ISRAEL JOURNAL OF MATHEMATICS **227** (2018), 163–188
DOI: 10.1007/s11856-018-1723-z

SHARP CONVERGENCE RATES FOR AVERAGED NONEXPANSIVE MAPS

BY

MARIO BRAVO*

Departamento de Matemática y Ciencia de la Computación
Universidad de Santiago de Chile
Alameda Libertador Bernardo O'higgins 3363, Santiago, Chile
e-mail: mario.bravo.g@usach.cl

AND

ROBERTO COMINETTI**

Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez
Diagonal Las Torres 2640, Santiago, Chile
e-mail: roberto.cominetti@uai.cl

ABSTRACT

We establish sharp estimates for the convergence rate of the Kranosel'skiĭ–Mann fixed point iteration in general normed spaces, and we use them to show that the optimal constant of asymptotic regularity is exactly $1/\sqrt{\pi}$. To this end we consider a nested family of optimal transport problems that provide a recursive bound for the distance between the iterates. We show that these bounds are tight by building a nonexpansive map $T : [0,1]^{\mathbb{N}} \to [0,1]^{\mathbb{N}}$ that attains them with equality, settling a conjecture by Baillon and Bruck. The recursive bounds are in turn reinterpreted as absorption probabilities for an underlying Markov chain which is used to establish the tightness of the constant $1/\sqrt{\pi}$.

* The work of Mario Bravo was partially funded by FONDECYT Grant 11151003 and the Núcleo Milenio Información y Coordinación en Redes ICM/FIC RC130003.

** Roberto Cominetti gratefully acknowledges the support provided by FONDECYT 1130564 and FONDECYT 1171501.

Received March 3, 2017 and in revised form August 16, 2017

1. Introduction

Let C be a bounded closed convex set in a normed space $(X, \|\cdot\|)$, and $T : C \to C$ a nonexpansive map so that $\|Tx - Ty\| \leq \|x - y\|$ for all $x, y \in C$. Given an initial point $x^0 \in C$ and scalars $\alpha_n \in]0,1[$, the Kranosel'skiĭ–Mann iteration [19, 18] generates a sequence $(x^n)_{n\in\mathbb{N}}$ that approximates a fixed point of T by the recursive averaging process

$$\text{(KM)} \qquad x^{n+1} = (1-\alpha_{n+1})\, x^n + \alpha_{n+1}\, Tx^n.$$

Besides providing a method to compute fixed points and to prove their existence, this iteration appears in other contexts such as the Prox and Douglas–Rachford algorithms for convex optimization (see [6]), as well as in the discretization of nonlinear semigroups $\frac{dx}{dt} + [I - T](x) = 0$ (see [4]).

The convergence of the KM iterates towards a fixed point has been established under various conditions. In 1955, Krasnosel'skiĭ [18] considered the case $\alpha_n \equiv 1/2$ proving that convergence holds if $T(C)$ is pre-compact and the space X is uniformly convex, obtaining as a by-product the existence of fixed points. Soon after, Schaefer [22] proved that convergence also holds when $\alpha_n \equiv \alpha$ with $\alpha \in]0,1[$, while Edelstein [12] showed that the uniform convexity of X could be relaxed to strict convexity. The weak convergence of the iterates was studied in [20], while extensions to hyperbolic metric spaces were considered in [21]. The asymptotic properties of the iterates, without assuming the existence of fixed points, were studied in [7].

A crucial step in these results was to show that $\|x^n - Tx^n\|$ tends to 0. This property was singled out by Browder and Petryshyn [8] under the name of asymptotic regularity, proving that it holds for $\alpha_n \equiv \alpha$ when X is uniformly convex and $\mathrm{Fix}(T) \neq \phi$, without any compactness assumption. The extension to non-constant sequences α_n with

$$\sum \alpha_n(1-\alpha_n) = \infty$$

was achieved in 1972 by Groetsch [15]. Until then, uniform convexity seemed essential to obtain rate of convergence estimates. For instance, in Hilbert spaces, Browder and Petryshin [9] proved that for $\alpha_n \equiv \alpha$ one has

$$\sum \|x^{n+1} - x^n\|^2 < \infty$$

with $\|x^{n+1} - x^n\|$ decreasing, which readily implies $\|x^n - Tx^n\| = o(1/\sqrt{n})$, as observed in [3].

In 1976, Ishikawa [16] proved that strict convexity was not required and asymptotic regularity holds in every Banach space as soon as $\sum \alpha_n = \infty$ and $\alpha_n \leq 1 - \epsilon$. The same result was obtained independently by Edelstein and O'Brien [13], who observed that the limit $\|x^n - Tx^n\| \to 0$ is uniform with respect to the initial point x^0. Later, Goebel and Kirk [14] noted that it is also uniform in T, namely, for each $\epsilon > 0$ we have $\|x^n - Tx^n\| \leq \epsilon$ for all $n \geq n_0$ with n_0 depending on ϵ and C but independent of x^0 and T, while Kohlenbach [17] showed that n_0 depends on C only through its diameter. For a thorough account of the historic development of iterations of nonexpansive maps until 1983 we refer to the excellent survey by Bruck [10].

The first rate of convergence in general normed spaces was obtained in 1992 by Baillon and Bruck [3]: for the case $\alpha_n \equiv \alpha$ they proved that

$$\|x^n - Tx^n\| = O(1/\log n)$$

and conjectured a faster rate of $O(1/\sqrt{n})$. In fact, based on numerical experiments they guessed a metric estimate that would imply all the previous results, namely, they claimed the existence of a universal constant κ such that

$$\|x^n - Tx^n\| \leq \kappa \frac{\operatorname{diam}(C)}{\sqrt{\sum_{i=1}^{n} \alpha_i(1-\alpha_i)}}. \tag{1}$$

In [4] this estimate was confirmed to hold with $\kappa = 1/\sqrt{\pi} \sim 0.5642$ for constant $\alpha_n \equiv \alpha$, while the extension to arbitrary sequences α_n was recently achieved in [11] with exactly the same constant. This bound not only subsumes the known results but it also provides an explicit estimate of the fixed point residual, making it possible to compute the number of iterations required to attain any prescribed accuracy.

The constant $1/\sqrt{\pi}$, which appears somewhat mysteriously in this context, arises from a subtle connection between the KM iteration and a certain random walk over the integers (see [11]). A natural question is whether this is an essential feature, or whether a sharper bound can be proved by other means. In this direction, when T is an affine map and $\alpha_n \equiv 1/2$ a tight bound with $\kappa = 1/\sqrt{2\pi} \sim 0.3989$ was proved in [3], while for general α_n's the optimal constant is $\kappa = \max_{x \geq 0} \sqrt{x} e^{-x} I_0(x) \sim 0.4688$, where $I_0(\cdot)$ is a modified Bessel function of the first kind (see [5, 11]). For nonlinear maps, no sharp bound seems to be available. In this paper we close this gap by proving

THEOREM 1.1: *The constant* $\kappa = 1/\sqrt{\pi}$ *in the bound* (1) *is tight. Specifically, for each* $\kappa < 1/\sqrt{\pi}$ *there exists a nonexpansive map* T *defined on the unit cube* $C = [0,1]^{\mathbb{N}} \subseteq \ell^\infty(\mathbb{N})$, *an initial point* $x^0 \in C$, *and a constant sequence* $\alpha_n \equiv \alpha$, *such that the corresponding KM iterates* x^n *satisfy for some* $n \in \mathbb{N}$

$$\|x^n - Tx^n\| > \kappa \frac{\operatorname{diam}(C)}{\sqrt{\sum_{i=1}^n \alpha_i(1-\alpha_i)}}. \tag{2}$$

Note that by rescaling the norm it can always be assumed that $\operatorname{diam}(C) = 1$, which we do from now on.

Our analysis builds upon previous results from [3] and [2], for which we first establish a link with optimal transport and thence with a related stochastic process connecting with the results in [11].

Since the proof is somewhat involved, we sketch the main steps. We begin by noting that

$$x^n - Tx^n = (x^n - x^{n+1})/\alpha_{n+1}$$

so we will search for the best possible estimates for $\|x^n - x^{n+1}\|$. This is achieved in Section 2 by considering a nested sequence of optimal transport problems that provide a recursive bound d_{mn} for the distance between the iterates $\|x^m - x^n\| \leq d_{mn}$. These bounds d_{mn} are universal and do not depend either on the operator T or the space X. Moreover, it turns out that the d_{mn}'s induce a metric on the integers that satisfies a special 4-point inequality, which implies that the optimal transports have a simple structure and can be computed by the inside-out algorithm, as conjectured in [3]. We then show that the bounds d_{mn} are the best possible by constructing a specific nonexpansive map T that attains them all as equalities

$$\|x^m - x^n\| = d_{mn},$$

which settles a conjecture formulated in [3]. In Section 3 we re-interpret the recursive bounds d_{mn} as absorption probabilities for an underlying Markov chain with transition matrix defined from the optimal transports. When $\alpha_n \equiv \alpha \geq \frac{1}{2}$ the optimal transports take a particularly simple form and the induced Markov chain turns out to be asymptotically equivalent for $\alpha \sim 1$ to the Markov chain used in [11]. These results are combined in Section 4 in order to prove that the constant $\kappa = 1/\sqrt{\pi}$ is optimal, establishing Theorem 1.1. Some final comments are presented in Section 5, while the more technical proofs are left to the Appendix.

