Krasnoselskii-Mann fixed point iterations: sharp convergence rates

Roberto Cominetti
Universidad Adolfo Ibáñez
roberto.cominetti@uai.cl

AMSI Optimise Workshop
June 29-30, 2017 — Melbourne, Australia

Outline

Warmup: Banach-Picard iteration

When $T : X \to X$ is a ρ-contraction

$$(BP) \qquad \boxed{x^{n+1} = Tx^n}$$

$$\|x^{n+1} - x^n\| = \|Tx^n - x^n\| \leq \rho^n \|Tx^0 - x^0\| \to 0$$

$$\Downarrow$$

convergence $+$ error estimates $+$ stopping rule

Krasnoselskii-Mann sequential averaging process

$T : C \to C$ non-expansive / C convex bounded in $(X, \|\cdot\|)$

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

Krasnoselskii-Mann sequential averaging process

$T : C \to C$ non-expansive / C convex bounded in $(X, \|\cdot\|)$

$$(KM) \qquad \boxed{x^{n+1} = (1-\alpha_{n+1})\, x^n + \alpha_{n+1} T x^n}$$

- algorithm to compute & prove existence of fixed points
- **convex optimization:** Prox, Gradient, Douglas-Rachford, ADMM, POCS
- **dynamic programming:** stochastic shortest paths, Q-learning
- **continuous semigroups:** discretization of $\frac{dx}{dt} + [I - T](x) = 0$

Krasnoselskii-Mann sequential averaging process

$T : C \to C$ non-expansive / C convex bounded in $(X, \|\cdot\|)$

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

- algorithm to compute & prove existence of fixed points
- **convex optimization:** Prox, Gradient, Douglas-Rachford, ADMM, POCS
- **dynamic programming:** stochastic shortest paths, Q-learning
- **continuous semigroups:** discretization of $\frac{dx}{dt} + [I-T](x) = 0$

Questions:

- a) Existence of fixed points ?
- b) Convergence of iterates ?
- c) $\|Tx^n - x^n\| \to 0$? (Asymptotic Regularity)
- d) How fast ? (Rate of Convergence)

If we have $\|Tx^n - x^n\| \to 0$ *then...*

- $\Rightarrow$ all strong/weak cluster points are fixed points of T
- $\Rightarrow$ existence of fixed points: Browder-Göhde-Kirk'65

If we have $\|Tx^n - x^n\| \to 0$ *then...*

$\Rightarrow$ all strong/weak cluster points are fixed points of T

$\Rightarrow$ existence of fixed points: Browder-Göhde-Kirk'65

and since $\|x^n - \bar{x}\|$ decreases for all $\bar{x} \in \mathrm{Fix}T$

$\Rightarrow$ x^n converges strong/weak to a fixed point

$\Rightarrow$ convergence results: Krasnoselskii'55, Shaefer'57, Browder-Petryshyn'66, Edelstein'70, Groetsch'72, Ishikawa'76, Edelstein-O'Brien'78, Reich'79...

Example: θ-rotation in $\mathbb{R}^2$

$$T(x,y) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} \quad \Rightarrow \quad \|Tx^n - x^n\| = 2\sin(\tfrac{\theta}{2})\cos^n(\tfrac{\theta}{2})\|x^0\|$$

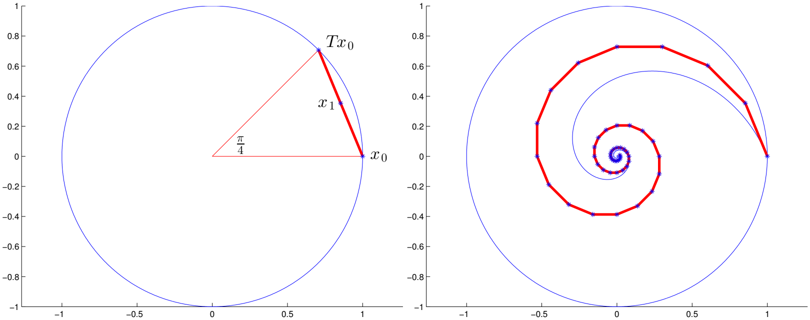


Figure: KM iterates for $\theta = \frac{\pi}{4}$ with $\alpha_n \equiv \frac{1}{2}$

Example: Right-shift on $\ell^1(\mathbb{N})$

$C = \{(p_i)_{i\in\mathbb{N}} \geq 0 : \sum_{i=0}^{\infty} p_i = 1\}$ is bounded with $\text{diam}(C) = 2$

$T(p_0, p_1, p_2, \ldots) = (0, p_0, p_1, p_2, \ldots)$ is an isometry

Example: Right-shift on $\ell^1(\mathbb{N})$

$C=\{(p_i)_{i\in\mathbb{N}}\geq 0:\sum_{i=0}^{\infty}p_i=1\}$ is bounded with $\mathrm{diam}(C)=2$

$T(p_0,p_1,p_2,\ldots)=(0,p_0,p_1,p_2,\ldots)$ is an isometry

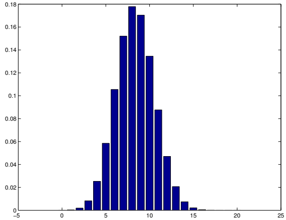
$$
\begin{aligned}
p^0 &= (1,0,0,0,\ldots)\\
p^1 &= (1-\alpha_1,\alpha_1,0,0,\ldots)\\
p^2 &= ((1-\alpha_2)(1-\alpha_1),(1-\alpha_2)\alpha_1+\alpha_2(1-\alpha_1),\alpha_2\alpha_1,0,\ldots)\\
p^3 &= \ldots
\end{aligned}
$$

Example: Right-shift on $\ell^1(\mathbb{N})$

$C = \{(p_i)_{i\in\mathbb{N}} \geq 0 : \sum_{i=0}^{\infty} p_i = 1\}$ is bounded with diam$(C) = 2$

$T(p_0, p_1, p_2, \ldots) = (0, p_0, p_1, p_2, \ldots)$ is an isometry

$$\begin{aligned}
p^0 &= (1, 0, 0, 0, \ldots) \\
p^1 &= (1-\alpha_1, \alpha_1, 0, 0, \ldots) \\
p^2 &= ((1-\alpha_2)(1-\alpha_1), (1-\alpha_2)\alpha_1 + \alpha_2(1-\alpha_1), \alpha_2\alpha_1, 0, \ldots) \\
p^3 &= \ldots
\end{aligned}$$



$p_k^n = \mathbb{P}(X_1 + \cdots + X_n = k)$

$X_i \sim$ Bernoulli(α_i)

