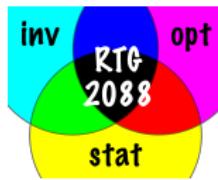


# ITERATED SELF-MAPPINGS IN NONLINEAR SPACES: THE CASE OF RANDOM FUNCTION ITERATIONS AND INCONSISTENT STOCHASTIC FEASIBILITY

Russell Luke

Universität Göttingen

One World Optimization Seminar  
Earth, May 4, 2020



# Outline

## 1 Announcements

## 2 Motivation

## 3 Case Study: Random Function Iterations

## 4 Quantitative Convergence Analysis: Fundamentals

- Almost  $\alpha$ -Firmly Nonexpansive Mappings in Metric Spaces
- Metric Subregularity and Rates

## 5 Examples and New-Old Results

## 6 References

# Open Positions at Göttingen



- PhD: Mathematical/Continuous Optimization (Luke)

# New Open Access/Electronic Journal

The screenshot shows a web browser window with the following details:

- Address Bar:** https://ojmo-test.centre-mersenne.org/page/editorial-team\_en/
- Title Bar:** Open Journal of Mathematical Optimization
- Header:** Open Journal of Mathematical Optimization (OJMO)
- Navigation Bar:** About the journal, Author guidelines, Editorial Team, Submit a paper
- Search Bar:** Search articles, authors...
- Content Area:**
  - EDITORIAL TEAM**
  - Steering Committee**
    - Dimitris Bertsimas [MIT, USA]
    - Martine Labb   [U.B, Belgium]
    - Eva K. Lee [Georgia Tech, USA]
    - Marc Teboulle [Tel-Aviv University, Israel]
  - Section Editors**
    - Continuous Optimization - David Russell Luke [Universit  t G  ttingen, Germany]
    - Discrete Optimization - Sebastian Rouletta [Georgia Tech, USA]
    - Optimization under Uncertainty - G黷tin Baykalson [Ohio State University, USA]
    - Computational aspects and applications - Bernard Gendron [Montr  al University, Canada]
  - Managing Editor**
    - Michael Poss [CNRS - LRMN, Montpellier, France]
  - Area Editors**
    - J  r  me Bolte [Toulouse School of Economics, France]
    - Christoph Buchheim [TU Dortmund, Germany]
    - C  d Cavalcini de Souza [University of Campinas, Brazil]
    - Fran  ois Clautiaux [Bordeaux University, France]
    - Antonio Frangioni [Pisa University, Italy]
    - Angelos Georgiou [McGill University, Canada]
    - Marc Goerigk [Siegen University, Germany]
    - Miguel Lejeune [The George Washington University, USA]
    - Adam Letchford [Jامعة Lancaster, UK]
    - J  r  me Malick [CNRS - LJK, Grenoble, France]
    - Sara Mattia [ASI-OR, Roma, Italy]
    - Cicomo Nannicini [IBM Watson, USA]
    - Jeremy Ormer [Institut Polytechnique de Paris, France]
    - Shoham Sabach [Technion, Israel]
    - Angelika Wiegele [Alpen-Adria-Universit  t Klagenfurt, Austria]



Published by  
Soci  t  e de Math  matiques Appliqu  es et Industrielles

ROADEF



Developed by  
Methodes



for open scientific publishing

# Outline

1 Announcements

2 Motivation

3 Case Study: Random Function Iterations

4 Quantitative Convergence Analysis: Fundamentals

- Almost  $\alpha$ -Firmly Nonexpansive Mappings in Metric Spaces
- Metric Subregularity and Rates

5 Examples and New-Old Results

6 References

# Motivation

$(H, d)$  is a **complete metric space** and  $\mathcal{T} : H \rightarrow H$ .

Find  $\bar{x} \in \text{Fix } \mathcal{T}$ .

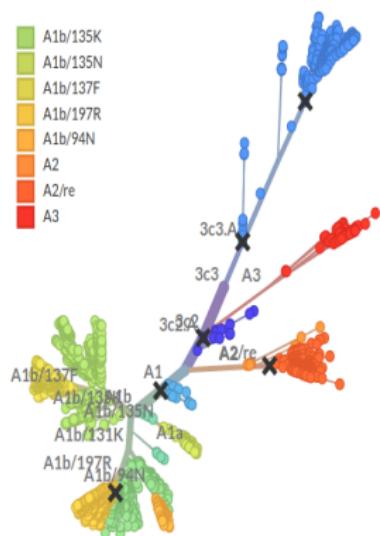
# Motivating Applications

- Computing mean trees:

Phylogeny

Clade ▾

3c2
3c2A
3c3
3c3A
A1
A1a
A1b
A1b/131K



Given trees  $T_1, T_2, \dots, T_M \in H_{BHV}$

$$\underset{T \in H_{BHV}}{\text{minimize}} F(T) := \frac{1}{M} \sum d^2(T, T_j)$$

(Owen [2011], Sturm [2009], Baçak [2014])

(<https://nextstrain.org/flu/>)

(joint work with Arian Berdellima and Florian Lauster)

# Motivating Applications

- $\infty$ -dimensional operator equations (Butnariu and Flåm [1995])
- distributed optimization (e.g. stochastic gradient descent)
- Markov chain Monte Carlo methods (Diaconis [2007, 2008])