2. Sharp recursive bounds based on optimal transport

Following [3], we look for the sharpest possible estimates d_{mn} for the distance between the KM iterates $\|x^m - x^n\| \leq d_{mn}$. To this end, we note that x^n is a weighted average between the previous iterate x^{n-1} and its image Tx^{n-1}, so it follows inductively that x^n is an average of $x^0, Tx^0, Tx^1, \ldots, Tx^{n-1}$. Explicitly, set $\alpha_0 = 1$ and consider the probability distribution

$$\pi_i^n = \alpha_i \prod_{k=i+1}^{n} (1-\alpha_k)$$

for $i = 0, \ldots, n$. Then, adopting the convention $Tx^{-1} = x^0$ we have

$$x^n = \sum_{i=0}^{n} \pi_i^n \, Tx^{i-1}.$$

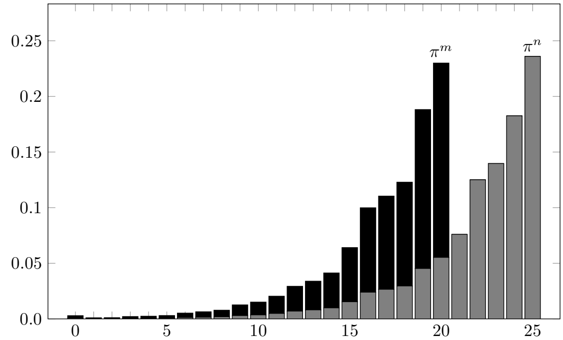

Figure 1. A plot of π^m and π^n $(m = 20, n = 25)$.

Since a similar formula holds for x^m with π^n replaced by π^m, we may try to match both expressions by solving an optimal transport problem between π^m and π^n (see Figure 1). Namely, let $\mathcal{N} = \{-1, 0, 1, \ldots\}$ and set $d_{-1,-1} = 0$ and $d_{-1,k} = d_{k,-1} = 1$ for all $k \in \mathbb{N}$, and then define recursively d_{mn} for $m, n \in \mathcal{N}$ by solving a sequence of linear optimal transport problems on the complete bipartite graph of Figure 2. Specifically, let us consider supplies π_i^m at the source nodes $i = 0, \ldots, m$ on the left, demands π_j^n at the destination nodes $j = 0, \ldots, n$ on the right, unit costs $d_{i-1,j-1}$ on each link (i, j), and define

$$(\mathcal{P}_{mn}) \qquad d_{mn} = \min_{z \in F^{mn}} C^{mn}(z) = \sum_{i=0}^{m} \sum_{j=0}^{n} z_{ij} \, d_{i-1,j-1}$$

where F^{mn} denotes the polytope of feasible transport plans z between π^m and π^n defined by

$$
\begin{aligned}
z_{ij} \geq& 0 \quad \text{for } i = 0, \ldots, m \text{ and } j = 0, \ldots, n, \\
\sum_{j=0}^{n} z_{ij} =& \pi_i^m \quad \text{for } i = 0, \ldots, m, \\
\sum_{i=0}^{m} z_{ij} =& \pi_j^n \quad \text{for } j = 0, \ldots, n.
\end{aligned}
\tag{3}
$$

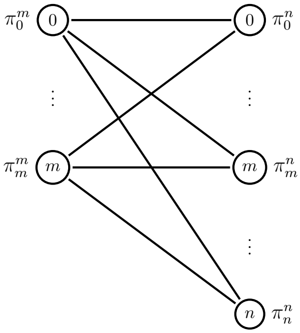

Figure 2. The optimal transport problem $(\mathcal{P}_{mn})$.

The problems $(\mathcal{P}_{mn})$ are equivalent to the recursive linear programs considered in [3], though the formulation in terms of optimal transport is new. The connection between the d_{mn}'s and the KM iterates is as follows.

THEOREM 2.1: *Let $T : C \to C$ be any nonexpansive map on a convex set C with $\mathrm{diam}(C) = 1$. Then the KM iterates satisfy $\|x^m - x^n\| \leq d_{mn}$ for all $m, n \in \mathbb{N}$.*

Proof. For each feasible transport plan $z \in F^{mn}$ we have

$$
\begin{aligned}
x^m - x^n =& \sum_{i=0}^{m} \pi_i^m \, Tx^{i-1} - \sum_{j=0}^{n} \pi_j^n \, Tx^{j-1} \\
=& \sum_{i=0}^{m} \sum_{j=0}^{n} z_{ij} \, Tx^{i-1} - \sum_{j=0}^{n} \sum_{i=0}^{m} z_{ij} \, Tx^{j-1} \\
=& \sum_{i=0}^{m} \sum_{j=0}^{n} z_{ij} (Tx^{i-1} - Tx^{j-1}).
\end{aligned}
$$

Now, suppose inductively that the bound $\left\|x^i - x^j\right\| \leq d_{ij}$ holds for all $i < m$ and $j < n$ (this is trivially the case for $m = n = 1$). Using the triangle inequality, the nonexpansivity of T, and the fact that $d_{-1,k} = 1 = \mathrm{diam}(C)$ to bound the terms $i = 0$ and $j = 0$, we get

$$\|x^m - x^n\| \leq \sum_{i=0}^{m}\sum_{j=0}^{n} z_{ij}\left\|Tx^{i-1} - Tx^{j-1}\right\| \leq \sum_{i=0}^{m}\sum_{j=0}^{n} z_{ij} d_{i-1,j-1}.$$

Minimizing the right-hand side over $z \in F^{mn}$ we get $\|x^m - x^n\| \leq d_{mn}$, completing the induction step. ∎

We stress that the d_{mn}'s are fully determined by the sequence α_n and do not depend on the map T, yet the bound $\|x^m - x^n\| \leq d_{mn}$ is valid for all KM iterates for every nonexpansive map $T : C \to C$ in any normed space X. Since the proof relies solely on the triangle inequality, a legitimate question is whether the bounds d_{mn} could be improved by using more sophisticated arguments. We will show that these bounds are in fact the best possible, as conjectured in [3].

Before proceeding with the proof we make some observations regarding the recursive bounds d_{mn}. Firstly, note that problem $(\mathcal{P}_{mn})$ is symmetric, so that $d_{mn} = d_{nm}$ and therefore it suffices to compute these quantities for $m \leq n$. On the other hand, any feasible transport z has non-negative entries that add up to one, so that $C^{mn}(z)$ is an average of previous values $d_{i-1,j-1}$ and then it follows inductively that $d_{mn} \in [0,1]$ for all $m, n \in \mathcal{N}$. It is also easy to prove inductively that $d_{nn} = 0$ for all $n \in \mathcal{N}$.

Informally, an optimal transport should satisfy the demand π_j^n at each node $j = 0, \ldots, n$ from the closest supply nodes $i = 0, \ldots, m$. In particular, for $i \leq m \leq n$ we have

$$\pi_i^n = \pi_i^m \prod_{k=m+1}^{n} (1-\alpha_k) \leq \pi_i^m$$

and since $d_{i-1,i-1} = 0$, the demand π_i^n at destination i should be fulfilled directly from the supply π_i^m available at the twin source node i. The proof of these intuitive facts is not so simple, though. The next Theorem, which is due to [2], summarizes some properties that will be relevant in the analysis of the KM iteration. Since the proofs in [2] are rather long and technical, in the Appendix we present much simpler proofs which strongly exploit the interpretation of d_{mn} in terms of optimal transport.

THEOREM 2.2 (Aygen-Satik [2]):

(a) *The function* $d : \mathcal{N} \times \mathcal{N} \to [0,1]$ *where* $d(m,n) = d_{mn}$ *defines a distance over the set* $\mathcal{N}$.

(b) *For* $m \leq n$ *there exists an optimal transport* z *for* d_{mn} *with* $z_{ii} = \pi_i^n$ *for all* $i = 0, \ldots, m$ *and a fortiori* $z_{ij} = 0$ *for all* $j \in \{0, \ldots, m\}, j \neq i$. *Such an optimal solution will be called simple.*

(c) *For fixed* m *we have that* $n \mapsto d_{mn}$ *is decreasing for* $n \leq m$ *and increasing for* $n \geq m$.

(d) *If* $\alpha_k \geq 1/2$ *for all* $k \in \mathbb{N}$, *then for all integers* $0 \leq i \leq k \leq j \leq l$ *we have* $d_{il} + d_{kj} \leq d_{ij} + d_{kl}$. *We refer to this as the 4-point inequality.*

Proof. See Appendix A. ∎

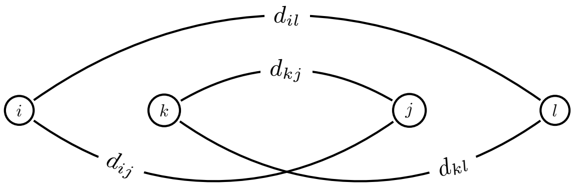

Figure 3. The 4-point inequality.

