Non-Euclidean Monotone Operator Theory and Applications

Journal of Machine Learning Research 25 (2024) 1-33 Submitted 6/23; Revised 4/24; Published 10/24

Non-Euclidean Monotone Operator Theory and Applications

Alexander Davydov[*] DAVYDOV@UCSB.EDU
Center for Control, Dynamical Systems, and Computation
University of California, Santa Barbara
Santa Barbara, CA 93106-5070, USA

Saber Jafarpour[*] SABER.JAFARPOUR@COLORADO.EDU
Department of Electrical, Computer, and Energy Engineering
University of Colorado, Boulder
Boulder, CO 80309-0020, USA

Anton V. Proskurnikov ANTON.P.1982@IEEE.ORG
Department of Electronics and Telecommunications
Politecnico di Torino
Turin, Italy

Francesco Bullo BULLO@UCSB.EDU
Center for Control, Dynamical Systems, and Computation
University of California, Santa Barbara
Santa Barbara, CA 93106-5070, USA

Editor: Silvia Villa

Abstract

While monotone operator theory is often studied on Hilbert spaces, many interesting problems in machine learning and optimization arise naturally in finite-dimensional vector spaces endowed with non-Euclidean norms, such as diagonally-weighted ℓ_1 or ℓ_∞ norms. This paper provides a natural generalization of monotone operator theory to finite-dimensional non-Euclidean spaces. The key tools are weak pairings and logarithmic norms. We show that the resolvent and reflected resolvent operators of non-Euclidean monotone mappings exhibit similar properties to their counterparts in Hilbert spaces. Furthermore, classical iterative methods and splitting methods for finding zeros of monotone operators are shown to converge in the non-Euclidean case. We apply our theory to equilibrium computation and Lipschitz constant estimation of recurrent neural networks, obtaining novel iterations and tighter upper bounds via forward-backward splitting.

Keywords: non-Euclidean norms, monotone operator theory, fixed point equations, nonexpansive maps

1. Introduction

Problem description and motivation: Monotone operator theory is a fertile field of nonlinear functional analysis that extends the notion of monotone functions on $\mathbb{R}$ to mappings on Hilbert spaces. Monotone operator methods are widely used to solve problems in machine learning (Combettes and Pesquet, 2020b; Winston and Kolter, 2020), data science (Combettes and Pesquet, 2021), optimization and control (Simonetto, 2017; Bernstein et al.,

[*]. The first two authors contributed equally

2019), game theory (Pavel, 2020), and systems analysis (Chaffey et al., 2021). A crucial part of this theory is the design of algorithms for computing zeros of monotone operators. This problem is central in convex optimization since (i) the subdifferential of any convex function is monotone and (ii) minimizing a convex function is equivalent to finding a zero of its subdifferential. To this end, there has been extensive research in the last decade in applying monotone operator methods to convex optimization; see, e.g., (Ryu and Boyd, 2016; Combettes, 2018; Ryu and Yin, 2022).

Existing monotone operator techniques are primarily based on inner-product spaces, while many problems are better-suited for analysis in more general normed spaces. For instance robustness analysis of artificial neural networks in machine learning often requires the use of the ℓ_∞ norm for high-dimensional input data such as images (Goodfellow et al., 2015). In distributed optimization, it is known that many natural conditions for the convergence of totally asynchronous algorithms are based upon contractions in an ℓ_∞ norm (Bertsekas and Tsitsiklis, 1997, Chapter 6, Section 3).

Motivated by problems in non-Euclidean spaces, we aim to extend monotone operator techniques for computing zeros of monotone operators to operators which are naturally "monotone" with respect to (w.r.t.) a non-Euclidean norm in a finite-dimensional space.

Literature review: The literature on monotone operators dates back to Minty and Browder (Minty, 1962; Browder, 1967) and the connection to convex analysis was drawn upon by Minty and Rockafellar (Minty, 1964; Rockafellar, 1966). Since these foundational works, the theory of monotone operators over Hilbert spaces and its connection with convex optimization continues to expand, especially in the last decade (Bauschke and Combettes, 2017; Ryu and Boyd, 2016; Ryu and Yin, 2022; Ryu et al., 2021). Despite these connections between convex optimization and monotone operators, many problems in machine learning involve monotone operators beyond gradients of convex functions. Examples of such problems include generative adversarial networks, adversarially robust training of models, and training of models under fairness constraints. Instead of minimizing a convex function, to address these problems, one must solve for variational inequalities, monotone inclusions, and game-theoretic equilibria. In each of these more general cases, monotone operator theory has played an essential role in their analyses.

In machine learning, monotone operators have been used in the training of generative adversarial networks (Gidel et al., 2019), in the design of novel neural network architectures (Winston and Kolter, 2020), in the analysis of equilibrium behavior (infinite-depth limit) of neural networks (Combettes and Pesquet, 2020b), in the estimation of Lipschitz constants of neural networks (Combettes and Pesquet, 2020a; Pabbaraju et al., 2021), and in normalizing flows (Ahn et al., 2022). Monotone operators have also been studied in the machine learning community in the context of variational inequality algorithms, stochastic monotone inclusions, and saddle-point problems; see e.g. (Balamurugan and Bach, 2016; Diakonikolas, 2020; Cai et al., 2022; Pethick et al., 2022; Zhang et al., 2022; Alacaoglu et al., 2023; Jordan et al., 2023; Yang et al., 2023) for recent works in this direction. See also the recent survey (Combettes and Pesquet, 2021) for applications in data science.

The theory of dissipative and accretive operators on Banach spaces largely parallels the theory of monotone operators on Hilbert spaces (Deimling, 1985). Despite these parallels, this theory has found far fewer direct applications to machine learning and data science; instead it is mainly applied for iterative solving integral equations and PDEs in L_p spaces for $p \neq 2$ (see the book, Chidume 2009, for iterative methods). Moreover, many works in Banach spaces focus on spaces that have a uniformly smooth or uniformly convex structure, which finite-dimensional ℓ_1 and ℓ_∞ spaces do not possess. In a similar vein, methods based on Bregman divergences utilize smoothness and strict convexity of the distance-generating convex functions (Bauschke et al., 2003). Connections between logarithmic norms and dissipative and accretive operators may be found in (Söderlind, 1986, 2006).

A concept similar to a monotone operator in a Hilbert space is that of a contracting vector field in dynamical systems theory (Lohmiller and Slotine, 1998). If the metric with respect to which the vector field is contracting is the standard Euclidean distance, the vector field, F, is strongly infinitesimally contracting if and only if the negative vector field $-\mathsf{F}$ is strongly monotone when thought of as on operator on $\mathbb{R}^n$. However, vector fields need not be contracting with respect to a Euclidean distance. Indeed, a vector field may be contracting w.r.t. a non-Euclidean norm but not a Euclidean one (Aminzare and Sontag, 2014). 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 w.r.t. a non-Euclidean norm. In this spirit, preliminary connections between contracting vector fields and monotone operators were made in (Bullo et al., 2021).

Contributions: Our contributions are as follows. First, to address the gap in applying monotone operator strategies to problems that arise in finite-dimensional non-Euclidean spaces, we propose a non-Euclidean monotone operator framework that is based on the theory of weak pairings (Davydov et al., 2022a) and logarithmic norms. We use weak pairings as a substitute for inner products and we demonstrate that many classic results from monotone operator theory are applicable to its non-Euclidean counterpart. In particular, we show that the resolvent and reflected resolvent operators of a non-Euclidean monotone mapping exhibit properties similar to those arising in Hilbert spaces. To ensure that the resolvent and reflected resolvents have full domain, we prove an extension of the classic Minty-Browder theorem (Minty, 1962; Browder, 1967) in Theorem 17.

Second, leveraging the non-Euclidean monotone operator framework, we show that traditional iterative algorithms such as the forward step method and proximal point method can be used to compute zeros of non-Euclidean monotone mappings. We provide convergence rate estimates for these iterative algorithms and the Cayley method in Theorems 26, 28, and 31 and demonstrate that for diagonally-weighted ℓ_1 and ℓ_∞ norms, they exhibit improved convergence rates compared to their Euclidean counterparts. Notably, we prove that for a Lipschitz mapping which is monotone w.r.t. a diagonally-weighted ℓ_1 or ℓ_∞ norm, the forward step method is guaranteed to converge for a sufficiently small step size, whereas convergence cannot be guaranteed if the mapping is monotone with respect to a Euclidean norm.

Third, we study operator splitting methods for mappings which are monotone w.r.t. diagonally-weighted ℓ_1 or ℓ_∞ norms. In Theorems 33 and 36 we prove that the forward-backward, Peaceman-Rachford, and Douglas-Rachford splitting algorithms are all guaranteed to converge, with some key differences compared to the classical theory. For instance, in the classical setting where two operators, F and G, are monotone w.r.t. a Euclidean norm, the forward-backward splitting algorithm will only converge if F is cocoercive. In contrast, when considering ℓ_1 or ℓ_∞ norms, Lipschitzness of F is sufficient for convergence.

Fourth, we present new insights into non-Euclidean properties of proximal operators and their impact on the study of special set-valued operator inclusions. Specifically, in Proposition 41, we demonstrate that when F is the subdifferential of a separable, proper, lower semicontinuous, convex function, its resolvent and reflected resolvent are nonexpansive with respect to an ℓ_∞ norm. To showcase the practical relevance of this result, we apply our non-Euclidean monotone operator theory to the equilibrium computation of a recurrent neural network (RNN). We extend the recent work of (Jafarpour et al., 2021) and show that our theory provides novel iterations and convergence criteria for RNN equilibrium computation.

Finally, we study the robustness of the RNN via its ℓ_∞ norm Lipschitz constant. In Theorem 43, we generalize the results from (Pabbaraju et al., 2021) to non-Euclidean norms and provide sharper estimates for the ℓ_∞ Lipschitz constant than were provided in the previous work (Jafarpour et al., 2021).

A preliminary version of this work appeared in (Davydov et al., 2022b). Compared to this preliminary version, this version (i) provides novel theoretical results on the analysis of nonsmooth operators which are monotone with respect to general norms, (ii) proves a novel generalization of the classical Minty-Browder theorem for these non-Euclidean monotone mappings, (iii) study special classes of set-valued inclusions by providing novel non-Euclidean properties of proximal operators, (iv) includes a more comprehensive application to RNNs, allowing for more general activation functions and studies the robustness of the neural network by providing a tighter Lipschitz estimate, and (v) includes proofs of all technical results. Finally, we provide further comparisons to monotone operator theory on Hilbert spaces. Other prior work, (Davydov et al., 2022a, 2024), focuses on continuous-time contracting dynamical systems with respect to non-Euclidean norms and their robustness properties. In contrast, this work instead uses weak pairings, developed in (Davydov et al., 2022a), to establish monotonicity properties of maps with respect to non-Euclidean norms and how we can find zeros of these maps using iterative methods. The prior works (Davydov et al., 2022a, 2024) do not consider these discrete-time iterations.

