

UNIVERSAL BOUNDS FOR FIXED POINT ITERATIONS VIA OPTIMAL TRANSPORT METRICS

MARIO BRAVO¹, THIERRY CHAMPION², ROBERTO COMINETTI^{1,*}

¹*Facultad de Ingeniería y Ciencias, Universidad Adolfo Ibáñez, Diagonal Las Torres 2640, Santiago, Chile.*

²*Laboratoire Imath, U.F.R. des Sciences et Techniques, Université de Toulon, Av. de l'Université, B.P. 20132, 83957 La Garde cedex, France.*

Dedicated to the memory of Professor Ronald E. Bruck

Abstract. We present a self-contained analysis of a particular family of metrics over the set of non-negative integers. We show that these metrics, which are defined through a nested sequence of optimal transport problems, provide tight estimates for general Krasnosel'skii-Mann fixed point iterations for non-expansive maps. We also describe some of their very special properties, including their monotonicity and the so-called *convex quadrangle inequality* that yields a greedy algorithm for computing them efficiently.

Keywords. Fixed-point iterations, non-expansive maps, error bounds, convergence rates, optimal transport metrics.

1. INTRODUCTION

This paper studies a special family of metrics over the set of non-negative integers, in which the distances $d_{m,n}$ between $m, n \in \mathbb{N}$ are defined recursively through a nested family of optimal transport problems. These metrics were first introduced in a remarkable paper by Baillon & Bruck [1] with the aim of establishing convergence rates for the Krasnosel'skii fixed-point iteration, although their metric nature and the connection with optimal transport were not noticed at that time. The metric properties of the $d_{m,n}$'s were studied extensively in Aygen-Satik's thesis [2], with a *tour-de-force* that required very long and highly technical proofs. Several of these results were revisited in [3, Bravo & Cominetti] with a much simpler approach based on optimal transport. The goal of this paper is to complete this program by presenting a full and self-contained analysis of these metrics, and their implications for fixed point iterations. In doing so, we expand the scope of [1, 2, 3] by considering a much larger family of iterations and metrics.

To get into the matter, throughout this paper we consider a fixed sequence $(\pi^n)_{n \in \mathbb{N}}$ where each $\pi^n = (\pi_i^n)_{i \in \mathbb{N}}$ is a discrete probability distribution on the set of non-negative integers, with support included in $\{0, \dots, n\}$ and with $\pi^n \neq \pi^m$ for $m \neq n$. Some results assume in addition that these distributions gradually drift their mass towards larger integers, namely (see Figure 1)

$$(H) \quad (\forall n \geq 1) \quad \pi_n^n > 0 \text{ and } 0 \leq \pi_i^n \leq \pi_i^{n-1} \text{ for } i < n.$$

*Corresponding author.

E-mail addresses: <mario.bravo@uai.cl>, <champion@univ-tln.fr>, <roberto.cominetti@uai.cl>

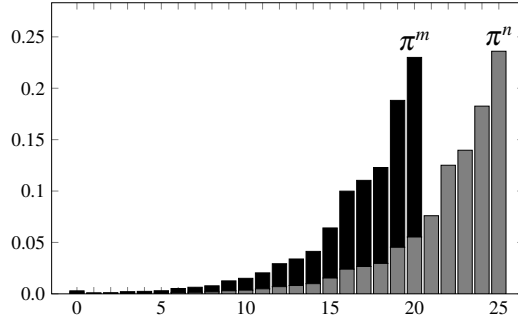


FIGURE 1. The distributions π^m and π^n ($m = 20, n = 25$).

1.1. Recursive optimal transports. Starting with $d_{-1,j} = d_{j,-1} = 1$ for $j \in \mathbb{N}$ and $d_{-1,-1} = 0$, we consider the double-indexed family of reals $d_{m,n}$ defined recursively for $m, n \in \mathbb{N}$ through the optimal transport problems

$$(\mathcal{P}_{m,n}) \quad d_{m,n} = \min_{z \in \mathcal{F}_{m,n}} C_{m,n}(z) \triangleq \sum_{i=0}^m \sum_{j=0}^n z_{i,j} d_{i-1,j-1}$$

where $\mathcal{F}_{m,n}$ is the polytope of transport plans sending π^m to π^n , that is, the set of all $z = (z_{i,j})_{i=0,\dots,m;j=0,\dots,n}$ such that $z_{i,j} \geq 0$ and

$$\sum_{j=0}^n z_{i,j} = \pi_i^m \quad \text{for all } i = 0, \dots, m;$$

$$\sum_{i=0}^m z_{i,j} = \pi_j^n \quad \text{for all } j = 0, \dots, n.$$

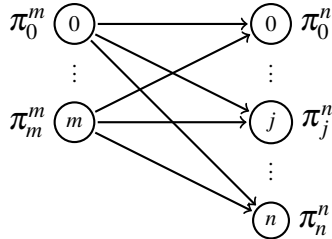


FIGURE 2. The optimal transports for $d_{m,n}$.

Each sequence $(\pi^n)_{n \in \mathbb{N}}$ produces a different set of $d_{m,n}$'s. In particular we have $d_{0,n} = 1 - \pi_0^n$, though in general the other $d_{m,n}$'s do not admit a simple explicit formula. We observe that a simple induction yields the symmetry $d_{m,n} = d_{n,m}$ so it suffices to compute $d_{m,n}$ for $m \leq n$. Also, since the transports $z_{i,j}$ add up to one, the cost $C_{m,n}(z)$ is a convex combination of the previous $d_{i-1,j-1}$'s and inductively we get $d_{m,n} \in [0, 1]$ with $d_{n,n} = 0$.

1.2. Fixed point iterations. The optimal transports $d_{m,n}$ arise in connection with fixed point iterations. Namely, let $T : C \rightarrow C$ be a non-expansive map on a bounded convex domain $C \subseteq X$ of a normed space $(X, \|\cdot\|)$. Notice that when $T : X \rightarrow X$ is defined on the full space and has some fixed point $x^* = Tx^*$, one may take $C = B(x^*, r)$ as any ball centered at x^* with radius $r \geq 0$. On the other hand, by rescaling the norm by a factor $1/\text{diam}(C)$ we may assume without loss of generality that $\text{diam}(C) = 1$, which we do from now on.

Starting from $x^0, y^0 \in C$, and setting by convention $Tx^{-1} = y^0$, the general Krasnosel'skii-Mann fixed point iteration (see [4, Mann] and [5, Krasnosel'skii]) recursively builds a sequence x^n as a convex combination of the images of the previous iterates

$$(KM) \quad x^n = \sum_{i=0}^n \pi_i^n T x^{i-1} \quad \forall n \geq 1.$$

The following straightforward result brings forward the connection with the recursive optimal transports, showing how they provide bounds for the distance between the (KM) iterates and for the fixed point residuals.

Theorem 1.1. *For all $m, n \in \mathbb{N}$ we have $\|x^m - x^n\| \leq d_{m,n}$ and*

$$\|x^n - T x^n\| \leq R_n \triangleq \sum_{i=0}^n \pi_i^n d_{i-1,n}. \quad (1.1)$$

Proof. Let us first show that $\|x^m - x^n\| \leq d_{m,n}$. The cases $m = 0$ and $n = 0$ follow directly from $\|x^0 - x^n\| \leq \text{diam}(C) = 1 = d_{-1,n}$. For the general case, we observe that each transport plan z from π^m to π^n yields the estimate

$$\begin{aligned} \|x^m - x^n\| &= \left\| \sum_{i=0}^m \pi_i^m T x^{i-1} - \sum_{j=0}^n \pi_j^n T x^{j-1} \right\| \\ &= \left\| \sum_{i=0}^m \sum_{j=0}^n z_{i,j} (T x^{i-1} - T x^{j-1}) \right\| \\ &\leq \sum_{i=0}^m \sum_{j=0}^n z_{i,j} \|T x^{i-1} - T x^{j-1}\| \\ &\leq \sum_{i=0}^m \sum_{j=0}^n z_{i,j} d_{i-1,j-1} \end{aligned}$$

where the last inequality uses the non-expansivity of T and assumes inductively that we already have $\|x^{i-1} - x^{j-1}\| \leq d_{i-1,j-1}$ for the previous iterates (for $i = 0$ or $j = 0$ use the convention $Tx^{-1} = y^0$ and the coarse estimate $\|y^0 - T x^k\| \leq \text{diam}(C) = 1 = d_{-1,k-1}$). Minimizing over z we get $\|x^m - x^n\| \leq d_{m,n}$ and then the proof is completed by induction. The bound (1.1) follows directly from the triangle inequality and non-expansivity, namely

$$\|x^n - T x^n\| = \left\| \sum_{i=0}^n \pi_i^n (T x^{i-1} - T x^n) \right\| \leq \sum_{i=0}^n \pi_i^n d_{i-1,n}. \quad \square$$

We emphasize that the bounds $d_{m,n}$ and R_n are universal in the sense that they only depend on the sequence $(\pi^n)_{n \in \mathbb{N}}$ and not on the particular map T being considered. Moreover, the estimate $\|x^m - x^n\| \leq d_{m,n}$ turns out to be tight and cannot be improved unless we restrict the class of maps or spaces. In addition, we will show that under (H) the bound $\|x^n - T x^n\| \leq R_n$ is also sharp, in which case R_n captures the exact convergence rate of (KM).