$\|Tp^n - p^n\|_1 = 2 \max_k p_k^n$

60 years of $\|Tx^n - x^n\| \to 0$

- 1955 **Krasnoselskii:** X uniformly convex, C compact, $\alpha_n \equiv 1/2$
- 1957 **Schaefer:** extension to $\alpha_n \equiv \alpha$
- 1966 **Ishikawa**: extension to X strictly convex
- 1966 **Browder-Petryshyn**: X uniformly convex, Fix$(T) \neq \emptyset$, $\alpha_n \equiv \alpha$
- 1976 **Ishikawa**: X general, C compact, $\sum \alpha_n = \infty$, $\alpha_n \leq 1 - \epsilon$
- 1978 **Edelstein-O'Brien**: $\|Tx^n - x^n\| \to 0$ uniformly w.r.t x^0
- 1983 **Goebel-Kirk**: uniformly w.r.t. x^0 and T (for C given)
- 1992 **Baillon-Bruck**: $\alpha_n \equiv \alpha \Rightarrow \|Tx^n - x^n\| = O(1/\log n)$
- 1996 **Baillon-Bruck**: $\alpha_n \equiv \alpha \Rightarrow \|Tx^n - x^n\| = O(1/\sqrt{n})$
- 2003 **Kohlenbach**: $\|Tx^n - x^n\| \to 0$ only depends on diam(C)

Conjecture (Baillon-Bruck 1992)

There exists a universal constant κ such that

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

Theorem (Baillon-Bruck 1996)

For $\alpha_n \equiv \alpha$ constant, (BB) holds with $\kappa = 1/\sqrt{\pi}$.

Conjecture (Baillon-Bruck 1992)

There exists a universal constant κ such that

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

Theorem (Baillon-Bruck 1996)

For $\alpha_n \equiv \alpha$ constant, (BB) holds with $\kappa = 1/\sqrt{\pi}$.

Our contributions (C-Soto-Vaisman 2014, Bravo-C 2016, Bravo-C-Pavez 2017)

1. (BB) holds for general α_n with $\kappa = 1/\sqrt{\pi} \sim 0.5642$
2. Nonlinear maps: the constant $\kappa = 1/\sqrt{\pi}$ is the best possible
3. Affine maps: tight bound with $\kappa = \max_z \sqrt{z}e^{-z}I_0(z) \sim 0.4688$
4. Inexact KM: $x^{n+1} = (1-\alpha_{n+1})\,x^n + \alpha_{n+1}(Tx^n + e^{n+1})$

$$\boxed{\|x^n - Tx^n\| \le \kappa \frac{\text{diam}(C)}{\sqrt{\sum_{i=1}^n \alpha_i(1-\alpha_i)}}} \qquad \text{(BB)}$$

In this talk...

1. (BB) holds with $\kappa = 1/\sqrt{\pi} \sim 0.5642$
2. The constant $\kappa = 1/\sqrt{\pi}$ is the best possible

Along the way we unveil the probabilistic structure behind (KM)...

and discover that Krasnoselskii was "almost" right in taking $\alpha_n \equiv \frac{1}{2}$.

Streamline

Algorithms for convex optimization

$\Downarrow$

Fixed point iterations

$\Downarrow$

Recursive bounds by optimal transport

$\Downarrow$

Markov processes and combinatorics

$\Downarrow$

Rates of convergence with optimal constants

$\Downarrow$

Selection of optimal parameters

Outline

A recursive bound $\|x^m - x^n\| \leq d_{mn}$

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

- We look for sharp bounds for $\|Tx^n - x^n\| = \left\|x^{n+1} - x^n\right\| / \alpha_{n+1}$
- We achieve this by bounding $\|x^m - x^n\| \leq d_{mn}$ for all $m \leq n$

A recursive bound $\|x^m - x^n\| \leq d_{mn}$

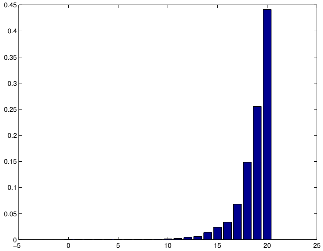
(KM) $$x^{n+1} = (1-\alpha_{n+1})\, x^n + \alpha_{n+1} T x^n$$

- We look for sharp bounds for $\|Tx^n - x^n\| = \left\|x^{n+1} - x^n\right\| / \alpha_{n+1}$
- We achieve this by bounding $\|x^m - x^n\| \leq d_{mn}$ for all $m \leq n$

Using the conventions $\alpha_0 = 1$ and $Tx^{-1} = x^0$, we have

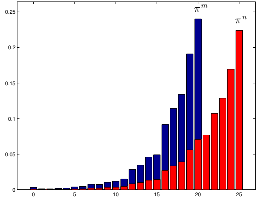
$$x^n = \sum_{i=0}^{n} \pi_i^n T x^{i-1}$$

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



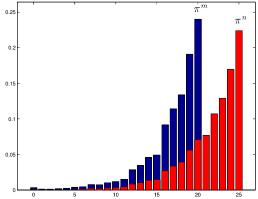
A recursive bound $\|x^m - x^n\| \leq d_{mn}$

$$x^m - x^n = \sum_{i=0}^{m} \pi_i^m T x^{i-1} - \sum_{j=0}^{n} \pi_j^n T x^{j-1}$$



A recursive bound $\|x^m - x^n\| \le d_{mn}$

$$x^m - x^n = \sum_{i=0}^{m} \pi_i^m T x^{i-1} - \sum_{j=0}^{n} \pi_j^n T x^{j-1}$$

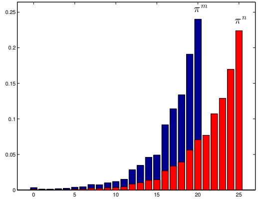


Let P_{mn} be the set of transport plans $z \ge 0$ taking π^m to π^n

$$\begin{aligned} \pi_i^m &= \textstyle\sum_{j=0}^{n} z_{ij} \\ \pi_j^n &= \textstyle\sum_{i=0}^{m} z_{ij} \end{aligned}$$

A recursive bound $\|x^m - x^n\| \leq d_{mn}$

$$x^m - x^n = \sum_{i=0}^{m}\sum_{j=0}^{n} z_{ij}Tx^{i-1} - \sum_{j=0}^{n}\sum_{i=0}^{m} z_{ij}Tx^{j-1}$$

