On the convergence of the Krasnoselskij iteration for strictly pseudocontractive operators

Diego Deplano, Sergio Grammatico, Mauro Franceschelli

***Abstract*— We study the convergence of the nonlinear Krasnoselskij iteration $x(k+1) = (1-\theta)x(k) + \theta T(x(k))$ in real vector spaces of finite dimension equipped with a p-norm, which is relevant for stability analysis and distributed computation in several discrete-time dynamical systems. Specifically, we provide sufficient conditions for the convergence of the Krasnoselskij iteration, derived via implications between the strict pseudocontractivity of the operator T and the nonexpansiveness of $(1-\theta)\mathrm{Id} + \theta T$. Interestingly, it turns out that strict pseudocontractivity of T is necessary for the Euclidean norm ($p = 2$) only; not necessary for non-Euclidean norms ($p \neq 2$); sufficient for any finite norm $p \in (1, \infty)$; not sufficient for the taxi-cab norm ($p = 1$) and the supremum norm ($p = \infty$). We numerically verify the above results in the context of recurrent neural networks and multi-agent systems with nonlinear Laplacian dynamics.**

***Index Terms*— Fixed-point iteration, Krasnoselskij iteration, Strict pseudocontractivity, Contractive systems.**

I. Introduction

CONSIDER the Banach-Picard iteration [1, Eq. (1.69)] in the form of discrete-time dynamical system

$$\boldsymbol{x}(k+1) = \mathsf{T}_\theta(\boldsymbol{x}(k)) = (1-\theta)\boldsymbol{x}(k) + \theta\mathsf{T}(\boldsymbol{x}(k)),\ k \in \mathbb{N}, \quad (1)$$

where $\theta \in (0, 1)$ and $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^n$ such that $\mathrm{fix}(\mathsf{T}) \neq \emptyset$.

One of the first convergence results dates back to 1955 and it is due to Krasnoselskii [2][3, Theorem 6.4.1], who proved convergence of $\boldsymbol{x}(k)$ to a fixed point when T is nonexpansive and $\theta = \frac{1}{2}$ for *uniformly convex* spaces [4, Definition 1.8]. More than 10 years later, Edelstein in [5] extended this result to $\theta \in (0, 1)$ and *strictly convex* spaces[1] [4, Definition 1.10]. In 1976, the convergence results for the Banach-Picard iteration in (1) in uniformly/strictly convex spaces were extended to general Banach spaces by Ishikawa [6, Theorem 1], see also [3, Theorem 6.4.3]. By limiting their analysis to Hilbert spaces, Marino and Xu in [7] proved that the iteration in (1) converges also when the map T is κ-strictly pseudocontractive and $\theta < 1 - \kappa$. Moreover, for linear maps in Hilbert spaces, it has been recently proven that the condition on T being κ-strictly pseudocontractive is both necessary and sufficient for the convergence of the Krasnoselkij iteration, given $\theta < 1 - \kappa$ [8, Theorem 1]. Marino and Xu in [7] also posed the currently open question: "Does this result hold also in Banach spaces which are uniformly convex?". Since then, many authors have provided different answers to this question by considering several iteration schemes and sets of assumptions [9]–[15].

From a general mathematical perspective, the convergence problem is a fixed-point problem [4], or equivalently, a zero finding problem [1]. For example, consensus in nonlinear multi-agent systems is equivalent to finding a collective state in the kernel of the nonlinear Laplacian operator [16]–[19]. Variations of the Krasnoselskij fixed-point iteration have also been adopted in aggregative game theory [20], [21], monotone operator splitting methods in distributed convex optimization [22]–[24], and monotone dynamical systems [25], [26].

The main contribution of this paper is showing what are the values of θ that ensure convergence of the Krasnoselskij iteration in real Banach spaces $\mathcal{S}_p = (\mathbb{R}^n, \|\cdot\|_p)$ of finite-dimension n equipped with a p-norm for $p \in [1, \infty]$. Note that spaces $\mathcal{S}_p$ are indeed $\max\{2, p\}$-uniformly convex for $p \in (1, \infty)$, but not for $p \in \{1, \infty\}$. More precisely, we provide an explicit tight upper bound for θ such that convergence holds true. Our technical results disprove those of Chidume and Shahzad in [13], Sahu and Petrusel in [27] and those of Cholamjiak and Suantai in [14], while they are consistent with the earlier work of Zhang and Su in [9].

II. Notation and Preliminaries

The set of real and integer numbers are denoted by $\mathbb{R}$ and $\mathbb{Z}$, and their restriction to nonnegative and positive values are denoted with $\mathbb{R}_{\geq 0}$, $\mathbb{N}$ and $\mathbb{R}_{>0}$, $\mathbb{N}_+$, respectively. Matrices $M \in \mathbb{R}^{n\times n}$ are denoted by uppercase letters, vectors $\boldsymbol{v} \in \mathbb{R}^n$ by bold letters, scalars $s \in \mathbb{R}$ by lowercase letters, while sets and spaces $\mathcal{S}$ are denoted by uppercase calligraphic letters. We denote by $\mathbf{0}_n$ and $\mathbf{1}_n$ the vector of zeros and ones of dimension n, respectively. Mappings $\mathsf{T} : \mathcal{X}_1 \to \mathcal{X}_2$ between two spaces $\mathcal{X}_1, \mathcal{X}_2$ are usually denoted with block capital letters; for instance, the linear operator associated to the identity matrix I is defined by $\mathsf{Id} : \boldsymbol{x} \mapsto I\boldsymbol{x}$. When $\mathcal{X}_2 \equiv \mathbb{R}$, block lowercase letters are used instead, e.g., $\mathsf{t} : \mathcal{X} \to \mathbb{R}$. Given a self-mapping $\mathsf{T} : \mathcal{X} \to \mathcal{X}$, $\mathrm{fix}(\mathsf{T}) = \{\boldsymbol{x} \in \mathcal{X} \mid \mathsf{T}(\boldsymbol{x}) = \boldsymbol{x}\}$ denotes the set of its fixed points and $\mathrm{zer}(\mathsf{T}) = \{\boldsymbol{x} \in \mathcal{X} \mid \mathsf{T}(\boldsymbol{x}) = \mathbf{0}\}$ denotes the set of its zeros.

Diego Deplano and Mauro Franceschelli are with the DIEE, University of Cagliari, 09123, Italy. Emails: {diego.deplano, mauro.franceschelli}@unica.it.

Sergio Grammatico is with the Delft Center for Systems and Control, TU Delft, The Netherlands. Email: s.grammatico@tudelft.nl.

[1]Uniform convexity and strict convexity are equivalent in finite-dimensional Banach spaces [4, Page 9, Comment 3]). Moreover, spaces $\mathcal{S}_p = (\mathbb{R}^n, \|\cdot\|_p)$ are uniformly/strictly convex if and only if $p \in (1, \infty)$, whereas the spaces S_1 and S_∞ are not.