Remark 2.3: Although our proof of the 4-point inequality (d) works only for $\alpha_k \geq 1/2$ (which suffices for our purposes), it is worth noting that in [2] it is proved for all $\alpha_k \in]0,1[$. A relevant consequence of this property is the existence of an optimal transport in which the flows do not cross. Namely, if z is a transport plan such that $z_{ij} > 0$ and $z_{kl} > 0$ with $i < k < j < l$, then we can reduce the cost by removing $\varepsilon = \min\{z_{ij}, z_{kl}\}$ from these lower arcs (see Figure 3) while augmenting the uncrossed flows z_{il} and z_{kj} by the same amount. This no flow-crossing property justifies in turn the inside-out algorithm which was conjectured in [3] to solve $(\mathcal{P}_{mn})$: set $z_{ii} = \pi_i^n$ for all $i = 0, \ldots, m$ and then consider sequentially the demands π_j^n for $j = m+1, \ldots, n$ and fulfill them from the closest supply nodes starting with $i = m$ and moving towards smaller indices $i = m-1, \ldots, 0$ when the residual supply $\pi_i^m - \pi_i^n$ of node i is exhausted.

Let us proceed to show that the bounds d_{mn} are sharp. The proof exploits properties (a), (b), (c) to construct a nonexpansive map T for which the KM iterates satisfy $\|x^m - x^n\| = d_{mn}$ for all m, n. Property (d) will only be used later in order to show that the constant $\kappa = 1/\sqrt{\pi}$ in (1) is sharp.

THEOREM 2.4: *Let a sequence $\alpha_n \in]0,1[$ be given and consider the corresponding bounds d_{mn}. Then there exists a nonexpansive map $T : C \to C$ defined on the unit cube $C = [0,1]^{\mathbb{N}} \subseteq \ell^\infty(\mathbb{N})$ and a corresponding KM sequence for which the equality $\|x^m - x^n\|_\infty = d_{mn}$ holds for all $m, n \in \mathbb{N}$.*

Proof. The map will be defined on the set $C = [0,1]^Q$ in the space $(\ell^\infty(Q), \|\cdot\|_\infty)$ where Q denotes the set of all integer pairs (m,n) with $0 \le m \le n < \infty$, which is isomorphic to $\ell^\infty(\mathbb{N})$.

Let us begin by observing that property (b) allows $(\mathcal{P}_{mn})$ to be reformulated as a min-cost flow on the complete graph G_n with node set $\{0,\ldots,n\}$ and costs $d_{i-1,j-1}$ on each arc (i,j), while considering only the residual demands $\pi_j^n - \pi_j^m$, namely

$$(\mathcal{Q}_{mn})\ \ d_{mn} = \min_{z\ge 0}\left\{\sum_{i,j=0}^{n} z_{ij} d_{i-1,j-1} : \sum_{i=0}^{n} z_{ij} - \sum_{i=0}^{n} z_{ji} = \pi_j^n - \pi_j^m \text{ for } j=0,\ldots,n\right\}.$$

The residual demands $\pi_j^n - \pi_j^m$ are balanced and add up to zero. They are negative for $j = 0,\ldots,m$ and positive for $j = m+1,\ldots,n$ (since $\pi_j^n < \pi_j^m$ for $j \le m$ and $\pi_j^m = 0$ for $j > m$). See Figure 4.

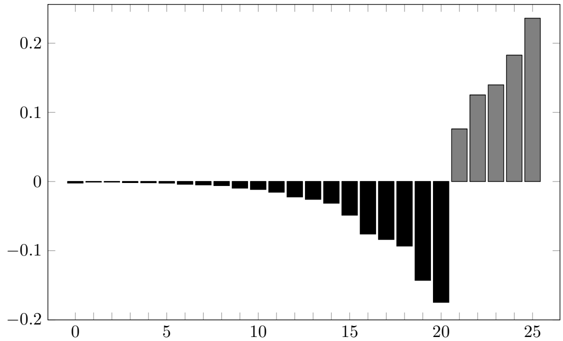

Figure 4. Residual demands $\pi^n - \pi^m$ $(m = 20, n = 25)$.

Now, for each $(m,n) \in Q$ let z^{mn} be an optimal solution for $(\mathcal{Q}_{mn})$ and u^{mn} optimal for the dual problem

$$(\mathcal{D}_{mn}) \qquad \max_{u}\left\{\sum_{j=0}^{n} (\pi_j^n - \pi_j^m) u_j : u_j \le u_i + d_{i-1,j-1} \text{ for all } i,j = 0,\ldots,n\right\}$$

so that by complementary slackness we have

$$z_{ij}^{mn}(u_j^{mn} - u_i^{mn}) = z_{ij}^{mn} d_{i-1,j-1} \quad \text{for all } i,j = 0,\dots,n. \tag{4}$$

Since $(\mathcal{D}_{mn})$ is invariant by addition of constants we can fix the value of one component, say $u_0^{mn} = 0$, from which we get $u_j^{mn} \in [0,1]$ for all $j = 0,\dots,n$. Indeed, feasibility implies $u_j^{mn} \le u_0^{mn} + d_{-1,j-1} = 1$. To prove non-negativity we note that $\pi_0^m > \pi_0^n$ so we can find $i > 0$ with $z_{0i}^{mn} > 0$ and complementarity gives $u_i^{mn} = u_0^{mn} + d_{-1,i-1} = 1$, which combined with feasibility and $d_{i-1,j-1} \le 1$ yields $0 \le u_i^{mn} - d_{i-1,j-1} \le u_j^{mn}$.

Note that the constraints in $(\mathcal{D}_{mn})$ can be written as

$$|u_j^{mn} - u_i^{mn}| \le d_{i-1,j-1} \quad \text{for all } i,j = 0,\dots,n. \tag{5}$$

If we extend the definition of u^{mn} by setting $u_j^{mn} = u_n^{mn}$ for $j > n$, the monotonicity in Theorem 2.2(c) implies that (5) remains valid for all $i,j \in \mathbb{N}$.

Now, let us consider the sequence $y^0, y^1, \dots, y^k, \dots \in [0,1]^Q$ defined as

$$y^k = (u_{k+1}^{mn})_{mn \in Q} \tag{6}$$

and build $x^k \in [0,1]^Q$ starting from $x^0 = 0 \in [0,1]^Q$ by the recursion

$$x^k = (1-\alpha_k)x^{k-1} + \alpha_k y^{k-1}, \quad k \ge 1. \tag{7}$$

If we define the operator T along this sequence by setting

$$Tx^k = y^k \quad k = 0,1,\dots \tag{8}$$

then (7) corresponds exactly to the Kranosel'skiĭ–Mann iterates for this map.

We will show that $\|x^m - x^n\|_\infty = d_{mn} = \|y^m - y^n\|_\infty$ for all $(m,n) \in Q$. This implies that T is an isometry and therefore, since $\ell^\infty(Q)$ is an hyperconvex space, by the Aronszajn–Panitchpakdi Theorem [1] it can be extended to a nonexpansive map $T : [0,1]^Q \to [0,1]^Q$, which completes the proof.

It remains to show that $\|x^m - x^n\|_\infty = d_{mn} = \|y^m - y^n\|_\infty$. Since x^k are the Kranosel'skiĭ–Mann iterates for the map $x^k \mapsto Tx^k = y^k$, as in Theorem 2.1 we have

$$x^m - x^n = \sum_{i=0}^{m}\sum_{j=0}^{n} z_{ij}^{mn}[y^{i-1} - y^{j-1}]. \tag{9}$$

From the definition of y^k and since (5) holds for all $i,j \in \mathbb{N}$, it follows that

$$\|y^{i-1} - y^{j-1}\|_\infty = \|(u_i^{m'n'} - u_j^{m'n'})_{m'n' \in Q}\|_\infty \le d_{i-1,j-1}$$

and then the triangle inequality implies

$$(10)\quad \|x^m - x^n\|_\infty \le \sum_{i=0}^{m}\sum_{j=0}^{n} z_{ij}^{mn}\|y^{i-1} - y^{j-1}\|_\infty \le \sum_{i=0}^{m}\sum_{j=0}^{n} z_{ij}^{mn} d_{i-1,j-1} = d_{mn}.$$

On the other hand, if we consider only the mn-coordinate in (9) we have

$$\begin{aligned} x_{mn}^m - x_{mn}^n &= \sum_{i=0}^{m}\sum_{j=0}^{n} z_{ij}^{mn}[y_{mn}^{i-1} - y_{mn}^{j-1}] \\ &= \sum_{i=0}^{m}\sum_{j=0}^{n} z_{ij}^{mn}[u_i^{mn} - u_j^{mn}] \\ &= \sum_{i=0}^{m}\sum_{j=0}^{n} z_{ij}^{mn} d_{i-1,j-1} = d_{mn} \end{aligned}$$

where the last equality follows from the complementary slackness (4).