## Random Function Iterations

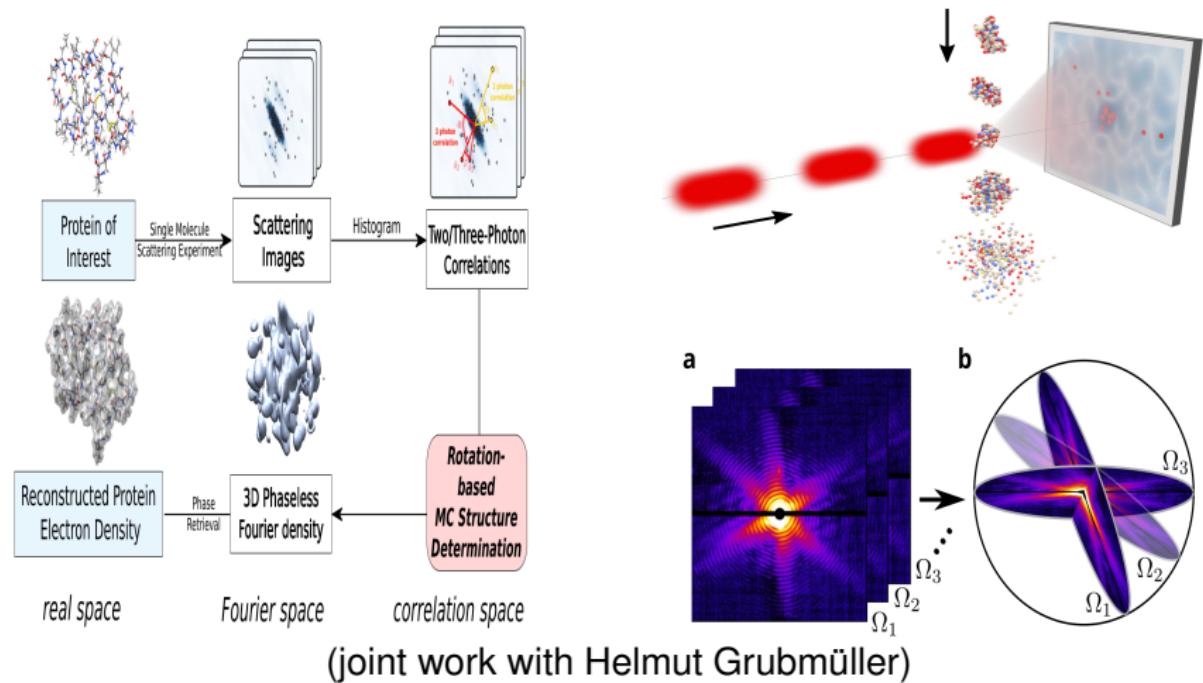
Given  $X^0 \sim \mu$ , generate

$$X^{k+1} = F_{\xi_k} X^k \quad (k = 0, 1, \dots)$$

(joint work with Anja Sturm and Neal Hermer [2019, 202?])

# Motivating Applications

- Single-shot Femtosecond X-ray Imaging



# Motivating Applications

## Linear Systems of Equations

Given  $A \in \mathbb{R}^{m \times n}$  and  $b_j \in \mathbb{R}^m$  solve

$$Ax = b$$

$$\iff$$

$$x \in \bigcap_{j=1,2,\dots,m} L_j$$

$$\text{where } L_j = \{y \mid \langle a_j, y \rangle = b_j\}$$

## Cyclic Projections

$$x^{k+1} = P_{L_1} P_{L_2} \cdots P_{L_m} x^k$$

# Motivating Applications

## Linear Systems of Equations

Given  $A \in \mathbb{R}^{m \times n}$  and  $b_j \in \mathbb{R}^m$  solve

$$Ax = b \iff x \in \bigcap_{j=1,2,\dots,m} L_j$$

where  $L_j = \{y \mid \langle a_j, y \rangle = b_j\}$

## Cyclic Projections

$$x^{k+1} = P_{L_1} P_{L_2} \cdots P_{L_m} x^k$$

# Motivating Applications

## Linear Systems of Equations

Given  $A \in \mathbb{R}^{m \times n}$  and  $b_j \in \mathbb{R}^m$  solve

$$Ax = b \iff x \in \bigcap_{j=1,2,\dots,m} L_j$$

where  $L_j = \{y \mid \langle a_j, y \rangle = b_j\}$

## Cyclic Projections

$$x^{k+1} = P_{L_1} P_{L_2} \cdots P_{L_m} x^k$$

# Motivating Applications

## Linear Systems of Equations

Given  $A \in \mathbb{R}^{m \times n}$  and  $b_j \in \mathbb{R}^m$  find

$$x \in L := \bigcap_{j=1,2,\dots,m} L_j \neq \emptyset \implies \text{Cyclic Projections:}$$

$$\text{where } L_j = \{y \mid \langle a_j, y \rangle = b_j\} \quad x^{k+1} = P_{L_1} P_{L_2} \cdots P_{L_m} x^k$$

## Convergence Theory

If  $A$  is full rank then cyclic projections converges either finitely or linearly to some  $\bar{x} \in L$  from any starting point.

# Motivating Applications

## Linear Systems of Equations

Given  $A \in \mathbb{R}^{m \times n}$  and  $b_j \in \mathbb{R}^m$  find

$$x \in L := \bigcap_{j=1,2,\dots,m} L_j \neq \emptyset \implies \text{Cyclic Projections:}$$

$$\text{where } L_j = \{y \mid \langle a_j, y \rangle = b_j\} \quad x^{k+1} = P_{L_1} P_{L_2} \cdots P_{L_m} x^k$$

## Convergence Theory

If  $A$  is full rank then cyclic projections converges either finitely or linearly to some  $\bar{x} \in L$  from any starting point.

