Non-Euclidean Monotone Operator Theory with Applications to Recurrent Neural Networks

(Article begins on next page)

Non-Euclidean Monotone Operator Theory with Applications to Recurrent Neural Networks

Alexander Davydov*, Saber Jafarpour*, Anton V. Proskurnikov, and Francesco Bullo

***Abstract*— We provide a novel transcription of monotone operator theory to the non-Euclidean finite-dimensional spaces ℓ_1 and ℓ_∞. We first establish properties of mappings which are monotone with respect to the non-Euclidean norms ℓ_1 or ℓ_∞. In analogy with their Euclidean counterparts, mappings which are monotone with respect to a non-Euclidean norm are amenable to numerous algorithms for computing their zeros. We demonstrate that several classic iterative methods for computing zeros of monotone operators are directly applicable in the non-Euclidean framework. We present a case-study in the equilibrium computation of recurrent neural networks and demonstrate that casting the computation as a suitable operator splitting problem improves convergence rates.**

I. Introduction

Monotone operator methods have become prevalent to solve problems in optimization and control [4], [20], game theory [17], systems analysis [5], and to better understand machine learning models [8], [22]. However, monotone operator techniques are primarily based on the theory of Hilbert and Euclidean spaces, while many problems are well-posed or better-suited for analysis in a Banach space or finite-dimensional non-Euclidean space. For example, in machine learning, it is known that robustness analysis of artificial neural networks is naturally performed via the ℓ_∞ norm and that such a norm is most appropriate for high-dimensional input data. Additionally, in the field of robust control, $\mathcal{H}_\infty$ techniques are naturally stated over an infinite-dimensional Banach space, so monotone operator techniques do not apply.

Problem description and motivation: In this paper, we aim to provide a natural transcription of many monotone operator techniques for computing zeros of monotone operators for operators which are naturally "monotone" with respect to an ℓ_1 or ℓ_∞ norm in a finite-dimensional space.

Monotone operator theory is a fertile field of nonlinear functional analysis that generalizes the notion of monotone functions on $\mathbb{R}$ to mappings on arbitrary Hilbert spaces and examines the properties of such maps. In particular, an integral component of monotone operator theory is the design of algorithms to compute zeros of monotone operators. This aspect makes monotone operator theory compatible with convex optimization since the subdifferential of any convex function is monotone and minimizing a convex function is synonymous with finding a zero of its subdifferential. To this end, there has been an extensive amount of work in the last decade in applying monotone operator theory to convex optimization; e.g., see [7], [18], [19].

Through the lens of duality theory, the theory of dissipative and accretive operators on Banach spaces mirrors monotone operators on Hilbert spaces [**?**]. Despite these parallels, the theory of dissipative and accretive operators has largely focused on iteratively computing solutions of integral equations and PDEs in L_p spaces for $p \neq 2$; see [6] for a relevant book. Moreover, many works in this direction focus on Banach spaces that additionally have a uniformly smooth or uniformly convex structure; this structure is not possesed by the finite-dimensional ℓ_1 and ℓ_∞ spaces. Ultimately, in contrast to monotone operator theory, the theory of dissipative and accretive operators has found far fewer direct applications to systems, control, and machine learning.

A notion similar to a monotone operator in a Hilbert space is that of a contracting vector field [16]. In fact, a vector field $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is contracting with respect to an ℓ_2 norm if and only if the negative vector field $-\mathsf{F}$ is monotone when thought of as an operator. However, vector fields are not restricted to being contracting with respect to a Euclidean norm. In general, a vector field may be contracting with respect to a non-Euclidean norm but not a Euclidean one [1]. Recently, there has been an increased interest in studying vector fields that are contracting with respect to the non-Euclidean norms ℓ_1 and ℓ_∞ [2], [10], [11]. Due to the connection between monotone operators and contracting vector fields, it is of interest to explore the properties of operators that may be thought of as monotone with respect to an ℓ_1 or ℓ_∞ norm.

Contributions: To facilitate the application of monotone operator theory techniques to problems naturally arising in non-Euclidean spaces, we propose a novel non-Euclidean monotone operator framework based on the theory of logarithmic norms [21]. We use the logarithmic norm as a substitute for inner-products in Hilbert spaces and we demonstrate that many results from monotone operator theory directly carry over to their non-Euclidean counterpart. Specifically, we show that the resolvent and reflected resolvent operators of a non-Euclidean monotone operator have properties analogous to those arising in Euclidean spaces.

Second, we demonstrate that classical iterative algorithms such as the forward step method and proximal point method

* These authors contributed equally.

This material is based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 2139319 and AFOSR grant FA9550-22-1-0059.

Alexander Davydov and Francesco Bullo are with the Center for Control, Dynamical Systems, and Computation, University of California, Santa Barbara, 93106-5070, USA. {davydov, bullo}@ucsb.edu.

Saber Jafarpour is with the School of Electrical and Computer Engineering, Georgia Institute of Technology. saber@gatech.edu.

Anton V. Proskurnikov is with the Department of Electronics and Telecommunications, Politecnico di Torino, Turin, Italy. anton.p.1982@ieee.org.

allow us to compute zeros of non-Euclidean monotone operators in a manner identical to the procedure for traditional monotone operators. We present estimates for Lipschitz constants of these iterative methods and demonstrate that, for diagonally weighted ℓ_1 and ℓ_∞ norms, these algorithms achieve improved rates of convergence compared to their Euclidean counterparts. As a clear distinction from the classical theory, we prove that the forward step method is convergent for an operator which is (weakly) monotone with respect to an ℓ_1 or ℓ_∞ norm, but that the method need not converge if the operator is monotone with respect to a Euclidean norm. This result is analogous to the result on weakly-contracting ODEs as in [13, Theorem 21].