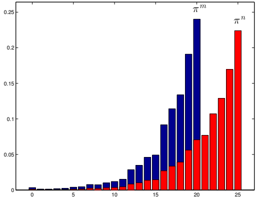


Let P_{mn} be the set of transport plans $z \geq 0$ taking π^m to π^n

$$\pi_i^m = \sum_{j=0}^{n} z_{ij}$$
$$\pi_j^n = \sum_{i=0}^{m} z_{ij}$$

A recursive bound $\|x^m - x^n\| \leq d_{mn}$

$$x^m - x^n = \sum_{i=0}^{m} \sum_{j=0}^{n} z_{ij} [Tx^{i-1} - Tx^{j-1}]$$

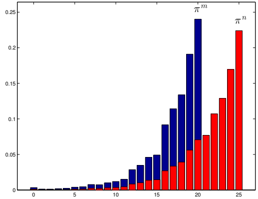


Let P_{mn} be the set of transport plans $z \geq 0$ taking π^m to π^n

$$\begin{aligned} \pi_i^m &= \textstyle\sum_{j=0}^n z_{ij} \\ \pi_j^n &= \textstyle\sum_{i=0}^m z_{ij} \end{aligned}$$

A recursive bound $\|x^m - x^n\| \leq d_{mn}$

$$\|x^m - x^n\| \leq \sum_{i=0}^{m} \sum_{j=0}^{n} z_{ij} \|x^{i-1} - x^{j-1}\|$$

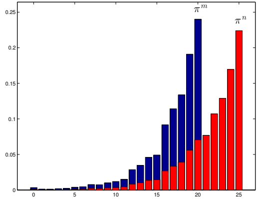


Let P_{mn} be the set of transport plans $z \geq 0$ taking π^m to π^n

$$\pi_i^m = \sum_{j=0}^{n} z_{ij}$$
$$\pi_j^n = \sum_{i=0}^{m} z_{ij}$$

A recursive bound $\|x^m - x^n\| \leq d_{mn}$

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

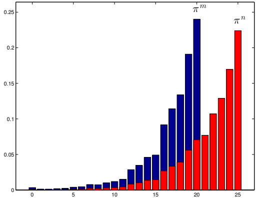


Let P_{mn} be the set of transport plans $z \geq 0$ taking π^m to π^n

$$\pi_i^m = \sum_{j=0}^{n} z_{ij}$$
$$\pi_j^n = \sum_{i=0}^{m} z_{ij}$$

A recursive bound $\|x^m - x^n\| \leq d_{mn}$

$$\|x^m - x^n\| \leq \sum_{i=0}^{m}\sum_{j=0}^{n} z_{ij}\, d_{i-1,j-1} \quad \longrightarrow \quad \min_z$$



Let P_{mn} be the set of transport plans $z \geq 0$ taking π^m to π^n

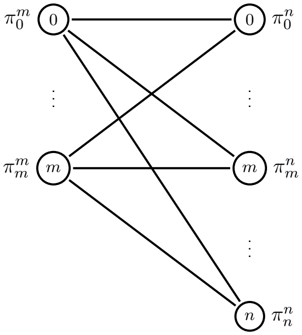
$$\begin{aligned} \pi_i^m &= \textstyle\sum_{j=0}^{n} z_{ij} \\ \pi_j^n &= \textstyle\sum_{i=0}^{m} z_{ij} \end{aligned}$$

Recursive optimal transports

Assume w.l.o.g. that diam(C)=1 and set $d_{-1,n} = 1$ for all $n \in \mathbb{N}$

Define d_{mn} inductively for $0 \leq m \leq n$ as

$$(R) \qquad d_{mn} \triangleq \min_{z \in P_{mn}} \sum_{i=0}^{m} \sum_{j=0}^{n} z_{ij}\, d_{i-1,j-1}$$



Property

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

The bounds d_{mn} are fully determined by the α_k's and are:

- Independent of the space X
- Independent of the operator $T : C \to C$

Property

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

The bounds d_{mn} are fully determined by the α_k's and are:

- Independent of the space X
- Independent of the operator $T : C \to C$

Theorem (Bravo-C 2016)

These bounds are the best possible: There exists a nonexpansive $T : C \to C$ on the unit cube $C = [0,1]^{\mathbb{N}} \subseteq \ell^\infty(\mathbb{N})$ and a corresponding KM sequence such that $\|x^m - x^n\|_\infty = d_{mn}$ for all $m, n \in \mathbb{N}$.

Proof: T built from dual solutions of the optimal transports. $\square$

Property

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

The bounds d_{mn} are fully determined by the α_k's and are:

- Independent of the space X
- Independent of the operator $T : C \to C$

Theorem (Bravo-C 2016)

These bounds are the best possible: There exists a nonexpansive $T : C \to C$ on the unit cube $C = [0,1]^{\mathbb{N}} \subseteq \ell^\infty(\mathbb{N})$ and a corresponding KM sequence such that $\|x^m - x^n\|_\infty = d_{mn}$ for all $m, n \in \mathbb{N}$.

Proof: T built from dual solutions of the optimal transports. □

$\Rightarrow$ The d_{mn}'s are the right object to obtain sharp bounds !

Restatement of (BB)

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

Restatement of (BB)

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

$$\downarrow$$

$$\frac{d_{n,n+1}}{\alpha_{n+1}} \leq \frac{1}{\sqrt{\pi}} \frac{1}{\sqrt{\sum_{k=1}^{n} \alpha_k(1-\alpha_k)}} \quad ?$$

Metric properties of the optimal transport bounds

Theorem (Aygen-Satik 1994, Bravo-C 2016)

$d(m,n) = d_{mn}$ defines a metric over the set $\mathcal{N} = \{-1, 0, 1, 2, 3, \ldots\}$.

Original proof is 30+ pages long. Short proof in Bravo-C 2016 (3 pages).

Metric properties of the optimal transport bounds

Theorem (Aygen-Satik 1994, Bravo-C 2016)

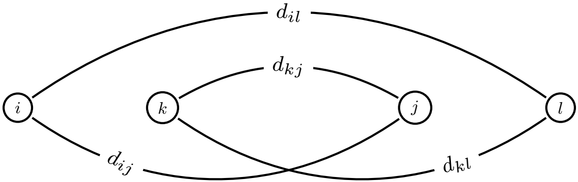
$d(m,n) = d_{mn}$ defines a metric over the set $\mathcal{N} = \{-1, 0, 1, 2, 3, \ldots\}$.

Original proof is 30+ pages long. Short proof in Bravo-C 2016 (3 pages).

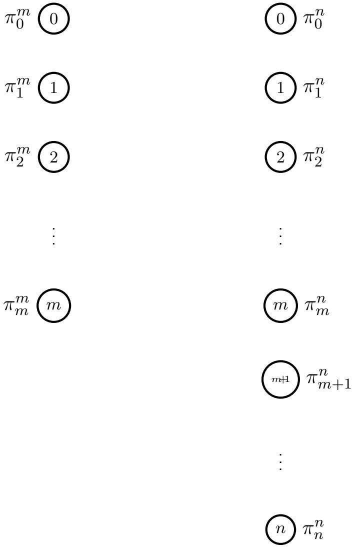
Theorem (Aygen-Satik 1994, Bravo-C. 2016)

For all $i \leq k \leq j \leq l$ we have $d_{il} + d_{kj} \leq d_{ij} + d_{kl}$.