The recursion (KM) is very general and includes among others the Krasnosel'skii iteration (see [5, Krasnosel'skii] and [6, Bruck])

$$x^n = (1 - \alpha_n) x^{n-1} + \alpha_n T x^{n-1}, \quad \alpha_n \in (0, 1),$$

as well as Halpern's method (see [7, Halpern] and [8, López *et al.*])

$$x^n = (1 - \beta_n) y^0 + \beta_n T x^{n-1}, \quad \beta_n \in (0, 1),$$

and the 2-step iteration of Ishikawa (see [9, Ishikawa])

$$\begin{cases} x^{2n+1} &= (1 - \beta_n) x^{2n} + \beta_n T x^{2n} \\ x^{2n+2} &= (1 - \alpha_n) x^{2n} + \alpha_n T x^{2n+1} \end{cases}, \quad 0 \leq \alpha_n \leq \beta_n \leq 1.$$

The Krasnosel'skii iteration is the case where $\pi^n = (1 - \alpha_n) \pi^{n-1} + \alpha_n \delta^n$ with δ^n a Dirac at n , for which we have explicitly $\pi_i^n = \alpha_i \prod_{k=i+1}^n (1 - \alpha_k)$ with $\alpha_0 = 1$ and $\pi_n^n = \alpha_n$. Its convergence

rate was studied in a series of papers [1, 2, 3, 10, 11]. A first result in [1, Baillon & Bruck] for $\alpha_n \equiv \alpha$ showed that $\|x^n - Tx^n\| \sim O(\frac{1}{\ln n})$. Shortly after, in [10, Baillon & Bruck] this rate was improved to $\|x^n - Tx^n\| \sim O(\frac{1}{\sqrt{n}})$, with an explicit upper bound recently extended to non-constant α_n 's in [11, Cominetti *et al.*]. The papers [1, 10] are two gems among the many lasting contributions of Professor Ronald E. Bruck to the field of fixed point theory. A more detailed discussion of these previous results, as well as for Halpern iteration, is postponed to section §4.

1.3. Summary of results. This paper investigates the metric properties of the optimal transport bounds $d_{m,n}$. Section §2 shows that $(m,n) \mapsto d_{m,n}$ defines a distance on $\overline{\mathbb{N}} \triangleq \mathbb{N} \cup \{-1\}$, which in turn leads to an alternative characterization via Kantorovich-Rubinstein duality.

This dual characterization is exploited in section §3 to prove that the estimates $\|x^m - x^n\| \leq d_{m,n}$ in Theorem 1.1 are the best possible, by building a particular non-expansive map T and a corresponding (KM) sequence that attains these bounds with equality. Moreover, we prove that under (H) the bounds $\|x^n - Tx^n\| \leq R_n$ for the residuals are also tight.

Section §4 illustrates these bounds in some specific iterations. For Halpern with $\beta_n = \frac{n}{n+1}$ we obtain $R_n = \frac{H_{n+1}}{n+1} \sim O(\frac{\ln n}{n})$ with $H_n = \sum_{k=1}^n \frac{1}{k}$ the n -th harmonic number, while for $\beta_n = \frac{n}{n+2}$ we have $R_n = \frac{4}{n+1}(1 - \frac{H_{n+2}}{n+2}) \sim O(\frac{1}{n})$, both rates being tight for the corresponding β_n 's.

In section §5 we show that under (H) the $d_{m,n}$'s enjoy two relevant additional properties: the monotonicity as m and n get farther apart, and the so-called *convex quadrangle inequality*

$$(Q) \quad d_{i,l} + d_{j,k} \leq d_{i,k} + d_{j,l} \quad \text{for all } i \leq j \leq k \leq l.$$

Note that for $j = k$ this is just a triangle inequality.

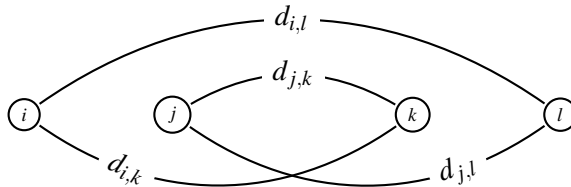


FIGURE 3. The convex quadrangle inequality.

A remarkable consequence of (Q) is the existence of optimal transports that are *nested* in the sense that the flows do not intersect. This yields an efficient greedy algorithm to compute $d_{m,n}$. This algorithm is used in §6 for various examples, revealing some particularly intriguing structures in the $d_{m,n}$'s. We mention that the quadrangle inequality is closely related to the *inverse Monge property* [12], and arises in various contexts with relevant algorithmic implications (see [13, 14, 15, 16, 17]).

2. THE RECURSIVE OPTIMAL TRANSPORT METRIC

Let us start by proving that the $d_{m,n}$'s define a distance. This extends a result by [2, Aygen-Satik] proved for the special case of the Krasnosel'skii iteration with $\pi^n = (1 - \alpha_n)\pi^{n-1} + \alpha_n\delta^n$. The original proof was highly technical, though it was greatly simplified in [3] for the special case where $\alpha_n \geq \frac{1}{2}$ for all $n \in \mathbb{N}$. This simpler proof is reproduced below, suitably adapted to the more general class of π^n 's considered here.

Theorem 2.1. *The map $(m,n) \mapsto d_{m,n}$ is a metric on $\overline{\mathbb{N}} = \mathbb{N} \cup \{-1\}$.*

Proof. We already observed that $d_{m,n} = d_{n,m}$ and $d_{n,n} = 0$, so we only need to establish the triangle inequality and that $d_{m,n} > 0$ for $m \neq n$. We show inductively that these properties hold for $m, n \leq \ell$ for each $\ell \in \overline{\mathbb{N}}$. The base case $\ell = -1$ is trivial. Suppose that both properties hold up to $\ell - 1$ and let us prove them for ℓ .

For $m, n \leq \ell$ with $m \neq n$ we have $d_{m,n} > 0$: Let z be an optimal transport for $d_{m,n}$. Since $\pi^m \neq \pi^n$ we can find $i \neq j$ with $z_{i,j} > 0$ and the induction hypothesis implies $d_{m,n} \geq z_{i,j} d_{i-1,j-1} > 0$.

For $m, n, p \leq \ell$ we have $d_{m,n} \leq d_{m,p} + d_{p,n}$: Let $z^{m,p}$ and $z^{p,n}$ be optimal transports for $d_{m,p}$ and $d_{p,n}$ respectively, and define

$$(\forall i = 0, \dots, m)(\forall j = 0, \dots, n) \quad z_{i,j} = \sum_{k=0}^p \omega_{i,k,j}$$

with

$$\omega_{i,k,j} = \begin{cases} \frac{z_{i,k}^{m,p} z_{k,j}^{p,n}}{\pi_k^p} & \text{if } \pi_k^p \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

A straightforward computation shows that

$$\begin{cases} \sum_{i=0}^m \omega_{i,k,j} & = z_{k,j}^{p,n} \\ \sum_{j=0}^n \omega_{i,k,j} & = z_{i,k}^{m,p} \end{cases} \quad (2.1)$$

from which it readily follows that $z \in \mathcal{F}_{m,n}$. Using the induction hypothesis and (2.1) we get

$$\begin{aligned} d_{m,n} &\leq \sum_{i=0}^m \sum_{j=0}^n z_{i,j} d_{i-1,j-1} \\ &= \sum_{i=0}^m \sum_{j=0}^n \sum_{k=0}^p \omega_{i,j,k} d_{i-1,j-1} \\ &\leq \sum_{i=0}^m \sum_{j=0}^n \sum_{k=0}^p \omega_{i,j,k} (d_{i-1,k-1} + d_{k-1,j-1}) \\ &= \sum_{i=0}^m \sum_{k=0}^p z_{i,k}^{m,p} d_{i-1,k-1} + \sum_{j=0}^n \sum_{k=0}^p z_{k,j}^{p,n} d_{k-1,j-1} \\ &= d_{m,p} + d_{p,n}. \end{aligned} \quad \square$$

Now, part of the demand π_i^n can be fulfilled at zero cost by shipping as much as possible from the corresponding supply node i . This suggests the following type of simple transport plans.

Definition 2.1. A transport z from π^m to π^n with $m \leq n$ is called *simple* if $z_{i,i} = \min\{\pi_i^m, \pi_i^n\}$ for all $i = 0, \dots, m$.

Remark 2.1. If z is a simple transport from π^m to π^n , then when $\pi_i^m \geq \pi_i^n$ we have $z_{i,i} = \pi_i^n$ and therefore $z_{j,i} = 0$ for $j \neq i$. Similarly, if $\pi_i^m \leq \pi_i^n$ then $z_{i,i} = \pi_i^m$ and $z_{i,k} = 0$ for $k \neq i$.

As a consequence of the triangle inequality, it turns out that we may always restrict ourselves to transports that are simple.

