Exotic error bounds, Karamata theory and convergence rate analysis

Bruno F. Lourenço
Institute of Statistical Mathematics, Japan
Joint work with Tianxiang Liu (Tokyo Institute of Technology)

24th of July of 2024

Common Fixed Point Problems

$T_1, \ldots, T_m$: α-averaged operators ($\alpha \in (0,1)$)

$$\text{find } x \in F := \bigcap_{i=1}^{m} \mathrm{Fix}\, T_i, \tag{CFP}$$

A particular case is when $T_i = P_{C_i}$ so that

$$\text{find } x \in F := \bigcap_{i=1}^{m} C_i, \tag{CFP}$$

There are many methods for both problems.

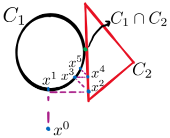


Figure: Cyclic projections

- Do these methods **converge**? **Typically yes, because of convexity**
- How **fast** do they converge? **Depends on the kind of regularity property that holds between operators**

Regularity properties

Hölderian error bound

C_1, C_2 satisfy a **uniform Hölderian error bound** $\overset{\text{def}}{\iff}$ there exists $\gamma \in (0, 1]$ such that for every bounded set B there exist $\theta_B > 0$

$$\operatorname{dist}(x, C_1 \cap C_2) \le \theta_B \max_{1 \le i \le 2} \operatorname{dist}^{\gamma}(x, C_i) \quad \forall\, x \in B.$$

If $\gamma = 1$, we call it a **Lipschitzian** error bound.

Hölder regularity

T is uniformly Hölder regular $\overset{\text{def}}{\iff}$ there exists $\gamma \in (0, 1]$ such that for every bounded set B there exist $\theta_B > 0$

$$\operatorname{dist}(x, \operatorname{Fix} T) \le \theta_B \|x - Tx\|^{\gamma} \quad \forall\, x \in B.$$

Lipschtizian (regularity + error bound) $\implies \operatorname{dist}(x^k, F) \le M\theta^k$ (**Linear** convergence)

Hölder (regularity + error bound) $\implies \operatorname{dist}(x^k, F) \le Mk^{-\alpha}$ (**Sublinear convergence**)

J. M. Borwein, G. Li, and M. K. Tam.
Convergence rate analysis for averaged fixed point iterations in common fixed point problems.
SIAM Journal on Optimization, 27(1):1–33, 2017.

The exponential cone

$$K_{\exp} := \left\{(x,y,z) \mid y > 0, z \geq ye^{x/y}\right\} \cup \{(x,y,z) \mid x \leq 0, z \geq 0, y = 0\}.$$

1. Applications to entropy optimization, logistic regression, geometric programming and etc.
2. Available in Alfonso, DDS, Hypatia, Mosek, SCS, ...
   `https://docs.mosek.com/modeling-cookbook/expo.html`.

V. Chandrasekaran, P. Shah
Relative entropy optimization and its applications.
Math. Program. 161, 2017

Scott B. Lindstrom; L and Ting Kei Pong
Error bounds, facial residual functions and applications to the exponential cone
Math. Program., 2023

Beyond Hölderian regularity - Exotic error bounds

1. If $C_1 = K_{\exp}$, $C_2 = \{(0,1,0)\}^{\perp}$, the error bound is of the form

$$\mathrm{dist}\,(x, C_1 \cap C_2) \le \kappa_B \mathfrak{g}_{-\infty}(\max_{1\le i\le 2}\{\mathrm{dist}\,(x, C_i)\})$$

where

$$\mathfrak{g}_{-\infty}(t) := -t\ln(t), \qquad \text{(for } t \text{ small)}$$

This is an **entropic error bound**.

2. If $C_1 = K_{\exp}$, $C_2 = \{(0,0,1)\}^{\perp}$, the error bound is of the form

$$\mathrm{dist}\,(x, C_1 \cap C_2) \le \kappa_B \mathfrak{g}_{\infty}(\max_{1\le i\le 2}\{\mathrm{dist}\,(x, C_i)\})$$

where

$$\mathfrak{g}_{\infty}(t) := -\frac{1}{\ln(t)}, \qquad \text{(for } t \text{ small)}$$

This is an **logarithmic error bound**.

3. Sets having **exponentials** and **logarithms** may have exotic error bounds.

Our goals

1. Prove convergence rates for algorithms for common fixed point problems in a context **as general as possible**.
2. Rates should be concrete: $\operatorname{dist}(x^k, F) \leq R(k)$, for a “reasonable” function R.

