On convergence of online learning in routing games

Walid Krichene   Benjamin Drighes   Alex Bayen

September 13, 2013

Outline

1. No-regret selfish routing
   - The routing game and Nash equilibria
   - No-regret routing
   - Weak convergence of no-regret routing
2. Discounted regret
   - Motivation for decreasing learning rates
   - Weak convergence of no-discounted-regret learning
3. Strong convergence
   - A continuous-time version of dynamics
   - The REP update rules
4. Open problems and extensions

Outline

1. No-regret selfish routing
   - The routing game and Nash equilibria
   - No-regret routing
   - Weak convergence of no-regret routing
2. Discounted regret
   - Motivation for decreasing learning rates
   - Weak convergence of no-discounted-regret learning
3. Strong convergence
   - A continuous-time version of dynamics
   - The REP update rules
4. Open problems and extensions

Routing game

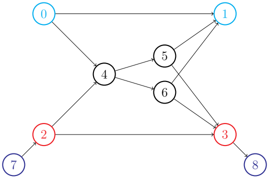

Figure : Example network

- Graph (V, E)
- source-sink pairs, (s_k, t_k): total flow F_k (cars/s, packets/s etc.), paths $\mathcal{P}_k$
- feasible flow: f such that for all k, $\sum_{p \in \mathcal{P}_k} f_p = F_k$

Routing game

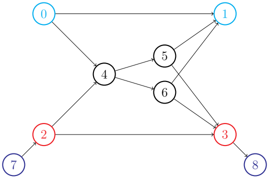

Figure : Example network

- Graph (V, E)
- source-sink pairs, (s_k, t_k): total flow F_k (cars/s, packets/s etc.), paths $\mathcal{P}_k$
- feasible flow: f such that for all k, $\sum_{p \in \mathcal{P}_k} f_p = F_k$
- Latency on edge e: $\ell_e : f_e \mapsto \ell_e(f_e)$, convex increasing

Routing game

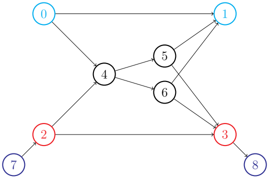

Figure : Example network

- Graph (V, E)
- source-sink pairs, (s_k, t_k): total flow F_k (cars/s, packets/s etc.), paths $\mathcal{P}_k$
- feasible flow: f such that for all k, $\sum_{p \in \mathcal{P}_k} f_p = F_k$
- Latency on edge e: $\ell_e : f_e \mapsto \ell_e(f_e)$, convex increasing
- Players choose a path $p \in \mathcal{P}_k$ selfishly,
  want to minimize personal latency $\ell_p(f) = \sum_{e \in p} \ell_e(f_e)$

player = infinitesimal amount of flow.
f = combined decision of all players.

More precisely

- Measurable set of players $(S_k, \mathcal{S}_k, m_k)$, atomless
- $F_k = m_k(S_k)$

More precisely

- Measurable set of players $(S_k, \mathcal{S}_k, m_k)$, atomless
- $F_k = m_k(S_k)$
- Path choice function

$$\begin{aligned} C_k : S_k &\to \mathcal{P}_k \\ x &\mapsto C_k(x) \end{aligned}$$

$$f_p^k = m_k(C_k^{-1}(\{p\}))$$

Selfish routing game

Nash equilibrium

f is a Nash equilibrium if for all k, for all $p \in \mathcal{P}_k$ with positive flow, $\ell_p(f)$ is minimal on $\mathcal{P}_k$
$(\ell_p(f) \leq \ell_{p'}(f)$ for all $p' \in \mathcal{P}_k)$.

Selfish routing game

Nash equilibrium

f is a Nash equilibrium if for all k, for all $p \in \mathcal{P}_k$ with positive flow, $\ell_p(f)$ is minimal on $\mathcal{P}_k$
$(\ell_p(f) \leq \ell_{p'}(f)$ for all $p' \in \mathcal{P}_k)$.

Equivalent to Nash equilibrium for almost every player.

