First-order algorithm with $\mathcal{O}(\log(1/\epsilon))$ convergence for ϵ-equilibrium in two-person zero-sum games

Javier Peña
Carnegie Mellon University

joint work with

A. Gilpin, T. Sandholm (Carnegie Mellon),
B. Mordukhovich (Wayne State), V. Roshchina (U of Évora)

IPAM, UCLA
Los Angeles, October 2010

The Problem

Nash equilibrium (two-person, zero-sum games)

$$\max_{x \in Q_1} \min_{y \in Q_2} x^{\mathsf{T}} A y = \min_{y \in Q_2} \max_{x \in Q_1} x^{\mathsf{T}} A y.$$

- Q_1, Q_2: sets of strategies of players 1 and 2 respectively
- A: player 1's payoff matrix
- Games in normal form: Q_1, Q_2 are simplices.
- Games in sequential form: Q_1, Q_2 are *treeplexes.*

For simplicity, assume $A \in \mathbb{R}^{m \times n}$, $Q_1 = \Delta_m$, $Q_2 = \Delta_n$.

Algorithms to compute Nash equilibria

Definition

Given $\epsilon > 0$, a point $(x, y) \in \Delta_m \times \Delta_n$ is an ϵ-equilibrium if

$$\max A^{\mathsf{T}} x - \min Ay \leq \epsilon.$$

Here $\max A^{\mathsf{T}} x = \max_{j=1,\ldots,n} (A^{\mathsf{T}} x)_j$. Likewise for $\min Ay$.

Algorithms to compute ϵ-equilibria

- Interior-point methods: $\mathcal{O}(\log(1/\epsilon))$ iteration complexity
- Subgradient methods: $\mathcal{O}(1/\epsilon^2)$ iteration complexity
- (Accelerated) first-order methods: $\mathcal{O}(1/\epsilon)$ iteration complexity

Main result

Tradeoff

One first-order iteration is far simpler than one interior-point iteration.

Main Theorem (Gilpin, P, Sandholm 2009)

First-order algorithm to compute an ϵ-equilibrium with $\mathcal{O}(\kappa(A)\log(1/\epsilon))$ iteration complexity.

- Same dependence on ϵ as interior-point methods.
- Same simplicity per iteration as first-order methods.
- Dependence on a *condition measure* $\kappa(A)$ of A.

Some Notation

Let $F : \Delta_m \times \Delta_n \to \mathbb{R}$ be defined by

$$F(x, y) := \max A^\mathsf{T} x - \min Ay.$$

Note: $F(w) = \max_{u \in \Delta_m \times \Delta_n} u^\mathsf{T} M w$ for $M := \begin{bmatrix} 0 & -A \\ A^\mathsf{T} & 0 \end{bmatrix}$.

Note: $w = (x, y)$ is an ϵ-equilibrium iff $F(x, y) \leq \epsilon$.

Let

$$\begin{aligned} S &:= \operatorname{Argmin}\{F(u) : u \in \Delta_m \times \Delta_n\} \\ &= \{w \in \Delta_m \times \Delta_n : F(w) = 0\}. \end{aligned}$$

Formulate equilibrium problem as

$$\min_{w \in \Delta_m \times \Delta_n} F(w).$$

Nesterov's Smoothing Algorithm for $\min_{w \in \Delta_m \times \Delta_n} F(w)$

For $\mu > 0$ let

$$F_\mu(w) := \max_{u \in \Delta_m \times \Delta_n} \left\{ u^\mathsf{T} M w - \frac{\mu}{2} \|u - \bar{u}\|^2 \right\}.$$

smoothing(A, w_0, ϵ)
Let $\mu = \frac{\epsilon}{4}$ and $z_0 = w_0$
For $k = 0, 1, \ldots$

