

# A Converse to Banach’s Fixed Point Theorem and its CLS Completeness

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

Christos Tzamos  
EECS and CSAIL, MIT  
tzamos@mit.edu

Manolis Zampetakis  
EECS and CSAIL, MIT  
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 [DP11] 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.

# 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 [BTN01, Nes13, BV04]. 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 [PP16, LSJR16] 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  $\ell_2$  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; see Appendix A.2 for a refresher. 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. [Har14].

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  $x_0$ .

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  $\epsilon$ .<sup>1</sup> 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  $f : \mathcal{D} \rightarrow \mathcal{D}$  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:

---

<sup>1</sup>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.

given an iterative map  $x_{t+1} \leftarrow f(x_t)$  for some  $f : \mathcal{D} \rightarrow \mathcal{D}$ , 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 required 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 Appendix B.

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 and in Appendix A.2, 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  $\ell_p$ -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 [DP11] 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 [JPY88] and PPAD [Pap94]. 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. [DGP09, CDT09, Rub16, FPT04, SV08, ER17, ABPW17] and their references. CLS, lying in the intersection of PLS and PPAD, contains computational problems whose existence of solutions is guaranteed by both a potential function and a fixed point argument.<sup>2</sup>

---

<sup>2</sup>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.

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 [DP11] for precise definitions of these problems and for references. Moreover, recent work has provided cryptographic hardness results for CLS [HY17] based on obfuscation, extending work which proved cryptographic hardness results for PPAD [BPR15, RSS16, KNY17].

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 violation of the contraction of  $f$  with respect to  $d$ . In Theorem 2 we show that BANACH is CLS-complete.<sup>3</sup>

**Further Related Work.** We note that contemporaneously and independently from our work, Fearnley et al. [FGMS17] 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.  $f : \mathcal{D} \rightarrow \mathcal{D}$ . For a selfmap  $f$  we use  $f^{[n]}$  to refer to the  $n$  times composition  $f$  with it self, i.e.  $\underbrace{f(f(\dots f(\cdot)))}_{n \text{ times}}$ .

We use  $\|\cdot\|_p$  to refer to the  $\ell_p$  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  $g : \mathcal{D}^2 \rightarrow \mathbb{R}$  is called *symmetric* if  $g(x, y) = g(y, x)$  and *anti-symmetric* if  $g(x, y) = -g(y, x)$ .

In Appendix A.1, the reader can find some well-known definitions that we are going to use in the rest of the paper. More precisely in the field of:

**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

---

<sup>3</sup>It is worth pointing out that, while some problems in CLS (e.g. Banach fixed points, simple stochastic games) have unique solutions, most do not. Given that contraction maps have unique fixed points, the way we bypass the potential oxymoron, is by accepting as solutions violations of contraction.

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  $d : \mathcal{D}^2 \rightarrow \mathbb{R}$  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

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

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 [Con14].

**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  $f : \mathcal{D} \rightarrow \mathcal{D}$  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 [Bes59].

**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 [Mey67].

**Meyers's Converse Theorem.** *Let  $(\mathcal{D}, d)$  be a complete metric space, where  $\mathcal{D}$  is compact, and suppose that  $f : \mathcal{D} \rightarrow \mathcal{D}$  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  $f^{[n]}(U) \rightarrow \{x^*\}$ . 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,  $f : \mathcal{D} \rightarrow \mathcal{D}$  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  $f^{[n]}(U) \rightarrow \{x^*\}$ .

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

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

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

**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  $f^{[n]}(x) \rightarrow x^*$  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 Appendix B, where we repeat also all the technical details proven by Meyers [Mey67].

*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  $\varepsilon$ -ball close to the fixed point.

While for the original proof of Meyer any such  $\varepsilon$ -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  $S$ . We show that such a set always exists in Lemma 5 in Appendix B.

This guarantees that  $\rho_{c,\varepsilon}(x, y) \geq d_M(x, y)$  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. We show that after this operation Properties (2a) and (2b). This is done in Lemma 9 and Lemma 10 in Appendix B.

□

### 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  $\varepsilon$  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  $\varepsilon$  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  $\varepsilon$  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  $\varepsilon$ . 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 explicit closed form of the tradeoff between the number of steps and  $\varepsilon$ , 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:

$$\mathbf{x}_{t+1} = \frac{A\mathbf{x}_t}{\|A\mathbf{x}_t\|_2}$$

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, presented in Appendix A.2, 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  $\ell_p$ -norm fails.



**Counterexamples for  $\|\cdot\|_p$ .** We show a counter example for  $\ell_2$  norm which directly extends to any  $\ell_p$  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  $f(\mathbf{x}) = \frac{(2x_1, x_2)}{\sqrt{4x_1^2 + x_2^2}}$ . We set  $\mathbf{x} = \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right)$  and  $\mathbf{y} = \left(\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\right)$ . We get that  $\|f(\mathbf{x}) - f(\mathbf{y})\|_2 = \left\| \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) - \left(\frac{2}{\sqrt{13}}, \frac{3}{\sqrt{13}}\right) \right\|_2 \geq 0.19$ . Also  $\|\mathbf{x} - \mathbf{y}\|_2 = \left\| \left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right) - \left(\frac{1}{\sqrt{10}}, \frac{3}{\sqrt{10}}\right) \right\|_2 \leq 0.14$  and therefore  $\|f(\mathbf{x}) - f(\mathbf{y})\|_2 > \|\mathbf{x} - \mathbf{y}\|_2$ .

Even though contraction is not achieved under any  $\ell_p$ -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  $(\lambda_1, \mathbf{v}_1), \dots, (\lambda_n, \mathbf{v}_n)$  such that  $\lambda_1 > \lambda_2 \geq \dots \geq \lambda_n$ . Then the power iteration,  $\mathbf{x}_{t+1} = f(\mathbf{x}_t) \triangleq \frac{A\mathbf{x}_t}{\|A\mathbf{x}_t\|}$  is contracting under the metric  $d(\mathbf{x}, \mathbf{y}) = \left\| \frac{\mathbf{x}}{\langle \mathbf{x}, \mathbf{v}_1 \rangle} - \frac{\mathbf{y}}{\langle \mathbf{y}, \mathbf{v}_1 \rangle} \right\|_2$  with contraction constant  $\lambda_2/\lambda_1$ , i.e. for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ :*

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

Moreover,  $t = \frac{\log(d(\mathbf{x}_0, \mathbf{v}_1)/\varepsilon)}{\log(\lambda_1/\lambda_2)}$  iterations suffice to have  $\|\mathbf{x}_t - \mathbf{v}_1\|_2 \leq d(\mathbf{x}_t, \mathbf{v}_1) \leq \varepsilon$ .

*Proof.* For any vector  $\mathbf{x}$ , it holds that  $\langle A\mathbf{x}, \mathbf{v}_1 \rangle = \lambda_1 \langle \mathbf{x}, \mathbf{v}_1 \rangle$ . We have that

$$\begin{aligned} d(f(\mathbf{x}), f(\mathbf{y})) &= \left\| \frac{A\mathbf{x}}{\langle A\mathbf{x}, \mathbf{v}_1 \rangle} - \frac{A\mathbf{y}}{\langle A\mathbf{y}, \mathbf{v}_1 \rangle} \right\|_2 \\ &= \frac{1}{\lambda_1} \left\| A \left( \frac{\mathbf{x}}{\langle \mathbf{x}, \mathbf{v}_1 \rangle} - \frac{\mathbf{y}}{\langle \mathbf{y}, \mathbf{v}_1 \rangle} \right) \right\|_2 \\ &\leq \frac{\lambda_1}{\lambda_1} \left\| \frac{\mathbf{x}}{\langle \mathbf{x}, \mathbf{v}_1 \rangle} - \frac{\mathbf{y}}{\langle \mathbf{y}, \mathbf{v}_1 \rangle} \right\|_2 = \frac{\lambda_2}{\lambda_1} d(\mathbf{x}, \mathbf{y}) \end{aligned}$$

where the inequality is true as the vector  $\frac{\mathbf{x}}{\langle \mathbf{x}, \mathbf{v}_1 \rangle} - \frac{\mathbf{y}}{\langle \mathbf{y}, \mathbf{v}_1 \rangle}$  is perpendicular to the principal eigenvector  $\mathbf{v}_1$ . 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  $\ell_2$  norm, we can see that at every step  $t > 0$ ,  $\|\mathbf{x}_t\|_2 = 1$ . If at some step  $t > 0$ , it holds that  $d(\mathbf{x}_t, \mathbf{v}_1) \leq \varepsilon$ , we get

$$\varepsilon^2 \geq d(\mathbf{x}_t, \mathbf{v}_1)^2 = \left\| \frac{\mathbf{x}_t}{\langle \mathbf{x}_t, \mathbf{v}_1 \rangle} - \mathbf{v}_1 \right\|_2^2 = \langle \mathbf{x}_t, \mathbf{v}_1 \rangle^{-2} - 1 \Rightarrow \langle \mathbf{x}_t, \mathbf{v}_1 \rangle \geq (1 + \varepsilon^2)^{-1/2}.$$