A. Operator-Theoretic definitions in real Banach spaces

A *normed vector space* is a pair $(\mathcal{X}, \|\cdot\|)$ where $\mathcal{X}$ is a vector space and $\|\cdot\|$ is a norm on $\mathcal{X}$, which induces in the natural way a metric, i.e., a notion of distance: the distance between two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$ is given by $\|\boldsymbol{x}-\boldsymbol{y}\|$. We focus on the real vector space $\mathcal{X} = \mathbb{R}^n$ of finite dimension $n \in \mathbb{N}$ equipped with a p-norm $\|\cdot\|_p$, for $p \in [1,\infty]$, defined as follows:

$$p \in [1,\infty): \qquad \|\boldsymbol{x}\|_p = \sqrt[p]{\sum_{i=1}^{n} |x_i|^p},$$
$$p = \infty: \qquad \|\boldsymbol{x}\|_\infty = \lim_{p\to\infty} \|\boldsymbol{x}\|_p = \max_{i=1,\ldots,n} |x_i|.$$

We denote these spaces with $\mathcal{S}_p = (\mathbb{R}^n, \|\cdot\|_p)$, which are Banach spaces since every finite-dimensional normed vector space is complete as in [4, Def. 1.5 and Rem. 2 on page 7], see also [28, Theorem 5.33]. The only Hilbert space is for $p = 2$, for which the inner product is well defined by $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = \boldsymbol{x}^\top \boldsymbol{x} = \|\boldsymbol{x}\|_2^2$. We now introduce some duality concepts of real Banach spaces.

Definition 1. *[4, Def. 1.11] The dual of $\mathcal{S}_p$ is denoted by $\mathcal{S}_p^* = ((\mathbb{R}^n)^*, \|\cdot\|_p^*)$ and it is defined as follows:*

- *The dual space $(\mathbb{R}^n)^*$ is the set of all continuous linear mappings $\mathsf{L}_{\boldsymbol{z}} : \mathbb{R}^n \to \mathbb{R}$ uniquely defined by a vector $\boldsymbol{z} \in \mathbb{R}^n$, such that $\mathsf{L}_{\boldsymbol{z}}(\boldsymbol{x}) = \boldsymbol{z}^\top \boldsymbol{x}$;*
- *The dual norm $\|\cdot\|_p^*$ is defined by*

$$\|\mathsf{L}_{\boldsymbol{z}}\|_p^* = \sup_{\|\boldsymbol{x}\|_p \le 1} \left|\boldsymbol{z}^\top \boldsymbol{x}\right|.$$

The concept of a duality mapping was introduced by Beurling and Livingston in [29]. We define its generalized form in the case of spaces $\mathcal{S}_p$ by means of the Holder's conjugate numbers.

Definition 2. *Two elements $p, q \in [1,\infty]$ are Holder's conjugate if $\frac{1}{p} + \frac{1}{q} = 1$ where, by convention, $1/\infty = 0$.*

Definition 3. *[9, Page 1, Paragraph 2] The generalized duality mapping for $\mathcal{S}_p = (\mathbb{R}^n, \|\cdot\|_p)$ is defined by*[2]

$$\mathsf{J}_r(\boldsymbol{x}) = \{\mathsf{L}_{\boldsymbol{z}} : \mathbb{R}^n \to \mathbb{R} \mid \ \boldsymbol{x}^\top \boldsymbol{z} = \|\boldsymbol{x}\|_p^r,\ \|\boldsymbol{x}\|_p^{r-1} = \|\boldsymbol{z}\|_q\}, \tag{2}$$

where $r \in [1,2]$ and $p, q \in [1,\infty]$ are Holder's conjugate.

We now provide some useful results.

Lemma 1. *Let $p, q \in [1,\infty]$ be Holder's conjugate, then the dual norm $\|\cdot\|_p^*$ is given by $\|\mathsf{L}_{\boldsymbol{z}}\|_p^* = \|\boldsymbol{z}\|_q$.*

Proof: See Appendix A. ∎

Lemma 2. *Consider a Banach space $\mathcal{S}_p = (\mathbb{R}^n, \|\cdot\|_p)$ with $p \in [1,\infty]$. Given $r = \min\{2,p\}$, the generalized duality mapping is not empty $\mathsf{J}_r(\boldsymbol{x}) \neq \emptyset$ for any $\boldsymbol{x} \in \mathcal{X}$ and consists of (at least) one linear mapping $\mathsf{L}_{\mathsf{j}_r(\boldsymbol{x})} \in \mathsf{J}_r(\boldsymbol{x})$ defined by*

$$\mathsf{j}_r(\boldsymbol{x}) := \begin{cases} \operatorname{sign}(\boldsymbol{x}) \circ |\boldsymbol{x}|^{p-1}/\|\boldsymbol{x}\|_p^{p-r} & \text{if} \quad p \in [1,\infty) \\ \boldsymbol{x} \circ \boldsymbol{x}_\infty / \mathbf{1}^\top \boldsymbol{x}_\infty & \text{if} \quad p = \infty \end{cases}, \tag{3}$$

where $\circ$ denotes the Hadamard product and where

$$\boldsymbol{x}_\infty = \begin{bmatrix} \vdots \\ x_{\infty,i} \\ \vdots \end{bmatrix} \quad \text{with} \quad x_{\infty,i} = \begin{cases} 1 & \text{if} \quad |x_i| = \max|\boldsymbol{x}| \\ 0 & \text{otherwise} \end{cases}.$$

Proof: See Appendix B. ∎

Lemma 3. *Consider a Banach space $\mathcal{S}_p = (\mathbb{R}^n, \|\cdot\|_p)$ with $p \in [1,\infty]$. Given $r = \min\{2,p\}$, the generalized duality mapping is single-valued $\mathsf{J}_r(\boldsymbol{x}) = \{\mathsf{L}_{\mathsf{j}_r(\boldsymbol{x})}\}$ for any $\boldsymbol{x} \in \mathcal{X}$ if and only if $p \in (1,\infty)$.*

Proof: See Appendix C. ∎

Among nonlinear mappings, the classes of nonexpansive mappings and pseudocontractions play a pivotal role. Let us define these properties in the context of Banach spaces $\mathcal{S}_p$.