This combined with (10) yields $\|x^m - x^n\|_\infty = d_{mn}$. Going back to (10), it follows that $\|y^{i-1} - y^{j-1}\|_\infty = d_{i-1,j-1}$ as soon as $z_{ij}^{mn} > 0$ for at least one pair $(m,n) \in Q$. Now, this strict inequality holds for $m = j-1$ and $n = j$: the optimal transport in that case is given by $z_{ii}^{j-1,j} = \pi_i^j$ and $z_{ij}^{j-1,j} = \pi_i^{j-1} - \pi_i^j > 0$ for all $i = 0, \ldots, j-1$. Therefore $\|y^{i-1} - y^{j-1}\|_\infty = d_{i-1,j-1} = \|x^{i-1} - x^{j-1}\|_\infty$, which establishes the isometry property and completes the proof. ∎

3. A related stochastic process

For the rest of the analysis we will exploit a probabilistic interpretation of the bounds d_{mn}. To this end, let us fix simple optimal transports z^{mn} for $(\mathcal{P}_{mn})$ and consider them as transition probabilities for a Markov chain $\mathcal{D}$ interpreted as a race between a fox and a hare: a hare located at m is being chased by a fox located at n and they sequentially jump from (m,n) to a new state $(i-1, j-1)$ with probability z_{ij}^{mn}. If the process ever attains a state (k,k) on the diagonal, then with probability one it will move to a new diagonal state $(i-1, i-1)$ so we can represent all diagonal nodes with an absorbing state f meaning that the fox captures the hare. Otherwise the process remains in the region $m < n$ and eventually reaches a state $(-1, k)$ with $k \ge 0$, in which case the hare can be said to escape by reaching the burrow at position -1. We represent all these latter nodes by an absorbing state h meaning that the hare escapes. Hence, the state space for the chain is $\mathcal{S} = \{(m,n) : 0 \le m < n\} \cup \{h, f\}$ (see Figure 5). Note that since $z_{ij}^{mn} = 0$ for $j < i$ the process remains in $\mathcal{S}$.

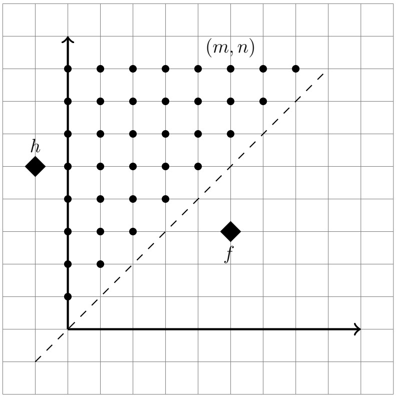

Figure 5. The state space $\mathcal{S}$.

For notational convenience we will write mn instead of (m,n). All the states mn are transient and in fact each iteration decreases m by at least one unit so that after at most m stages the chain must attain one of the absorbing states f or h. Now, the basic observation is that the recursion

$$d_{mn} = \sum_{i=0}^{m} \sum_{j=0}^{n} z_{ij}^{mn} d_{i-1,j-1}$$

characterizes the probability that the hare escapes conditional on the fact that the process starts in state mn, that is to say $d_{mn} = \mathbb{P}(\mathcal{D} \text{ attains } h | mn)$.

A similar Markov chain $\mathcal{C}$ was considered in [11], except that the transition probabilities were defined by the following sub-optimal transport plans

$$\tilde{z}_{ij}^{mn} = \begin{cases} \pi_i^m \pi_j^n & \text{for } i = 0, \ldots, m \text{ and } j = m+1, \ldots, n, \\ \pi_i^n & \text{for } i = j = 0, \ldots, m, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

These transport plans produced the bounds $c_{mn} = \mathbb{P}(\mathcal{C} \text{ attains } h|mn)$ which satisfy the analog recursion

$$c_{mn} = \sum_{i=0}^{m} \sum_{j=0}^{n} \tilde{z}_{ij}^{mn} c_{i-1,j-1}$$

with boundary conditions $c_{-1,-1} = 0$ and $c_{-1,k} = c_{k,-1} = 1$ for $k \in \mathbb{N}$. Since d_{mn} is obtained from the optimal transports while c_{mn} uses a sub-optimal one, it follows inductively that $d_{mn} \le c_{mn}$. Note also that in both recursions the inner sum can be replaced by $\sum_{j=m+1}^{n}$ since the only nonzero flows for $j \le m$ are z_{jj} and $\tilde{z}_{jj}$ which do not contribute to the sums as they are multiplied by $d_{j-1,j-1} = 0$ and $c_{j-1,j-1} = 0$.

Let $\mathbf{C}$ and $\mathbf{D}$ denote the transition matrices corresponding to $\mathcal{C}$ and $\mathcal{D}$, so that $\mathbf{C}(s,s')$ and $\mathbf{D}(s,s')$ are the transition probabilities from state $s \in \mathcal{S}$ to $s' \in \mathcal{S}$. Recalling that when the process starts from mn it attains one of the absorbing states after at most m stages, the probabilities d_{mn} and c_{mn} can be computed as the (mn,h) entries of the m-th stage transition matrices $\mathbf{D}^m$ and $\mathbf{C}^m$, hence

$$|c_{mn} - d_{mn}| = |\mathbf{D}^m(mn,h) - \mathbf{C}^m(mn,h)| \le |||\mathbf{D}^m - \mathbf{C}^m|||_\infty \tag{12}$$

where $|||A|||_\infty = \max_{s\in\mathcal{S}} \sum_{s'\in\mathcal{S}} |A(s,s')|$ is the operator norm induced by $\|\cdot\|_\infty$. This estimate, combined with the inside-out algorithm in Remark 2.3, allows us to show that for $\alpha_n \equiv \alpha \sim 1$ both processes are similar and $|c_{mn} - d_{mn}|$ becomes negligible. More precisely,

PROPOSITION 3.1: *If $\alpha_n \equiv \alpha \in [\frac{1}{2}, 1[$ then we have*

$$0 \le c_{mn} - d_{mn} \le 4m(1-\alpha)^2$$

for all $0 \le m \le n$.

Proof. See Appendix B. ∎

4. Proof of Theorem 1.1

We now have all the ingredients to prove Theorem 1.1. Recall that the recursive bounds d_{mn} and c_{mn} were introduced with the aim of finding a sharp estimate of $\|x^n - Tx^n\| = \|x^{n+1} - x^n\|/\alpha_{n+1}$, for which the relevant terms are $d_{n,n+1}$ and $c_{n,n+1}$. Let us consider a constant sequence $\alpha_n \equiv \alpha$ (with α to be chosen later) and denote $c_{mn}(\alpha), d_{mn}(\alpha)$ the corresponding bounds. Let us also define

$$\kappa_n(\alpha) = \sqrt{n\,\alpha(1-\alpha)}\; d_{n,n+1}(\alpha)/\alpha,$$
$$\tilde{\kappa}_n(\alpha) = \sqrt{n\,\alpha(1-\alpha)}\; c_{n,n+1}(\alpha)/\alpha.$$

Considering the map T and the KM iterates as given by Theorem 2.4 we have

$$\sqrt{n\,\alpha(1-\alpha)}\;\|x^n - Tx^n\| = \kappa_n(\alpha),$$

so that Theorem 1.1 will be proved if we show that $\kappa_n(\alpha)$ is arbitrarily close to $1/\sqrt{\pi}$ for appropriate values of n and α. We claim that $\kappa_n(\alpha)$ and $\tilde{\kappa}_n(\alpha)$ are very similar for $\alpha \sim 1$ and close to $1/\sqrt{\pi}$ for large n. As an illustration, Figure 6 compares these quantities for 4 different α's, where each plot shows the curves $\tilde{\kappa}_n(\alpha)$ (upper) and $\kappa_n(\alpha)$ (lower) as a function of n for $n = 1, \ldots, 300$.

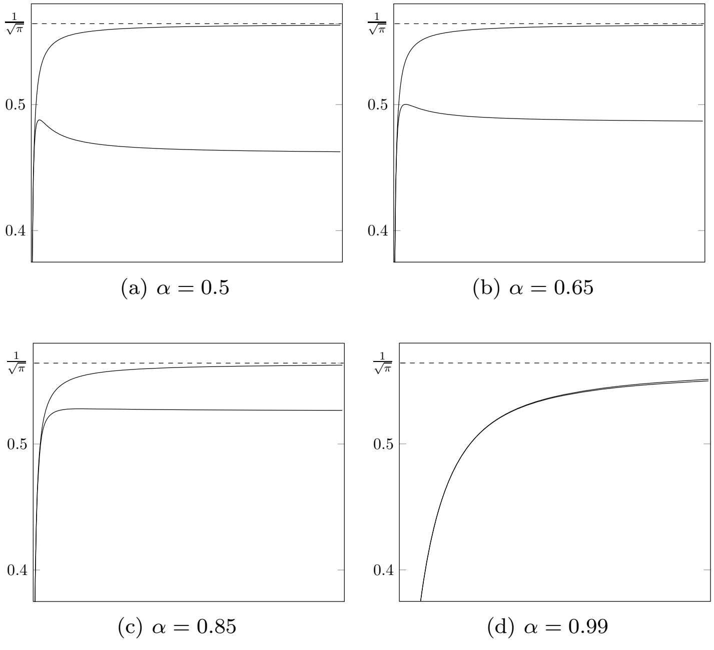

Figure 6. Comparison of $\kappa_n(\alpha)$ and $\tilde{\kappa}_n(\alpha)$ for $n = 1, \ldots, 300$.