Third, we study operator splitting methods. We prove that the forward-backward, Peaceman-Rachford, and Douglas-Rachford splitting algorithms all apply in our framework and that improved convergence may be achieved for these non-Euclidean norms compared to their Euclidean counterparts.

Fourth, as an application, we present methods to compute equilibria for recurrent neural networks. We extend the recent work of [**?**], [14] to demonstrate that our non-Euclidean monotone operator theory is readily applicable and can provide accelerated convergence of iterations when viewing the problem of computing an equilibrium as an appropriate operator splitting problem. Finally, we present numerical simulations presenting rates of convergence of the different iterations when applied to this problem.

II. Preliminaries

A. Notations

For differentiable $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$, we let $D\mathsf{F}(x) := \frac{\partial \mathsf{F}(x)}{\partial x} \in \mathbb{R}^{n\times n}$ denote its Jacobian evaluated at x. For an arbitrary mapping F, we let $\mathrm{Dom}(\mathsf{F})$ be its domain. For $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$, we let $\mathrm{Zero}(\mathsf{F}) := \{x \in \mathbb{R}^n \mid \mathsf{F}(x) = 0\}$ and $\mathrm{Fix}(\mathsf{F}) = \{x \in \mathbb{R}^n \mid \mathsf{F}(x) = x\}$ be the sets of zeros of F and fixed points of F, respectively. We let $\mathsf{Id} : \mathbb{R}^n \to \mathbb{R}^n$ be the identity map and $I_n \in \mathbb{R}^{n\times n}$ be the $n \times n$ identity matrix.

B. Norms and Logarithmic Norms

Definition 1 (Logarithmic norm)**.** *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ and its corresponding induced norm on $\mathbb{R}^{n\times n}$. The* logarithmic norm *of a matrix $A \in \mathbb{R}^{n\times n}$ is*

$$\mu(A) := \lim_{h\to 0^+} \frac{\|I_n + hA\| - 1}{h}. \tag{1}$$

It is well-known that this limit is well posed. We refer to [12] for properties enjoyed by log norms.

We will be specifically interested in diagonally weighted ℓ_1 and ℓ_∞ norms defined by

$$\|x\|_{1,[\eta]} = \sum_i \eta_i |x_i| \qquad \text{and} \qquad \|x\|_{\infty,[\eta]^{-1}} = \max_i \frac{1}{\eta_i}|x_i|,$$

where, given a positive vector $\eta \in \mathbb{R}^n_{>0}$, we use $[\eta]$ to denote the diagonal matrix with diagonal entries η. For $A \in \mathbb{R}^{n\times n}$, the corresponding induced and log norms are[1]

$$\|A\|_{\infty,[\eta]^{-1}} = \max_{i\in\{1,\dots,n\}} \sum_{j=1}^n \frac{\eta_j}{\eta_i}|a_{ij}|,$$
$$\mu_{\infty,[\eta]^{-1}}(A) = \max_{i\in\{1,\dots,n\}} \Big(a_{ii} + \sum\nolimits_{j=1,j\neq i}^n |a_{ij}| \frac{\eta_j}{\eta_i}\Big),$$
$$\|A\|_{1,[\eta]} = \|A^\top\|_{\infty,[\eta]^{-1}}, \quad \mu_{1,[\eta]}(A) = \mu_{\infty,[\eta]^{-1}}(A^\top).$$

C. Contractions, nonexpansive maps, Banach-Picard and Krasnosel'skii–Mann iterations

For the remainder of the paper, we assume all mappings are continuously differentiable unless otherwise stated.

Definition 2 (Lipschitz continuity)**.** *Let $\|\cdot\|$ be a norm and $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be a map. F is* Lipschitz continuous *with constant* $\mathrm{Lip}(\mathsf{F}) \in \mathbb{R}_{\geq 0}$ *if for all $x_1, x_2 \in \mathbb{R}^n$*

$$\|\mathsf{F}(x_1) - \mathsf{F}(x_2)\| \leq \mathrm{Lip}(\mathsf{F})\|x_1 - x_2\|. \tag{2}$$

Definition 3 (Contractions and nonexpansive maps)**.** *Let $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz with respect to a norm $\|\cdot\|$. We say*

(i) T *is a* contraction *if* $\mathrm{Lip}(\mathsf{T}) \in [0,1[$,
(ii) T *is* nonexpansive *if* $\mathrm{Lip}(\mathsf{T}) = 1$.

Definition 4 (Krasnosel'skii–Mann iterations [3, Section 5.2])**.** *Let $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^n$ be nonexpansive with respect to a norm $\|\cdot\|$. For $\theta \in \,]0,1[$, the* Krasnosel'skii–Mann iterations *applied to T defines the sequence $\{x_k\}_{k=0}^\infty$ by*

$$x_{k+1} = (1-\theta)x_k + \theta\mathsf{T}(x_k). \tag{3}$$

Lemma 5 (Convergence of Krasnosel'skii–Mann iterations [9])**.** *Let $\mathsf{T} : \mathbb{R}^n \to \mathbb{R}^n$ be nonexpansive with respect to a norm $\|\cdot\|$ and consider the Krasnosel'skii–Mann iterations as in (3). Suppose* $\mathrm{Fix}(\mathsf{T}) \neq \emptyset$ *and let $x^* \in \mathrm{Fix}(\mathsf{T})$. Then*

$$\|x_k - \mathsf{T}(x_k)\| \leq \frac{2\|x_0 - x^*\|}{\sqrt{k\pi\theta(1-\theta)}}. \tag{4}$$

In particular, $\|x_k - \mathsf{T}(x_k)\| \to 0$ as $k \to \infty$.

