
8. Dampened Gradient Approximation: We take the mirror map $Q_i(y_i) = exp(y_i)$. Then, letting $Y_n = logX_n$, we get

$$Y_{n+1} = Y_n + log(1 + (u_i(X_{n+1/2}) - u_i(X_n))W_{i,n})$$

DGA can be viewed as a special case of MRM with $\gamma_n = 1/n$ and gradient signals
$\hat{v}_{i,n} = n.log(1 + (u_i(X_{n+1/2}) - u_i(X_n))W_{i,n})$

9. EXP3 Algorithm: Same as in EW algorithm, but with gradients given by the Importance Weighted Estimator, and pure strategies $\hat{\alpha}_n$ chosen according to $\hat{X}_n$.

Algorithm	Actions ($\mathcal{X}_i$)	Mirror Map (Q)	Feedback	Bias (B_n)	Magnitude (M_n)
SGA	$\mathbb{R}^{d_i}$	y	oracle	0	$\mathcal{O}(1)$
SeqGA	$\mathbb{R}^{d_i}$	y	oracle	$\mathcal{O}(1/n^p)$	$\mathcal{O}(1)$
EG \ OG	$\mathbb{R}^{d_i}$	y	oracle	$\mathcal{O}(1/n^p)$	$\mathcal{O}(1)$
EW	$\Delta(\mathcal{A}_i)$	$\Lambda(y)$	oracle	0	$\mathcal{O}(1)$
MP	general	general	oracle	$\mathcal{O}(1/n^p)$	$\mathcal{O}(1)$
SPSA	$\mathbb{R}^{d_i}$	y	payoff	$\mathcal{O}(1/n^r)$	$\mathcal{O}(n^r)$
DGA	$[0,\infty)]$	$exp(y)$	payoff	$\mathcal{O}(1/n)$	$\mathcal{O}(1)$
EXP3	$\Delta(\mathcal{A}_i)$	$\Lambda(y)$	payoff	$\mathcal{O}(1/n^r)$	$\mathcal{O}(n^r)$

Table 2: Algorithms that can be modelled using MRM with appropriate Q *mirror-maps*

4.2 Stochastic Approximation and Mean Dynamics

MRM can be assumed to be a noisy discretization of Mean Dynamics (MD) of the continuous state space game. The Mean dynamics are

$$\dot{y} = v(x) \qquad x = Q(y)$$

$\dot{y}$ represents the continuous time limit of the finite difference quotient $(Y_{n+1} - Y_n)/\gamma_n$. The linear interpolation of discretized updates to continuous time scale gives an *asymptotic trajectory* of the *mean dynamics*.

$$Y(t) = Y_n + \frac{t-\tau_n}{\tau_{n+1}-\tau_n}(Y_{n+1} - Y_n)$$

We now have the notion of asymptotic closeness, which is delineated by the concept of *Asymptotic Pseudo-Trajectories* (APT) of MD if $\Psi(Y(t))$ is the mean dynamics flow)

$$lim_{t\to\infty} sup_{0\le s\le T} ||Y(t+s) - \Psi(Y(t))|| \text{ for all } T > 0$$

Proposition If we run the MRM algorithm with step sizes γ_n such that

1. $lim_{n\to\infty}\gamma_n = 0$
2. $lim_{n\to\infty}B_n = 0$ and $\Sigma_n \gamma_n^{1+q/2} M_n^q < \infty$

Then the sequence $X_n = Q(Y_n)$ is an Asymptotic Pseudo Trajectory of Mean Dynamics w.p. 1.

5 General Convergence Analysis

5.1 Primal-Dual Space Considerations

Any stochastic update algorithm and subsequent convergence analysis holds only in the space where gradients are calculated and accumulated. Hence, for the MRM algorithm, convergence in the original space $\mathcal{X}_i$ is not automatically guaranteed

when iterates converge in the dual space $\mathcal{Y}_i$. Such asymptotic convergence depends on the boundary behaviour of the regularizer $h(x)$. For analyzing the convergence of such systems, we characterize two mirror maps-

1. *Surjective mirror maps* : In such mirror maps, $\mathcal{X}_h = im\ Q = \mathcal{X}$. The mean dynamics' primal orbit $x(t) = Q(y(t))$ does not capture the state of the systems, and we cannot capture the moments when $x(t)$ enters or exits the boundary of $\mathcal{X}$.

