A Converse to Banach's Fixed Point Theorem and Its CLS-Completeness

Constantinos Daskalakis
EECS and CSAIL, MIT
USA
costis@mit.edu

Christos Tzamos
Microsoft Research
USA
chtzamos@microsoft.com

Manolis Zampetakis
EECS and CSAIL, MIT
USA
mzampet@mit.edu

ABSTRACT

Banach's fixed point theorem for contraction maps has been widely used to analyze the convergence of iterative methods in non-convex problems. It is a common experience, however, that iterative maps fail to be globally contracting under the natural metric in their domain, making the applicability of Banach's theorem limited. We explore how generally we can apply Banach's fixed point theorem to establish the convergence of iterative methods when pairing it with carefully designed metrics.

Our first result is a strong converse of Banach's theorem, showing that it is a *universal analysis tool* for establishing global convergence of iterative methods to unique fixed points, and for bounding their convergence rate. In other words, we show that, whenever an iterative map globally converges to a unique fixed point, there exists a metric under which the iterative map is contracting and which can be used to bound the number of iterations until convergence. We illustrate our approach in the widely used power method, providing a new way of bounding its convergence rate through contraction arguments.

We next consider the computational complexity of Banach's fixed point theorem. Making the proof of our converse theorem constructive, we show that computing a fixed point whose existence is guaranteed by Banach's fixed point theorem is CLS-complete. We thus provide the first natural complete problem for the class CLS, which was defined in [9] to capture the complexity of problems such as P-matrix LCP, computing KKT-points, and finding mixed Nash equilibria in congestion and network coordination games.

CCS CONCEPTS

• **Theory of computation** → **Problems, reductions and completeness**; *Complexity classes*; *Continuous optimization*;

KEYWORDS

CLS-completeness, contraction maps, power method, Banach's Theorem, Bessaga's Theorem

1 INTRODUCTION

Several widely used computational methods are fixed point iteration methods. These include gradient descent, the power iteration method, alternating optimization, the expectation-maximization algorithm, k-means clustering, and others. In several important applications, we have theoretical guarantees for the convergence of these methods. For example, convergence to a unique solution can be guaranteed when the method is explicitly, or can be related to, gradient descent on a convex function [2, 5, 21]. More broadly, convergence to a stationary point can be guaranteed when the method is, or can be related to, gradient descent; for some interesting recent work on the limit points of gradient descent, see [19, 22] and their references.

Another, more general, style of analysis for proving convergence of fixed point iteration methods is via a potential (a.k.a. Lyapunov) function. For example, analyzing the power iteration method amounts to showing that, as time progresses, the unit vector maintained by the algorithm places more and more of its ℓ_2 energy on the principle eigenvector of the matrix used in the iteration, if it is unique, or, anyways, on the eigenspace spanned by the principal eigenvectors. In passing, it should also be noted that the power iteration method itself is commonly used as a tool for establishing the convergence of other fixed point iteration methods, such as alternating optimization; e.g. [15].

Ultimately, all fixed point iteration methods aim at converging to a fixed point of their iteration map. For global convergence to a unique solution, it should also be the case that the fixed point of the iteration map is unique. It is, thus, unsurprising that another widely used approach for establishing convergence of these methods is by appealing to Banach's fixed point theorem. To recall, consider an iteration map $x_{t+1} \leftarrow f(x_t)$, where $f : \mathcal{D} \rightarrow \mathcal{D}$, and suppose that there is a distance metric d such that $(\mathcal{D}, d)$ is a complete metric space and f is contracting with respect to d, i.e. for some constant $c < 1$, $d(f(x), f(y)) \leq c \cdot d(x, y)$, for all $x, y \in \mathcal{D}$. Under this condition, Banach's fixed point theorem guarantees that there is a *unique* fixed point $x^* = f(x^*)$. Moreover, iterating f is bound to converge to x^*. Specifically, the t-fold composition, $f^{[t]}$, of f with itself satisfies: $d(f^{[t]}(x_0), x^*) \leq c^t d(x_0, x^*)$, for any starting point x_0.

Constantinos Daskalakis, Christos Tzamos, and Manolis Zampetakis

Given Banach's theorem, if you established that your iteration method is contracting under some distance metric d, you would also have immediately proven that your method converges and that it may only converge to a unique point. Moreover, you can predict how many steps you need from any starting point x_0 to reach an approximate fixed point x satisfying $d(f(x), x) < \epsilon$ for some accuracy ϵ.[1] Alas, several widely used fixed point iteration methods are not generally contracting, or only contracting in a small neighborhood around their fixed points and not the entire domain where they are defined. At least, this is typically the case for the metric d under which approximate fixed points, $d(f(x), x) < \epsilon$, are sought. There is also quite an important reason why they may not be contracting: several of these methods may in fact have multiple fixed points.

Given the above motivation, our goal in this paper is to *investigate the extent to which Banach's fixed point theorem is a* universal analysis tool *for establishing that a fixed point iteration method both converges and globally converges to a unique fixed point.* More precisely, our question is the following: if an iterative map $x_{t+1} \leftarrow f(x_t)$ for some $f : \mathcal{D} \to \mathcal{D}$ converges to a unique fixed point x^* from any starting point, is there always a way to prove this using Banach's fixed point theorem? Additionally, can we always use Banach's fixed point theorem to compute how many iterations we would need to find an approximate fixed point x of f satisfying $d(x, f(x)) < \epsilon$, for some distance metric d and accuracy $\epsilon > 0$?