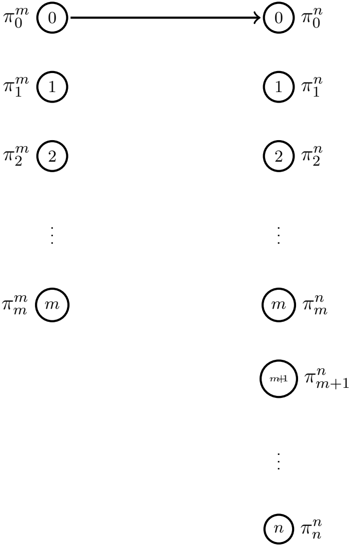
Original proof is 80+ pages long but holds for $\alpha_n \in [0,1]$.
We found a 2-page proof for the case $\alpha_n \geq 1/2$.



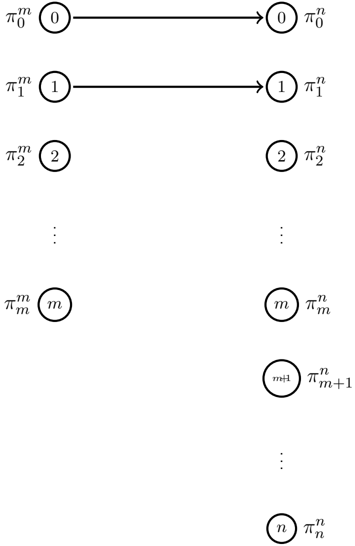
Optimal transport for $\alpha_n \geq 1/2$



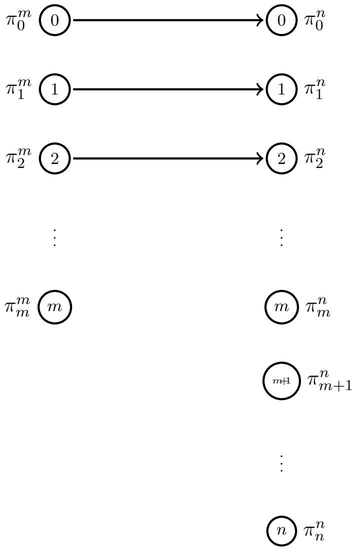
Optimal transport for $\alpha_n \geq 1/2$



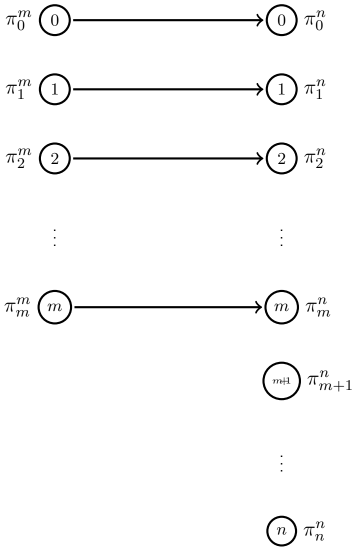
Optimal transport for $\alpha_n \geq 1/2$



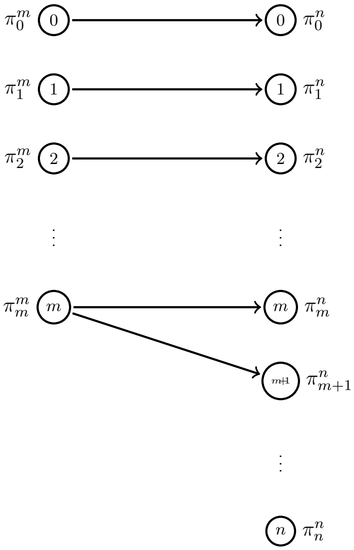
Optimal transport for $\alpha_n \geq 1/2$



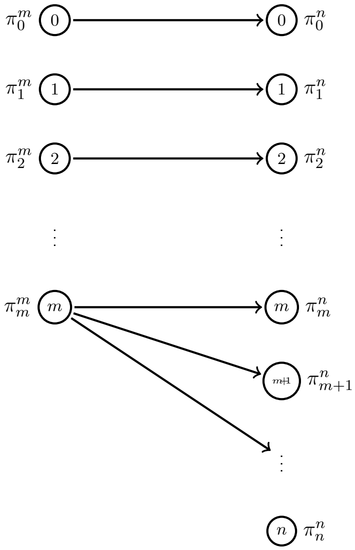
Optimal transport for $\alpha_n \geq 1/2$



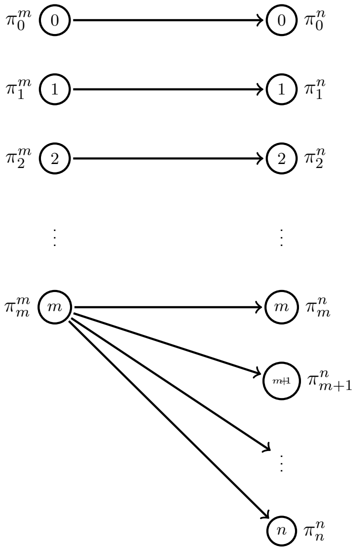
Optimal transport for $\alpha_n \geq 1/2$



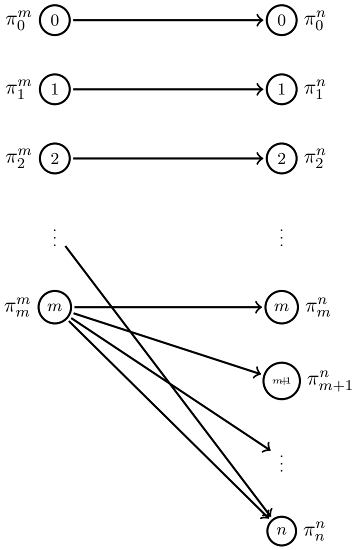
Optimal transport for $\alpha_n \geq 1/2$



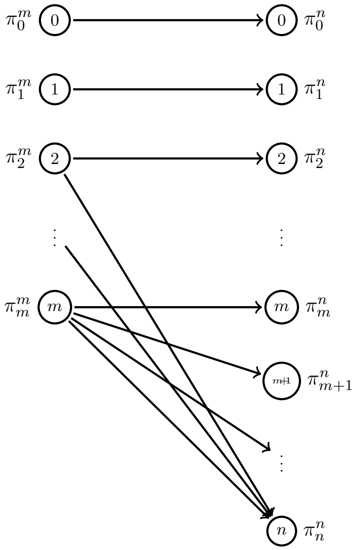
Optimal transport for $\alpha_n \geq 1/2$