Definition 4. *Given a Banach space $\mathcal{S}_p$, if a mapping $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^n$ satisfies*

$$\|\mathsf{T}(\boldsymbol{x}) - \mathsf{T}(\boldsymbol{y})\|_p \le \ell\, \|\boldsymbol{x}-\boldsymbol{y}\|_p, \tag{4}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, then it is called:

- *ℓ-contractive (ℓ-C) if $\ell \in (0,1)$;*
- *nonexpansive (NE) if $\ell = 1$.*

Definition 5. *Given the Banach space $\mathcal{S}_p$, if a mapping $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^n$ satisfies*

$$\begin{aligned} \mathsf{L}_r(\mathsf{T}(\boldsymbol{x}) - \mathsf{T}(\boldsymbol{y})) \le &\ \|\boldsymbol{x}-\boldsymbol{y}\|_p^r \\ &- \frac{1-\kappa}{r}\|\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y}))\|_p^r, \end{aligned} \tag{5}$$

for all $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$, with $r = \min\{2,p\}$ and $\mathsf{L}_r \in \mathsf{J}_r(\boldsymbol{x}-\boldsymbol{y})$, then it is called ([9]):

- *κ-strictly pseudocontractive (κ-sPC) if $\kappa \in (0,1)$;*
- *pseudocontractive (PC) if $\kappa = 1$.*

We note that it holds: ℓ-C $\Rightarrow$ NE $\Rightarrow$ κ-sPC $\Rightarrow$ PC.

III. Main results

Our first main result in Theorem 1 characterizes the relation between the nonexpansiveness of the Krasnoselkij iteration operator and the strict pseudocontractivity of the corresponding mapping, which is instrumental to obtain sufficient conditions for its convergence, our second main result, Theorem 2. With this aim, we make use of Reich's inequality [30], which is given in the following Lemma 4 in the special case of $\mathcal{S}_p$ spaces, by exploiting the results of Honh-Kun Xu in [31].

Lemma 4. *[30][31, Eqs. (3.5)' and (3.8)' in Corollary 2] Consider the Banach space $\mathcal{S}_p$ with $p \in (1,\infty)$. Given two vectors $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ and the dual linear mapping $\mathsf{L}_{\mathsf{j}_r(\boldsymbol{x})} \in \mathsf{J}_r(\boldsymbol{x})$, then it holds that*

$$\|\boldsymbol{x}+\boldsymbol{y}\|_p^r \le \|\boldsymbol{x}\|_p^r + r\mathsf{L}_{\mathsf{j}_r(\boldsymbol{x})}(\boldsymbol{y}) + c_p\|\boldsymbol{y}\|_p^r, \tag{6}$$

where $r = \min\{p, 2\}$ and

$$c_p = \begin{cases} p-1 & \text{if } p \ge 2 \\ (1+t_p^{p-1})(1+t_p)^{1-p} & \text{if } p \in (1,2) \end{cases},$$

with t_p being the unique solution of the following equation

$$(p-2)t^{p-1} + (p-1)t^{p-2} = 1.$$

Moreover, the constant c_p as given above is the best possible.

[2] This definition of generalized duality mapping is a special case of that in [9] for spaces $\mathcal{S}_p$, for which holds that $\|\mathsf{L}_{\boldsymbol{z}}\|_p^* = \|\boldsymbol{z}\|_q$, as shown in Lemma 1.

Example 1. *This example illustrates Lemma 4, showing that for $p=\infty$, it holds $c_p=\infty$. Let $a>1$ and consider the following pair of vectors:*

$$\boldsymbol{x}=\begin{bmatrix}a+1 & a\end{bmatrix}^\top, \quad \boldsymbol{y}=\begin{bmatrix}-1 & 1\end{bmatrix}^\top.$$

Given $r=\min\{p,2\}$, one can compute the generalized duality mapping J_2 of $\boldsymbol{x}$, which is given by

$$\mathsf{J}_2(\boldsymbol{x})=\left\{\mathsf{L}_{\boldsymbol{z}}:\mathbb{R}^n\to\mathbb{R}\mid \boldsymbol{z}=\begin{bmatrix}a+1 & 0\end{bmatrix}^\top\right\}.$$

Note that $\mathsf{J}_2(\boldsymbol{x})$ consists of only one element and is in line with Lemma 2. One can verify that the above is correct by

$$\mathsf{j}_2(\boldsymbol{x})\in\mathsf{J}_2(\boldsymbol{x}) \;\Rightarrow\; \boldsymbol{x}^\top\mathsf{j}_2(\boldsymbol{x})=\|\boldsymbol{x}\|_\infty^2=\|\mathsf{j}_2(\boldsymbol{x})\|_1^2=a+1.$$

Let us now compute the following norms:

$$\|\boldsymbol{x}\|_\infty=a+1, \quad \|\boldsymbol{x}+\boldsymbol{y}\|_\infty=a+1, \quad \|\boldsymbol{y}\|_\infty=1$$

and also

$$\mathsf{L}_{\mathsf{j}_2(\boldsymbol{x})}(\boldsymbol{y})=\boldsymbol{y}^\top\mathsf{j}_2(\boldsymbol{x})=-(a+1).$$

By substituting the above into inequality (6) *of Lemma 4, we obtain $(a+1)^2\le(a+1)^2-2(a+1)+c_p$, yielding*

$$c_p\ge 2(a+1).$$

Thus, as $a\to\infty$ we have $c_p\to\infty$. This proves that for $p=\infty$, there is not a finite lower bound to c_p that holds for any pair $\boldsymbol{x},\boldsymbol{y}$ in the whole space $\mathbb{R}^n$ with $n\ge 2$. ■

Our main result about the relation between the pseudo-contractivity of a mapping and the nonexpansiveness of the Krasnoselkij iteration map is given next.

Theorem 1. *Consider a Banach space $\mathcal{S}_p$ with $p\in[1,\infty]$, a mapping $\mathsf{T}:\mathcal{X}\to\mathcal{X}$ and the following properties:*
- (a) *T is κ-sPC for some $\kappa\in(0,1)$;*
- (b) *$\mathsf{T}_\theta=(1-\theta)\mathsf{Id}+\theta\mathsf{T}$ is NE for some $\theta\in(0,1]$;*

Given $r=\min\{2,p\}$, the following statements hold:
- $(s1)$ *$(a)\Leftarrow(b)$ with $\theta\ge 1-\kappa$ if and only if $p=2$;*
- $(s2)$ *$(a)\Rightarrow(b)$ with $\theta^{r-1}\le(1-\kappa)/c_p$ when $p\in(1,\infty)$;*
- $(s3)$ *$(a)\not\Rightarrow(b)$ for any $\theta\in(0,1)$ when $p\in\{1,\infty\}$.*

Proof: We recall that a continuous linear mapping $\mathsf{L}\in\mathsf{J}(\boldsymbol{x})$ is such that the following properties hold:
- $\mathsf{L}(\theta\boldsymbol{x})=\theta\mathsf{L}(\boldsymbol{x})$ for all $\theta\in\mathbb{R}$ and $\boldsymbol{x}\in\mathbb{R}^n$;
- $\mathsf{L}(\boldsymbol{x}\pm\boldsymbol{y})=\mathsf{L}(\boldsymbol{x})\pm\mathsf{L}(\boldsymbol{y})$ for all $\boldsymbol{x},\boldsymbol{y}\in\mathbb{R}^n$.