2. *Interior-valued mirror maps* : Here $\mathcal{X}_h = im\ Q = ri\ \mathcal{X}$, where ri $\mathcal{X}$ denotes the relative interior of set $\mathcal{X}$.

5.2 Convergence to Internally Chain Transitive Sets

For any flow map of mean dynamics $\Psi : \mathbb{R} \times \mathcal{Y} \to \mathcal{Y}$, and a nonempty convex subset $\mathcal{D}$ of $\mathcal{Y}$,

1. $\mathcal{D}$ is invariant under MD if $\Psi_t(\mathcal{D}) = \mathcal{D}$ for all $t \in \mathbb{R}$.

2. $\mathcal{D}$ is an attractor for MD if it admits a neighbourhood $\mathcal{D} \subseteq \mathcal{Y}$ such that $dist(\Psi_t(y), \mathcal{D}) \to 0$ uniformly in $y \in \mathcal{D}$ as $t \to \infty$.

3. $\mathcal{D}$ is internally chain transitive (ICT) if it is invariant and $\Psi|_{\mathcal{D}}$ has no attractors except $\mathcal{D}$.

Convergence of Asymptotic Pseudo Trajectories to an ICT is guaranteed when $sup_n||Y_n||_* < \infty$. Thus, for games having solutions confined in $\mathcal{X}$, boundedness of Y_n is essential for convergence. To account for this, we define the notion of *subcoercivity* for any game $\mathcal{G}$.

Subcoercivity: We say that $\mathcal{G}$ is subcoercive if there exists a compact set $\mathcal{K} \in$ ri $\mathcal{X}$ and a reference point $p \in \mathcal{K}$ such that

$$\langle v(x), x - p \rangle \leq 0 \text{ for all } x \in \mathcal{X} \backslash \mathcal{K}$$

The notion of subcoercivity suggests that all strongly attracting points are contained in $\mathcal{K}$. Apart from the 'attractive' domain $\mathcal{K}$, other points in $\mathcal{X}$ do not 'drift-away' from $\mathcal{K}$. Consequently, given that an MRM is run with step-size and gradient signal sequences satisfying the boundedness properties, and game $\mathcal{G}$ is subcoercive then the sequence of iterates are bounded w.p. 1.
The above definition of subcoercivity naturally gives rise to the following corollaries:

- If $\mathcal{G}$ admits a subcoercive potential, X_n converges to a component of critical points of $\mathcal{G}$ w.p.1. Additional, X_n converges to a set of Nash Equilibria of $\mathcal{G}$ if potential is concave.

- Suppose that $\mathcal{G}$ is a strictly convex-concave min-max game with an interior equilibrium $x^* \in ri\ \mathcal{X}$. Then X_n converges to x^* w.p.1.

The idea of subcoercivity by itself given us reasoning regarding why there cannot be a consistent drifting point away from $\mathcal{K}$, which gives a reasonable deduction that

Y_n is also not repelled to infinity either. To control the stochasticity of Y_n and put our reasoning precisely, we define the following conditions on finite summability of bias, variance and gradient signal magnitudes -

$$\Sigma_n \gamma_n B_n < \infty \qquad \Sigma_n \gamma_n^2 \sigma_n^2 < \infty \qquad \Sigma_n \gamma_n^2 M_n^2 < \infty$$

6 Convergence to Primal Attractors

The general convergence analysis provided in the last section does not suffice to explain what types of sets could be attracting under given MRM setup, and also how to interpret the stability of boundary equilibrium that attracts all mean dynamics in the primal space even if the dual space iterates blow up to infinity. Usually attractors are associated with local Lyapunov Functions Φ that are smooth, zero at the attractor point, positive everywhere else and strictly decreasing along every nearby trajectory that does not belong to the attractor ($\dot{\Phi} < 0$). But for our analysis, we are interested in judging the attracting properties of subsets of primal space $\mathcal{X} \subseteq \mathcal{V}$ whereas the dynamics evolve in the dual space $\mathcal{Y} = \mathcal{V}^*$.

We define the notion of a primal attractor below, which addresses such questions.