III. Non-Euclidean Monotone Operators

A. Definitions and Properties

Definition 6 (Non-Euclidean monotone operator)**.** *A continuously differentiable operator $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is strongly monotone with monotonicity parameter $c > 0$ with respect to a norm $\|\cdot\|$ on $\mathbb{R}^n$ provided for all $x \in \mathbb{R}^n$,*

$$-\mu(-D\mathsf{F}(x)) \geq c. \tag{5}$$

If the inequality holds with $c = 0$, we say F is monotone[2] *(or weakly monotone) with respect to $\|\cdot\|$.*

Remark 7 (Comparison to the Euclidean case)**.** *For an operator $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$, let $\|\cdot\|_2$ be the Euclidean norm with corresponding inner product $\langle\cdot,\cdot\rangle$. Then following [3, Definition 20.1], F is monotone with respect to $\|\cdot\|_2$ if*

$$\langle \mathsf{F}(x) - \mathsf{F}(y), x - y\rangle \geq 0, \qquad \textit{for all } x, y \in \mathbb{R}^n. \tag{6}$$

[1] We note also that for the Euclidean norm $\|\cdot\|_2$, the corresponding log norm is $\mu_2(A) = \frac{1}{2}\lambda_{\max}(A + A^\top)$.

[2] If F is only locally Lipschitz, we ask that (5) holds almost everywhere.

If F is continuously differentiable, (6) *is known to be equivalent to (e.g., [18]) $D\mathsf{F}(x) + D\mathsf{F}(x)^\top \succeq 0$, or equivalently $-\mu_2(-D\mathsf{F}(x)) \geq 0$, which coincides with Definition 6.*

By subadditivity of μ, a sum of operators which are monotone with respect to the same norm is also monotone. Additionally, if F is (strongly) monotone with monotonicity parameter $c \geq 0$, then for any $\alpha \geq 0$, $\mathsf{Id} + \alpha\mathsf{F}$ is strongly monotone with monotonicity parameter $1 + \alpha c$.

Remark 8 (Connection with contracting vector fields [16])**.** *A mapping $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is strongly contracting with rate $c > 0$ with respect to a norm $\|\cdot\|$ on $\mathbb{R}^n$ provided $\mu(D\mathsf{F}(x)) \leq -c$, for all $x \in \mathbb{R}^n$. If $c = 0$, we say F is weakly contracting with respect to $\|\cdot\|$. Clearly, F is (strongly) monotone if and only if $-\mathsf{F}$ is (strongly) contracting.*

Lemma 9. *Let $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be globally Lipschitz with respect to a diagonally-weighted ℓ_1 or ℓ_∞ norm $\|\cdot\|$ with constant $\mathrm{Lip}(\mathsf{F}) = \ell$. If F is (possibly strongly) monotone with respect to $\|\cdot\|$ with monotonicity parameter $c \geq 0$, then*

$$\mathrm{Lip}(\mathsf{Id} - \alpha\mathsf{F}) = 1 - \alpha c, \qquad \textit{for all } \alpha \in \left]0, (\mathrm{diagL}(\mathsf{F}))^{-1}\right],$$

where $\mathrm{diagL}(\mathsf{F}) := \sup_{x\in\mathbb{R}^n} \max_{i\in\{1,\dots,n\}} (D\mathsf{F}(x))_{ii} \leq \ell$.

Note that for Euclidean norms, if F is monotone, but not strongly monotone, then $(\mathsf{Id} - \alpha\mathsf{F})$ need not be nonexpansive for any $\alpha > 0$. Indeed, consider $\mathsf{F}(x) = \left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right) x$, which is monotone with respect to the ℓ_2 norm, but $(\mathsf{Id} - \alpha\mathsf{F})$ is expansive for every $\alpha > 0$.

B. Resolvent and reflected resolvent operators

Definition 10 (Resolvent and reflected resolvent)**.** *Let $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be a mapping and $\alpha > 0$. The resolvent of $\alpha\mathsf{F}$ is defined as*

$$\mathsf{J}_{\alpha\mathsf{F}} := (\mathsf{Id} + \alpha\mathsf{F})^{-1}. \tag{7}$$

The reflected resolvent of $\alpha\mathsf{F}$ is $\mathsf{R}_{\alpha\mathsf{F}} := 2\mathsf{J}_{\alpha\mathsf{F}} - \mathsf{Id}$.

Note for any $\alpha > 0$, $\mathrm{Zero}(\mathsf{F}) = \mathrm{Fix}(\mathsf{J}_{\alpha\mathsf{F}}) = \mathrm{Fix}(\mathsf{R}_{\alpha\mathsf{F}})$.

Lemma 11 (Lipschitz constant of the resolvent operator)**.** *Suppose $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is (strongly) monotone with parameter $c \geq 0$. Then for every $\alpha > 0$,*

$$\mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) = \frac{1}{1 + \alpha c}. \tag{8}$$

Theorem 12 (Lipschitz constant of the reflected resolvent)**.** *Suppose $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is globally Lipschitz with respect to a diagonally weighted ℓ_1 or ℓ_∞ norm $\|\cdot\|$. Moreover, suppose F is (strongly) monotone with respect to $\|\cdot\|$ with monotonicity parameter $c \geq 0$. Then for $\alpha \in \left]0, (\mathrm{diagL}(\mathsf{F}))^{-1}\right[$,*

$$\mathrm{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) = \frac{1 - \alpha c}{1 + \alpha c} \leq 1. \tag{9}$$

IV. Finding Zeros of Non-Euclidean Monotone Operators

Consider the problem of finding an $x \in \mathbb{R}^n$ that satisfies $\mathsf{F}(x) = 0$, where F is monotone. We present several well-known algorithms for finding zeros of monotone operators (see, e.g., [18]) and show how the non-Euclidean monotone operator framework allows the same algorithms to compute zeros of non-Euclidean monotone operators.

