## The Complexity of Gradient Descent: CLS = PPAD $\cap$ PLS

## ALEXANDROS HOLLENDER

joint work with JOHN FEARNLEY, PAUL GOLDBERG and RAHUL SAVANI

## Some interesting computational problems

## Some interesting computational problems

## NASH:

Find a mixed Nash equilibrium of a game.

## Some interesting computational problems

## NASH:

Find a mixed Nash equilibrium of a game.
FACTORING:
Find a prime factor of a number $n \geq 2$.

## Some interesting computational problems

## NASH:

Find a mixed Nash equilibrium of a game.
FACTORING:
Find a prime factor of a number $n \geq 2$.

## BROUWER:

Find a fixpoint of a continuous function $f:[0,1]^{3} \rightarrow[0,1]^{3}$.

## Some interesting computational problems

## NASH:

Find a mixed Nash equilibrium of a game.
FACTORING:
Find a prime factor of a number $n \geq 2$.

## BROUWER:

Find a fixpoint of a continuous function $f:[0,1]^{3} \rightarrow[0,1]^{3}$.
CONTRACTION:
Find the unique fixpoint of a contraction $f:[0,1]^{n} \rightarrow[0,1]^{n}$.

## Some interesting computational problems

## NASH:

Find a mixed Nash equilibrium of a game.
FACTORING:
Find a prime factor of a number $n \geq 2$.

## BROUWER:

Find a fixpoint of a continuous function $f:[0,1]^{3} \rightarrow[0,1]^{3}$.
CONTRACTION:
Find the unique fixpoint of a contraction $f:[0,1]^{n} \rightarrow[0,1]^{n}$.

## PURE-CONGESTION:

Find a pure Nash equilibrium of a congestion game.

## Some interesting computational problems

## NASH:

Find a mixed Nash equilibrium of a game.
What do these problems have in common?
2
FACTORING:
Find a prime factor of a number $n \geq 2$.

## BROUWER:

Find a fixpoint of a continuous function $f:[0,1]^{3} \rightarrow[0,1]^{3}$.
CONTRACTION:
Find the unique fixpoint of a contraction $f:[0,1]^{n} \rightarrow[0,1]^{n}$.

## PURE-CONGESTION:

Find a pure Nash equilibrium of a congestion game.

## Some interesting computational problems

## NASH:

Find a mixed Nash equilibrium of a game.

## FACTORING:

Find a prime factor of a number $n \geq 2$.

## BROUWER:

Find a fixpoint of a continuous function $f:[0,1]^{3} \rightarrow[0,1]^{3}$.

## CONTRACTION:

Find the unique fixpoint of a contraction $f:[0,1]^{n} \rightarrow[0,1]^{n}$.

## PURE-CONGESTION:

Find a pure Nash equilibrium of a congestion game.

## Some interesting computational problems

## NASH:

Find a mixed Nash equilibrium of a game.

## FACTORING:

Find a prime factor of a number $n \geq 2$.

## BROUWER:

Find a fixpoint of a continuous function $f:[0,1]^{3} \rightarrow[0,1]^{3}$. CONTRACTION:
Find the unique fixpoint of a contraction $f:[0,1]^{n} \rightarrow[0,1]^{n}$.

What do these problems have in common?
$\longrightarrow$
They are NP Total Search (TFNP) problems!

- Total: there is always a solution
- NP: it is easy to verify solutions



## PURE-CONGESTION:

Find a pure Nash equilibrium of a congestion game.

## Some interesting computational problems

## NASH:

Find a mixed Nash equilibrium of a game.

## FACTORING:

Find a prime factor of a number $n \geq 2$.

## BROUWER:

Find a fixpoint of a continuous function $f:[0,1]^{3} \rightarrow[0,1]^{3}$.

## CONTRACTION:

Find the unique fixpoint of a contraction $f:[0,1]^{n} \rightarrow[0,1]^{n}$.

## PURE-CONGESTION:

Find a pure Nash equilibrium of a congestion game.

What do these problems have in common?
$\longrightarrow$
They are NP Total Search (TFNP) problems!