2. Preliminaries

2.1 Notation

We let $\mathbb{R}_{\geq 0}$ be the set of nonnegative real numbers and $\mathbb{R}_{>0}$ be the set of positive real numbers. For a set $\mathcal{S}$, let $2^{\mathcal{S}}$ denote its power set. For a complex number z, let $\mathrm{Re}(z)$ denote its real part. For a vector $\eta \in \mathbb{R}^n$, let $[\eta]$ denote the diagonal matrix satisfying $[\eta]_{ii} = \eta_i$, where $\eta_1, \dots, \eta_n$ are the components of η. Given a matrix $A \in \mathbb{R}^{n \times n}$, let $\mathrm{spec}(A)$ denote its spectrum. For a mapping $\mathsf{F} : \mathcal{X} \to \mathcal{Y}$ where $\mathcal{X} \subseteq \mathbb{R}^n, \mathcal{Y} \subseteq \mathbb{R}^m$, let $\mathrm{Dom}(\mathsf{F})$ be its domain. If F is differentiable, let $D\mathsf{F}(x) := \frac{\partial \mathsf{F}(x)}{\partial x}$ denote its Jacobian evaluated at x. For a mapping $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$, let $\mathrm{Zero}(\mathsf{F}) := \{x \in \mathbb{R}^n \mid \mathsf{F}(x) = \mathbb{0}_n\}$ 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 mapping and $I_n \in \mathbb{R}^{n \times n}$ be the $n \times n$ identity matrix. For a vector $y \in \mathbb{R}^n$, define the mapping $\mathrm{sign} : \mathbb{R}^n \to \{-1, 0, 1\}^n$ by $(\mathrm{sign}(y))_i = y_i/|y_i|$ if $y_i \neq 0$ and zero otherwise. For a convex function $f : \mathbb{R}^n \to \,]-\infty, +\infty]$, let $\partial f(x) = \{g \in \mathbb{R}^n \mid f(x) - f(y) \geq g^\top (x - y), \forall y \in \mathbb{R}^n\}$ be its subdifferential at x and, if f is differentiable, $\nabla f(x)$ be its gradient at x.

2.2 Logarithmic Norms and Weak Pairings

Instrumental to the theory of non-Euclidean monotone operators are logarithmic norms (also referred to as matrix measures, or going forward, log norms) discovered by Dahlquist and Lozinskii in 1958 (Dahlquist, 1958; Lozinskii, 1958).

Definition 1 (Logarithmic norm) *Let $\|\cdot\|$ denote a norm on $\mathbb{R}^n$ and also the induced operator norm on the set of matrices $\mathbb{R}^{n\times n}$. The* logarithmic norm *of a matrix $A \in \mathbb{R}^{n\times n}$ is*

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

The log norm of a matrix A is the one-sided directional derivative of the induced norm in the direction of A evaluated at the identity matrix. It is well known that this limit exists for any norm and matrix (Lozinskii, 1958, Lemma 1). Properties of log norms include positive homogeneity, subadditivity, convexity, and $\mathrm{Re}(\lambda) \le \mu(A) \le \|A\|$ for all $\lambda \in \mathrm{spec}(A)$, (Lozinskii, 1958). Note that unlike the induced matrix norm, the log norm of a matrix may be negative. For more details on log norms, see the influential survey (Söderlind, 2006) and the recent monograph (Bullo, 2022, Chapter 2).

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

$$\|x\|_{1,[\eta]} = \|[\eta]x\|_1 = \sum_{i=1}^n \eta_i |x_i| \qquad \text{and} \qquad \|x\|_{\infty,[\eta]^{-1}} = \|[\eta]^{-1}x\|_\infty = \max_{i\in\{1,\dots,n\}} \frac{1}{\eta_i}|x_i|. \tag{2}$$

The formulas for the corresponding induced norms and log norms are provided in (Bullo, 2023, Eq. (2.36)-(2.38)) and are

$$\|A\|_{1,[\eta]} = \max_{j\in\{1,\dots,n\}} \sum_{i=1}^n \frac{\eta_i}{\eta_j}|a_{ij}|, \qquad \mu_{1,[\eta]}(A) = \max_{j\in\{1,\dots,n\}} \Big(a_{jj} + \sum_{i=1, i\neq j}^n |a_{ij}|\frac{\eta_i}{\eta_j}\Big),$$

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

We additionally review the notion of a weak pairing (WP) on $\mathbb{R}^n$ from (Davydov et al., 2022a) which generalizes inner products to non-Euclidean spaces.

Definition 2 (Weak pairing) *A* weak pairing *is a map $[\![\cdot,\cdot]\!] : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ satisfying:*

1. *(sub-additivity and continuity of first argument) $[\![x_1 + x_2, y]\!] \le [\![x_1, y]\!] + [\![x_2, y]\!]$, for all $x_1, x_2, y \in \mathbb{R}^n$ and $[\![\cdot,\cdot]\!]$ is continuous in its first argument,*
2. *(weak homogeneity) $[\![\alpha x, y]\!] = [\![x, \alpha y]\!] = \alpha [\![x, y]\!]$ and $[\![-x, -y]\!] = [\![x, y]\!]$, for all $x, y \in \mathbb{R}^n, \alpha \ge 0$,*
3. *(positive definiteness) $[\![x, x]\!] > 0$, for all $x \neq \mathbb{0}_n$,*
4. *(Cauchy-Schwarz inequality) $|\,[\![x, y]\!]\,| \le [\![x, x]\!]^{1/2}\,[\![y, y]\!]^{1/2}$, for all $x, y \in \mathbb{R}^n$.*

For every norm $\|\cdot\|$ on $\mathbb{R}^n$, there exists a (possibly not unique) compatible WP $[\![\cdot,\cdot]\!]$ such that $\|x\|^2 = [\![x, x]\!]$, for every $x \in \mathbb{R}^n$. If the norm is induced by an inner product, the WP coincides with the inner product.

Definition 3 (Deimling's inequality and curve norm derivative formula) *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ with compatible WP $[\![\cdot,\cdot]\!]$.*

1. *The WP $[\![\cdot,\cdot]\!]$ satisfies* Deimling's inequality *if*

$$[\![x,y]\!] \leq \|y\| \lim_{h\to 0^+} h^{-1}(\|y+hx\| - \|y\|), \qquad \text{for all } x,y\in\mathbb{R}^n. \tag{3}$$

2. *The WP $[\![\cdot,\cdot]\!]$ satisfies the* curve norm derivative formula *if for all differentiable $x :]a,b[\to \mathbb{R}^n$, $\|x(t)\|D^+\|x(t)\| = [\![\dot{x}(t), x(t)]\!]$ holds for almost every $t\in]a,b[$, where D^+ denotes the upper right Dini derivative.*[1]

For every norm, there exists at least one WP that satisfies the properties in Definition 3.[2] Thus, going forward, we assume that WPs satisfy these additional properties. Indeed, due to Deimling's inequality, we have the following useful relationship between WPs and log norms.

Lemma 4 (Lumer's equality, Davydov et al. 2022a, Theorem 18) *Let $\|\cdot\|$ be a norm on $\mathbb{R}^n$ with compatible WP $[\![\cdot,\cdot]\!]$. Then for every $A\in\mathbb{R}^{n\times n}$,*

$$\mu(A) = \sup_{\|x\|=1} [\![Ax,x]\!] = \sup_{x\neq \mathbb{0}_n} \frac{[\![Ax,x]\!]}{\|x\|^2}. \tag{4}$$

We will focus on WPs corresponding to diagonally-weighted ℓ_1 and ℓ_∞ norms. Specifically, from (Davydov et al., 2022a, Table III), we introduce the WPs $[\![\cdot,\cdot]\!]_{1,[\eta]}, [\![\cdot,\cdot]\!]_{\infty,[\eta]^{-1}} : \mathbb{R}^n\times\mathbb{R}^n \to \mathbb{R}$ defined by

$$[\![x,y]\!]_{1,[\eta]} = \|y\|_{1,[\eta]} \operatorname{sign}(y)^\top [\eta] x \qquad \text{and} \qquad [\![x,y]\!]_{\infty,[\eta]^{-1}} = \max_{i\in I_\infty([\eta]^{-1}y)} \eta_i^{-2} y_i x_i. \tag{5}$$

where $I_\infty(x) = \{i\in\{1,\dots,n\} \mid |x_i| = \|x\|_\infty\}$. One can show that both of these WPs satisfy Deimling's inequality and the curve-norm derivative formula. Formulas for more general ℓ_p norms are available in (Davydov et al., 2022a).

2.3 Contractions, Nonexpansive Maps, and Iterations

Definition 5 (Lipschitz continuity) *Let $\|\cdot\|$ be a norm and $\mathsf{F}:\mathbb{R}^n\to\mathbb{R}^n$ be a mapping. F is* Lipschitz continuous *with constant $\ell\in\mathbb{R}_{\geq 0}$ if*

$$\|\mathsf{F}(x_1) - \mathsf{F}(x_2)\| \leq \ell\|x_1 - x_2\| \qquad \text{for all } x_1,x_2\in\mathbb{R}^n. \tag{6}$$

Moreover we define $\mathrm{Lip}(\mathsf{F})$ *to be the minimal (or infimum) constant which satisfies* (6).

If two mappings $\mathsf{F},\mathsf{G}:\mathbb{R}^n\to\mathbb{R}^n$ are Lipschitz continuous w.r.t. the same norm, then the composition $\mathsf{F}\circ\mathsf{G}$ has Lipschitz constant $\mathrm{Lip}(\mathsf{F}\circ\mathsf{G}) \leq \mathrm{Lip}(\mathsf{F})\,\mathrm{Lip}(\mathsf{G})$.

---

1. The definition and properties of Dini derivatives are presented in (Giorgi and Komlósi, 1992).
2. Indeed, given a norm, the map $[\![\cdot,\cdot]\!] : \mathbb{R}^n\times\mathbb{R}^n\to\mathbb{R}$ given by $[\![x,y]\!] = \|y\| \lim_{h\to 0^+} h^{-1}(\|y+hx\| - \|y\|)$ defines a WP that satisfies all of these properties. For more discussions about properties of this pairing, we refer to (Deimling, 1985, Section 13) and (Davydov et al., 2022a, Appendix A).

Definition 6 (One-sided Lipschitz mappings, Davydov et al. 2022a) *Given a norm* $\|\cdot\|$ *with compatible WP* $[\![\cdot,\cdot]\!]$*, a map* $\mathsf{F}:\mathbb{R}^n\to\mathbb{R}^n$ *is* one-sided Lipschitz *with constant* $c\in\mathbb{R}$ *if*

$$[\![\mathsf{F}(x_1)-\mathsf{F}(x_2),x_1-x_2]\!]\leq c\|x_1-x_2\|^2 \qquad \text{for all } x_1,x_2\in\mathbb{R}^n. \tag{7}$$

Moreover we define $\mathrm{osL}(\mathsf{F})$ *to be the minimal (or infimum) constant which satisfies* (7)*.*

As was proved in (Davydov et al., 2022a, Theorem 27), if $\mathsf{F},\mathsf{G}:\mathbb{R}^n\to\mathbb{R}^n$ are one-sided Lipschitz w.r.t. the same WP, then $\mathrm{osL}(\alpha\mathsf{F})=\alpha\,\mathrm{osL}(\mathsf{F})$, $\mathrm{osL}(\mathsf{F}+\mathsf{G})\leq\mathrm{osL}(\mathsf{F})+\mathrm{osL}(\mathsf{G})$, and $\mathrm{osL}(\mathsf{F}+c\mathsf{Id})=\mathrm{osL}(\mathsf{F})+c$ for all $\alpha\geq0$, $c\in\mathbb{R}$. Note that (i) the one-sided Lipschitz constant is upper bounded by the Lipschitz constant, (ii) a Lipschitz continuous map is always one-sided Lipschitz, and (iii) the one-sided Lipschitz constant may be negative. Moreover, if F is locally Lipschitz *continuous*, we have an alternative characterization of $\mathrm{osL}(\mathsf{F})$.

