Rates of convergence for the Krasnoselskii-Mann fixed point iteration

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

Based on joint work with

Mario Bravo, Matías Pavez-Signé, José Soto, José Vaisman

One-World Optimization Seminar

February 8th, 2021

Outline

Krasnoselskii-Mann sequential averaging process

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

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

- algorithm to compute & prove existence of fixed points
- **convex optimization:** Gradient, Prox, Douglas-Rachford, ADMM, POCS...
- **stochastic processes:** stochastic shortest path, Q-learning, random walks
- **evolution equations:** discretization of $\frac{dx}{dt} + [I - T](x) = 0$

Questions:
$$\begin{cases} \text{a)} \quad \|x^n - Tx^n\| \to 0 \ ? & \text{(Asymptotic Regularity)} \\ \text{b)} \quad \text{How fast ?} & \text{(Rate of Convergence)} \end{cases}$$

Example: θ-rotation in $\mathbb{R}^2$

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

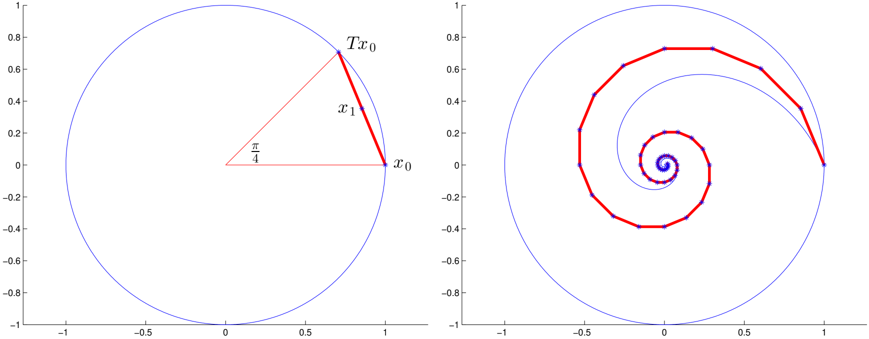


Figure: KM iterates for $\theta = \frac{\pi}{4}$ with $\alpha_n \equiv \frac{1}{2}$

a) If $\|x^n - Tx^n\| \to 0$ then...

- every strong/weak cluster point is a fixed point of T
- if C is strong/weak compact $\Rightarrow$ existence of fixed points (Browder-Göhde-Kirk'65)
- moreover $\|x^n - \bar{x}\|$ decreases for all $\bar{x} \in \mathrm{Fix}\, T$, hence x^n converges strong/weak to a fixed point (Krasnoselskii'55, Shaefer'57, Browder-Petryshyn'66, Edelstein'70, Groetsch'72, Ishikawa'76, Edelstein-O'Brien'78, Reich'79...)

b) How fast?

- convergence rates
- non-asymptotic error bounds

Example: Gradient descent

$f : \mathcal{H} \to \mathbb{R}$ convex with L-Lipschitz gradient, $\mathcal{H}$ Hilbert

$$(G) \qquad \boxed{x^{n+1} = x^n - \gamma_{n+1} \nabla f(x^n)}$$

From Baillon-Haddad's Theorem

$$\nabla f \text{ is } L\text{-Lipschitz} \quad \Leftrightarrow \quad Tx \triangleq x - \tfrac{2}{L} \nabla f(x) \text{ is non-expansive}$$

and (G) can be written as

$$x^{n+1} = (1 - \tfrac{\gamma_{n+1} L}{2}) x^n + \tfrac{\gamma_{n+1} L}{2} T x^n$$

Remarks:

- Fix$\,(T) = \operatorname{Argmin}_{x \in \mathcal{H}} f(x)$
- Residuals $|x^n - Tx^n| = \frac{2}{L} |\nabla f(x^n)|$

Example: Proximal point method

$\min_{x\in\mathcal{H}} f(x)$ where $f \in \Gamma_0(\mathcal{H})$ with $\mathcal{H}$ a Hilbert space

$(Prox)$ $$\boxed{x^{n+1} = \mathrm{Prox}_{\lambda f}(x^n) = \mathrm{argmin}_{y\in\mathcal{H}}\, f(y) + \tfrac{1}{2\lambda}|y - x^n|^2}$$

Theorem (Moreau 1965)

$f_\lambda(x) \triangleq \min_{y\in\mathcal{H}} f(y) + \frac{1}{2\lambda}|y - x|^2$ is a smooth convex minorant of f
$\nabla f_\lambda(x) = \frac{1}{\lambda}[x - \mathrm{Prox}_{\lambda f}(x)]$ is $\frac{1}{\lambda}$-Lipschitz.