This implies that  $\|\mathbf{x}_t - \mathbf{v}_1\|_2^2 = 2(1 - \langle \mathbf{x}_t, \mathbf{v}_1 \rangle) \leq 2(1 - (1 + \varepsilon^2)^{-1/2}) \leq \varepsilon^2$ . This guarantees that bounding the norm  $d$  by  $\varepsilon$  implies a bound of  $\varepsilon$  on the  $\ell_2$  norm between the principal eigenvector and the current iterate  $\mathbf{x}_t$ .

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

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  $\ell_2$  norm.

## 5 Banach is Complete for CLS

As discussed in Section 1, the complexity class CLS was defined in [DP11] 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  $f : [0, 1]^3 \rightarrow [0, 1]^3$ ,  $p : [0, 1]^3 \rightarrow [0, 1]$ , both represented as arithmetic circuits, and two rational positive constants  $\varepsilon$  and  $\lambda$ . 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  $\lambda$ -Lipschitz continuity of  $f$ , i.e.
 
$$|f(x) - f(x')|_1 > \lambda|x - x'|_1.$$
- (CO3) two points  $x, x'$  violating the  $\lambda$ -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 [DP11], both the choice of domain  $[0, 1]^3$  and the use of  $\ell_1$  norm in the definition of the above problem are not crucial, and high-dimensional polytopes as well as other  $\ell_p$  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 [DP11], [DGP09] and [EY07]. 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** ([DP11]). CONTRACTION MAP takes as input a function  $f : [0, 1]^3 \rightarrow [0, 1]^3$  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  $\lambda$ -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  $f : [0, 1]^3 \rightarrow [0, 1]^3$  and  $d : [0, 1]^3 \times [0, 1]^3 \rightarrow \mathbb{R}$  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  $\lambda$ -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  $\lambda$ -Lipschitz continuity of  $d(\cdot, \cdot)$ , i.e.  $|d(x_1, x_2) - d(y_1, y_2)| > \lambda(|x_1 - y_1|_1 + |x_2 - y_2|_1)$ .

**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 contracting 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 Appendix C.

*Proof Sketch.* Since the inclusion to CLS is a simple argument very similar to the argument from [DP11] that shows that CONTRACTION MAP belongs to CLS, we focus here on the hardness proof.

We are given two functions  $f : [0, 1]^3 \rightarrow [0, 1]^3$ ,  $p : [0, 1]^3 \rightarrow [0, 1]$  and we want to find a contraction  $d : [0, 1]^3 \times [0, 1]^3 \rightarrow \mathbb{R}$  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  $\varepsilon$ -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 than 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 than  $\text{poly}(1/\varepsilon)$  steps which cannot hold given that  $\text{CLS} \neq \text{FP}$ .  $\square$

## Acknowledgements

The authors were supported by NSF CCF-1551875, CCF-1617730, CCF-1650733, and a Simons Graduate Research Fellowship.

## References

- [ABPW17] Omer Angel, Sébastien Bubeck, Yuval Peres, and Fan Wei. Local max-cut in smoothed polynomial time. In *Symposium on Theory of Computing (STOC)*, 2017.
- [Bes59] C. Bessaga. On the converse of banach "fixed-point principle". *Colloquium Mathematicae*, 7(1):41–43, 1959.
- [BPR15] Nir Bitansky, Omer Paneth, and Alon Rosen. On the cryptographic hardness of finding a nash equilibrium. In *Foundations of Computer Science (FOCS), 2015 IEEE 56th Annual Symposium on*, pages 1480–1498. IEEE, 2015.
- [BTN01] Aharon Ben-Tal and Arkadi Nemirovski. *Lectures on modern convex optimization: analysis, algorithms, and engineering applications*. SIAM, 2001.
- [BV04] Stephen Boyd and Lieven Vandenberghe. *Convex optimization*. Cambridge university press, 2004.
- [CDT09] Xi Chen, Xiaotie Deng, and Shang-Hua Teng. Settling the complexity of computing two-player nash equilibria. *Journal of the ACM (JACM)*, 56(3):14, 2009.
- [Con14] Keith Conrad. The contraction mapping theorem. *Expository paper. University of Connecticut, College of Liberal Arts and Sciences, Department of Mathematics*, 2014.
- [DGP09] Constantinos Daskalakis, Paul W Goldberg, and Christos H Papadimitriou. The complexity of computing a nash equilibrium. *SIAM Journal on Computing*, 39(1):195–259, 2009.
- [DP11] Constantinos Daskalakis and Christos H. Papadimitriou. 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*, pages 790–804, 2011.
- [ER17] Michael Etscheid and Heiko Röglin. Smoothed analysis of local search for the maximum-cut problem. *ACM Transactions on Algorithms (TALG)*, 2017.
- [EY07] Kousha Etessami and Mihalis Yannakakis. On the complexity of nash equilibria and other fixed points. In *Foundations of Computer Science (FOCS)*. IEEE, 2007.
- [FGMS17] John Fearnley, Spencer Gordon, Ruta Mehta, and Rahul Savani. Cls: New problems and completeness. *arXiv preprint arXiv:1702.06017*, 2017.

- [FPT04] Alex Fabrikant, Christos Papadimitriou, and Kunal Talwar. The complexity of pure nash equilibria. In *Symposium on Theory of Computing (STOC)*. ACM, 2004.
- [Har14] Moritz Hardt. Understanding alternating minimization for matrix completion. In *Foundations of Computer Science (FOCS), 2014 IEEE 55th Annual Symposium on*, pages 651–660. IEEE, 2014.
- [HY17] Pavel Hubacek and Eylon Yogev. Hardness of continuous local search: Query complexity and cryptographic lower bounds. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 1352–1371. SIAM, 2017.
- [JPY88] David S. Johnson, Christos H. Papadimitriou, and Mihalis Yannakakis. How easy is local search? *J. Comput. Syst. Sci.*, 37(1):79–100, 1988.
- [KNY17] Ilan Komargodski, Moni Naor, and Eylon Yogev. White-box vs. black-box complexity of search problems: Ramsey and graph property testing. In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 24, page 15, 2017.
- [Kör10] T Körner. Metric and topological spaces, 2010.
- [LSJR16] Jason D. Lee, Max Simchowitz, Michael I. Jordan, and Benjamin Recht. Gradient descent only converges to minimizers. In *Proceedings of the 29th Conference on Learning Theory, COLT 2016, New York, USA, June 23-26, 2016*, pages 1246–1257, 2016.
- [Mey67] Philip R. Meyers. 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):73, apr 1967.
- [Nes13] Yurii Nesterov. *Introductory lectures on convex optimization: A basic course*, volume 87. Springer Science & Business Media, 2013.
- [Pap94] Christos H. Papadimitriou. On the complexity of the parity argument and other inefficient proofs of existence. *J. Comput. Syst. Sci.*, 48(3):498–532, 1994.
- [PP16] Ioannis Panageas and Georgios Piliouras. Gradient descent converges to minimizers: The case of non-isolated critical points. *CoRR*, abs/1605.00405, 2016.
- [RSS16] Alon Rosen, Gil Segev, and Ido Shahaf. Can ppad hardness be based on standard cryptographic assumptions? In *Electronic Colloquium on Computational Complexity (ECCC)*, volume 23, page 59, 2016.
- [Rub16] Aviad Rubinfeld. Settling the complexity of computing approximate two-player nash equilibria. In *Foundations of Computer Science (FOCS)*. IEEE, 2016.
- [SV08] Alexander Skopalik and Berthold Vöcking. Inapproximability of pure nash equilibria. In *Proceedings of the fortieth annual ACM symposium on Theory of computing*, pages 355–364. ACM, 2008.