# Motivating Applications

## Linear Systems of Equations

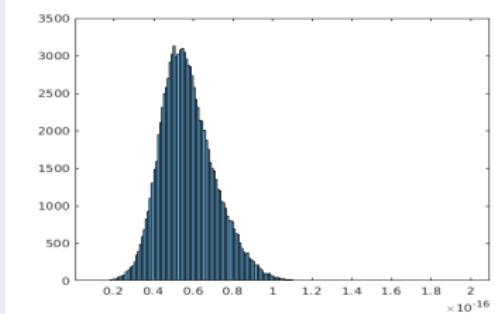
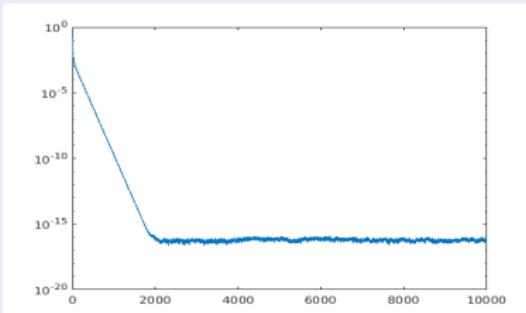
Given  $A \in \mathbb{R}^{m \times n}$  and  $b_j \in \mathbb{R}^m$  find

$$x \in L := \bigcap_{j=1,2,\dots,m} L_j \implies \text{Cyclic Projections:}$$

$$\text{where } L_j = \{y \mid \langle a_j, y \rangle = b_j\}$$

$$x^{k+1} = P_{L_1} P_{L_2} \cdots P_{L_m} x^k$$

$m = 50, n = 60$



# Motivating Applications

## Linear Systems of Equations

Given  $A \in \mathbb{R}^{m \times n}$  and  $b_j \in \mathbb{R}^m$  find

$$x \in L := \bigcap_{j=1,2,\dots,m} L_j \implies \text{Cyclic Projections with errors:}$$

$$\text{where } L_j = \{y \mid \langle a_j, y \rangle = b_j\} \quad x^{k+1} = P_{L_1} P_{L_2} \cdots P_{L_m} x^k + \epsilon_k$$

## Convergence Theory

If  $A$  is full rank then cyclic projections with vanishing errors converges.

# Motivating Applications

## Linear Systems of Equations

Given  $A \in \mathbb{R}^{m \times n}$  and  $b_j \in \mathbb{R}^m$  find

$$x \in L := \bigcap_{j=1,2,\dots,m} L_j \implies \text{Cyclic Projections with errors:}$$

$$\text{where } L_j = \{y \mid \langle a_j, y \rangle = b_j\} \quad x^{k+1} = P_{L_1} P_{L_2} \cdots P_{L_m} x^k + \epsilon_k$$

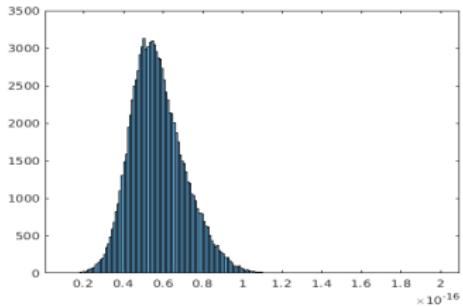
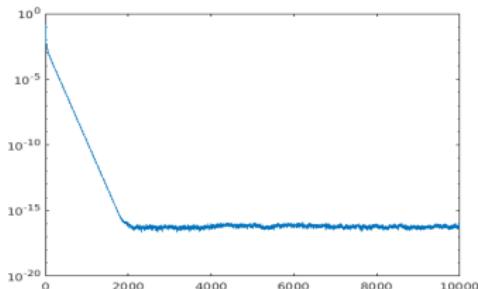
## Convergence Theory

If  $A$  is full rank then cyclic projections with vanishing errors converges.

# Motivating Applications

Does not explain:

$$m = 50, n = 60$$



# Outline

1 Announcements

2 Motivation

3 Case Study: Random Function Iterations

4 Quantitative Convergence Analysis: Fundamentals

- Almost  $\alpha$ -Firmly Nonexpansive Mappings in Metric Spaces
- Metric Subregularity and Rates

5 Examples and New-Old Results

6 References

Let  $F_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $i \in I$ , where  $I$  is an arbitrary index set,  $\xi$  is an  $I$ -valued random variable.

---

### Algorithm 1 RFI: Random Function Iterations

---

**Initialization:**  $X^0 \sim \mu$

**for**  $k = 0, 1, 2, \dots$  **do**

$X^{k+1} = F_{\xi_k} X^k$

**return**  $\{X^k\}_{k \in \mathbb{N}}$

---

### Markov Chain

A sequence of random variables  $(X^k)_{k \in \mathbb{N}_0}$ ,  $X^k : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$  is called Markov chain with transition kernel  $p$  if for all  $k \in \mathbb{N}_0$  and  $A \in \mathcal{B}(\mathbb{R}^n)$   $\mathbb{P}$ -a.s. the following hold:

- (I)  $\mathbb{P}(X^{k+1} \in A | X^0, X^1, \dots, X^k) = \mathbb{P}(X^{k+1} \in A | X^k);$
- (II)  $\mathbb{P}(X^{k+1} \in A | X^k) = p(X^k, A).$

# RFI is a Markov Chain

## Markov Chain

A sequence of random variables  $(X^k)_{k \in \mathbb{N}_0}$ ,  $X^k : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (G, \mathcal{B}(\mathbb{R}^n))$  is called Markov chain with transition kernel  $p$  if for all  $k \in \mathbb{N}_0$  and  $A \in \mathcal{B}(\mathbb{R}^n)$   $\mathbb{P}$ -a.s. the following hold:

- (i)  $\mathbb{P}(X^{k+1} \in A | X^0, X^1, \dots, X^k) = \mathbb{P}(X^{k+1} \in A | X^k);$
- (ii)  $\mathbb{P}(X^{k+1} \in A | X^k) = p(X^k, A).$