More formally, using Proposition 3.1 we get

$$\begin{aligned} 0 \le \tilde{\kappa}_n(\alpha) - \kappa_n(\alpha) &\le \sqrt{n\,\alpha(1-\alpha)}\; 4n(1-\alpha)^2/\alpha \\ &= 4n^{3/2}(1-\alpha)^{5/2}/\sqrt{\alpha} \end{aligned} \tag{13}$$

while a formula for $\tilde{\kappa}_n(\alpha)$ involving the hypergeometric function ${}_2F_1(-n, \frac{1}{2}; 2; \cdot)$ obtained in [11] gives explicitly (see the proof of [11, Proposition 8])

$$\tilde{\kappa}_n(\alpha) = \frac{1}{\pi} \int_0^{4n\alpha(1-\alpha)} \sqrt{\frac{1}{s} - \frac{1}{4n\alpha(1-\alpha)}} \left(1 - \frac{s}{n}\right)^n ds. \tag{14}$$

Now, select a sequence $\theta_n \in]0, 1[$ with

$$4n^{3/2}(1-\theta_n)^{5/2}/\sqrt{\theta_n} \to 0 \quad \text{and} \quad 4n\theta_n(1-\theta_n) \to \infty$$

(e.g., $\theta_n = 1 - \ln n/n$). Taking the constant α equal to θ_n it follows from (13) that $\tilde{\kappa}_n(\theta_n) - \kappa_n(\theta_n) \to 0$ and then (14) gives

$$\lim_{n\to+\infty} \kappa_n(\theta_n) = \lim_{n\to+\infty} \tilde{\kappa}_n(\theta_n) = \frac{1}{\pi}\Gamma\left(\frac{1}{2}\right) = \frac{1}{\sqrt{\pi}}$$

which completes the proof of Theorem 1.1. ∎

5. Final comments

According to Theorem 1.1 the best uniform constant in (1) is $\kappa = 1/\sqrt{\pi}$. Since this value is attained for $\alpha_n \equiv \alpha \sim 1$, it would be interesting to find the asymptotic regularity constant $\kappa = \kappa(\alpha)$ as a function of α for constant sequences $\alpha_n \equiv \alpha$, and to discover which α yields the best convergence rate. From Theorems 2.1 and 2.4 we know that the recursive bounds d_{mn} provide sharp estimates for arbitrary sequences $\alpha_n \in]0, 1[$, so that

$$\kappa(\alpha) = \sup_{n\in\mathbb{N}} \kappa_n(\alpha).$$

Hence, defining

$$\gamma(\alpha) = \frac{\kappa(\alpha)}{\sqrt{\alpha(1-\alpha)}} = \sup_{n\in\mathbb{N}} \sqrt{n}\, d_{n,n+1}(\alpha)/\alpha \tag{15}$$

we get the convergence rate

$$\|x^n - Tx^n\| \le \gamma(\alpha)/\sqrt{n}.$$

Using the inside-out algorithm in Remark 2.3 we computed $\gamma(\alpha)$ to obtain the plot in Figure 7. This simple looking graph masks the complexity of the analytic expression of $\gamma(\alpha)$ which involves the maximum of the terms $\sqrt{n}\, d_{n,n+1}(\alpha)/\alpha$.

Note that the graph of $\gamma(\cdot)$ is slightly asymmetrical with respect to $\alpha = 1/2$, which is due to the fact that the optimal transports computed by the inside-out algorithm have a more complex structure for $\alpha < 1/2$ than for $\alpha \geq 1/2$. In fact, from the inside-out algorithm it follows that $d_{n,n+1}(\alpha)/\alpha$ is a polynomial for $\alpha \geq 0$ whereas it is piecewise polynomial for $\alpha < 1/2$. Hence $\gamma(\alpha)$ is formed by polynomial pieces which are glued together quite smoothly.

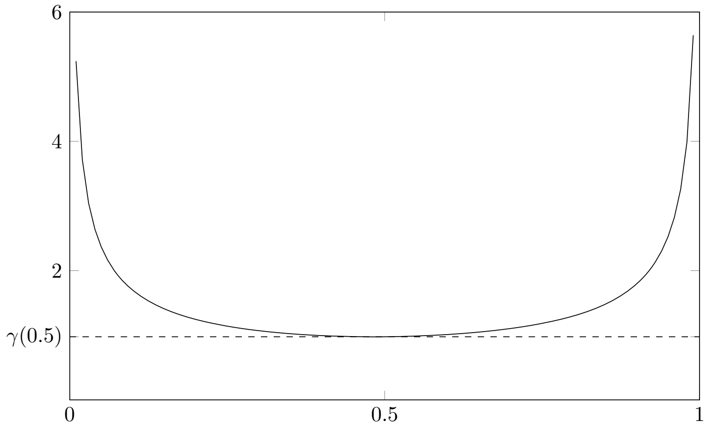

Figure 7. The rate $\gamma(\alpha)$ as a function of α (for $0.01 \leq \alpha \leq 0.99$).

For the original iteration of Krasnosel'skiĭ [18] with $\alpha = 1/2$, the supremum in (15) seems to be attained at $n = 8$ (see Figure 6(a)). If this is the case, then for $\alpha \geq 1/2$ the inside-out algorithm gives

$$\begin{aligned} d_{8,9}(\alpha)/\alpha = & 1 - 8\,\alpha + 64\,\alpha^2 - 448\,\alpha^3 + 2835\,\alpha^4 - 16008\,\alpha^5 \\ & + 79034\,\alpha^6 - 334908\,\alpha^7 + 1201873\,\alpha^8 \\ & - 3622324\,\alpha^9 + 9129380\,\alpha^{10} - 19214722\,\alpha^{11} \\ & + 33796129\,\alpha^{12} - 49776610\,\alpha^{13} + 61566687\,\alpha^{14} \\ & - 64152608\,\alpha^{15} + 56488500\,\alpha^{16} - 42133404\,\alpha^{17} \\ & + 26651679\,\alpha^{18} - 14288252\,\alpha^{19} + 6472429\,\alpha^{20} \\ & - 2462126\,\alpha^{21} + 778478\,\alpha^{22} - 201354\,\alpha^{23} \\ & + 41584\,\alpha^{24} - 6604\,\alpha^{25} + 758\,\alpha^{26} - 56\,\alpha^{27} + 2\,\alpha^{28}, \end{aligned}$$

which would yield the following exact value for the sharp rate in Krasnosel'skiĭ's iteration:

$$\gamma(0.5) = \frac{46302245}{67108864}\sqrt{2} = \frac{5 \cdot 11 \cdot 841859}{2^{26}}\sqrt{2} \sim 0.9757468.$$

Figure 7 could give the impression that this rate is optimal. However, a more careful look around $\alpha = 1/2$ reveals that this is not the case, and numerically we get smaller values such as $\gamma(0.48121) \sim 0.974637$. The difference is quite small, though, and Karsnosel'skiĭ's original iteration seems close to optimal.

While throughout this paper we focused on finding the best uniform bounds of the form (1), it would also be interesting to study the asymptotic rates by determining whether the limit $\kappa_\infty(\alpha) = \lim_{n\to\infty} \kappa_n(\alpha)$ exists. We computed $\kappa_n(\alpha)$ for different values of α and up to $n = 50000$, which seems to provide a reasonable approximation for $\kappa_\infty(\alpha)$. In particular, we found $\kappa_\infty(0.5) \sim \sqrt{2/3\pi}$ which is consistent with an observation in [3]. A formal treatment of these questions remains open.

Appendix A. Proof of Theorem 2.2

(a) *The function $d(m,n) = d_{mn}$ defines a distance on the set $\mathcal{N}$.*

Proof. We already noted that $d_{mn} \geq 0$ and $d_{mn} = d_{nm}$ for all $n, m \in \mathcal{N}$, so we only have to prove the triangle inequality and that $d_{mn} = 0$ iff $m = n$. We will show inductively that these properties hold for $m, n \leq \ell$ for each $\ell \in \mathcal{N}$. The base case $\ell = -1$ is trivial. Suppose that both properties hold for $\ell - 1$ and let us prove them for ℓ.