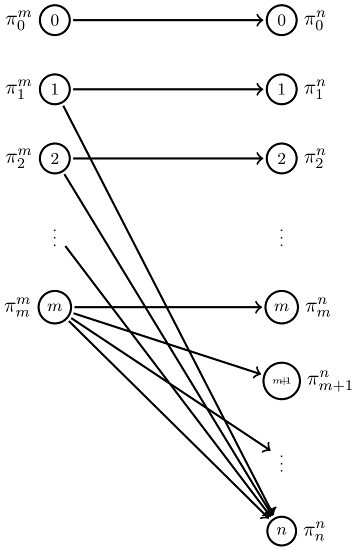
Optimal transport for $\alpha_n \geq 1/2$



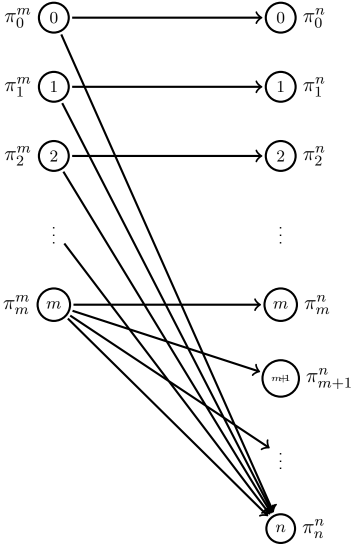
Optimal transport for $\alpha_n \geq 1/2$



Optimal transport for $\alpha_n \geq 1/2$



Optimal transport for $\alpha_n \geq 1/2$



Optimal transport for $\alpha_n \geq 1/2$

The optimal solution is

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

and the recursion *"simplifies"* to

$$d_{mn} = \sum_{j=m+1}^{n-1} \pi_j^n d_{m-1,j-1} + \sum_{i=0}^{m-1} (\pi_i^m - \pi_i^n) d_{i-1,n-1} + (\pi_m^m - \sum_{j=m}^{n-1} \pi_j^n) d_{m-1,n-1}$$

$$\alpha_n \equiv \alpha \geq \frac{1}{2}$$

$$\begin{aligned} d_{6,10}(\alpha) =& \alpha(4 - 36\alpha + 328\alpha^2 - 2671\alpha^3 + 19853\alpha^4 - 132880\alpha^5 + 785003\alpha^6 \\ & - 4016624\alpha^7 + 17541102\alpha^8 - 64796454\alpha^9 + 201809157\alpha^{10} \\ & - 530670200\alpha^{11} + 1183318617\alpha^{12} - 2250818306\alpha^{13} + 3675506816\alpha^{14} \\ & - 5184593492\alpha^{15} + 6352439437\alpha^{16} - 6792441644\alpha^{17} + 6361687020\alpha^{18} \\ & - 5232669869\alpha^{19} + 3785701567\alpha^{20} - 2409974375\alpha^{21} + 1348858198\alpha^{22} \\ & - 662337623\alpha^{23} + 284299971\alpha^{24} - 106102624\alpha^{25} + 34171973\alpha^{26} \\ & - 9400913\alpha^{27} + 2178730\alpha^{28} - 417352\alpha^{29} + 64328\alpha^{30} \\ & - 7667\alpha^{31} + 663\alpha^{32} - 37\alpha^{33} + \alpha^{34}) \end{aligned}$$

...?????????

Outline

Upper estimate: $d_{mn} \leq c_{mn}$

Consider the feasible non-optimal transport plan

$$\hat{z}_{ij} = \begin{cases} \pi_j^n & \text{for } i = j \leq m \\ \pi_i^m \pi_j^n & \text{for } j = m+1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

Upper estimate: $d_{mn} \leq c_{mn}$

Consider the feasible non-optimal transport plan

$$\hat{z}_{ij} = \begin{cases} \pi_j^n & \text{for } i = j \leq m \\ \pi_i^m \pi_j^n & \text{for } j = m+1, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

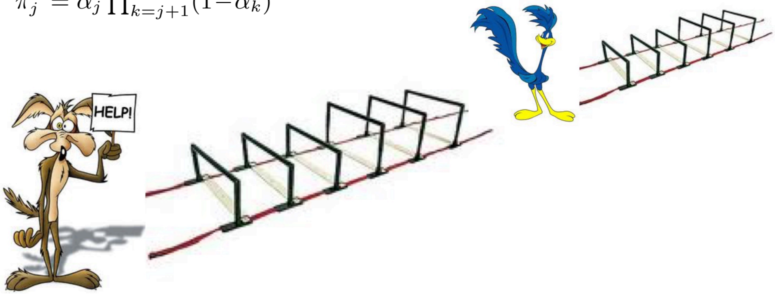
Setting $c_{-1,n} = 1$ we get inductively

$$\|x^m - x^n\| \leq d_{mn} \leq c_{mn} \triangleq \sum_{i=0}^{m} \sum_{j=m+1}^{n} \pi_i^m \pi_j^n c_{i-1,j-1}$$

Probabilistic interpretation of c_{mn}

$$\mathbb{P}[C_i = 1] = \mathbb{P}[R_i = 1] = \alpha_i$$

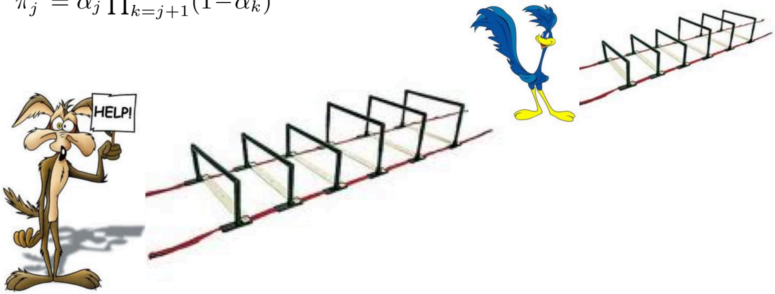
$$\pi_j^n = \alpha_j \prod_{k=j+1}^{n} (1-\alpha_k)$$