We study this question from both a mathematical and a computational perspective. On the mathematical side, we show a strong converse of Banach's fixed point theorem, saying the following: given an iterative map $x_{t+1} \leftarrow f(x_t)$ for some $f : \mathcal{D} \to \mathcal{D}$, some distance metric d on $\mathcal{D}$ such that $(\mathcal{D}, d)$ is a complete and proper metric space, and some accuracy $\epsilon > 0$, if f has a unique fixed point that the f-iteration converges to from any starting point, then for any constant $c \in (0, 1)$, there exists a distance metric d_c on $\mathcal{D}$ such that:

(1) d_c certifies uniqueness and convergence to the fixed point, by satisfying $d_c(f(x), f(y)) \leq c \cdot d_c(x, y)$, for all $x, y \in \mathcal{D}$;

(2) d_c allows an analyst to predict how many iterations of f would suffice to arrive at an approximate fixed point x satisfying $d(x, f(x)) < \epsilon$; notice in particular that we are interested in finding an approximate fixed point with respect to the original distance metric d (and not the constructed one d_c).

Our converse theorem is formally stated as Theorem 1 in Section 3. In the same section we discuss its relationship to other known converses of Banach's theorem known in the literature, in particular Bessaga's and Meyers's converse theorems. The improvement over these converses is that our constructed metric d_c is such that it allows us to bound the number of steps requied to reach an approximate fixed point according to the metric of interest d and not just d_c; namely Property 2 above holds. We discuss this further in Section 3.3. Section 3.2 provides a sketch of the proof, and the complete details can be found in the full version of the paper [10].

While the proof of Theorem 1 is non-constructive, it does imply that Banach's fixed point theorem is a universal analysis tool for establishing global convergence of fixed point iteration methods

to unique solutions. Namely, it implies that one can *always* find a witnessing metric. We illustrate this by studying an important such method: power iteration. The power iteration method is a widely-used and well-understood method for computing the eigenvalues and eigenvectors of a matrix. It is well known that if a matrix A has a unique principal eigenvector, then the power method starting from a vector non-perpendicular to the principal eigenvector will converge to it. This is shown using a potential function argument outlined above which also pins down the rate of convergence.

Our converse to Banach's theorem, guarantees that, besides the potential function argument, there must also exist a distance metric under which the power iteration is a contraction map. Such a distance metric is not obvious, as contraction under any ℓ_p-norm fails; we provide counter-examples in Section 4. To illustrate our theorem, we identify a new distance metric under which the power method is indeed contracting at the optimal rate. See Proposition 1. Our distance metric serves as an alternative proof for establishing that the power iteration converges and for pinning down its convergence rate.

We close the circle by studying Banach's fixed point theorem from a computational standpoint. Recent work of Daskalakis and Papadimitriou [9] has identified a complexity class, CLS, where finding a Banach fixed point lies. CLS, defined formally in Section 5, is a complexity class at the intersection of PLS [17] and PPAD [23]. Roughly speaking, PLS contains total problems whose existence of solutions is guaranteed by a potential function argument, while PPAD contains total problems whose existence of solutions is guaranteed by Brouwer's fixed point theorem. Lots of interesting work has been done on both classes in the past two decades; for a small sample see e.g. [1, 6, 8, 12, 13, 25, 26] and their references. CLS, lying in the intersection of PLS and PPAD, contains compututational problems whose existence of solutions is guaranteed by both a potential function and a fixed point argument.[2]

Unsurprisingly CLS contains several interesting problems, whose complexity is not known to lie in P, but which also are unlikely to be complete for PPAD or PLS. One of these problems is finding a Banach fixed point. Others include the P-matrix Linear Complementarity Problem, finding mixed Nash equilibria of network coordination and congestion games, computational problems related to finding KKT points, and solving Simple Stochastic Games; see [9] for precise definitions of these problems and for references. Moreover, recent work has provided cryptographic hardness results for CLS [16] based on obfuscation, extending work which proved cryptographic hardness results for PPAD [4, 18, 24].

Ultimately, the definition of CLS was inspired by a vast range of total problems that could not be properly classified as complete in PPAD or PLS due to the nature of their totality arguments. However, no natural complete problem for this class has been identified, besides CONTINUOUS LOCALOPT, through which the class was defined. By making our converse to Banach's fixed point theorem constructive, we show that finding a Banach fixed point is CLS-complete. More precisely, in Section 5 we define problem BANACH, whose input is a continuous function f and a continuous metric d, and whose goal is to either output an approximate fixed point of f or a

---

[1]Indeed, it can be easily shown that $d(f^{[t+1]}(x_0), f^{[t]}(x_0)) \leq c^t d(x_1, x_0)$. So $t = \log_{1/c} \frac{d(x_1, x_0)}{\epsilon}$ steps suffice.

[2]More precisely, it contains all problems reducible to CONTINUOUS LOCALOPT, defined in Section 5, and which doesn't necessarily capture the whole intersection of PPAD and PLS.

violation of the contraction of f with respect to d. In Theorem 2 we show that BANACH is CLS-complete.[3]

Further Related Work. We note that contemporaneously and independently from our work, Fearnley et al. [14] have also identified a CLS-complete problem related to Banach's fixed point theorem. Their problem, called METAMETRICCONTRACTION, takes as input a function f and a metametric d, and asks to find an approximate fixed point of f, or a violation of the contraction of f with respect to d. In comparison to our CLS-completeness results, the CLS-hardness of BANACH in our paper is stronger than that of METAMETRICCONTRACTION as the input to BANACH is a metric. On the other hand, the containment of METAMETRICCONTRACTION into CLS is stronger than the containment of BANACH, as BANACH is polynomial-time reducible to METAMETRICCONTRACTION.