Proposition 2.1. *Each $d_{m,n}$ admits a simple optimal transport.*

Proof. Let z be an optimal transport. If $z_{i,i} < \min\{\pi_i^m, \pi_i^n\}$ we must have $z_{i,k} > 0$ for some $k \neq i$ and $z_{j,i} > 0$ for some $j \neq i$. Decreasing $z_{j,i}$ and $z_{i,k}$ by ε while increasing $z_{i,i}$ and $z_{j,k}$ by the same amount (see Figure 4), the modified transport is still feasible and the cost is reduced by

$$[d_{i-1,i-1} + d_{j-1,k-1} - d_{j-1,i-1} - d_{i-1,k-1}] \varepsilon \leq 0$$

so it remains optimal. Thus we can increase each $z_{i,i}$ up to $\min\{\pi_i^m, \pi_i^n\}$. \square

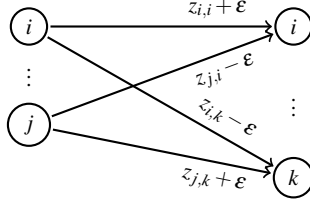
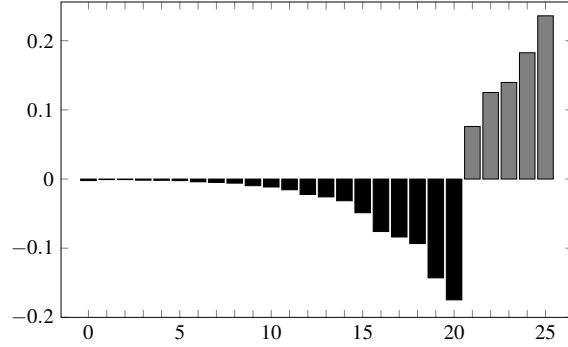


FIGURE 4. Redistribution of flow for simple optimal transports

FIGURE 5. Residual demands $\delta_j^{mn} = \pi_j^n - \pi_j^m$ ($m = 20, n = 25$).

A further consequence of the fact that the transport cost $(i, j) \mapsto d_{i-1, j-1}$ is a distance, is the following Kantorovich-Rubinstein dual characterization for all $m \leq n$ (see *e.g.* [18, Villani, Section 1.2]) where $\delta_j^{mn} = \pi_j^n - \pi_j^m$ denote the residual demands (see Figure 5)

$$(\mathcal{D}_{m,n}) \quad d_{m,n} = \max_{u \in \mathbb{R}^{n+1}} \left\{ \sum_{j=0}^n \delta_j^{mn} u_j : u_j \leq u_i + d_{i-1, j-1} \quad \forall i, j = 0, \dots, n \right\}.$$

This is a refinement of linear programming duality, and each pair of primal-dual optimal solutions z^{mn} and u^{mn} satisfy the complementary slackness

$$z_{i,j}^{mn} (u_j^{mn} - u_i^{mn}) = z_{i,j}^{mn} d_{i-1, j-1} \quad \text{for all } i, j = 0, \dots, n. \quad (2.2)$$

Remark 2.2. Every feasible $(u_i)_{i=0}^n$ in $(\mathcal{D}_{m,n})$ can be extended by setting $u_i = \min_{0 \leq k \leq n} u_k + d_{k-1, i-1}$ for $i > n$, so that the triangle inequality yields

$$|u_i - u_j| \leq d_{i-1, j-1} \quad \text{for all } i, j \in \mathbb{N}. \quad (2.3)$$

This is a special case of the MacShane-Whitney extension of Lipschitz functions. In particular all the u_i 's are within a distance at most 1 and, since the objective function in $(\mathcal{D}_{m,n})$ is invariant under translation, we may further assume that $u_i \in [0, 1]$ for all $i \in \mathbb{N}$.

Remark 2.3. Under (H) there is a unique simple optimal transport from π^n to π^{n+1} with $z_{i,i} = \pi_i^{n+1}$ and $z_{i, n+1} = \pi_i^n - \pi_i^{n+1}$ for $0 \leq i \leq n$. Moreover, setting $u_i = 1 - d_{i-1, n} \in [0, 1]$ for all $0 \leq i \leq n+1$, the triangle inequality implies that u is a feasible solution for $(\mathcal{D}_{n, n+1})$ which is also optimal because

$$\sum_{i=0}^{n+1} \delta_i^{n, n+1} u_i = \sum_{i=0}^n z_{i, n+1} d_{i-1, n} = d_{n, n+1}.$$

3. TIGHTNESS OF THE OPTIMAL TRANSPORT BOUNDS

Exploiting the dual characterization of the distances $d_{m,n}$ we proceed to establish that the optimal transport bounds in Theorem 1.1 are the best possible estimates for the Krasnosel'skii-Mann iterates. Our results build upon similar ideas in [1, 3], suitably adapted to the more general π^n 's.

Theorem 3.1. *Let \mathcal{S} be the set of all pairs of integers (m, n) with $0 \leq m \leq n$, and consider the unit cube $C = [0, 1]^{\mathcal{S}}$ in the space $(\ell^\infty(\mathcal{S}), \|\cdot\|_\infty)$. Then, for each sequence $(\pi^n)_{n \in \mathbb{N}}$ there exists a non-expansive map $T : C \rightarrow C$ and a corresponding (KM) sequence with $\|x^m - x^n\|_\infty = d_{m,n}$ for all $m, n \in \mathbb{N}$. Moreover, under (H) we also have the tight bound $\|x^n - Tx^n\|_\infty = R_n$.*

Proof. For each $(m, n) \in \mathcal{S}$ consider an optimal solution u^{mn} for $(\mathcal{D}_{m,n})$ and its extension as in Remark 2.2 so that $u_i^{mn} \in [0, 1]$ and (2.3) hold for all $i, j \in \mathbb{N}$.

For every integer $k \in \mathbb{N}$ define $y^k \in C = [0, 1]^{\mathcal{S}}$ as

$$y_{m,n}^k = u_k^{mn} \quad \forall (m, n) \in \mathcal{S}, \quad (3.1)$$

and a corresponding sequence $x^k \in C$ for $k \geq 0$ given by

$$x^k = \sum_{i=0}^k \pi_i^k y^i \quad (3.2)$$

so that in particular $x^0 = y^0 = (u_0^{mn})_{(m,n) \in \mathcal{S}}$.

We claim that $\|y^{m+1} - y^{n+1}\|_\infty \leq d_{m,n} = \|x^m - x^n\|_\infty$ for all $m, n \in \mathbb{N}$. Indeed, from (2.3) we get $|u_{m+1}^{m'n'} - u_{n+1}^{m'n'}| \leq d_{m,n}$ for all $(m', n') \in \mathcal{S}$, which implies

$$\|y^{m+1} - y^{n+1}\|_\infty = \|(u_{m+1}^{m'n'} - u_{n+1}^{m'n'})_{(m',n') \in \mathcal{S}}\|_\infty \leq d_{m,n}.$$

Also, selecting an optimal transport z^{mn} for $(\mathcal{D}_{m,n})$ and proceeding as in the proof of Theorem 1.1 we have

$$x^m - x^n = \sum_{i=0}^m \sum_{j=0}^n z_{i,j}^{mn} (y^i - y^j) \quad (3.3)$$

so that the triangle inequality yields

$$\|x^m - x^n\|_\infty \leq \sum_{i=0}^m \sum_{j=0}^n z_{i,j}^{mn} d_{i-1,j-1} = d_{m,n} \quad (3.4)$$

while considering the (m, n) -coordinate in (3.3), the complementary slackness (2.2) gives

$$\begin{aligned} |x_{m,n}^m - x_{m,n}^n| &= \left| \sum_{i=0}^m \sum_{j=0}^n z_{i,j}^{mn} (y_{m,n}^i - y_{m,n}^j) \right| \\ &= \left| \sum_{i=0}^m \sum_{j=0}^n z_{i,j}^{mn} (u_i^{mn} - u_j^{mn}) \right| \\ &= \sum_{i=0}^m \sum_{j=0}^n z_{i,j}^{mn} d_{i-1,j-1} = d_{m,n} \end{aligned}$$

which combined with (3.4) yields $\|x^m - x^n\|_\infty = d_{m,n}$ as claimed.

Define $T : S \rightarrow C$ on the set $S = \{x^k : k \in \mathbb{N}\} \subseteq C$ by $Tx^k = y^{k+1}$, so that T is non-expansive. Since $\ell^\infty(\mathcal{S})$ as well as the unit cube C are hyperconvex, by Theorem 4 in [19, Aronszajn & Panitchpakdi], T can be extended to a non-expansive map $T : C \rightarrow C$ and then (3.2) is precisely a (KM) sequence which attains all the bounds $\|x^m - x^n\|_\infty = d_{m,n}$ with equality.