- Total: there is always a solution
- NP: it is easy to verify solutions


Not unless co-NP = NP...

## The class TFNP [Megiddo-Papadimitriou, 1991]

Total NP search problems:

- "search" : looking for a solution, not just YES or NO
- "NP": any solution can be checked efficiently
- "total": there always exists at least one solution


## The class TFNP [Megiddo-Papadimitriou, 1991]

Total NP search problems:

- "search" : looking for a solution, not just YES or NO
- "NP": any solution can be checked efficiently
- "total": there always exists at least one solution

TFNP lies between P and NP (search versions)

## The class TFNP [Megiddo-Papadimitriou, 1991]

Total NP search problems:

- "search" : looking for a solution, not just YES or NO
- "NP": any solution can be checked efficiently
- "total": there always exists at least one solution

How do we show that a TFNP-problem is hard:

## The class TFNP [Megiddo-Papadimitriou, 1991]

## Total NP search problems:

- "search" : looking for a solution, not just YES or NO
- "NP": any solution can be checked efficiently
- "total": there always exists at least one solution

How do we show that a TFNP-problem is hard:

- No TFNP-problem can be NP-hard, unless NP = coNP...


## The class TFNP [Megiddo-Papadimitriou, 1991]

## Total NP search problems:

- "search" : looking for a solution, not just YES or NO
- "NP": any solution can be checked efficiently
- "total": there always exists at least one solution

How do we show that a TFNP-problem is hard:

- No TFNP-problem can be NP-hard, unless NP = coNP...

3-SAT $\leq$ NASH $\quad \Rightarrow$ certificate for unsatisfiable 3-SAT formulas

## The class TFNP [Megiddo-Papadimitriou, 1991]

## Total NP search problems:

- "search" : looking for a solution, not just YES or NO
- "NP": any solution can be checked efficiently
- "total": there always exists at least one solution

How do we show that a TFNP-problem is hard:

- No TFNP-problem can be NP-hard, unless NP = coNP...


## The class TFNP [Megiddo-Papadimitriou, 1991]

## Total NP search problems:

- "search" : looking for a solution, not just YES or NO
- "NP": any solution can be checked efficiently
- "total": there always exists at least one solution

How do we show that a TFNP-problem is hard:

- No TFNP-problem can be NP-hard, unless NP = coNP...
- Believed that no TFNP-complete problems exists...


## The TFNP landscape



## The TFNP landscape



## The TFNP landscape



## The TFNP landscape



## The TFNP landscape



## The TFNP landscape



## TFNP subclasses

What reasons are there to believe that PPAD $\neq P, P L S \neq P$, etc?

## TFNP subclasses

What reasons are there to believe that PPAD $\neq P, P L S \neq P$, etc?

- many seemingly hard problems lie in PPAD, PLS etc...


## TFNP subclasses

What reasons are there to believe that PPAD $\neq P, P L S \neq P$, etc?

- many seemingly hard problems lie in PPAD, PLS etc...
- oracle separations between the classes (in particular PPAD $=$ PLS)


## TFNP subclasses

What reasons are there to believe that PPAD $\neq P, P L S \neq P$, etc?

- many seemingly hard problems lie in PPAD, PLS etc...
- oracle separations between the classes (in particular PPAD $=$ PLS)
- hard under cryptographic assumptions









## PPAD $\cap$ PLS seems unnatural...

## PPAD $\cap$ PLS seems unnatural...

Problem $A$ : PPAD-complete
Problem B: PLS-complete

## PPAD $\cap$ PLS seems unnatural...

Problem $A$ : PPAD-complete
Problem $B$ : PLS-complete

EITHER-SOLUTION $(A, B)$ :
Input: instance $I_{A}$ of $A$, instance $I_{B}$ of $B$
Goal: find a solution of $I_{A}$, or a solution of $I_{B}$

## PPAD $\cap$ PLS seems unnatural...

Problem $A$ : PPAD-complete
Problem B: PLS-complete

EITHER-SOLUTION $(A, B)$ :
Input: instance $I_{A}$ of $A$, instance $I_{B}$ of $B$
Goal: find a solution of $I_{A}$, or a solution of $I_{B}$
$\rightarrow$ EITHER-SOLUTION $(A, B)$ is (PPAD $\cap$ PLS)-complete!