2 NOTATION AND PRELIMINARIES

Basic Notation We use $\mathbb{R}_+$ to refer to set of non-negative real numbers and $\mathbb{N}_1$ is the set of natural numbers except 0. We call a function f *selfmap* if it maps a domain $\mathcal{D}$ to itself, i.e. $f : \mathcal{D} \to \mathcal{D}$. For a selfmap f we use $f^{[n]}$ to refer to the n times composition f with it self, i.e. $\underbrace{f(f(\ldots f(\cdot)))}_{n \text{ times}}$.

We use $\|\cdot\|_p$ to refer to the ℓ_p norm of a vector in $\mathbb{R}^n$. We use $\mathcal{D}/\!\sim$ to refer to the set of equivalence classes of the equivalence relation $\sim$ on a set $\mathcal{D}$. Finally, we use S^* to refer to the Kleene star of a set S.

A real valued function $g : \mathcal{D}^2 \to \mathbb{R}$ is called *symmetric* if $g(x, y) = g(y, x)$ and *anti-symmetric* if $g(x, y) = -g(y, x)$.

Topological Spaces, we define the notion of: topology, topological spaces, open sets, closed sets, interior of a set A, denoted $\text{Int}(A)$, closure of a set A, denoted $\text{Clos}(A)$.

Metric Spaces, we define the notion of: distance metric, metric space, diameter, bounded metric space, continuous function, open and closed sets in a metric space, compact set, locally compact metric space, proper metric space, open and closed balls, Cauchy sequence, complete metric space, equivalent metrics, continuity, Lipschitz continuity, contraction property, fixed point.

Because of its importance for the rest of the paper we also give here the definition of a distance metric and metric space.

DEFINITION 1. *Let $\mathcal{D}$ be a set and $d : \mathcal{D}^2 \to \mathbb{R}$ a function with the following properties:*
  (i) $d(x, y) \geq 0$ for all $x, y \in \mathcal{D}$.
  (ii) $d(x, y) = 0$ if and only if $x = y$.
  (iii) $d(x, y) = d(y, x)$ for all $x, y \in \mathcal{D}$.
  (iv) $d(x, y) \leq d(x, z) + d(z, x)$ for all $x, y, z \in \mathcal{D}$. *This is called triangle inequality.*
Then we say that d is a metric *on $\mathcal{D}$, and $(\mathcal{D}, d)$ is a* metric space.

Basic Iterative Procedure. If a selfmap f has a fixed point and is continuous, we can define the following sequence of points $x_{n+1} = f(x_n)$ where the starting point x_0 can be picked arbitrarily. If (x_n) converges to a point $\bar{x}$ then

$$\lim_{n \to \infty} x_{n+1} = \lim_{n \to \infty} f(x_n) \Rightarrow \lim_{n \to \infty} x_{n+1} = f\left(\lim_{n \to \infty} x_n\right) \Rightarrow \bar{x} = f(\bar{x}).$$

This observation implies that a candidate procedure for computing a fixed point of a selfmap f is to *iteratively* apply the function f starting from an arbitrary point x_0. If this procedure converges then the limit is a fixed point x^* of f. We will refer to this method of computing fixed points as the *Basic Iterative Procedure.*

Arithmetic Circuits. In Section 5 we work with functions from continuous domains to continuous domains represented as *arithmetic circuits*. An arithmetic circuit is defined by a directed acyclic graph (DAG). The inputs to the circuit are in-degree 0 nodes, and the outputs are out-degree 0 nodes. Each non-input node is a gate from the set $\{+, -, *, \max, \min, >\}$, performing an operation on the outputs of its in-neighbors. The meaning of the ">" gate is $> (x, y) = 1$ if $x > y$ and 0 otherwise. We also allow "output a rational constant" gates. These are gates without any inputs, which output a rational constant.

3 CONVERSE BANACH FIXED POINT THEOREMS

We start, in Section 3.1, with an overview of known converses to Banach's fixed point theorem. We also explain why these converses are not enough to prove that Banach's fixed point theorem is a universal tool for analyzing the convergence of iterative algorithms. Then, in Section 3.2, we prove a stronger converse theorem that demonstrates the universality of Banach's fixed point theorem for the analysis of iterative algorithms. Before beginning, we formally state Banach's fixed Point Theorem. A useful survey of the applications of this theorem can be found in [7].

BANACH'S FIXED POINT THEOREM. *Suppose d is a distance metric function such that $(\mathcal{D}, d)$ is a complete metric space, and suppose that $f : \mathcal{D} \to \mathcal{D}$ is a contraction map according to d, i.e.*

$$d(f(x), f(y)) \leq c \cdot d(x, y), \forall x, y, \text{ for some } c < 1. \tag{1}$$

Then f has a unique fixed point x^* and the convergence rate of the Basic Iterative Procedure with respect to d is c. That is, $d(f^{[n]}(x_0), x^*) < c^n \cdot d(x_0, x^*)$, for all x_0.

3.1 Known Converses to Banach's Fixed Point Theorem

The first known converse to Banach's fixed point theorem is the following [3].

BESSAGA'S CONVERSE THEOREM. *Let f be a map from $\mathcal{D}$ to itself, and suppose that $f^{[n]}$ has unique fixed point for every $n \in \mathbb{N}_1$. Then, for every constant $c \in (0, 1)$, there exists a distance metric d_c such that $(\mathcal{D}, d_c)$ is a complete metric space and f is a contraction map with respect to d_c with contraction constant c.*

The implication of the above theorem is that, if we want to prove existence and uniqueness of fixed points of $f^{[n]}$ for all n, then Banach's fixed point theorem is a universal way to do it. Moreover,

---

there is a potential function of the form $p(x) = d_c(x, f(x))$, where d_c is a distance metric, that decreases under successive applications of f, and successive applications of f starting from any point x_0 are bound to converge to the unique fixed point of f.