Lemma 7 ($\mathrm{osL}(\mathsf{F})$ for locally Lipschitz continuous F, Davydov et al. 2024, Theorem 16) *Suppose the map* $\mathsf{F}:\mathbb{R}^n\to\mathbb{R}^n$ *is locally Lipschitz continuous. Then* F *is one-sided Lipschitz with constant* $c\in\mathbb{R}$ *if and only if*[3]

$$\mu(D\mathsf{F}(x))\leq c \qquad \text{for almost every } x\in\mathbb{R}^n. \tag{8}$$

Definition 8 (Contractions and nonexpansive maps) *Let* $\mathsf{T}:\mathbb{R}^n\to\mathbb{R}^n$ *be Lipschitz continuous w.r.t.* $\|\cdot\|$*. We say* T *is a* contraction *if* $\mathrm{Lip}(\mathsf{T})<1$*, and* T *is* nonexpansive *if* $\mathrm{Lip}(\mathsf{T})\leq1$*.*

Definition 9 (Picard iteration) *Let* $\mathsf{T}:\mathbb{R}^n\to\mathbb{R}^n$ *be a contraction w.r.t. a norm* $\|\cdot\|$ *with* $\mathrm{Lip}(\mathsf{T})<1$*. The* Picard iteration *applied to* T *with initial condition* x^0 *defines the sequence* $\{x^k\}_{k=0}^{\infty}$ *by*

$$x^{k+1}=\mathsf{T}(x^k). \tag{9}$$

By the Banach fixed-point theorem T has a unique fixed point, x^*, and the Picard iteration applied to T satisfy $\|x^k-x^*\|\leq\mathrm{Lip}(\mathsf{T})^k\|x^0-x^*\|$, for any initial condition x^0.

If T is nonexpansive with $\mathrm{Fix}(\mathsf{T})\neq\emptyset$, Picard iteration may fail to find a fixed point of T. Such situations can be addressed by the following iteration and convergence result, initially proved in (Ishikawa, 1976) and with rate given in (Cominetti et al., 2014).

Definition 10 (Krasnosel'skii–Mann iteration) *Let* $\mathsf{T}:\mathbb{R}^n\to\mathbb{R}^n$ *be nonexpansive w.r.t. a norm* $\|\cdot\|$*. The* Krasnosel'skii–Mann iteration[4] *applied to* T *with initial condition* x^0 *and* $\theta\in\,]0,1[$ *defines the sequence* $\{x^k\}_{k=0}^{\infty}$ *by*

$$x^{k+1}=(1-\theta)x^k+\theta\mathsf{T}(x^k). \tag{10}$$

Lemma 11 (Asymptotic regularity and convergence of Krasnosel'skii–Mann iteration, Cominetti et al. 2014; Ishikawa 1976) *Let* $\mathsf{T}:\mathbb{R}^n\to\mathbb{R}^n$ *be nonexpansive*

3. Note that for locally Lipschitz continuous F, $D\mathsf{F}(x)$ exists for almost every x by Rademacher's theorem.
4. The Krasnosel'skii–Mann iteration may be defined with step sizes $\theta_k\in\,]0,1[$ which vary for each iteration. In this document, we will only work with constant step sizes.

w.r.t. a norm $\|\cdot\|$ *and consider the Krasnosel'skii–Mann iteration as in* (10)*. Suppose* $\mathrm{Fix}(\mathsf{T}) \neq \emptyset$*. Then for any initial condition* x^0*,*

$$\|x^k - \mathsf{T}(x^k)\| \leq \frac{2\inf_{x^* \in \mathrm{Fix}(\mathsf{T})} \|x^0 - x^*\|}{\sqrt{k\pi\theta(1-\theta)}} = \mathcal{O}(1/\sqrt{k}). \tag{11}$$

Moreover, the sequence of iterates, $\{x_k\}_{k=0}^{\infty}$*, converges to a fixed point of* T*.*

3. Non-Euclidean Monotone Operators

3.1 Definitions and Properties

Definition 12 (Non-Euclidean monotone mapping) *A mapping* $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ *is strongly monotone with monotonicity parameter* $c > 0$ *w.r.t. a norm* $\|\cdot\|$ *on* $\mathbb{R}^n$ *if there exists a compatible WP* $[\![\cdot,\cdot]\!]$ *and if for all* $x, y \in \mathbb{R}^n$*,*

$$-[\![-(\mathsf{F}(x) - \mathsf{F}(y)), x - y]\!] \geq c\|x - y\|^2. \tag{12}$$

If the inequality holds with $c = 0$*, we say* F *is monotone w.r.t.* $\|\cdot\|$*.*

In the language of Banach spaces, such a function F is called strongly accretive (Chidume, 2009, Definition 8.10). Note that Definition 12 is equivalent to $-\mathrm{osL}(-\mathsf{F}) \geq c$.

In the case of a Euclidean norm, the WP corresponds to the inner product and Definition 12 corresponds to the usual definition of a monotone operator as in (Minty, 1962) and (Bauschke and Combettes, 2017, Definition 20.1).

By properties of osL, if $\mathsf{F}, \mathsf{G} : \mathbb{R}^n \to \mathbb{R}^n$ are both monotone w.r.t. the same norm (and WP), then $-\mathrm{osL}(-\mathsf{F} - \mathsf{G}) \geq -\mathrm{osL}(-\mathsf{F}) - \mathrm{osL}(-\mathsf{G})$ and thus a sum of mappings which are monotone w.r.t. the same norm are monotone. Additionally, if F is monotone with monotonicity parameter $c \geq 0$, then for any $\alpha \geq 0$, $-\mathrm{osL}(-\mathsf{Id} - \alpha\mathsf{F}) = 1 - \alpha\,\mathrm{osL}(-\mathsf{F})$ and thus $\mathsf{Id} + \alpha\mathsf{F}$ is strongly monotone with monotonicity parameter $1 + \alpha c$.

Remark 13 (Connection with contracting vector fields) *A mapping* $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ *is strongly infinitesimally contracting with rate* $c > 0$ *w.r.t. a norm* $\|\cdot\|$ *on* $\mathbb{R}^n$ *provided* $\mathrm{osL}(\mathsf{F}) \leq -c$ *(Davydov et al., 2022a). If* $c = 0$*, we say* F *is weakly infinitesimally contracting w.r.t.* $\|\cdot\|$*. Clearly* F *is strongly monotone if and only if* $-\mathsf{F}$ *is strongly infinitesimally contracting. Vector fields which are strongly infinitesimally contracting w.r.t. a norm generate flows which are contracting with respect to the same norm. In the case of weakly infinitesimally contracting vector fields, their flows are nonexpansive.*

Lemma 14 (Monotonicity for locally Lipschitz continuous mappings) *Let* $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ *be locally Lipschitz continuous.* F *is (strongly) monotone with monotonicity parameter* $c \geq 0$ *w.r.t. a norm* $\|\cdot\|$ *if and only if* $-\mu(-D\mathsf{F}(x)) \geq c$ *for almost every* $x \in \mathbb{R}^n$*.*

Proof Lemma 14 is a straightforward application of Lemma 7. ■

We can see the application of Lemma 14 more explicitly in the context of continuously differentiable monotone operators in Euclidean norms. To be specific, for an operator

$\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$, let $\|\cdot\|_2$ be the Euclidean norm with corresponding inner product $\langle\!\langle \cdot, \cdot \rangle\!\rangle$. Then, following (Minty, 1962), F is monotone with respect to $\|\cdot\|_2$ if

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

If F is continuously differentiable, this condition is known to be equivalent to (see e.g., (Ryu and Boyd, 2016)) $D\mathsf{F}(x) + D\mathsf{F}(x)^\top \succeq 0$, or equivalently $-\mu_2(-D\mathsf{F}(x)) \geq 0$ or $\frac{1}{2}\lambda_{\min}(D\mathsf{F}(x) + D\mathsf{F}(x)^\top) \geq 0$, where $\mu_2(A) = \frac{1}{2}\lambda_{\max}(A + A^\top)$ is the log norm corresponding to the norm $\|\cdot\|_2$. This result coincides with what was demonstrated in Lemma 14.

Example 1 *An affine function $\mathsf{F}(x) = Ax + b$ is monotone if and only if $-\mu(-A) \geq 0$ and strongly monotone with parameter c if and only if $-\mu(-A) \geq c$. This condition implies that the spectrum of A lies in the portion of the complex plane given by $\{z \in \mathbb{C} \mid \mathrm{Re}(z) \geq c\}$.*

3.2 Resolvent, Reflected Resolvents, Forward Step Operators, and Lipschitz Estimates

Monotone operator theory transforms the problem of finding a zero of a monotone operator into finding a fixed point of a suitably defined operator. Monotone operator theory on Hilbert spaces studies the resolvent and reflected resolvent, operators dependent on the original operator, with fixed points corresponding to zeros of the original monotone operator. In this subsection we study these same two operators and also the forward step operator in the context of operators which are monotone w.r.t. a non-Euclidean norm. In particular, we characterize the Lipschitz constants of these operators, first providing Lipschitz upper bounds for arbitrary norms and then specializing to diagonally-weighted ℓ_1 and ℓ_∞ norms.

Definition 15 (Resolvent and reflected resolvent) *Let $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be a monotone mapping w.r.t. some norm. The resolvent of F with parameter $\alpha > 0$ denoted by $\mathsf{J}_{\alpha\mathsf{F}} : \mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{F}}) \to \mathbb{R}^n$ and defined by*

$$\mathsf{J}_{\alpha\mathsf{F}} = (\mathsf{Id} + \alpha\mathsf{F})^{-1}. \tag{13}$$

The reflected resolvent of F with parameter $\alpha > 0$ is denoted by $\mathsf{R}_{\alpha\mathsf{F}} : \mathrm{Dom}(\mathsf{R}_{\alpha\mathsf{F}}) \to \mathbb{R}^n$ and defined by

$$\mathsf{R}_{\alpha\mathsf{F}} = 2\mathsf{J}_{\alpha\mathsf{F}} - \mathsf{Id}. \tag{14}$$

Definition 16 (Forward step operator) *Let $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be a mapping and $\alpha \in \mathbb{R}$. The forward step of F with parameter $\alpha > 0$ is denoted by $\mathsf{S}_{\alpha\mathsf{F}} : \mathbb{R}^n \to \mathbb{R}^n$ and defined by*

$$\mathsf{S}_{\alpha\mathsf{F}} = \mathsf{Id} - \alpha\mathsf{F}. \tag{15}$$

Note that for any $\alpha > 0$, we have $\mathsf{F}(x) = \mathbb{0}_n$ if and only if $x = \mathsf{J}_{\alpha\mathsf{F}}(x) = \mathsf{R}_{\alpha\mathsf{F}}(x) = \mathsf{S}_{\alpha\mathsf{F}}(x)$, i.e., $\mathrm{Zero}(\mathsf{F}) = \mathrm{Fix}(\mathsf{J}_{\alpha\mathsf{F}}) = \mathrm{Fix}(\mathsf{R}_{\alpha\mathsf{F}}) = \mathrm{Fix}(\mathsf{S}_{\alpha\mathsf{F}})$. Note that under the assumption that F is monotone, both $\mathsf{J}_{\alpha\mathsf{F}}$ and $\mathsf{R}_{\alpha\mathsf{F}}$ are single-valued mappings.

We have deliberately not been specific with the domains of the resolvent and reflected resolvent operators. As we will show in the following theorem, under mild assumptions (continuity and monotonicity), both of their domains are all of $\mathbb{R}^n$.

Theorem 17 (A non-Euclidean Minty-Browder theorem) *Suppose* $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ *is continuous and monotone. Then for every* $\alpha > 0$, $\mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{F}}) = \mathrm{Dom}(\mathsf{R}_{\alpha\mathsf{F}}) = \mathbb{R}^n$.