# A Preliminaries

## A.1 Basic Definitions

**Topological Spaces** Let  $\mathcal{D}$  be a set and  $\tau$  a collection of subsets of  $\mathcal{D}$  with the following properties.

- (a) The empty set  $\emptyset \in \tau$  and the space  $\mathcal{D} \in \tau$ .
- (b) If  $U_a \in \tau$  for all  $a \in A$  then  $\bigcup_{a \in A} U_a \in \tau$ .
- (c) If  $U_j \in \tau$  for all  $1 \leq j \leq n \in \mathbb{N}$ , then  $\bigcap_{j=1}^n U_j \in \tau$ .

Then we say that  $\tau$  is a *topology* on  $\mathcal{D}$  and that  $(\mathcal{D}, \tau)$  is a *topological space*. We call *open sets* the members of  $\tau$ . Also a subset  $C$  of  $\mathcal{D}$  is called *closed* if  $\mathcal{D} \setminus C$  is an open set, i.e. belongs to  $\tau$ . Let  $(\mathcal{D}, \tau)$  be a topological space and  $A$  a subset of  $\mathcal{D}$ . We write

$$\begin{aligned} \text{Int}(A) &= \bigcup \{U \in \tau \mid U \subseteq A\} \\ \text{Clos}(A) &= \bigcap \{U \text{ closed} \mid A \subseteq U\} \end{aligned}$$

and we call  $\text{Clos}(A)$  the *closure* of  $A$  and  $\text{Int}(A)$  the *interior* of  $A$ . We now give a basic lemma without proof. A proof can be found in [Kör10].

**Lemma 1.** (i)  $\text{Int}(A) = \{x \in A \mid \exists U \in \tau \text{ with } x \in U \subseteq A\}$ .  
(ii)  $\text{Clos}(A) = \{x \in \mathcal{D} \mid \forall U \in \tau \text{ with } x \in U, \text{ we have } U \cap A \neq \emptyset\}$ .

**Metric Spaces** The *diameter* of a set  $W \subseteq \mathcal{D}$  according to the metric  $d$  is defined as

$$\text{diam}_d[W] = \max_{x, y \in W} d(x, y)$$

A metric space  $(\mathcal{D}, d)$  is called *bounded* if  $\text{diam}_d[\mathcal{D}]$  is finite. We define  $d_S : \mathcal{D}^2 \rightarrow \mathbb{R}$  by

$$d_S(x, y) = \begin{cases} 0 & x = y \\ 1 & x \neq y \end{cases}$$

then  $d_S$  is called the *discrete metric* on  $\mathcal{D}$ .

**Remark.** It is very easy to see that discrete metric is indeed a metric, i.e. it satisfies the conditions (i)-(iv).

Let  $(\mathcal{D}, d)$  and  $(\mathcal{X}, d')$  be metric spaces. A function  $f : \mathcal{D} \rightarrow \mathcal{X}$  is called *continuous* if, given  $x \in \mathcal{D}$  and  $\varepsilon > 0$ , we can find a  $\delta(x, \varepsilon) > 0$  such that

$$d'(f(x), f(y)) < \varepsilon \text{ whenever } d(x, y) < \delta(x, \varepsilon)$$

We say that a subset  $E \subseteq \mathcal{D}$  is *open* in  $\mathcal{D}$  if, whenever  $e \in E$ , we can find a  $\delta > 0$  (depending on  $e$ ) such that

$$x \in E \text{ whenever } d(x, e) < \delta$$

The next lemma connects the definition of open sets according to some metric with the definition of open sets in a topological space.

**Lemma 2.** *If  $(\mathcal{D}, d)$  is a metric space, then the collection of open sets forms a topology.*

We define the *open ball* of radius  $r$  around  $x$  to be  $B(x, r) = \{y \in \mathcal{D} \mid d(x, y) < r\}$ .

**Closed Sets for Metric Spaces** Consider a sequence  $(x_n)$  in a metric space  $(\mathcal{D}, d)$ . If  $x \in \mathcal{D}$  and, given  $\varepsilon > 0$ , we can find an integer  $N \in \mathbb{N}_1$  (depending maybe on  $\varepsilon$ ) such that

$$d(x_n, x) < \varepsilon \text{ for all } n \geq N$$

then we say that  $x_n \rightarrow x$  as  $n \rightarrow \infty$  and that  $x$  is the *limit* of the sequence  $(x_n)$ .

A set  $G \subseteq \mathcal{D}$  is said to be *closed* if, whenever  $x_n \in G$  and  $x_n \rightarrow x$  then  $x \in G$ . A proof of the following lemma can be found in [Kör10].

**Lemma 3.** *Let  $(\mathcal{D}, d)$  be a metric space and  $A$  a subset of  $\mathcal{D}$ . Then  $\text{Clos}(A)$  consists of all those  $x \in \mathcal{D}$  such that we can find  $(x_n)$  with  $x_n \in A$  with  $d(x_n, x) \rightarrow 0$ .*

We define the *closed ball* of radius  $r$  around  $x$  to be  $\bar{B}(x, r) = \{y \in \mathcal{D} | d(x, y) \leq r\}$ .

A subset  $G$  of a metric space  $(\mathcal{D}, d)$  is called *compact* if  $G$  is closed and every sequence in  $G$  has a convergent subsequence. A metric space  $(\mathcal{D}, d)$  is called compact if  $\mathcal{D}$  is compact, *locally compact* if for any  $x \in \mathcal{D}$ ,  $x$  has a neighborhood that is compact and *proper* if every closed ball is compact.

**Complete Metric Spaces** We say that a sequence  $(x_n)$  in  $\mathcal{D}$  is *Cauchy sequence* (or *d-Cauchy sequence* if the distance metric is not clear from the context) if, given  $\varepsilon > 0$ , we can find  $N(\varepsilon) \in \mathbb{N}_1$  with

$$d(x_n, x_m) < \varepsilon \text{ whenever } n, m \geq N(\varepsilon)$$

A metric space  $(\mathcal{D}, d)$  is *complete* if every Cauchy sequence converges.

Two metrics  $d, d'$  of the same set  $\mathcal{D}$  are called *topologically equivalent* (or just *equivalent*) if for every sequence  $(x_n)$  in  $\mathcal{D}$ ,  $(x_n)$  is  $d$ -Cauchy sequence if and only if it is  $d'$ -Cauchy sequence.

**Definition 5.** *Let  $(\mathcal{D}, d)$  be a metric spaces. A function  $f : \mathcal{D} \rightarrow \mathcal{D}$  is called *continuous with respect to  $d$* , if given  $x \in \mathcal{D}$  and  $\varepsilon > 0$ , we can find a  $\delta(x, \varepsilon)$  such that*

$$d'(f(x), f(y)) < \varepsilon \text{ whenever } d(x, y) < \delta(x, \varepsilon)$$

**Lipschitz Continuity** Let  $(\mathcal{D}, d)$  and  $(\mathcal{X}, d')$  be metric spaces. A function  $f : \mathcal{D} \rightarrow \mathcal{X}$  is *Lipschitz continuous* (or  $(d, d')$ -*Lipschitz continuous* if the distance metric is not clear from the context or  $d$ -*Lipschitz continuous* if  $d = d'$ ) if there exists a positive constant  $\lambda \in \mathbb{R}_+$  such that for all  $x, y \in \mathcal{D}$

$$d'(f(x), f(y)) \leq \lambda d(x, y)$$

**Lemma 4.** *If a function  $f : \mathcal{D} \rightarrow \mathcal{X}$  is Lipschitz continuous then it is continuous.*

**Definition 6.** *Let  $(\mathcal{D}, d)$  and  $(\mathcal{X}, d')$  be metric spaces. A function  $f : \mathcal{D} \rightarrow \mathcal{X}$  is *contraction* (or  $(d, d')$ -*contraction* or  $d$ -*contraction* if  $d = d'$ ) if there exists a positive constant  $1 > c \in \mathbb{R}_+$  such that for all  $x, y \in \mathcal{D}$*

$$d'(f(x), f(y)) \leq cd(x, y)$$

*If  $c = 1$  then we call  $f$  non-expansion.*

A *fixed point* of a selfmap  $f$  is any point  $x^* \in \mathcal{D}$  such that  $f(x^*) = x^*$ .