- $u_k = \frac{2}{k+2} z_k + \frac{k}{k+2} w_k$
- $w_{k+1} =$
  $\operatorname{argmin} \left\{ \langle \nabla F_\mu(u_k), w - u_k \rangle + \frac{\|A\|^2}{2\mu} \|w - u_k\|^2 : w \in \Delta_m \times \Delta_n \right\}$
- If $F(w_{k+1}) < \epsilon$ return w_{k+1}
- $z_{k+1} =$
  $\operatorname{argmin} \left\{ \sum_{i=0}^{k} \frac{i+1}{2} \langle \nabla F_\mu(u_i), z - u_i \rangle + \frac{\|A\|^2}{2\mu} \|z - w_0\|^2 : z \in \Delta_m \times \Delta_n \right\}$

Nesterov's Smoothing Algorithm for $\min_{w \in \Delta_m \times \Delta_n} F(w)$

Theorem (Lan, Lu, Monteiro 2006 & Nesterov 2004)

Algorithm **smoothing** *finishes in at most*

$$\frac{4 \cdot \|A\| \cdot \mathrm{dist}(w_0, \mathrm{S})}{\epsilon}$$

first-order iterations.

Here $\mathrm{dist}(w, \mathrm{S}) := \min\{\|w - u\| : u \in \mathrm{S}\}$.

Iterated Smoothing Algorithm

Let $\gamma > 1$ be fixed.

iterated$(A, x_0, y_0, \gamma, \epsilon)$

(1) Let $\epsilon_0 = F(x_0, y_0)$

(2) For $i = 0, 1, \ldots$

- $\epsilon_{i+1} = \frac{\epsilon_i}{\gamma}$
- $(x_{i+1}, y_{i+1}) =$ **smoothing**$(A, x_i, y_i, \epsilon_{i+1})$
- If $F(x_{i+1}, y_{i+1}) < \epsilon$, return (x_{i+1}, y_{i+1})

Main Theorem

Condition Measure $\delta(A)$

$$\delta(A) := \sup\left\{\delta : \mathrm{dist}((x,y),\mathrm{S}) \le \frac{F(x,y)}{\delta} \ \ \forall\,(x,y) \in \Delta_m \times \Delta_n\right\}.$$

Main Theorem (Gilpin, P, Sandholm 2009)

Algorithm **iterated** *finishes after at most*

$$\frac{4\cdot\gamma\cdot\|A\|\cdot\log(2\|A\|/\epsilon)}{\log(\gamma)\cdot\delta(A)}$$

first-order iterations.

Proof of Main Theorem

Claim 1

Each call to **smoothing** in Algorithm **iterated** halts in at most

$$\frac{4 \cdot \|A\| \cdot \gamma}{\delta(A)}$$

first-order iterations.

Proof of Main Theorem

Claim 1

Each call to **smoothing** in Algorithm **iterated** halts in at most

$$\frac{4 \cdot \|A\| \cdot \gamma}{\delta(A)}$$

first-order iterations.

Proof.

For $i = 0, 1, \ldots$ we have $\text{dist}((x_i, y_i), \text{S}) \leq \frac{F(x_i, y_i)}{\delta(A)} \leq \frac{\epsilon_i}{\delta(A)} = \frac{\gamma \cdot \epsilon_{i+1}}{\delta(A)}$.
Next, apply Lan et al./Nesterov's Theorem: i-th call to **smoothing** will halt after

$$\frac{4 \cdot \|A\| \cdot \text{dist}((x_i, y_i), \text{S})}{\epsilon_{i+1}} \leq \frac{4 \cdot \|A\| \cdot \gamma}{\delta(A)}.$$

first-order iterations. $\square$

Proof of Main Theorem

Claim 2

Algorithm **iterated** halts in at most

$$\frac{\log(2\|A\|/\epsilon)}{\log(\gamma)}$$

outer iterations.

Proof of Main Theorem

Claim 2

Algorithm **iterated** halts in at most

$$\frac{\log(2\|A\|/\epsilon)}{\log(\gamma)}$$

outer iterations.

Proof.

After N outer iterations, we get $(x_N, y_N) \in \Delta_m \times \Delta_n$ with

$$F(x_N, y_N) < \epsilon_N = \frac{\epsilon_0}{\gamma^N} = \frac{F(x_0, y_0)}{\gamma^N} \leq \frac{2\|A\|}{\gamma^N}.$$

Thus, $F(x_N, y_N) < \epsilon$ for $N = \frac{\log(2\|A\|/\epsilon)}{\log(\gamma)}$. $\square$

What if $\delta(A)$ is tiny?