T. Liu and L.
Convergence analysis under consistent error bounds
Foundations of Computational Mathematics Vol. 24, 2024, pp. 429-479

T. Liu and L.
Concrete convergence rates for common fixed point problems under Karamata regularity
https://arxiv.org/abs/2407.13234.

Regular Variation (Karamata Theory)

Figure: Jovan Karamata (1902–1967) - pioneer of regularly varying functions. Photo from wikipedia.

N. H. Bingham, C. M. Goldie, and J. L. Teugels.
Regular Variation.
Encyclopedia of Mathematics and its Applications. Cambridge University Press, 1987.

E. Seneta.
Regularly Varying Functions.
Lecture Notes in Mathematics. Springer Berlin Heidelberg, 1976.

Functions of regular variation

$f : [a, \infty) \to (0, \infty)$ is **regularly varying at** ∞ with index ρ if

$$\lim_{x \to \infty} \frac{f(\lambda x)}{f(x)} = \lambda^\rho, \quad \lambda > 0.$$

In this case we write $f \in \mathrm{RV}_\rho$

$f : (0, a] \to (0, \infty)$ is **regularly varying at** 0 with index ρ if

$$\lim_{x \to 0_+} \frac{f(\lambda x)}{f(x)} = \lambda^\rho, \quad \lambda > 0.$$

In this case we write $f \in \mathrm{RV}^0_\rho$

Examples of RV^0 functions:

- t^α has index α
- $-t\ln(t)$ has index 1.
- $-\frac{1}{\ln(t)}$ has index 0.
- $-\sqrt{t}\ln(t)$ has index $1/2$.

Non-example: $e^{-1/t}$.

Helpful properties of regular variation

Asymptotic equivalence up to a constant

$$f(t) \overset{c}{\sim} h(t) \text{ as } t \to a \overset{\text{def}}{\iff} \lim_{t\to a} \frac{f(t)}{h(t)} = \mu > 0$$

- For $f \in \mathrm{RV}_\rho, \rho > -1$

$$\int_a^x f(t)dt \quad \sim \quad \frac{x}{\rho+1} f(x) \quad \text{as} \quad x \to \infty.$$

- For $f, h \in \mathrm{RV}_\rho,\ \rho > 0$

$$f(t) \overset{c}{\sim} h(t) \text{ as } t \to \infty \ \Rightarrow f^{-1}(t) \overset{c}{\sim} h^{-1}(t) \text{ as } t \to \infty$$

$$f(t) = o(h(t)) \text{ as } t \to \infty \ \Rightarrow h^{-1}(t) = o(f^{-1}(t)) \text{ as } t \to \infty$$

Karamata regularity

Joint Karamata regularity

- T_i : operators with $F := \bigcap_{i=1}^m \mathrm{Fix}\, T_i \neq \emptyset$
- B: bounded subset

The T_i are **jointly Karamata regular (JKR) over** B if there exists $\psi_B : \mathbb{R}_+ \to \mathbb{R}_+$ such that.

1. $\mathrm{dist}\,(x, F) \leq \psi_B\left(\max_{1\leq i\leq n} \|x - T_i(x)\|\right), \quad \forall\, x \in B.$
2. ψ_B is nondecreasing and $\lim_{t\to 0_+} \psi_B(t) = \psi_B(0) = 0.$
3. $\psi_B \in \mathrm{RV}_\rho^0$ with $\rho \in [0, 1]$.

- Encompasses Hölderian error bounds, Hölder regular operators and all the previous examples of non-Hölder behavior.

Consistent error bounds

$C_1, \ldots, C_m \subseteq \mathbb{R}^n$: closed convex sets
$C = \cap_{i=1}^m C_i$.

Consistent error bound functions - Liu, L.' 24

$\psi : \mathbb{R}_+ \times \mathbb{R}_+ \to \mathbb{R}_+$ is a **consistent error bound function** for $C_1, \ldots, C_m$ if:

(i)
$$\mathrm{dist}\,(x, C) \leq \psi\left(\max_{1 \leq i \leq m} \mathrm{dist}\,(x, C_i), \|x\|\right) \quad \forall\, x \in \mathbb{R}^n;$$

(ii) $\forall b \geq 0$, $\psi(\cdot, b)$ is monotone nondecreasing, right-continuous at 0 and $\psi(0, b) = 0$;

(iii) $\forall a \geq 0$, $\psi(a, \cdot)$ is monotone nondecreasing.