It remains to prove that under (H) we have $\|x^n - Tx^n\|_\infty = R_n$. The upper bound follows from (1.1) or by direct computation

$$\|x^n - Tx^n\|_\infty = \left\| \sum_{i=0}^n \pi_i^n (y^i - y^{n+1}) \right\|_\infty \leq \sum_{i=0}^n \pi_i^n d_{i-1,n} = R_n.$$

For the reverse inequality, let $u_{n+1}^{n,n+1}$ be chosen as in Remark 2.3, so that $u_{n+1}^{n,n+1} - u_i^{n,n+1} = d_{i-1,n}$ for each $0 \leq i \leq n$, and therefore

$$\begin{aligned} \|x^n - Tx^n\|_\infty &= \|\sum_{i=0}^n \pi_i^n (y^i - y^{n+1})\|_\infty \\ &\geq |\sum_{i=0}^n \pi_i^n (y_{n,n+1}^i - y_{n,n+1}^{n+1})| \\ &= |\sum_{i=0}^n \pi_i^n (u_i^{n,n+1} - u_{n+1}^{n,n+1})| \\ &= \sum_{i=0}^n \pi_i^n d_{i-1,n} = R_n \end{aligned}$$

which completes the proof. \square

At this point we do not know if R_n is still a sharp bound for the residuals when (H) fails, or whether a tighter bound might hold in that case. On the other hand, we observe that the unit cube $C = [0, 1]^{\mathcal{J}}$ is not only hyperconvex but also bounded so that, from a result by [20, Sine] and [21, Soardi], the map T in Theorem 3.1 has fixed points, that is, $\text{Fix}(T)$ is nonempty. For a survey of the extremely rich structure and results on hyperconvex spaces we refer to [22, Espínola & Khamsi].

4. KRASNOSEL'SKII AND HALPERN ITERATIONS

Let us illustrate the previous tight estimates for the classical Krasnosel'skii and Halpern iterations. For the Krasnosel'skii iteration

$$x^n = (1 - \alpha_n)x^{n-1} + \alpha_n Tx^{n-1}, \quad \alpha_n \in (0, 1)$$

we have $\pi^n = (1 - \alpha_n)\pi^{n-1} + \alpha_n\delta^n$ and (H) holds automatically. In view of Remark 2.3 the unique simple optimal transport from π^n to π^{n+1} gives

$$d_{n,n+1} = \sum_{i=0}^n (\pi_i^n - \pi_i^{n+1}) d_{i-1,n} = \alpha_{n+1} \sum_{i=0}^n \pi_i^n d_{i-1,n} = \alpha_{n+1} R_n$$

which yields a simpler expression for the fixed point residual bound

$$\|x^n - Tx^n\| \leq R_n = \frac{d_{n,n+1}}{\alpha_{n+1}}.$$

This coincides with the tight bound established in [3, Bravo & Cominetti]. Moreover, in view of [11, Cominetti *et al.*] (see also [10, Baillon & Bruck] for the case when $\alpha_k \equiv \alpha$), we have

$$R_n = \frac{d_{n,n+1}}{\alpha_{n+1}} \leq \frac{\text{diam}(C)}{\sqrt{\pi \sum_{k=1}^n \alpha_k (1 - \alpha_k)}} \quad (4.1)$$

which converges to 0 as soon as $\sum_{k=1}^{\infty} \alpha_k (1 - \alpha_k) = \infty$. In particular, for $\alpha_k \equiv \alpha$ we recover the rate $R_n \sim O(1/\sqrt{n})$ proved in [10].

Let us mention that for $\alpha_k \equiv \alpha \approx 1$ the bound (4.1) becomes asymptotically tight and the proportionality constant $1/\sqrt{\pi}$ is the best possible (see [3]). Recall also that when X is uniformly convex with Fréchet differentiable norm, the property $\sum_{k=1}^{\infty} \alpha_k (1 - \alpha_k) = \infty$ implies weak convergence of the iterates x^n to a fixed point, provided $\text{Fix}(T)$ is nonempty (see [23, Reich]).

Let us consider next Halpern's iteration

$$x^n = (1 - \beta_n)y^0 + \beta_n Tx^{n-1}, \quad \beta_n \in (0, 1).$$

In this case we have the simpler structure $\pi_0^n = 1 - \beta_n$ and $\pi_i^n = \beta_n$ so that

$$R_n = \sum_{i=0}^n \pi_i^n d_{i-1,n} = (1 - \beta_n) + \beta_n d_{n-1,n}. \quad (4.2)$$

Here (H) amounts to β_n being non-decreasing, in which case Remark 2.3 yields

$$d_{n-1,n} = (\beta_n - \beta_{n-1}) + \beta_{n-1}d_{n-2,n-1}.$$

This combined with (4.2) implies $R_n = (1 - \beta_n)^2 + \beta_n R_{n-1}$ which has an explicit solution: letting $B_i^n = \prod_{k=i}^n \beta_k$ with $\beta_0 = 0$ and $B_{n+1}^n = 1$, we get

$$R_n = \sum_{i=0}^n (1 - \beta_i)^2 B_{i+1}^n.$$

For the particular case $\beta_n = \frac{n}{n+1}$ a direct calculation gives

$$\|x^n - Tx^n\| \leq R_n = \frac{H_{n+1}}{n+1} \sim O\left(\frac{\ln n}{n}\right) \quad (4.3)$$

with $H_n = \sum_{k=1}^n \frac{1}{k}$ the n -th harmonic number. From Theorem 3.1 we know that this bound is sharp and cannot be improved. However, [24, Lieder] recently proved that when $(X, |\cdot|)$ is a Hilbert space and $T : X \rightarrow X$ has a fixed point $x^* = Tx^*$, the choice $\beta_n = \frac{n}{n+1}$ yields the faster rate

$$|x^n - Tx^n| \leq \frac{2|x^0 - x^*|}{n+1} \sim O\left(\frac{1}{n}\right).$$

In order to compare with (4.3), consider the rescaled norm $\|\cdot\| = \frac{1}{r_0}|\cdot|$ with $r_0 = 2|x^0 - x^*|$ so that $C = B(x^*, |x^0 - x^*|)$ has diameter 1. Then Lieder's bound reads $\|x^n - Tx^n\| \leq \frac{1}{n+1}$ which is clearly smaller than (4.3). This does not contradict Theorem 3.1: it simply shows that the tight bounds are attained in a non-Hilbert setting such as the space $\ell^\infty(\mathcal{I})$.

On the other hand, a minor modification of Halpern with $\beta_n = \frac{n}{n+2}$ yields the faster rate $R_n = \frac{4}{n+1}(1 - \frac{H_{n+2}}{n+2}) \sim O(\frac{1}{n})$ in general normed spaces. This bound is also tight and slightly improves the estimate $\|x^n - Tx^n\| \leq \frac{4}{n+1}$ established in [25, Sabach & Shtern, Lemma 5]. For the record, we observe that although $\beta_n = \frac{n}{n+2}$ achieves a faster rate than $\beta_n = \frac{n}{n+1}$, the slow growth of H_n implies that for small n both sharp bounds coincide within a moderate multiplicative factor (less than 2.5 up to $n = 10.000$).

In general it remains open to find conditions under which $R_n \rightarrow 0$. The forthcoming paper [26, Contreras & Cominetti] further investigates various algorithms that fit the general framework (KM), and determines the optimal choices for the weights π^n that guarantee the faster rate $R_n \sim O(1/n)$ with the best proportionality constants.

5. MONOTONICITY, CONVEX QUADRANGLE INEQUALITY, AND THE INSIDE-OUT GREEDY ALGORITHM

In this section we discuss some further properties of the optimal transport metric under the additional assumption (H). A first consequence is that the distances $d_{m,n}$ increase as m and n get farther apart: for any fixed m the map $n \mapsto d_{m,n}$ decreases from a value of 1 at $n = -1$ down to 0 at $n = m$, after which it increases for $n \geq m$. Intuitively, this comes from the fact that the distribution π^n drifts gradually towards larger integers, so that for $n \geq m$ it is more expensive to transport π^m towards π^{n+1} than it is to satisfy the demands π^n which are located somewhat closer to π^m . A similar argument can be made for $n < m$. We formally prove this monotonicity by exploiting both the primal and dual characterizations of $d_{m,n}$.

Theorem 5.1. *Assume (H). Then in the region $m \leq n$ we have that $d_{m,n}$ decreases with m and increases with n .*

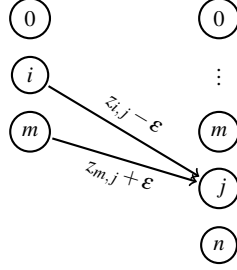


FIGURE 6. Redistribution of flows to prove monotonicity.

Proof. Let us first show that $d_{m,n} \leq d_{m-1,n}$ for $0 \leq m \leq n$. Since for $n = 0$ this holds trivially, we proceed by induction assuming that these inequalities hold up to $n - 1$. To establish the property for n we make a second induction on m . For $m = 0$ we clearly have $d_{0,n} \leq d_{-1,n} = 1$. Now, take z a simple optimal transport for $d_{m-1,n}$ so that for $i = 0, \dots, m - 1$ we have $z_{i,i} = \pi_i^n$ with the excess supply $\pi_i^{m-1} - \pi_i^n$ shipped to nodes $j \geq m$. We transform z into a feasible transport for $d_{m,n}$ while reducing the cost. To this end, for each $i = 0, \dots, m - 1$ with $\pi_i^m < \pi_i^{m-1}$ we take some positive flow $z_{i,j} > 0$ with $m \leq j \leq n$, from which we remove $\epsilon > 0$ and increase $z_{m,j}$ by the same amount (see Figure 6). Since $i - 1 < m - 1 \leq j - 1 \leq n - 1$, the induction hypothesis implies that the cost decreases by $[d_{m-1,j-1} - d_{i-1,j-1}]\epsilon \leq 0$. We repeat these flow transfers until the outflow $\sum_{j=0}^n z_{i,j}$ is reduced to π_i^m . At this point a fraction $\pi_i^{m-1} - \pi_i^m$ of the excess supply of each $i \leq m - 1$ has been fully transferred to m and therefore

$$\sum_{j=0}^n z_{m,j} = \sum_{i=0}^{m-1} (\pi_i^{m-1} - \pi_i^m) = 1 - \sum_{i=0}^{m-1} \pi_i^m = \pi_m^m.$$

Hence, the final flow is feasible for $d_{m,n}$ and since the cost was decreased at every step we conclude $d_{m,n} \leq C_{m,n}(z) \leq d_{m-1,n}$.