Unfortunately, d_c cannot provide any information about the number of steps that the Basic Iterative Procedure needs before computing an approximate fixed point under some metric d of interest. The reason is that, after $\log_c \varepsilon$ steps of the Basic Iterative Procedure, we only have $d_c(x_n, f(x_n)) \leq \varepsilon$. However, d_c might not have any relation to d, hence an approximate fixed point under d_c may not be one for d. So Bessaga's theorem is not useful for bounding the running time of iterative methods for approximate fixed point computation.

Given the above discussion, it is reasonable to expect that a converse to Banach's theorem that is useful for bounding the running time of approximate fixed point computation methods should take into account, besides the function f and its domain $\mathcal{D}$, the distance metric d under which we are interested in computing approximate fixed points. One step in this direction has already been made by Meyers [20].

Meyers's Converse Theorem. *Let $(\mathcal{D}, d)$ be a complete metric space, where $\mathcal{D}$ is compact, and suppose that $f : \mathcal{D} \to \mathcal{D}$ is continuous with respect to d. Suppose further that f has a unique fixed point x^*, that the Basic Iterative Method converges to x^* from any starting point, and that there exists an open neighborhood U of x^* such that $f^{[n]}(U) \to \{x^*\}$. Then, for any $c \in (0, 1)$, there exists a distance metric d_c, which is topologically equivalent to d, such that $(\mathcal{D}, d_c)$ is a complete metric space and f is a contraction map with respect to d_c with contraction c.*

Compared to Bessaga's theorem, the improvement offered by Meyer's Theorem is that, instead of the existence of an arbitrary metric, it proves the existence of a metric that is topologically equivalent to the metric d. However, this is still not enough to bound the number of steps needed by the Basic Iterative Procedure in order to arrive at a point x_n such that $d(x_n, f(x_n)) \leq \varepsilon$. Our goal in the next section is to close this gap. We will also replace the compactness assumption with the assumption that $(\mathcal{D}, d)$ is proper, so that the converse holds for unbounded spaces.

3.2 A New Converse to Banach's Fixed Point Theorem

The main technical idea behind our converse to Banach's fixed point theorem is to adapt the proof of Meyers's theorem to get a distance metric d_c with the property $d_c(x, y) \geq d(x, y)$ everywhere, except maybe for the region $d(x, x^*) \leq \varepsilon$. This implies that, if we guarantee that $d_c(x_n, x^*) \leq \varepsilon$, then $d(x_n, x^*) \leq \varepsilon$.

Theorem 1. *Suppose $(\mathcal{D}, d)$ is a complete, proper metric space, $f : \mathcal{D} \to \mathcal{D}$ is continuous with respect to d and the following hold:*

(1) f has a unique fixed point x^*;
(2) for every $x \in \mathcal{D}$, the sequence $(f^{[n]}(x))$ converges to x^* with respect to d; moreover there exists an open neighborhood U of x^* such that $f^{[n]}(U) \to \{x^*\}$.

Then, for every $c \in (0, 1)$ and $\varepsilon > 0$, there exists a distance metric function $d_{c,\varepsilon}$ that is topologically equivalent to d and is such that

$(\mathcal{D}, d_{c,\varepsilon})$ *is a complete metric space and*

$$\forall x, y \in \mathcal{D} : d_{c,\varepsilon}(f(x), f(y)) \leq c \cdot d_{c,\varepsilon}(x, y); \tag{2a}$$

$$\forall x, y \in \mathcal{D} : d_{c,\varepsilon}(x, y) \leq \varepsilon \implies \min\{d(x^*, x), d(x^*, y), d(x, y)\} \leq 2\varepsilon. \tag{2b}$$

Remark. Notice that the continuity of f is a necessary assumption for the above statement to hold, as (2a) implies continuity given that $d_{c,\varepsilon}$ and d are topologically equivalent. Also the condition 2. of the theorem is implied by the existence of $d_{c,\varepsilon}$ and it is not true even if $f^{[n]}(x) \to x^*$ for any $n \in \mathbb{N}$, since counter examples exist. Therefore this assumption is also necessary for our theorem to hold.

The proof of our Theorem 1 adapts the construction of Meyers's proof, to ensure that (2b) is satisfied. We give here a proof sketch postponing the complete details to the full version of the paper [10], where we repeat also all the technical details proven by Meyers [20].

Proof Sketch. The construction of the metric d_c follows is done in three steps:

I. Starting from the original metric d, a non-expanding closure of d is defined as the metric

$$d_M(x, y) = \sup_{i \geq 0} d(f^{(i)}(x), f^{(i)}(y)).$$

This is topologically equivalent to d, but ensures that the images of any two points are at least as close in d_M as the original two points (non-expanding property).

Notice that as $d_M(x, y) \geq d(x, y)$ for all points $x, y \in \mathcal{D}$, if we ensure that Property (2b) holds with respect to d_M for the final constructed metric $d_{c,\varepsilon}$, it will also hold with respect to the original metric d.

II. Given d_M, the construction proceeds by defining a function $\rho_{c,\varepsilon}$ which satisfies (2a). This function achieves contraction by a constant $c < 1$ by counting the number of steps required to reach an ε-ball close to the fixed point.