Probabilistic interpretation of c_{mn}

$$\mathbb{P}[C_i = 1] = \mathbb{P}[R_i = 1] = \alpha_i$$

$$\pi_j^n = \alpha_j \prod_{k=j+1}^{n} (1-\alpha_k)$$

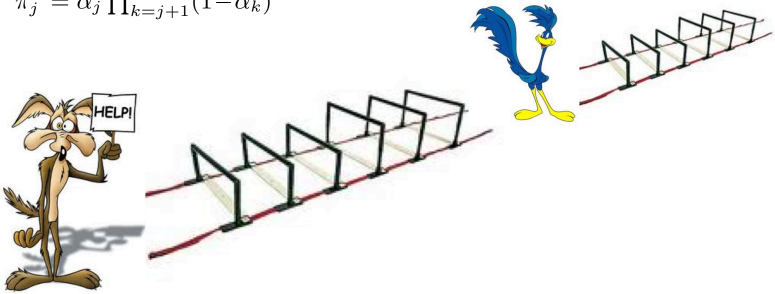


$$\tilde{c}_{mn} = \mathbb{P}[\text{roadrunner escapes}]$$

Probabilistic interpretation of c_{mn}

$\mathbb{P}[C_i = 1] = \mathbb{P}[R_i = 1] = \alpha_i$

$\pi_j^n = \alpha_j \prod_{k=j+1}^n (1-\alpha_k)$

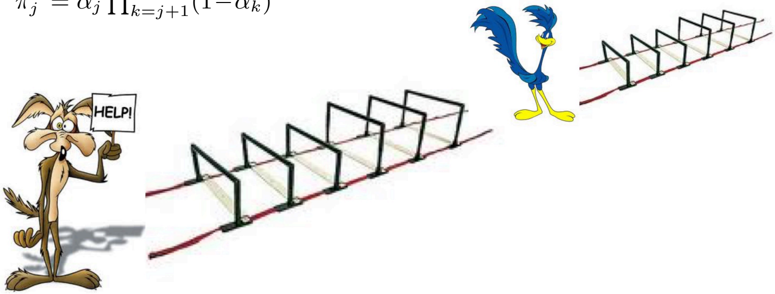


$$\tilde{c}_{mn} = \sum_{i=0}^{m} \sum_{j=m+1}^{n} \pi_i^m \pi_j^n \tilde{c}_{i-1,j-1}$$

Probabilistic interpretation of c_{mn}

$$\mathbb{P}[C_i = 1] = \mathbb{P}[R_i = 1] = \alpha_i$$

$$\pi_j^n = \alpha_j \prod_{k=j+1}^n (1-\alpha_k)$$



$$c_{mn} = \mathbb{P}\left(\sum_k^n C_i > \sum_k^m R_i, \forall k = m+1, \ldots, 1\right)$$

Coyote must fall more often than Roadrunner

The random walk and the gambler's ruin appear...

Using probabilistic arguments we get an explicit formula for the bound

$$\|Tx^n - x^n\| \leq \frac{c_{n,n+1}}{\alpha_{n+1}} = \mathbb{E}[F(M)]$$

where M is a sum of independent nonhomogeneous Bernoulli trials

$$M = M_1 + \ldots + M_n \quad ; \quad \mathbb{P}(M_i = 1) = p_i \triangleq 2\alpha_i(1-\alpha_i)$$

and $F(m)$ is the probability that a standard random walk in $\mathbb{Z}$ remains non-negative for the first m stages

$$F(m) = \binom{m}{\lfloor m/2 \rfloor} 2^{-m}.$$

Upper estimate

Thus (BB) has been reduced to show that

$$\mathbb{E}[F(M)] \leq \frac{1}{\sqrt{\pi \sum_{i=1}^{n} \alpha_i(1-\alpha_i)}}$$

Upper estimate

Thus (BB) has been reduced to show that

$$\mathbb{E}[F(M)] \leq \frac{1}{\sqrt{\pi \sum_{i=1}^{n} \alpha_i(1-\alpha_i)}}$$

Since $p_i = 2\alpha_i(1-\alpha_i)$ this is equivalent to

$$\underbrace{\sqrt{\tfrac{\pi}{2}(p_1 + \ldots + p_n)}\, \mathbb{E}[F(M_1 + \ldots + M_n)]}_{R_n(p)} \leq 1$$

Upper estimate

Thus (BB) has been reduced to show that

$$\mathbb{E}[F(M)] \leq \frac{1}{\sqrt{\pi \sum_{i=1}^{n} \alpha_i(1-\alpha_i)}}$$

Since $p_i = 2\alpha_i(1-\alpha_i)$ this is equivalent to

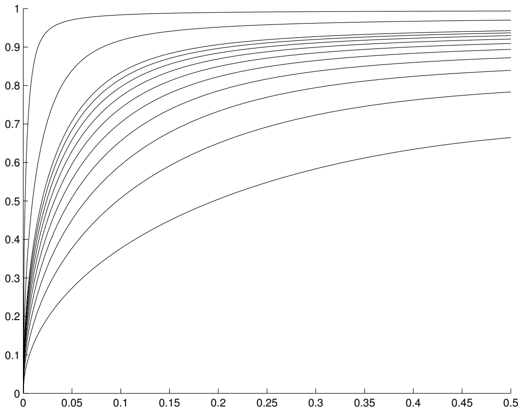
$$\underbrace{\sqrt{\tfrac{\pi}{2}(p_1 + \ldots + p_n)}\, \mathbb{E}[F(M_1 + \ldots + M_n)]}_{R_n(p)} \leq 1$$

Lemma

$R_n(p)$ *is maximal when* $p_i \in \{u, \frac{1}{2}\}$ *for some* $0 < u < \frac{1}{2}$

Upper estimate – Case 1: all $p_i = u$

$$R_n(p) = \sqrt{\tfrac{\pi}{2} n u}\, \mathbb{E}[F(B(n,u))] = \sqrt{\tfrac{\pi}{2} n u}\; {}_2F_1(-n, \tfrac{1}{2}; 2; 2u) \leq 1$$



Upper estimate – Case 2: some $p_i = \frac{1}{2}$

Suppose $p_1 = \frac{1}{2}$ and let $S = M_2 + \ldots + M_n$. Conditioning on M_1

$$\mathbb{E}[F(M)] = \mathbb{E}[G(S)]$$

where $G(k) = \frac{1}{2}[F(k) + F(k+1)]$ is convex so we may use the following Hoeffding-type inequality

Theorem (C.-Soto-Vaisman'2014, *Israel J. Math.*)

$\mathbb{E}[G(S)] \leq \mathbb{E}[G(Z)]$ *where* $Z \sim$ *Poisson*(z) *with* $z = \mathbb{E}(S)$.