6.1 Primal Attractors

Energy Functions: A Lipschitz continuous and smooth function $E : \mathcal{Y} \to R$ is a local energy function for MD if $\sup\{\dot{E}(y) : E_- < E(y) < E_+\} < 0$ for all sufficiently small $E_+ > E_- > inf\ E$.

Primal Attractors: A nonempty compact subset S of $\mathcal{X}$ is said to be a *primal attractor* of ME if it admits an energy function E with $\inf E > -\infty$ and such that $Q(y) \to S$ whenever $E(y) \to inf\ E$.

7 Sharper Convergence: The Role of Coherence

7.1 Coherence

A nonempty compact subset S of X will be called *coherent* if it admits a (finite) set of deviation directions $Z = \{z_1, ..., z_m\} \subseteq V$ such that

1. $\langle v(x), z\rangle < 0$ for all $x \in \mathcal{S}$ and all $z \in Z$
2. $Q(y) \to \mathcal{S}$ whenever $max_{z\in\mathcal{Z}}\langle y, z\rangle \to -\infty$

The first condition of coherence suggests that there is no strictly profitable deviation available to any player. The gradient $v(x)$ along any direction $z \in \mathcal{Z}$ is directed towards the equilibrium, thus disincentivizing deviations. The second condition posits that the elements of $\mathcal{Z}$ are sufficient to identify $\mathcal{S}$ by acting as a 'primal-dual' support for $\mathcal{S}$ under $\mathcal{Q}$.

7.2 Convergence Analysis

We observe that coherent sets are primal attractors. Let us understand the concept of coherent sets using the following example for energy function.

$$E(y) = log(1 + \Sigma_{z \in \mathcal{Z}} exp\langle y, x \rangle)$$

We check that two conditions required for coherent sets. Firstly, if $E(y) \rightarrow inf\ E = 0$, we need to have $\langle y, x \rangle \rightarrow -\infty$ for all $z \in \mathcal{Z}$. Thus, the limit of $Q(y) \rightarrow \mathcal{S}$ is guaranteed by definition. Also, for all such y, we have that $\nabla E(y) = \Sigma_{z \in \mathcal{Z}} \langle v(x), z \rangle e^{\langle y,z \rangle} / (1 + \Sigma_{z \in \mathcal{Z}} e^{\langle y,z \rangle}) < 0$ by the continuity of v.

7.3 Requirements for Technical Proofs on Convergence

We state some corollaries, lemmas and proposition which are required for the technical proofs of convergence bounds. We omit the proofs for the sake of brevity.

Corollary: Suppose that $\mathcal{S}$ is coherent, and let X_n ve the sequence of MRM iterates with appropriate considerations for boundedness of noise and gradients. Then *(i)* if $\mathcal{S}$ is globally coherent, then X_n converges to $\mathcal{S}$ w.p. 1; and *(ii)* if $\mathcal{S}$ is locally coherent, X_n converges locally to $\mathcal{X}$ with probability atleast $1 - \rho$ if γ is small enough relative to ρ (ρ is the ϵ-confidence level for unboundedness of noise and gradients).

Lemma If we have the set $\mathcal{S} \subseteq \mathcal{X}$ as coherent, and the energy function $E_z(y) = \langle y, z \rangle$ for $y \in Y, z \in \mathcal{Z}$. Then the iterates $E = E_z(Y_n)$ of the energy function E_z satisfy the inequality

$$E_{n+1} \leq E_n + \gamma_n \langle v(X_n), z \rangle + \gamma_n \xi_n + \gamma_n \chi_n$$

and the error terms ξ_n and χ_n are given by

$$\xi_n = \langle U_n, z \rangle \qquad \text{and} \qquad \chi_n = max_{z \in Z} ||z||.B_n$$

Proof We set $y \leftarrow Y_{n+1}$ in $E_z(y)$ and invoke the definition of MRM.

Proposition Suppose that $\mathcal{S}$ is globally coherent, and let $X_n = Q(Y_n)$ be the sequence of play generated by MRM. If the aggregate error processes are sublinear in τ_n, then X_n converges to $\mathcal{S}$ w.p. 1.