Proof Note that $\mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{F}}) = \mathbb{R}^n$ provided that for every $u \in \mathbb{R}^n$, there exists $x \in \mathbb{R}^n$ such that $(\mathsf{Id} + \alpha\mathsf{F})(x) = u$. To establish this fact, consider the differential equation

$$\dot{x} = -x - \alpha\mathsf{F}(x) + u =: G(x). \tag{16}$$

Note that any equilibrium, x^*, of (16) satisfies $(\mathsf{Id} + \alpha\mathsf{F})(x^*) = u$. Thus it suffices to show that the differential equation (16) has an equilibrium. First we note that for all $x, y \in \mathbb{R}^n$,

$$[\![G(x) - G(y), x - y]\!] \leq [\![-(x - y), x - y]\!] + \alpha\, [\![-(\mathsf{F}(x) - \mathsf{F}(y)), x - y]\!] \leq -\|x - y\|^2. \tag{17}$$

Thus, we conclude that $\mathrm{osL}(G) \leq -1$. In line with Remark 13, we conclude that G is strongly infinitesimally contracting which ensures uniqueness of solutions to (16) (see (Davydov et al., 2022a, Theorem 31)). Let $\phi(t, x_0)$ denote the flow of the dynamics (16) at time $t \geq 0$ from initial condition $x(0) = x_0$. Then by (Davydov et al., 2022a, Theorem 31), we conclude that

$$\|\phi(t, x_0) - \phi(t, y_0)\| \leq e^{-t}\|x_0 - y_0\|$$

for all $x_0, y_0 \in \mathbb{R}^n$ and for all $t \geq 0$. In other words, for a fixed $t > 0$, the map $x \mapsto \phi(t, x)$ is a contraction. By the Banach fixed point theorem, for $\tau > 0$, there exists unique x^* such that $x^* = \phi(\tau, x^*)$. Then either x^* is an equilibrium point of (16) or it is part of a periodic orbit with period τ. If x^* were part of a periodic orbit, then every other point on the periodic orbit would be a fixed point of $\phi(\tau, \cdot)$, contradicting the uniqueness of the fixed point from the Banach fixed point theorem. Thus, we conclude that x^* is an equilibrium point of (16) and thus verifies $(\mathsf{Id} + \alpha\mathsf{F})(x^*) = u$. This proves that $\mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{F}}) = \mathbb{R}^n$. The proof for $\mathsf{R}_{\alpha\mathsf{F}}$ is a consequence of $\mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{F}}) = \mathbb{R}^n$. ■

We have the following corollary about inverses of strongly monotone mappings.

Corollary 18 (Lipschitz constants of inverses of strongly monotone operators) *Suppose* $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ *is continuous and strongly monotone with monotonicity parameter* $c > 0$. *Then* $\mathsf{F}^{-1} : \mathbb{R}^n \to \mathbb{R}^n$ *is a Lipschitz continuous mapping with Lipschitz constant estimate* $\mathrm{Lip}(\mathsf{F}^{-1}) \leq 1/c$.

Proof To see this fact, note that

$$\|\mathsf{F}(x) - \mathsf{F}(y)\|\|x - y\| \geq -\,[\![-(\mathsf{F}(x) - \mathsf{F}(y)), x - y]\!] \geq c\|x - y\|^2, \tag{18}$$

where the left hand inequality is the Cauchy-Schwarz inequality for WPs. So if $\mathsf{F}(x) = \mathsf{F}(y)$, then necessarily $x = y$, which implies that F^{-1} is a single-valued mapping. The fact that $\mathrm{Dom}(\mathsf{F}^{-1}) = \mathbb{R}^n$ follows the same argument as in Theorem 17 instead studying the differential equation $\dot{x} = -\mathsf{F}(x)$. Choosing $u, v \in \mathbb{R}^n$ and substituting $x = \mathsf{F}^{-1}(u), y = \mathsf{F}^{-1}(v)$ into (18), we conclude

$$\|u - v\| \geq c\|x - y\| = c\|\mathsf{F}^{-1}(u) - \mathsf{F}^{-1}(v)\|, \tag{19}$$

which shows that $\mathrm{Lip}(\mathsf{F}^{-1}) \leq 1/c$. ■

For each of $\mathsf{J}_{\alpha\mathsf{F}}, \mathsf{R}_{\alpha\mathsf{F}}$ and, $\mathsf{S}_{\alpha\mathsf{F}}$ we have now established that each of their domains is all of $\mathbb{R}^n$ and that fixed points of these operators correspond to zeros of F. In order to compute zeros of F, we aim to provide estimates of the Lipschitz constants of $\mathsf{J}_{\alpha\mathsf{F}}, \mathsf{R}_{\alpha\mathsf{F}}$, and $\mathsf{S}_{\alpha\mathsf{F}}$ as a function of α and the norm and show when these maps are either contractions or nonexpansive. The following lemmas characterize these Lipschitz estimates.

Lemma 19 (Lipschitz estimates of the forward step operator) *Let $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz continuous w.r.t. the norm $\|\cdot\|$ with constant $\mathrm{Lip}(\mathsf{F}) = \ell$.*

(i) Suppose F is monotone w.r.t. $\|\cdot\|$ with monotonicity parameter $c \geq 0$, then

$$\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq e^{-\alpha c} + e^{\alpha\ell} - 1 - \alpha\ell, \qquad \textit{for all } \alpha > 0. \tag{20}$$

(ii) Alternatively suppose $\|\cdot\|$ is a diagonally weighted ℓ_1 or ℓ_∞ norm and F is monotone w.r.t. $\|\cdot\|$ with monotonicity parameter $c \geq 0$, then

$$\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq 1 - \alpha c \leq 1, \qquad \textit{for all } \alpha \in \left]0, \frac{1}{\mathrm{diagL}(\mathsf{F})}\right], \tag{21}$$

where $\mathrm{diagL}(\mathsf{F}) := \sup_{x \in \mathbb{R}^n \setminus \Omega_\mathsf{F}} \max_{i \in \{1,\dots,n\}} (D\mathsf{F}(x))_{ii} \leq \ell$, where Ω_F is the measure zero set of points where F is not differentiable.

Proof Regarding item (i), we recall the inequality (Dahlquist, 1958, pp. 14), (Söderlind, 2006, Prop. 2.1)

$$\|e^{\alpha A}\| \leq e^{\alpha\mu(A)}, \qquad \text{for all } \alpha \geq 0, A \in \mathbb{R}^{n \times n}. \tag{22}$$

We additionally note that since F is Lipschitz continuous, $\mathsf{S}_{\alpha\mathsf{F}}$ is as well and $\mathsf{S}_{\alpha\mathsf{F}}$ has $\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq L$ if and only if $\|D\mathsf{S}_{\alpha\mathsf{F}}(x)\| \leq L$ for almost every $x \in \mathbb{R}^n$. Also we have that $D\mathsf{S}_{\alpha\mathsf{F}}(x) = I_n - \alpha D\mathsf{F}(x)$ everywhere it exists and that $D\mathsf{F}(x)$ satisfies $-\mu(-D\mathsf{F}(x)) \geq c$ and $\|D\mathsf{F}(x)\| \leq \ell$. In what follows, when we write $D\mathsf{F}(x)$, we mean for all x for which the Jacobian exists.

To derive an upper bound on $\|I_n - \alpha D\mathsf{F}(x)\|$, we define

$$S(x) := \sum_{i=2}^{\infty} \frac{(-\alpha)^i D\mathsf{F}(x)^i}{i!} = e^{-\alpha D\mathsf{F}(x)} - I_n + \alpha D\mathsf{F}(x)$$

and it is straightforward to see that $\|S(x)\| \leq \sum_{i=2}^{\infty} \frac{\alpha^i \|D\mathsf{F}(x)\|^i}{i!} \leq e^{\alpha\ell} - 1 - \alpha\ell$. Moreover,

$$\begin{aligned} \|I_n - \alpha D\mathsf{F}(x)\| &\leq \|e^{-\alpha D\mathsf{F}(x)}\| + \|S(x)\| \leq e^{\alpha\mu(-D\mathsf{F}(x))} + e^{\alpha\ell} - 1 - \alpha\ell \\ &\leq e^{-\alpha c} + e^{\alpha\ell} - 1 - \alpha\ell. \end{aligned} \tag{23}$$

Since this bound holds for all x for which $D\mathsf{S}_{\alpha\mathsf{F}}(x)$ exists, the result is proved.

Regarding item (ii), for every $x \in \mathbb{R}^n$ for which $D\mathsf{F}(x)$ exists,

$$\|I_n - \alpha D\mathsf{F}(x)\|_{\infty,[\eta]^{-1}} = \max_{i\in\{1,\dots,n\}} |1-\alpha(D\mathsf{F}(x))_{ii}| + \sum_{j=1,j\neq i}^{n} |-\alpha(D\mathsf{F}(x))_{ij}|\frac{\eta_j}{\eta_i} \tag{24}$$

$$= \max_{i\in\{1,\dots,n\}} 1-\alpha(D\mathsf{F}(x))_{ii} + \sum_{j=1,j\neq i}^{n} |-\alpha(D\mathsf{F}(x))_{ij}|\frac{\eta_j}{\eta_i} \tag{25}$$

$$= 1 + \alpha\mu(-D\mathsf{F}(x)) \leq 1 - \alpha c, \tag{26}$$

where (25) holds because $0 < \alpha \leq \frac{1}{\mathrm{diagL}(\mathsf{F})}$ so that $1 - \alpha(D\mathsf{F}(x))_{ii} \geq 0$ for all $x \in \mathbb{R}^n, i \in \{1,\dots,n\}$ and (26) is due to the formula for $\mu_{\infty,[\eta]^{-1}}$. The proof for $\mu_{1,[\eta]}$ is analogous, replacing row sums by column sums, and is omitted. ■

Remark 20 *If $c > 0$, then for small enough $\alpha > 0$, one can make the upper bound on $\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}})$ in (20) less than unity. In particular, one can show that minimizing the upper bound (20) yields the optimal step size $\alpha_{\mathrm{opt}} = \frac{1}{\ell}\ln(s(\gamma))$ and contraction factor $s(\gamma) + s(\gamma)^{-\gamma} - 1 - \ln(s(\gamma))$, where $\gamma = c/\ell \leq 1$ and $s(\gamma)$ is the unique solution to the transcendental equation $s - 1 - \gamma s^{-\gamma} = 0$.*

Remark 21 *Note that for general norms, if F is monotone, but not strongly monotone, then $\mathsf{S}_{\alpha\mathsf{F}}$ 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 w.r.t. the ℓ_2 norm, but $\mathsf{S}_{\alpha\mathsf{F}}$ is not nonexpansive for any $\alpha > 0$. On the other hand, Lemma 19(ii) implies that if F is monotone w.r.t. a diagonally weighted ℓ_1 or ℓ_∞ norm, then $\mathsf{S}_{\alpha\mathsf{F}}$ is nonexpansive for sufficiently small α.*

We plot the upper bounds on the estimates of $\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}})$ as a function of α and choice of norm for fixed parameters c and ℓ in Figure 1.