## A.2 Introduction to Power Method

Let  $A \in \mathbb{R}^{n \times n}$ . Recall that if  $q$  is an eigenvector for  $A$  with eigenvalue  $\lambda$ , then  $Aq = \lambda q$ , and in general,  $A^k q = \lambda^k q$  for all  $k \in \mathbb{N}$ . This observation is the foundation of the *power iteration method*.

Suppose that the set  $\{q_i\}$  of unit eigenvectors of  $A$  forms a basis of  $\mathbb{R}^n$ , and has corresponding set of real eigenvalues  $\{\lambda_i\}$  such that  $|\lambda_1| > |\lambda_2| > \dots > |\lambda_n|$ . Let  $v_0$  be an arbitrary initial vector, not perpendicular to  $q_1$ , with  $\|v_0\| = 1$ . We can write  $v_0$  as a linear combination of the eigenvectors of  $A$  for some  $c_1, \dots, c_n \in \mathbb{R}$  we have that

$$v_0 = c_1 q_1 + c_2 q_2 + \dots + c_n q_n$$

and since we assumed that  $v_0$  is not perpendicular to  $q_1$  we have that  $c_1 \neq 0$ .

Also

$$Av_0 = c_1 \lambda_1 q_1 + c_2 \lambda_2 q_2 + \dots + c_n \lambda_n q_n$$

and therefore

$$\begin{aligned} Av_k &= c_1 \lambda_1^k q_1 + c_2 \lambda_2^k q_2 + \dots + c_n \lambda_n^k q_n \\ &= \lambda_1^k \left( c_1 q_1 + c_2 \left( \frac{\lambda_2}{\lambda_1} \right)^k q_2 + \dots + c_n \left( \frac{\lambda_n}{\lambda_1} \right)^k q_n \right) \end{aligned}$$

Since the eigenvalues are assumed to be real, distinct, and ordered by decreasing magnitude, it follows that

$$\lim_{k \rightarrow \infty} \left( \frac{\lambda_i}{\lambda_1} \right)^k = 0$$

So, as  $k$  increases,  $A^k v_0$  approaches  $c_1 \lambda_1^k q_1$ , and thus for large values,

$$\frac{A^k v_0}{\|A^k v_0\|} \rightarrow q_1 \text{ as } k \rightarrow \infty$$

The power iteration method is simple and elegant, but suffers some drawbacks. Except from a measure 0 of initial conditions, the method returns a single eigenvector, corresponding to the eigenvalue of largest magnitude. In addition, convergence is only guaranteed if the eigenvalues are distinct—in particular, the two eigenvalues of largest absolute value must have distinct magnitudes. The rate of convergence primarily depends upon the ratio of these magnitudes, so if the two largest eigenvalues have similar sizes, then the convergence will be slow.

In spite of its drawbacks, the power method is still used in many applications, since it works well on large, sparse matrices when only a single eigenvector is needed. However, there are other methods overcoming some of the issues with the power iteration method.

## B Proof of Theorem 1

### B.1 Meyer's Construction and Our Contribution

The proof of Theorem 1, follows the construction of [Mey67]. We give a complete step by step description of this construction. For every step of this construction we explain what it was already proven by Meyers and we additionally prove some properties that are needed in order to satisfy our additional condition (2b).

The construction of  $d_{c,\varepsilon}$  starts with an open neighborhood of  $x^*$  with some desired properties. Meyers starts with an arbitrary open neighborhood  $W$  such that  $f(W) \subset W$ , whereas we also need that  $\text{diam}_d[W] \leq \varepsilon$ .



**Lemma 5.** *There exists an open neighborhood  $W$  of  $x^*$  such that*

$$f(W) \subseteq W \tag{3a}$$

$$\text{diam}_d[W] \leq \varepsilon \tag{3b}$$

*Proof.* From the hypothesis of the theorem there exists an open neighborhood  $U$  such that  $f^{[n]}(U) \rightarrow \{x^*\}$ . This implies that any open subset  $V$  of  $U$  satisfies  $f^{[n]}(V) \rightarrow \{x^*\}$ . Therefore we can choose a  $V = \text{Int}(\bar{B}(x^*, \varepsilon))$  such that  $\text{diam}_d[V] \leq \varepsilon$  and  $f^{[n]}(V) \rightarrow \{x^*\}$ . For simplicity of the notation we just assume refer to  $V$  as  $U$  and so  $\text{diam}_d[U] \leq \varepsilon$ .

Starting from  $U$  we prove the existence of  $W$ . For this, we will prove that there exists an open neighborhood  $W$  of  $x^*$  such that  $f(W) \subset W$  and  $W \subset U$ . The latter implies  $f^{[n]}(W) \rightarrow \{x^*\}$  and  $\text{diam}_d[W] \leq \varepsilon$ .

Since  $f^{[n]}(U) \rightarrow \{x^*\}$ , there is an integer  $k$  such that  $f^{[k]}(U) \subseteq U$ . Let

$$W = \bigcap_{j=0}^{k-1} f^{[-j]}(U) \subseteq U$$

Then for any  $x \in W$  and for any  $1 \leq j \leq k-1$  it holds that  $x \in f^{[-j]}(U)$  and thus  $f(x) \in f^{[-(j-1)]}(U)$ . Moreover  $x \in U$ , so that  $f^{[k]}(x) \in f^{[k]}(U) \subset U$  and thus  $f(x) \in f^{[-(k-1)]}(U)$ . Hence  $x \in W$  implies  $f(x) \in W$ , which was to be shown. The diameter of  $W$  can be bounded by the diameter of  $U$  and hence is less than  $\varepsilon$ . ■

We now proceed to the main line of the proof. The construction follows three steps:

- I. We first construct a metric  $d_M$ , which is topologically equivalent to  $d$ , and with respect to which  $f$  is non-expanding. It also holds that  $d_M(x, y) \geq d(x, y)$  for all  $x, y \in \mathcal{D}$  and therefore Property (2b) can be transferred from  $d_M$  to  $d$ .
- II. Given  $d_M$ , we proceed to construct a “distance” function  $\rho_{c, \varepsilon}$ , which satisfies (2a) and all the metric properties except maybe for the triangle inequality. Moreover  $\rho_{c, \varepsilon}$  satisfies that  $\rho_{c, \varepsilon}(x, y) \geq d_M(x, y)$  if  $\max\{d(x^*, x), d(x^*, y)\} \geq \varepsilon$ , and therefore (2b) is preserved.
- 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. Given the properties of  $\rho_{c, \varepsilon}$  and the definition of  $d_{c, \varepsilon}$ , we can prove that  $d_{c, \varepsilon}$  is a metric and Properties (2a) and (2b) hold.

### I. Construction of $d_M$

In the first step of the construction we define an metric  $d_M$  as

$$d_M(x, y) = \sup_{n \in \mathbb{N}} \{d(f^{[n]}(x), f^{[n]}(y))\}$$

and we show that  $f$  is non-expanding with respect to  $d_M$ .

**Lemma 6** ([Mey67]). *For the  $d_M$  function defined above we have that:*

1.  $d_M$  is well defined and satisfies all the metric properties (see Definition 1).
2.  $d_M$  is topologically equivalent with  $d$ .

The proof of Lemma 6 can be found in Section B.2 where for completeness we keep the proofs that were already proved by Meyers’s [Mey67].

For our purposes we also observe that by the definition of  $d_M$  the following holds

$$d_M(x, y) \geq d(x, y)$$

and hence if  $d_M(x, y) \leq \varepsilon$  then also  $d(x, y) \leq \varepsilon$ , therefore  $d_M$  satisfies (2b).