$(Prox)$ can be rewritten as a Gradient and/or a (KM) iteration

$$x^{n+1} = x^n - \lambda\nabla f_\lambda(x^n) = \tfrac{1}{2}x^n + \tfrac{1}{2}R_{\lambda f}(x^n)$$

for the non-expansive *reflection operator* $R_{\lambda f}(x) = 2\,\mathrm{Prox}_{\lambda f}(x) - x$.

Remark: $\mathrm{Fix}\,(R_{\lambda f}) = \mathrm{Fix}\,(\mathrm{Prox}_{\lambda f}) = \mathrm{Argmin}\, f_\lambda = \mathrm{Argmin}\, f$.

Example: Douglas-Rachford

$\min_{x\in\mathcal{H}} f(x)+g(x)$ with $f,g\in\Gamma_0(\mathcal{H})$ and $\mathcal{H}$ Hilbert

$$(DR) \qquad \boxed{x^{n+1}=(1-\alpha)\,x^n+\alpha\,R_{\lambda f}\circ R_{\lambda g}(x^n)}$$

Special cases:

- Alternating Direction Method of Multipliers (ADMM)
- Spingarn's partial inverse method
- Decomposition of separable problems

Example: Forward-Backward

$\min_{x\in\mathcal{H}} f(x)+g(x)$ with $f:\mathcal{H}\to\mathbb{R}$ convex with L-Lipschitz gradient, $g\in\Gamma_0(\mathcal{H})$

$$(FB) \qquad \boxed{x^{n+1}=\mathrm{Prox}_{\lambda g}(x^n-\gamma\nabla f(x^n))}.$$

Recall: $S:\mathcal{H}\to\mathcal{H}$ is α-averaged if $S(x)=\alpha x+(1-\alpha)Tx$ with T nonexpansive.

- $x-\gamma\nabla f(x)$ is α-averaged with $\alpha=(1-\frac{\gamma L}{2})$
- $\mathrm{Prox}_{\lambda g}$ is $\frac{1}{2}$-averaged

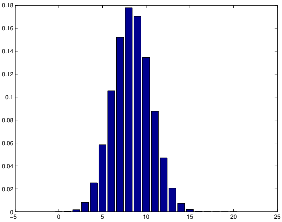
Hence, (FB) is again a (KM) iteration since:

Lemma: If S_i are α_i-averaged for $i=1,2$ then $S_1\circ S_2$ is $(\alpha_1\alpha_2)$-averaged.

Example: Right-shift in $C=\{x\in\ell^1_+(\mathbb{N}):\sum_{i=0}^\infty x_i=1\}$

$T(x_0,x_1,x_2,\ldots)=(0,x_0,x_1,x_2,\ldots)$ is an ℓ^1-isometry

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



$x^n_k=\mathbb{P}(X_1+\cdots+X_n=k)$

$X_i\sim$ Bernoulli(α_i)

$\|x^n-Tx^n\|_1=2\max_k x^n_k$

Example: "Lazy" random walks

A random walk on a metric space (X, d) is described by $\mu^{n+1} = T\mu^n$, with the evolution operator T induced by a Markov transition kernel $(m_x)_{x\in X}$

$$T\mu(A) \triangleq \int_X m_x(A)\, d\mu(x) \qquad \forall A \in \mathcal{B}(X), \forall \mu \in \mathcal{P}_1(X).$$

- Fix (T) is the set of invariant measures
- T is nonexpansive in total variation $\|\cdot\|_{\text{tv}}$
- residuals $\|\mu^n - T\mu^n\|_{\text{tv}}$

For *lazy random walks* with kernel

$$m_x^\alpha = (1-\alpha)\delta_x + \alpha\, m_x$$

the evolution is a (KM) iteration.