Selfish routing game

Nash equilibrium

f is a Nash equilibrium if for all k, for all $p \in \mathcal{P}_k$ with positive flow, $\ell_p(f)$ is minimal on $\mathcal{P}_k$
($\ell_p(f) \leq \ell_{p'}(f)$ for all $p' \in \mathcal{P}_k$).

Equivalent to Nash equilibrium for almost every player.

- How to compute Nash equilibria?

Selfish routing game

Nash equilibrium

f is a Nash equilibrium if for all k, for all $p \in \mathcal{P}_k$ with positive flow, $\ell_p(f)$ is minimal on $\mathcal{P}_k$
($\ell_p(f) \leq \ell_{p'}(f)$ for all $p' \in \mathcal{P}_k$).

Equivalent to Nash equilibrium for almost every player.

- How to compute Nash equilibria?
  Convex formulation

Rosenthal potential function

f is a Nash equilibrium iff it minimizes a potential function

$$\text{minimize}_{f \geq 0, \phi} \quad \sum_e \int_0^{\phi_e} \ell_e(u)du$$

$$\text{subject to} \quad \forall e, \sum_{p \ni e} f_p = \phi_e \quad \sum_p f_p = F$$

Motivation for a learning model

- How do players find a Nash equilibrium?

Motivation for a learning model

- How do players find a Nash equilibrium?
  Ideally: distributed, and has minimal information requirements.

Motivation for a learning model

- How do players find a Nash equilibrium?
  Ideally: distributed, and has minimal information requirements.
- Need a model of dynamics to apply control

Outline

1. No-regret selfish routing
   - The routing game and Nash equilibria
   - No-regret routing
   - Weak convergence of no-regret routing
2. Discounted regret
   - Motivation for decreasing learning rates
   - Weak convergence of no-discounted-regret learning
3. Strong convergence
   - A continuous-time version of dynamics
   - The REP update rules
4. Open problems and extensions

The hedge algorithm

Fix one player.
Player maintains a probability distribution $\mu(t)$ over paths, draw path according to $\mu(t)$.

Multiplicative Weights

- distribution $\mu(t)$ over paths p on day t
- update the distribution according to observed loss $\mu_p(t+1) \propto \mu_p(t) e^{-\gamma \ell_p(t)}$

Regret Bound

- Assume losses are in $[0, \rho]$.
- Expected loss is $\ell_{alg}(t) = \sum_p \mu_p(t)\ell_p(t)$

$$R(T) = \sum_{t=1}^{T} \ell_{alg}(t) - \min_p \sum_{t=1}^{T} \ell_p(t)$$

Regret Bound

- Assume losses are in $[0, \rho]$.
- Expected loss is $\ell_{alg}(t) = \sum_p \mu_p(t)\ell_p(t)$

$$R(T) = \sum_{t=1}^{T} \ell_{alg}(t) - \min_p \sum_{t=1}^{T} \ell_p(t)$$

Regret of the expected loss

$$\frac{R(T)}{T} \leq \frac{\rho \ln |\mathcal{P}|}{T\gamma} + \rho\gamma$$

No-regret routing at the population level

- Assume all players apply the same learning algorithm.

No-regret routing at the population level

- Assume all players apply the same learning algorithm.
- For any player x, $C(x,t)$ is a random variable with $P(C(x,t) = p) = \mu_p(t)$

No-regret routing at the population level

- Assume all players apply the same learning algorithm.
- For any player x, $C(x, t)$ is a random variable with $P(C(x, t) = p) = \mu_p(t)$
- The flows $f_p(t) = m(C(\cdot, t)^{-1}(\{p\}))$ is a random variable

No-regret routing at the population level

- Assume all players apply the same learning algorithm.
- For any player x, $C(x, t)$ is a random variable with $P(C(x, t) = p) = \mu_p(t)$
- The flows $f_p(t) = m(C(\cdot, t)^{-1}(\{p\}))$ is a random variable
- $f_p(t) = \mu_p(t)$ a.s. (Fubini's theorem)

Outline

1. No-regret selfish routing
   - The routing game and Nash equilibria
   - No-regret routing
   - Weak convergence of no-regret routing
2. Discounted regret
   - Motivation for decreasing learning rates
   - Weak convergence of no-discounted-regret learning
3. Strong convergence
   - A continuous-time version of dynamics
   - The REP update rules
4. Open problems and extensions