While for the original proof of Meyer any such ε-ball suffices, in order to guarantee Property (2b), our proof requires a set S of points with small diameter with respect to d such that performing an iteration of f on any one of them results in a point still in the set S. In the full version of the paper [10], we show that such a set always exists. This guarantees that

$$\rho_{c,\varepsilon}(x, y) \geq d_M(x, y), \text{ if } \max\{d(x^*, x), d(x^*, y)\} \geq \varepsilon,$$

and therefore Property (2b) is preserved.

The function $\rho_{c,\varepsilon}$ satisfies all required properties other than triangle inequality and thus is not a metric. However, it can be converted into one.

III. Given $\rho_{c,\varepsilon}$, we construct the sought after metric $d_{c,\varepsilon}$ by taking it equal to the $\rho_{c,\varepsilon}$-geodesic distance (metric closure of $\rho_{c,\varepsilon}$). This directly converts $\rho_{c,\varepsilon}$ into a metric. In the full version of the paper [10], we show that after this operation Properties (2a) and (2b). □

3.3 Corollaries of Theorem 1

Property (2b) of the metric output by Theorem 1 has some interesting corollaries that we would not be able to get using the known converses to Banach's theorem discussed in Section 3.1. The first

one is that we can now compute, from $d_{c,\varepsilon}$, the number of iterations needed in order to get to within ε of the fixed point x^* of f from any starting point $x_0 \in \mathcal{D}$.

COROLLARY 1. *Under the assumptions of Theorem 1, starting from a point $x_0 \in \mathcal{D}$, and for any constant $c \in (0,1)$, the Basic Iterative Procedure finds a point x such that $d(x, x^*) \leq \varepsilon$ after*

$$\frac{\log(d_{c,\varepsilon/2}(x_0, f(x_0))) + \log((2 - 2c)/\varepsilon)}{\log(1/c)}$$

iterations, where $d_{c,\varepsilon/2}$ is the metric guaranteed by Theorem 1.

In Corollary 1, for any given ε of interest, we have to identify a different distance metric $d_{c,\varepsilon/2}$, guaranteed by Theorem 1, to bound the number of steps required by the Basic Iterative Procedure to get to within ε from the fixed point. Sometimes we are interested in the explicit tradeoff between the number of steps required to get to the proximity of the fixed point and the amount of proximity ε. To find such a tradeoff we have to make additional assumptions on f. A mild assumption that is commonly satisfied by iterative procedures for non-convex problems is that the Basic Iterative Procedure *locally converges* to the fixed point x^*. That is, if x_0 is appropriately close to x^*, then the Basic Iterative Procedure converges. A common way of proving local convergence is to prove that f is a contraction with respect to d locally for $x, y \in \bar{B}(x^*, \varepsilon)$. Theorem 1 provides a way to extend this local contraction property to the whole domain $\mathcal{D}$ and get an an explicit closed form of the tradeoff between the number of steps and ε, as implied by the following result.

COROLLARY 2. *Under the assumptions of Theorem 1, and the assumption that there exists $0 < c < 1, \delta > 0$ such that*

$$d(f(x), f(y)) \leq c \cdot d(x, y) \text{ for all } x, y \in \bar{B}(x^*, \delta),$$

starting from any point $x_0 \in \mathcal{D}$, the Basic Iterative Procedure finds a point x such that $d(x, x^*) \leq \varepsilon$ after

$$\frac{\log(d_{c,\delta/2}(x_0, f(x_0))) + \log(1/\varepsilon) + \log(1 - c) + 1}{\log(1/c)} + 1$$

iterations, where $d_{c,\delta/2}$ is the metric guaranteed by Theorem 1.

4 EXAMPLE: THE POWER ITERATION AS A CONTRACTION MAP

The results of the previous section imply that Banach's fixed point theorem is a universal analysis tool for establishing global convergence of fixed point iteration methods to unique solutions. While the proof of Theorem 1 is non-constructive, it does imply that one can *always* find a witnessing metric under which the iterative map is contracting.

In this section, we illustrate this possibility by studying an important iterative method, the power iteration. The power iteration method is a widely-used and well-understood method for computing the eigenvalues and eigenvectors of a matrix. For a given matrix A, it is defined as:

$$x_{t+1} = \frac{Ax_t}{\|Ax_t\|_2}$$

It is well known that if a matrix A has a unique principal eigenvector, then the power method starting from a vector non-perpendicular to the principal eigenvector will converge to it. This is shown using

a potential function argument which also pins down the rate of convergence.

Our converse to Banach's theorem, guarantees that, besides the potential function argument, there must also exist a distance metric under which the power iteration is a contraction map. To illustrate our theorem, we identify a new distance metric under which the power method is indeed contracting at the optimal rate.

Such a distance metric is not obvious. As the following counterexample shows, contraction under any ℓ_p-norm fails.

Counterexamples for $\|\cdot\|_p$. We show a counter example for ℓ_2 norm which directly extends to any ℓ_p norm. In particular, let $n = 2$, $\lambda_1 = 2, \lambda_2 = 1$ and the corresponding eigenvectors be $e_1 = (1, 0)$ and $e_2 = (0, 1)$. The power iteration is given by $f(x) = \frac{(2x_1, x_2)}{\sqrt{4x_1^2 + x_2^2}}$.

We set $x = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)$ and $y = \left(\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\right)$. We get that

$$\|f(x) - f(y)\|_2 = \left\|\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) - \left(\frac{2}{\sqrt{13}}, \frac{3}{\sqrt{13}}\right)\right\|_2 \geq 0.19.$$

and

$$\|x - y\|_2 = \left\|\left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right) - \left(\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\right)\right\|_2 \leq 0.14$$

and therefore $\|f(x) - f(y)\|_2 > \|x - y\|_2$.