From PEP (back) to optimal transport

- PEP techniques, introduced by Drori & Teboulle around 2012, yield convergence rates for convex optimization, either in terms of the function values $f(x^n) - f(x^*)$ or the gradient norms $|\nabla f(x^n)|$. These techniques seem inherently restricted to Hilbert spaces.
- For fixed point iterations the natural metric is the residual $\|x^n - Tx^n\|$. An approach to get convergence rates was proposed by Baillon & Bruck in 1992-1996, and completed by C-Soto-Vaisman in 2014. The principle is similar to PEP except that it uses optimal transport instead of semidefinite programming. It works on general normed spaces.

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

Conjecture (Baillon-Bruck 1992)

There exists a universal constant κ such that

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

Theorem (Baillon-Bruck 1996)

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

Our contribution (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 tight
3. Affine maps: tight bound with $\kappa = \max_z \sqrt{z}e^{-z}I_0(z) \sim 0.4688$
4. Extension to inexact KM: $x^{n+1} = (1-\alpha_{n+1})\,x^n + \alpha_{n+1}(Tx^n + \varepsilon^{n+1})$
5. Rates of convergence for nonlinear semigroups in continuous time

Outline

A recursive bound

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

- We look for sharp bounds for $\|x^n - Tx^n\| = \left\|x^{n+1} - x^n\right\| / \alpha_{n+1}$
- This is achieved by bounding $\|x^m - x^n\|$ for all $m \leq n$

A recursive bound

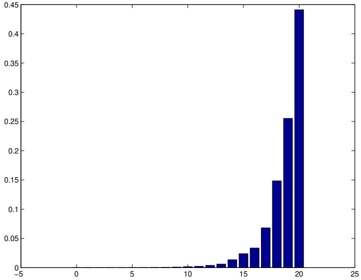
(KM) $$\boxed{x^{n+1} = (1-\alpha_{n+1})\, x^n + \alpha_{n+1}\, Tx^n}$$

- We look for sharp bounds for $\|x^n - Tx^n\| = \left\|x^{n+1} - x^n\right\| / \alpha_{n+1}$
- This is achieved by bounding $\|x^m - x^n\|$ for all $m \leq n$

Set $\alpha_0 = 1$ and $Tx^{-1} = x^0$

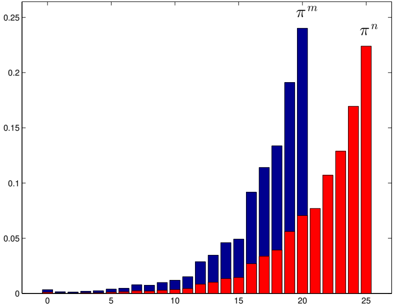
$$\boxed{x^n = \sum_{i=0}^{n} \pi_i^n \, Tx^{i-1}}$$

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



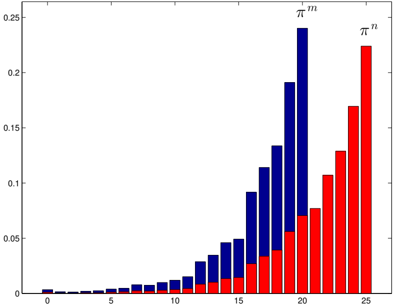
A recursive bound $\|x^m - x^n\| \leq d_{mn}$