## II. Construction of $\rho_c$

We begin by defining  $K_n$  to be the closure of  $f^n(W)$  for  $n \geq 0$ , in particular we have that  $K_0 = W$  and hence by Lemma 5 we have that  $\text{diam}_d[K_0]$ . Also we define  $K_{(-n)} = f^{[-n]}(K_0)$ , so that our assumption  $f^{[n]}(W) \rightarrow \{x^*\}$  implies

$$K_n \rightarrow \{x^*\} \quad \text{as } n \rightarrow \infty. \quad (4)$$

For  $x \in K_0 \setminus \{x^*\}$  we set  $n(x) = \max_{x \in K_n} \{n\} \geq 0$ . The fact that  $n(x)$  is finite is guaranteed by (4). To see this assume that there is an infinite sequence  $n_1, n_2, \dots$  such that  $x \in K_{n_i}$  which implies that  $x \in \bigcap_{i=1}^{\infty} K_i$  which definitely contradicts (4). We define also  $n(x^*) = \infty$ , and for  $x \in \mathcal{D} \setminus K_0$  set  $n(x) = -\min_{f^{[m]}(x) \in K_0} \{m\} = \max_{x \in K_n} \{n\} < 0$  which again is finite because of condition 2. Let  $\kappa(x, y) = \min\{n(x), n(y)\}$ , we define  $\rho_c$  to be

$$\rho_c(x, y) = c^{\kappa(x, y)} d_M(x, y).$$

**Lemma 7** ([Mey67]). *For the  $\rho_c$  function defined above we have that:*

1.  $\rho_c$  is well defined and satisfies the metric properties (see Definition 1), except the triangle inequality (iv).
2.  $f$  is a contraction map with respect to  $\rho_c$  with contraction constant  $c$ .

$$\rho_c(f(x), f(y)) \leq c \cdot \rho_c(x, y). \quad (5)$$

The proof of Lemma 7 is almost immediate from the definition of  $\rho_c$ , but for a detailed explanation we refer to the initial proof by Meyers [Mey67].

## III. Construction of $d_c$

In this last step what we do is that we assign the distance between two points to be the length of the shortest path that connects these two points, with the lengths computed according to  $\rho_c$ . Then the distance satisfies the triangle inequality because of the shortest path property.

Formally, denote by  $S_{xy}$  the set of chains  $s_{xy} = (x = x_0, x_1, \dots, x_m = y)$  from  $x$  to  $y$  with associated lengths  $L_c(s_{xy}) = \sum_{i=1}^m \rho_c(x_i, x_{i-1})$ . We define

$$d_c(x, y) = \inf\{L_c(s_{xy}) \mid s_{xy} \in S_{xy}\}. \quad (6)$$

**Lemma 8** ([Mey67]). *For the  $d_c$  function defined above we have that:*

1.  $d_c$  is well defined and satisfies all the metric properties (see Definition 1).
2.  $f$  is a contraction map with respect to  $d_c$  with contraction constant  $c$ .

$$d_c(f(x), f(y)) \leq c \cdot d_c(x, y). \quad (7)$$

3.  $d_c$  is topologically equivalent with  $d$  and hence  $(\mathcal{D}, d_c)$  is a complete metric space.

The proof of Lemma 8 can be found in Section B.2.

We now need to prove two lemmas to prove that Meyers's construction also satisfies (2b).

**Lemma 9.** *Consider any  $x \neq x^*$  and  $y \neq x, x^*$  and assume that  $y \notin K_0$ , then*

$$d_c(x, y) \geq \min\{d_M(x, y), d_M(x, K_0)\} > 0. \quad (8)$$

*Proof of Lemma 9:* By definition any chain  $s_{xy}$  either lies in  $\mathcal{D} \setminus K_0$  entirely, or has a last link which leaves  $K_0$ . If  $s_{xy}$  lies in  $\mathcal{D} \setminus K_0$  entirely then  $n(x), n(y) < 0$  and hence  $d_c(x, y) \geq d_M(x, y)$ .

Otherwise we consider the last link that leaves  $K_0$  and we have that the length between  $x$  and  $y$  according to  $d_c$  is greater than the distance with respect to  $d_M$  of  $x$  from  $K_0$  which gives that

$$d_M(x, K_0) \leq d_c(x, y).$$

□

The final step is to prove (2b).

**Lemma 10.**

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

*Proof of Lemma 10:* Let  $A = \text{diam}_d [\bar{B}(x^*, 2\varepsilon)]$  and without loss of generality  $d(x, x^*) \geq d(y, x^*)$ .

If either  $d_M(x, x^*) \leq \varepsilon$  or  $d_M(y, x^*) \leq \varepsilon$  then we are done since as we have seen in the construction of  $d_M$ ,  $d_M(x, y) \geq d(x, y)$ , thus either  $d(x, x^*) \leq \varepsilon$  or  $d(y, x^*) \leq \varepsilon$  and (2b) is satisfied. So we may assume that  $d(x, x^*) \geq \varepsilon$  and  $d(y, x^*) \geq \varepsilon$ . Therefore  $x, y \in \mathcal{D} \setminus K_0$  and which translates to  $n(x), n(y) < 0$ . So now using Lemma 9 and we get

$$d_c(x, y) \geq \min\{d_M(x, y), d_M(x, K_0)\}. \quad (9)$$

Now we consider two cases according to the value of  $d_M(x, K_0)$ . If  $d_M(x, K_0) \geq \varepsilon$  then

$$d_c(x, y) \leq \varepsilon \implies d_M(x, y) \leq \varepsilon \leq A \implies d(x, y) \leq \varepsilon.$$

Otherwise if  $d_M(x, K_0) \leq \varepsilon$  then  $d(x, K_0) \leq \varepsilon$  and by triangle inequality  $d(x, x^*) \leq 2\varepsilon$ . By our assumption for the relative position of  $x$  and  $y$  we also get  $d(y, x^*) \leq 2\varepsilon$  and therefore  $x, y \in \bar{B}(x^*, 2\varepsilon)$ . Thus,  $d(x, y) \leq \text{diam}_d [\bar{B}(x^*, 2\varepsilon)]$ . □

□

## B.2 Omitted Proof of Lemmas proven by Meyers in [Mey67]

*Proof of Lemma 6:* The fact that this maximum is finite can be proved using the condition 2. of the theorem. Indeed, since  $d(f^{[n]}(x), x^*) \rightarrow 0$  and  $d(f^{[n]}(y), x^*) \rightarrow 0$ , for any  $\delta > 0$  there is a number  $N \in \mathbb{N}$  such that  $d(f^{[n]}(x), x^*) \leq \delta$  and  $d(f^{[n]}(y), x^*) \leq \delta$  for all  $n > N$ . Now if let  $\delta = d(x, y)$  we get that  $\max_{n \geq N} \{d(f^{[n]}(x), f^{[n]}(y))\} \leq d(x, y)$  and therefore  $\max_{n \in \mathbb{N}} \{d(f^{[n]}(x), f^{[n]}(y))\} = \max_{0 \leq n \leq N} \{d(f^{[n]}(x), f^{[n]}(y))\}$ . Hence the maximum has a finite value. Observe also that by definition it holds that

$$d_M(f(x), f(y)) \leq d_M(x, y)$$

and hence  $f$  is a non-expansion according to  $d_M$ . It only remains to prove that  $d_M$  satisfies the properties of a metric function. The positive definiteness and symmetry of  $d_M$  follow immediately from the corresponding properties of  $d$ . The fact that  $d_M(x, y) \neq 0$  for  $x \neq y$  follows from the fact that  $d(x, y) \leq d_M(x, y)$ , which follows directly from the definition of  $d_M$  since  $f^{[0]}(x) = x$ . It remains to prove the triangle inequality. For this we observe that by the definition of  $d_M$  and using the fact that the maximum in the definition of  $d_M$  for any  $x, y \in \mathcal{D}$  is finite, there exists an  $n \in \mathbb{N}$  such that

$$\begin{aligned} d_M(x, z) &= d(f^{[n]}(x), f^{[n]}(z)) \leq \\ &\leq d(f^{[n]}(x), f^{[n]}(y)) + d(f^{[n]}(y), f^{[n]}(z)) \leq \\ &\leq d_M(x, y) + d_M(y, z). \end{aligned}$$

