Subgame Perfect Algorithms for Smooth Convex Optimization

Alex L. Wang, Daniels School of Business, Purdue University

Based on joint work with: **Benjamin Grimmer**, **Kevin Shu**

PURDUE UNIVERSITY

Mitchell E. Daniels, Jr. School of Business

Talk Outline

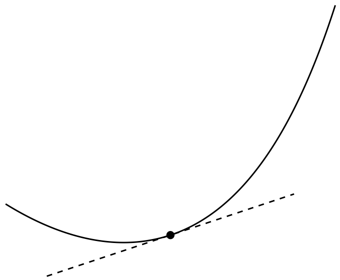
- Game theoretic view of smooth convex optimization
- Optimized Gradient Method **(OGM)**
  - Algorithm structure, guarantees
  - Maximin optimal strategy
- Subgame Perfect Gradient Method **(SPGM)**
  - Subgame perfect strategy

[Drori Teboulle 12] [Kim Fessler 16] [Grimmer Shu Wang 24]

Game Theoretic View of Smooth Convex Optimization

Smooth Convex Optimization

- **Task**: Solve $\displaystyle \min_{x\in\mathbb{R}^d} f(x)$
- A priori, we know
  - f is 1-smooth: $\|\nabla f(x) - \nabla f(y)\| \leq \|x-y\| \quad \forall x, y$
  - f is convex
  - f has a minimizer
- Let $\mathcal{F}$ be the set of all instances



Algorithms for Smooth Convex Optimization

- Black-box model of optimization $f \in \mathcal{F}$
- First-order oracle: $x \mapsto (f(x), \nabla f(x))$

Algorithm. N-query first-order method

- For $n = 0, \ldots, N$:
  - **Choose** $x_n \in \operatorname{span}\{g_0, \ldots, g_{n-1}\}$
  - **Query first-order oracle**

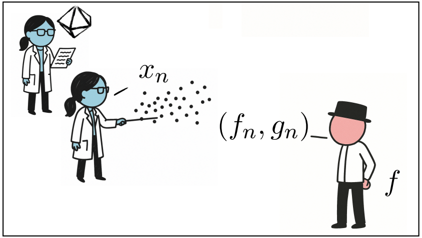
$$f_n = f(x_n) \qquad g_n = \nabla f(x_n)$$

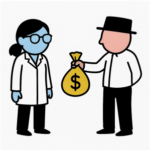
- Output x_N

The Convex Optimization Game v1

- **Alice** (algorithm), **Bob** (oracle), N number of rounds
- **Bob** secretly chooses $f \in \mathcal{F}$ with $(x_\star, f_\star)$
- Round $n = 0, \ldots, N$
  - **Alice** chooses $x_n \in \operatorname{span}\{g_0, \ldots, g_{n-1}\}$
  - **Bob** returns $f_n = f(x_n)$ and $g_n = \nabla f(x_n)$
- **Bob** pays **Alice** $\frac{\|x_0 - x_\star\|^2}{f_N - f_\star}$

e.g., payoff = $1000 means
$f(x_N) - f_\star = \frac{1}{1000}\|x_0 - x_\star\|^2$



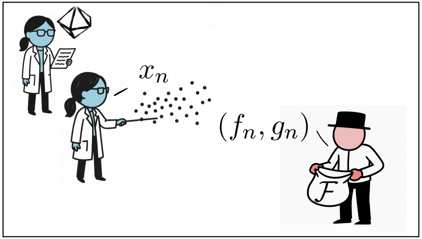


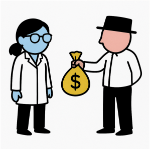
The convex optimization game v2

- **Alice** (algorithm), **Bob** (oracle), N number of rounds
- Round $n = 0, \ldots, N$
  - **Alice** chooses $x_n \in \operatorname{span}\{g_0, \ldots, g_{n-1}\}$
  - **Bob** returns $(f_n, g_n) \in \mathbb{R} \times \mathbb{R}^d$
- **Bob** specifies $f \in \mathcal{F}$ with $(x_\star, f_\star)$

$$f(x_i) = f_i \quad \text{and} \quad \nabla f(x_i) = g_i \quad \text{for all } i \in [0, N]$$