Proposition (Gilpin, P, Sandholm, 2010)

Algorithm **iterated** *finishes in at most*

$$\frac{16 \cdot \gamma^2 \cdot \|A\|}{(\gamma - 1) \cdot \epsilon}$$

first-order iterations.

What if $\delta(A)$ is tiny?

Proposition (Gilpin, P, Sandholm, 2010)

Algorithm **iterated** *finishes in at most*

$$\frac{16 \cdot \gamma^2 \cdot \|A\|}{(\gamma - 1) \cdot \epsilon}$$

first-order iterations.

Proof.

1. The i-th call to **smoothing** halts in at most $\frac{16 \cdot \|A\| \cdot \gamma^{i+1}}{\epsilon_0}$ first-order iterations.
2. Algorithm **iterated** halts in at most N outer iterations, where N is such that $\epsilon_0/\gamma^N = \epsilon_N \leq \epsilon < \epsilon_{N-1} = \epsilon_0/\gamma^{N-1}$.

$\square$

Similar result for sequential games

Definition

- A simplex is a treeplex.
- If $Q_1, \ldots, Q_k$ treeplexes then

$$\{(u^0, u^1, \ldots, u^k) : u^0 \in \Delta_k, u^i \in u_i^0 \cdot Q_i, i = 1, \ldots, k\}$$

  is a treeplex.
- If $Q_1, \ldots, Q_k$ treeplexes then $Q_1 \times \cdots \times Q_k$ is a treeplex.

Nash equilibrium for sequential games

$$\max_{x \in Q_1} \min_{y \in Q_2} x^\mathsf{T} A y = \min_{y \in Q_2} \max_{x \in Q_1} x^\mathsf{T} A y,$$

where Q_1, Q_2 treeplexes.

Similar result for sequential games

Modify Algorithm **iterated** in a straightforward way.

Theorem

$$\frac{2\sqrt{2D} \cdot \gamma \cdot \|A\| \cdot \log(2\|A\|/\epsilon)}{\log(\gamma) \cdot \delta(A)}$$

first-order iterations.

Similar result for sequential games

Modify Algorithm **iterated** in a straightforward way.

Theorem

$$\frac{2\sqrt{2D} \cdot \gamma \cdot \|A\| \cdot \log(2\|A\|/\epsilon)}{\log(\gamma) \cdot \delta(A)}$$

first-order iterations.

Caveats

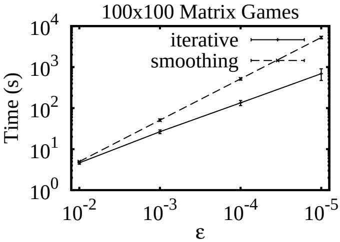
- Dependence on $D := \max\left\{\frac{\|u-\bar{u}\|^2}{2} : u \in Q_1 \times Q_2\right\}$.
- Need to compute the projection

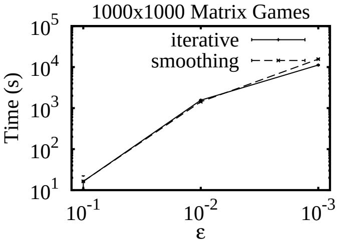
$$\min\left\{\langle g, u\rangle + \frac{\|u\|^2}{2} : u \in Q_1 \times Q_2\right\}$$

easily at each first-order iteration.

Numerical experiments

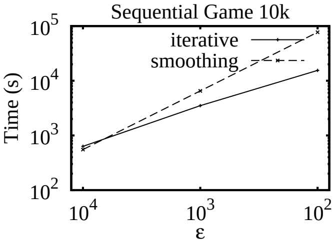
Randomly generated matrix games

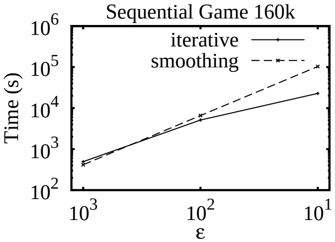




Numerical experiments

Two benchmark instances of sequential games





What do we know about $\delta(A)$?