Proof of statement $(s1)$: For each $p\in(1,\infty)$, let $r=\min\{2,p\}$ and $\mathsf{L}_r\in\mathsf{J}_r(\boldsymbol{x}-\boldsymbol{y})$, then the following chain of inequalities holds:

$$\begin{aligned}
&\mathsf{L}_r(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y}))=\\
&\quad=\frac{1}{\theta}\mathsf{L}_r\big(\mathsf{T}_\theta(\boldsymbol{x})-(1-\theta)\boldsymbol{x}-\mathsf{T}_\theta(\boldsymbol{y})+(1-\theta)\boldsymbol{y}\big)\\
&\quad\overset{(i)}{=}\frac{1}{\theta}\Big[\mathsf{L}_r\big(\mathsf{T}_\theta(\boldsymbol{x})-\mathsf{T}_\theta(\boldsymbol{y})\big)-(1-\theta)\mathsf{L}_r\big(\boldsymbol{x}-\boldsymbol{y}\big)\Big]\\
&\quad\overset{(ii)}{=}-\frac{1}{\theta}\Big[\mathsf{L}_r\big(\mathsf{T}_\theta(\boldsymbol{y})-\mathsf{T}_\theta(\boldsymbol{x})\big)+(1-\theta)\|\boldsymbol{x}-\boldsymbol{y}\|_p^r\Big]\\
&\quad\overset{(iii)}{\le}-\frac{1}{\theta}\Big[\frac{1}{r}\Big(\|\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}_\theta(\boldsymbol{x})-\mathsf{T}_\theta(\boldsymbol{y}))\|_p^r-\|\boldsymbol{x}-\boldsymbol{y}\|_p^r\\
&\qquad\qquad -c_p\|\mathsf{T}_\theta(\boldsymbol{x})-\mathsf{T}_\theta(\boldsymbol{y})\|_p^r\Big)+(1-\theta)\|\boldsymbol{x}-\boldsymbol{y}\|_p^r\Big]\\
&\quad\overset{(iv)}{\le}-\frac{1}{\theta}\Big[(1-\theta-\frac{1+c_p}{r})\|\boldsymbol{x}-\boldsymbol{y}\|_p^r+\frac{\theta^r}{r}\|\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y}))\|_p^r\\
&\quad=(1-\frac{1}{\theta}+\frac{1+c_p}{\theta r})\|\boldsymbol{x}-\boldsymbol{y}\|_p^r-\frac{\theta^{r-1}}{r}\|\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y}))\|_p^r\\
&\quad\overset{(v)}{\le}\|\boldsymbol{x}-\boldsymbol{y}\|_p^r-\frac{\theta^{r-1}}{r}\|\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y}))\|_p^r\\
&\quad\overset{(vi)}{\le}\|\boldsymbol{x}-\boldsymbol{y}\|_p^r-\frac{1-\kappa}{r}\|\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y}))\|_p^r,
\end{aligned}$$

where (i) holds by linearity of L; (ii) follows by Definition 3; (iii) follows by Lemma 4; (iv) follows by nonexpansiveness of T_θ; (v) holds if and only if $c_p\le r-1$, which holds if and only if $p=r=2$; (vi) holds for $\theta^{r-1}\ge 1-\kappa$. This proves statement $(s1)$. We now prove statement $(s2)$:

$$\begin{aligned}
&||\mathsf{T}_\theta(\boldsymbol{x})-\mathsf{T}_\theta(\boldsymbol{y})||_p^r=\\
&\quad=\|(1-\theta)(\boldsymbol{x}-\boldsymbol{y})+\theta(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y}))\|_p^r\\
&\quad=\|\boldsymbol{x}-\boldsymbol{y}-\theta(\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y})))\|_p^r\\
&\quad\overset{(i)}{\le}\|\boldsymbol{x}-\boldsymbol{y}\|_p^r-r\theta\mathsf{L}_r(\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y})))\\
&\qquad\quad+c_p\|\theta(\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y})))\|_p^r\\
&\quad=(1-r\theta)\|\boldsymbol{x}-\boldsymbol{y}\|_p^r+r\theta\mathsf{L}_{\mathsf{j}_p(\boldsymbol{x}-\boldsymbol{y})}(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y}))\\
&\qquad\quad+\theta^r c_p\|\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y}))\|_p^r\\
&\quad\overset{(ii)}{\le}\|\boldsymbol{x}-\boldsymbol{y}\|_p^r-\theta(1-\kappa)\|\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y}))\|_p^r\\
&\qquad\quad+\theta^r c_p\|\boldsymbol{x}-\boldsymbol{y}-(\mathsf{T}(\boldsymbol{x})-\mathsf{T}(\boldsymbol{y}))\|_p^r\\
&\quad\overset{(iii)}{\le}\|\boldsymbol{x}-\boldsymbol{y}\|_p^2
\end{aligned}$$

where (i) holds by Lemma 4; (ii) holds since map T is κ-sPC; (iii) holds for $-\theta(1-\kappa)+\theta^r c_p\le 0$ which implies $\theta^{r-1}\le(1-\kappa)/c_p$. This proves statement $(s2)$. On the other hand, in the limit of $p\to 1^+$, it holds that $r=p$ and $c_p\to 2^-$, and therefore $1\le(1-\kappa)/2$, which is in contrast with $k\ge 0$ impossible. Moreover, in the limit of $p\to\infty$, it holds that $r=2$ and $c_p\to\infty$, and therefore $\theta\le 0$, which is a contradiction. This proves statement $(s3)$. ■

Corollary 1. *For the Euclidean norm, i.e. for $p=2$, statements $(s1)$ and $(s2)$ of Theorem 1 imply*[1] *that*

$$(a)\Leftrightarrow(b) \text{ with } \theta=1-\kappa.$$

We are now in the position to prove our second main result for the convergence of the Krasnoselsij iteration.

Theorem 2. *Consider a Banach space $\mathcal{S}_p$ and a map $\mathsf{T}:\mathbb{R}^n\to\mathbb{R}^n$ such that $\mathsf{fix}(\mathsf{T})\neq\emptyset$. Given $r=\min\{2,p\}$, the following statements hold:*
- *The iteration in* (1) *converges if T is* NE *w.r.t. $\|\cdot\|_p$ for $p\in[1,\infty]$ and $\theta<1$;*
- *The iteration in* (1) *converges if T is κ-sPC w.r.t. $\|\cdot\|_p$ for $p\in(1,\infty)$ and $\theta^{r-1}<(1-\kappa)/c_p$.*

Proof: The first statement has been proved by Ishikawa in [6, Theorem 1]. By our Theorem 1, for $p\in(1,\infty)$ it holds that for $\theta^*=((1-\kappa)/c_p)^{1/(r-1)}\in(0,1)$ the map

[1] A similar result has been recently proven for linear maps in the Hilbert space $(\mathbb{R}^n,\|\cdot\|_{2,P})$ where $\|\cdot\|_{2,P}=\sqrt{x^\top Px}$ and where P is a symmetric and positive definite matrix [8, Lemma 5].