- **Bob** pays **Alice** $\frac{\|x_0 - x_\star\|^2}{f_N - f_\star}$





Classic (textbook) examples

- **Gradient descent**: $x_n = \boxed{x_{n-1}^+ := x_{n-1} - g_{n-1}}$

  For all $f \in \mathcal{F}$,

  $$f(x_N) - f(x_\star) \leq \frac{1}{4N+2} \|x_0 - x_\star\|^2$$

  Alice can guarantee payoff $\geq 4N + 2$

- **Fast Gradient Method**: $x_n = x_{n-1}^+ +$ (momentum)

  For all $f \in \mathcal{F}$,

  $$f(x_N) - f(x_\star) \leq \frac{2}{(N+1)^2} \|x_0 - x_\star\|^2$$

  Alice can guarantee payoff $\geq (N+1)^2/2$

[Nesterov 83]

The Optimized Gradient Method

[Drori Teboulle 12], [Kim, Fessler 16]

OGM initialization

- OGM maintains a pair of points in each iteration $(x_0, z_0), \ldots, (x_N, z_N) \in \mathbb{R}^d \times \mathbb{R}^d$
- Performance stated in terms of $\tau_0, \ldots, \tau_N \in \mathbb{R}$

Algorithm. OGM

- Initialization
  - **Define** $x_0 = 0$ and $\tau_0 = 2$
  - **Query** (f_0, g_0)
  - **Define** $z_0 = x_0 - 2g_0$
- For $n = 1, \ldots, N$
  - $\ldots$

Lemma.

OGM's initialization satisfies

$$2\tau_0(f(x_0^+) - f_\star) + \|z_0 - x_\star\|^2 \leq \|x_0 - x_\star\|^2$$

i.e., either $f(x_0^+) \approx f_\star$ or $z_0 \approx x_\star$

[Kim Fessler 16]

OGM induction

Lemma.

Suppose (x, z, τ) and $f \in \mathcal{F}$ satisfy

$$2\tau(f(x^+) - f_\star) + \|z - x_\star\|^2 \leq \|x_0 - x_\star\|^2$$

Then,

$$\hat{x} = \text{some convex combination of } x^+ \text{ and } z$$

$$(\hat{\tau}, \hat{\delta}) = \ldots$$

satisfy

$$2\hat{\tau}(f(\hat{x}^+) - f_\star) + \|\underbrace{z - \hat{\delta}\nabla f(\hat{x})}_{\hat{z}} - x_\star\|^2 \leq \|x_0 - x_\star\|^2$$

Algorithm. Optimized Gradient Method

- **Define** $x_0 = 0$, $\tau_0 = 2$, **query** (f_0, g_0), **define** $z_0 = x_0 - 2g_0$:

$$2\tau_0(f(x_0^+) - f_\star) + \|z_0 - x_\star\|^2 \leq \|x_0 - x_\star\|^2$$

- For $n = 1, \dots, N-1$
  - **Apply Lemma** to $(x_{n-1}, z_{n-1}, \tau_{n-1})$ to get (x_n, τ_n, δ_n)
  - **Reveal** (f_n, g_n)
  - **Set** $z_n = z_{n-1} - \delta_n g_n$

$$2\tau_n(f(x_n^+) - f_\star) + \|z_n - x_\star\|^2 \leq \|x_0 - x_\star\|^2$$

- Slight modification for iteration N ...
- Output x_N

[Kim Fessler 16]

Optimality of OGM

Theorem.

For any $f \in \mathcal{F}$, the output x_N of OGM satisfies

$$f(x_N) - f_\star \leq \frac{1}{2\tau_N} \|x_0 - x_\star\|^2$$

Using the OGM strategy, Alice can guarantee payoff $\geq 2\tau_N$

Theorem.

If $d \geq N+2$, there exists $f \in \mathcal{F}$ so that the output x_N for any FOM satisfies

$$f(x_N) - f_\star \geq \frac{1}{2\tau_N} \|x_0 - x_\star\|^2$$

No strategy for Alice can guarantee payoff $> 2\tau_N$

- $2\tau_N \approx N^2$ is faster than Nesterov's FGM by ≈ 2
- OGM is maximin optimal!