We now use the dual characterization $(\mathcal{D}_{m,n})$ to prove $d_{m,n} \leq d_{m,n+1}$ by induction on n . This holds trivially for $n = m$. Assume that it holds up to $n - 1$ and let us prove it for n . Take u^{mn} optimal for $(\mathcal{D}_{m,n})$ and consider the vector u with $u_i = u_i^{mn}$ for $i = 0, \dots, n$ and $u_{n+1} = u_{j_0}^{mn}$ where $u_{j_0}^{mn}$ is the maximal coordinate in u^{mn} . By construction $|u_i - u_j| \leq d_{i-1,j-1}$ for all $0 \leq i, j \leq n$ while the induction hypothesis gives

$$|u_i - u_{n+1}| = |u_i^{mn} - u_{j_0}^{mn}| \leq d_{i-1,j_0-1} \leq d_{i-1,n}$$

so that u is feasible in $(\mathcal{D}_{m,n+1})$. On the other hand, $\delta_j^{m,n} = \delta_j^{m,n+1} + \epsilon_j$ with $\epsilon_j = \pi_j^n - \pi_j^{n+1} \geq 0$ for $j = 0, \dots, n$ from which we get

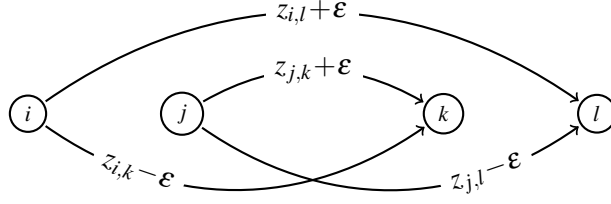
$$d_{m,n} = \sum_{j=0}^n \delta_j^{m,n} u_j = \sum_{j=0}^n \delta_j^{m,n+1} u_j + \sum_{j=0}^n \epsilon_j u_j.$$

Since $u_j \leq u_{n+1}$ this last sum can be majorized by $\sum_{j=0}^n \epsilon_j u_{n+1}$, which combined with the equality $\sum_{j=0}^n \epsilon_j = \pi_{n+1}^{n+1} = \delta_{n+1}^{m,n+1}$ implies

$$d_{m,n} \leq \sum_{j=0}^{n+1} \delta_j^{m,n+1} u_j \leq d_{m,n+1}.$$

completing the induction step. \square

REMARK 4. The informal argument just before Theorem 5.1 might suggest that monotonicity could hold under the weaker condition of first order stochastic dominance $\pi^n \preceq_{st} \pi^{n+1}$, that is

FIGURE 7. Recirculation of ϵ along the cycle $i \rightarrow l \rightarrow j \rightarrow k \rightarrow i$.

$\sum_{i \geq k} \pi_i^n \leq \sum_{i \geq k} \pi_i^{n+1}$ for all $k \in \mathbb{N}$. It turns out that this is not always the case as shown by

$$\begin{aligned} \pi^0 &= (1, 0, 0, 0, \dots) \\ \pi^1 &= \left(\frac{4}{5}, \frac{1}{5}, 0, 0, \dots\right) \\ \pi^2 &= \left(\frac{1}{2}, \frac{1}{4}, \frac{1}{4}, 0, 0, \dots\right) \\ \pi^3 &= \left(\frac{1}{2}, \frac{1}{4}, 0, \frac{1}{4}, 0, 0, \dots\right) \\ \pi^4 &= \left(0, \frac{1}{2}, \frac{1}{4}, 0, \frac{1}{4}, 0, 0, \dots\right) \\ \pi^5 &= \left(0, 0, \frac{1}{2}, 0, 0, \frac{1}{2}, 0, 0, \dots\right) \end{aligned}$$

for which we have $d_{3,5} = \frac{22}{40}$ and $d_{4,5} = \frac{23}{40}$ so that $d_{3,5} < d_{4,5}$.

We next show that under (H) the distances $d_{m,n}$ satisfy the convex quadrangle inequality (Q). The proof proceeds inductively, by considering the quadrangle inequality up to n , namely

$$(Q_n) \quad d_{i,l} + d_{j,k} \leq d_{i,k} + d_{j,l} \quad \text{for all } -1 \leq i \leq j \leq k \leq l \leq n.$$

A key consequence of the latter inequalities is the existence of optimal transports that are not only simple but also nested in the sense that the flows do not intersect.

Definition 5.1. A transport plan z is called *nested* if there are no $i < j < k < l$ with $z_{i,k} > 0$ and $z_{j,l} > 0$.

Proposition 5.1. Assume (Q_{n-1}) . Then each $d_{m,n}$ for $m \leq n$ admits an optimal transport that is both simple and nested.

Proof. Let z be a simple optimal transport for $d_{m,n}$ and suppose that $z_{i,k} > 0$ and $z_{j,l} > 0$ for some $0 \leq i < j < k < l \leq n$. Removing $\epsilon = \min\{z_{i,k}, z_{j,l}\}$ from $z_{i,k}$ and $z_{j,l}$ while increasing $z_{i,l}$ and $z_{j,k}$ by the same amount (see Figure 7), the modified transport remains feasible and (Q_{n-1}) implies a cost reduction of

$$[d_{i-1,l-1} + d_{j-1,k-1} - d_{i-1,k-1} - d_{j-1,l-1}] \epsilon \leq 0.$$

Hence, the modified transport remains optimal with $z_{i,k} = 0$ or $z_{j,l} = 0$. Inductively we can eliminate all flow intersections and obtain an optimal transport that is simple and nested. \square

Under (H) the properties of being simple and nested determine a *unique transport plan* through the following greedy procedure. Imagine the supply nodes as a set of buckets S_0, \dots, S_m initially filled with volumes π_i^m for $0 \leq i \leq m$, and a set of empty demand buckets D_0, \dots, D_n with capacities π_j^n for $0 \leq j \leq n$. Considering the supplies S_m, S_{m-1}, \dots, S_0 in reverse order, we use the volume π_i^m in each S_i to fill D_i and distribute the surplus $\sigma_i = \pi_i^m - \pi_i^n$ among the unmet

demands D_k for $k > m$ in increasing order. Thus, S_m fills $D_m, D_{m+1}, \dots, D_{k_m}$ where D_{k_m} is the bucket at which the volume π_m^m is depleted. Next S_{m-1} fills D_{m-1} and resumes at the bucket D_{k_m} partially filled in the previous round and up to some $D_{k_{m-1}}$ at which π_{m-1}^m is depleted, and so on.

More formally, the procedure is as follows:

Algorithm 1 Inside-Out — input π^m, π^n with $m \leq n$

- 1: For each $i = 0, \dots, m$ set $z_{i,i} = \pi_i^n$ and $z_{j,i} = 0$ for $j \neq i$, and compute the residual supply $\sigma_i = \pi_i^m - \pi_i^n$.
 - 2: Use σ_m to fill the demands $\pi_{m+1}^n, \pi_{m+2}^n, \dots$ up to the node k_m at which σ_m is depleted with $\pi_{k_m}^n$ only partially filled. This sets the flows $z_{m,j} = \pi_j^n$ for $j \in \{m+1, \dots, k_m-1\}$ and $z_{m,k_m} = \sigma_m - \sum_{j=m+1}^{k_m-1} \pi_j^n$.
 - 3: Use σ_{m-1} to fill any unmet demand at k_m as well as the subsequent demands up to a node k_{m-1} at which σ_{m-1} is depleted. This determines the nonzero flows $z_{m-1,j}$ for $j \in \{k_m, \dots, k_{m-1}\}$.
 - 4: Proceed sequentially using σ_i for $i = m-2, \dots, 0$ to fulfill the unmet demands at the nodes $j \in \{k_{i+1}, \dots, k_i\}$, which determine the corresponding flows $z_{i,j}$.
-

As it turns out, the validity of the quadrangle inequality and that of the inside-out algorithm are closely related, and will be proved simultaneously. For the special case where $\pi_i^n = \alpha_i \prod_{k=i+1}^n (1 - \alpha_k)$, the optimality of this algorithm was conjectured in [1], and later confirmed in [2] with a long and technical proof. Below we present a simpler argument that exploits the optimal transport structure of the $d_{m,n}$'s, and which applies to the more general distributions π^n 's considered here. The proof uses Proposition 5.1 inductively to show that (Q_n) — or more precisely, its equivalent form (\tilde{Q}_n) below — holds for all $n \in \mathbb{N}$.