$\mathsf{T}_{\theta^*} = (1-\theta^*)\mathsf{Id} + \theta^* \mathsf{T}$ is NE. Consequently, map T_θ ruling the iteration in (1) can be equivalently written as

$$\mathsf{T}_\theta = (1 - \frac{\theta}{\theta^*})\mathsf{Id} + \frac{\theta}{\theta^*}\mathsf{T}_{\theta^*}.$$

Thus, T_θ can be seen as the Krasnoselskij iteration of the nonexpansive map T_{θ^*} with coefficient θ/θ^*, which is known to converge for $\theta/\theta^* < 1$ by [6, Theorem 1], i.e.,

$$\frac{\theta}{\theta^*} < 1 \Rightarrow \frac{\theta}{((1-\kappa)/c_p)^{1/(r-1)}} < 1 \Rightarrow \frac{\theta^{r-1}}{(1-\kappa)/c_p} < 1,$$

completing the proof. ■

The following corollary directly follows from Theorem 2 and [1, Theorem 5.14(iii)], which ensures the convergence of the Krasnoselksij-Mann iteration given by

$$\boldsymbol{x}(k+1) = (1-\theta_k)\boldsymbol{x}(k) + \theta_k \mathsf{T}(\boldsymbol{x}(k)), \tag{7}$$

when T is sPC and where the sequence $(\theta_k)_{k\in\mathbb{N}}$ is such that $0 \leq \theta_k \leq \theta_{\text{MAX}} < \infty$ for all $k \in \mathbb{N}$, for some θ_{MAX}, and such that $\lim_{k\to\infty} \theta_k = 0$ with $\sum_0^\infty \theta_k = 0$.

Corollary 2. *Consider a Banach space $\mathcal{S}_p$ and a map $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^n$ such that $\mathrm{fix}(\mathsf{T}) \neq \emptyset$. Given $r = \min\{2, p\}$, the following statements hold:*

- *The iteration in (7) converges if T is NE w.r.t. $\|\cdot\|_p$ for $p \in [1,\infty]$;*
- *The iteration in (7) converges if T is κ-sPC w.r.t. $\|\cdot\|_p$ for $p \in (1,\infty)$.*

A. Comparison with the state-of-the-art

Let us first show that our results disprove the correctness of the results in [13], [14], [27]. Let $p \geq 2$ and let us recall the general form of the Reich inequality in (6), as originally formulated in [30] and then recalled in [13, Lemma 2.3] and [14, Lemma 1.5]:

$$\|\boldsymbol{x}+\boldsymbol{y}\|_p^2 \leq \|\boldsymbol{x}\|_p^2 + 2\mathsf{L}_{\mathsf{j}_r(\boldsymbol{x})}(\boldsymbol{y}) + \max\{\|\boldsymbol{x}\|_p, 1\}\|\boldsymbol{y}\|_p \beta(\|\boldsymbol{y}\|_p),$$

where $\beta : \mathbb{R}_{>0} \to \mathbb{R}_{>0}$ is a continuous function such that

$$\lim_{t\to 0^+} \beta(t) = 0, \qquad \beta(ct) \leq c\beta(t), \quad \forall c \geq 1.$$

Chidume and Su in [13, Lemma 3.2] and also Sahu and Petrusel [27] have based their result on the following assumption, with $p \geq 2$:

$$\beta(t) \leq t. \tag{8}$$

For vectors $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|_p \leq 1$ and under the assumption in (8), the Rich inequality becomes

$$\|\boldsymbol{x}+\boldsymbol{y}\|_p^2 \leq \|\boldsymbol{x}\|_p^2 + 2\mathsf{L}_{\mathsf{j}_r(\boldsymbol{x})}(\boldsymbol{y}) + \|\boldsymbol{y}\|_p^2.$$

Comparing the above with eq. (6) in Lemma 4, one can verify that it holds if and only if $c_p = p - 1 = 1$, i.e., $p = 2$. Therefore, the results in [13], [27] hold for the Hilbert space $\mathcal{S}_2$ only and not for generic Banach spaces $\mathcal{S}_p$ with $p \neq 2$. This fact was also noticed by Cholamjiak and Suntai in [14], who made instead the following assumption, with $p \geq 2$:

$$\beta(t) \leq 2t. \tag{9}$$

For vectors $\boldsymbol{x}$ such that $\|\boldsymbol{x}\|_p \leq 1$ and under the assumption in (8), the Rich inequality becomes

$$\|\boldsymbol{x}+\boldsymbol{y}\|_p^2 \leq \|\boldsymbol{x}\|_p^2 + 2\mathsf{L}_{\mathsf{j}_r(\boldsymbol{x})}(\boldsymbol{y}) + 2\|\boldsymbol{y}\|_p^2.$$

Comparing the above with eq. (6) in Lemma 4, one can verify that it holds only if $c_p = p - 1 \leq 2$, i.e., $p \in [2,3]$. Therefore, the results in [13], [27] hold only for $p \in [2,3]$ and not for $p \in [1,2) \cup (3,\infty]$. On the other hand, our results are consistent with those of Zhang in [9] where, however, the explicit values of the constant c_p determining the upper bound on the feasible value of θ are not given.

B. Application example: Recurrent Neural Networks

Consider the following continuous-time recurrent neural-network, usually referred to as the *firing-rate* model:

$$\dot{\boldsymbol{x}}(t) = \overbrace{-\boldsymbol{x}(t) + \underbrace{\Phi(A\boldsymbol{x}(t) + \boldsymbol{b})}_{\mathsf{T}(\boldsymbol{x}(t))}}^{-\mathsf{F}(\boldsymbol{x}(t))} \tag{10}$$

where $\boldsymbol{x}, \boldsymbol{b} \in \mathbb{R}^n$, $A \in \mathbb{R}^{n\times n}$ and $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is an activation mapping applied entrywise, i.e., $\Phi(\boldsymbol{x}) = (\phi(x_1), \ldots, \phi(x_n))$ with $\phi : \mathbb{R} \to \mathbb{R}$. In this example, we consider the case that ϕ is a LeakyReLU activation function, i.e.,

$$\phi(x) = \max\{x, ax\}, \qquad a = 0.1.$$

Stationary point of F are also fixed points of T, i.e., $\mathrm{zer}(\mathsf{F}) \equiv \mathrm{fix}(\mathsf{T}(k))$. In order to find a stationary point, one can apply the forward step method $\boldsymbol{x}(k+1) = (\mathsf{Id} - \theta\mathsf{F})\boldsymbol{x}(k)$, which leads to the iteration in (1):

$$\boldsymbol{x}(k+1) = \mathsf{T}_\theta \boldsymbol{x}(k) = (1-\theta)\boldsymbol{x}(k) + \theta\mathsf{T}(\boldsymbol{x}(k)).$$

Let us consider the following matrix:

$$A = \begin{bmatrix} -0.12 & -0.63 & -0.33 & +0.21 \\ +0.12 & +0.15 & +0.03 & +0.09 \\ -0.63 & -0.30 & +0.36 & +1.65 \\ -0.90 & -5.79 & +0.45 & -6.39 \end{bmatrix}.$$

Note that for $\theta = 1$ the iteration surely does not converge because the matrix A has an eigenvalue outside the unitary circle, that is $\lambda = -6.4428$.