## PPAD $\cap$ PLS seems unnatural...

## BROUWER:

Input: a continuous function $f:[0,1]^{n} \rightarrow[0,1]^{n}$
Goal: find a fixpoint $x$

$$
f(x)=x
$$

## PPAD $\cap$ PLS seems unnatural...

## BROUWER:

Input: a continuous function $f:[0,1]^{n} \rightarrow[0,1]^{n}$, precision $\varepsilon>0$
Goal: find an approximate fixpoint $x$

$$
\|f(x)-x\| \leq \varepsilon
$$

## PPAD $\cap$ PLS seems unnatural...

## BROUWER:

Input: a continuous function $f:[0,1]^{n} \rightarrow[0,1]^{n}$, precision $\varepsilon>0$
Goal: find an approximate fixpoint $x$

$$
\|f(x)-x\| \leq \varepsilon
$$

## REAL-LOCAL-OPT:

Input:

- a continuous function $p:[0,1]^{n} \rightarrow[0,1]$
- a (possibly non-continuous) function $g:[0,1]^{n} \rightarrow[0,1]^{n}$


## PPAD $\cap$ PLS seems unnatural...

## BROUWER:

Input: a continuous function $f:[0,1]^{n} \rightarrow[0,1]^{n}$, precision $\varepsilon>0$
Goal: find an approximate fixpoint $x$

$$
\|f(x)-x\| \leq \varepsilon
$$

## REAL-LOCAL-OPT:

Input:

- a continuous function $p:[0,1]^{n} \rightarrow[0,1]$
- a (possibly non-continuous) function $g:[0,1]^{n} \rightarrow[0,1]^{n}$

Goal: find a local minimum of $p$ with respect to $g$

$$
p(g(x)) \geq p(x)
$$

## PPAD $\cap$ PLS seems unnatural...

## BROUWER:

Input: a continuous function $f:[0,1]^{n} \rightarrow[0,1]^{n}$, precision $\varepsilon>0$
Goal: find an approximate fixpoint $x$

$$
\|f(x)-x\| \leq \varepsilon
$$

## REAL-LOCAL-OPT:

Input:

- a continuous function $p:[0,1]^{n} \rightarrow[0,1]$
- a (possibly non-continuous) function $g:[0,1]^{n} \rightarrow[0,1]^{n}$ Goal: find a local minimum of $p$ with respect to $g$

$$
p(g(x)) \geq p(x)-\varepsilon
$$

## $P P A D \cap P L S$ seems unnatural...

## BROUWER:

Input: a continuous function $f:[0,1]^{n} \rightarrow[0,1]^{n}$, precision $\varepsilon>0$
Goal: find an approximate fixpoint $x$

$$
\|f(x)-x\| \leq \varepsilon
$$

## REAL-LOCAL-OPT:

Input:

- a continuous function $p:[0,1]^{n} \rightarrow[0,1]$
- a (possibly non-continuous) function $g:[0,1]^{n} \rightarrow[0,1]^{n}$

Goal: find a local minimum of $p$ with respect to $g$

$$
p(g(x)) \geq p(x)-\varepsilon
$$

$\rightarrow$ EITHER-SOLUTION(BROUWER,LOCAL-OPT) is (PPAD $\cap$ PLS)-complete.

## Continuous Local Search

But EITHER-SOLUTION(BROUWER,LOCAL-OPT) is not very natural...

## Continuous Local Search

But EITHER-SOLUTION(BROUWER,LOCAL-OPT) is not very natural...

## CONTINUOUS-LOCAL-OPT:

Input: continuous functions $g:[0,1]^{n} \rightarrow[0,1]^{n}$ and $p:[0,1]^{n} \rightarrow[0,1]$
Goal: find $x$ such that

$$
p(g(x)) \geq p(x)-\varepsilon
$$

## Continuous Local Search

But EITHER-SOLUTION(BROUWER,LOCAL-OPT) is not very natural...