Proposition Suppose that $\mathcal{S}$ is locally coherent, then for any given confidence level ρ and iterate sequence X_n generated by MRM, there exists an open set of initialization $\mathcal{D} \subseteq \mathcal{Y}$ of initializations such that (assuming the previous statements on boundedness hold)

$$\mathbb{P}(X_n \text{ converges to } \mathcal{S} | Y_i \in \mathcal{D}) \geq 1 - (m+1)\rho$$

Theorem Let $X_n = Q(Y_n)$ be the sequence of play generated by MRM with step size and gradient signal as per section (4.1.1). Then:

1. If $\mathcal{S}$ is globally coherent, X_n converges to $\mathcal{S}$ w.p. 1.

2. If $\mathcal{S}$ is locally coherent and *(i)* $p - s \geq 1/2$ or; *(ii)* $0 \leq p < q/(2+q)$ and $s < 1/2 - 1/q$, there exists an open set $\mathcal{D} \subseteq \mathcal{Y}$ of initializations such that for any $\rho > 0$

$$\mathbb{P}(X_n \text{ converges to } \mathcal{S} | Y_i \in \mathcal{D}) \geq 1 - \rho$$

Refer Giannou et al [8] for convergence rate analysis.

Theorem Suppose that the mirror map $Q : \mathcal{Y} \to \mathcal{X}$ of MRM is surjective. If $\mathcal{S}$ is coherent, then, w.p. 1, every trajectory $X_n = Q(Y_n)$ that converges to $\mathcal{S}$ does so in a finite number of iterations. In other words, there exists some n_0 such that $X_n \in \mathcal{S}$ for all $n \geq n_0$.

8 Appendix. Regularizers and Mirror Maps

Extensions to the idea of mirror map $Q(\cdot)$ and regularizers $h(\cdot)$ are presented in this appendix. Firstly, recollect that the subderivative of of h at any $x \in \mathcal{X}$ is defined as $\partial h := \{y \in \mathcal{Y} : h(x') \geq h(x) + \langle y, x - x' \rangle \text{ for all } x' \in \mathcal{V}\}$. Also, the domain of subdifferentiability of h is defined as $h^*(y) = max_{x \in \mathcal{X}}\{\langle y, x \rangle - h(x)\}$ for all $y \in \mathcal{Y}$. We can now discuss the following results on regularizers and mirror map.

Lemma Let h be a regularizer on $\mathcal{X}$, and let $Q : \mathcal{Y} \to \mathcal{X}$ bet its induced mirror map. Then:

1. Q is single valued on $\mathcal{Y}$: in particular, for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$, we have $x = Q(y) \iff y \in \partial h(x)$.
2. The *prox-domain* $\mathcal{X}_h := im\ Q$ of h satisfies $ri\ \mathcal{X} \subseteq \mathcal{X}_h \subseteq \mathcal{X}$
3. Q is (1/K) Lipschitz continuous and $Q = \nabla h^*$.

Lemma Let h be a regularizer on $\mathcal{X}$ with induced mirror map $Q : \mathcal{Y} \to \mathcal{X}$, and let $F(p, y) = h(p) + h^*(y) - \langle y, p \rangle$ for $p \in \mathcal{X}$, $y \in \mathcal{Y}$. Then for all $y' \in \mathcal{Y}$, we have

1. $F(p, y) \geq (1/2)K||Q(y) - p||^2$
2. $F(p, y') \leq F(p, y) + \langle y' - y, Q(y) - p \rangle + \frac{1}{2}||y' - y||_*^2$

In particular if $h(0) = 0$, we have

$$(K/2)||Q(y)||^2 \leq h^*(y) \leq -min\ h + \langle y, Q(y) \rangle + (2/K)||y||_*^2 \text{ for all } y \in \mathcal{Y}$$

Lemma Let h be a regularizer on the simplex $\Delta\mathcal{A} \subseteq \mathbb{R}^{\mathcal{A}}$. If $y_\alpha - y_\beta \to -\infty$ then $Q_\alpha(y) \to 0$.