Weak convergence

Weak convergence of no-regret routing

If an update rule satisfies the regret bound, then for all $\epsilon > 0$, there exists $\gamma > 0$ such that no-regret learning with rate γ converges in the sense

$$\lim_{T \to \infty} \frac{1}{T} \sum_{t=1}^{T} \mu(t) \in \mathcal{N}_\epsilon$$

$\mathcal{N}_\epsilon$: ϵ-approximate Nash equilibrium.

Recall the regret bound

$$\frac{R(T)}{T} \leq \frac{\rho \ln |\mathcal{P}|}{T\gamma} + \rho\gamma$$

Outline

1. No-regret selfish routing
   - The routing game and Nash equilibria
   - No-regret routing
   - Weak convergence of no-regret routing
2. Discounted regret
   - Motivation for decreasing learning rates
   - Weak convergence of no-discounted-regret learning
3. Strong convergence
   - A continuous-time version of dynamics
   - The REP update rules
4. Open problems and extensions

Hedge as a regularized greedy algorithm

Can show the hedge update rule is solution to

Greedy algorithm, regularized by the K-L divergence

$$\text{minimize}_{\mu \geq 0} \quad \sum_p \mu_p \ell_p(t-1) + \frac{1}{\gamma} D(\mu || \mu(t-1))$$

$$\text{subject to} \quad \sum_p \mu_p = 1$$

Hedge as a regularized greedy algorithm

Can show the hedge update rule is solution to

Greedy algorithm, regularized by the K-L divergence

$$\text{minimize}_{\mu \geq 0} \quad \sum_p \mu_p \ell_p(t-1) + \frac{1}{\gamma} D(\mu \| \mu(t-1))$$

$$\text{subject to} \quad \sum_p \mu_p = 1$$

- $D(\mu \| \mu(t-1)) = \sum_p \mu_p \ln \frac{\mu_p}{\mu_p(t-1)}$
- $\ell_p(t-1)$ loss on the previous day.

Hedge as a regularized greedy algorithm

Can show the hedge update rule is solution to

Greedy algorithm, regularized by the K-L divergence

$$\text{minimize}_{\mu \geq 0} \quad \sum_p \mu_p \ell_p(t-1) + \frac{1}{\gamma} D(\mu || \mu(t-1))$$

$$\text{subject to} \quad \sum_p \mu_p = 1$$

- $D(\mu || \mu(t-1)) = \sum_p \mu_p \ln \frac{\mu_p}{\mu_p(t-1)}$
- $\ell_p(t-1)$ loss on the previous day.

Limit cases:

- $\gamma \to \infty$, greedy algorithm
- $\gamma \to 0$, static distribution

Discounting the losses

Put weights on time $(\gamma(t))_{t\in\mathbb{N}}$: players care more about present than future.
The sequence of discounting factors γ is universal.
Assumption: γ positive, non-summable, $\longrightarrow 0$.

Discounting the losses

Put weights on time $(\gamma(t))_{t\in\mathbb{N}}$: players care more about present than future.
The sequence of discounting factors γ is universal.
Assumption: γ positive, non-summable, $\longrightarrow 0$.

Definition (Discounted regret)

$$R(T) = \sum_{t=0}^{T} \gamma(t) \sum_{p} \mu_p(t)\ell_p(\mu(t)) - \min_p \sum_{t=0}^{T} \gamma(t)\ell_p(\mu(t))$$

No-regret if

$$\frac{1}{\sum_{t=0}^{T} \gamma(t)} R(T) \xrightarrow{T\to\infty} 0$$

Discounted hedge algorithm

Regret bound

$$R(T) \leq \rho \log |\mathcal{P}| + \rho \sum_{t \leq T} \gamma(t)^2/8$$

Discounted hedge algorithm

Regret bound

$$R(T) \leq \rho \log |\mathcal{P}| + \rho \sum_{t \leq T} \gamma(t)^2/8$$

Consequence: if γ is square-summable, discounted Hedge achieves no-regret.