For $m, n \leq \ell$ we have $d_{mn} = 0$ iff $m = n$: From the induction hypothesis it suffices to prove that $d_{m\ell} > 0$ when $m < \ell$ and $d_{\ell\ell} = 0$. For $m < \ell$ take an optimal transport plan $z \in F^{m\ell}$ so that

$$\sum_{i=0}^{m} z_{i\ell} = \pi_\ell^\ell = \alpha_\ell > 0$$

and we find $i \in \{0, \ldots, m\}$ with $z_{i\ell} > 0$ and then $d_{m\ell} \geq z_{i\ell} d_{i-1,\ell-1} > 0$. Now, for $m = \ell$ we consider the feasible transport plan $z \in F^{\ell\ell}$ with $z_{ii} = \pi_i^\ell$ and $z_{ij} = 0$ for $j \neq i$ so that using the induction hypothesis we get

$$0 \leq d_{\ell\ell} \leq \sum_{i=0}^{\ell} \sum_{j=0}^{\ell} z_{ij} d_{i-1,j-1} = \sum_{i=0}^{\ell} \pi_i^\ell d_{i-1,i-1} = 0.$$

For $m,n,p \leq \ell$ we have $d_{mn} \leq d_{mp}+d_{pn}$: Let z^{mp} and z^{pn} be optimal transport plans for d_{mp} and d_{pn} respectively, and define for $i=0,\dots,m$ and $j=0,\dots,n$

$$z_{ij}=\sum_{k=0}^{p}\frac{z_{ik}^{mp}z_{kj}^{pn}}{\pi_k^p}.$$

A straightforward computation shows that $z\in F^{mn}$ and therefore using the induction hypothesis we get

$$\begin{aligned}
d_{mn} \leq& \sum_{i=0}^{m}\sum_{j=0}^{n} z_{ij}d_{i-1,j-1}\\
=&\sum_{i=0}^{m}\sum_{j=0}^{n}\sum_{k=0}^{p}\frac{z_{ik}^{mp}z_{kj}^{pn}}{\pi_k^p}d_{i-1,j-1}\\
\leq&\sum_{i=0}^{m}\sum_{j=0}^{n}\sum_{k=0}^{p}\frac{z_{ik}^{mp}z_{kj}^{pn}}{\pi_k^p}[d_{i-1,k-1}+d_{k-1,j-1}]\\
=&\sum_{i=0}^{m}\sum_{k=0}^{p}\sum_{j=0}^{n}\frac{z_{ik}^{mp}z_{kj}^{pn}}{\pi_k^p}d_{i-1,k-1}+\sum_{j=0}^{n}\sum_{k=0}^{p}\sum_{i=0}^{m}\frac{z_{ik}^{mp}z_{kj}^{pn}}{\pi_k^p}d_{k-1,j-1}\\
=&\sum_{i=0}^{m}\sum_{k=0}^{p} z_{ik}^{mp}d_{i-1,k-1}+\sum_{j=0}^{n}\sum_{k=0}^{p} z_{kj}^{pn}d_{k-1,j-1}\\
=&d_{mp}+d_{pn}. \qquad \blacksquare
\end{aligned}$$

(b) *For $m\leq n$ there exists a simple optimal transport plan z for d_{mn} with $z_{ii}=\pi_i^n$ for all $i=0,\dots,m$.*

Proof. Let z be an optimal solution for d_{mn} and suppose that $z_{ii}<\pi_i^n$ for some $i\in\{0,\dots,m\}$. Choose the smallest of such i's and take $j\in\{0,\dots,m\}$ with $j\neq i$ and $z_{ji}>0$. Now, take $k\in\{0,\dots,n\}$ the smallest $k\neq i$ with $z_{ik}>0$ and consider a modified transport plan $\tilde{z}$ identical to z except for (see Figure 8)

$$\tilde{z}_{ii}=z_{ii}+\varepsilon,$$

$$\tilde{z}_{ji}=z_{ji}-\varepsilon,$$

$$\tilde{z}_{ik}=z_{ik}-\varepsilon,$$

$$\tilde{z}_{jk}=z_{jk}+\varepsilon.$$

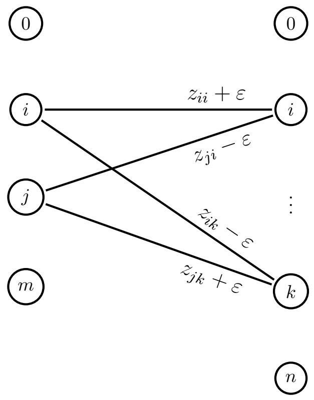

Figure 8. Redistribution of flow to construct a simple optimal transport

Using the triangle inequality we get the difference in transport cost as

$$C^{mn}(\tilde{z}) - C^{mn}(z) = \varepsilon(d_{i-1,i-1} + d_{j-1,k-1} - d_{j-1,i-1} - d_{i-1,k-1}) \leq 0$$

so that taking $\varepsilon = \min\{z_{ji}, z_{ik}\}$ we get a new optimal transport plan $\tilde{z}$ with either $\tilde{z}_{ik} = 0$ or $\tilde{z}_{ji} = 0$ (or both).

Updating $z \leftarrow \tilde{z}$ and repeating this process iteratively we note that the pair (i, k) increases lexicographically and can remain constant for at most m iterations. Indeed, for any given i the same k can occur only as long as $\tilde{z}_{ik}$ remains strictly positive, which implies that $\tilde{z}_{ji} = 0$. This can occur in at most $m-1$ iterations (one of each $j \neq i$), after which we necessarily have either $\tilde{z}_{ik} = 0$ and k must increase in the next iteration, or $\tilde{z}_{ii} = \pi_i^n$, in which case i increases in the next iteration. It follows that the process is finite and we eventually reach an optimal solution with $z_{ii} = \pi_i^n$ for all $i = 0, \ldots, m$. ∎

(c) *For fixed m we have that $n \mapsto d_{mn}$ increases for $n \geq m$ and decreases for $n \leq m$.*

Proof. Let us prove by induction that $d_{mn} \leq d_{m,n+1}$ for all $n \geq m$. The base case $n = m$ is trivial. For the induction step take $z \in F^{m,n+1}$, a simple optimal solution for $d_{m,n+1}$. We observe that for $j = 0, \ldots, n$ we have

$$\sum_{i=0}^{m} z_{ij} = \pi_j^{n+1} = (1-\alpha_{n+1})\pi_j^n < \pi_j^n$$

whereas

$$\sum_{i=0}^{m} z_{i,n+1} = \pi_{n+1}^{n+1} = \alpha_{n+1}.$$

We will transform z into a feasible solution for d_{mn} by redirecting the inflow α_{n+1} of node $n+1$ towards all other nodes $j \leq n$ in such a way that the inflows $\sum_{i=0}^{m} z_{ij}$ are increased from π_j^{n+1} to π_j^n, while maintaining the flow balance at the source nodes $\sum_{j=0}^{n+1} z_{ij} = \pi_i^m$. In the process, we make sure that all flow shifts reduce the cost.

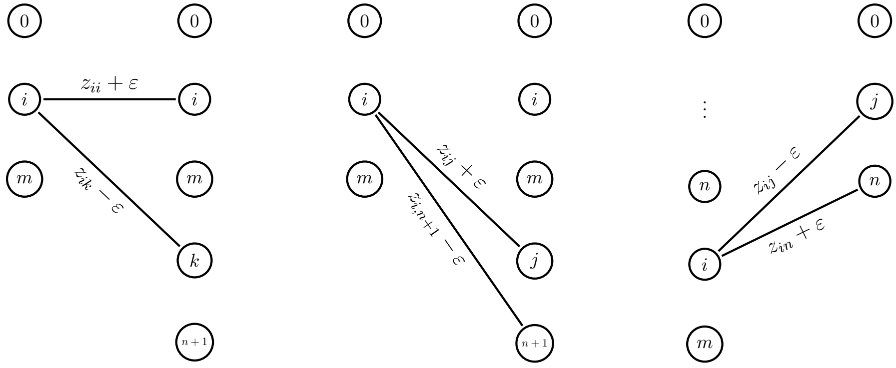

Figure 9. Redistribution of flows to prove monotonicity.

We proceed in two stages. Firstly, for each $i = 0, \ldots, m$ we choose any $k > m$ with $z_{ik} > 0$ from which we remove some amount ε and re-route it to destination i by augmenting z_{ii}, as shown in Figure 9 left. This increases the inflow to destination node i while keeping the flow balance at the source node i, and the cost is reduced by $\varepsilon(d_{i-1,i-1} - d_{i-1,k-1}) < 0$. We repeat this process until the inflow to i reaches π_i^n as required. In the second stage, for each node $j = m+1, \ldots, n$ we increase its inflow up to π_j^n by reducing the remaining flows $z_{i,n+1} > 0$ for $i = 0, \ldots, m$ and augmenting z_{ij}, as shown in Figure 9 center. This changes the transportation cost by $\epsilon(d_{i-1,j-1} - d_{i-1,n})$. Since $i-1 \leq j-1 \leq n$ the induction hypothesis implies that this cost difference is again negative. After completing these flow transfers we have $\sum_{i=0}^{m} z_{ij} = \pi_j^n$ for all $j = 0, \ldots, n$ and a fortiori $\sum_{i=0}^{m} z_{i,n+1} = 0$. Hence the transformed z is feasible for d_{mn} and therefore $d_{mn} \leq C^{mn}(z)$. Since along the transformation the cost was reduced from its initial value $d_{m,n+1}$, we conclude $d_{mn} \leq d_{m,n+1}$, completing the induction step.