Lemma Let h be a regularizer on $\mathcal{X}$, let y_n, $n = 1, 2, \cdots$ be a sequence in $\mathcal{Y}$, and fix some $x \in \mathcal{X}$. If $\langle y_n, z \rangle \to -\infty$ for every nonzero $z \in TC(x)$ (TC(x) is the tangent cone of x), we have $Q(y_n) \to x$.
Proof Assume that $limsup_n||x_n - x|| > 0$. We are also given that $y_n \in \partial h(x_n)$, so if we replace $z_n = (x_n - x)/||x_n - x||$, we get that $h(x) \geq h(x_n) + \langle y_n, x - x_n \rangle \geq$

$h(x_n) - \langle y_n, z_n \rangle ||x_n - x||$. If we assume that z_n converges to a unit sphere in the normed space, hen there exists some $z \in TC(x)$ with $||z|| = 1$ and such that $\langle y_n, x_n \rangle \leq (1+\epsilon)\langle y_n, z \rangle$ for some $\epsilon > 0$. Thus taking the limsup of the above estimate gives $h(x) \geq \infty$, which is a contradiction. Thus our claim is proved.

Lemma Let h be a regularizer on a convex polytope $\mathcal{P}$ of $\mathcal{V}$, let $\mathcal{S}$ be a face of $\mathcal{P}$, and let $\mathcal{Z} = \{z_1, \cdots, z_n\}$ be a set of unit vectors of $\mathcal{V}$ such that every point $x \in \mathcal{P} \backslash \mathcal{S}$ cam be written as $x = p + \lambda z$ for some $p \in \mathcal{S}$, $z \in \mathcal{Z}$ and $\lambda > 0$. If $max_{z \in \mathcal{Z}} \langle y, z \rangle \to -\infty$, then $Q(y) \to \mathcal{S}$.
Proof We can assume that $x_n = Q(y_n)$ converges to some $x \in \mathcal{P}$ due to the compactness property of $\mathcal{P}$. If $x \notin \mathcal{S}$, then there exists $p \in \mathcal{S}, z \in \mathcal{Z}$ and $\lambda > 0$ such that $x = p + \lambda z$. After taking z_n to be the normed difference vector of x_n, i.e. $z_n = (x_n - p)/||x_n - p||$, we get $h(p) \geq h(x_n) + \langle y_n, p - x_n \rangle \geq h(x_n) - \langle y_n, z_n \rangle ||x_n - p||$. As $z_n \to z$, taking $n \to \infty$ we get $h(p) \geq \infty$, which is a contradiction. thus our claim that $x = lim\ x_n \in \mathcal{S}$, as claimed.

References

[1] P. Mertikopoulos, Y.-P. Hsieh, and V. Cevher, "A unified stochastic approximation framework for learning in games," 2023.

[2] C. Daskalakis, A. Ilyas, V. Syrgkanis, and H. Zeng, "Training gans with optimism," *CoRR*, vol. abs/1711.00141, 2017.

[3] P. Mertikopoulos, H. Zenati, B. Lecouat, C. Foo, V. Chandrasekhar, and G. Piliouras, "Mirror descent in saddle-point problems: Going the extra (gradient) mile," *CoRR*, vol. abs/1807.02629, 2018.

[4] P. Kamalaruban, Y. Huang, Y. Hsieh, P. Rolland, C. Shi, and V. Cevher, "Robust reinforcement learning via adversarial training with langevin dynamics," *CoRR*, vol. abs/2002.06063, 2020.

[5] V. Vovk, "Aggregating strategies," in *Proceedings of the Third Annual Workshop on Computational Learning Theory* (M. Fulk and J. Case, eds.), pp. 371–383, Morgan Kaufmann, 1990.

[6] M. Bravo, D. Leslie, and P. Mertikopoulos, "Bandit learning in concave n-person games," in *Advances in Neural Information Processing Systems* (S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, eds.), vol. 31, Curran Associates, Inc., 2018.

[7] S. Bervoets, M. Bravo, and M. Faure, "Learning with minimal information in continuous games," *Theoretical Economics*, vol. 15, no. 4, pp. 1471–1508, 2020.

[8] A. Giannou, E.-V. Vlatakis-Gkaragkounis, and P. Mertikopoulos, "On the rate of convergence of regularized learning in games: From bandits and uncertainty to optimism and beyond," in *Advances in Neural Information Processing Systems* (M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, eds.), vol. 34, pp. 22655–22666, Curran Associates, Inc., 2021.