Algorithm 13 (Forward step method)**.** *The* forward step method *corresponds to the fixed point iteration*

$$x_{k+1} = (\mathsf{Id} - \alpha\mathsf{F})(x_k). \tag{10}$$

Theorem 14 (Forward step method convergence)**.** *Let $x_0 \in \mathbb{R}^n$. Suppose $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is globally Lipschitz with respect to a diagonally-weighted ℓ_1 or ℓ_∞ norm $\|\cdot\|$ and*

(i) F is strongly monotone with respect to $\|\cdot\|$ with monotonicity parameter $c > 0$. Then the sequence generated by (10) *converges to the unique $x^* \in \mathrm{Zero}(\mathsf{F})$ for every $\alpha \in \left]0, (\mathrm{diagL}(\mathsf{F}))^{-1}\right]$. Moreover, for every $k \in \mathbb{Z}_{\geq 0}$,*

$$\|x_{k+1} - x^*\| \leq (1 - \alpha c)\|x_k - x^*\|,$$

with convergence rate optimized at $\alpha = 1/\mathrm{diagL}(\mathsf{F})$.

(ii) F is monotone with respect to $\|\cdot\|$. Then if $\mathrm{Zero}(\mathsf{F}) \neq \emptyset$, (10) *converges to an element of $\mathrm{Zero}(\mathsf{F})$ for every $\alpha \in \left]0, (\mathrm{diagL}(\mathsf{F}))^{-1}\right[$.*

Algorithm 15 (Proximal point method)**.** *The* proximal point method *corresponds to the fixed point iteration*

$$x_{k+1} = \mathsf{J}_{\alpha\mathsf{F}}(x_k) = (\mathsf{Id} + \alpha\mathsf{F})^{-1}(x_k). \tag{11}$$

Theorem 16 (Proximal point method convergence)**.** *Let $x_0 \in \mathbb{R}^n$. Suppose $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is*

(i) strongly monotone with respect to $\|\cdot\|$ with monotonicity parameter $c > 0$. Then for any $x_0 \in \mathbb{R}^n$, the sequence generated by (11) *converges to the unique $x^* \in \mathrm{Zero}(\mathsf{F})$ for every $\alpha \in \left]0, \infty\right[$. Moreover, for every $k \in \mathbb{Z}_{\geq 0}$,*

$$\|x_{k+1} - x^*\| \leq \frac{1}{1 + \alpha c}\|x_k - x^*\|.$$

(ii) monotone and globally Lipschitz with respect to a diagonally weighted ℓ_1 or ℓ_∞ norm. Then if $\mathrm{Zero}(\mathsf{F}) \neq \emptyset$, (11) *converges to an element of $\mathrm{Zero}(\mathsf{F})$ for every $\alpha \in \left]0, \infty\right[$ and $x_0 \in \mathbb{R}^n$.*

Algorithm 17. *The* Cayley method *corresponds to the fixed point iteration*

$$x_{k+1} = \mathsf{R}_{\alpha\mathsf{F}}(x_k) = 2(\mathsf{Id} + \alpha\mathsf{F})^{-1}(x_k) - x_k. \tag{12}$$

Theorem 18 (Cayley method convergence)**.** *Let $x_0 \in \mathbb{R}^n$. Suppose $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is globally Lipschitz with respect to a diagonally-weighted ℓ_1 or ℓ_∞ norm $\|\cdot\|$ and*

(i) F is strongly monotone with respect to $\|\cdot\|$ with monotonicity parameter $c > 0$. Then for any $x_0 \in \mathbb{R}^n$, the sequence generated by (12) *converges to the unique $x^* \in \mathrm{Zero}(\mathsf{F})$ for every $\alpha \in \left]0, (\mathrm{diagL}(\mathsf{F}))^{-1}\right]$. Moreover, for every $k \in \mathbb{Z}_{\geq 0}$,*

$$\|x_{k+1} - x^*\| \leq \frac{1 - \alpha c}{1 + \alpha c}\|x_k - x^*\|,$$

with convergence rate optimized at $\alpha = 1/\mathrm{diagL}(\mathsf{F})$.

Algorithm	F strongly monotone and globally Lipschitz			
	ℓ_2		Diagonally weighted ℓ_1 or ℓ_∞	
	α range	Optimal Lip	α range	Optimal Lip
Forward step	$\left]0, \frac{2c}{\ell^2}\right[$	$1 - \frac{1}{2\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right)$	$\left]0, \frac{1}{\operatorname{diagL}(\mathsf{F})}\right]$	$1 - \frac{1}{\kappa_\infty}$
Proximal point	$]0, \infty[$	N/A	$]0, \infty[$	N/A
Cayley method	$]0, \infty[$	$1 - \frac{1}{2\kappa} + \mathcal{O}\left(\frac{1}{\kappa^2}\right)$	$\left]0, \frac{1}{\operatorname{diagL}(\mathsf{F})}\right]$	$1 - \frac{2}{\kappa_\infty} + \mathcal{O}\left(\frac{1}{\kappa_\infty^2}\right)$

TABLE I

STEP SIZE RANGES AND LIPSCHITZ CONSTANTS FOR ALGORITHMS FOR FINDING ZEROS OF MONOTONE OPERATORS. FOR F STRONGLY MONOTONE, LET c BE ITS MONOTONICITY PARAMETER, ℓ ITS APPROPRIATE LIPSCHITZ CONSTANT, AND $\operatorname{diagL}(\mathsf{F}) := \sup_{x\in\mathbb{R}^n} \max_{i\in\{1,\dots,n\}}(D\mathsf{F}(x))_{ii} \leq \ell$. ADDITIONALLY, $\kappa := \ell/c \geq 1$ AND $\kappa_\infty := \operatorname{diagL}(\mathsf{F})/c \in [1, \kappa]$. RANGES OF α AND OPTIMAL LIPSCHITZ CONSTANTS FOR THE EUCLIDEAN CASE ARE PROVIDED IN [18]. WE DO NOT ASSUME THAT THE STRONGLY MONOTONE F IS THE GRADIENT OF A STRONGLY CONVEX FUNCTION.