Upper estimate – Case 2: some $p_i = \frac{1}{2}$

Suppose $p_1 = \frac{1}{2}$ and let $S = M_2 + \ldots + M_n$. Conditioning on M_1

$$\mathbb{E}[F(M)] = \mathbb{E}[G(S)]$$

where $G(k) = \frac{1}{2}[F(k) + F(k+1)]$ is convex so we may use the following Hoeffding-type inequality

Theorem (C.-Soto-Vaisman'2014, *Israel J. Math.*)

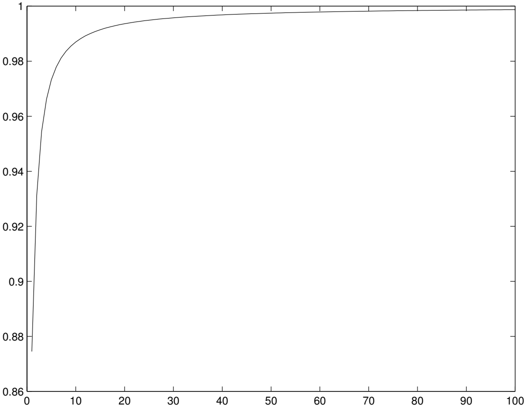
$\mathbb{E}[G(S)] \leq \mathbb{E}[G(Z)]$ *where* $Z \sim$ *Poisson*(z) *with* $z = \mathbb{E}(S)$*.*

$$\Rightarrow \mathbb{E}[F(M)] \leq \mathbb{E}[G(Z)] = I_0(z) + (1 - \tfrac{1}{2z})I_1(z)$$

with $I_0(z), I_1(z)$ modified Bessel functions

Upper estimate – Case 2: some $p_i = \frac{1}{2}$

$$R(p) \leq \sqrt{\tfrac{\pi}{2}(\tfrac{1}{2}+z)}\,[I_0(z) + (1-\tfrac{1}{2z})I_1(z)] \leq 1$$



Conclusion: (BB) holds with $\kappa = 1/\sqrt{\pi} \sim 0.5642$

Both cases combined yield

Theorem (C.-Soto-Vaisman'2014, *Israel J. Math.*)

$$\|Tx^n - x^n\| \leq \frac{diam(C)}{\sqrt{\pi \sum_{k=1}^{n} \alpha_k (1-\alpha_k)}}$$

Conclusion: (BB) holds with $\kappa = 1/\sqrt{\pi} \sim 0.5642$

Both cases combined yield

Theorem (C.-Soto-Vaisman'2014, *Israel J. Math.*)

$$\|Tx^n - x^n\| \leq \frac{\mathit{diam}(C)}{\sqrt{\pi \sum_{k=1}^{n} \alpha_k(1-\alpha_k)}}$$

Remarks: (Baillon-C.-Vaisman'2016, *Comb. Prob. & Computing*)

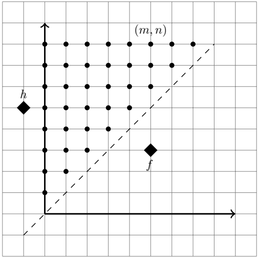
- If T is affine (BB) holds with $\kappa = \max_{u\geq 0} \sqrt{u}\, e^{-u} I_0(u) \sim 0.4688.$
- This bound is tight and is attained by the right shift of Example 2.

Outline

A related stochastic process

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

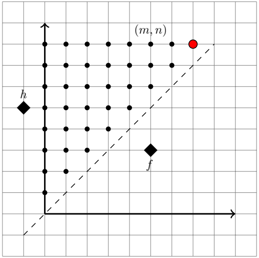
d_{mn} can be interpreted as the absorption probability of a finite Markov chain $\mathcal{D}$ with state space $\mathcal{S} = \{(m,n) : 0 \leq m < n\} \cup \{h, f\}$.



A related stochastic process

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

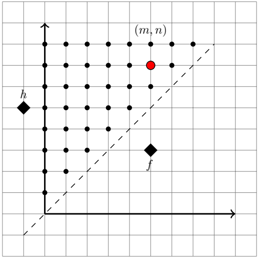
d_{mn} can be interpreted as the absorption probability of a finite Markov chain $\mathcal{D}$ with state space $\mathcal{S} = \{(m,n) : 0 \leq m < n\} \cup \{h, f\}$.



A related stochastic process

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

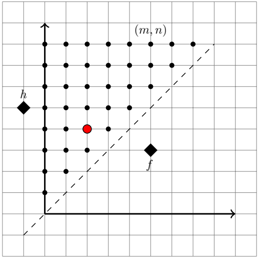
d_{mn} can be interpreted as the absorption probability of a finite Markov chain $\mathcal{D}$ with state space $\mathcal{S} = \{(m,n) : 0 \leq m < n\} \cup \{h, f\}$.



A related stochastic process

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

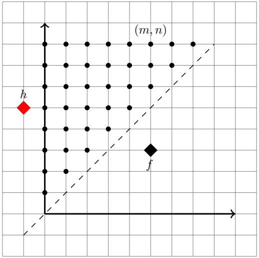
d_{mn} can be interpreted as the absorption probability of a finite Markov chain $\mathcal{D}$ with state space $\mathcal{S} = \{(m,n) : 0 \leq m < n\} \cup \{h, f\}$.



A related stochastic process

$$d_{mn} = \sum_{i=0}^{m} \sum_{j=0}^{n} z_{ij}^{mn} d_{i-1,j-1} = \mathbb{P}(\mathcal{D} \text{ attains } h | mn)$$

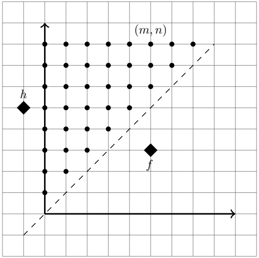
d_{mn} can be interpreted as the absorption probability of a finite Markov chain $\mathcal{D}$ with state space $\mathcal{S} = \{(m,n) : 0 \leq m < n\} \cup \{h, f\}$.



A related stochastic process

$$d_{mn} = \sum_{i=0}^{m} \sum_{j=0}^{n} z_{ij}^{mn} d_{i-1,j-1} = \mathbb{P}(\mathcal{D} \text{ attains } h|mn)$$

d_{mn} can be interpreted as the absorption probability of a finite Markov chain $\mathcal{D}$ with state space $\mathcal{S} = \{(m,n) : 0 \le m < n\} \cup \{h, f\}$.