For $p = 2$, one can verify that the operator T is not κ-strictly pseudocontractive w.r.t. $\|\cdot\|_2$ for any $\kappa \in (0,1)$ by the following choice of vectors:

$$\boldsymbol{x} = \begin{bmatrix} +3.34 \\ -4.82 \\ +4.87 \\ +1.05 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} +3.42 \\ -1.86 \\ +0.18 \\ -1.25 \end{bmatrix},$$

which leads to a value of $\kappa > 1.5$. Instead, for $p = 4$ we have empirically verified that the operator T is κ-strictly pseudocontractive w.r.t. $\|\cdot\|_4$ with $\kappa \approx 0.972$. Thus, the forward step method converges for $\theta < (1-\kappa)/p \approx 0.007$ according to Theorem 2.

C. Application example: nonlinear Laplacian dynamics

Consider a network of n agents with discrete-time dynamics seeking consensus via the following nonlinear protocol

$$\begin{aligned}\boldsymbol{x}(k+1) &= \boldsymbol{x}(k) - \theta_k f(L\boldsymbol{x}(k)), \\ &= (1-\theta_k)\boldsymbol{x}(k) + \theta_k \underbrace{(-\boldsymbol{x}(k) + f(L\boldsymbol{x}(k)))}_{\mathsf{T}(\boldsymbol{x}(t))},\end{aligned}$$

where $L \in \mathbb{R}^{n\times n}$ is the Laplacian matrix associated to the graph describing the interactions among the agents, and $f : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear operator such that $f = [\dots, f_i, \dots]$ with $f_i : \mathbb{R} \to \mathbb{R}$. In this example, we consider the case all f_i for $i = 1, \dots, n$ are the same saturating function given by

$$f_i(x) = \frac{1 - e^{-mx}}{1 + e^{-mx}}, \qquad m \geq 0.$$

Note that for $m = 2$ the above reduces to the hyperbolic tangent function and for $m \to \infty$ it approximates the sign function; from now on we consider $m = 10$. For $p = 2$, one can verify that the operator T is not κ-strictly pseudocontractive w.r.t. $\|\cdot\|_2$ for any $\kappa \in (0, 1)$ by the following choice of vectors:

$$\boldsymbol{x} = \begin{bmatrix} +0.2 \\ -1.0 \\ +1.0 \\ +0.3 \end{bmatrix}, \qquad \boldsymbol{y} = \begin{bmatrix} -0.1 \\ +1.3 \\ -0.6 \\ -0.6 \end{bmatrix},$$

Instead, for $p > 5$ we have empirically verified that the operator T is κ-strictly pseudocontractive w.r.t. $\|\cdot\|_p$ for some $\kappa \in (0, 1)$. Thus, the agents could employ a vanishing time-varying sequence θ_k as in eq. (7) and converge to a consensus according to Corollary 2. In this case, it is not necessary that the agents know the constant κ of strict pseudocontractivity of the operator T, i.e., they do not need to know (as typical is the case) the Laplacian matrix L.

IV. Discussion and Future Directions

The class of κ-strictly pseudocontractive operators, a superclass of nonexpansive operators, has attracted peculiar attention because it leads to generalized convergence results of fixed-point iterations, such as the Krasnoselksij iteration. In this work, we have provided the tightest condition for the convergence of the Krasnoselskij iteration on strict pseudocontractive operators. Our analysis holds for spaces $\mathcal{S}_p = (\mathbb{R}^n, \|\cdot\|_p)$ with $p \in [0, \infty]$, that are all Banach spaces, with $\mathcal{S}_2$ being the only Hilbert space. Notably, for $p = 2$, κ-strict pseudocontractivity of a mapping T is necessary and sufficient for nonexpansiveness of the averaged mapping $\mathsf{T}_\theta = (1-\theta)\mathsf{Id} + \theta\mathsf{T}$ with $\theta = 1 - \kappa$. On the other hand, for $p \neq 2$ this is not anymore the case, but there exists a sufficiently small θ such that T_θ becomes nonexpansive given that T is strict pseudocontractive. Finally, for $p \in \{1, \infty\}$, there exists no such a θ. In our opinion, the fact that moving from the Hilbert space $\mathcal{S}_2$ to a Banach space $\mathcal{S}_p$ with $p \neq 2$ the relation between strict pseudocontractivity of T and nonexpansiveness of T_θ becomes progressively less strong, and that for $p \in \{1, \infty\}$ it is completely lost, suggests two different lines of research:

1) Consider weighted norms $\|\cdot\|_{p,P}$ where P is a symmetric and positive semidefinite matrix;
2) Search for alternative properties to strict pseudocontractivity in Banach spaces.

References

[1] H. H. Bauschke and P. L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, 2nd ed. Cham, Switzerland: Springer, 2017.

[2] M. A. Krasnosel'skii, "Two comments on the method of successive approximations," *Usp. Math. Nauk*, vol. 10, 1955.

[3] R. P. Agarwal, D. O'Regan, and D. Sahu, *Fixed point theory for Lipschitzian-type mappings with applications*. Springer, 2009, vol. 6.

[4] V. Berinde and F. Takens, *Iterative approximation of fixed points*. Springer, 2007, vol. 1912.

[5] M. Edelstein, "A remark on a theorem of ma krasnoselski," *Amer. Math. Monthly*, vol. 73, 1966.

[6] S. Ishikawa, "Fixed points and iteration of a nonexpansive mapping in a Banach space," *Proceedings of the American Mathematical Society*, vol. 59, no. 1, 1976.

[7] G. Marino and H.-K. Xu, "Weak and strong convergence theorems for strict pseudo-contractions in hilbert spaces," *Journal of Mathematical Analysis and Applications*, vol. 329, no. 1, 2007.

[8] G. Belgioioso, F. Fabiani, F. Blanchini, and S. Grammatico, "On the convergence of discrete-time linear systems: A linear time-varying mann iteration converges iff its operator is strictly pseudocontractive," *IEEE Control Systems Letters*, vol. 2, no. 3, 2018.

[9] H. Zhang and Y. Su, "Convergence theorems for strict pseudo-contractions in q-uniformly smooth Banach spaces," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 71, no. 10, 2009.

[10] H. Zhang and Y. Su, "Strong convergence theorems for strict pseudo-contractions in q-uniformly smooth Banach spaces," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 70, no. 9, 2009.

[11] G. Cai and C. song Hu, "Strong convergence theorems of a general iterative process for a finite family of λ-strict pseudo-contractions in q-uniformly smooth Banach spaces," *Computers & Mathematics with Applications*, vol. 59, no. 1, 2010.