(ii) F *is monotone with respect to* $\|\cdot\|$. *Then if* $\operatorname{Zero}(\mathsf{F}) \neq \emptyset$, *the averaged iterations*

$$x_{k+1} = \tfrac{1}{2}x_k + \tfrac{1}{2}\mathsf{R}_{\alpha\mathsf{F}}(x_k)$$

converge to an element of $\operatorname{Zero}(\mathsf{F})$ *for every* $\alpha \in]0, \infty[$.

We provide a comparison of the range of step sizes and Lipschitz constants as provided by the classical monotone operator theory [18] and Theorems 14, 16, and 18 in Table I.

V. FINDING ZEROS OF A SUM OF NON-EUCLIDEAN MONOTONE OPERATORS

In many instances, one may wish to execute the proximal point method, Algorithm 15, to compute a zero of a monotone operator $\mathsf{N} : \mathbb{R}^n \to \mathbb{R}^n$. However, in general, the implementation of the iteration (11) may be hindered by the difficulty in evaluating $\mathsf{J}_{\alpha\mathsf{N}}$. To remedy this issue, it is often assumed that N can be expressed as the sum of two monotone operators F and G where $\mathsf{J}_{\alpha\mathsf{G}}$ may be easy to compute and F satisfies some regularity condition. Alternatively, in some situations, decomposing $\mathsf{N} = \mathsf{F}+\mathsf{G}$ and finding $x \in \mathbb{R}^n$ such that $(\mathsf{F}+\mathsf{G})(x) = 0$ provides additional flexibility in choice of algorithm and may improve convergence rates.

Motivated by the above, we consider the problem of finding an $x \in \mathbb{R}^n$ such that $(\mathsf{F}+\mathsf{G})(x) = 0$, where $\mathsf{F}, \mathsf{G} : \mathbb{R}^n \to \mathbb{R}^n$ are monotone with respect to a diagonally weighted ℓ_1 or ℓ_∞ norm.

Algorithm 19 (Forward-backward splitting)**.** *Assume* $\alpha > 0$. *Then by [18, Section 7.1]*

$$(\mathsf{F}+\mathsf{G})(x) = 0 \quad \Longleftrightarrow \quad x = (\mathsf{J}_{\alpha\mathsf{G}} \circ (\mathsf{Id} - \alpha\mathsf{F}))(x).$$

The forward-backward splitting method *corresponds to the fixed point iteration*

$$x_{k+1} = (\mathsf{J}_{\alpha\mathsf{G}} \circ (\mathsf{Id} - \alpha\mathsf{F}))(x_k). \tag{13}$$

Additionally, if both F *and* G *are monotone, define the* averaged forward-backward splitting iterations

$$x_{k+1} = \tfrac{1}{2}x_k + \tfrac{1}{2}(\mathsf{J}_{\alpha\mathsf{G}} \circ (\mathsf{Id} - \alpha\mathsf{F}))(x_k). \tag{14}$$

Theorem 20 (Forward-backward splitting convergence)**.** *Let* $x_0 \in \mathbb{R}^n$. *Suppose* $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ *is globally Lipschitz with respect to a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm* $\|\cdot\|$ *and* $\mathsf{G} : \mathbb{R}^n \to \mathbb{R}^n$ *is monotone with respect to the same norm.*

(i) If F *is strongly monotone with respect to* $\|\cdot\|$ *with monotonicity parameter* $c > 0$, *then the sequence generated by* (13) *converges to the unique* $x^* \in \operatorname{Zero}(\mathsf{F}+\mathsf{G})$ *for every* $\alpha \in]0, \frac{1}{\operatorname{diagL}(\mathsf{F})}]$. *Moreover, for every* $k \in \mathbb{Z}_{\geq 0}$,

$$\|x_{k+1} - x^*\| \leq (1 - \alpha c)\|x_k - x^*\|,$$

with convergence rate optimized at $\alpha = 1/\operatorname{diagL}(\mathsf{F})$.

(ii) If F *is monotone with respect to* $\|\cdot\|$ *and* $\operatorname{Zero}(\mathsf{F}+\mathsf{G}) \neq \emptyset$, *then* (14) *converges to an element of* $\operatorname{Zero}(\mathsf{F}+\mathsf{G})$ *for every* $\alpha \in]0, (\operatorname{diagL}(\mathsf{F}))^{-1}]$.

Algorithm 21 (Peaceman-Rachford and Douglas-Rachford splitting)**.** *Let* $\alpha > 0$. *Then by [18, Section 7.3],*

$$(\mathsf{F}+\mathsf{G})(x) = 0 \quad \Longleftrightarrow \quad (\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}})z = z \text{ and } x = \mathsf{J}_{\alpha\mathsf{G}}z. \tag{15}$$

The Peaceman-Rachford splitting method *corresponds to the fixed point iteration*

$$\begin{aligned} x_{k+1} &= \mathsf{J}_{\alpha\mathsf{G}}(z_k), \\ z_{k+1} &= z_k + 2\mathsf{J}_{\alpha\mathsf{F}}(2x_{k+1} - z_k) - 2x_{k+1}. \end{aligned} \tag{16}$$

If both F *and* G *are monotone, the term* $\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}}$ *in* (15) *is averaged to yield the fixed point equation*

$$\tfrac{1}{2}(\mathsf{Id} + \mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}})z = z \text{ and } x = \mathsf{J}_{\alpha\mathsf{G}}z. \tag{17}$$

The fixed point iteration corresponding to (17) *is called the* Douglas-Rachford splitting method *and is given by*

$$\begin{aligned} x_{k+1} &= \mathsf{J}_{\alpha\mathsf{G}}(z_k), \\ z_{k+1} &= z_k + \mathsf{J}_{\alpha\mathsf{F}}(2x_{k+1} - z_k) - x_{k+1}. \end{aligned} \tag{18}$$