The proof of $d_{m,n-1} \geq d_{mn}$ for $n \leq m$ is similar. The base case $n = 0$ is trivial since $d_{m,-1} = 1 \geq d_{m0}$. For the induction step, take $z \in F^{m,n-1}$ a simple optimal solution for $d_{m,n-1}$. Let us transform z into a feasible solution for d_{mn} while reducing the cost. To this end, for each $j = 0, \ldots, n-1$ we take some positive flow $z_{ij} > 0$ with $i \geq n$ from which we remove a positive amount ε and reroute it to node n by increasing z_{in}, as shown in Figure 9 right. As a result, the cost changes by $\varepsilon(d_{i-1,n-1} - d_{i-1,j-1})$, which is negative by the induction hypothesis since $j-1 \leq n-1 \leq i-1$. We repeat these flow transfers until the inflow $\sum_{i=0}^{m} z_{ij}$ is reduced from π_j^{n-1} to π_j^n. The difference $\pi_j^{n-1} - \pi_j^n = \alpha_n \pi_j^{n-1}$ is now redirected to node n and therefore

$$\sum_{i=0}^{m} z_{in} = \sum_{j=0}^{n-1} \alpha_n \pi_j^{n-1} = \alpha_n = \pi_n^n.$$

It follows that the transformed flow is feasible for d_{mn} and therefore $d_{mn} \leq C^{mn}(z) \leq d_{m,n-1}$. ∎

In order to establish the 4-point inequality (d) we will exploit inductively the fact that the inside-out algorithm in Remark 2.3 computes an optimal transport. More formally, we will use the following

Lemma A.1: *Suppose that $\alpha_n \geq \frac{1}{2}$ for all $n \in \mathbb{N}$. Let $n \in \mathbb{N}$ and assume that the 4-point inequality holds for all $0 \leq i \leq k \leq j \leq l < n$. Then for all $m \leq n$ the inside-out procedure computes an optimal transport $z = z^{mn}$ given by*

$$\begin{aligned} z_{ii} &= \pi_i^n && \text{for } i = 0, \ldots, m, \\ z_{mj} &= \pi_j^n && \text{for } j = m+1, \ldots, n-1, \\ z_{in} &= \pi_i^m - \pi_i^n && \text{for } i = 0, \ldots, m-1, \\ z_{mn} &= \pi_m^m - \sum_{j=m}^{n-1} \pi_j^n. \end{aligned}$$

Proof. Let $0 \leq m \leq n$. Since the costs in $(\mathcal{P}_{mn})$ are the distances $d_{i-1,j-1}$ with the nodes shifted by -1, we have the 4-point inequalities required to argue, as in Remark 2.3, that there is an optimal transport in which the flows do not cross, which is precisely the one computed by the inside-out algorithm. Now, this procedure sets $z_{ii} = \pi_i^n$ for $i = 0, \ldots, m$. Next, we note that $z_{mj} = \pi_j^n$ holds for all $j = m, \ldots, n-1$ if and only if the supply of node m suffices to satisfy the demands of all nodes $j = m, \ldots, n-1$, that is, $\pi_m^m \geq \sum_{j=m}^{n-1} \pi_j^n$. This is precisely the case, since $\pi_m^m = \alpha_m \geq \frac{1}{2}$ while $\sum_{j=m}^{n-1} \pi_j^n \leq 1 - \pi_n^n = 1 - \alpha_n \leq \frac{1}{2}$. Hence we have $z_{mj} = \pi_j^n$ for $j = m+1, \ldots, n-1$ and thus the remaining supplies are sent to node n, namely $z_{in} = \pi_i^m - \pi_i^n$ for $i = 0, \ldots, m-1$ and $z_{mn} = \pi_m^m - \sum_{j=m}^{n-1} \pi_j^n$. ∎

Remark A.2: In the constant case $\alpha_n \equiv \alpha$ the condition $\alpha \geq \frac{1}{2}$ is not only sufficient but also necessary for the optimal transport to be as decribed in the Lemma. Indeed, in the constant case the condition $\pi_m^m \geq \sum_{j=m}^{n-1} \pi_j^n$ is equivalent to $\alpha \geq \beta - \beta^{n-m+1}$ which holds for all $0 \leq m \leq n$ iff $\alpha \geq \frac{1}{2}$.

We now proceed to establish the 4-point inequality.

(d) *If* $\alpha_n \geq 1/2$ *for all* $n \in \mathbb{N}$, *then for all integers* $0 \leq i \leq k \leq j \leq l$ *we have* $d_{il} + d_{kj} \leq d_{ij} + d_{kl}$.

Proof. For $k = j$ this is the triangle inequality which was already proved, so let us consider the case $k < j$. Let us fix $i \in \mathbb{N}$ and let us prove by induction that for all $l > i$ we have

$$(I_l) \qquad d_{il} + d_{kj} \leq d_{ij} + d_{kl} \quad \text{for all } i \leq k \leq j \leq l.$$

The base case $l = i+1$ reads $d_{i,i+1} + d_{kj} \leq d_{ij} + d_{k,i+1}$ with $k = i$ and $j = i+1$, which holds trivially. For the induction step, we will exploit the following equivalence.

Claim A.3: *For each* $\ell \in \mathbb{N}$ *the following are equivalent:*

(a) *For all integers* $0 \leq i \leq k < j \leq l \leq \ell$ *we have*

$$d_{il} + d_{kj} \leq d_{ij} + d_{kl}.$$

(b) *For all integers* $0 \leq m < n < \ell$ *we have*

$$d_{m,n+1} + d_{m+1,n} \leq d_{m,n} + d_{m+1,n+1}.$$

Proof of Claim A.3. Clearly (b) is the special case of (a) with $i=m$, $k=m+1$, $j = n$, $l = n+1$. In order to show that, conversely, (b) implies (a), take $0 \leq i \leq k < j \leq l \leq \ell$. By telescoping we have $d_{il} - d_{ij} = \sum_{n=j}^{l-1} d_{i,n+1} - d_{in}$ and since from (b) we have that the difference $d_{i,n+1} - d_{in}$ increases with i, we deduce $d_{il} - d_{ij} \leq \sum_{n=j}^{l-1} d_{k,n+1} - d_{kn} = d_{kl} - d_{kj}$ which yields (a). ∎

Induction step: Let us assume that (I_l) holds. In view of Claim A.3, in order to prove (I_{l+1}) it suffices to show that for $n = l+1$ one has

$$\Delta := d_{m,n} + d_{m+1,n+1} - d_{m,n+1} - d_{m+1,n} \geq 0.$$

Since $\alpha_n \geq \frac{1}{2}$ and thanks to the induction hypothesis (I_l), we can use Lemma A.1 to obtain the following explicit expressions for each of the 4 terms appearing in Δ:

$$\begin{aligned}
d_{m,n} &= \sum_{j=m+1}^{n-1} \pi_j^n d_{m-1,j-1} + \sum_{i=0}^{m-1} (\pi_i^m - \pi_i^n) d_{i-1,n-1} + \left(\pi_m^m - \sum_{j=m}^{n-1} \pi_j^n\right) d_{m-1,n-1} \\
&= \sum_{j=m+1}^{n} \pi_j^n [d_{m-1,j-1} - d_{m-1,n-1}] + \sum_{i=0}^{m} (\pi_i^m - \pi_i^n) d_{i-1,n-1}, \\
d_{m+1,n+1} &= \sum_{j=m+2}^{n} \pi_j^{n+1} [d_{m,j-1} - d_{m,n}] + \sum_{i=0}^{m+1} (\pi_i^{m+1} - \pi_i^{n+1}) d_{i-1,n} \\
&= \sum_{j=m+2}^{n} \pi_j^{n+1} [d_{m,j-1} - d_{m,n-1}] + \sum_{i=0}^{m+1} (\pi_i^{m+1} - \pi_i^{n+1}) d_{i-1,n} \\
&\quad + \sum_{j=m+2}^{n} \pi_j^{n+1} [d_{m,n-1} - d_{m,n}] \\
&= \sum_{j=m+1}^{n} \pi_j^{n+1} [d_{m,j-1} - d_{m,n-1}] + \sum_{i=0}^{m} (\pi_i^{m+1} - \pi_i^{n+1}) d_{i-1,n} \\
&\quad + \sum_{j=m+1}^{n} \pi_j^{n+1} [d_{m,n-1} - d_{m,n}] + \pi_{m+1}^{m+1} d_{m,n}, \\
d_{m,n+1} &= \sum_{j=m+1}^{n} \pi_j^{n+1} [d_{m-1,j-1} - d_{m-1,n}] + \sum_{i=0}^{m} (\pi_i^m - \pi_i^{n+1}) d_{i-1,n} \\
&= \sum_{j=m+1}^{n} \pi_j^{n+1} [d_{m-1,j-1} - d_{m-1,n-1}] + \sum_{i=0}^{m} (\pi_i^m - \pi_i^{n+1}) d_{i-1,n} \\
&\quad + \sum_{j=m+1}^{n} \pi_j^{n+1} [d_{m-1,n-1} - d_{m-1,n}], \\
d_{m+1,n} &= \sum_{j=m+2}^{n-1} \pi_j^n [d_{m,j-1} - d_{m,n-1}] + \sum_{i=0}^{m+1} (\pi_i^{m+1} - \pi_i^n) d_{i-1,n-1} \\
&= \sum_{j=m+1}^{n} \pi_j^n [d_{m,j-1} - d_{m,n-1}] + \sum_{i=0}^{m} (\pi_i^{m+1} - \pi_i^n) d_{i-1,n-1} + \pi_{m+1}^{m+1} d_{m,n-1}.
\end{aligned}$$