Even though contraction is not achieved under any ℓ_p-norm, it is possible to construct a metric under which power iteration is contracting even at the optimal rate which is given by the ratio of the two largest eigenvalues of matrix A. Our next theorem constructs such a metric.

PROPOSITION 1. *Let $A \in \mathbb{R}^{n \times n}$ be a matrix with left eigenvector-eigenvalue pairs $(\lambda_1, v_1), ..., (\lambda_n, v_n)$ such that $\lambda_1 > \lambda_2 \geq ... \geq \lambda_n$. Then the power iteration, $x_{t+1} = f(x_t) \triangleq \frac{Ax_t}{\|Ax_t\|}$ is contracting under the metric $d(x, y) = \left\|\frac{x}{\langle x, v_1\rangle} - \frac{y}{\langle y, v_1\rangle}\right\|_2$ with contraction constant λ_2/λ_1, i.e. for all $x, y \in \mathbb{R}^n$:*

$$d(f(x), f(y)) \leq \frac{\lambda_2}{\lambda_1} d(x, y).$$

Moreover, $t = \frac{\log(d(x_0, v_1)/\varepsilon)}{\log(\lambda_1/\lambda_2)}$ iterations suffice to have $\|x_t - v_1\|_2 \leq d(x_t, v_1) \leq \varepsilon$.

PROOF. For any vector x, it holds that $\langle Ax, v_1\rangle = \lambda_1\langle x, v_1\rangle$. We have that

$$\begin{aligned} d(f(x), f(y)) &= \left\|\frac{Ax}{\langle Ax, v_1\rangle} - \frac{Ay}{\langle Ay, v_1\rangle}\right\|_2 \\ &= \frac{1}{\lambda_1}\left\|A\left(\frac{x}{\langle x, v_1\rangle} - \frac{y}{\langle y, v_1\rangle}\right)\right\|_2 \\ &\leq \frac{\lambda_1}{\lambda_1}\left\|\frac{x}{\langle x, v_1\rangle} - \frac{y}{\langle y, v_1\rangle}\right\|_2 \\ &= \frac{\lambda_2}{\lambda_1}d(x, y) \end{aligned}$$

where the inequality is true as the vector $\frac{x}{\langle x, v_1\rangle} - \frac{y}{\langle y, v_1\rangle}$ is perpendicular to the principal eigenvector v_1. This shows that f is contracting with respect to d as required.

To convert a bound on the d metric to a bound on the error with respect to the ℓ_2 norm, we can see that at every step $t > 0$, $\|x_t\|_2 = 1$. If at some step $t > 0$, it holds that $d(x_t, v_1) \leq \varepsilon$, we get

$$\varepsilon^2 \geq d(x_t, v_1)^2 = \left\| \frac{x_t}{\langle x_t, v_1 \rangle} - v_1 \right\|_2^2 = \langle x_t, v_1 \rangle^{-2} - 1 \Rightarrow$$

$$\langle x_t, v_1 \rangle \geq \left(1 + \varepsilon^2 \right)^{-1/2}.$$

This implies that

$$\|x_t - v_1\|_2^2 = 2(1 - \langle x_t, v_1 \rangle) \leq 2 \left(1 - \left(1 + \varepsilon^2 \right)^{-1/2} \right) \leq \varepsilon^2.$$

This guarantees that bounding the norm d by ε implies a bound of ε on the ℓ_2 norm between the principal eigenvector and the current iterate x_t.

Using these observations and following the same approach as in Corollaries 1-2 we get the required bound on the number of iterations. □

Notice, that the definition of the metric in Proposition 1 depends on the principal eigenvector but not on any of the other eigenvectors. When applied to show global convergence of Markov chains, the principal eigenvector corresponds to the stationary distribution. For a symmetric Markov chain whose stationary distribution is uniform Proposition 1 implies that the iterations are contracting directly with respect to the ℓ_2 norm.

5 BANACH IS COMPLETE FOR CLS

As discussed in Section 1, the complexity class CLS was defined in [9] to capture problems in the intersection of PPAD and PLS, such as P-matrix LCP, mixed Nash equilibria of congestion and multi-player coordination games, finding KKT points, etc. It also contains computational variants of finding fixed points whose existence is guaranteed by Banach's fixed point theorem. In this section, we close the circle by proposing two variants of Banach fixed point computation that are both CLS-complete. Our CLS completeness results are obtained by making our proof of Theorem 1 constructive. We start with a formal definition of CLS, which is defined in terms of the problem CONTINUOUS LOCALOPT.

DEFINITION 2. CONTINUOUS LOCALOPT takes as input two functions $f : [0,1]^3 \to [0,1]^3$, $p : [0,1]^3 \to [0,1]$, both represented as arithmetic circuits, and two rational positive constants ε and λ. The desired output is any of the following:
(CO1) a point $x \in [0,1]^3$ such that $p(f(x)) \geq p(x) - \varepsilon$.
(CO2) two points $x, x' \in [0,1]^3$ violating the λ-Lipschitz continuity of f, i.e.
$|f(x) - f(x')|_1 > \lambda |x - x'|_1$.
(CO3) two points x, x' violating the λ-Lipschitz continuity of p, i.e.
$|p(x) - p(x')| > \lambda |x - x'|_1$.
The class CLS is the set of search problems that can be reduced to CONTINUOUS LOCALOPT.