Theorem 22 (Peaceman-Rachford and Douglas-Rachford splitting convergence)**.** *Let* $x_0 \in \mathbb{R}^n$. *Suppose both* $\mathsf{F}, \mathsf{G} : \mathbb{R}^n \to \mathbb{R}^n$ *are globally Lipschitz with respect to a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm* $\|\cdot\|$ *and (without loss of generality)* G *is monotone with respect to the same norm.*

(i) If F *is strongly monotone with respect to* $\|\cdot\|$ *with monotonicity parameter* $c > 0$, *then the sequence generated by* (16) *converges to the unique* $x^* \in \operatorname{Zero}(\mathsf{F}+\mathsf{G})$

for every $\alpha \in \left]0, \min\left\{(\mathrm{diagL}(\mathsf{F}))^{-1}, (\mathrm{diagL}(\mathsf{G}))^{-1}\right\}\right]$. Moreover, for every $k \in \mathbb{Z}_{\geq 0}$,

$$\|x_{k+1} - x^*\| \leq \frac{1-\alpha c}{1+\alpha c}\|x_k - x^*\|,$$

with convergence rate optimized at $\alpha = \min\left\{(\mathrm{diagL}(\mathsf{F}))^{-1}, (\mathrm{diagL}(\mathsf{G}))^{-1}\right\}$.

(ii) If F is monotone with respect to $\|\cdot\|$ and $\mathrm{Zero}(\mathsf{F}+\mathsf{G}) \neq \emptyset$, then (18) converges to an element of $\mathrm{Zero}(\mathsf{F}+\mathsf{G})$ for every $\alpha \in \left]0, \min\left\{(\mathrm{diagL}(\mathsf{F}))^{-1}, (\mathrm{diagL}(\mathsf{G}))^{-1}\right\}\right]$.

VI. Application to Recurrent Neural Networks

A. Analysis and various iterations

Consider the continuous-time recurrent neural network

$$\dot{x} = -x + \Phi(Ax + Bu + b) =: F(x, u), \tag{19}$$

where $x \in \mathbb{R}^n, u \in \mathbb{R}^m, A \in \mathbb{R}^{n\times n}, B \in \mathbb{R}^{n\times m}, b \in \mathbb{R}^n$, and $\Phi : \mathbb{R}^n \to \mathbb{R}^n$ is an activation function applied entrywise, i.e., $\Phi(x) = (\phi(x_1), \dots, \phi(x_n))^\top$. We consider the case that ϕ is a LeakyReLU activation function, i.e., $\phi(x) = \max\{x, ax\}$ for some $a \in]0, 1[$. In **[?]**, it was shown that a sufficient condition for the contractivity of (19) is the existence of $\eta \in \mathbb{R}^n_{>0}$ such that $\mu_{\infty,[\eta]^{-1}}(A) < 1$. If this condition holds, then (19) is contracting with respect to $\|\cdot\|_{\infty,[\eta]^{-1}}$ with rate $1 - \phi(\mu_{\infty,[\eta]^{-1}}(A))$. In what follows, we define $\gamma := \mu_{\infty,[\eta]^{-1}}(A) < 1$.

Suppose that, for fixed u, we are interested in efficiently computing the unique equilibrium point $x^*(u)$ of $F(x, u)$. Since $F(x, u)$ is contracting with respect to $\|\cdot\|_{\infty,[\eta]^{-1}}$, $-F(x, u)$ is strongly monotone with monotonicity parameter $1 - \phi(\gamma)$. As a consequence, applying the forward step method, Algorithm 10 to compute $x^*(u)$ yields the iteration

$$x_{k+1} = (1-\alpha)x_k + \alpha\Phi(Ax_k + Bu + b), \tag{20}$$

which is the iteration proposed in [14]. This iteration is guaranteed to converge for every $\alpha \in \left]0, (1 - \min_{i\in\{1,\dots,n\}} \min\{a \cdot (A)_{ii}, (A)_{ii}\})^{-1}\right]$ with contraction factor $1 - \alpha(1 - \phi(\gamma))$.

However, rather than viewing finding an equilibrium of (19) as finding a zero of a non-Euclidean monotone operator, it is also possible to view it as an operator splitting problem. In particular, in the spirit of [22, Theorem 1], we prove that finding a fixed point of $\Phi(Ax + Bu + b)$ corresponds to an appropriate operator splitting problem under suitable assumptions on Φ. However, first we must define the proximal operator.

Definition 23 (Proximal operator [3, Definition 12.23])**.** *Let $f : \mathbb{R}^n \to]-\infty, \infty]$ be a convex closed proper function. Then the* proximal operator of f *evaluated at $x \in \mathbb{R}^n$ is*

$$\mathrm{prox}_f(x) = \arg\min_{z\in\mathbb{R}^n} \tfrac{1}{2}\|x - z\|_2^2 + f(z). \tag{21}$$

Proposition 24. *Suppose ϕ is the proximal operator of a continuously differentiable convex function f. Then finding an equilibrium point $x^*(u)$ of (19) is equivalent to the operator splitting problem $(\mathsf{F} + \mathsf{G})(x^*(u)) = 0$, where*

$$\mathsf{F}(z) = (I_n - A)z - (Bu + b), \qquad \mathsf{G}(z) = df(z), \tag{22}$$

where we denote $df(z) = (f'(z_1), \dots, f'(z_n))^\top$.

For the LeakyReLU activation function, it is known that the f corresponding to ϕ is given by $f(z_i) = \frac{1-a}{2a}\min\{z_i, 0\}^2$, [15, Table 1] which is continuously differentiable with derivative $df(z) = \frac{1-a}{a}\min\{z, 0\}$. Moreover, df is Lipschitz with constant $(1-a)/a$. Now we will show that under the sufficient condition $\gamma < 1$, F is strongly monotone and G is monotone with respect to $\|\cdot\|_{\infty,[\eta]^{-1}}$.