Lemma 22 (Lipschitz constant of the resolvent operator) *Suppose $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and monotone with monotonicity parameter $c \geq 0$. Then,*

$$\mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) \leq \frac{1}{1+\alpha c}, \qquad \textit{for all } \alpha > 0. \tag{27}$$

Proof We observe that $\mathsf{Id} + \alpha\mathsf{F}$ is strongly monotone with parameter $1 + \alpha c$. Then by Corollary 18, the result holds. ■

Lemma 23 (Lipschitz constant of the reflected resolvent) *Suppose $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with constant ℓ w.r.t. a norm $\|\cdot\|$.*

(i) Suppose F is monotone w.r.t. $\|\cdot\|$ with monotonicity parameter $c \geq 0$. Then

$$\mathrm{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) \leq \frac{e^{-\alpha c} + e^{\alpha\ell} - 1 - \alpha\ell}{1+\alpha c}, \qquad \textit{for all } \alpha > 0. \tag{28}$$

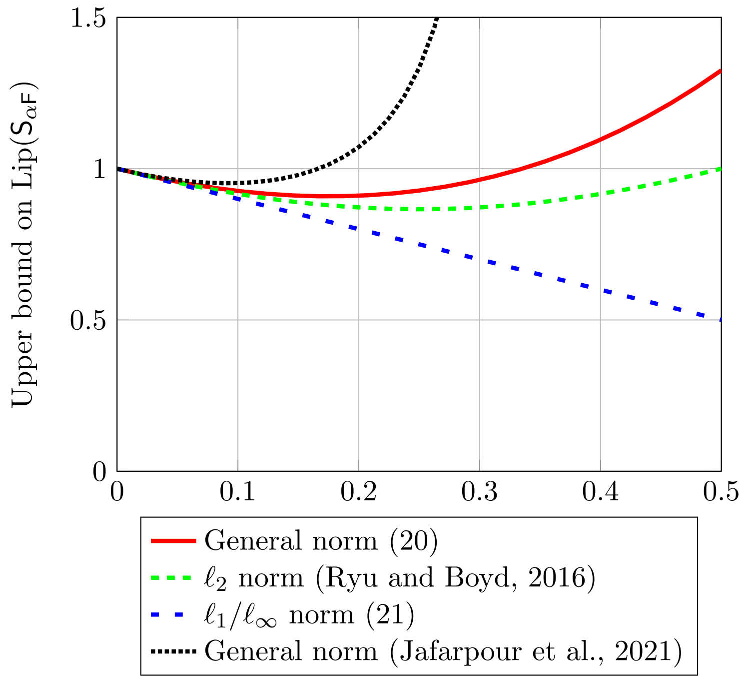

Figure 1: Plots of upper bounds of $\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}})$ with respect to different norms. We fix parameters $c = 1, \ell = 2$ and vary the choice of norm. The solid red curve corresponds to the Lipschitz estimate (20) for arbitrary norms, the densely dashed green curve corresponds to the estimate $\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq \sqrt{1 - 2\alpha c + \alpha^2\ell^2}$ from (Ryu and Boyd, 2016, pp. 16) for the ℓ_2 norm, the loosely dashed blue curve corresponds to the estimate (21) for diagonally-weighted ℓ_1/ℓ_∞ norms which is valid on the interval $]0, \frac{1}{\mathrm{diagL}(\mathsf{F})}]$. Finally, the dotted black curve corresponds to the estimate $\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq \big(1 + \alpha c - \frac{\alpha^2\ell^2}{1-\alpha\ell}\big)^{-1}$ previously established in (Jafarpour et al., 2021, Theorem 1). We see that the estimate (20) is a tighter estimate than the estimate from (Jafarpour et al., 2021) and that Lipschitz upper bounds are least conservative in the case of diagonally-weighted ℓ_1/ℓ_∞ norms.

(ii) Alternatively suppose $\|\cdot\|$ is a diagonally weighted ℓ_1 or ℓ_∞ norm. Moreover, suppose F is monotone w.r.t. $\|\cdot\|$ with monotonicity parameter $c \geq 0$. Then,

$$\mathrm{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) \leq \frac{1 - \alpha c}{1 + \alpha c} \leq 1, \qquad \text{for all } \alpha \in \Big]0, \frac{1}{\mathrm{diagL}(\mathsf{F})}\Big]. \tag{29}$$

Proof Recall from (Ryu and Boyd, 2016, pp. 21) that since F is monotone and continuous, we have that $\mathsf{R}_{\alpha\mathsf{F}} = \mathsf{S}_{\alpha\mathsf{F}} \circ \mathsf{J}_{\alpha\mathsf{F}}$. Both results then follow from $\mathrm{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) \leq \mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}})\,\mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{F}})$ and the bounds on $\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}})$ from Lemma 19 and on $\mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{F}})$ from Lemma 22. ■

Lemma 23 stands in striking contrast with results on monotone operators in Hilbert spaces which says that for any maximally monotone operator,[5] F, the reflected resolvent of F with parameter $\alpha > 0$ is nonexpansive for every $\alpha > 0$. Indeed, in the non-Euclidean case, this property cannot be recovered as is demonstrated in the following example.

Example 2 *Consider the linear mapping* $\mathsf{F}(x) = Ax = \left(\begin{smallmatrix} 2 & -2 \\ 1 & 1 \end{smallmatrix}\right) x$. F *is monotone w.r.t. the* ℓ_∞ *norm since* $-\mu_\infty(-A) = -\mu_\infty \left(\begin{smallmatrix} -2 & 2 \\ -1 & -1 \end{smallmatrix}\right) = 0$. *For* $\alpha = 2$*, we compute*

$$\mathsf{J}_{\alpha\mathsf{F}}(x) = \begin{pmatrix} 3/23 & 4/23 \\ -2/23 & 5/23 \end{pmatrix} x, \qquad \mathsf{R}_{\alpha\mathsf{F}}(x) = \begin{pmatrix} -17/23 & 8/23 \\ -4/23 & -13/23 \end{pmatrix} x.$$

Thus, $\mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) = 7/23$ *and* $\mathrm{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) = 25/23$. *In other words, for* $\alpha = 2$, $\mathsf{J}_{\alpha\mathsf{F}}$ *is a contraction and* $\mathsf{R}_{\alpha\mathsf{F}}$ *is not nonexpansive.*

Despite this key divergence from the classical theory, we will still be able to prove convergence of iterative algorithms involving the reflected resolvent under suitable assumptions on the parameter $\alpha > 0$.

4. Finding Zeros of Non-Euclidean Monotone Operators

For a mapping $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ which is continuous and monotone, consider the problem of finding an $x \in \mathbb{R}^n$ that satisfies

$$\mathsf{F}(x) = \mathbb{0}_n. \tag{30}$$

Without further assumptions on F, this problem may have no solutions or nonunique solutions. First we provide a preliminary sufficient condition for existence and uniqueness of a solution.

Lemma 24 (Uniqueness of zeros of strongly monotone maps) *Suppose* $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ *is continuous and strongly monotone. Then* $\mathrm{Zero}(\mathsf{F})$ *is a singleton.*

Proof We have that $\mathrm{Zero}(\mathsf{F}) = \mathrm{Fix}(\mathsf{J}_{\alpha\mathsf{F}})$ for $\alpha > 0$. By Lemma 22, we have that $\mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) \leq 1/(1 + \alpha c) < 1$, where $c > 0$ is the monotonicity parameter of F. Then by the Banach fixed point theorem, $\mathsf{J}_{\alpha\mathsf{F}}$ has a unique fixed point and thus $\mathrm{Zero}(\mathsf{F})$ is a singleton. ■

Alternatively, if F is continuous and monotone, then we study fixed points of the nonexpansive map $\mathsf{J}_{\alpha\mathsf{F}}$, which may or may not exist and may or may not be unique. In what follows, we will study the case where it is known a priori that zeros of F exist but need not be unique.

We show that the most known algorithms for finding zeros of monotone operators on Hilbert spaces (see, e.g., (Ryu and Boyd, 2016)) can be generalized to non-Euclidean monotone operators using our framework and, furthermore, explicitly estimate the convergence rate of these methods.

5. Recall that in monotone operator theory on Hilbert spaces, a set-valued mapping $\mathsf{F} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is maximally monotone if it is monotone and there does not exist another monotone operator, G, whose graph properly contains the graph of F. See (Bauschke and Combettes, 2017, Sec. 20.2) for more details.

4.1 The Forward Step Method

Algorithm 25 (Forward step method) *The* forward step method *corresponds to the fixed point iteration*

$$x^{k+1} = \mathsf{S}_{\alpha\mathsf{F}}(x^k) = x^k - \alpha\mathsf{F}(x^k). \tag{31}$$

Theorem 26 (Convergence guarantees for the forward step method) *Let $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is Lipschitz continuous with constant ℓ w.r.t. a norm $\|\cdot\|$ and let $x^0 \in \mathbb{R}^n$ be arbitrary.*

(i) Suppose F is strongly monotone w.r.t. $\|\cdot\|$ with monotonicity parameter $c > 0$. Then the iteration (31) *converges to the unique zero, x^*, of F for every $\alpha \in \,]0, \alpha^*[$. Moreover, for every $k \in \mathbb{Z}_{\geq 0}$,*

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

where α^* is the unique positive value of α that satisfies $e^{-\alpha^* c} + e^{\alpha^*\ell} = 2 + \alpha^*\ell$.

(ii) Alternatively suppose $\|\cdot\|$ is a diagonally-weighted ℓ_1 or ℓ_∞ norm and F is strongly monotone w.r.t. $\|\cdot\|$ with monotonicity parameter $c > 0$. Then the iteration (31) *converges to the unique zero, x^*, of F for every $\alpha \in \,]0, \frac{1}{\mathrm{diagL}(\mathsf{F})}]$. Moreover, for every $k \in \mathbb{Z}_{\geq 0}$,*

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

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

(iii) Alternatively suppose $\|\cdot\|$ is a diagonally weighted ℓ_1 or ℓ_∞ norm and F is monotone w.r.t. $\|\cdot\|$. Then $\mathrm{Zero}(\mathsf{F}) \neq \emptyset$ *implies the iteration* (31) *converges to an element of* $\mathrm{Zero}(\mathsf{F})$ *for every $\alpha \in \,]0, \frac{1}{\mathrm{diagL}(\mathsf{F})}[$.*

Proof Regarding statement (i), from Lemma 19(i), we have that $\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq e^{-\alpha c} + e^{\alpha\ell} - 1 - \alpha\ell$. It is straightforward to compute that at $\alpha = \alpha^*$, $\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq 1$ and for $\alpha \in \,]0, \alpha^*[$ we have that $\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) < 1$. Thus, $\mathsf{S}_{\alpha\mathsf{F}}$ is a contraction and fixed points of $\mathsf{S}_{\alpha\mathsf{F}}$ correspond to zeros of F. Then by the Banach fixed point theorem, the result follows.

Regarding statement (ii), Lemma 19(ii) implies that $\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) = 1 - \alpha c < 1$ for all $\alpha \in \,]0, 1/\operatorname{diagL}(\mathsf{F})]$. The result is then a consequence of the Banach fixed point theorem.

Regarding statement (iii), since F is monotone w.r.t. a diagonally weighted ℓ_1 or ℓ_∞ norm, $\mathsf{S}_{\alpha\mathsf{F}}$ is nonexpansive for $\alpha \in \,]0, 1/\operatorname{diagL}(\mathsf{F})]$ by Lemma 19(ii). Moreover, for every $\alpha \in \,]0, 1/\operatorname{diagL}(\mathsf{F})[$, there exists $\theta \in \,]0, 1[$ such that $\mathsf{S}_{\alpha\mathsf{F}} = (1-\theta)\mathsf{Id} + \theta\mathsf{S}_{\tilde{\alpha}\mathsf{F}}$, for some $\tilde{\alpha} \in \,]0, 1/\operatorname{diagL}(\mathsf{F})]$. Therefore the iteration (31) is the Krasnosel'skii–Mann iteration of the nonexpansive operator $\mathsf{S}_{\tilde{\alpha}\mathsf{F}}$ and Lemma 11 implies the result. ■

Note that Theorem 26(iii) is a direct consequence of the fact that the forward step operator is nonexpansive for suitable $\alpha > 0$ when the mapping is monotone w.r.t. a diagonally-weighted ℓ_1 or ℓ_∞ norm, a fact which need not hold when the mapping is monotone w.r.t. a different norm, e.g., a Hilbert one. See the relevant discussion in Remark 21 for an example of a mapping, F, which is monotone with respect to the ℓ_2 norm but $\mathsf{S}_{\alpha\mathsf{F}}$ is not nonexpansive for any $\alpha > 0$.