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



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

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

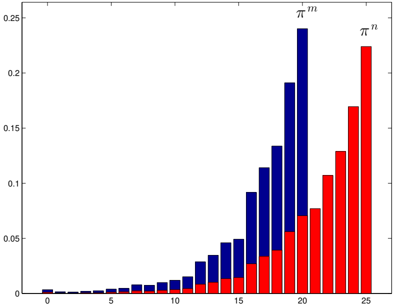


Let Z^{mn} be the set of transport plans $z \geq 0$ taking π^m to π^n

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

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

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

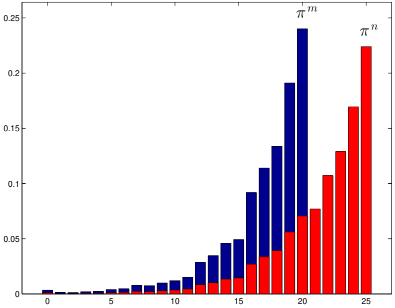


Let Z^{mn} be the set of transport plans $z \ge 0$ taking π^m to π^n

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

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

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

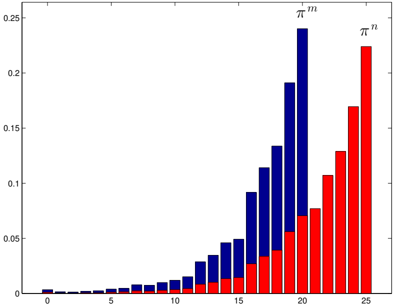


Let Z^{mn} be the set of transport plans $z \geq 0$ taking π^m to π^n

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

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

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

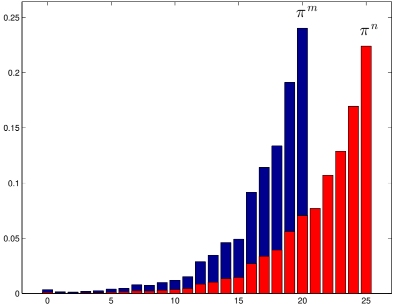


Let Z^{mn} be the set of transport plans $z \geq 0$ taking π^m to π^n

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

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

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

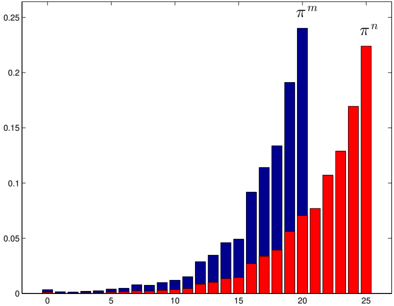


Let Z^{mn} be the set of transport plans $z \geq 0$ taking π^m to π^n

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

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

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



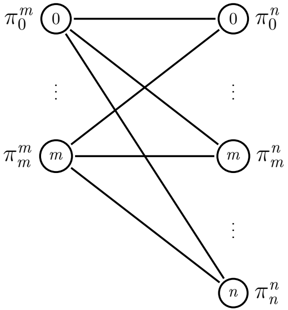
Let Z^{mn} be the set of transport plans $z \ge 0$ taking π^m to π^n

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

Recursive optimal transports

Suppose w.l.o.g. diam(C)=1. Set $d_{-1,n} = 1$ and define d_{mn} inductively as

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



Property

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

The bounds d_{mn} are:

- 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 $\mathrm{diam}(C) = 1$. Then the KM iterates satisfy $\|x^m - x^n\| \leq d_{mn}$ for all $m, n \in \mathbb{N}$.

The bounds d_{mn} are:

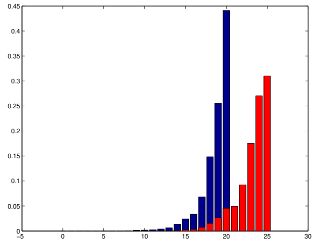
- Independent of the space X
- Independent of the operator $T : C \to C$

Theorem (Bravo-C 2016)

These bounds are uniformly tight: 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 $m, n \in \mathbb{N}$.

$\Rightarrow$ The d_{mn}'s are the right object to get sharp convergence rates !

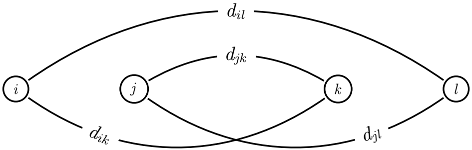
Metric properties of optimal transport bounds



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

Theorem (Aygen-Satik'1994, Bravo-C.'2016, Bravo-Champion-C.'2018)

The d_{mn}'s define a metric in $\mathbb{N}$ and $d_{il}+d_{jk} \leq d_{ik}+d_{jl}$ for all $i<j<k<l$.