- If $\psi(\cdot, b) \in \mathrm{RV}_\rho^0$, CEBs become a particular case of Karamata regularity.

Main convergence result

$T_1, \ldots, T_m$: **JKR** α-averaged operators ($\alpha \in (0,1)$). $F := \bigcap_{i=1}^m \mathrm{Fix}\, T_i \neq \emptyset$.
$\{x^k\}$: sequence generated by some **reasonable** algorithm.
ψ_B: regularity function over a bounded set B containing $\{x^k\}$

Define $\phi(u) := \psi_B^2(\sqrt{\kappa u})$

$$\Phi_\phi(u) := \int_u^1 \frac{1}{\phi^-(t)} dt, \ \ u > 0.$$

Then, the convergence of $\{x^k\}$ to $x^* \in F$ is either finite or $\exists \tau > 0$,

$$\mathrm{dist}\,(x^k, F) \leq \sqrt{(\Phi_\phi)^{-1}\Big(L - \tau k\Big)} \quad \forall\ k,$$

where $L = \Phi_\phi\left(\mathrm{dist}^{\,2}\,(x^0, F)\right)$.

Gee... that looks like hard to compute. How practical is that?

It isn't. :(
But regular variation helps bypass most of the misery and pain.

Index of regular variation and convergence rates - Liu, L.'24

Let ρ denote the index of ψ_B.

1. $\rho = 1 \Rightarrow$ convergence rate is **almost linear**.
   - i.e., faster than k^{-r} for any $r > 0$.
2. $\rho \in (0, 1) \Rightarrow$ convergence rate almost the same of as being Hölder with exponent ρ.
   - i.e., faster than $k^{-r/2}$ for any $r > 0$ such that $r < \rho/(1-\rho)$.
3. $\rho = 0 \Rightarrow k^{-r} = o(\Phi_\phi^{-1}(k))$ as $k \to \infty$, i.e., $\Phi_\phi^{-1}(k)$ goes to 0 slower than any **sublinear rate**.
   - If ψ_B is logarithmic, the rate is $\frac{1}{\ln(k)}$.

Reminder: f has index $\rho \overset{\text{def}}{\iff} \lim_{x \to 0_+} \frac{f(\lambda x)}{f(x)} = \lambda^\rho, \quad \lambda > 0$.

Better rates with more effort

$f \in \mathrm{RV}_\rho^0$ with $\rho \in [0, 1]$, nondecreasing with $\lim_{x \to 0_+} f(x) = 0$.

$$\Phi_f(x) := \int_x^1 \frac{1}{f^-(t)} dt, \quad x > 0,$$

Let $g(x) := \frac{1}{x f^-(1/x)}$

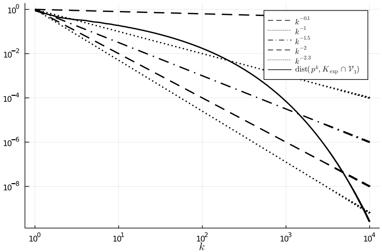
Better rates - Liu, L.'24

1. $\rho = 1$ and $f(t) \overset{c}{\sim} t$ as $t \to 0_+ \Rightarrow \tau_1 c_1^s \le \Phi_f^{-1}(s) \le \tau_2 c_2^s$ whenever s is large enough.
2. $\rho = 1$ and $t = o(f(t))$ as $t \to 0_+ \Rightarrow \Phi_f^{-1}(s) = \frac{1}{o(g^{\leftarrow}(s))}$ as $s \to \infty$.
3. $\rho \in (0, 1) \Rightarrow \Phi_f^{-1}(s) \overset{c}{\sim} \frac{1}{g^{\leftarrow}(s)}$ as $s \to \infty$
4. $\rho = 0$ and $\ln(g) \in \mathrm{RV}_q$ with $q > 0 \Rightarrow$ then for $\widehat{g}(x) := x g(x) = \frac{1}{f^-(1/x)}$, we have $\Phi_f^{-1}(s) \sim \frac{1}{\widehat{g}^{\leftarrow}(s)}$ as $s \to \infty$.

