More Efficient Internal-Regret-Minimizing Algorithms

Amy Greenwald Zheng Li Warren Schudy

July 11, 2008

Outline

- Introduction
- Our algorithms
- Analysis highlights
- Experiments
- Conclusions

External Regret

- Rock-paper scissors history:

	1	2	3	4	5
Them	P	R	R	P	R
Us	R	P	S	R	P

- We won 2, lost 3 (net -1).

External Regret

- Rock-paper scissors history:

	1	2	3	4	5
Them	P	R	R	P	R
Us	R	P	S	R	P

- We won 2, lost 3 (net -1).
- If we had always played P, we would have won 3, lost 0 (net 3). External Regret: $3 - (-1) = 4$.

External Regret

- Rock-paper scissors history:

	1	2	3	4	5
Them	P	R	R	P	R
Us	R	P	S	R	P

- We won 2, lost 3 (net -1).
- If we had always played P, we would have won 3, lost 0 (net 3). External Regret: $3-(-1)=4$.
- No-external-regret: $\lim_{t\to\infty}\max(\frac{\text{largest regret}}{t},0)=0$ where t is number of rounds played.
- Efficient no-external-regret algorithms exist, e.g. Freund and Schapire (1997).

Model of Online Decision Problems

- n available actions
- for $t = 1, 2, \ldots, \infty$
  - Play mixed action q_t (a probability distribution row vector)
  - Receive reward $q_t \pi_t$ (a dot product)
  - Update action
- Assume $0 \leq (\pi_t)_i \leq 1$ for all actions i.
- Can express external regret in matrix form:

$$\phi = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{bmatrix}$$

For any q, $\quad q\phi^P = (0, 1, 0)$

- Regret not playing P: $\quad \sum_{\tau=1}^{t} q_\tau \phi^P \pi_\tau - q_\tau \pi_\tau$.

Internal Regret

- Transform $P \longrightarrow S$ leaving R and S unchanged:

$$\phi^{P\rightarrow S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

if $q = (.1, .3, .6)$ then $q\phi^{P\rightarrow S} = (.1, 0, .9)$

Internal Regret

- Transform $P \to S$ leaving R and S unchanged:

$$\phi^{P\to S} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

  if $q = (.1, .3, .6)$ then $q\phi^{P\to S} = (.1, 0, .9)$

- Let $\Phi_{\text{INT}} = \{\, \phi^{a\to b} \; : \; a \neq b, 1 \leq a, b \leq n \,\}$ where

$$(\phi^{a\to b})_{ij} = \begin{cases} 1 & \text{if } i \neq a \wedge i = j \\ 1 & \text{if } i = a \wedge j = b \\ 0 & \text{otherwise} \end{cases}$$

- Internal regret vector R_t:

$$(R_t)_\phi = \sum_{\tau=1}^{t} q_\tau \phi \pi_\tau - q_\tau \pi_\tau \quad \text{where } \phi \in \Phi_{\text{INT}}.$$

  No-internal-regret: $\lim_{t\to\infty} \max(\frac{(R_t)_\phi}{t}, 0) = 0$

  (Implies convergence to correlated equilibria)

Internal-Regret-Minimizing Algorithms

	Runtime	Average Regret
Previous	$O(n^3)$	$O(\sqrt{\frac{n}{t}})$ or $O(\sqrt{\frac{\log n}{t}})$
Our “Power Iteration” (PI)	$O(n^2)$	$O(\sqrt{n}t^{-1/10})$
Our “Multithreaded” (MT)	$O(n^3/p)$	$O\left(\sqrt{\frac{np}{t}}\right)$

- Notation:
  - n actions available to us
  - t rounds played so far
  - p is a tunable parameter
- Many previous algorithms achieve the stated bounds; e.g. Foster and Vohra (1999), Greenwald et al. (To Appear).

Internal-Regret-Minimizing Algorithms

	Runtime	Average Regret
Previous	$O(n^3)$	$O(\sqrt{\frac{n}{t}})$ or $O(\sqrt{\frac{\log n}{t}})$
Our “Power Iteration” (PI)	$O(n^2)$	$O(\sqrt{n}t^{-1/10})$
Our “Multithreaded” (MT)	$O(n^3/p)$	$O\left(\sqrt{\frac{np}{t}}\right)$

- Notation:
  - n actions available to us
  - t rounds played so far
  - p is a tunable parameter
- Many previous algorithms achieve the stated bounds; e.g. Foster and Vohra (1999), Greenwald et al. (To Appear).
- Young (2004) stated an algorithm similar to our PI algorithm but did not analyze it rigorously.
- The first and third stated results are for the “natural” $O(n^3)$ matrix inversion algorithms; $O(n^{2.36})$ algorithms are known.