[12] H. Y. Zhou, "Convergence theorems for λ-strict pseudo-contractions in q-uniformly smooth Banach spaces," *Acta Mathematica Sinica, English Series*, vol. 26, no. 4, 2010.

[13] C. Chidume and N. Shahzad, "Weak convergence theorems for a finite family of strict pseudocontractions," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 72, no. 3-4, 2010.

[14] P. Cholamjiak and S. Suantai, "Weak convergence theorems for a countable family of strict pseudocontractions in Banach spaces," *Fixed Point Theory and Applications*, 2010.

[15] P. Cholamjiak and S. Suantai, "Strong convergence for a countable family of strict pseudocontractions in q-uniformly smooth Banach spaces," *Computers & Mathematics with Applications*, vol. 62, no. 2, 2011.

[16] D. Deplano, M. Franceschelli, and A. Giua, "Lyapunov-free analysis for consensus of nonlinear discrete- time multi-agent systems," in *IEEE Conference on Decision and Control (CDC)*, 2018.

[17] V. Srivastava, J. Moehlis, and F. Bullo, "On bifurcations in nonlinear consensus networks," *Journal of Nonlinear Science*, vol. 21, 2011.

[18] D. Deplano, M. Franceschelli, and A. Giua, "A nonlinear perron–frobenius approach for stability and consensus of discrete-time multi-agent systems," *Automatica*, vol. 118, 2020.

[19] R. Bonetto and H. J. Kojakhmetov, "Nonlinear laplacian dynamics: Symmetries, perturbations, and consensus," *arXiv preprint arXiv:2206.04442*, 2022.

[20] S. Grammatico, F. Parise, M. Colombino, and J. Lygeros, "Decentralized convergence to nash equilibria in constrained deterministic mean field control," *IEEE Transactions on Automatic Control*, vol. 61, no. 11, 2015.

[21] S. Grammatico, "Dynamic control of agents playing aggregative games with coupling constraints," *IEEE Transactions on Automatic Control*, vol. 62, no. 9, 2017.

[22] P. Giselsson and S. Boyd, "Linear convergence and metric selection for douglas-rachford splitting and admm," *IEEE Transactions on Automatic Control*, vol. 62, no. 2, 2016.

[23] L. Pavel, "Distributed gne seeking under partial-decision information over networks via a doubly-augmented operator splitting approach," *IEEE Transactions on Automatic Control*, vol. 65, no. 4, 2019.

[24] L. Guo, X. Shi, S. Yang, and J. Cao, "Disa: A dual inexact splitting algorithm for distributed convex composite optimization," *IEEE Transactions on Automatic Control*, 2023.

[25] A. Davydov, S. Jafarpour, A. V. Proskurnikov, and F. Bullo, "Non-euclidean monotone operator theory and applications," *arXiv preprint arXiv:2303.11273*, 2023.

[26] D. Deplano, M. Franceschelli, and A. Giua, "Novel stability conditions for nonlinear monotone systems and consensus in multi-agent networks," *IEEE Transactions on Automatic Control*, 2023.

[27] D. Sahu and A. Petruşel, "Strong convergence of iterative methods by strictly pseudocontractive mappings in Banach spaces," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 74, no. 17, 2011.

[28] J. K. Hunter and B. Nachtergaele, *Applied analysis*. World Scientific Publishing Company, 2001.

[29] A. Beurling and A. Livingston, "A theorem on duality mappings in Banach spaces," *Arkiv för Matematik*, vol. 4, no. 5, 1962.

[30] S. Reich, "An iterative procedure for constructing zeros of accretive sets in Banach spaces," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 2, no. 1, 1978.

[31] H.-K. Xu, "Inequalities in Banach spaces with applications," *Nonlinear Analysis: Theory, Methods & Applications*, vol. 16, no. 12, 1991.

APPENDIX

A. Proof of Lemma 1

By means of the Hölder's inequality it holds,

$$|\boldsymbol{y}^\top \boldsymbol{x}| = \left|\sum_i^n x_i y_i\right| \le \sum_i^n |x_i y_i| \le \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q,$$

which leads to an upper bound to the norm,

$$\|\mathsf{L}_{\boldsymbol{y}}\|_p^* = \sup_{\|\boldsymbol{x}\|_p \le 1} |\mathsf{L}_{\boldsymbol{y}}(\boldsymbol{x})| \le \sup_{\|\boldsymbol{x}\|_p \le 1} \|\boldsymbol{x}\|_p \|\boldsymbol{y}\|_q \le \|\boldsymbol{y}\|_q.$$

To prove the converse inequality, we should estimate the supremum from below. We start by considering $p \in (1, \infty)$ and the vector $\tilde{\boldsymbol{x}} = [\tilde{x}_1, \dots, \tilde{x}_n]^\intercal$ defined component-wise by

$$\tilde{x}_i = \frac{|y_i|^{q-2} y_i}{\|\boldsymbol{y}\|_q^{q-1}}.$$

By simple manipulation one can verify that $\|\tilde{\boldsymbol{x}}\|_p = 1$, i.e., $\tilde{\boldsymbol{x}}$ belongs to the constraint set $\|\boldsymbol{x}\|_p \le 1$ of the supremum function, and thus we can write

$$\|\mathsf{L}_{\boldsymbol{y}}\|_p^* \ge |\boldsymbol{y}^\top \tilde{\boldsymbol{x}}| = \left|\sum_{i=1}^n \tilde{x}_i y_i\right| = \frac{\sum_{i=1}^n |y_i|^q}{\|\boldsymbol{y}\|_q^{q-1}} = \frac{\|\boldsymbol{y}\|_q^q}{\|\boldsymbol{y}\|_q^{q-1}} = \|\boldsymbol{y}\|_q.$$

Whereas, for $p = \infty$ we consider the vector of ones $\tilde{\boldsymbol{x}} = \mathbf{1}$ which is such that $\|\tilde{\boldsymbol{x}}\|_\infty = 1$. Thus, we get

$$\|\mathsf{L}_{\boldsymbol{y}}\|_p^* \ge |\boldsymbol{y}^\top \tilde{\boldsymbol{x}}| = \left|\sum_{i=1}^n \tilde{x}_i y_i\right| = \sum_{i=1}^n \tilde{x}_i |y_i| = \left|\sum_{i=1}^n y_i\right| = \|\boldsymbol{y}\|_1.$$

Finally, for $p = 1$ we let I be the set of indexes such that $|x_i| = \max |\boldsymbol{x}|$, i.e., $I = \{i \in [1, n] : |y_i| = \max |\boldsymbol{y}|\}$, we let $|I|$ be the cardinality of the set and consider the vector $\tilde{\boldsymbol{x}}$ defined component-wise by

$$\tilde{x}_i = \begin{cases} 1/|I| & \text{if} \quad i \in I \\ 0 & \text{otherwise} \end{cases},$$

which clearly satisfies $\|\tilde{\boldsymbol{x}}\|_1 = 1$. Thus, we get

$$\|\mathsf{L}_{\boldsymbol{y}}\|_p^* \ge |\boldsymbol{y}^\top \tilde{\boldsymbol{x}}| = \left|\sum_{i=1}^n \tilde{x}_i y_i\right| = \left|\sum_{i \in I} \frac{y_i}{|I|}\right| = \sum_{i \in I} \frac{|y_i|}{|I|} = \|\boldsymbol{y}\|_\infty.$$

Since $\|\boldsymbol{y}\|_q \le \|\mathsf{L}_{\boldsymbol{y}}\|_p^* \le \|\boldsymbol{y}\|_q$, for all p, the proof is completed.