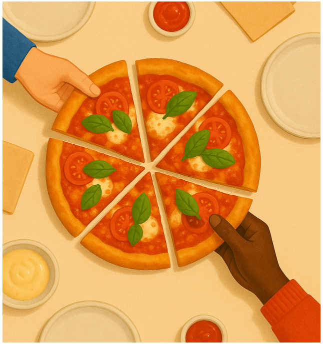
$$\max_{\text{Alice's strategy}} \; \min_{\text{Bob's strategy}} \left(\text{Alice's payoff}\right)$$

[Kim Fessler 16] [Drori 17]

Maximin Optimal Strategies vs. Subgame Perfect Strategies

The Seminar Pizza Line Game

- **Bob** and **Alice** are in line to grab pizza
- There are 6 slices remaining
- **Bob** goes first and may take $b \leq 3$ slices
- **Alice** goes second and may take $a \leq 6 - b$ slices
- **Alice**'s payoff $= a$ and **Bob**'s payoff $= b$
- **Silly strategy**: **Alice** takes $a = 3$ slices
  - This strategy guarantees $a \geq 3$
  - No strategy for **Alice** can guarantee $a > 3$
- The silly strategy is maximin optimal



Subgame Perfect Strategies

- In a sequential game, should demand a subgame perfect strategy
- A subgame is the remainder of game after some actions have been fixed
- Can identify subgame withwith sequence of committed actions
- In the seminar pizza line game:
  - $\{\}$
  - $\{b = 0\}, \ldots, \{b = 3\}$
- A strategy for Alice is **subgame perfect** if for every subgame:

$$\text{Alice's payoff in subgame} = \max_{\text{Alice's remaining actions}} (\text{Alice's payoff in subgame})$$

- The only subgame perfect strategy for Alice is $a = 6 - b$

OGM is *NOT* Subgame Perfect

- Consider $f(x) = (x-1)^2/2$
- This is a worst-case function for OGM:

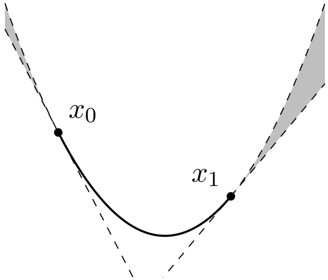
$$f(x_N) - f_\star = \frac{1}{2\tau_N} \|x_0 - x_\star\|^2$$

- At iteration $n = 2$, **Alice** is in the subgame

$$\begin{aligned}\{&(x_0, f_0, g_0) = (0, 1/2, -1),\\ &(x_1, f_1, g_1) = (1.618, 0.191, 0.618)\}\end{aligned}$$

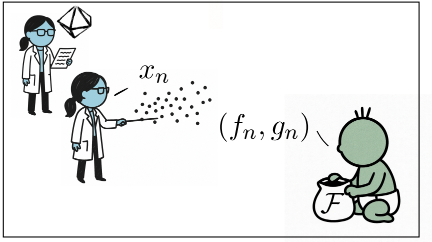
- In this subgame, it is possible to guarantee an infinite payoff:

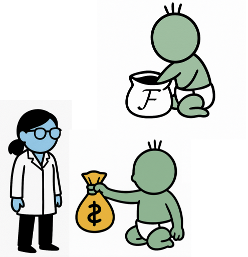
$$x_N = x_{N-1} = \cdots = x_2 = x_1^+$$



Why Does Subgame Perfect Matter?

- Traditional maximin view assumes an adversarial oracle
- In reality, function is *fixed* and there is no adversary
- We want to exploit any suboptimal play by oracle, while guarding against possible adversarial play (by chance)
- Numerical evidence





Subgame Perfect Gradient Method

[Grimmer Shu Wang 24]