Outline

- Introduction
- Our algorithms
- Analysis highlights
- Experiments
- Conclusions

NIR Algorithm Basics

Define:

$$(R_t^+)_\phi \equiv \max((R_t)_\phi, 0)$$

$$D_t = \sum_{\phi \in \Phi_{\text{INT}}} (R_t^+)_\phi \quad \text{and} \quad N_t = \sum_{\phi \in \Phi_{\text{INT}}} (R_t^+)_\phi \phi$$

For example, if $(R_t)_{R \to S} = 3$, $(R_t)_{S \to P} = 2$, and all other components are non-positive, then:

$$D_t = 3 + 2 = 5 \quad \text{and} \quad N_t = \begin{bmatrix} 2 & 0 & 3 \\ 0 & 5 & 0 \\ 0 & 2 & 3 \end{bmatrix}$$

NIR Algorithm Basics

Define:

$$(R_t^+)_\phi \equiv \max((R_t)_\phi, 0)$$

$$D_t = \sum_{\phi \in \Phi_{\text{INT}}} (R_t^+)_\phi \quad \text{and} \quad N_t = \sum_{\phi \in \Phi_{\text{INT}}} (R_t^+)_\phi \phi$$

For example, if $(R_t)_{R \to S} = 3$, $(R_t)_{S \to P} = 2$, and all other components are non-positive, then:

$$D_t = 3 + 2 = 5 \quad \text{and} \quad N_t = \begin{bmatrix} 2 & 0 & 3 \\ 0 & 5 & 0 \\ 0 & 2 & 3 \end{bmatrix}$$

N_t / D_t is stochastic, and therefore has a fixed point:

$$(0, 1, 0) \begin{bmatrix} 2/5 & 0 & 3/5 \\ 0 & 5/5 & 0 \\ 0 & 2/5 & 3/5 \end{bmatrix} = (0, 1, 0)$$

A Traditional NIR Algorithm

- Initialize q_1 to be an arbitrary mixed action.
- During each round $t = 1, 2, 3, \ldots$:

1. Play mixed action q_t.
2. Update the regret vector R_t based on observed rewards π_t.
3. Set the mixed action q_{t+1} to a fixed-point of $\frac{N_t}{D_t}$.

Foster and Vohra (1999), Greenwald et al. (To Appear)

A Traditional NIR Algorithm

- Initialize q_1 to be an arbitrary mixed action.
- During each round $t = 1, 2, 3, \ldots$:

1. Play mixed action q_t.
2. Update the regret vector R_t based on observed rewards π_t.
3. Set the mixed action q_{t+1} to a fixed-point of $\frac{N_t}{D_t}$.

Foster and Vohra (1999), Greenwald et al. (To Appear)

Theorem 2 *[Greenwald et al. (2006)] This algorithnm has per-round runtime $O(LS(n))$ and regret bound*

$$\left\| \frac{R_t^+}{t} \right\|_\infty \leq \sqrt{\frac{(n-1)}{t}}$$

where $LS(n)$ is the time required to invert an $n \times n$ matrix.

“Power Iteration” Algorithm

- Initialize q_1 to be an arbitrary mixed action.
- During each round $t = 1, 2, 3, \ldots$:

1. Play mixed action q_t.
2. Update the regret vector R_t based on observed rewards π_t.
3. Set the mixed action $q_{t+1} \leftarrow q_t \frac{N_t}{D_t}$.

“Power Iteration” Algorithm

- Initialize q_1 to be an arbitrary mixed action.
- During each round $t = 1, 2, 3, \ldots$:

1. Play mixed action q_t.
2. Update the regret vector R_t based on observed rewards π_t.
3. Set the mixed action $q_{t+1} \leftarrow q_t \frac{N_t}{D_t}$.

Theorem 4 *PI has per-round runtime $O(n^2)$ and regret bound*

$$\left\| \frac{R_t^+}{t} \right\|_\infty \leq O(\sqrt{n} t^{-1/10})$$

We prove this natural and fast algorithm) (which has previously been used in practice without proof) has internal regret tending to zero.