## RFI as a Markov Chain

$X^{k+1} = F_{\xi_k} X^k := \Phi(X^k, \xi_k)$  ( $k \in \mathbb{N}_0$ ) as a time homogeneous Markov chain with transition kernel  $p$ :

$$(x \in \mathbb{R}^n)(A \in \mathcal{B}(\mathbb{R}^n)) \quad p(x, A) := \mathbb{P}(\Phi(x, \xi) \in A) = \mathbb{P}(F_\xi x \in A)$$

# Example: Deterministic optimization

$$\underset{x}{\text{minimize}} \quad f_1(x) + f_2(x)$$

$X^0 \sim \delta(x^0)$ ,  $I = \{1\}$ , then  $T_\xi = T_1$  for  $\xi \sim \delta(1)$  and

$$X^{k+1} = T_{\xi_k} X^k \quad \text{where } T_1 = \begin{cases} \text{prox}_{f_1, \alpha_1} \text{ prox}_{f_2, \alpha_2} & \text{alt prox} \\ \frac{1}{2} (R_{f_1, \alpha_1} R_{f_2, \alpha_2} + \text{Id}) & \text{Douglas-Rachford} \\ \text{prox}_{f_1, \alpha} (\text{Id} - \frac{1}{\alpha} \nabla f_2) & \text{forward-backward} \\ \vdots & \end{cases}$$

## Example: Randomized algorithms

$$\begin{array}{ll} \text{minimize}_{(x,y) \in \mathbb{R}^n \times \mathbb{R}^m} & f_0(x) + \sum_{j=1}^M f_j(y_j) \\ \text{subject to} & x = -\mathcal{A}^T y \end{array} . \quad (\tilde{\mathcal{D}})$$

where  $\mathcal{A}^T = [A_1^T, A_2^T, \dots, A_M^T]$

---

**Algorithm 2** Random Block-coordinate Primal-Dual Algorithm (L. Malitsky 2018)

---

**Initialization:** Choose  $\tau > 0$ ,  $\sigma = (\sigma_0, \sigma_1, \sigma_2, \dots, \sigma_M) \in \mathbb{R}_{++}^M$ . Choose  $y^1 = 0 \in \mathbb{R}^m$ , and set  $x^1 = u^1 = -\mathcal{A}^T y^1 = 0 \in \mathbb{R}^n$ .

**for**  $k = 1, 2, \dots$  **do**

Choose  $i \in \{0, 1, 2, \dots, M\}$  uniformly at random

Update:

$$y_i^{k+1} = \text{prox}_{\sigma_i f_i}(y_i^k - \sigma_i A_i x^k);$$

$$\delta^{k+1} = A_i^T (y_i^{k+1} - y_i^k);$$

$$u^{k+1} = u^k + \frac{\tau}{M} \delta^{k+1};$$

$$x^{k+1} = u^{k+1} + x^k + \tau \delta^{k+1};$$

**return**  $((x^k, y^k))_{k \in \mathbb{N}}$

# Markov Operators and the Stochastic Fixed Point Problem

## Markov operator $\mathcal{T}$

- (dual) The Markov operator  $\mathcal{T} : \mathcal{P}(\mathbb{R}^n) \rightarrow \mathcal{P}(\mathbb{R}^n)$  acting on a measure  $\mu \in \mathcal{P}(\mathbb{R}^n)$  is defined via

$$(A \in \mathcal{B}(\mathbb{R}^n)) \quad \mathcal{T}(A)\mu := \int_{\mathbb{R}^n} p(x, A)\mu(dx).$$

## Stochastic Fixed Point Problem

Given a Markov operator  $\mathcal{T}$ ,

Find  $\bar{\mu} \in \text{inv } \mathcal{T} := \{\mu \mid \mathcal{T}\mu = \mu\} (= \text{Fix } \mathcal{T})$

# Outline

1 Announcements

2 Motivation

3 Case Study: Random Function Iterations

## 4 Quantitative Convergence Analysis: Fundamentals

- Almost  $\alpha$ -Firmly Nonexpansive Mappings in Metric Spaces
- Metric Subregularity and Rates

5 Examples and New-Old Results

6 References

# Basic Quantitative Convergence

## Linear Convergence of Fixed Point Iterations in Nonlinear Spaces

$(H, d)$  is a **separable complete metric space** and  $\mathcal{T} : H \rightarrow H$ . For  $\mu \in H$  define  $\Psi$  to be a **reasonable surrogate function**. Suppose that  $\text{Fix } \mathcal{T} \neq \emptyset$  and

- (a) for a fixed  $\alpha \in (0, 1)$  and all  $\epsilon > 0$ , there is a  $\delta > 0$  such that  $\mathcal{T}$  is **pointwise almost  $\alpha$ -firmly nonexpansive** at all points in  $S \subset \text{Fix } \mathcal{T}$  with constant  $\alpha \in (0, 1)$  and violation  $\epsilon$  on  $S_\delta := \bigcup_{\bar{\mu} \in \text{Fix } \mathcal{T}} \{\mu \mid d(\mu, \bar{\mu}) \leq \delta\}$ ,
- (b)  $\Psi$  is **metrically subregular** for 0 on  $S_\delta \setminus \text{Fix } \mathcal{T}$  with constant  $\kappa$ .

Then, for any  $\mu^0$  close enough to  $\text{Fix } \mathcal{T}$ , the iterates  $\mu^{k+1} = \mathcal{T}\mu^k$  converge R-linearly in the metric  $d$  to some  $\bar{\mu} \in \text{Fix } \mathcal{T}$  with

$$d(\mu^{k+1}, \text{Fix } \mathcal{T}) \leq c d(\mu^k, \text{Fix } \mathcal{T}) \quad \forall k \in \mathbb{N},$$

where  $c := \sqrt{1 + \epsilon - (\frac{1-\alpha}{\kappa^2 \alpha})} < 1$ .