4.2 The Proximal Point Method

Algorithm 27 (Proximal point method) *The* proximal point method *corresponds to the fixed point iteration*

$$x^{k+1} = \mathsf{J}_{\alpha\mathsf{F}}(x^k) = (\mathsf{Id} + \alpha\mathsf{F})^{-1}(x^k). \tag{32}$$

Theorem 28 (Convergence guarantees for the proximal point method) *Suppose* $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ *is continuous and let* $x^0 \in \mathbb{R}^n$ *be arbitrary.*

(i) Suppose F *is strongly monotone w.r.t. a norm* $\|\cdot\|$ *with monotonicity parameter* $c > 0$*. Then the iteration* (32) *converges to the unique zero,* x^**, of* F *for every* $\alpha \in\,]0, \infty[$*. Moreover, for every* $k \in \mathbb{Z}_{\geq 0}$*, the iteration satisfies*

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

(ii) Alternatively suppose F *is monotone and globally Lipschitz continuous w.r.t. a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm* $\|\cdot\|$ *and* $\mathrm{diagL}(\mathsf{F}) \neq 0$*. Then if* $\mathrm{Zero}(\mathsf{F}) \neq \emptyset$*, the iteration* (32) *converges to an element of* $\mathrm{Zero}(\mathsf{F})$ *for every* $\alpha \in\,]0, \infty[$*.*

Proof Regarding statement (i), Lemma 22 provides the Lipschitz estimate $\mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) \leq \frac{1}{1+\alpha c} < 1$ for all $\alpha > 0$. Thus $\mathsf{J}_{\alpha\mathsf{F}}$ is a contraction and since fixed points of $\mathsf{J}_{\alpha\mathsf{F}}$ correspond to zeros of F, the Banach fixed point theorem implies the result.

Regarding statement (ii), we will demonstrate that the iteration (32) is a Krasnosel'skii–Mann iteration of a suitably-defined nonexpansive mapping. To this end, let $\theta \in\,]0, 1[$ be arbitrary and consider the auxiliary mapping $\overline{\mathsf{R}}^\theta_{\alpha\mathsf{F}} : \mathbb{R}^n \to \mathbb{R}^n$ given by $\overline{\mathsf{R}}^\theta_{\alpha\mathsf{F}} := \frac{\mathsf{J}_{\alpha\mathsf{F}}}{\theta} - \frac{1-\theta}{\theta}\mathsf{Id}$. Then it is straightforward to compute

$$\begin{aligned}
\overline{\mathsf{R}}^\theta_{\alpha\mathsf{F}} &= \frac{(\mathsf{Id} + \alpha\mathsf{F})^{-1}}{\theta} - \frac{1-\theta}{\theta}(\mathsf{Id} + \alpha\mathsf{F}) \circ (\mathsf{Id} + \alpha\mathsf{F})^{-1} \\
&= \Big(\frac{\mathsf{Id}}{\theta} - \frac{1-\theta}{\theta}(\mathsf{Id} + \alpha\mathsf{F})\Big) \circ (\mathsf{Id} + \alpha\mathsf{F})^{-1} = \Big(\mathsf{Id} - \frac{(1-\theta)\alpha}{\theta}\mathsf{F}\Big) \circ \mathsf{J}_{\alpha\mathsf{F}} = \mathsf{S}_{\frac{1-\theta}{\theta}\alpha\mathsf{F}} \circ \mathsf{J}_{\alpha\mathsf{F}}.
\end{aligned}$$

Moreover, $\mathsf{J}_{\alpha\mathsf{F}}$ is nonexpansive by Lemma 22, and by Lemma 19(ii), $\mathrm{Lip}(\mathsf{S}_{\frac{1-\theta}{\theta}\alpha\mathsf{F}}) \leq 1$, for all $\alpha \in\,]0, \frac{1-\theta}{\theta\,\mathrm{diagL}(\mathsf{F})}]$. We conclude that $\mathrm{Lip}(\overline{\mathsf{R}}^\theta_{\alpha\mathsf{F}}) \leq \mathrm{Lip}(\mathsf{S}_{\frac{1-\theta}{\theta}\alpha\mathsf{F}})\,\mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{F}}) \leq 1$ for $\alpha \in\,]0, \frac{1-\theta}{\theta\,\mathrm{diagL}(\mathsf{F})}]$ which implies that $\overline{\mathsf{R}}^\theta_{\mathsf{F}}$ is nonexpansive for all α in this range.

Let $\alpha > 0$ be arbitrary. Then for any[6] $\theta \leq \frac{1}{1+\alpha\,\mathrm{diagL}(\mathsf{F})} \in\,]0, 1[$, we have that $\mathsf{J}_{\alpha\mathsf{F}} = (1-\theta)\mathsf{Id} + \theta\overline{\mathsf{R}}^\theta_{\alpha\mathsf{F}}$, and $\overline{\mathsf{R}}^\theta_{\mathsf{F}}$ is nonexpansive since $\alpha \in\,]0, \frac{1-\theta}{\theta\,\mathrm{diagL}(\mathsf{F})}]$. Thus, the iteration (32) is the Krasnosel'skii–Mann iteration for $\overline{\mathsf{R}}^\theta_{\alpha\mathsf{F}}$ and the result follows from Lemma 11. ■

6. Note that $\frac{1}{1+\alpha\,\mathrm{diagL}(\mathsf{F})} \in\,]0, 1[$ holds under the assumption $\mathrm{diagL}(\mathsf{F}) \neq 0$ since $\mathrm{diagL}(\mathsf{F}) \geq 0$ for any monotone F.

Remark 29 *Theorem 28(ii) is an analog of the classical result in monotone operator theory on Hilbert spaces which states that the resolvent of a maximally monotone operator is firmly nonexpansive (Minty, 1962) and (Bauschke and Combettes, 2017, Prop. 23.8). This firm nonexpansiveness is a consequence of the fact that the reflected resolvent of a maximally monotone operator with respect to a Euclidean norm is always nonexpansive and* $\mathsf{J}_{\alpha\mathsf{F}} = \frac{1}{2}\mathsf{Id} + \frac{1}{2}\mathsf{R}_{\alpha\mathsf{F}}$. *Note that this property need not hold when* F *is monotone with respect to a non-Euclidean norm but we are able to show that in the case of diagonally-weighted* ℓ_1/ℓ_∞ *norms, a similar result holds.*

4.3 The Cayley Method

Algorithm 30 (Cayley method) *The* Cayley method *corresponds to the iteration*

$$x^{k+1} = \mathsf{R}_{\alpha\mathsf{F}}(x^k) = 2(\mathsf{Id} + \alpha\mathsf{F})^{-1}(x^k) - x^k. \tag{33}$$

Theorem 31 (Convergence guarantees for the Cayley method) *Suppose* $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ *is Lipschitz continuous with constant* ℓ *w.r.t. a norm* $\|\cdot\|$ *and let* $x^0 \in \mathbb{R}^n$ *be arbitrary.*

(i) Suppose F *is strongly monotone w.r.t.* $\|\cdot\|$ *with monotonicity parameter* $c > 0$. *Then the iteration* (33) *converges to the unique zero,* x^*, *of* F *for sufficiently small* $\alpha > 0$. *Moreover, for every* $k \in \mathbb{Z}_{\geq 0}$, *the iteration satisfies*

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

(ii) Alternatively suppose $\|\cdot\|$ *is a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm and* F *is strongly monotone w.r.t.* $\|\cdot\|$ *with monotonicity parameter* $c > 0$. *Then the iteration* (33) *converges to the unique zero,* x^*, *of* F *for every* $\alpha \in \left]0, \frac{1}{\operatorname{diagL}(\mathsf{F})}\right]$. *Moreover, for every* $k \in \mathbb{Z}_{\geq 0}$, *the iteration satisfies*

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

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

(iii) Alternatively suppose $\|\cdot\|$ *is a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm and* F *is monotone w.r.t.* $\|\cdot\|$. *Then if* $\operatorname{Zero}(\mathsf{F}) \neq \emptyset$, *the Krasnosel'skii–Mann iteration with* $\theta = 1/2$

$$x^{k+1} = \frac{1}{2}x^k + \frac{1}{2}\mathsf{R}_{\alpha\mathsf{F}}(x^k)$$

correspond to the proximal point iteration (32), *which is guaranteed to converge to an element of* $\operatorname{Zero}(\mathsf{F})$ *for every* $\alpha \in \left]0, \infty\right[$.

Proof Regarding statement (i), from Lemma 23(i), we have that $\operatorname{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) \leq (e^{-\alpha c} + e^{\alpha\ell} - 1 - \alpha\ell)/(1 + \alpha c)$ which is less than unity for small enough $\alpha > 0$. Thus, for small enough α, $\mathsf{R}_{\alpha\mathsf{F}}$ is a contraction and fixed points of $\mathsf{R}_{\alpha\mathsf{F}}$ correspond to zeros of F. Thus, by the Banach fixed point theorem, the result follows.

Algorithm, Iterated map	F strongly monotone and globally Lipschitz continuous					
	ℓ_2		General norm		Diagonally weighted ℓ_1 or ℓ_∞	
	α range	Optimal Lip	α range	Optimal Lip	α range	Optimal Lip
Forward step, $\mathsf{S}_{\alpha\mathsf{F}}$	$\left]0, \frac{2c}{\ell^2}\right[$	$1 - \frac{1}{2\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right)$	$]0, \alpha^*[$	$1 - \frac{1}{2\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right)$	$\left]0, \frac{1}{\operatorname{diagL}(\mathsf{F})}\right]$	$1 - \frac{1}{\kappa_\infty}$
Proximal point, $\mathsf{J}_{\alpha\mathsf{F}}$	$]0, \infty[$	A.S.	$]0, \infty[$	A.S.	$]0, \infty[$	A.S.
Cayley, $\mathsf{R}_{\alpha\mathsf{F}}$	$]0, \infty[$	$1 - \frac{1}{2\kappa} + \mathcal{O}\left(\frac{1}{\kappa^2}\right)$	$]0, \alpha^*[$	$1 - \frac{2}{\kappa^2} + \mathcal{O}\left(\frac{1}{\kappa^3}\right)$	$\left]0, \frac{1}{\operatorname{diagL}(\mathsf{F})}\right]$	$1 - \frac{2}{\kappa_\infty} + \mathcal{O}\left(\frac{1}{\kappa_\infty^2}\right)$

Table 1: Table of step size ranges and Lipschitz constants for three algorithms for finding zeros of monotone operators with respect to arbitrary norms. For F strongly monotone and Lipschitz continuous, let c be its monotonicity parameter (with respect to the appropriate norm), ℓ its appropriate Lipschitz constant, and $\operatorname{diagL}(\mathsf{F}) := \sup_{x \in \mathbb{R}^n \setminus \Omega_\mathsf{F}} \max_{i \in \{1,\dots,n\}} (D\mathsf{F}(x))_{ii} \leq \ell$. Additionally, $\kappa := \ell/c \geq 1$, $\kappa_\infty := \operatorname{diagL}(\mathsf{F})/c \in [1, \kappa]$, and α^* is the unique positive solution to $e^{-\alpha^* c} + e^{\alpha^* \ell} = 2 + \alpha^* \ell$. *A.S.* means the Lipschitz constant can be made arbitrarily small. We do not assume that the strongly monotone F is the gradient of a strongly convex function.

Regarding statement (ii), Lemma 23(ii) implies that $\operatorname{Lip}(\mathsf{R}_{\alpha\mathsf{F}}) \leq (1 - \alpha c)/(1 + \alpha c) < 1$ for $\alpha \in]0, 1/\operatorname{diagL}(\mathsf{F})]$. The result is then a consequence of the Banach fixed point theorem.

Statement (iii) holds since $\frac{1}{2}\mathsf{Id} + \frac{1}{2}(2\mathsf{J}_{\alpha\mathsf{F}} - \mathsf{Id}) = \mathsf{J}_{\alpha\mathsf{F}}$, and convergence follows by Theorem 28(ii) since $\operatorname{Zero}(\mathsf{F}) \neq \emptyset$. ■

Table 1 summarizes and compares the range of step sizes and Lipschitz estimates as provided by the classical monotone operator theory for the ℓ_2 norm (Ryu and Boyd, 2016, pp. 16 and 20) and by Theorems 26, 28, and 31 for general and diagonally-weighted ℓ_1/ℓ_∞ norms.