Recall

$$\delta(A) = \sup \left\{ \delta : \mathrm{dist}((x,y), \mathrm{S}) \leq \frac{F(x,y)}{\delta} \;\; \forall\, (x,y) \in \Delta_m \times \Delta_n \right\},$$

where

$$\mathrm{S} = \mathrm{Argmin}\{F(x,y) : (x,y) \in \Delta_m \times \Delta_n\} = F^{-1}(0) \cap \Delta_m \times \Delta_n.$$

Consider $\kappa(A) = 1/\delta(A)$. Notice

$$\kappa(A) = \inf \left\{ \kappa : \mathrm{dist}((x,y), \mathrm{S}) \leq \kappa \cdot F(x,y) \;\; \forall\, (x,y) \in \Delta_m \times \Delta_n \right\}.$$

Based on convenience, we will use $\delta(A)$ or $\kappa(A) = 1/\delta(A)$.

Metric Regularity

Definition

A set-valued mapping $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is *metrically regular* around $(\bar{x}, \bar{z}) \in \mathrm{gph}(G)$ if there exists μ such that

$$\mathrm{dist}\big(x, G^{-1}(z)\big) \leq \mu \cdot \mathrm{dist}\big(z, G(x)\big) \tag{1}$$

for (x, z) in a neighborhood of $(\bar{x}, \bar{z})$.

Regularity modulus of G at $\bar{x}$ for $\bar{z}$

$$\mathrm{reg}\, G(\bar{x}, \bar{z}) := \text{infimum of } \mu \text{ satisfying (1).}$$

Connection between $\kappa(A)$ and metric regularity

Define $\Phi : \mathbb{R}^{m+n} \rightrightarrows \mathbb{R}$ as

$$\Phi(w) := \begin{cases} \left[F(w), \infty\right) & \text{if } \; w \in \Delta_m \times \Delta_n, \\ \\ \emptyset & \text{otherwise.} \end{cases}$$

Connection between $\kappa(A)$ and metric regularity

Define $\Phi : \mathbb{R}^{m+n} \rightrightarrows \mathbb{R}$ as

$$\Phi(w) := \begin{cases} \left[F(w), \infty\right) & \text{if } \; w \in \Delta_m \times \Delta_n, \\ \emptyset & \text{otherwise.} \end{cases}$$

Theorem (Mordukhovich, P, Roshchina, 2010)

(a) *Assume* $\Delta_m \times \Delta_n \setminus \mathrm{S} \neq \emptyset$. *Then*

$$\kappa(A) = \max_{w \in \Delta_m \times \Delta_n \setminus \mathrm{S}} \operatorname{reg} \Phi\big(w, F(w)\big)$$

(b) *Assume* $w \in \Delta_m \times \Delta_n \setminus \mathrm{S}$. *Then*

$$\operatorname{reg} \Phi\big(w, F(w)\big) = \frac{1}{\operatorname{dist}\big(0, \partial F(w) + N_{\Delta_m \times \Delta_n}(w)\big)}$$

Proof of part (a)

Key technical lemma:

Lemma

Assume $\Delta_m \times \Delta_n \setminus \mathsf{S} \neq \emptyset$*. Then there exists* $\bar{w} \in \Delta_m \times \Delta_n \setminus \mathsf{S}$ *such that*

$$\kappa(A) = \frac{\operatorname{dist}(\bar{w}, \mathsf{S})}{F(\bar{w})}.$$

Proof of part (a)

Key technical lemma:

Lemma

Assume $\Delta_m \times \Delta_n \setminus \mathrm{S} \neq \emptyset$. Then there exists $\bar{w} \in \Delta_m \times \Delta_n \setminus \mathrm{S}$ such that

$$\kappa(A) = \frac{\operatorname{dist}(\bar{w}, \mathrm{S})}{F(\bar{w})}.$$

Part (a) reduces to showing that

$$\kappa(A) = \operatorname{reg}\Phi\big(\bar{w}, F(\bar{w})\big) = \max_{w \in \Delta_m \times \Delta_n \setminus \mathrm{S}} \operatorname{reg}\Phi\big(w, F(w)\big).$$

$\square$

Proof of part (b)

Rely on some tools from variational analysis.