### Proof:

- 1 Apply the definition of metric subregularity.
- 2 Apply the definition of almost  $\alpha$ -firm nonexpansiveness
- 3 Rearrange the inequality.

# Recent Deterministic Results in Euclidean Spaces

## Convergence of Prox algorithms

Cyclic projections converges linearly to Fix  $T_{CP}$  in the following instances:

- $\Omega_j$  convex, empty intersection (global) [L. Tam, Thao, 2018]  
[L. Teboulle, Thao 2020]
- Sparse affine feasibility -  $\Omega_1$  affine,  $\Omega_2$  a sparsity set (global)  
[Hesse, L., Neumann 2014]
- $\Omega_j$  nonintersecting rings (local)  
[L. Tam, Thao, 2018]
- real world phase retrieval (local) [L. Tam, Thao, 2018]
- Douglas-Rachford converges linearly to Fix  $T_{DR}$  for consistent feasibility of reasonable sets with nice intersection [Hesse, L. 2013; Phan 2016]
- relaxed Douglas-Rachford converges linearly to Fix  $T_{DR\lambda}$  for inconsistent feasibility for nontangential circles and cones [L., Martins 2020]
- ADMM dual sequence converges Q-linearly for piecewise linear-quadratic objectives and linear equality constraints [Aspelmeier, L., Charitha 2016].

# Almost $\alpha$ -firmly Nonexpansive Mappings in Metric Spaces

[Berdellima, Lauster, L. (2020), L. Tam Thao (2018)]

Let  $D \subset H$  nonempty closed, and let  $T : D \rightarrow H$ .

The mapping  $T$  is said to be **pointwise almost  $\alpha$ -firmly nonexpansive at  $x_0 \in D$  on  $D$**  whenever there exists a constant  $\alpha \in (0, 1)$  such that

$$d^2(Tx, Tx_0) + (1 - (2 + \epsilon)\alpha)d^2(x, x_0) \leq 2(1 - \alpha)\Delta_T(x, x_0) \quad \forall x \in D$$

where

$$\Delta_T(x, x_0) := d^2(Tx, x_0) + d^2(x, Tx_0) - d^2(Tx, x) - d^2(Tx_0, x_0)$$

When the above inequality holds for all  $x_0 \in D$  then  $T$  is said to be **almost  $\alpha$ -firmly nonexpansive on  $D$** . When  $\epsilon = 0$ , drop “almost”.

This is a generalization of **averaged operators** introduced and studied by Krasnoselsky [1955] and Mann [1953], Edelstein [1966], Gurin, Polyak&Raik [1967], Nussbaum [1972], Baillon, Bruck& Reich [1978]. See Bauschke& Combettes [2017].

# $\alpha$ -firmly Nonexpansive Mappings in Hilbert Spaces

$\mathcal{H}$  is a Hilbert space

Let  $D \subset \mathcal{H}$  nonempty closed, and let  $T : D \rightarrow \mathcal{H}$ .

The mapping  $T$  is  $\alpha$ -firmly nonexpansive on  $D$  whenever there exists a constant  $\alpha \in (0, 1)$  such that

$$\|Tx - Ty\|^2 + (1 - 2\alpha)\|x - x_0\|^2 \leq 2(1 - \alpha)\Delta_T(x, x_0) \quad \forall x, y \in D$$

$\iff$

$$\|Tx - Ty\|^2 \leq \|x - x_0\|^2 - \frac{1-\alpha}{\alpha}\psi_T(x, x_0) \quad \forall x, y \in D$$

where

$$\Delta_T(x, x_0) := \langle x - y, Tx - Ty \rangle$$

$$\begin{aligned} \psi_T(x, y) &:= d^2(Tx, x) + d^2(Ty, y) + d^2(Tx, Ty) + d^2(x, y) \\ &\quad - d^2(Tx, y) - d^2(x, Ty) \end{aligned}$$

$$= \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.$$

$\alpha$ -firm nonexpansive mappings are nonexpansive.

# Almost $\alpha$ -firmly Nonexpansive Mappings

## Examples

- projectors onto convex sets (Hadamard spaces):  $\alpha$ -firmly nonexpansive with  $\alpha = 1/2$ ;
- projectors onto sets that are **super-regular at a distance** (Euclidean spaces [L. Martins, 2020]): almost  $\alpha$ -firmly nonexpansive for each  $\epsilon > 0$  on neighborhoods.
- steepest descent directions for sufficiently smooth functions (Euclidean): almost  $\alpha$ -firmly nonexpansive for each  $\epsilon > 0$  on neighborhoods.

# Calculus of Almost $\alpha$ -firmly Nonexpansive Mappings: consistent

Let

- $(H, d)$  be a Hadamard space;
- $T_i : H \rightarrow H$  be pointwise  $\alpha$ -firmly nonexpansive on  $H$  at all points in Fix  $T_i$  with constant  $\alpha_i \in [0, 1)$  ( $i = 1, 2$ );
- Fix  $T_1 \cap \text{Fix } T_2 \neq \emptyset$  (consistent)

## Compositions [Berdellima, Lauster, L. (2020)]

Define  $\mathcal{T} := T_1 \circ T_2$ . Then  $\mathcal{T}$  is pointwise  $\alpha$ -firmly nonexpansive on  $H$  at all points in Fix  $\mathcal{T} \neq \emptyset$  with constant

$$\gamma = \frac{\alpha_1 + \alpha_2 - 2\alpha_1\alpha_2}{1 - \alpha_1\alpha_2}$$

## Convex Combinations [Berdellima, Lauster, L. (2020)]

For  $\lambda \in [0, 1]$  the operator  $\mathcal{T} = \lambda T_1 + (1 - \lambda)T_2$  is pointwise  $\alpha$ -firm on  $H$  at all points in Fix  $T_1 \cap \text{Fix } T_2$  with constant  $\gamma = \max\{\alpha_1, \alpha_2\}$ .