5. Finding Zeros of a Sum of Non-Euclidean Monotone Operators

In many instances, one may wish to execute the proximal point method, Algorithm 27, to compute a zero of a continuous monotone mapping $\mathsf{N} : \mathbb{R}^n \to \mathbb{R}^n$. However, the implementation of the iteration (32) 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 mappings 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 $(\mathsf{F} + \mathsf{G})(x) = \mathbb{0}_n$ 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

$$(\mathsf{F} + \mathsf{G})(x) = \mathbb{0}_n, \tag{34}$$

where $\mathsf{F}, \mathsf{G} : \mathbb{R}^n \to \mathbb{R}^n$ are continuous and monotone w.r.t. a diagonally weighted ℓ_1 or ℓ_∞ norm.[7] In particular, we focus on the forward-backward, Peaceman-Rachford, and Douglas-Rachford splitting algorithms. For some extensions of the theory to set-valued mappings, we refer to Section 6.1.

5.1 Forward-Backward Splitting

Algorithm 32 (Forward-backward splitting) *Assume $\alpha > 0$. Then in (Ryu and Boyd, 2016, Section 7.1) it is shown that*

$$(\mathsf{F} + \mathsf{G})(x) = \mathbb{0}_n \quad \iff x = (\mathsf{J}_{\alpha\mathsf{G}} \circ \mathsf{S}_{\alpha\mathsf{F}})(x).$$

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

$$x^{k+1} = \mathsf{J}_{\alpha\mathsf{G}}(x^k - \alpha\mathsf{F}(x^k)). \tag{35}$$

If both F *and* G *are monotone, define the* averaged forward-backward splitting iteration

$$x^{k+1} = \frac{1}{2}x^k + \frac{1}{2}\mathsf{J}_{\alpha\mathsf{G}}(x^k - \alpha\mathsf{F}(x^k)). \tag{36}$$

Theorem 33 (Convergence guarantees for forward-backward splitting) *Let $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ be Lipschitz continuous w.r.t. a diagonally weighted ℓ_1 or ℓ_∞ norm $\|\cdot\|$, $\mathsf{G} : \mathbb{R}^n \to \mathbb{R}^n$ be continuous and monotone w.r.t. the same norm, and let $x^0 \in \mathbb{R}^n$ be arbitrary.*

(i) Suppose F is strongly monotone w.r.t. $\|\cdot\|$ with monotonicity parameter $c > 0$, then the iteration (35) *converges to the unique zero, x^*, of $\mathsf{F} + \mathsf{G}$ for every $\alpha \in \left]0, \frac{1}{\mathrm{diagL}(\mathsf{F})}\right]$. Moreover, for every $k \in \mathbb{Z}_{\geq 0}$, the iteration satisfies*

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

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

(ii) If F is monotone w.r.t. $\|\cdot\|$ and $\mathrm{Zero}(\mathsf{F} + \mathsf{G}) \neq \emptyset$, *then the iteration* (36) *converges to an element of* $\mathrm{Zero}(\mathsf{F} + \mathsf{G})$ *for every $\alpha \in \left]0, \frac{1}{\mathrm{diagL}(\mathsf{F})}\right]$.*

Proof Regarding statement (i), Lemmas 19(ii) and 22 together imply that $\mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{G}} \circ \mathsf{S}_{\alpha\mathsf{F}}) \leq \mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{G}})\,\mathrm{Lip}(\mathsf{S}_{\alpha\mathsf{F}}) \leq 1 - \alpha c < 1$ for all $\alpha \in \left]0, 1/\mathrm{diagL}(\mathsf{F})\right]$. Then since $\mathrm{Fix}(\mathsf{J}_{\alpha\mathsf{F}} \circ \mathsf{S}_{\alpha\mathsf{F}}) = \mathrm{Zero}(\mathsf{F} + \mathsf{G})$, the result is then a consequence of the Banach fixed point theorem.

Statement ii follows from $\mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{G}} \circ \mathsf{S}_{\alpha\mathsf{F}}) \leq 1$ and that the iteration (36) is the Krasnosel'skii–Mann iteration of the nonexpansive mapping $\mathsf{J}_{\alpha\mathsf{G}} \circ \mathsf{S}_{\alpha\mathsf{F}}$ with $\theta = 1/2$. ■

Remark 34 (Comparison with standard forward-backward splitting convergence criteria)

7. The results that follow can also be generalized to arbitrary norms using the Lipschitz estimates derived for $\mathsf{J}_{\alpha\mathsf{F}}, \mathsf{R}_{\alpha\mathsf{F}}$, and $\mathsf{S}_{\alpha\mathsf{F}}$ in Section 3.2.

- *Compared to the Hilbert case, in the non-Euclidean setting, if both* F *and* G *are monotone, then iteration* (36) *must be applied to compute a zero of* $\mathsf{F}+\mathsf{G}$. *In the Hilbert case, iteration* (35) *may be used instead since the composition of averaged operators is averaged. For non-Hilbert norms, the composition of two averaged operators need not be averaged.*

- *In monotone operator theory on Hilbert spaces, Lipschitz continuity of* F *is not sufficient for the convergence of the iteration* (35). *Instead, a standard sufficient condition for convergence is cocoercivity of* F, *see (Mercier, 1980) and (Bauschke and Combettes, 2017, Theorem 26.14). In the case of diagonally-weighted* ℓ_1/ℓ_∞ *norms, Lipschitz continuity is sufficient for convergence. This fact is due to the nonexpansiveness of* $\mathsf{S}_{\alpha\mathsf{F}}$ *for* ℓ_1/ℓ_∞ *monotone* F *and small enough* $\alpha>0$ *as discussed in Remark 21.*

5.2 Peaceman-Rachford and Douglas-Rachford Splitting

Algorithm 35 (Peaceman-Rachford and Douglas-Rachford splitting) *Let* $\alpha>0$. *Then in (Ryu and Boyd, 2016, Section 7.3), it is shown that*

$$(\mathsf{F}+\mathsf{G})(x)=\mathbb{0}_n \quad\Longleftrightarrow\quad (\mathsf{R}_{\alpha\mathsf{F}}\circ\mathsf{R}_{\alpha\mathsf{G}})(z)=z \text{ and } x=\mathsf{J}_{\alpha\mathsf{G}}(z). \tag{37}$$

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

$$\begin{aligned} x^{k+1} &= \mathsf{J}_{\alpha\mathsf{G}}(z^k),\\ z^{k+1} &= z^k + 2\mathsf{J}_{\alpha\mathsf{F}}(2x^{k+1}-z^k) - 2x^{k+1}. \end{aligned} \tag{38}$$

If both F *and* G *are monotone, the term* $\mathsf{R}_{\alpha\mathsf{F}}\circ\mathsf{R}_{\alpha\mathsf{G}}$ *in* (37) *is averaged to yield*

$$(\mathsf{F}+\mathsf{G})(x)=\mathbb{0}_n \quad\Longleftrightarrow\quad \Big(\frac{1}{2}\mathsf{Id}+\frac{1}{2}\mathsf{R}_{\alpha\mathsf{F}}\circ\mathsf{R}_{\alpha\mathsf{G}}\Big)(z)=z \text{ and } x=\mathsf{J}_{\alpha\mathsf{G}}(z). \tag{39}$$

The fixed point iteration corresponding to (39) *is called the* Douglas-Rachford splitting method *and is given by*

$$\begin{aligned} x^{k+1} &= \mathsf{J}_{\alpha\mathsf{G}}(z^k),\\ z^{k+1} &= z^k + \mathsf{J}_{\alpha\mathsf{F}}(2x^{k+1}-z^k) - x^{k+1}. \end{aligned} \tag{40}$$

Theorem 36 (Convergence guarantees for Peaceman-Rachford and Douglas-Rachford splitting) *Let both* $\mathsf{F}:\mathbb{R}^n\to\mathbb{R}^n$ *and* $\mathsf{G}:\mathbb{R}^n\to\mathbb{R}^n$ *be Lipschitz continuous w.r.t. a diagonally weighted* ℓ_1 *or* ℓ_∞ *norm* $\|\cdot\|$, *let* G *be monotone w.r.t. the same norm, and let* $x^0\in\mathbb{R}^n$.

(i) Suppose F *is strongly monotone w.r.t.* $\|\cdot\|$ *with monotonicity parameter* $c>0$. *Then the sequence of* $\{x_k\}_{k=0}^\infty$ *generated by the iteration* (38) *converges to the unique zero,* x^*, *of* $\mathsf{F}+\mathsf{G}$ *for every* $\alpha\in\left]0,\min\left\{\frac{1}{\mathrm{diagL}(\mathsf{F})},\frac{1}{\mathrm{diagL}(\mathsf{G})}\right\}\right]$. *Moreover, for every* $k\in\mathbb{Z}_{\geq0}$, *the iteration satisfies*

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

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

(ii) *Alternatively suppose* F *is monotone w.r.t.* $\|\cdot\|$ *and* $\mathrm{Zero}(\mathsf{F}+\mathsf{G}) \neq \emptyset$. *Then the sequence of* $\{x_k\}_{k=0}^{\infty}$ *generated by the iteration* (40) *converges to an element of* $\mathrm{Zero}(\mathsf{F}+\mathsf{G})$ *for every* $\alpha \in \left]0, \min\left\{\frac{1}{\mathrm{diagL}(\mathsf{F})}, \frac{1}{\mathrm{diagL}(\mathsf{G})}\right\}\right]$.

Proof Regarding statement (i), by Lemma 23(ii), we have that

$$\mathrm{Lip}(\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}}) \leq \mathrm{Lip}(\mathsf{R}_{\alpha\mathsf{F}})\,\mathrm{Lip}(\mathsf{R}_{\alpha\mathsf{G}}) \leq \frac{1-\alpha c}{1+\alpha c} < 1$$

for $\alpha \in \left]0, \min\{1/\mathrm{diagL}(\mathsf{F}), 1/\mathrm{diagL}(\mathsf{G})\}\right]$. Then since $\mathrm{Lip}(\mathsf{J}_{\alpha\mathsf{G}})$ is nonexpansive, the Banach fixed point theorem implies the result.

Statement (ii) holds because Lemma 23(ii) implies $\mathrm{Lip}(\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}}) \leq 1$. Then the iteration (40) converges because of Lemma 11. ∎

Compared to classical criteria for the convergence of the Douglas-Rachford iteration, Theorem 36 requires Lipschitz continuity of F and G in order to utilize the Lipschitz estimates for the reflected resolvents $\mathsf{R}_{\alpha\mathsf{F}}$ and $\mathsf{R}_{\alpha\mathsf{G}}$. Moreover, the parameter $\alpha > 0$ must be chosen small enough in the non-Euclidean setting whereas convergence is guaranteed for any choice of α in the Hilbert case. This is because the reflected resolvent is only nonexpansive for a certain range of α when the norm is not a Hilbert one, see Example 2.

6. Set-Valued Inclusions and an Application to Recurrent Neural Networks

6.1 Set-Valued Inclusions and Non-Euclidean Properties of Proximal Operators

In many instances one may wish to solve an inclusion problem of the form

$$\text{Find } x \in \mathbb{R}^n \text{ such that } \quad \mathbb{0}_n \in (\mathsf{F}+\mathsf{G})(x), \tag{41}$$

where $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is a single-valued continuous monotone mapping but $\mathsf{G} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ is a set-valued mapping. In monotone operator theory on Hilbert spaces, leveraging the fact that $\mathsf{J}_{\alpha\mathsf{G}}$ is single-valued and nonexpansive for every $\alpha > 0$ when G is maximally monotone, algorithms such as the forward-backward splitting and Douglas-Rachford splitting may be used to solve (41) under suitable assumptions on F.