Lemma 5.1. *Let $\Delta_{m,j} \triangleq d_{m,j+1} - d_{m,j}$. Then (Q_n) is equivalent to*

$$(\tilde{Q}_n) \quad \text{For all } m < j < n \text{ we have } \Delta_{m,j} \leq \Delta_{m+1,j}.$$

Proof. The inequality in (Q_n) can be written as $d_{i,l} - d_{i,k} \leq d_{j,l} - d_{j,k}$ which amounts to the fact that $d_{m,l} - d_{m,k}$ increases with m (for $m \leq k \leq l \leq n$). Since $d_{m,l} - d_{m,k} = \sum_{j=k}^{l-1} \Delta_{m,j}$ this is in turn equivalent to (\tilde{Q}_n) . \square

Theorem 5.2. *Assume (H). Then, the recursive optimal transport metric $d_{m,n}$ satisfies the convex quadrangle inequality (Q) , and therefore the inside-out algorithm computes an optimal transport for all $m \leq n$.*

Proof. Since (\tilde{Q}_0) holds trivially, it suffices to prove that $(\tilde{Q}_n) \Rightarrow (\tilde{Q}_{n+1})$. In view of the induction hypothesis (\tilde{Q}_n) , in order to prove (\tilde{Q}_{n+1}) it suffices to consider $j = n$ and to show that $\Delta_{m,n} \leq \Delta_{m+1,n}$ for all $m < n$.

Using (\tilde{Q}_n) , and according to Proposition 5.1 and Lemma 5.1, it follows that the four problems $d_{m+1,n+1}$, $d_{m+1,n}$, $d_{m,n+1}$, $d_{m,n}$, admit simple and nested optimal transports given by the inside-out algorithm. Let z and z' be such optimal transports for $d_{m+1,n+1}$ and $d_{m+1,n}$ respectively. We will modify z and z' simultaneously, in such a way that $\Delta_{m+1,n}$ decreases and the transports are progressively transformed into simple and nested transports for $d_{m,n+1}$ and $d_{m,n}$. By Proposition 5.1 these modified transports are optimal for $d_{m,n+1}$ and $d_{m,n}$, which yields the conclusion $\Delta_{m,n} \leq \Delta_{m+1,n}$.

The modification of z and z' is done in a sequence of stages. The informal idea is as follows. Since $\pi_{m+1}^{m+1} > \pi_{m+1}^m = 0$, the initial transport z has a surplus of flows $z_{m+1,j}$ shipped out from

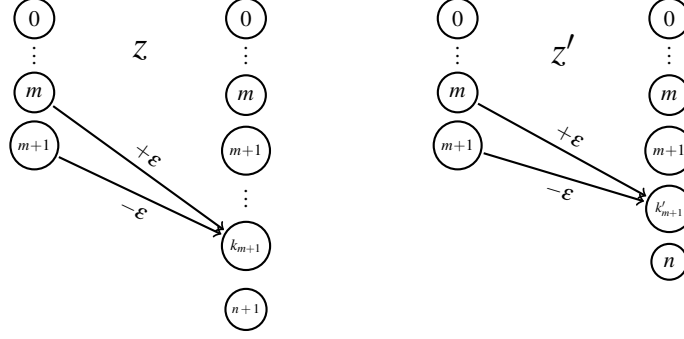


FIGURE 8. Phase 1 of redistribution of flows.

$m + 1$, so that to regain feasibility for $d_{m,n+1}$ these must be substituted with flows $z_{i,j}$ from sources $i \leq m$. The same is needed to transform z' into a feasible transport for $d_{m,n}$. We do this in a series of flow shifts that progressively eliminate the outflow surplus at the tail supply nodes, pushing the imbalance towards sources closer to 0 and iterating until all nodes dispatch the correct supplies π_i^m . Along the process we make sure that the transports remain simple and nested, while the quantity $\Delta_{m+1,n}$ is reduced at each step.

In order to formalize this idea, let $V_i = \sum_{k=i}^{m+1} \pi_k^{m+1}$ be the cumulative supply from i to $m + 1$. Then the node $k_i > m + 1$ at which V_i is depleted in the inside-out procedure for $d_{m+1,n+1}$ is characterized by the inequalities

$$\sum_{j=i}^{k_i-1} \pi_j^{n+1} < V_i \leq \sum_{j=i}^{k_i} \pi_j^{n+1}.$$

Similarly, the corresponding node $k'_i > m + 1$ for $d_{m+1,n}$ is such that

$$\sum_{j=i}^{k'_i-1} \pi_j^n < V_i \leq \sum_{j=i}^{k'_i} \pi_j^n.$$

From this it follows that $k'_i \leq k_i$. Indeed, this inequality is trivial if $k_i = n + 1$, whereas if $k_i \leq n$ we may use the inequalities $\pi_j^{n+1} \leq \pi_j^n$ to obtain

$$\sum_{j=i}^{k'_i-1} \pi_j^n < V_i \leq \sum_{j=i}^{k_i} \pi_j^{n+1} \leq \sum_{j=i}^{k_i} \pi_j^n \quad (5.1)$$

from where we deduce that $k'_i - 1 < k_i$, and therefore $k'_i \leq k_i$ as claimed.

In a first phase of the transformation of z we consider the supply π_{m+1}^{m+1} that is shipped from $m + 1$ towards the demand nodes $m + 1, \dots, k_{m+1}$. We transfer a small amount ε from $m + 1$ to m (see Figure 8), which implies a change in the cost of $[d_{m-1,k_{m+1}-1} - d_{m,k_{m+1}-1}]\varepsilon$. An analog flow transfer in z' induces a corresponding cost change of $[d_{m-1,k'_{m+1}-1} - d_{m,k'_{m+1}-1}]\varepsilon$. Both changes combined imply that $\Delta_{m+1,n}$ is decreased by

$$[d_{m-1,k_{m+1}-1} - d_{m,k_{m+1}-1} - d_{m-1,k'_{m+1}-1} + d_{m,k'_{m+1}-1}]\varepsilon \leq 0$$

where negativity follows from (\tilde{Q}_n) since $m - 1 \leq m \leq k'_{m+1} - 1 \leq k_{m+1} - 1$. Let $z_\varepsilon, z'_\varepsilon$ and $\Delta_{m+1,n}^\varepsilon = C_{m+1,n+1}(z_\varepsilon) - C_{m+1,n}(z'_\varepsilon)$ denote the modified transports and value.

As we proceed with these flow transfers the residual supply $\pi_{m+1}^{m+1} - \varepsilon$ of node $m+1$ decreases and as a consequence the nodes k_{m+1} and k'_{m+1} at which this residual supply is depleted will eventually decrease (or move upwards in Figure 8). However, the same argument as in (5.1) with V_{m+1} replaced by $V_{m+1} - \varepsilon$, implies that they remain ordered as $k'_{m+1} \leq k_{m+1}$. Hence, we can continue the flow transfers until the full supply π_{m+1}^{m+1} is moved from $m+1$ to m . This completes the first phase, along which $\Delta_{m+1,n}^\varepsilon$ decreases.

At this point, the supply at m has been increased to $V_m = \pi_m^{m+1} + \pi_{m+1}^{m+1}$, and is being shipped to $j = m, \dots, k_m$ in z_ε and to $j = m, \dots, k'_m$ in z'_ε . We further observe that

$$\pi_m^{m+1} + \pi_{m+1}^{m+1} = 1 - \sum_{i=0}^{m-1} \pi_i^{m+1} \geq 1 - \sum_{i=0}^{m-1} \pi_i^m = \pi_m^m.$$

We then start the second phase in which we reduce the supply $\pi_m^{m+1} + \pi_{m+1}^{m+1}$ down to π_m^m by transferring some flow ε from node m to $m-1$. As we proceed with these flow transfers once again the nodes k_m and k'_m will decrease, but remain in the order $k'_m \leq k_m$ as it follows from (5.1) with V_m replaced by $V_m - \varepsilon$. As in the first phase, these flow transfers further reduce $\Delta_{m+1,n}^\varepsilon$. The second phase stops as soon as the supply $\pi_m^{m+1} + \pi_{m+1}^{m+1}$ has been reduced to π_m^m .

At this point, the supply of $m-1$ is $\pi_{m-1}^{m+1} + \pi_m^{m+1} + \pi_{m+1}^{m+1} - \pi_m^m$, which is larger than π_{m-1}^m as it follows from

$$\pi_{m-1}^{m+1} + \pi_m^{m+1} + \pi_{m+1}^{m+1} = 1 - \sum_{i=0}^{m-2} \pi_j^{m+1} \geq 1 - \sum_{i=0}^{m-2} \pi_j^m = \pi_{m-1}^m + \pi_m^m.$$

We note that the sum of the modified supplies of $m-1$ and m continues to be equal to V_{m-1} and is shipped to $m-1, \dots, k_{m-1}$ in z_ε and to $m-1, \dots, k'_{m-1}$ in z'_ε . We may then proceed as before by moving a flow ε from $m-1$ to $m-2$, until the supply of $m-1$ is decreased to π_{m-1}^m .