The Subgame Perfect Gradient Method

Algorithm. Subgame Perfect Gradient Method

- **Define** $x_0 = 0$, $\tau_0 = 2$, **query** (f_0, g_0), **define** $z_0 = x_0 - 2g_0$:

$$2\tau_0(f(x_0^+) - f_\star) + \|z_0 - x_\star\|^2 \leq \|x_0 - x_\star\|^2$$

- For $n = 1, \ldots, N-1$
  - ♠ **Solve a subproblem** to get $(\tilde{x}, \tilde{z}, \tilde{\tau})$ so that

$$2\tilde{\tau}(f(\tilde{x}^+) - f_\star) + \|\tilde{z} - x_\star\|^2 \leq \|x_0 - x_\star\|^2$$

  - **Apply Lemma** to $(\tilde{x}, \tilde{z}, \tilde{\tau})$ to get (x_n, τ_n, δ_n)
  - **Reveal** (f_n, g_n)
  - **Set** $z_n = z_{n-1} - \delta_n g_n$

$$2\tau_n(f(x_n^+) - f_\star) + \|z_n - x_\star\|^2 \leq \|x_0 - x_\star\|^2$$

- Slight modification for iteration N, Output x_N

SPGM Dynamically Reoptimizes Inductive Hypothesis

- At round n, we have observed

$$\mathcal{H} = \{(x_0, f_0, g_0), \ldots, (x_{n-1}, f_{n-1}, g_{n-1})\}$$

  and know

$$f \in \mathcal{F}_{\mathcal{H}} := \left\{ f \in \mathcal{F} : \begin{array}{ll} f(x_i) = f_i & \forall i \in [0, n-1] \\ \nabla f(x_i) = g_i & \forall i \in [0, n-1] \end{array} \right\}$$

- **Goal**: find $(\tilde{x}, \tilde{z}, \tilde{\tau})$ to solve

$$\max_{\tilde{x}, \tilde{z} \in \mathbb{R}^d,\, \tilde{\tau} \in \mathbb{R}} \left\{ \tilde{\tau} : \quad 2\tilde{\tau}(f(\tilde{x}^+) - f_\star) + \|\tilde{z} - x_\star\|^2 \leq \|x_0 - x_\star\|^2 \quad \forall f \in \mathcal{F}_{\mathcal{H}} \right\}$$

- Nonconvex, nonsmooth, infinite dimensional problem
- After reformulations, this becomes a second-order cone program (dual strictly feasible, strong duality, primal bounded)

Optimality of SPGM

Theorem.

Let $0 \leq n \leq N$ and suppose SPGM is in a subgame

$$\mathcal{H} = \{(x_0, f_0, g_0), \ldots, (x_{n-1}, f_{n-1}, g_{n-1})\}$$

The output x_N of $N - n + 1$ additional iterations of SPGM satisfies

$$f(x_N) - f_\star \leq \frac{1}{2\tau_N(\mathcal{H})} \|x_0 - x_\star\|^2$$

Theorem.

Let $0 \leq n \leq N$ and suppose $d \geq N + 2$.
Consider a subgame
$\mathcal{H} = \{(x_0, f_0, g_0), \ldots, (x_{n-1}, f_{n-1}, g_{n-1})\}$
encountered by SPGM.
There exists $f \in \mathcal{F}_\mathcal{H}$ so that the output of any FOM for the subgame satisfies

$$f(x_N) - f_\star \geq \frac{1}{2\tau_N(\mathcal{H})} \|x_0 - x_\star\|^2$$

- SPGM is subgame perfect!

[Grimmer Shu Wang 24]

Numerical Results

Experiments

- Regularized regression problems

$$f(x) = \frac{1}{m} \|Ax - b\|_2^2 + \sum_{i=1}^{d} h_\alpha(|x_i|) \approx \frac{1}{m} \|Ax - b\|_2^2 + \alpha \|x\|_1$$