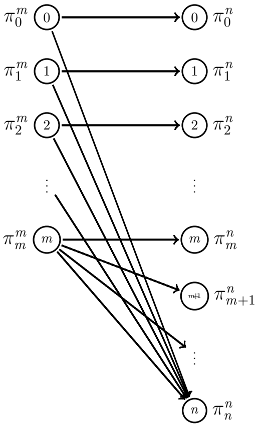


No flow-crossing in optimal transport

$\Downarrow$

Greedy algorithm

Example: $\alpha_n \geq \frac{1}{2}$



Optimal transport

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

This yields the *"simpler"* explicit recursion

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

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

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

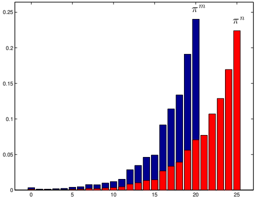
$\boxed{\text{...?????????}}$

Outline

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

Consider the sub-optimal transport plan

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



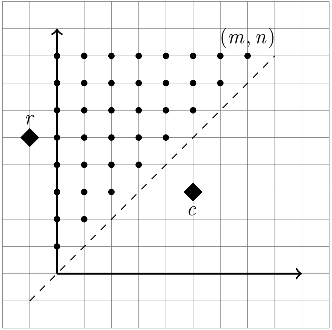
Setting $c_{-1,n} = 1$ for all $n \in \mathbb{N}$ we get inductively

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

Two Markov chains

Both c_{mn} and d_{mn} can be interpreted as the absorption probability of two Markov chains with state space $\mathcal{S} = \{(m,n) : 0 \leq m < n\} \cup \{r, c\}$.

$$\text{transition probabilities} \left\{ \begin{array}{ll} Q_{ij}^{mn} = \hat{z}_{ij}^{mn} & \text{(sub-optimal transports for } c_{mn}\text{)} \\ P_{ij}^{mn} = z_{ij}^{mn} & \text{(optimal transports for } d_{mn}\text{)} \end{array} \right.$$

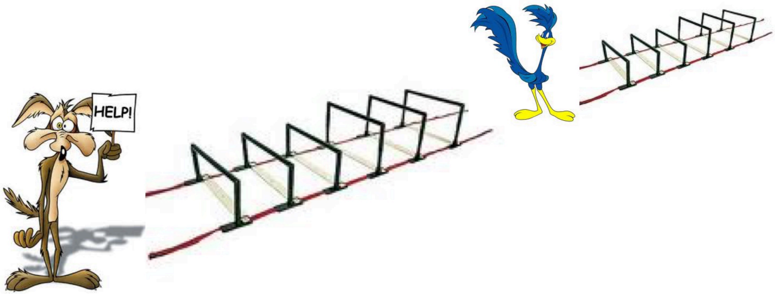


$$\begin{array}{rcl} c_{mn} & = & \mathbb{P}_Q(\text{absorbed at } r|mn) \\ d_{mn} & = & \mathbb{P}_P(\text{absorbed at } r|mn) \end{array}$$

Probabilistic interpretation of c_{mn}

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

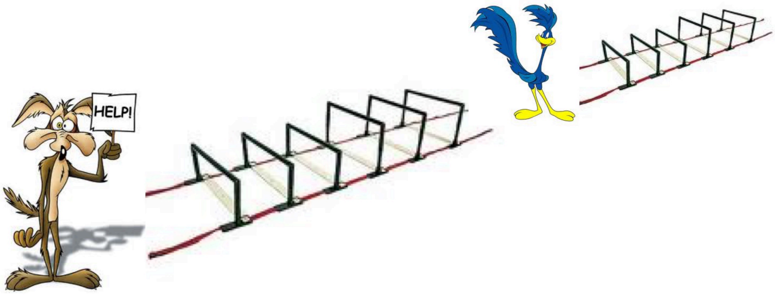
$\pi_j^n = \alpha_j \prod_{k=j+1}^{n}(1-\alpha_k)$