We continue in this manner with $i = m-2, m-3, \dots, 0$ until the transport plans z_ε and z'_ε are shipping the correct supplies π_i^m from each source node $i = 0, \dots, m$. This is possible since at the end of each phase the supply of node i is equal to $\sum_{j=i}^{m+1} \pi_j^{m+1} - \sum_{j=i+1}^m \pi_j^m$ which is larger than π_i^m as it follows as before from the inequality

$$\sum_{j=i}^{m+1} \pi_j^{m+1} = 1 - \sum_{j=0}^{i-1} \pi_j^{m+1} \geq 1 - \sum_{j=0}^{i-1} \pi_j^m = \sum_{j=i}^m \pi_j^m.$$

The final transport plans z_ε and z'_ε are therefore feasible for $d_{m,n+1}$ and $d_{m,n}$ respectively. Since they are also simple and nested, it follows from the induction hypothesis (\tilde{Q}_n) and Proposition 5.1 that these transformed flows are optimal, so the final value of $\Delta_{m+1,n}^\varepsilon$ coincides with $\Delta_{m,n}$ and therefore $\Delta_{m,n} = \Delta_{m+1,n}^\varepsilon \leq \Delta_{m+1,n}$, completing the induction step. \square

REMARK 5. The example in Remark 4 also shows that the convex quadrangle inequality does not hold in general, not even under stochastic dominance. In that example (Q) fails for several combinations $i < j < k < l$, for instance with $d_{1,5} + d_{2,4} = \frac{139}{100} > \frac{135}{100} = d_{1,4} + d_{2,5}$.

6. SOME FINAL EXAMPLES AND COMMENTS

This section presents some interesting examples that expose the (sometimes) surprisingly convoluted structure of the bounds $d_{m,n}$.

Example 6.1. The case where each $\pi^n = \delta^n$ is a Dirac mass at $n \in \mathbb{N}$, induces the trivial distance $d_{m,n} = \mathbb{1}_{m \neq n}$ in $\overline{\mathbb{N}}$.

Example 6.2. The next simplest case is probably when $\pi^n = (1 - \alpha)\delta^{n-1} + \alpha\delta^n$ with $\alpha \geq \frac{1}{2}$. Since $\pi_{n-1}^n = 1 - \alpha \leq \alpha = \pi_n^n$, condition (H) is satisfied and we may run the Inside-Out algorithm from §5 to obtain $d_{m,n}(\alpha)$ as polynomials in α . Figure 9 shows these polynomials for $m = 7$. An interesting feature is that in the range $\lfloor m/2 \rfloor \leq n < 2(m+1)$ with $m \neq n$ the polynomial $d_{m,n}(\alpha)$ has degree $\min\{m, n\} + 1$ with some peculiar integer coefficients, while outside this range $d_{m,n} \equiv 1$. Naturally, for $\alpha = 1$ we recover $d_{m,n} = \mathbb{1}_{m \neq n}$ as in the simple Dirac case of Example 6.1. However, it is amazing how a minor tweak in the π^n 's induces a drastic change in the structure of the recursive bounds $d_{m,n}$.

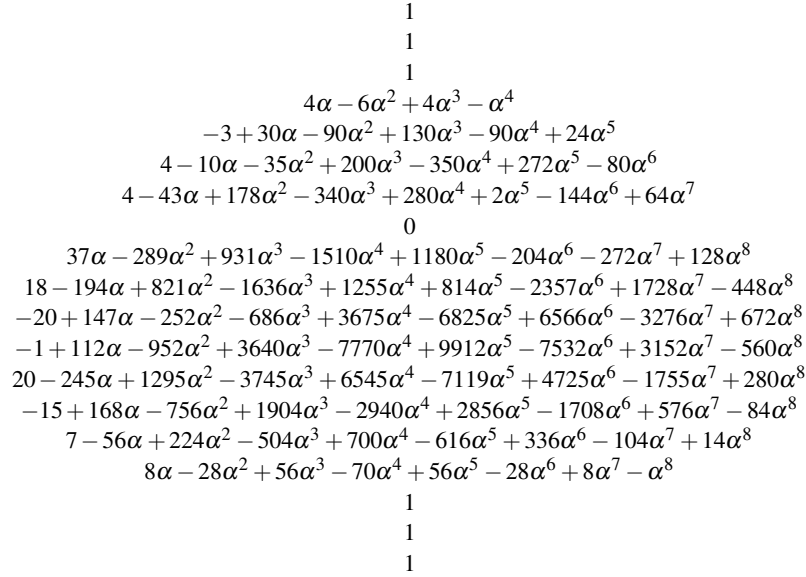


FIGURE 9. $\{d_{7,n}(\alpha)\}_{0 \leq n \leq 18}$ for $\pi^n = (1 - \alpha)\delta^{n-1} + \alpha\delta^n$, $\alpha \geq \frac{1}{2}$.

Example 6.3. Let us consider next the classical Krasnosel'skii iteration with constant $\alpha_n \equiv \alpha$ and $\pi^n = (1 - \alpha)\pi^{n-1} + \alpha\delta^n$. Since (H) holds automatically, Theorem 3.1 and §4 imply the tight estimate $\|x^n - Tx^n\| \leq R_n = d_{n,n+1}/\alpha$, which can be computed by the Inside-Out algorithm. For $\alpha \geq \frac{1}{2}$ the algorithm takes a particularly simple form, and yields a polynomial expression for R_n . Figure 10 shows some of these polynomials as a function of $\alpha \geq \frac{1}{2}$. Again, we know very little about these polynomials $R_n(\alpha)$, except that the first two leading terms can be proved to be $1 - n\alpha$, while the quadratic and cubic terms seem to be $n^2\alpha^2$ and $(n^2 - n^3)\alpha^3$. On the other hand, by considering a family of sub-optimal transport plans in $(\mathcal{P}_{m,n})$, [11] established a non-trivial bound expressed in terms of Gauss' hypergeometric function

$$R_n(\alpha) \leq {}_2F_1\left(-n, \frac{1}{2}; 2; 4\alpha(1 - \alpha)\right) \quad \text{for all } n \geq 1.$$

Interestingly, the following observation hints at a possible underlying combinatorial structure of the polynomials R_n . The sequence of degrees of R_n

$$\{2, 4, 7, 10, 14, 18, 23, 28, 34, 40, 47, 54, 62, 70, 79, 88, 98, 108, 119, 130, \dots\}$$

$$\begin{aligned}
R_1 &= 1 - \alpha + \alpha^2 \\
R_2 &= 1 - 2\alpha + 4\alpha^2 - 4\alpha^3 + 2\alpha^4 \\
R_3 &= 1 - 3\alpha + 9\alpha^2 - 18\alpha^3 + 25\alpha^4 - 21\alpha^5 + 9\alpha^6 - \alpha^7 \\
R_4 &= 1 - 4\alpha + 16\alpha^2 - 48\alpha^3 + 112\alpha^4 - 192\alpha^5 + 230\alpha^6 - 180\alpha^7 \\
&\quad + 84\alpha^8 - 20\alpha^9 + 2\alpha^{10} \\
R_5 &= 1 - 5\alpha + 25\alpha^2 - 100\alpha^3 + 331\alpha^4 - 876\alpha^5 + 1795\alpha^6 - 2762\alpha^7 \\
&\quad + 3106\alpha^8 - 2482\alpha^9 + 1366\alpha^{10} - 500\alpha^{11} + 117\alpha^{12} - 16\alpha^{13} + \alpha^{14} \\
R_6 &= 1 - 6\alpha + 36\alpha^2 - 180\alpha^3 + 775\alpha^4 - 2806\alpha^5 + 8324\alpha^6 - 19778\alpha^7 \\
&\quad + 37023\alpha^8 - 53948\alpha^9 + 60623\alpha^{10} - 52122\alpha^{11} + 34044\alpha^{12} - 16770\alpha^{13} \\
&\quad + 6163\alpha^{14} - 1652\alpha^{15} + 308\alpha^{16} - 36\alpha^{17} + 2\alpha^{18} \\
R_7 &= 1 - 7\alpha + 49\alpha^2 - 294\alpha^3 + 1562\alpha^4 - 7222\alpha^5 + 28408\alpha^6 - 93187\alpha^7 \\
&\quad + 251365\alpha^8 - 552678\alpha^9 + 985643\alpha^{10} - 1422448\alpha^{11} + 1660135\alpha^{12} \\
&\quad - 1567511\alpha^{13} + 1198337\alpha^{14} - 741914\alpha^{15} + 371352\alpha^{16} - 149443\alpha^{17} \\
&\quad + 47802\alpha^{18} - 11909\alpha^{19} + 2233\alpha^{20} - 297\alpha^{21} + 25\alpha^{22} - \alpha^{23}
\end{aligned}$$

FIGURE 10. $R_n(\alpha)$ for Krasnosel'skii iteration with $\alpha_n \equiv \alpha \geq \frac{1}{2}$.