REMARK 1. *As discussed in [9], both the choice of domain $[0,1]^3$ and the use of ℓ_1 norm in the definition of the above problem are not crucial, and high-dimensional polytopes as well as other ℓ_p norms can also be used in the definition without any essential effect to the complexity of the problem. Moreover, instead of the functions f and p being provided in the input as arithmetic circuits there is a canonical way to provide them in the input as binary circuits that define the values of f and p on all points of some finite bit complexity, and (implicitly) extend to the full domain via continuous interpolation. In this way, we can syntactically guarantee the Lipschitz continuity of both f and p and can remove (CO2) and (CO3) from the above definition. For more details, please see [9], [8] and [11]. This remark applies to all definitions in this section.*

The variant of Banach's theorem that is known to belong to CLS is CONTRACTION MAP, defined as follows:

DEFINITION 3 ([9]). CONTRACTION MAP takes as input a function $f : [0,1]^3 \to [0,1]^3$ represented as an arithmetic circuit and three rational positive constants $\varepsilon, \lambda, c < 1$. The desired output is any of the following (where d represents Euclidean distance):
(Oa) a point $x \in [0,1]^3$ such that $d(x, f(x)) \leq \varepsilon$
(Ob) two points $x, x' \in [0,1]^3$ disproving the contraction of f w.r.t. d with constant c, i.e.
$d(f(x), f(x')) > c \cdot d(x, x')$
(Oc) two points $x, x' \in [0,1]^3$ disproving the λ-Lipschitz continuity of f, i.e.
$|f(x) - f(x')|_1 > \lambda |x - x'|_1$.

CONTRACTION MAP targets fixed points whose existence is guaranteed by Banach's fixed point theorem when f is a contraction map with respect to the Euclidean distance. However, it doesn't capture the full generality of Banach's theorem, since the latter can be applied to any complete metric space. We thus define a more general problem, BANACH that: (i) still lies inside CLS, (ii) captures the generality of Banach's theorem, (iii) and in fact tightly captures the complexity of the class CLS, by being CLS-complete. This problem is defined as follows:

DEFINITION 4. BANACH takes as input two functions $f : [0,1]^3 \to [0,1]^3$ and $d : [0,1]^3 \times [0,1]^3 \to \mathbb{R}$ represented as arithmetic circuits, where d is promised to be a metric that is topologically equivalent to the Euclidean distance and satisfy that $([0,1]^3, d)$ is a complete metric space, and three rational positive constants $\varepsilon, \lambda, c < 1$. The desired output is any of the following:
(Oa) a point $x \in [0,1]^3$ such that $d(x, f(x)) \leq \varepsilon$
(Ob) two points $x, x' \in [0,1]^3$ disproving the contraction of f w.r.t. d with constant c, i.e.
$d(f(x), f(x')) > c \cdot d(x, x')$
(Oc) two points $x, x' \in [0,1]^3$ disproving the λ-Lipschitz continuity of f, i.e.
$|f(x) - f(x')|_1 > \lambda |x - x'|_1$.
(Od) four points $x_1, x_2, y_1, y_2 \in [0,1]^3$ with $x_1 \neq x_2$ and $y_1 \neq y_2$ disproving the λ-Lipschitz continuity of $d(\cdot, \cdot)$, i.e.

$$|d(x_1, x_2) - d(y_1, y_2)| > \lambda \left(|x_1 - y_1|_1 + |x_2 - y_2|_1 \right).$$

REMARK 2. *We remark that BANACH is tightly related to CONTRACTION MAP defined above, with the following differences. First, instead of Euclidean distance, the metric with respect to which f is purportedly contacting is provided as part of the input and it is promised to be a metric. Second, we need to add an extra type of accepted solution (Od), which is a violation of the Lipschitz property of that metric. This is necessary to guarantee that the above problem has a solution of polynomial length for any possible input, and in particular*

is needed to place the above problem in CLS. (It is not needed for the CLS-hardness.)

Our main result is the following:

THEOREM 2. BANACH is CLS-complete.

We give here a sketch of the proof of Theorem 2 and we present the full proof in the full version of the paper [10].

PROOF SKETCH. Since the inclusion to CLS is a simple argument very similar to the argument from [9] that shows that CONTRACTION MAP belongs to CLS, we focus here on the hardness proof.

We are given two functions $f : [0, 1]^3 \to [0, 1]^3$, $p : [0, 1]^3 \to [0, 1]$ and we want to find a contraction $d : [0, 1]^3 \times [0, 1]^3 \to \mathbb{R}$ such that f is a contraction map with respect to d and the points where $p(f(x)) \geq p(x) - \varepsilon$ are approximate fixed points of f with respect to d.

The inspiration of this proof is to make the proof of Theorem 1 constructive in polynomial time. We therefore follow the steps of the proof sketch of Theorem 1 as presented in Section 3.

Step I. Since we don't have the strong requirement of Theorem 1 to output a metric that is topologically equivalent with some given metric we can use in place of d_M any metric d' such that f is non-expanding with respect to d'. Hence we can easily observe that the *discrete metric* can be used as d_M.

Step II. The construction of Theorem 1 uses in the definition of $d(x, y)$ the number of times $n(x)$, that we have to apply f on x in order for $f^{[n(x)]}(x)$ to come ε-close to the fixed point x^* of f. Of course $n(x)$ is not a quantity that can be computed in polynomial time. Instead we show that it suffices to use an upper bound on $n(x)$ which we can get using the potential function, namely $p(x)/\varepsilon$. Of course the operations that we are allowed to use to describe d as an arithmetic circuit are limited and this step appears to need more expressive power that the simple arithmetic operations that we are allowed to use. We give a careful construction that bypasses these difficulties and completes this step of the proof.