## CONTINUOUS-LOCAL-OPT:

Input: continuous functions $g:[0,1]^{n} \rightarrow[0,1]^{n}$ and $p:[0,1]^{n} \rightarrow[0,1]$
Goal: find $x$ such that

$$
p(g(x)) \geq p(x)-\varepsilon
$$

$\rightarrow$ class Continuous Local Search (CLS) [Daskalakis-Papadimitriou, 2011]

## PPAD $\cap$ PLS

EITHER-SOLUTION $(A, B)$


## PPAD $\cap$ PLS



## PPAD $\cap$ PLS



## Motivation behind the classes

## Motivation behind the classes

PPAD: "all problems that can be solved by a path-following algorithm" (Lemke-Howson algorithm for NASH)

## Motivation behind the classes

PPAD: "all problems that can be solved by a path-following algorithm" (Lemke-Howson algorithm for NASH)

PLS: "all problems that can be solved by a local search algorithm"

## Motivation behind the classes

PPAD: "all problems that can be solved by a path-following algorithm" (Lemke-Howson algorithm for NASH)

PLS: "all problems that can be solved by a local search algorithm"
CLS: "all problems that can be solved by a continuous local search algorithm"

## Motivation behind the classes

PPAD: "all problems that can be solved by a path-following algorithm" (Lemke-Howson algorithm for NASH)

PLS: "all problems that can be solved by a local search algorithm"
CLS: "all problems that can be solved by a continuous local search algorithm"

GD: "all problems that can be solved by gradient descent"

## Gradient Descent Problems

GD: "all problems that can be solved by gradient descent"

## Gradient Descent Problems

GD: "all problems that can be solved by gradient descent"
Input: $C^{1}$-function $f:[0,1]^{n} \rightarrow[0,1]$, step size $\eta>0$, precision $\varepsilon>0$

## Gradient Descent Problems

GD: "all problems that can be solved by gradient descent"
Input: $C^{1}$-function $f:[0,1]^{n} \rightarrow[0,1]$, step size $\eta>0$, precision $\varepsilon>0$

$$
x_{k+1} \longleftarrow x_{k}-\eta \nabla f\left(x_{k}\right)
$$

## Gradient Descent Problems

GD: "all problems that can be solved by gradient descent"
Input: $C^{1}$-function $f:[0,1]^{n} \rightarrow[0,1]$, step size $\eta>0$, precision $\varepsilon>0$

$$
x_{k+1} \leftarrow x_{k}-\eta \nabla f\left(x_{k}\right)
$$

Goal: find a point where gradient descent terminates

## Gradient Descent Problems

GD: "all problems that can be solved by gradient descent"
Input: $C^{1}$-function $f:[0,1]^{n} \rightarrow[0,1]$, step size $\eta>0$, precision $\varepsilon>0$

$$
x_{k+1} \leftarrow x_{k}-\eta \nabla f\left(x_{k}\right)
$$

Goal: find a point where gradient descent terminates

$$
\left[x^{\prime}:=x-\eta \nabla f(x)\right]
$$

GD-Local-Search:
Goal: find $x$ such that $f\left(x^{\prime}\right) \geq f(x)-\varepsilon$
(the next iterate decreases $f$ by at most $\varepsilon$ )

## Gradient Descent Problems

GD: "all problems that can be solved by gradient descent"
Input: $C^{1}$-function $f:[0,1]^{n} \rightarrow[0,1]$, step size $\eta>0$, precision $\varepsilon>0$

$$
x_{k+1} \leftarrow x_{k}-\eta \nabla f\left(x_{k}\right)
$$

Goal: find a point where gradient descent terminates

$$
\left[x^{\prime}:=x-\eta \nabla f(x)\right]
$$

GD-Local-Search:
Goal: find $x$ such that $f\left(x^{\prime}\right) \geq f(x)-\varepsilon \quad$ (the next iterate decreases $f$ by at most $\varepsilon$ )
$\rightarrow$ in CLS: $\quad p(x):=f(x)$ and $g(x):=x-\eta \nabla f(x)$