Replacing these expressions in Δ and rearranging terms we can rewrite $\Delta = A + B + C$ with

$$A = \sum_{j=m+1}^{n} (\pi_j^n - \pi_j^{n+1})[d_{m-1,j-1} - d_{m-1,n-1} - d_{m,j-1} + d_{m,n-1}] \geq 0,$$

$$B = \sum_{i=0}^{m} (\pi_i^m - \pi_i^{m+1})[d_{i-1,n-1} - d_{i-1,n} + d_{m-1,n} - d_{m-1,n-1}] \geq 0,$$

$$\begin{aligned} C = &\sum_{i=0}^{m} (\pi_i^m - \pi_i^{m+1})[d_{m-1,n-1} - d_{m-1,n}] \\ &+ \sum_{j=m+1}^{n} \pi_j^{n+1}[d_{m,n-1} - d_{m,n}] + \pi_{m+1}^{m+1} d_{m,n} \\ &- \sum_{j=m+1}^{n} \pi_j^{n+1}[d_{m-1,n-1} - d_{m-1,n}] - \pi_{m+1}^{m+1} d_{m,n-1} \\ = &\left(\pi_{m+1}^{m+1} - \sum_{j=m+1}^{n} \pi_j^{n+1}\right)[d_{m-1,n-1} - d_{m-1,n} - d_{m,n-1} + d_{m,n}] \geq 0, \end{aligned}$$

where the positivity of the three quantities A, B, C follows from the induction hypothesis. ∎

Appendix B. Proof of Proposition 3.1

For $\alpha_n \equiv \alpha \in [\frac{1}{2}, 1[$ *we have* $0 \leq c_{mn} - d_{mn} \leq 4m(1-\alpha)^2$.

Proof. Starting from the estimate (12), a simple application of the Mean Value Theorem yields

$$|c_{mn} - d_{mn}| \leq m \|\mathbf{D} - \mathbf{C}\|_\infty \tag{16}$$

so it suffices to prove that $\|\mathbf{D} - \mathbf{C}\|_\infty \leq 4(1-\alpha)^2$. Hence, denoting $\beta = 1 - \alpha$ we must show that for each $s \in \mathcal{S}$ the sum $\Gamma(s) = \sum_{s' \in \mathcal{S}} |\mathbf{D}(s, s') - \mathbf{C}(s, s')|$ is at most $4\beta^2$.

For the absorbing states we have $\Gamma(f) = \Gamma(h) = 0$ so we just consider a transient state $s = mn \in \mathcal{S}$. For $s' = f$ we have

$$\mathbf{D}(mn, f) = \sum_{i=0}^{m} \pi_i^n = \mathbf{C}(mn, f),$$

whereas for $s' = h$ we have

$$\mathbf{D}(mn,h) = \pi_0^m - \pi_0^n = \pi_0^m \sum_{j=m+1}^{n} \pi_j^n = \mathbf{C}(mn,h).$$

It follows that in the sum $\Gamma(mn)$ we can ignore the terms $s' = f, h$ so

$$\Gamma(mn) = \sum_{kl\in\mathcal{S}} |\mathbf{D}(mn,kl) - \mathbf{C}(mn,kl)| = \sum_{1\le i\le m<j\le n} |z_{ij}^{mn} - \tilde{z}_{ij}^{mn}|, \tag{17}$$

where z^{mn} is the optimal transport described in Lemma A.1 and $\tilde{z}^{mn}$ is given by (11).

Using the expressions of z^{mn} and $\tilde{z}^{mn}$ we get explicitly

$$|z_{ij}^{mn} - \tilde{z}_{ij}^{mn}| = \begin{cases} \alpha^2\beta^{m-i+n-j} & \text{if } i = 1,\dots,m-1 \\ & \text{and } j = m+1,\dots,n-1, \\ \alpha\beta^{m-i+1}(1-\beta^{n-m-1}) & \text{if } i = 1,\dots,m-1 \text{ and } j = n, \\ \alpha\beta^{n-j+1} & \text{if } i = m \text{ and } j = m+1,\dots,n-1, \\ \beta^2(1-\beta^{n-m-1}) & \text{if } i = m \text{ and } j = n, \\ 0 & \text{otherwise,} \end{cases}$$

which replaced into (17) yield the conclusion

$$\begin{aligned}\Gamma(mn) =& \alpha^2 \sum_{i=1}^{m-1}\sum_{j=m+1}^{n-1} \beta^{m-i+n-j} + \alpha(1-\beta^{n-m-1}) \sum_{i=1}^{m-1} \beta^{m-i+1} \\ &+ \alpha \sum_{j=m+1}^{n-1} \beta^{n-j+1} + \beta^2(1-\beta^{n-m-1}) \le 4\beta^2. \qquad \blacksquare\end{aligned}$$

References

[1] N. Aronszajn and P. Panitchpakdi, *Extension of uniformly continuous transformations and hyperconvex metric spaces*, Pacific Journal of Mathematics **6** (1956), 405–439.

[2] Y. Aygen-Satik, *Optimal Bounds of Asymptotic Regularity*, Ph.D. Thesis, University of Southern California, 1994.

[3] J. B. Baillon and R. E. Bruck, *Optimal rates of asymptotic regularity for averaged nonexpansive mappings*, In *Fixed Point Theory and Applications (Halifax, NS, 1991)*, World Scientific, River Edge, NJ, 1992, pp. 27–66.

[4] J. B. Baillon and R. E. Bruck, *The rate of asymptotic regularity is* $O(1/\sqrt{n})$, in *Theory and Applications of Nonlinear Operators of Accretive and Monotone Type*, Lecture Notes in Pure and Applied Mathematics, Vol. 178, Dekker, New York, 1996, pp. 51–81.

[5] J.-B. Baillon, R. Cominetti and J. Vaisman, *A sharp uniform bound for the distribution of sums of Bernoulli trials*, Combinatorics, Probability and Computing **25** (2016), 352–361.

[6] H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, New York, 2011.

[7] J. Borwein, S. Reich and I. Shafrir, *Krasnosel'skiĭ-Mann iterations in normed spaces*, Canadian Mathematical Bulletin **35** (1992), 21–28.

[8] F. E. Browder and W. V. Petryshyn, *The solution by iteration of nonlinear functional equations in Banach spaces*, Bulletin of the American Mathematical Society **72** (1966), 571–575.

[9] F. E. Browder and W. V. Petryshyn, *Construction of fixed points of nonlinear mappings in Hilbert space*, Journal of Mathematical Analysis and Applications **20** (1967), 197–228.

[10] R. E. Bruck, *Asymptotic behavior of nonexpansive mappings*, in *Fixed Points and Nonexpansive Mappings (Cincinnati, Ohio, 1982)*, Contemporary Mathematics, Vol. 18, American Mathematical Society, Providence, RI, 1983, pp. 1–47.

[11] R. Cominetti, J. Soto and J. Vaisman, *On the rate of convergence of Krasnosel'skiĭ–Mann iterations and their connection with sums of Bernoullis*, Israel Journal of Mathematics **199** (2014), 757–772.

[12] M. Edelstein, *A remark on a theorem of M. A. Krasnosel'skiĭ*, American Mathematical Monthly **73** (1966), 509–510.

[13] M. Edelstein and R. C. O'Brien, *Nonexpansive mappings, asymptotic regularity and successive approximations*, Journal of the London Mathematical Society **17** (1978), 547–554.

[14] K. Goebel and W. A. Kirk, *Iteration processes for nonexpansive mappings*, in *Topological Methods in Nonlinear Functional Analysis (Toronto, ON, 1982)*, Contemporary Mathematics, Vol. 21, American Mathematical Society, Providence, RI, 1983, pp. 115–123.

[15] C. W. Groetsch, *A note on segmenting Mann iterates*, Journal of Mathematical Analysis and Applications **40** (1972), 369–372.

[16] S. Ishikawa, *Fixed points and iterations of a nonexpansive mapping in a Banach space*, Proceedings of the American Mathematical Society **59** (1976), 65–71.

[17] U. Kohlenbach, *Uniform asymptotic regularity for Mann iterates*, Journal of Mathematical Analysis and Applications **279** (2003), 531–544.

[18] M. A. Krasnosel'skiĭ, *Two remarks on the method of successive approximations*, Uspekhi Matematicheskikh Nauk **10** (1955), 123–127.

[19] W. R. Mann, *Mean value methods in iteration*, Proceedings of the American Mathematical Society **4** (1953), 506–510.

[20] S. Reich, *Weak convergence theorems for nonexpansive mappings in Banach spaces*, Journal of Mathematical Analysis and Applications **67** (1979), 274–276.

[21] S. Reich and I. Shafrir, *Nonexpansive iterations in hyperbolic spaces*, Nonlinear Analysis **15** (1990), 537–558.

[22] H. Schaefer, *Über die Methode sukzessiver Approximationen*, Jahresbericht der Deutschen Mathematiker-Vereinigung **59** (1957), 131–140.