Thus  $d_M$  is indeed a metric. We now show that  $d_M$  is topologically equivalent to  $d$ . From the inequality  $d(x, y) \leq d_M(x, y)$  it follows that any  $d_M$ -convergent sequence is also  $d$ -convergent, with the same limit point. To prove the implication in the opposite direction, note that condition 2. of the hypotheses of the theorem implies the existence for each  $\eta > 0$  of an  $N$  such that

$$\text{diam}_d \left[ f^{[n]}(W) \right] < \eta \quad \text{for } n > N.$$

For each  $x \in \mathcal{D}$ , it follows from 2. that

$$\nu(x) = \min_{n \in \mathbb{N}, f^{[n]}(x) \in W} \{n\} \tag{10}$$

is finite. Since  $f$  is continuous, there is an  $\delta > 0$  so small that  $d(x, y) < \delta$  implies

$$f^{[\nu(x)]}(y) \in W \text{ and } d(f^{[j]}(x), f^{[j]}(y)) < \eta \text{ for } 0 \leq j \leq N + \nu(x). \tag{11}$$

By (3a) we have  $f^{[n+N+\nu(x)]}(x) \in f^{[n+N]}(W)$  and  $f^{[n+N+\nu(x)]}(y) \in f^{[n+N]}(W)$  for all  $n > 0$ , so that the (11) implies

$$d(f^{[j]}(x), f^{[j]}(y)) < \eta \quad \text{for } j > N + \nu(x).$$

Thus  $d(x, y) \leq \delta$  implies  $d_M(x, y) \leq \eta$ . This shows that a sequence which is  $d$ -convergent to  $x$  is also  $d_M$ -convergent to  $x$ , completing the proof of topological equivalence. Finally since  $d$  and  $d_M$  are topologically equivalent and  $d$  is complete for  $\mathcal{D}$  it follows that  $d_M$  is also complete for  $\mathcal{D}$ .  $\square$

*Proof of Lemma 8:* We will prove that  $d_c$  is the desired metric. That  $f$  is a contraction with constant  $c$  with respect to  $d_c$  follows by applying (7) to the links  $[x_{i-1}, x_i]$  of any chain  $s_{xy}$ . Clearly  $d_c$  is symmetric and  $d_c(x, x) = 0$ . The triangle law holds since following a  $s_{xy}$  with a  $s_{yz}$  yields a  $s_{xz}$ . It remains to show that it is positive definite.

Consider any  $x \neq x^*$  and  $y \neq x$  and assume  $n(x) \leq n(y)$  without loss of generality. If  $y \neq x^*$ , any chain  $s_{xy}$  either lies in  $\mathcal{D} \setminus K_{n(y)+1}$ , or has a last link which leaves  $K_{n(y)+1}$ , so that

$$d_c(x, y) \geq c^{n(y)} \min\{d_M(x, y), d_M(x, K_{n(y)+1})\} > 0. \tag{12}$$

The remaining case,  $y = x^*$  is covered by

$$d_c(x, y) \geq c^{n(x)} d_M(x, K_{n(x)+1}) > 0. \tag{13}$$

Thus  $d_c$  is a distance metric. We now have to prove that  $d_c$  is equivalent to  $d_M$ . Let  $B_\nu = \mathcal{D} \setminus f^{[-\nu]}(W)$  for  $\nu \geq 0$ , so that the definition of  $\nu(x)$  (10) implies  $d_M(x, B_{\nu(x)}) > 0$  and  $n(x) \geq -\nu(x)$ . For any  $x \neq x^*$ , if  $y$  obeys

$$d_M(x, y) < \delta(x) = \min\{d_M(x, K_{n(x)+1}), d_M(x, B_{\nu(x)})\} \tag{14}$$

then  $n(x) \geq -\nu(x)$ , so that (6) and (12), the last with  $x$  and  $y$  interchanged, imply

$$c^{n(x)} d_M(x, y) \leq d_c(x, y) \leq \rho_c(x, y) \leq c^{-\nu(x)} d_M(x, y). \tag{15}$$

Now choose  $k(x) > \max\{0, n(x)\}$  such that  $z \in K_{k(x)}$  implies  $d_M(z, x^*) < d_c(x, x^*)/2$ . Then  $d_c(x, K_{k(x)}) \geq d_c(x, x^*)/2$ , so that if  $y$  obeys

$$d_c(x, y) < d_c(x, x^*)/2 \tag{16}$$

then only chains disjoint from  $K_{k(x)}$  need enter (6), implying

$$d_c(x, y) \geq c^{k(x)} d_M(x, y). \quad (17)$$

In particular, if

$$d_c(x, y) < \min\{d_c(x, x^*)/2, c^{k(x)} \delta(x)\}$$

then with (16) and (17) this implies (14) and hence (15) applies. Thus  $d_c(x_n, x) \rightarrow 0$  whenever  $d_M(x_n, x) \rightarrow 0$ .

Now if  $x = x^*$ , note first that if  $d_M(x^*, y) < d_M(x^*, B_0)$ , then

$$d_c(x^*, y) \leq \rho_c(x^*, y) \leq d_M(x^*, y). \quad (18)$$

Also note that for any  $\eta > 0$ ,  $f^{[n]}(W) \rightarrow \{x^*\}$  guarantees an  $N(\eta) > 0$  such that  $d_M(x^*, z) < \eta/2$  for all  $z \in K_{N(\eta)}$ . Then  $d_M(x^*, y) > \eta$  implies that  $d_M(y, K_{N(\eta)}) \geq \eta/2$  and thus that

$$d_c(x^*, y) \geq d_c(K_{N(\eta)}, y) \geq c^{N(\eta)} \eta/2$$

Hence  $d_c(x_n, x^*) \rightarrow 0$  if and only if  $d_M(x_n, x^*) \rightarrow 0$ .

To show that  $d_M$ -completeness is preserved, assume that  $(x_n)$  is a  $d_c$ -Cauchy sequence and that  $(X, d_M)$  is complete. If  $(x_n)$  does not converge to  $x^*$  then since  $d_c$  and  $d_M$  are equivalent, for some  $N \in \mathbb{N}$  and all sufficiently large  $n$ ,  $n(x_n) < N$ .

Now exactly as above choose  $k((x_n)) = P > \max 0, N$  such that  $z \in K_{k((x_n))}$  implies

$$d_M(x^*, z) < \inf_{i \in \mathbb{N}} \left\{ \frac{d_c(x_i, x^*)}{2} \right\} = \frac{R}{2}$$

then since  $(x_n)$  is a Cauchy sequence there is an  $i \in \mathbb{N}_1$  such that

$$d_c(x_p, x_{p+j}) < \frac{R}{2}$$

for all  $p > i$ , and using (17) with  $k(x) = P$ , we have

$$c^{-P} d_c(x_p, x_{p+j}) \geq d_M(x_p, x_{p+j})$$

so that  $(x_n)$  is a  $d_M$ -Cauchy sequence. Therefore since  $(\mathcal{D}, d_M)$  is complete, the topological space  $(\mathcal{D}, d_c)$  is complete too.  $\square$

## C Proof of Theorem 2

On route to establishing the CLS-completeness of BANACH, we will define an intermediate, syntactic problem METRICBANACH, which is similar to BANACH except that the function  $d$  given in the input is not promised to be a metric, and hence a violation of the metricity of  $d$  is accepted as a solution.

**Definition 7.** METRICBANACH takes as input two functions  $f : [0, 1]^3 \rightarrow [0, 1]^3$ ,  $d : [0, 1]^3 \times [0, 1]^3 \rightarrow \mathbb{R}$ , both represented as arithmetic circuits, 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  $\lambda$ -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  $\lambda$ -Lipschitz continuity of  $d(\cdot, \cdot)$ , i.e.  $|d(x_1, x_2) - d(y_1, y_2)| > \lambda(|x_1 - y_1| + |x_2 - y_2|)$ .
- (Oe) points  $x, y, z \in [0, 1]^3$  violating any of the metric properties of  $d$  ((i)-(iv) of Definition 1).

Notice that METRICBANACH is syntactic, namely for any input there exists a solution. We proceed to show that the problem is CLS-complete.

**Theorem 3.** METRICBANACH is CLS-complete.

*Proof of Theorem 3.* We first show that METRICBANACH belongs to CLS even when we disallow (Oe). Starting from an instance  $(f, d, \varepsilon, \lambda, c)$  of CONTINUOUS LOCALOPT we create the following instance

$$(f'(x) = f(x), p(x) = d(x, f(x)), \varepsilon' = (1 - c) \cdot \varepsilon, \lambda' = \lambda)$$

Now we have to show that any output of the CONTINUOUS LOCALOPT with input  $(f, p, \varepsilon', \lambda)$  will give us a output of METRICBANACH with input  $(f, d, \varepsilon, \lambda, c)$ .