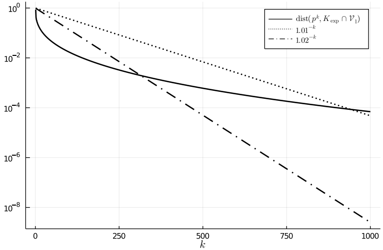
Entropic error bound ($-t\ln t$)

$$\text{find} \quad p \in K_{\exp} \cap \{(0,1,0)\}^{\perp}$$

Consider the cyclic projections algorithm. Starting point is $(1,1,1)$. Theory says the convergence is **almost linear**.



(a) Log-log plot of $\mathrm{dist}\,(p^k, K_{\exp} \cap \{(0,1,0)\}^{\perp})$. Dashed and dotted lines correspond to k^{-r} for a few values of r.



(b) Plot of $\mathrm{dist}\,(p^k, K_{\exp} \cap \{(0,1,0)\}^{\perp})$, where only the y-axis is in log scale. Functions of the form c^{-k} appear as straight lines.

$$\sqrt{k}c^{-\sqrt{k}} = o\left(\sqrt{\Phi_{\phi}^{-1}(k)}\right) \quad \text{as} \quad s \to \infty.$$

Logarithmic error bound $(-\frac{1}{\ln t})$

$$\text{find} \quad p \in K_{\exp} \cap \{(0,0,1)\}^{\perp}$$

Consider the cyclic projections algorithm. Starting point is $(1,1,1)$. Theory says the convergence is **logarithmic**.

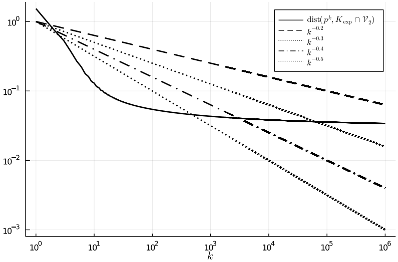


Figure: Log-log plot of $\mathrm{dist}\,(p^k, K_{\exp} \cap \{(0,1,0)\}^{\perp})$. Dashed and dotted lines correspond to k^{-r} for a few values of r.

A Douglas-Rachford example

- We constructed two sets C_1, C_2 for which T_{DR} satisfies:

$$\operatorname{dist}\left(w, \operatorname{Fix} T_{\mathrm{DR}}\right) \le \kappa\psi_B\Big(\,\|T_{\mathrm{DR}}(w) - w\|\,\Big),$$

  with

$$\psi_B(t) = -\sqrt{t}\ln(t), \qquad \text{(for } t \text{ small)}$$

  and $\psi_B \in \mathrm{RV}^0_{1/2}$ is the "best" possible regularity function.

Bounds for the convergence rate of $\{w^k\}$ generated by DR:

- Faster than $k^{-r/2}$ for any $r<1$.
- $\sqrt{\Phi_\phi^{-1}(k)} \overset{c}{\sim} \left[W_0(\sqrt{k})\right]^2 k^{-1/2}$ as $s \to \infty$.

Definable operators and joint Karamata regularity

- Pick your favourite *o*-minimal structure: semialgebraic sets, global subanalytic sets, log-exp structure...

If

- $T_1, \ldots, T_m$ are definable continuous quasi-nonexpansive operators with $F := \bigcap_{i=1}^m \mathrm{Fix}\, T_i \neq \emptyset$;
- B is a bounded set

then

$T_1, \ldots, T_m$ are jointly Karamata regular.

- Error bounds between definable convex sets can always be described by regularly varying functions.
- Definable convex sets admit consistent error bound functions that are regularly varying.

Conclusion

- 1. Joint Karamata regularity:
  $\mathrm{dist}\,(x, F) \leq \psi_B\Big(\max_{1\leq i\leq n} \|x - T_i(x)\|\Big), \quad \forall\ x \in B.$
  2. Convergence rates: $\mathrm{dist}\,(x^k, F) \leq \sqrt{(\Phi_\phi)^{-1}\Big(L - \tau k\Big)}$, where
  $\Phi_\phi(u) := \int_u^1 \frac{1}{\phi^-(t)} dt$ and $\phi(u) := \psi_B^2(\sqrt{\kappa u})$.
  3. $(\Phi_\phi)^{-1}$ is hard to compute, but asymptotic analysis can be done with **regular variation**. (index is **easy** to compute)
- Concrete convergence rates for exotic regularity.

T. Liu and L.
Convergence analysis under consistent error bounds
Foundations of Computational Mathematics Vol. 24, 2024, pp. 429-479

T. Liu and L.
Concrete convergence rates for common fixed point problems under Karamata regularity
https://arxiv.org/abs/2407.13234.