Definition

Given $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$, the *coderivative* of G at $(\bar{x}, \bar{z}) \in \operatorname{gph} G$ is the mapping $D^* G(\bar{x}, \bar{z}) : \mathbb{R}^m \rightrightarrows \mathbb{R}^n$ defined by

$$D^* G(\bar{x}, \bar{z})(v) := \{u : (u, v) \in N_{\operatorname{gph} G}(\bar{x}, \bar{z})\}.$$

Theorem (Mordukhovich, 1984)

Suppose $G : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ has closed graph around $(\bar{x}, \bar{z}) \in gph G$. Then G is metrically regular around $(\bar{x}, \bar{z})$ iff

$$\ker D^* G(\bar{x}, \bar{z}) = \{0\}.$$

In this case

$$\operatorname{reg} G(\bar{x}, \bar{z}) = \|D^* G(\bar{x}, \bar{z})^{-1}\|.$$

Proof of part (b)

Proposition (Dontchev, Lewis, Rockafellar, 2003)

Assume $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ is positively homogeneous. Then

$$\|M^{-1}\| = \sup_{\|v\|=1} \frac{1}{\mathrm{dist}(0, Mv)}.$$

Proof of part (b)

Proposition (Dontchev, Lewis, Rockafellar, 2003)

Assume $M : \mathbb{R}^n \rightrightarrows \mathbb{R}^m$ *is positively homogeneous. Then*

$$\|M^{-1}\| = \sup_{\|v\|=1} \frac{1}{\mathrm{dist}(0, Mv)}.$$

To prove (b), apply the above results to Φ:

$$\begin{aligned} \operatorname{reg} \Phi(w, F(w)) &= \|D^*\Phi(w, F(w))^{-1}\| \\ &= \sup_{|v|=1} \frac{1}{\mathrm{dist}\,(0, D^*\Phi(w, F(w))(v))} \\ &= \frac{1}{\mathrm{dist}(0, D^*\Phi(w, F(w))(1))}. \end{aligned}$$

To finish, compute

$$D^*\Phi(w, F(w))(1) = \partial F(w) + N_{\Delta_m \times \Delta_n}(w).$$

Characterization of $\delta(A)$

A bit more notation

Let $A = \begin{bmatrix} a_1 & \cdots & a_n \end{bmatrix} = \begin{bmatrix} -b_1^\mathsf{T} \\ \vdots \\ -b_m^\mathsf{T} \end{bmatrix}$.

For $p \in \mathbb{Z}_+$, let $\mathbf{1}_p := \begin{bmatrix} 1 & \ldots & 1 \end{bmatrix}^\mathsf{T} \in \mathbb{R}^p$.

Given $(x, y) \in \Delta_m \times \Delta_n$, let

$$I(x) := \left\{ \bar{\imath} \in \{1, \ldots, n\} : \; a_{\bar{\imath}}^\mathsf{T} x = \max_{i \in \{1,\ldots,n\}} a_i^\mathsf{T} x \right\},$$
$$K(y) := \left\{ \bar{k} \in \{1, \ldots, m\} : \; b_{\bar{k}}^\mathsf{T} y = \max_{k \in \{1,\ldots,m\}} b_k^\mathsf{T} y \right\},$$
$$J(x, y) := \left\{ j \in \{1, \ldots, m\} : \; x_j = 0 \right\} \bigcup \left\{ j = m + p : \; y_p = 0 \right\}.$$

Characterization of $\delta(A)$

Theorem (Mordukhovich, P, Roshchina, 2010)

Assume $\Delta_m \times \Delta_n \setminus \mathrm{S} \neq \emptyset$. *Then*

$$\delta(A) = \min_{(x,y)\in\Delta_m\times\Delta_n\setminus \mathrm{S}} \Big[\operatorname{dist}\Big(0, \operatorname{conv}\big\{(a_i, b_k) : \ i \in I(x),\ k \in K(y)\big\} \\ + \operatorname{span}\{\mathbf{1}_m\} \times \operatorname{span}\{\mathbf{1}_n\} - \operatorname{coco}\big\{e_j : \ j \in J(x,y)\big\}\Big)\Big].$$