# Calculus of Almost $\alpha$ -firmly Nonexpansive Mappings: inconsistent stochastic feasibility

Summary of key objects

- $F_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$  for  $i \in I$ ;  $\mathbf{F} : I \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ :  $\mathbf{F}(i, x) := F_i(x)$
- $\nu$  a probability measure on  $(I, \mathcal{I})$ , and  $\xi \sim \nu$   $I$ -valued r.v.
- $X \sim \mu_x \in \mathcal{P}(\mathbb{R}^n)$  and  $Y \sim \mu_y \in \mathcal{P}(\mathbb{R}^n)$
- $X \perp\!\!\!\perp \xi$ ,  $Y \perp\!\!\!\perp \xi$  and  $X \perp\!\!\!\perp Y$ .
- $\mathcal{T}$  the Markov operator with transition kernel  $p(X, A) := \mathbb{P}(F_\xi X \in A)$
- $\text{inv } \mathcal{T}$  nonempty and closed
- 

$$d(\mu, \nu) := \left( \inf_{\gamma \in \Gamma(\mu, \nu)} \int_{\mathbb{R}^n \times \mathbb{R}^n} \|x - y\|^2 d\gamma(x, y) \right)^{1/2}$$

where  $\Gamma(\mu, \nu)$  is the set of couplings of  $\mu$  and  $\nu$ .

- 

$$\begin{aligned} \psi_{F_\xi}(x, y) := & d^2(F_\xi x, x) + d^2(F_\xi y, y) + d^2(F_\xi x, F_\xi y) + d^2(x, y) \\ & - d^2(F_\xi x, y) - d^2(x, F_\xi y) \end{aligned}$$

- $S_\delta := \bigcup_{\bar{\mu} \in \text{inv } \mathcal{T}} \{\mu \mid d(\mu, \bar{\mu}) \leq \delta\}$  for a nonnegative real  $\delta$ .

# Calculus of Almost $\alpha$ -firmly Nonexpansive Mappings: inconsistent stochastic feasibility

Let

- $F_i : \mathbb{R}^n \rightarrow \mathbb{R}^n$  be  $\alpha$ -firmly nonexpansive in expectation on  $\mathbb{R}^n$  with constant  $\alpha_i \leq \alpha \in [0, 1]$  ( $i \in I$ , possibly uncountable);
- $\mathcal{T} : \mathcal{P}(\mathbb{R}) \rightarrow \mathcal{P}(\mathbb{R})$  be the Markov operator corresponding to the random function iteration with  $F_\xi$  and the metric  $d$  the Wasserstein metric;
- $\text{inv } \mathcal{T} \neq \emptyset$  (but possibly  $\cap_{i \in I} F_i = \emptyset \rightarrow \text{inconsistent}$ )

$\alpha$ -firm Random Function Iterations  $\implies$   $\alpha$ -firm Markov Operators  
[Hermer, L., Sturm (2020)]

For any  $\bar{\mu} \in \text{inv } \mathcal{T}$ , the Markov operator  $\mathcal{T}$  satisfies

$$d^2(\mathcal{T}\mu, \bar{\mu}) \leq d^2(\mu, \bar{\mu}) - \frac{1-\alpha}{\alpha} \int_{G \times G} \mathbb{E}_\xi [\psi_\xi(x, y)] \gamma(dx, dy)$$
$$(\forall \bar{\mu} \in \text{inv } \mathcal{T})(\forall \mu \in \mathcal{P}(\mathbb{R}^n))$$

# $\alpha$ -firmly Nonexpansive Mappings in Hilbert Spaces

Compare the above to:

The mapping  $T$  is  $\alpha$ -firmly nonexpansive on  $D$  whenever there exists a constant  $\alpha \in (0, 1)$  such that

$$\|Tx - Ty\|^2 \leq \|x - x_0\|^2 - \frac{1-\alpha}{\alpha} \psi_T(x, x_0) \quad \forall x, y \in D$$

where

$$\begin{aligned}\psi_T(x, y) &:= d^2(Tx, x) + d^2(Ty, y) + d^2(Tx, Ty) + d^2(x, y) \\ &\quad - d^2(Tx, y) - d^2(x, Ty) \\ &\stackrel{\mathcal{H}}{=} \|(\text{Id} - T)x - (\text{Id} - T)y\|^2.\end{aligned}$$

$\alpha$ -firm nonexpansive mappings are nonexpansive.

# Regularity at Fixed Points

## Metric (Sub)regularity

Let  $(G, d_G)$  and  $(M, d_M)$  be metric spaces and let  $\Psi : G \rightrightarrows M$ ,  $D \subset G$ ,  $V \subset M$ . Let  $\mu$  be some gauge function. The mapping  $\Psi$  is called **metrically regular with gauge  $\mu$  on  $U \times V$  relative to  $\Lambda \subset G$**  if

$$\inf_{z \in \Psi^{-1}(y) \cap \Lambda} d_G(x, z) \leq \mu(d_M(y, \Psi(x)))$$

holds for all  $x \in D \cap \Lambda$  and  $y \in V$  with  $0 < \kappa d_M(y, \Psi(x))$  where  $\Psi^{-1}(y) := \{z \mid \Psi(z) = y\}$ . When the set  $V$  consists of a single point,  $V = \{\bar{y}\}$ , then  $\Psi$  is said to be **metrically subregular for  $\bar{y}$  on  $U$  with gauge  $\mu$  relative to  $\Lambda \subset G$** .