Probabilistic interpretation of c_{mn}

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

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

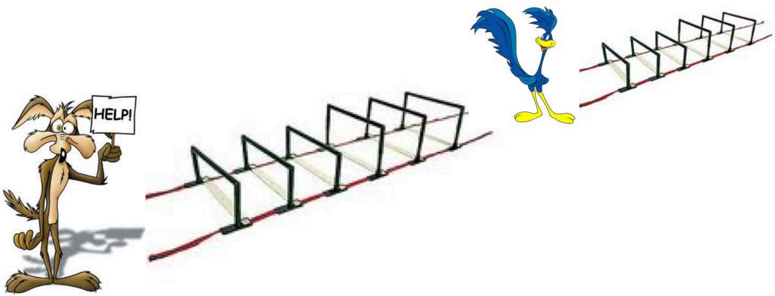


$$\tilde{c}_{mn} = \mathbb{P}[\text{roadrunner escapes}]$$

Probabilistic interpretation of c_{mn}

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

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

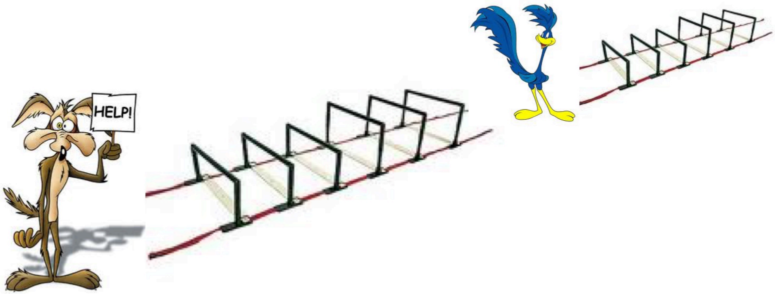


$$\tilde{c}_{mn} = \sum_{i=0}^{m} \sum_{j=m+1}^{n} \pi_i^m \pi_j^n \tilde{c}_{i-1,j-1}$$

Probabilistic interpretation of c_{mn}

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

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



$$c_{mn} = \mathbb{P}[\textstyle\sum_k^n C_i > \sum_k^m R_i, \forall k = m+1, \ldots, 1]$$

Coyote must fall more often than Roadrunner

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

Probabilistic arguments yield an explicit formula for the bound

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

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 the 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

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

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

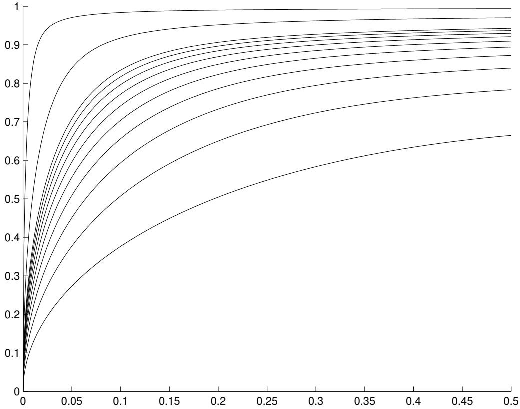
$$\underbrace{\sqrt{\tfrac{\pi}{2}(p_1 + \ldots + p_n)}\, \mathbb{E}[F(M_1 + \ldots + M_n)]}_{R_n(p)} \leq 1$$

Lemma

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

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

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



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.*)

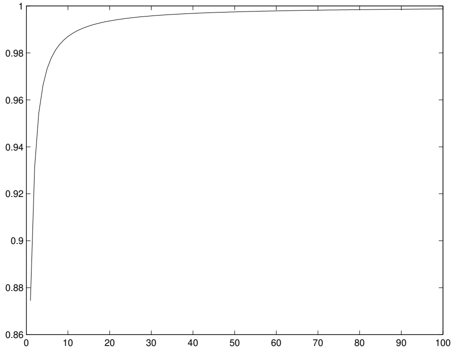
$\mathbb{E}[G(S)] \leq \mathbb{E}[G(Z)]$ *where* $Z \sim$ *Poisson*(z) *with* $z = \mathbb{E}(S)$*.*