## Gradient Descent Problems

GD: "all problems that can be solved by gradient descent"
Input: $C^{1}$-function $f:[0,1]^{n} \rightarrow[0,1]$, step size $\eta>0$, precision $\varepsilon>0$

$$
x_{k+1} \leftarrow x_{k}-\eta \nabla f\left(x_{k}\right)
$$

Goal: find a point where gradient descent terminates

$$
\left[x^{\prime}:=x-\eta \nabla f(x)\right]
$$

GD-Local-Search:
Goal: find $x$ such that $f\left(x^{\prime}\right) \geq f(x)-\varepsilon$
(the next iterate decreases $f$ by at most $\varepsilon$ )

## Gradient Descent Problems

GD: "all problems that can be solved by gradient descent"
Input: $C^{1}$-function $f:[0,1]^{n} \rightarrow[0,1]$, step size $\eta>0$, precision $\varepsilon>0$

$$
x_{k+1} \leftarrow x_{k}-\eta \nabla f\left(x_{k}\right)
$$

Goal: find a point where gradient descent terminates

$$
\left[x^{\prime}:=x-\eta \nabla f(x)\right]
$$

## GD-Local-Search:

Goal: find $x$ such that $f\left(x^{\prime}\right) \geq f(x)-\varepsilon$
(the next iterate decreases $f$ by at most $\varepsilon$ )

## GD-Fixed-Point:

Goal: find $x$ such that $\left\|x^{\prime}-x\right\| \leq \varepsilon$

## Gradient Descent Problems

GD: "all problems that can be solved by gradient descent"
Input: $C^{1}$-function $f:[0,1]^{n} \rightarrow[0,1]$, step size $\eta>0$, precision $\varepsilon>0$

$$
x_{k+1} \leftarrow x_{k}-\eta \nabla f\left(x_{k}\right)
$$

Goal: find a point where gradient descent terminates

$$
\left[x^{\prime}:=x-\eta \nabla f(x)\right]
$$

## GD-Local-Search:

Goal: find $x$ such that $f\left(x^{\prime}\right) \geq f(x)-\varepsilon$
(the next iterate decreases $f$ by at most $\varepsilon$ )

## GD-Fixed-Point:

Goal: find $x$ such that $\left\|x^{\prime}-x\right\| \leq \varepsilon$
(the next iterate is $\varepsilon$-close)
$\rightarrow$ polynomial-time equivalent!

## PPAD $\cap$ PLS



## PPAD $\cap$ PLS

EITHER-SOLUTION $(A, B)$


## PPAD $\cap$ PLS



## PPAD $\cap$ PLS

EITHER-SOLUTION $(A, B)$




## Consequences

## Consequences

- PPAD $\cap$ PLS is an interesting class!


## Consequences

- PPAD $\cap$ PLS is an interesting class!
- It captures continuous local search, and even gradient descent


## Consequences

- PPAD $\cap$ PLS is an interesting class!
- It captures continuous local search, and even gradient descent
- CLS and GD are robust with respect to:
$>$ dimension
$>$ domain
$>$ arithmetic circuits
> ...

Proof Sketch

## PPAD

Canonical complete problem: END-OF-LINE

## PPAD

Canonical complete problem: END-OF-LINE
Input: directed graph of paths and cycles, and a source

## PPAD

Canonical complete problem: END-OF-LINE
Input: directed graph of paths and cycles, and a source


## PPAD

Canonical complete problem: END-OF-LINE
Input: directed graph of paths and cycles, and a source


## PPAD

Canonical complete problem: END-OF-LINE
Input: directed graph of paths and cycles, and a source
Goal: find a sink, or another source


## PPAD

Canonical complete problem: END-OF-LINE
Input: directed graph of paths and cycles, and a source
Goal: find a sink, or another source