Similarly, for the Markov chain $\mathcal{C}$ defined by the sub-optimal transports

$$c_{mn} = \sum_{i=0}^{m} \sum_{j=m+1}^{n} \pi_i^m \pi_j^n c_{i-1,j-1} = \mathbb{P}(\mathcal{C} \text{ attains } h|mn)$$

Baillon-Bruck bound for constant $\alpha_n \equiv \alpha \in [1/2, 1[$

$$\sqrt{n\alpha(1-\alpha)}\, \|x^n - Tx^n\| \leq \kappa \,?$$

Baillon-Bruck bound for constant $\alpha_n \equiv \alpha \in [1/2, 1[$

$$\boxed{\sqrt{n\alpha(1-\alpha)}\,\|x^n - Tx^n\| \leq \kappa\ ?}$$

Consider the map T that attains the bounds d_{mn}, then

$$\sqrt{n\alpha(1-\alpha)}\,\|x^n - Tx^n\| = \kappa_n(\alpha) \leq \tilde{\kappa}_n(\alpha)$$

where

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

Baillon-Bruck bound for constant $\alpha_n \equiv \alpha \in [1/2, 1[$

$$\sqrt{n\alpha(1-\alpha)}\,\|x^n - Tx^n\| \leq \kappa\ ?$$

Consider the map T that attains the bounds d_{mn}, then

$$\sqrt{n\alpha(1-\alpha)}\,\|x^n - Tx^n\| = \kappa_n(\alpha) \leq \tilde{\kappa}_n(\alpha)$$

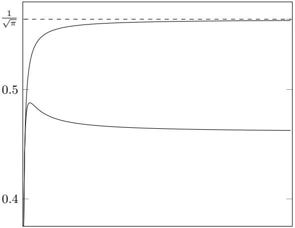
where

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

Moreover, the previous analysis of the c_{mn}'s gives explicitely

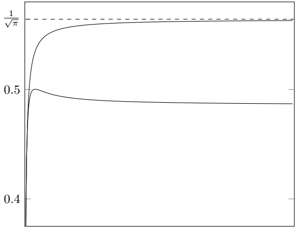
$$\tilde{\kappa}_n(\alpha) = \frac{1}{\pi}\int_0^{4n\alpha(1-\alpha)} \sqrt{\tfrac{1}{s} - \tfrac{1}{4n\alpha(1-\alpha)}}\,\left(1-\tfrac{s}{n}\right)^n ds \nearrow \frac{1}{\pi}\Gamma\left(\tfrac{1}{2}\right) = \frac{1}{\sqrt{\pi}}.$$

Difference between $\kappa_n(\alpha)$ *and* $\tilde{\kappa}_n(\alpha)$ *for* $n = 1, \ldots, 300$



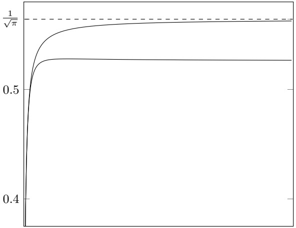
$\alpha = 0.5.$

Difference between $\kappa_n(\alpha)$ *and* $\tilde{\kappa}_n(\alpha)$ *for* $n = 1, \ldots, 300$



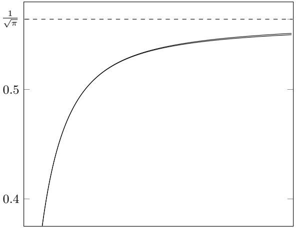
$\alpha = 0.65.$

Difference between $\kappa_n(\alpha)$ *and* $\tilde{\kappa}_n(\alpha)$ *for* $n = 1, \ldots, 300$



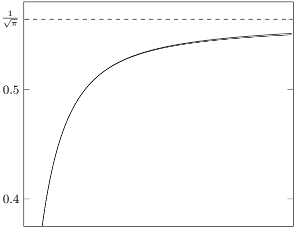
$\alpha = 0.85.$

Difference between $\kappa_n(\alpha)$ *and* $\tilde{\kappa}_n(\alpha)$ *for* $n = 1, \ldots, 300$



$\alpha = 0.99.$

Difference between $\kappa_n(\alpha)$ *and* $\tilde{\kappa}_n(\alpha)$ *for* $n = 1, \ldots, 300$

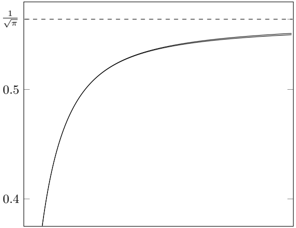


$\alpha = 0.99.$

Lemma

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

Difference between $\kappa_n(\alpha)$ *and* $\tilde{\kappa}_n(\alpha)$ *for* $n = 1, \ldots, 300$



$\alpha = 0.99.$

Lemma

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

Theorem (Bravo-C.'2016)

The constant $\kappa = \frac{1}{\sqrt{\pi}}$ *is the best possible.*

Proof. Taking $\alpha_n = 1 - \ln(n)/n$ we get $\kappa_n(\alpha_n) \sim \tilde{\kappa}_n(\alpha_n) \to \frac{1}{\sqrt{\pi}}$. □

Outline

Best constant stepsize $\alpha_n \equiv \alpha$?

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

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

The minimum seems to be attained at $\alpha = 0.4623...$!

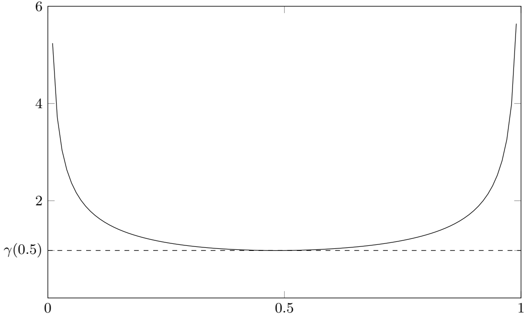


Figure: The rate $\gamma(\alpha)$ as a function of α (for $0.01 \leq \alpha \leq 0.99$).

Krasnoselskii's original iteration $\alpha_n \equiv \frac{1}{2}$

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

For $\alpha = \frac{1}{2}$ the sup seems to be attained at $n = 8$

Krasnoselskii's original iteration $\alpha_n \equiv \frac{1}{2}$

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

For $\alpha = \frac{1}{2}$ the sup seems to be attained at $n = 8$

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

$\Rightarrow$ the sharp rate in Krasnoselskii's iteration would be

$$\gamma(\tfrac{1}{2}) = \frac{46302245}{67108864}\sqrt{2} \sim 0.9757 \qquad \text{(smaller than } \tfrac{2}{\sqrt{\pi}} \sim 1.1284\text{)}$$

Sharp rate in Hilbert spaces?

- For $\alpha_n \equiv \alpha$, Browder & Petryshin proved $\sum \|x^{n+1} - x^n\|^2 < \infty$
- Since $\|x^{n+1} - x^n\|$ is decreasing, this readily gives

$$\|x^n - Tx^n\| = o(1/\sqrt{n})$$

- Which is the exact rate of asymptotic regularity in the Hilbert case?

What about errors?

Inexact KM:

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

- **Controlled errors**:

  Recently solved with M. Bravo and M. Pavez-Signé (available on arXiv).

- **Stochastic errors**:

Thanks !

Preprints available at

https://sites.google.com/site/cominettiroberto/