Since $\gamma < 1$, we see that F is strongly monotone with monotonicity parameter $1-\gamma$ since $-\mu_{\infty,[\eta]^{-1}}(-(I_n - A)) = 1 - \gamma > 0$. Moreover, checking that G is monotone is straightforward since df is Lipschitz and $Ddf(z)$ is diagonal for every $z \in \mathbb{R}^n$ for which it exists and has diagonal entries in $[0, (1-a)/a]$. As a consequence, for almost every $z \in \mathbb{R}^n$, $\mu_{\infty,[\eta]^{-1}}(-Ddf(z)) \leq 0$, which implies monotonicity of G with respect to $\|\cdot\|_{\infty,[\eta]^{-1}}$.

Therefore, we can consider different operator splitting algorithms to compute the equilibrium of (19). First, the forward-backward splitting method may be applied:

$$x_{k+1} = \mathrm{prox}_{\alpha f}((1-\alpha)x_k + \alpha(Ax_k + Bu + b)). \tag{23}$$

Since F is Lipschitz, this iteration is guaranteed to converge to the unique fixed point of (19). Moreover, the contraction factor for this iteration is $1 - \alpha(1-\gamma)$ for $\alpha \in]0, \frac{1}{1-\min_i(A)_{ii}}]$, with contraction factor being maximized at $\alpha^* = \frac{1}{1-\min_i(A)_{ii}}$. Note that compared to the iteration (20), the forward-backward iteration has a larger allowable range of step sizes and improved contraction factor at the expense of computing a proximal operator at each iteration.

Alternatively, the fixed point may be computed by means of the Peaceman-Rachford splitting algorithm, which is

$$\begin{aligned} x_{k+1} &= (I_n + \alpha(I_n - A))^{-1}(z_k + \alpha(Bu + b)), \\ z_{k+1} &= z_k + 2\mathrm{prox}_{\alpha f}(2x_{k+1} - z_k) - 2x_{k+1}. \end{aligned} \tag{24}$$

Since both F and G are Lipschitz, this iteration converges to the unique fixed point of (19). Moreover, the contraction factor is $\frac{1-\alpha(1-\gamma)}{1+\alpha(1-\gamma)}$ for $\alpha \in \left]0, \min\left\{\frac{1}{1-\min_i(A)_{ii}}, \frac{a}{1-a}\right\}\right]$, which comes from the Lipschitz constants of F and G. In other words, the contraction factor is improved for Peaceman-Rachford compared to forward-backward splitting, but the stepsize is limited by the Lipschitz constant of df.

B. Numerical implementations

To assess the iterations in (20), (23), and (24), we generated A, B, b, u in (19) and applied the iterations to compute the equilibrium. We generate $A \in \mathbb{R}^{200\times 200}, B \in \mathbb{R}^{200\times 50}, u \in \mathbb{R}^{50}, b \in \mathbb{R}^{200}$ with entries normally distributed as $A_{ij}, B_{ij}, b_i \sim \mathcal{N}(0, 1/\sqrt{200})$ and $u_i \sim \mathcal{N}(0, 1/\sqrt{50})$. To ensure that $A \in \mathbb{R}^{200\times 200}$ satisfies the constraint $\mu_{\infty,[\eta]^{-1}}(A) < 1$ for some $\eta \in \mathbb{R}^n_{>0}$, we pick $[\eta] = I_n$ and project A onto the convex polytope $\{A \in \mathbb{R}^{n\times n} \mid \mu_\infty(A) \leq 0.99\}$. We additionally computed $\mu_2(A) \approx 1.0034$, so F is not strongly monotone with respect to $\|\cdot\|_2$.

For all iterations, we initialize x_0 at the origin and for the Peaceman-Rachford iteration, we initialize z_0 at the origin. We set $a = 0.1$ in LeakyReLU and for each iteration pick

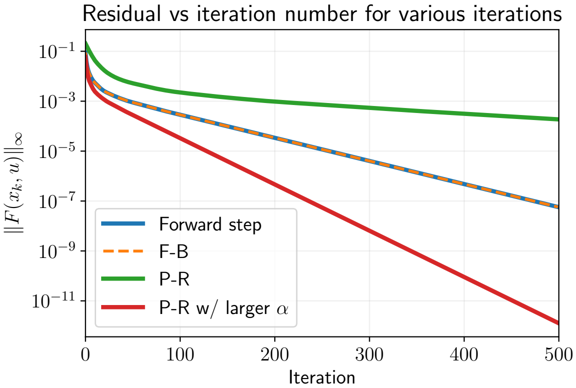

Fig. 1. Residual versus number of iterations for forward-step method (20), forward-backward (F-B) splitting (23), and Peaceman-Rachford (P-R) splitting (24) for computing the equilibrium of the recurrent neural network (19). Curves for the forward-step method and forward-backward splitting are directly on top of one another.

the largest theoretically allowable stepsize, which for the forward-step method and forward-backward splitting was $\frac{1}{1-\min_i(A)_{ii}} \approx 0.9015$. For Peaceman-Rachford splitting, the largest theoretically allowable stepsize was $a/(1-a) \approx 0.1111$, but we also simulated using $\alpha = 0.9015$. The plots of the residual $\|x_k - \Phi(Ax_k + Bu + b)\|_\infty = \|F(x_k, u)\|_\infty$ versus the number of iterations is shown in Figure 1.