(CO1)  $\implies$  If  $d(f(x), f(f(x))) > c \cdot d(x, f(x))$  then  $(x, f(x))$  satisfies (Ob). Otherwise

$$\begin{aligned} p(f(x)) &\geq p(x) - \varepsilon' \implies d(f(x), f(f(x))) \geq d(x, f(x)) - \varepsilon' \implies \\ c \cdot d(x, f(x)) &\geq d(f(x), f(f(x))) \geq d(x, f(x)) - \varepsilon' \implies \\ c \cdot d(x, f(x)) &\geq d(x, f(x)) - (1 - c) \cdot \varepsilon \implies \\ (1 - c) \cdot d(x, f(x)) &\leq (1 - c) \cdot \varepsilon \implies \\ d(x, f(x)) &\leq \varepsilon \end{aligned}$$

Therefore  $x$  satisfies (Oa) and therefore is a solution of METRICBANACH.

(CO2)  $\implies$  (Oc).

(CO3)  $\implies$  Without loss of generality let  $\|x - f(x)\| \leq \|y - f(y)\|$ . If  $x = f(x)$  then if  $d(x, f(x)) = 0$  we immediately satisfy (Oa) otherwise we satisfy (Oe). Otherwise we can give  $x_1 = x, x_2 = f(x), y_1 = y, y_2 = f(y)$  and since  $x_1 \neq x_2, y_1 \neq y_2$  we satisfy (ii) of (Od).

This implies that any output of CONTINUOUS LOCALOPT at the instance  $(f', p, \varepsilon', \lambda')$  can produce an output to the instance  $(f, d, \varepsilon, \lambda, c)$  of the METRICBANACH problem. Therefore METRICBANACH  $\in$  CLS.

Now we are going to show the opposite direction and reduce CONTINUOUS LOCALOPT to METRICBANACH. Starting from an instance  $(f, p, \varepsilon, \lambda)$  of CONTINUOUS LOCALOPT we define for any  $x, y \in [0, 1]^3$ ,

$$\kappa(x, y) = \min \left\{ -\frac{p(x)}{\varepsilon}, -\frac{p(y)}{\varepsilon} \right\}$$

We also remind the reader the definition of the *discrete metric*

$$d_S(x, y) = 1 \text{ if } x \neq y \text{ and } d_S(x, x) = 0$$

Finally we define the *smooth interpolation function* for  $w \leq 0$

$$B(w) = (1 - ([w] - w))c^{\lceil w \rceil} + (([w] - w))c^{\lceil w \rceil + 1}$$

The basic observation about  $B(\cdot)$  since  $c < 1$  is that  $c^{\lceil \kappa(x, y) \rceil + 1} \leq B(\kappa(x, y)) \leq c^{\lceil \kappa(x, y) \rceil}$ .

Based on these definitions we create the following instance of METRICBANACH

$$f' = f, d(x, y) = B(\kappa(x, y)) \cdot d_S(x, y), \varepsilon' = \frac{1}{c}, \lambda' = \max \left\{ \lambda, \left[ c^{-1/\varepsilon} \lambda \frac{\ln(1/c)}{\varepsilon} \right] \right\}, c = 1 - 0.1\varepsilon$$

As in the previous reduction we have to show that any result of the METRICBANACH with input  $(f, d, \varepsilon', \lambda, c)$  will give us a result of CONTINUOUS LOCALOPT with input  $(f, p, \varepsilon, \lambda)$ .

(Oa)  $\implies$  If  $p(f(x)) \geq p(x)$  then  $x$  satisfies (CO1). Otherwise we can see that  $\kappa(x, f(x)) = -p(x)/2\varepsilon$  and  $x \neq f(x)$  so

$$\begin{aligned} d(x, f(x)) \leq \varepsilon' \implies B(\kappa(x, y)) \leq \varepsilon' \implies \left(\frac{p(x)}{\varepsilon}\right) \log(1/c) \leq \log(\varepsilon') \implies \frac{p(x)}{\varepsilon} \leq \frac{\log(\varepsilon')}{\log(1/c)} \\ \implies p(x) \leq \varepsilon \end{aligned}$$

so  $p(f(x)) \geq 0 \geq p(x) - \varepsilon$  and so  $x$  satisfies (CO1).

(Ob)  $\implies$  As in the previous case we may assume that  $p(f(x)) \leq p(x) - \varepsilon$  and that  $p(f(y)) \leq p(y) - \varepsilon$ . Without loss of generality we can assume that  $p(x) > p(y)$ . If also  $p(f(x)) \geq p(f(y))$  then  $\kappa(x, y) = -p(x)/\varepsilon$  and  $\kappa(f(x), f(y)) = -p(f(x))/\varepsilon$ . Therefore

$$d(x, y) = B(\kappa(x, y)), \quad d(f(x), f(y)) = B(\kappa(f(x), f(y)))$$

Now if (Ob) is satisfied then

$$\begin{aligned} c^{-\lfloor \frac{p(f(x))}{\varepsilon} \rfloor} \geq d(f(x), f(y)) = B(\kappa(f(x), f(y))) > c \cdot B(\kappa(x, y)) = c \cdot d(x, y) \geq c^2 \cdot c^{-\lfloor \frac{p(x)}{\varepsilon} \rfloor} \\ \implies \left\lfloor \frac{p(f(x))}{\varepsilon} \right\rfloor \geq \left\lfloor \frac{p(x)}{\varepsilon} \right\rfloor - 2 \implies p(f(x)) \geq p(x) - 2\varepsilon \quad 4 \end{aligned}$$

Therefore  $x$  satisfies (CO1).

Now similarly if  $p(f(y)) > p(f(x))$  then  $p(f(y)) > p(x) - \varepsilon$ . But by our assumption that  $p(x) > p(y)$  we get  $p(f(y)) > p(y) - \varepsilon$ . Therefore  $y$  satisfies (CO1).

(Oc)  $\implies$  (CO2).

(Od)  $\implies$  We will analyze the function  $h(x) = c^{-x}$  when  $x \in [0, 1/\varepsilon]$ . By the mean value theorem we have that the Lipschitz constant  $\ell_h$  of  $h$  is less than  $\max_{x \in [0, 1/\varepsilon]} h'(x)$ . But

$$h'(x) = \left(e^{-x \ln c}\right)' = \ln(1/c) c^{-x}$$

and because  $c < 1$  we have that  $\max_{x \in [0, 1/\varepsilon]} h'(x) = c^{-1/\varepsilon} \ln(1/c)$ .

Let now  $\kappa(x_1, x_2) = -p(x_1)/\varepsilon$  and  $\kappa(y_1, y_2) = -p(y_1)/\varepsilon$ . Since  $x_1 \neq x_2$  and  $y_1 \neq y_2$  we have  $d(x_1, x_2) = B(c^{-p(x_1)/\varepsilon})$  and  $d(y_1, y_2) = B(c^{-p(y_1)/\varepsilon})$ . Since  $B(\kappa(x, y))$  is just an linear interpolation of points that belong to  $c^{\kappa x, y}$  using the Mean Value Theorem we have that  $|B(\kappa(x_1, x_2)) - B(\kappa(y_1, y_2))| \leq \max_{x \in [0, 1/\varepsilon]} h'(x) \left| \frac{p(x_1)}{\varepsilon} - \frac{p(y_1)}{\varepsilon} \right|$

$$\begin{aligned} |d(x_1, x_2) - d(y_1, y_2)| = |B(\kappa(x_1, x_2)) - B(\kappa(y_1, y_2))| \leq \left( \max_{x \in [0, 1/\varepsilon]} h'(x) \right) \left| \frac{p(x_1)}{\varepsilon} - \frac{p(y_1)}{\varepsilon} \right| \\ \implies |d(x_1, x_2) - d(y_1, y_2)| \leq c^{-1/\varepsilon} \frac{\ln(1/c)}{\varepsilon} |p(x_1) - p(y_1)| \end{aligned}$$

Now if  $|p(x_1) - p(y_1)| > \lambda|x_1 - y_1|$  then  $x_1, y_1$  satisfy (CO3) and we have a solution for CONTINUOUS LOCALOPT. So  $|p(x_1) - p(y_1)| \leq \lambda|x_1 - y_1|$  and from the last inequality we have that

$$|d(x_1, x_2) - d(y_1, y_2)| \leq c^{-1/\varepsilon} \lambda \frac{\ln(1/c)}{\varepsilon} |x_1 - y_1|$$

---

<sup>4</sup>At this point we should have set  $\kappa(x, y) = \min\{-2p(x)/\varepsilon, 2p(y)/\varepsilon\}$  to get the inequality  $p(f(x)) \geq p(x) - \varepsilon$  but this would complicate the calculations in the rest of the cases. It is clear though that we could scale every parameter so that  $\varepsilon$  becomes  $2\varepsilon$  and nothing changes.