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

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

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

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



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

Both cases combined yield

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

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

Outline

(BB) with constant $\alpha_n \equiv \alpha \geq \frac{1}{2}$

For the operator T on the cube $[0,1]^{\mathbb{N}}$ that attains the bounds d_{mn} we have

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

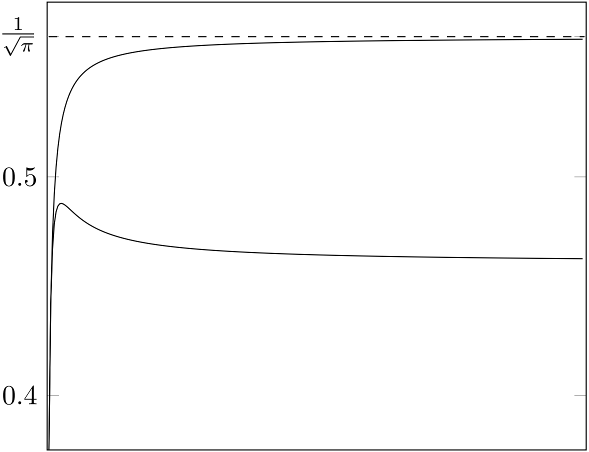
where

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

Explicit formula (C-Soto-Vaisman 2014)

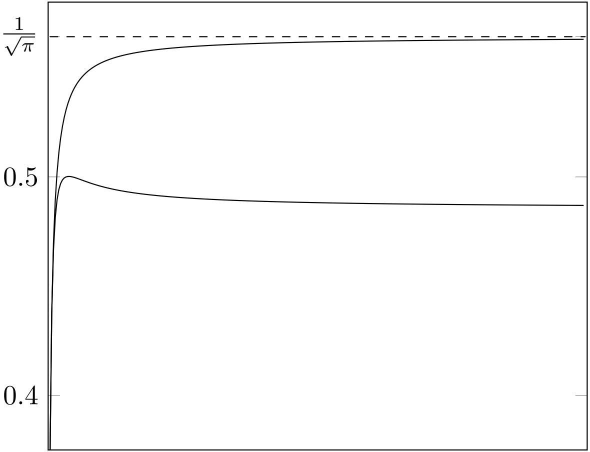
$$\tilde{\kappa}_n(\alpha) = \frac{1}{\pi}\int_0^{4n\alpha(1-\alpha)} \sqrt{\tfrac{1}{s} - \tfrac{1}{4n\alpha(1-\alpha)}}\,\left(1-\tfrac{s}{n}\right)^n ds \nearrow \frac{1}{\pi}\Gamma\left(\tfrac{1}{2}\right) = \frac{1}{\sqrt{\pi}}.$$