Both forward-step and forward-backward splitting methods for computing the equilibrium of (19) converge at the same rate. This result agrees with the theory since $\phi(\gamma) = \gamma$ and the estimated contraction factor for both the forward step method and forward-backward splitting is $1 - \alpha(1-\gamma) \approx 0.9910$. For the Peaceman-Rachford splitting method, for the theoretically largest allowable $\alpha = 1/9$, the estimated contraction factor is $\frac{1-\alpha(1-\gamma)}{1+\alpha(1-\gamma)} \approx 0.9978$, which is very close to 1 and thus justifies the slow rate of convergence for the iterations in this case. However, if we let $\alpha = 0.9015$ as in the other methods, we observe a significant acceleration in the convergence of these iterations.

VII. Conclusion

We develop a non-Euclidean monotone operator framework with an emphasis on operators which are monotone with respect to ℓ_1 and ℓ_∞ norms. Classical algorithms for computing zeros of monotone operators and splitting methods are applicable in our framework and can exhibit improved convergence rates compared to their corresponding algorithms in Euclidean spaces. We apply our results to RNN equilibrium computation and demonstrate that applying splitting methods yields improved rates of convergence to the equilibria as compared to other methods.

Topics of future research include (i) tightening the Lipschitz estimates of the operator splitting techniques, (ii) extending the results to include infinite-dimensional Banach spaces and set-valued operators F, and (iii) applying this framework for robustness analysis of control systems and machine learning models.

References

[1] Z. Aminzare and E. D. Sontag. Contraction methods for nonlinear systems: A brief introduction and some open problems. In *IEEE Conf. on Decision and Control*, pages 3835–3847, December 2014. doi:10.1109/CDC.2014.7039986.

[2] Z. Aminzare and E. D. Sontag. Synchronization of diffusively-connected nonlinear systems: Results based on contractions with respect to general norms. *IEEE Transactions on Network Science and Engineering*, 1(2):91–106, 2014. doi:10.1109/TNSE.2015.2395075.

[3] H. H. Bauschke and P. L. Combettes. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*. Springer, 2 edition, 2017, ISBN 978-3-319-48310-8.

[4] A. Bernstein, E. Dall'Anese, and A. Simonetto. Online primal-dual methods with measurement feedback for time-varying convex optimization. *IEEE Transactions on Signal Processing*, 67(8):1978–1991, 2019. doi:10.1109/TSP.2019.2896112.

[5] T. Chaffey, F. Forni, and R. Sepulchre. Scaled relative graphs for system analysis. In *IEEE Conf. on Decision and Control*, pages 3166–3172, 2021. doi:10.1109/CDC45484.2021.9683092.

[6] C. Chidume. *Geometric Properties of Banach Spaces and Nonlinear Iterations*. Springer, 2009, ISBN 978-1-84882-189-7.

[7] P. L. Combettes. Monotone operator theory in convex optimization. *Mathematical Programming*, 170:177–206, 2018. doi:10.1007/s10107-018-1303-3.

[8] P. L. Combettes and J.-C. Pesquet. Deep neural network structures solving variational inequalities. *Set-Valued and Variational Analysis*, 28(3):491–518, 2020. doi:10.1007/s11228-019-00526-z.

[9] R. Cominetti, J. A. Soto, and J. Vaisman. On the rate of convergence of Krasnosel'skii-Mann iterations and their connection with sums of Bernoullis. *Israel Journal of Mathematics*, 199(2):757–772, 2014. doi:10.1007/s11856-013-0045-4.

[10] S. Coogan. A contractive approach to separable Lyapunov functions for monotone systems. *Automatica*, 106:349–357, 2019. doi:10.1016/j.automatica.2019.05.001.

[11] A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, July 2021. Conditionally accepted as Paper. URL: https://arxiv.org/abs/2103.12263.

[12] C. A. Desoer and H. Haneda. The measure of a matrix as a tool to analyze computer algorithms for circuit analysis. *IEEE Transactions on Circuit Theory*, 19(5):480–486, 1972. doi:10.1109/TCT.1972.1083507.

[13] S. Jafarpour, P. Cisneros-Velarde, and F. Bullo. Weak and semi-contraction for network systems and diffusively-coupled oscillators. *IEEE Transactions on Automatic Control*, 67(3):1285–1300, 2022. doi:10.1109/TAC.2021.3073096.

[14] S. Jafarpour, A. Davydov, A. V. Proskurnikov, and F. Bullo. Robust implicit networks via non-Euclidean contractions. In *Advances in Neural Information Processing Systems*, December 2021. URL: http://arxiv.org/abs/2106.03194.

[15] J. Li, C. Fang, and Z. Lin. Lifted proximal operator machines. In *AAAI Conference on Artificial Intelligence*, pages 4181–4188, 2019. doi:10.1609/aaai.v33i01.33014181.

[16] W. Lohmiller and J.-J. E. Slotine. On contraction analysis for non-linear systems. *Automatica*, 34(6):683–696, 1998. doi:10.1016/S0005-1098(98)00019-3.

[17] L. Pavel. Distributed GNE seeking under partial-decision information over networks via a doubly-augmented operator splitting approach. *IEEE Transactions on Automatic Control*, 65(4):1584–1597, 2020. doi:10.1109/TAC.2019.2922953.

[18] E. K. Ryu and S. Boyd. Primer on monotone operator methods. *Applied Computational Mathematics*, 15(1):3–43, 2016.

[19] E. K. Ryu and W. Yin. *Large-Scale Convex Optimization via Monotone Operators*. Cambridge, 2022.

[20] A. Simonetto. Time-varying convex optimization via time-varying averaged operators, 2017. ArXiv e-print:1704.07338. URL: https://arxiv.org/abs/1704.07338.

[21] G. Söderlind. The logarithmic norm. History and modern theory. *BIT Numerical Mathematics*, 46(3):631–652, 2006. doi:10.1007/s10543-006-0069-9.

[22] E. Winston and J. Z. Kolter. Monotone operator equilibrium networks. In *Advances in Neural Information Processing Systems*, 2020. URL: https://arxiv.org/abs/2006.08591.