Steps III. This step of Theorem 1 is highly non-constructive and hence we cannot hope to replicate it in polynomial time. But we prove that our carefully designed metric already has the triangle inequality and hence the transitive closure step is not necessary.

The last part of our proof is to show that the constructed circuit of d is actually Lipschitz with a relatively small Lipschitz constant if the potential function p is Lipschitz. That is, we have to show that the circuit of d does not need some time exponentially many bits with respect to the size of the circuits of p and the magnitude of the constant $1/\varepsilon$. Not surprisingly we observe that in order to succeed to this task we have to set approximately $c = 1 - \varepsilon$. This is natural to expect, since if we could set a much lower contraction constant then we could find the approximate fixed point of f in much less that $\text{poly}(1/\varepsilon)$ steps which cannot hold given that CLS $\neq$ FP. □

ACKNOWLEDGMENTS

REFERENCES

[1] Omer Angel, Sébastien Bubeck, Yuval Peres, and Fan Wei. 2017. Local max-cut in smoothed polynomial time. In *Symposium on Theory of Computing (STOC)*.

[2] Aharon Ben-Tal and Arkadi Nemirovski. 2001. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications.* SIAM.

[3] C. Bessaga. 1959. On the converse of Banach "fixed-point principle". *Colloquium Mathematicae* 7, 1 (1959), 41–43.

[4] Nir Bitansky, Omer Paneth, and Alon Rosen. 2015. On the cryptographic hardness of finding a nash equilibrium. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*. IEEE, 1480–1498.

[5] Stephen Boyd and Lieven Vandenberghe. 2004. *Convex optimization.* Cambridge university press.

[6] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. 2009. Settling the complexity of computing two-player Nash equilibria. *Journal of the ACM (JACM)* 56, 3 (2009), 14.

[7] Keith Conrad. 2014. The contraction mapping theorem. *Expository paper. University of Connecticut, College of Liberal Arts and Sciences, Department of Mathematics* (2014).

[8] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. 2009. The complexity of computing a Nash equilibrium. *SIAM J. Comput.* 39, 1 (2009), 195–259.

[9] Constantinos Daskalakis and Christos H. Papadimitriou. 2011. Continuous Local Search. In *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011.* 790–804. https://doi.org/10.1137/1.9781611973082.62

[10] Constantinos Daskalakis, Christos Tzamos, and Manolis Zampetakis. 2017. A Converse to Banach's Fixed Point Theorem and its CLS Completeness. *arXiv preprint arXiv:1702.07339* (2017).

[11] Kousha Etessami and Mihalis Yannakakis. 2007. On the Complexity of Nash Equilibria and Other Fixed Points. In *Foundations of Computer Science (FOCS)*. IEEE.

[12] Michael Etscheid and Heiko Röglin. 2017. Smoothed analysis of local search for the maximum-cut problem. *ACM Transactions on Algorithms (TALG)* (2017).

[13] Alex Fabrikant, Christos Papadimitriou, and Kunal Talwar. 2004. The complexity of pure Nash equilibria. In *Symposium on Theory of Computing (STOC)*. ACM.

[14] John Fearnley, Spencer Gordon, Ruta Mehta, and Rahul Savani. 2017. CLS: New Problems and Completeness. *arXiv preprint arXiv:1702.06017* (2017).

[15] Moritz Hardt. 2014. Understanding alternating minimization for matrix completion. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*. IEEE, 651–660.

[16] Pavel Hubacek and Eylon Yogev. 2017. Hardness of continuous local search: Query complexity and cryptographic lower bounds. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*. SIAM, 1352–1371.

[17] David S. Johnson, Christos H. Papadimitriou, and Mihalis Yannakakis. 1988. How Easy is Local Search? *J. Comput. Syst. Sci.* 37, 1 (1988), 79–100.

[18] Ilan Komargodski, Moni Naor, and Eylon Yogev. 2017. White-Box vs. Black-Box Complexity of Search Problems: Ramsey and Graph Property Testing.. In *Electronic Colloquium on Computational Complexity (ECCC)*, Vol. 24. 15.

[19] Jason D. Lee, Max Simchowitz, Michael I. Jordan, and Benjamin Recht. 2016. Gradient Descent Only Converges to Minimizers. In *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA, June 23-26, 2016.* 1246–1257.

[20] Philip R. Meyers. 1967. A converse to Banach's contraction theorem. *Journal of Research of the National Bureau of Standards Section B Mathematics and Mathematical Physics* 71B, 2 and 3 (apr 1967), 73.

[21] Yurii Nesterov. 2013. *Introductory lectures on convex optimization: A basic course.* Vol. 87. Springer Science & Business Media.

[22] Ioannis Panageas and Georgios Piliouras. 2016. Gradient Descent Converges to Minimizers: The Case of Non-Isolated Critical Points. *CoRR* abs/1605.00405 (2016).

[23] Christos H. Papadimitriou. 1994. On the Complexity of the Parity Argument and Other Inefficient Proofs of Existence. *J. Comput. Syst. Sci.* 48, 3 (1994), 498–532.

[24] Alon Rosen, Gil Segev, and Ido Shahaf. 2016. Can PPAD Hardness be Based on Standard Cryptographic Assumptions?. In *Electronic Colloquium on Computational Complexity (ECCC)*, Vol. 23. 59.

[25] Aviad Rubinstein. 2016. Settling the complexity of computing approximate two-player Nash equilibria. In *Foundations of Computer Science (FOCS)*. IEEE.

[26] Alexander Skopalik and Berthold Vöcking. 2008. Inapproximability of pure Nash equilibria. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*. ACM, 355–364.