“Multi-Threaded” Algorithm

Action thread: During each round $t = 1, 2, 3, \ldots$:

- Play mixed action q_t, the most recent fixed point computed by the compute thread.
- Update the regret vector R_t based on observed rewards π_t.

Compute thread: Repeat forever:

- Wait until the action thread updates the regret vector R_τ.
- Compute a fixed point of N_τ / D_τ.

“Multi-Threaded” Algorithm

Action thread: During each round $t = 1, 2, 3, \ldots$:

- Play mixed action q_t, the most recent fixed point computed by the compute thread.
- Update the regret vector R_t based on observed rewards π_t.

Compute thread: Repeat forever:

- Wait until the action thread updates the regret vector R_τ.
- Compute a fixed point of N_τ / D_τ.

Theorem 6 *For any number of time-steps per fixed-point $p \geq 1$, MT has per-round run time $O(LS(n)/p + \log n + \alpha)$ where α is the time required to update the regret and regret bound*

$$\left\| \frac{R_t^+}{t} \right\|_\infty \leq \sqrt{\frac{(n-1)(4p-3)}{t}}$$

Outline

- Introduction
- Our algorithms
- Analysis highlights
- Experiments
- Conclusions

A Lemma

Lemma 7 *For any online learning algorithm and any function* $w(\cdot) > 0$*, we have the following inequality: for all times* $t > 0$*,*

$$\left\| R_t^+ \right\|_2^2 \leq 2 \sum_{\tau=1}^{t} q_\tau \left(N_{\tau - w(\tau)} - D_{\tau - w(\tau)} I \right) \pi_t + (n-1) \sum_{\tau=1}^{t} (2w(\tau) - 1))$$

where I *is the identity matrix.*

- First term “fixed point” quality (zero if $q_\tau (N_{\tau - w(\tau)} / D_{\tau - w(\tau)}) = q_\tau$)
- Second term “fixed point” age
- Proof straightforward.

MT Analysis

Let $w(\tau)$ be such that q_τ is fixed point of $N_{\tau-w(\tau)}/D_{\tau-w(\tau)}$. One can bound $w(\tau)$ above by $2p-1$.

MT Analysis

Let $w(\tau)$ be such that q_τ is fixed point of $N_{\tau-w(\tau)}/D_{\tau-w(\tau)}$. One can bound $w(\tau)$ above by $2p-1$. Apply Lemma 7:

$$\begin{aligned} \left\|R_t^+\right\|_2^2 &\leq 2\sum_{\tau=1}^{t} 0 + (n-1)\sum_{\tau=1}^{t}(2(2p-1)-1) \\ &= (n-1)t(4p-3) \end{aligned}$$

MT Analysis

Let $w(\tau)$ be such that q_τ is fixed point of $N_{\tau-w(\tau)}/D_{\tau-w(\tau)}$. One can bound $w(\tau)$ above by $2p-1$. Apply Lemma 7:

$$\begin{aligned} \left\|R_t^+\right\|_2^2 &\leq 2\sum_{\tau=1}^{t} 0 + (n-1)\sum_{\tau=1}^{t}(2(2p-1)-1) \\ &= (n-1)t(4p-3) \end{aligned}$$

Therefore:

$$\left\|\frac{R_t^+}{t}\right\|_\infty \leq \left\|\frac{R_t^+}{t}\right\|_2 \leq \sqrt{\frac{(n-1)(4p-3)}{t}}$$

PI Analysis (Sketch)

- Also uses Lemma 7
- Set $w(\tau) = \tau^{2/5}$
- The following Lemma, which generalizes a Lemma used in Hart and Mas-Colell (2000), is key:

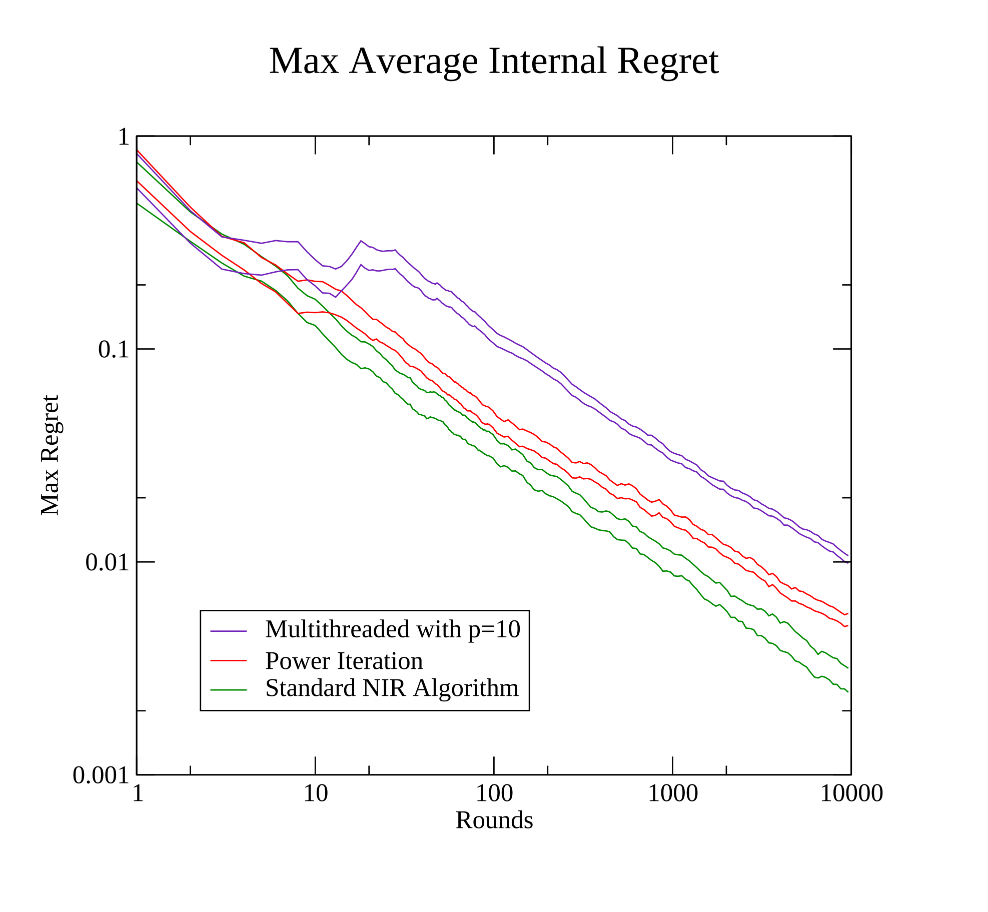
Lemma 8 *For all $z > 0$, if P is n-dimensional stochastic matrix that is close to the identity matrix in the sense that $\sum_{i=1}^{n} P_{ii} \geq n - 1$, then $\left\| q(P^z - P^{z-1}) \right\|_1 = O(1/\sqrt{z})$ for all n-dimensional vectors q with $\|q\|_1 = 1$.*

Outline

- Introduction
- Our algorithms
- Analysis highlights
- Experiments
- Conclusions

Experiments

Shapley Game: $\begin{bmatrix} 0,0 & 0,1 & 1,0 \\ 1,0 & 0,0 & 0,1 \\ 0,1 & 1,0 & 0,0 \end{bmatrix}$



Outline

- Introduction
- Our algorithms
- Analysis highlights
- Experiments
- Conclusions

Conclusions and Future Work

- Two new internal-regret-minimizing algorithms with tradeoff between runtime and convergence rate
- Open questions:
  - More sophisticated iterative method than power iteration such as bi-conjugate gradient?
  - Can regret bound for Power Iteration by improved to better match experiments?
  - Other link/potential functions to improve regret from $O(\sqrt{n}t^{-c})$ to $O(\sqrt{\log n}t^{-c})$?

Acknowledgements

Thanks to:

- Dean Foster
- Casey Marks
- Yuval Peres
- John Wicks

References

D. Foster and R. Vohra. Regret in the on-line decision problem. *Games and Economic Behavior*, 29:7–35, 1999

Y. Freund and R. E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computer and System Sciences*, 55: 119–139, 1997

A. Greenwald, A. Jafari, and C. Marks. A general class of no-regret algorithms and game-theoretic equilibria. In Amitabha Gupta, Johan van Benthem, and Eric Pacuit, editors, *Logic at the Crossroads: An Interdisciplinary View*, volume 2. Allied Publishers, To Appear

A. Greenwald, Z. Li, and C. Marks. Bounds for regret-matching algorithms. In *Proceedings of the Ninth International Symposium on Artificial Intelligence and Mathematics*, 2006

P. Young. *Strategic Learning and its Limits*. Oxford University Press, Oxford, 2004