Outline

1. No-regret selfish routing
   - The routing game and Nash equilibria
   - No-regret routing
   - Weak convergence of no-regret routing
2. Discounted regret
   - Motivation for decreasing learning rates
   - Weak convergence of no-discounted-regret learning
3. Strong convergence
   - A continuous-time version of dynamics
   - The REP update rules
4. Open problems and extensions

Weak convergence to Nash equilibria

Theorem

Under a discounted no-regret routing algorithm, $(\mu(t))_t$ converges to Nash equilibria on a subset of days of density one.

- subsequence $(\mu_{t_k})_k$ converges
- $\frac{\sum_{t_k \leq T} \gamma(t_k)}{\sum_{t \leq T} \gamma(t)} \xrightarrow{T \to \infty} = 1$

Weak convergence to Nash equilibria

Theorem

Under a discounted no-regret routing algorithm, $(\mu(t))_t$ converges to Nash equilibria on a subset of days of density one.

- subsequence $(\mu_{t_k})_k$ converges
- $\frac{\sum_{t_k \le T} \gamma(t_k)}{\sum_{t \le T} \gamma(t)} \xrightarrow{T \to \infty} = 1$

Proof.

- Convexity:

$$V(\mu(t)) - V(\mu) \le \nabla V(\mu(t))^T(\mu(t) - \mu) = \sum_{k=1}^{K} F_k \sum_{p \in \mathcal{P}_k} \ell_p(\mu(t))(\mu_p(t) - \mu_p)$$

Weak convergence to Nash equilibria

Theorem

Under a discounted no-regret routing algorithm, $(\mu(t))_t$ converges to Nash equilibria on a subset of days of density one.

- subsequence $(\mu_{t_k})_k$ converges
- $\frac{\sum_{t_k \leq T} \gamma(t_k)}{\sum_{t \leq T} \gamma(t)} \xrightarrow{T \to \infty} = 1$

Proof.

- Convexity:

$$V(\mu(t)) - V(\mu) \leq \nabla V(\mu(t))^T (\mu(t) - \mu) = \sum_{k=1}^{K} F_k \sum_{p \in \mathcal{P}_k} \ell_p(\mu(t))(\mu_p(t) - \mu_p)$$

- Discounted no-regret:

$$\sum_{t \leq T} \gamma(t)(V(\mu(t)) - V(\mu)) \leq \sum_{k=1}^{K} F_k R_k(T)$$

Weak convergence to Nash equilibria

Theorem

Under a discounted no-regret routing algorithm, $(\mu(t))_t$ converges to Nash equilibria on a subset of days of density one.

- subsequence $(\mu_{t_k})_k$ converges
- $\frac{\sum_{t_k \leq T} \gamma(t_k)}{\sum_{t \leq T} \gamma(t)} \xrightarrow{T\to\infty} = 1$

Proof.

- Convexity:

$$V(\mu(t)) - V(\mu) \leq \nabla V(\mu(t))^T(\mu(t) - \mu) = \sum_{k=1}^{K} F_k \sum_{p \in \mathcal{P}_k} \ell_p(\mu(t))(\mu_p(t) - \mu_p)$$

- Discounted no-regret:

$$\sum_{t \leq T} \gamma(t)(V(\mu(t)) - V(\mu)) \leq \sum_{k=1}^{K} F_k R_k(T)$$

- Absolute Cesaro convergence implies convergence on a subset of density one.

Outline

Replicator dynamics

Replicator equation

In the update equation $\mu_p(t+1) \propto \mu_p(t) e^{-\gamma \ell_p(t)}$, let $\gamma \to 0$.
We obtain the autonomous ODE:

$$\begin{cases} \mu(0) \in \mathring{\Delta} \\ \forall p \in \mathcal{P}_k, \frac{d\mu_p}{dt} = \mu_p(\bar{\ell}^k(\mu) - \ell_p(\mu))/\rho \end{cases} \quad (1)$$

Also in evolutionary game theory.