B. Proof of Lemma 2

Given $r = \min\{2, p\}$, we are going to prove that $\boldsymbol{y} = \mathsf{j}_r(\boldsymbol{x})$ as in eq. (3) belongs to $\mathsf{J}_r(\boldsymbol{x})$ for any $\boldsymbol{x} \in \mathcal{X}$. To do so, we need to verify the following two conditions:

a) $\|\boldsymbol{x}\|_p^r = \boldsymbol{x}^\intercal \boldsymbol{y}$;

b) $\|\boldsymbol{x}\|_p^{r-1} = \|\boldsymbol{y}\|_q$.

We go through the proof case by case:

- *Condition a) for* $p \in [1, \infty)$]:

$$\boldsymbol{x}^\intercal \boldsymbol{y} = \frac{\boldsymbol{x}^\intercal(\text{sign}(\boldsymbol{x}) \circ |\boldsymbol{x}|^{p-1})}{\|\boldsymbol{x}\|_p^{p-r}} = \frac{|\boldsymbol{x}|^\intercal |\boldsymbol{x}|^{p-1}}{\|\boldsymbol{x}\|_p^{p-r}} = \\ = \frac{\sum_{i=1}^n |x_i|^p}{\|\boldsymbol{x}\|_p^{p-r}} = \frac{\|\boldsymbol{x}\|_p^p}{\|\boldsymbol{x}\|_p^{p-r}} = \|\boldsymbol{x}\|_p^r.$$

- *Condition b) for* $p \in [1, \infty)$:

$$\|\boldsymbol{y}\|_q^r = \left\|\frac{\text{sign}(\boldsymbol{x}) \circ |\boldsymbol{x}|^{p-1}}{\|\boldsymbol{x}\|_p^{p-r}}\right\|_q^r = \\ = \left[\frac{\|\boldsymbol{x}^{p-1}\|_q}{\|\boldsymbol{x}\|_p^{p-r}}\right]^r = \left[\frac{\left(\sum_{i=1}^n |x_i|^{(p-1)q}\right)^{1/q}}{\|\boldsymbol{x}\|_p^{p-r}}\right]^r = \\ = \left[\frac{\left(\sum_{i=1}^n |x_i|^p\right)^{\frac{p-1}{p}}}{\|\boldsymbol{x}\|_p^{p-r}}\right]^r = \left[\frac{\|\boldsymbol{x}\|_p^{p-1}}{\|\boldsymbol{x}\|_p^{p-r}}\right]^r = \|\boldsymbol{x}\|_p^r.$$

- *Condition a) for* $p = \infty$ such that $r = 2$: let n_∞ be the number of entries of $\boldsymbol{x}$ such that $|x_i| = \max |\boldsymbol{x}|$, then it holds

$$\boldsymbol{x}^\intercal \boldsymbol{y} = \frac{\boldsymbol{x}^\intercal(\boldsymbol{x} \circ \boldsymbol{x}_\infty)}{\mathbf{1}^\intercal \boldsymbol{x}_\infty} = \frac{(\boldsymbol{x} \circ \boldsymbol{x})^\intercal \boldsymbol{x}_\infty}{\mathbf{1}^\intercal \boldsymbol{x}_\infty} = \frac{(\boldsymbol{x}^2)^\intercal \boldsymbol{x}_\infty}{\mathbf{1}^\intercal \boldsymbol{x}_\infty} = \\ = \frac{n_\infty \max \boldsymbol{x}^2}{n_\infty} = \|\boldsymbol{x}\|_\infty^2,$$

- *Condition b) for* $p = \infty$ such that $r = 2$: let n_∞ be the number of entries of $\boldsymbol{x}$ such that $|x_i| = \max |\boldsymbol{x}|$, then

$$\|\boldsymbol{y}\|_1^2 = \left\|\frac{\boldsymbol{x} \circ \boldsymbol{x}_\infty}{\mathbf{1}^\intercal \boldsymbol{x}_\infty}\right\|_1^2 = \frac{1}{n_\infty^2} \|\boldsymbol{x} \circ \boldsymbol{x}_\infty\|_1^2 = \\ = \frac{1}{n_\infty^2} \left[\sum_{i=1}^n |x_i x_{\infty,i}|\right]^2 = \frac{n_\infty^2}{n_\infty^2} \max \boldsymbol{x}^2 = \|\boldsymbol{x}\|_\infty^2.$$

This completes the proof.

C. Proof of Lemma 3

We first note that for $p \in (1, \infty)$ the generalized duality mapping J_r mapping is in one-to-tone relation with the so-called normalized duality mapping J_2 since $\mathsf{J}_r(\boldsymbol{x}) = \|\boldsymbol{x}\|_p^{p-2} \mathsf{J}_2(\boldsymbol{x})$ (cfr. [31]). Thus, the sufficiency of the statement is due to the strict convexity of the Banach space S_p (cfr. with Definition 1.10 and Remark 1 on Page 9 of [4]). The necessity follows from the next two counter examples:

(Case $p = 1$ *and* $r = 1$*)* Let $\boldsymbol{x} = [1, 0]^\intercal \in \mathbb{R}^2$, then all points $\boldsymbol{y} = [1, \theta]^\intercal \in \mathbb{R}^2$ with $|\theta| \le 1$ belongs to $\mathsf{J}_1(\boldsymbol{x})$, i.e., $\boldsymbol{y}^\intercal \boldsymbol{x} = \|\boldsymbol{x}\|_1 = \|\boldsymbol{y}\|_\infty = 1$.

(Case $p = \infty$ *and* $r = 2$*)* Let $\boldsymbol{x} = [2, 0, 2]^\intercal \in \mathbb{R}^2$, then all points $\boldsymbol{y} = [\theta, 0, 2 - \theta]^\intercal \in \mathbb{R}^2$ with $\theta \in [0, 2]$ belongs to $\mathsf{J}_2(\boldsymbol{x})$, i.e., $\boldsymbol{y}^\intercal \boldsymbol{x} = \|\boldsymbol{x}\|_\infty^2 = \|\boldsymbol{y}\|_1^2 = 4$.