coincides with $\lfloor (n^2 + 6n + 1)/4 \rfloor$ at least up to $n = 40$. This matches the sequence A014616 in the *Online Encyclopedia of Integer Sequences* (oeis.org), which is the solution of the *postage stamp problem* with n stamps and $k = 2$ denominations. The postage stamp problem is as follows: an envelope can carry at most n stamps with values chosen from k possible denominations $\{a_1, \dots, a_k\}$. The problem is to find denominations that allow to make up all consecutive postage values $p = 1, 2, \dots, N$ with N as large as possible. For instance, when $n = 3$ and $k = 2$ the solution is $N = 7$ and is attained with $a_1 = 1, a_2 = 3$ as follows

$$1, 1 + 1, 1 + 1 + 1, 1 + 3, 1 + 1 + 3, 3 + 3, 1 + 3 + 3.$$

Example 6.4. Another interesting common case is when π^n is chosen as a uniform distribution over $\{0, \dots, n\}$ with $\pi_i^n = \frac{1}{n+1}$ for all $0 \leq i \leq n$. This is again a classical Krasnosel'skii iteration since $\pi_i^n = (1 - \alpha_n)\pi_i^{n-1} + \alpha_n \delta_i^n$ with $\alpha_n = \frac{1}{n+1}$. Explicitly, in this case the iteration (KM) reads

$$x^n = \frac{n}{n+1}x^{n-1} + \frac{1}{n+1}Tx^{n-1}, \quad n \geq 1.$$

The matrix below presents the exact values of $d_{m,n}$ for $0 \leq m, n \leq 6$.

$$d_{m,n} = \begin{pmatrix} 0 & \frac{1}{2} & \frac{2}{3} & \frac{3}{4} & \frac{4}{5} & \frac{5}{6} & \frac{6}{7} \\ \frac{1}{2} & 0 & \frac{1}{4} & \frac{3}{8} & \frac{7}{15} & \frac{19}{36} & \frac{97}{168} \\ \frac{2}{3} & \frac{1}{4} & 0 & \frac{23}{144} & \frac{47}{180} & \frac{1}{3} & \frac{47}{120} \\ \frac{3}{4} & \frac{3}{8} & \frac{23}{144} & 0 & \frac{329}{2880} & \frac{1681}{8640} & \frac{1733}{6720} \\ \frac{4}{5} & \frac{7}{15} & \frac{47}{180} & \frac{329}{2880} & 0 & \frac{7609}{86400} & \frac{7793}{50400} \\ \frac{5}{6} & \frac{19}{36} & \frac{1}{3} & \frac{1681}{8640} & \frac{7609}{86400} & 0 & \frac{257219}{3628800} \\ \frac{6}{7} & \frac{97}{168} & \frac{47}{120} & \frac{1733}{6720} & \frac{7793}{50400} & \frac{257219}{3628800} & 0 \end{pmatrix}$$

Regarding the bounds $R_n = d_{n,n+1}/\alpha_{n+1}$, from (4.1) we readily obtain

$$R_n = \frac{d_{n,n+1}}{\alpha_{n+1}} \leq \frac{\text{diam}(C)}{\sqrt{\pi \sum_{k=1}^n \alpha_k (1-\alpha_k)}} \sim O\left(\frac{1}{\sqrt{\ln n}}\right).$$

Now, as mentioned earlier, the tightness of the inequality (4.1) proved in [3] occurs with $\alpha_n \equiv \alpha \approx 1$, and since here $\alpha_n = \frac{1}{n+1} \rightarrow 0$ one could expect a faster rate for R_n . However, numerical calculations suggest that this is not the case and the bound $O\left(1/\sqrt{\ln n}\right)$ seems to capture accurately the asymptotic behavior of R_n . Figure 11 displays $\varphi_n = R_n \sqrt{\ln(n+1)}$ for $1 \leq n \leq 500$. For the sake of comparison we also include $\phi_n = R_n \ln(n+1)$.

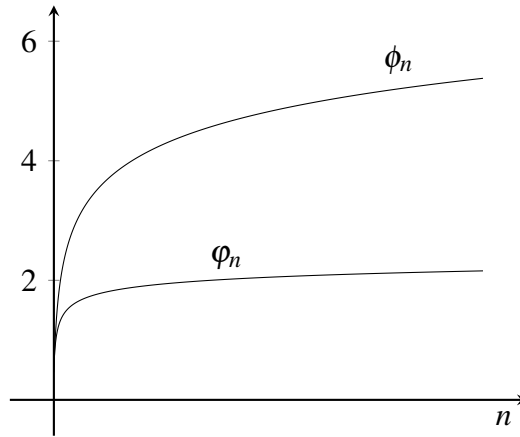


FIGURE 11. The quantities ϕ_n and φ_n , $1 \leq n \leq 500$.

Acknowledgments

The work of M. Bravo was partially funded by FONDECYT 1191924. R. Cominetti gratefully acknowledges the support provided by FONDECYT 1171501.

REFERENCES

- [1] J.B. Baillon, R.E. Bruck, *Optimal rates of asymptotic regularity for averaged non-expansive mappings*, In: Proceedings of the Second International Conference on Fixed Point Theory and Applications (K.K. Tan, ed.), World Scientific Press, London (1992), 27-66.
- [2] Y. Aygen-Satik, *Optimal Bounds of Asymptotic Regularity*, Ph.D. Thesis, University of Southern California, (1994).
- [3] M. Bravo, R. Cominetti, *Sharp convergence rates for averaged non-expansive maps*, Israel Journal of Mathematics 17(1) (2018), 163-188.
- [4] W.R. Mann, *Mean value methods in iteration*, Proc. Amer. Math. Soc. 4(3) (1953), 506-510.
- [5] M.A. Krasnosel'skii, *Two remarks on the method of successive approximations*, Uspekhi Mat. Nauk 10:1(63) (1955), 123-127.
- [6] R.E. Bruck, *Asymptotic behavior of nonexpansive mapping*, Contemporary Mathematics 18 (1983), 1-47.
- [7] B. Halpern, *Fixed points of nonexpanding maps*, Bull. Amer. Math. Soc. 73(6) (1967), 957-961.
- [8] G. López, V. Martín-Márquez, X. Hong-Kun, *Halpern's iteration for nonexpansive mappings*, Contemporary Mathematics 513 (2010), 211-231.
- [9] S. Ishikawa, *Fixed points by a new iteration method*. Proc. Amer. Math. Soc. 44 (1974), 147-150.

- [10] J.B. Baillon, R.E. Bruck, *The rate of asymptotic regularity is $O(1/\sqrt{n})$* , Lecture Notes in Pure and Applied Mathematics 178 (1996), 51-81.
- [11] R. Cominetti, J. Soto, J. Vaisman, *On the rate of convergence of Krasnosel'skiĭ-Mann iterations and their connection with sums of Bernoullis*, Israel J. Math. 199(2) (2014), 757-772.
- [12] R.E. Burkard, B. Klinz, R. Rudolf, *Perspectives of Monge properties in optimization*, Discrete Applied Mathematics 70 (1996), 95-161.
- [13] A. Aggarwal, A. Bar-Noy, S. Khuller, D. Kravets, B. Schieber, *Efficient minimum cost matching and transportation using the quadrangle inequality*. J. Algorithms 19(1) (1992), 116-143.
- [14] W.W. Bein, *Advanced Techniques for Dynamic Programming*. Handbook of Combinatorial Optimization, (2013), 41-92.
- [15] W.W. Bein, M.J. Golin, L.L. Larmore, Y. Zhang, *The Knuth-Yao Quadrangle-Inequality Speedup is a Consequence of Total-Monotonicity*. Proc. 17th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2006, ACM Transactions on Algorithms 6(1) (2009), 31-40.
- [16] D.E. Knuth, *Optimum binary search trees*. Acta Informatica 1 (1971), 14-25.
- [17] F.F. Yao, *Speed-up in dynamic programming*, SIAM J. Algebraic Discrete Methods 3 (1982), 532-540.
- [18] C. Villani, *Topics in optimal transportation*, Graduate Studies in Mathematics 58, American Mathematical Society (2003).
- [19] N. Aronszajn, P. Panitchpakdi, *Extension of uniformly continuous transformations and hyperconvex metric spaces*. Pacific J. Math. 6(3) (1956), 405-439.
- [20] R.C. Sine, *On linear contraction semigroups in sup norm spaces*, Nonlinear Anal. 3 (1979), 885-890.
- [21] P. Soardi, *Existence of fixed points for nonexpansive mappings in certain Banach lattices*, Proc. Amer. Math. Soc. 73 (1979), 25-29.
- [22] R. Espínola, M.A. Khamsi, *Introduction to Hyperconvex Spaces*. In: Kirk W.A., Sims B. (eds) Handbook of Metric Fixed Point Theory. Springer, Dordrecht, (2001).
- [23] S. Reich, *Weak convergence theorems for non-expansive mappings in Banach spaces*, J. Math. Anal. Appl. 67 (1979), 274-276.
- [24] F. Lieder, *On the convergence rate of the Halpern-iteration*, Optimization Letters 15 (2021), 405-418.
- [25] S. Sabach, S. Shtern, *A first order method for solving convex bilevel optimization problems*, SIAM Journal on Optimization 27(2) (2017), 640-660.
- [26] J.P. Contreras, R. Cominetti, *Optimal error bounds for nonexpansive fixed point iterations in normed spaces*. arXiv:2108.10969v1 (2021), 1-29.