Restricted Nash equilibria are stationary points, partitioned into:

- Nash equilibria
- non-Nash equilibria

Replicator dynamics

Replicator equation

In the update equation $\mu_p(t+1) \propto \mu_p(t) e^{-\gamma \ell_p(t)}$, let $\gamma \to 0$.
We obtain the autonomous ODE:

$$\begin{cases} \mu(0) \in \mathring{\Delta} \\ \forall p \in \mathcal{P}_k, \frac{d\mu_p}{dt} = \mu_p(\bar{\ell}^k(\mu) - \ell_p(\mu))/\rho \end{cases} \quad (1)$$

Also in evolutionary game theory.

Restricted Nash equilibria are stationary points, partitioned into:

- Nash equilibria
- non-Nash equilibria unstable

Replicator dynamics

Replicator equation

In the update equation $\mu_p(t+1) \propto \mu_p(t) e^{-\gamma \ell_p(t)}$, let $\gamma \to 0$.
We obtain the autonomous ODE:

$$\begin{cases} \mu(0) \in \mathring{\Delta} \\ \forall p \in \mathcal{P}_k, \frac{d\mu_p}{dt} = \mu_p(\bar{\ell}^k(\mu) - \ell_p(\mu))/\rho \end{cases} \quad (1)$$

Also in evolutionary game theory.

Restricted Nash equilibria are stationary points, partitioned into:

- Nash equilibria
- non-Nash equilibria unstable

Theorem

Every solution of the ODE (1) converges to the set of restricted Nash equilibria of the routing game.

Replicator dynamics

Replicator equation

In the update equation $\mu_p(t+1) \propto \mu_p(t) e^{-\gamma \ell_p(t)}$, let $\gamma \to 0$.
We obtain the autonomous ODE:

$$\begin{cases} \mu(0) \in \mathring{\Delta} \\ \forall p \in \mathcal{P}_k, \frac{d\mu_p}{dt} = \mu_p(\bar{\ell}^k(\mu) - \ell_p(\mu))/\rho \end{cases} \quad (1)$$

Also in evolutionary game theory.

Restricted Nash equilibria are stationary points, partitioned into:

- Nash equilibria
- non-Nash equilibria unstable

Theorem

Every solution of the ODE (1) converges to the set of restricted Nash equilibria of the routing game.

Proof: V is a Lyapunov function for $\mathcal{RN}$.

Outline

1. No-regret selfish routing
   - The routing game and Nash equilibria
   - No-regret routing
   - Weak convergence of no-regret routing
2. Discounted regret
   - Motivation for decreasing learning rates
   - Weak convergence of no-discounted-regret learning
3. Strong convergence
   - A continuous-time version of dynamics
   - The REP update rules
4. Open problems and extensions

REP algorithms

Discretization of the replicator dynamics (REP):

$$\begin{cases} \mu(0) \in \mathring{\Delta} \\ \mu_p(t+1) - \mu_p(t) = \gamma(t)\mu_p(t)\left(\bar{\ell}^k(\mu(t)) - \ell_p(\mu(t))\right)/\rho + \gamma(t)U_p(t+1) \end{cases}$$

$(U(t))_{t\geq 1}$ deterministic or stochastic perturbations that satisfy for all $T > 0$,

$$\lim_{\tau\to\infty} \max\left\{ \left\| \sum_{t=\tau}^{\tau'-1} \gamma(t)U(t+1) \right\| : \tau' = \{\tau+1, \cdots, \sup\{t \geq 0 : t \geq T_t + T\}\} \right\} = 0$$

In particular for $U = 0$, we obtain a new update rule

$$\begin{cases} \mu(0) \in \mathring{\Delta} \\ \mu_p(t+1) - \mu_p(t) = \gamma(t)\mu_p(t)\left(\bar{\ell}^k(\mu(t)) - \ell_p(\mu(t))\right)/\rho \end{cases}$$

In particular for $U = 0$, we obtain a new update rule

$$\begin{cases} \mu(0) \in \mathring{\Delta} \\ \mu_p(t+1) - \mu_p(t) = \gamma(t)\mu_p(t)\left(\bar{\ell}^k(\mu(t)) - \ell_p(\mu(t))\right)/\rho \end{cases}$$

- Discounted no-regret algorithm.