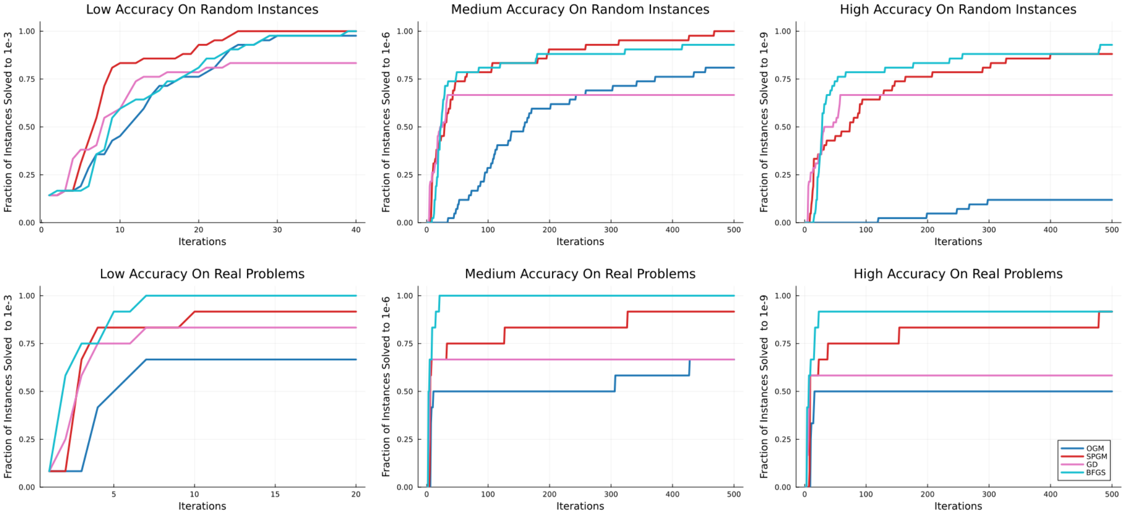
- Regularized logistic regression problems

$$f(x) = \frac{1}{m} \sum_{i=1}^{m} \log(1 + \exp(b_i \cdot a_i^\intercal x)) + \frac{1}{2m} \|x\|_2^2$$

- Combination of simulated data and real data (LIBSVM)

Results I

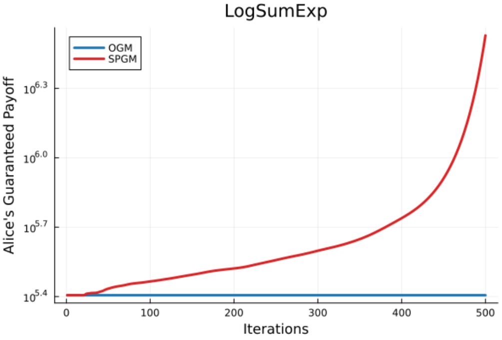
Fraction of instances solved: $f(x_n) - f_\star \leq \{10^{-3}, 10^{-6}, 10^{-9}\} \cdot \frac{1}{2}\|x_0 - x_\star\|^2$



Results II

- Value of $\tau_N(\mathcal{H})$ over course of algorithm
- Alice's payoff is $\geq 2\tau_N(\mathcal{H})$ where

$$\mathcal{H} = \{(x_0, f_0, g_0), \ldots, (x_{n-1}, f_{n-1}, g_{n-1})\}$$



Summary

Summary + Pointers

- Optimized Gradient Method **(OGM)** is maximin optimal
- Subgame perfect vs. Maximin optimal
- Subgame Perfect Gradient Method **(SPGM)** is subgame perfect
- See paper: **limited-memory variants of SPGM**
- Upcoming work:
  - Subgame perfect methods for convex **nonsmooth** optimization **[Grimmer Shu Wang** forthcoming**]**
  - **Adaptive** variants of SPGM **[Luner Grimmer** forthcoming**]**

`https://arxiv.org/abs/2412.06731`

Thank you for listening. Questions?

References I

Drori, Y. (2017). The exact information-based complexity of smooth convex minimization. *Journal of Complexity*, 39:1–16.

Drori, Y. and Teboulle, M. (2012). Performance of first-order methods for smooth convex minimization: a novel approach. *Mathematical Programming*, 145:451–482.

Grimmer, B., Shu, K., and Wang, A. L. (2024). Beyond minimax optimality: A subgame perfect gradient method. *arXiv preprint arXiv:2412.06731*.

Kim, D. and Fessler, J. A. (2016). Optimized first-order methods for smooth convex minimization. *Math. Program.*, 159(1–2):81–107.

Nesterov, Y. (1983). A method for solving the convex programming problem with convergence rate $O(1/k^2)$. *Proceedings of the USSR Academy of Sciences*, 269:543–547.