Characterization of $\delta(A)$

Theorem (Mordukhovich, P, Roshchina, 2010)

Assume $\Delta_m \times \Delta_n \setminus \mathrm{S} \neq \emptyset$. *Then*

$$\delta(A) = \min_{(x,y)\in\Delta_m\times\Delta_n\setminus \mathrm{S}} \Big[\operatorname{dist}\Big(0, \operatorname{conv}\big\{(a_i, b_k) : \ i \in I(x), \ k \in K(y)\big\} + \operatorname{span}\{\mathbf{1}_m\} \times \operatorname{span}\{\mathbf{1}_n\} - \operatorname{coco}\big\{e_j : \ j \in J(x,y)\big\}\Big)\Big].$$

Recall previous theorem:

(a) If $\Delta_m \times \Delta_n \setminus \mathrm{S} \neq \emptyset$ then $\kappa(A) = \max_{w\in(\Delta_m\times\Delta_n)\setminus \mathrm{S}} \operatorname{reg}\Phi\big(w, F(w)\big)$.

(b) If $w \in \Delta_m \times \Delta_n \setminus \mathrm{S}$ then

$$\operatorname{reg}\Phi\big(w, F(w)\big) = \frac{1}{\operatorname{dist}\big(0, \partial F(w) + N_{\Delta_m\times\Delta_n}(w)\big)}.$$

Characterization of $\delta(A)$

Proof.

Put $w := (x, y)$. From previous theorem we get

$$\delta(A) = \min_{w \in (\Delta_m \times \Delta_n) \setminus S} \frac{1}{\operatorname{reg} \Phi\big(w, F(w)\big)}$$
$$= \min_{w \in (\Delta_m \times \Delta_n) \setminus S} \operatorname{dist}\big(0, \partial F(w) + N_{\Delta_m \times \Delta_n}(w)\big).$$

To finish, use elementary non-smooth calculus to compute

$$\partial F(x, y) = \operatorname{conv}\big\{(a_i, b_k) : \; i \in I(x), \; k \in K(y)\big\},$$

and

$$N_{\Delta_m \times \Delta_n}(w) = \operatorname{span}\{\mathbf{1}_m\} \times \operatorname{span}\{\mathbf{1}_n\} - \operatorname{coco}\big\{e_j : \; j \in J(x, y)\big\}.$$

$\square$

Concluding remarks

- Algorithm that finds ϵ-solution to

$$\max_{x \in \Delta_m} \min_{y \in \Delta_n} x^{\mathsf{T}} A y = \min_{y \in \Delta_n} \max_{x \in \Delta_m} x^{\mathsf{T}} A y$$

  in $\mathcal{O}(\kappa(A) \cdot \log(1/\epsilon))$ first-order iterations

- Connection between condition measure $\kappa(A)$ and metric regularity.

- Similar results for more general equilibrium of sequential games

$$\max_{x \in Q_1} \min_{y \in Q_2} x^{\mathsf{T}} A y = \min_{y \in Q_2} \max_{x \in Q_1} x^{\mathsf{T}} A y.$$

  In this case Q_1, Q_2 are *treeplexes*.

Current & future work

- Reliance on Euclidean distance function and on saddle-point problem over polytopes. Do similar results hold for other prox-functions and/or other problems?
- Classes of well-conditioned problems.
- Average case analysis of $\kappa(A)$.
- Connection with other measures of conditioning.

References

- A. Gilpin, J. Peña, and T. Sandholm, "First-order algorithm with $O(\log(1/\epsilon))$ convergence for ϵ-equilibrium in two-person zero-sum games," To Appear in *Mathematical Programming.*
- B. Mordukhovich, J. Peña, and V. Roshchina, "Applying Metric Regularity to Compute a Condition Measure of a Smoothing Algorithms for Matrix Games," To Appear in *SIAM Journal on Optimization.*