In particular for $U = 0$, we obtain a new update rule

$$\begin{cases} \mu(0) \in \mathring{\Delta} \\ \mu_p(t+1) - \mu_p(t) = \gamma(t)\mu_p(t)\left(\bar{\ell}^k(\mu(t)) - \ell_p(\mu(t))\right)/\rho \end{cases}$$

- Discounted no-regret algorithm.
- Solution to regularized optimization:

$$\mu(t) \in \arg\min_{\mu \in \Delta} \sup_p \mu_p \ell_p(\mu(t-1))/\rho + \frac{1}{\gamma(t)} R(\mu \| \mu(t-1))$$

where $R(x||y) = \frac{1}{2} \sum_p y_p \left(\frac{x_p}{y_p} - 1\right)^2$

In particular for $U = 0$, we obtain a new update rule

$$\begin{cases} \mu(0) \in \mathring{\Delta} \\ \mu_p(t+1) - \mu_p(t) = \gamma(t)\mu_p(t)\left(\bar{\ell}^k(\mu(t)) - \ell_p(\mu(t))\right)/\rho \end{cases}$$

- Discounted no-regret algorithm.
- Solution to regularized optimization:

$$\mu(t) \in \arg\min_{\mu \in \Delta} \sup_p \mu_p \ell_p(\mu(t-1))/\rho + \frac{1}{\gamma(t)} R(\mu \| \mu(t-1))$$

where $R(x||y) = \frac{1}{2}\sum_p y_p \left(\frac{x_p}{y_p} - 1\right)^2$

The discounted hedge algorithm is also a REP algorithm.

Convergence to Nash equilibria

Theorem

Under any discounted no-regret REP algorithm, the sequence $\mu(t)$ converges to the set of Nash equilibria.

Convergence to Nash equilibria

Theorem

Under any discounted no-regret REP algorithm, the sequence $\mu(t)$ converges to the set of Nash equilibria.

Proof:

- Let X be the affine interpolation of the sequence $\mu(t)$. X is an Asymptotic Pseudo Trajectory for the ODE.

Convergence to Nash equilibria

Theorem

Under any discounted no-regret REP algorithm, the sequence $\mu(t)$ converges to the set of Nash equilibria.

Proof:

- Let X be the affine interpolation of the sequence $\mu(t)$. X is an Asymptotic Pseudo Trajectory for the ODE.
- Let $L(X)$ be the set limit points of X.

Convergence to Nash equilibria

Theorem

Under any discounted no-regret REP algorithm, the sequence $\mu(t)$ converges to the set of Nash equilibria.

Proof:

- Let X be the affine interpolation of the sequence $\mu(t)$. X is an Asymptotic Pseudo Trajectory for the ODE.
- Let $L(X)$ be the set limit points of X.
- V is constant over $L(X)$.

Convergence to Nash equilibria

Theorem

Under any discounted no-regret REP algorithm, the sequence $\mu(t)$ converges to the set of Nash equilibria.

Proof:

- Let X be the affine interpolation of the sequence $\mu(t)$. X is an Asymptotic Pseudo Trajectory for the ODE.
- Let $L(X)$ be the set limit points of X.
- V is constant over $L(X)$.
- Use weak convergence to conclude that constant value is minimum of V.

Open problems and extensions

- Conjecture: if $\mu(0) \in \mathring{\Delta}$, the replicator dynamics converge to $\mathcal{N}$

Open problems and extensions

- Conjecture: if $\mu(0) \in \mathring{\Delta}$, the replicator dynamics converge to $\mathcal{N}$
- Relax assumption that all players "learn in the same way" (universal discount sequence $\gamma(t)$, universal initial distribution $\mu(0)$).

Open problems and extensions

- Conjecture: if $\mu(0) \in \mathring{\Delta}$, the replicator dynamics converge to $\mathcal{N}$
- Relax assumption that all players "learn in the same way" (universal discount sequence $\gamma(t)$, universal initial distribution $\mu(0)$).
- Apply control to the system: e.g. tolling.

Thank you.