## PPAD

Canonical complete problem: END-OF-LINE
Input: directed graph of paths and cycles, and a source
Goal: find a sink, or another source

Reduction: high level

## Reduction: high level

Goal: reduction from EITHER-SOLUTION(END-OF-LINE, LOCAL-OPT) to 2D-GD-FIXED-POINT

## Reduction: high level

Goal: reduction from EITHER-SOLUTION(END-OF-LINE, LOCAL-OPT) to 2D-GD-FIXED-POINT
$\rightarrow$ Construct a continuously differentiable function $f:[0,1]^{2} \rightarrow \mathbb{R}$ such that any gradient descent fixed point yields a solution to the EITHER-SOLUTION instance

## Reduction: high level

Goal: reduction from EITHER-SOLUTION(END-OF-LINE, LOCAL-OPT) to 2D-GD-FIXED-POINT
$\rightarrow$ Construct a continuously differentiable function $f:[0,1]^{2} \rightarrow \mathbb{R}$ such that any gradient descent fixed point yields a solution to the EITHER-SOLUTION instance


## Reduction: high level

Goal: reduction from EITHER-SOLUTION(END-OF-LINE, LOCAL-OPT) to 2D-GD-FIXED-POINT
$\rightarrow$ Construct a continuously differentiable function $f:[0,1]^{2} \rightarrow \mathbb{R}$ such that any gradient descent fixed point yields a solution to the EITHER-SOLUTION instance


## Reduction: high level

Goal: reduction from EITHER-SOLUTION(END-OF-LINE, LOCAL-OPT) to 2D-GD-FIXED-POINT
$\rightarrow$ Construct a continuously differentiable function $f:[0,1]^{2} \rightarrow \mathbb{R}$ such that any gradient descent fixed point yields a solution to the EITHER-SOLUTION instance


## Reduction: high level

Goal: reduction from EITHER-SOLUTION(END-OF-LINE, LOCAL-OPT) to 2D-GD-FIXED-POINT
$\rightarrow$ Construct a continuously differentiable function $f:[0,1]^{2} \rightarrow \mathbb{R}$ such that any gradient descent fixed point yields a solution to the EITHER-SOLUTION instance



Warm up: Monotone-End-of-Line

Warm up: Monotone-End-of-Line
0
(1)
(2)
3
(4)
5
(6)
(7)

## Warm up: Monotone-End-of-Line



Special case of END-OF-LINE: No backward edges allowed!

## Warm up: Monotone-End-of-Line



Special case of END-OF-LINE: No backward edges allowed!

## Warm up: Monotone-End-of-Line



Special case of END-OF-LINE: No backward edges allowed!










Locally computable!
[Hubáček-Yogev, 2017] for CLS

Back to standard End-of-Line

## Back to standard End-of-Line

0
(1)
(2)
(3)
(4)
(5)
(6)
(7)

Back to standard End-of-Line


## Back to standard End-of-Line


green edges: forward
red edges: backwards










$\rightarrow$ to find a gradient descent fixed point, we have to solve the PPAD problem or the PLS problem


## Future Directions

- are there other intersections of classes that are interesting?


## Future Directions

- are there other intersections of classes that are interesting?
- candidates for (PPAD $\cap$ PLS)-completeness:
> CONTRACTION
$>$ TARSKI
> POLYNOMIAL-KKT
> MIXED-CONGESTION


## Future Directions

- are there other intersections of classes that are interesting?
- candidates for (PPAD $\cap$ PLS)-completeness:
> CONTRACTION
$>$ TARSKI
> POLYNOMIAL-KKT
> MIXED-CONGESTION $\square$ Solved!


## Future Directions

- are there other intersections of classes that are interesting?
- candidates for (PPAD $\cap$ PLS)-completeness:
> CONTRACTION
> TARSKI
> POLYNOMIAL-KKT
> MIXED-CONGESTION
Solved!
[Babichenko-Rubinstein, 2020]
2D-GD-FIXED-POINT $\leq$ MIXED-CONGESTION $\leq$ POLYNOMIAL-KKT

Thank You!