But this contradicts with (Od) since  $\lambda' = \max \left\{ \lambda, \left\lceil c^{-1/\varepsilon} \lambda \frac{\ln(1/c)}{\varepsilon} \right\rceil \right\}$ .

Finally it is easy to see that the size of the arithmetic circuits that we used for this reduction is polynomial in the size of the input. The only function that needs for explanation is that of  $d$  and  $\lambda'$ . We start with the observation that both  $c, c^{-1}$  are given and have descriptions of size only linear in the description of  $\varepsilon$ , since  $\varepsilon$  is a rational constant. The difficult term in the description of  $d$  is the term  $B(\kappa(x, y))$ . For this, we need to bound the size of  $\kappa(x, y)$ , let this bound be  $A$ . Then we can have precomputed the possible digits of  $\lceil \kappa(x, y) \rceil$  using  $\log(A)$  arithmetic circuits. Finally a final circuit combines the digits in order to get  $\lceil \kappa(x, y) \rceil$ . Now to compute  $c^{\lceil \kappa(x, y) \rceil}$  for each  $m_i 2^i$  of the  $\log(A)$  digits of  $\kappa(x, y)$  we compute the corresponding power  $2^i$  with repeated squaring using  $O(i)$  arithmetic gates. Then we combine the results such to compute  $c^{\lceil \kappa(x, y) \rceil}$ . This whole process needs  $O(\log^2 A)$  arithmetic gates. Since  $A \leq 1/\varepsilon$  the overall circuit for  $d$  needs  $\text{poly}(1/\varepsilon)$  arithmetic gates. For  $\lambda'$  we can also do a similar process but we have to bound  $c^{-\kappa(x, y)}$ . We can see that that  $c^{-\kappa(x, y)} \leq c^{-1/\varepsilon} = (1 - 0.1\varepsilon)^{-1/\varepsilon} \leq e^{10}$ . Therefore the size of  $c^{-1/\varepsilon}$  is bounded its ceil can be computed using a polynomial sized circuit.  $\square$

By inspecting the proof of the CLS-completeness of METRICBANACH we realize that, in the CLS-hardness part of the proof, we can actually guarantee that  $d$  is a metric. We can thus also establish the CLS-completeness of BANACH.

*Proof of Theorem 2.* Obviously because of Theorem 3, BANACH belongs to CLS.

For the opposite direction, we use the same reduction as in the proof of Theorem 3. We then prove that  $d$  satisfies the desired properties. We remind that we used the following instance of BANACH for the reduction

$$f' = f, d(x, y) = B(\kappa(x, y)) \cdot d_S(x, y), \varepsilon' = \frac{1}{\sqrt{c}}, \lambda' = \max \left\{ \lambda, \left\lceil c^{-1/\varepsilon} \lambda \frac{\ln(1/c)}{\varepsilon} \right\rceil \right\}, c = 1 - 0.1\varepsilon$$

We first prove that  $d$  is always a distance metric.

- (i) Obvious from the definition of  $d$ .
- (ii) If  $x \neq y$  then  $d_S(x, y) > 0$ . Also always  $c^{\kappa(x, y)} > 0$ , therefore  $d(x, y) > 0$ . Now since  $d_S(x, x) = 0$  we also have  $d(x, x) = 0$ .
- (iii) It is obvious from the definition of  $\kappa$  that  $\kappa(x, y) = \kappa(y, x)$  and since  $d_S$  is a distance metric, the same is true for the  $d_S$  and thus for  $d$ .
- (iv) Without loss of generality we assume that  $p(x) \geq p(y)$ . We consider the following cases

$\mathbf{p(x)} \geq \mathbf{p(z)}$  then we have  $d(x, y) = d(x, z)$  and therefore obviously  $d(x, y) \leq d(x, z) + d(z, y)$ .

$\mathbf{p(x)} \leq \mathbf{p(z)}$  then we have  $d(x, y) = B\left(-\frac{p(x)}{2\varepsilon}\right)$ ,  $d(x, z) = d(z, y) = B\left(-\frac{p(z)}{2\varepsilon}\right)$  but since  $p(x) \leq p(z)$  obviously  $2B\left(-\frac{p(z)}{2\varepsilon}\right) \geq B\left(-\frac{p(x)}{2\varepsilon}\right)$ .

Finally we will show the completeness of  $([0, 1]^3, d)$ . We first observe that for all  $x \neq y$ ,  $d(x, y) > 1$ , this comes from the fact that  $c < 1$  and so  $c^{-p(x)/\varepsilon} > 1$ .

Now let  $(x_n)$  be a Cauchy sequence then  $\forall \delta > 0, \exists N \in \mathbb{N}$  such that  $\forall n, m > N, d(x_n, x_m) \leq \delta$ . We set  $\delta = 1/2$  then there exists  $N \in \mathbb{N}$  such that  $\forall n, m > N, d(x_n, x_m) < 1/2$ . But from the previous observation this implies  $d(x_n, x_m) = 0$  and since  $d$  defines a metric we get  $x_n = x_m$ . Therefore  $(x_n)$  is constant for all  $n > N$  and obviously converges. This means that every Cauchy sequence converges and so  $([0, 1]^3, d)$  is a complete metric space.  $\square$



## D Proofs of Corollaries 1 2

*Proof of Corollary 1:* Let  $d_{c,\varepsilon/2}$  be the distance metric guaranteed by Theorem 1 with parameters  $c, \varepsilon/2$ . Let also  $(x_n)$  be the sequence produced by the Basic Iterative Procedure. Since  $f$  is a contraction with respect to  $d_{c,\varepsilon/2}$ , we have  $d_{c,\varepsilon/2}(x_n, x^*) \leq \frac{c^n}{1-c} d_{c,\varepsilon/2}(x_0, x_1)$ . If we make sure that  $d_{c,\varepsilon/2}(x_n, x_{n+1}) \leq \varepsilon/2$  then according to Theorem 1  $d(x_n, x^*) \leq \varepsilon$ . So the number of steps that are needed are:

$$\frac{c^n}{1-c} d_{c,\varepsilon/2}(x_0, x_1) \leq \frac{\varepsilon}{2} \Leftrightarrow n \geq \frac{\log(d_{c,\varepsilon/2}(x_0, x_1)) + \log((2-2c)/\varepsilon)}{\log(1/c)}.$$

□

*Proof of Corollary 2:* Using Corollary 1, we get that after  $n = \frac{\log(d_{c,\delta/2}(x_0, f(x_0))) + \log((2-2c)/\delta)}{\log(1/c)}$  iterations we will have  $d(x_n, x^*) \leq \delta$  or  $d(x_{n+1}, x^*) \leq \delta$ . Since in  $\bar{B}(x^*, \delta)$ ,  $f$  is a contraction with respect to  $d$ , it certainly must be that  $d(x_{n+1}, x^*) \leq \delta$ . By the same token,  $d(x_{n+1+m}, x^*) \leq c^m d(x_{n+1}, x^*)$ , for all  $m > 0$ . Therefore, to guarantee  $d(x_{n+1+m}, x^*) \leq \varepsilon$ , it suffices to take  $m \geq \frac{-\log(1/\delta) + \log(1/\varepsilon)}{\log(1/c)}$ . So in total we need  $n + 1 + m$  iterations, implying the number of iterations stated in the statement of the corollary. □