[Aze (06), Klatte&Kummer (09), Ioffe (11, 13), Ngai& Théra (04, 08)],  
Dontchev&Rockafellar (14), Kruger, L. & Thao (16)]

# Necessity of Metric Subregularity

A brief detour:

## Metric Subregularity is necessary for linear convergence

- Metric subregularity with linear gauge is necessary for linear convergence of fixed point iterations (L.Teboulle-Thao [2019]).
- The KL property and existence of error bounds are equivalent to metric subregularity with some gauge  $\mu$ . Lots of other equivalent notions (see Kruger-L.-Thao [2018]).
- Polyhedrality and isolated fixed points imply (linear) metric subregularity.

# Outline

1 Announcements

2 Motivation

3 Case Study: Random Function Iterations

4 Quantitative Convergence Analysis: Fundamentals

- Almost  $\alpha$ -Firmly Nonexpansive Mappings in Metric Spaces
- Metric Subregularity and Rates

5 Examples and New-Old Results

6 References

# Example: Randomized algorithms (distributed computing - nonseparable)

$$\underset{x}{\text{minimize}} \quad f(x) = \sum_{i \in I} f_i(x)$$

## Stochastic gradient descent

- $f_i : \mathbb{R}^n \rightarrow \mathbb{R}$  strongly convex with constant  $\mu_i \geq \mu > 0$ , bounded below and has with  $L$ -Lipschitz continuous gradient  $\nabla f_i \forall i \in I$ .
- $F := \mathbb{E}[f_\xi]$  has a minimum at  $\bar{x}$
- For  $T_\xi := \text{Id} - \lambda \nabla f_\xi$  with fixed step size  $\lambda \in (0, \frac{2\mu}{L^2})$ , given  $X^0 \sim \mu$  generate the sequence

$$X^{k+1} = T_{\xi_k} X^k \tag{1}$$

# Example: Randomized algorithms (distributed computing - nonseparable)

## Stochastic gradient descent

For  $T_\xi := \text{Id} - \lambda \nabla f_\xi$  with fixed step size  $\lambda \in (0, \frac{2\mu}{L^2})$ , given  $X^0 \sim \mu$  generate the sequence

$$X^{k+1} = T_{\xi_k} X^k \tag{2}$$

- There exists an invariant measure (actually, existence of the invariant measure holds for step sizes up to  $1/L$ )
- $T_i$  is  $\alpha$ -firmly nonexpansive in expectation with averaging constant  $\alpha = \lambda L^2 / 2\mu$ .
- For any  $\bar{\mu} \in \text{inv } \mathcal{T}$ , the Markov operator  $\mathcal{T}$  satisfies

$$\begin{aligned} d^2(\mathcal{T}\mu, \bar{\mu}) &\leq d^2(\mu, \bar{\mu}) - \frac{1-\alpha}{\alpha} \int_{G \times G} \mathbb{E}_\xi [\psi_\xi(x, y)] \gamma(dx, dy) \\ (\forall \bar{\mu} \in \text{inv } \mathcal{T})(\forall \mu \in \mathcal{P}(\mathbb{R}^n)) \end{aligned}$$

# Example: Randomized algorithms (distributed computing - nonseparable)

## Stochastic gradient descent

For  $T_\xi := \text{Id} - \lambda \nabla f_\xi$  with fixed step size  $\lambda \in (0, \frac{2\mu}{L^2})$ , given  $X^0 \sim \mu$  generate the sequence

$$X^{k+1} = T_{\xi_k} X^k \quad (3)$$

- For any  $\bar{\mu} \in \text{inv } \mathcal{T}$ , the Markov operator  $\mathcal{T}$  satisfies

$$d^2(\mathcal{T}\mu, \bar{\mu}) \leq d^2(\mu, \bar{\mu}) - \frac{1-\alpha}{\alpha} \int_{G \times G} \mathbb{E}_\xi [\psi_\xi(x, y)] \gamma(dx, dy) \\ (\forall \bar{\mu} \in \text{inv } \mathcal{T})(\forall \mu \in \mathcal{P}(\mathbb{R}^n))$$

- If

$$\Psi(\mu) := \left( \int_{G \times G} \mathbb{E}_\xi [\psi_\xi(x, y)] \gamma(dx, dy) \right)^{1/2}, \quad \gamma \in \mathcal{C}_*(\mu, \pi_\mu)$$

is metrically subregular (KL, etc), then convergence can be quantified.

# Example: inconsistent feasibility

Recall:

## Linear Systems of Equations

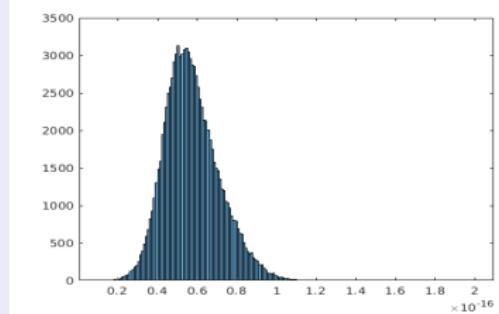
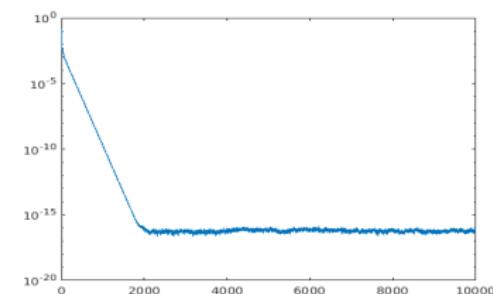
Given  $A \in \mathbb{R}^{m \times n}$  and  $b_j \in \mathbb{R}^m$  find

$$x \in L := \bigcap_{j=1,2,\dots,m} L_j \implies \text{Cyclic Projections:}$$

where  $L_j = \{y \mid \langle a_j, y \rangle = b_j\}$

$$x^{k+1} = P_{L_1} P_{L_2} \cdots P_{L_m} x^k$$

$m = 50, n = 60$



# Example: inconsistent feasibility

Careful how you model noise:

Let

$$P_{\xi_j}x = P_{L_j}x + \xi_j \quad (j = 1, 2, \dots, m)$$

Define

$$T_\xi := P_{\xi_1} \cdots \cdots P_{\xi_m} \quad \xi = (\xi_1, \dots, \xi_m).$$

Suppose the noise  $\xi_j$  is **isotropic**, in particular,

$$\mathbb{P}(\langle h, \xi_j \rangle > 0) = \beta > 0 \quad (j = 1, 2, \dots, m) \quad \forall h \in \bigcap_j L_j \text{ with } \|h\| = 1.$$

Then there does not exist an invariant measure for the Markov chain.

The projections are not exact with random error, but rather exact projections onto randomly selected sets.

# Example: inconsistent feasibility

Better noise model:

Let

$$L_j^{(\xi, \zeta)} := \{x \in \mathbb{R}^n \mid \langle a_j + \xi, x \rangle = b_j + \zeta\}$$

where  $(\xi, \zeta) \in \mathbb{R}^n \times \mathbb{R}$  is random and independent. Define

$$P_j^{(\xi, \zeta)} := x - \frac{\langle a_j + \xi, x \rangle - (b_j + \zeta)}{\|a_j + \xi\|^2} (a_j + \xi)$$

Let  $\xi$  be isotropic and  $\|\xi\| \leq \|a_j\|$  and  $\zeta$  have bounded variance.

Then

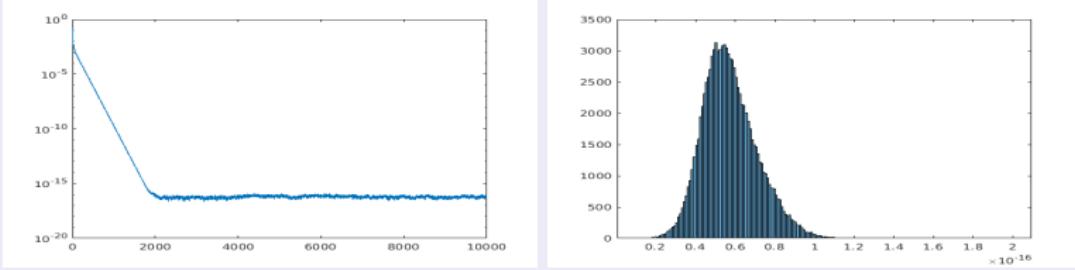
- $T_{(\xi, \zeta)} := P_1^{(\xi_1, \zeta_1)} \dots P_m^{(\xi_m, \zeta_m)}$     $\xi = (\xi_1, \dots, \xi_m)$ .
- the Markov operator possesses a unique invariant measure.
- the surrogate function  $\Psi(\mu)$  is (linearly) metrically subregular on  $\mathcal{P}(\mathbb{R}^n)$  for 0 with constant  $\kappa$
- the distributions  $\mu \mathcal{P}^k \rightarrow \pi$  Q-linearly with rate

$$c = \sqrt{1 - \frac{1-\alpha}{\kappa^2 \alpha}}$$

# Example: inconsistent feasibility

This seems to address

$$m = 50, n = 60$$



...assuming that the roundoff error can be modeled by this random process.

# Outline

1 Announcements

2 Motivation

3 Case Study: Random Function Iterations

4 Quantitative Convergence Analysis: Fundamentals

- Almost  $\alpha$ -Firmly Nonexpansive Mappings in Metric Spaces
- Metric Subregularity and Rates

5 Examples and New-Old Results

6 References

# References I

-  T. Aspelmeier, Charitha, D. R. Luke,  
Local Linear Convergence of the ADMM/Douglas–Rachford Algorithms without  
Strong Convexity and Application to Statistical Imaging  
*SIAM J. on Imaging Sciences* (2016)
-  N. Hermer, D. R. Luke, and A. K. Sturm,  
Random Function Iterations for Stochastic Fixed Point Problems  
(in preparation)
-  N. Hermer, D. R. Luke, and A. K. Sturm,  
Random Function Iterations for Consistent Stochastic Feasibility  
*Numerical Functional Analysis and Optimization* 40(4):386–420 (2019)  
DOI: 10.1080/01630563.2018.1535507.
-  Luke, D.R., Malitsky, Y.:  
Block-coordinate primal-dual method for the nonsmooth minimization over linear  
constraints.  
In: P. Giselsson, A. Rantzer (eds.) *Distributed and Large-Scale Optimization*.  
Springer Verlag (2018)

## References II

-  Luke, D.R., Martins, A.L.:  
Convergence Analysis of the Relaxed Douglas-Rachford Algorithm.  
*SIAM Journal on Optimization* (to appear).  
<https://arxiv.org/abs/1811.11590>
-  D. R. Luke, M. K. Tam and N. H. Thao,  
Quantitative convergence analysis of iterated expansive, set-valued mappings  
*Mathematics of Operations Research* 43(4):1143–1176 (2018).
-  Luke, D.R., Sabach, S., Teboulle, M.:  
Optimization on Spheres: Models and Proximal Algorithms with Computational Performance Comparisons.  
*SIAM Journal on Mathematical Data Science*, 1(3), 408–445 (2019).
-  Luke, D.R., Teboulle, M., Thao, N.H.:  
Necessary conditions for linear convergence of iterated expansive, set-valued mappings.  
*Mathematical Programming, A* (2020)180:1–31.  
<https://doi.org/10.1007/s10107-018-1343-8>