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



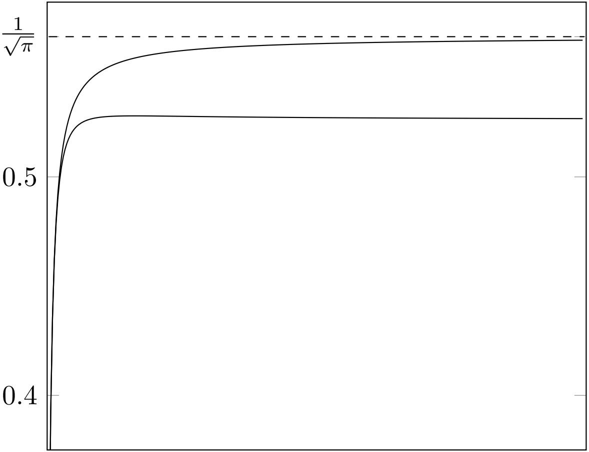
$\alpha = 0.5.$

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



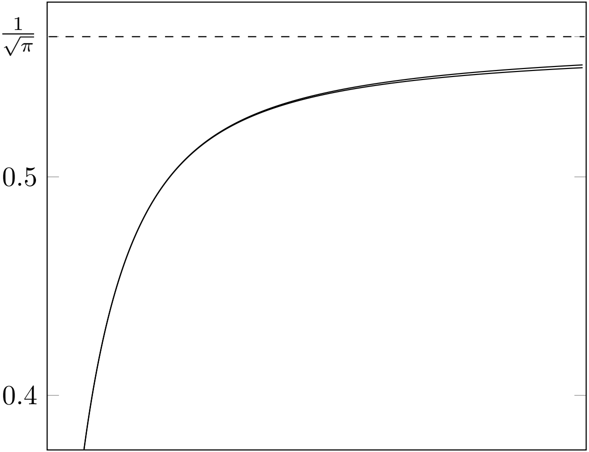
$\alpha = 0.65.$

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 $n = 1, \ldots, 300$



$\alpha = 0.99.$

Conclusion

Theorem (Bravo-C.'2016)

The constant $\kappa = \frac{1}{\sqrt{\pi}}$ is the best possible in (BB).

Proof. For constant $\alpha \geq \frac{1}{2}$ we have

$$0 \leq c_{n,n+1}(\alpha) - d_{n,n+1}(\alpha) \leq 4n(1-\alpha)^2$$

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

Inexact KM

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

Suppose $x_n \in C$ and $\|\,Tx_n - x_0\| \leq \kappa$ for some $\kappa \geq 0$ and all $\forall\; n \in \mathbb{N}$.
Let $\sigma(y) = \min\{1, 1/\sqrt{\pi y}\}$ and $\tau_n = \sum_{k=1}^{n} \alpha_k(1-\alpha_k)$. Then

$$\|x_n - Tx_n\| \leq \kappa\,\sigma(\tau_n) + \sum_{i=1}^{n} 2\,\alpha_i \|\varepsilon^i\|\,\sigma(\tau_n - \tau_i) + 2\,\|\varepsilon^{n+1}\|.$$

Inexact KM

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

Suppose $x_n \in C$ and $\|Tx_n - x_0\| \leq \kappa$ for some $\kappa \geq 0$ and all $\forall\, n \in \mathbb{N}$.
Let $\sigma(y) = \min\{1, 1/\sqrt{\pi y}\}$ and $\tau_n = \sum_{k=1}^n \alpha_k(1-\alpha_k)$. Then

$$\|x_n - Tx_n\| \leq \kappa\, \sigma(\tau_n) + \sum_{i=1}^{n} 2\,\alpha_i \|\varepsilon^i\|\, \sigma(\tau_n - \tau_i) + 2\,\|\varepsilon^{n+1}\|.$$

Corollary

Suppose $\|\varepsilon^n\| = O(1/n^a)$ *with* $a \geq \frac{1}{2}$, *and* α_n *bounded away from* 0 *and* 1.

a) *If* $\frac{1}{2} \leq a < 1$ *then* $\|x_n - Tx_n\| = O(1/n^{a-1/2})$.

b) *If* $a = 1$ *then* $\|x_n - Tx_n\| = O(\log n/\sqrt{n})$.

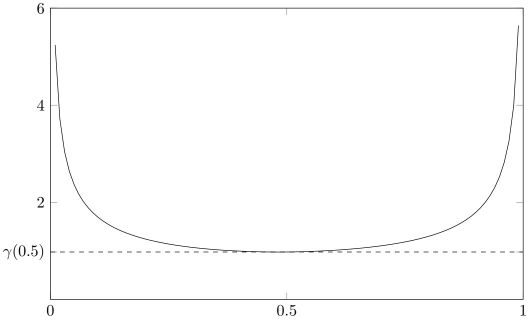
c) *If* $a > 1$ *then* $\|x_n - Tx_n\| = O(1/\sqrt{n})$.

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 of $\gamma(\alpha)$ is attained near $\alpha = 0.4623$

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

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

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

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

Which is the best α if n is fixed, say $n = 150$?

Sharp rates in Hilbert spaces?

- For $\alpha_n \equiv \alpha$, Browder and Petryshin'66 proved that

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

- Since $\|x^{n+1} - x^n\|$ is decreasing, this readily gives a faster rate (already observed in Baillon-Bruck'92, if not earlier)

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

- Which is the exact rate? $O(1/\sqrt{n \log n})$?

Papers available at

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