In this section we aim to prove similar results in the non-Euclidean case. We will specialize to the case that G is the subdifferential of a *separable, proper lower semicontinuous (l.s.c.), convex function.* To start we must define the proximal operator of a l.s.c. convex function.

Definition 37 (Proximal operator, Moreau 1962 and Bauschke and Combettes 2017, Def. 12.23) *Let* $g : \mathbb{R}^n \to \left]-\infty, +\infty\right]$ *be a proper l.s.c. convex function. The* proximal operator of g *evaluated at* $x \in \mathbb{R}^n$ *is the map* $\mathrm{prox}_g : \mathbb{R}^n \to \mathbb{R}^n$ *defined by*

$$\mathrm{prox}_g(x) = \operatorname*{arg\,min}_{z \in \mathbb{R}^n} \frac{1}{2}\|x - z\|_2^2 + g(z). \tag{42}$$

Since $g : \mathbb{R}^n \to]-\infty, +\infty]$ is proper, l.s.c., and convex, we can see that for $\alpha > 0$ and fixed $x \in \mathbb{R}^n$, the map $z \mapsto \frac{1}{2}\|x - z\|_2^2 + \alpha g(z)$ is strongly convex and thus has a unique minimizer, so for each $x \in \mathbb{R}^n$, $\mathrm{prox}_{\alpha g}(x)$ is single-valued. Moreover, we have the following connection between proximal operators and resolvents of subdifferentials.

Proposition 38 (Rockafellar 1976 and Bauschke and Combettes 2017, Example 23.3) *Suppose $g : \mathbb{R}^n \to]-\infty, +\infty]$ is proper, l.s.c., and convex. Then for every $\alpha > 0$, $\mathsf{J}_{\alpha \partial g}(x) = \mathrm{prox}_{\alpha g}(x)$.*

In the case of scalar functions, one can exactly capture the set of functions which are proximal operators of some proper l.s.c. convex functions.

Proposition 39 (Bauschke and Combettes 2017, Proposition 24.31) *Let $\phi : \mathbb{R} \to \mathbb{R}$. Then ϕ is the proximal operator of a proper l.s.c. convex function $g : \mathbb{R} \to]-\infty, +\infty]$, i.e., $\phi = \mathrm{prox}_g$ if and only if ϕ satisfies*

$$0 \leq \frac{\phi(x) - \phi(y)}{x - y} \leq 1, \quad \textit{for all } x, y \in \mathbb{R}, x \neq y. \tag{43}$$

A list of examples of scalar functions satisfying (43) and their corresponding proper l.s.c. convex function is provided in (Li et al., 2019, Table 1).

To prove non-Euclidean properties of proximal operators, we will leverage a well-known property, which we highlight in the following proposition.

Proposition 40 (Proximal operator of separable convex functions, Parikh and Boyd 2014, Section 2.1) *For $i \in \{1, \dots, n\}$, let $g_i : \mathbb{R} \to]-\infty, +\infty]$, be proper, l.s.c., and convex. Define $g : \mathbb{R}^n \to]-\infty, +\infty]$ by $g(x) = \sum_{i=1}^n g_i(x_i)$. Then g is proper, l.s.c., and convex and for all $\alpha > 0$,*

$$\mathrm{prox}_{\alpha g}(x) = (\mathrm{prox}_{\alpha g_1}(x_1), \dots, \mathrm{prox}_{\alpha g_n}(x_n)) \in \mathbb{R}^n.$$

If g satisfies $g(x) = \sum_{i=1}^n g_i(x_i)$ with each g_i proper, l.s.c., and convex, we call g separable.

In the following novel proposition, we showcase that when g is separable, $\mathrm{prox}_{\alpha g}$ and $2\mathrm{prox}_{\alpha g} - \mathsf{Id}$ are nonexpansive w.r.t. non-Euclidean norms.

Proposition 41 (Nonexpansiveness of proximal operators of separable convex maps) *For $i \in \{1, \dots, n\}$, let each $g_i : \mathbb{R} \to]-\infty, +\infty]$ be proper, l.s.c., and convex. Define $g : \mathbb{R}^n \to]-\infty, +\infty]$ by $g(x) = \sum_{i=1}^n g_i(x_i)$. For every $\alpha > 0$ and for any $\eta \in \mathbb{R}^n_{>0}$, both $\mathsf{J}_{\alpha \partial g} = \mathrm{prox}_{\alpha g}$ and $\mathsf{R}_{\alpha \partial g} = 2\mathrm{prox}_{\alpha g} - \mathsf{Id}$ are nonexpansive w.r.t. $\|\cdot\|_{\infty, [\eta]^{-1}}$.*[8]

Proof By Proposition 40 we have $\mathrm{prox}_{\alpha g}(x) = (\mathrm{prox}_{\alpha g_1}(x_1), \dots, \mathrm{prox}_{\alpha g_n}(x_n))$. Moreover, each $\mathrm{prox}_{\alpha g_i}$ is nonexpansive and monotone by Proposition 39 and thus satisfies

$$0 \leq (\mathrm{prox}_{\alpha g_i}(x_i) - \mathrm{prox}_{\alpha g_i}(y_i))(x_i - y_i) \leq (x_i - y_i)^2, \qquad \text{for all } x_i, y_i \in \mathbb{R}. \tag{44}$$

8. More generally, $\mathrm{prox}_{\alpha g}$ and $2\mathrm{prox}_{\alpha g} - \mathsf{Id}$ are nonexpansive with respect to any monotonic norm.

We then conclude

$$\begin{aligned}\|\mathrm{prox}_{\alpha g}(x) - \mathrm{prox}_{\alpha g}(y)\|_{\infty,[\eta]^{-1}} &= \max_{i\in\{1,\dots,n\}} \frac{1}{\eta_i}|\mathrm{prox}_{\alpha g_i}(x_i) - \mathrm{prox}_{\alpha g_i}(y_i)| \\ &\leq \max_{i\in\{1,\dots,n\}} \frac{1}{\eta_i}|x_i - y_i| = \|x-y\|_{\infty,[\eta]^{-1}}.\end{aligned}$$

Regarding $\mathsf{R}_{\alpha\partial g}$, we note that (44) implies for all $x_i, y_i \in \mathbb{R}$

$$\begin{aligned}-(x_i-y_i)^2 \leq ((2\mathrm{prox}_{\alpha g_i}(x_i) - x_i) - (2\mathrm{prox}_{\alpha g_i}(y_i) - y_i))(x_i - y_i) \leq (x_i-y_i)^2, \\ \implies |(2\mathrm{prox}_{\alpha g_i}(x_i) - x_i) - (2\mathrm{prox}_{\alpha g_i}(y_i) - y_i)| \leq |x_i - y_i|.\end{aligned}$$

Following the same reasoning as for $\mathrm{prox}_{\alpha g}$, we conclude that $2\mathrm{prox}_{\alpha g} - \mathsf{Id}$ is nonexpansive w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$. ■

We recall from monotone operator theory on Hilbert spaces that if $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ is continuous and $\mathsf{G} : \mathbb{R}^n \to 2^{\mathbb{R}^n}$ satisfies $\mathrm{Dom}(\mathsf{J}_{\alpha\mathsf{G}}) = \mathbb{R}^n$ and $\mathsf{J}_{\alpha\mathsf{G}}(x)$ is single-valued for all $x \in \mathbb{R}^n, \alpha > 0$, then the following equivalences hold: (i) $\mathbb{0}_n \in (\mathsf{F}+\mathsf{G})(x)$, (ii) $x = (\mathsf{J}_{\alpha\mathsf{G}} \circ \mathsf{S}_{\alpha\mathsf{F}})(x)$, and (iii) $z = (\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}})(z)$ and $x = \mathsf{J}_{\alpha\mathsf{G}}(z)$ (Ryu and Boyd, 2016, pp. 25 and 28). In other words, even if G is a set-valued mapping, forward-backward and Peaceman-Rachford splitting methods may be applied to compute zeros of the inclusion problem (41).

When $\mathsf{F} : \mathbb{R}^n \to \mathbb{R}^n$ in (41) is Lipschitz continuous and strongly monotone w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ with monotonicity parameter $c > 0$ and $\mathsf{G} = \partial g$ for a separable proper, l.s.c., convex mapping $g : \mathbb{R}^n \to]-\infty, +\infty]$, by Proposition 41, the composition $\mathrm{prox}_{\alpha g} \circ \mathsf{S}_{\alpha\mathsf{F}}$ is a contraction w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ for small enough $\alpha > 0$. Therefore, the forward-backward splitting method, Algorithm 32, may be applied to find a zero of the splitting problem (41). Analogously, for small enough $\alpha > 0$, $\mathsf{R}_{\alpha\mathsf{F}} \circ \mathsf{R}_{\alpha\mathsf{G}}$ is a contraction w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ and Peaceman-Rachford splitting, Algorithm 35, may be applied to find a zero of the problem (41). In the following section, we present an application of the above theory to recurrent neural networks.

6.2 Iterations for Recurrent Neural Network Equilibrium Computation

Consider the continuous-time recurrent neural network

$$\dot{x} = -x + \Phi(Ax + Bu + b) =: F(x,u), \tag{45}$$

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 a separable activation function, i.e., it acts entry-wise in the sense that $\Phi(x) = (\phi(x_1), \dots, \phi(x_n))^\top$. In this section we consider activation functions $\phi : \mathbb{R} \to \mathbb{R}$ satisfying slope bounds of the form

$$d_1 = \inf_{x,y\in\mathbb{R}, x\neq y} \frac{\phi(x) - \phi(y)}{x-y} \geq 0, \qquad d_2 = \sup_{x,y\in\mathbb{R}, x\neq y} \frac{\phi(x) - \phi(y)}{x-y} \leq 1. \tag{46}$$

Most standard activation functions used in machine learning satisfy these bounds. In (Davydov et al., 2024, Theorem 23), it was shown that a sufficient condition for the strong infinitesimal contractivity of the map $x \mapsto F(x,u)$ is the existence of weights $\eta \in \mathbb{R}^n_{>0}$ such that $\mu_{\infty,[\eta]^{-1}}(A) < 1$; if this condition holds, the recurrent neural network (45) is strongly infinitesimally contracting w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ with rate $1 - \max\{d_1\gamma, d_2\gamma\}$, where 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 $F(x,u)$. Note that equilibrium points x_u^* satisfy $x_u^* = \Phi(Ax_u^* + Bu + b)$ which corresponds to an implicit neural network (INN), which have recently gained significant attention in the machine learning community (Bai et al., 2019; Winston and Kolter, 2020; El Ghaoui et al., 2021). In this regard, computing equilibrium points of (45) corresponds to computing the forward pass of an INN.

Since the map $x \mapsto F(x,u)$ is strongly infinitesimally contracting w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$, the map $x \mapsto -F(x,u)$ is strongly monotone with monotonicity parameter $1 - \max\{d_1\gamma, d_2\gamma\}$ (see Remark 13). As a consequence, applying the forward step method, Algorithm 25, to compute x_u^* yields the iteration

$$x^{k+1} = (1-\alpha)x^k + \alpha\Phi(Ax^k + Bu + b), \tag{47}$$

which is the iteration proposed in (Jafarpour et al., 2021). This iteration is guaranteed to converge for every $\alpha \in]0, \frac{1}{1-\min_{i\in\{1,\dots,n\}}\min\{d_1\cdot(A)_{ii}, d_2\cdot(A)_{ii}\}}]$ with contraction factor $1-\alpha(1-\max\{d_1\gamma, d_2\gamma\})$ by Theorem 26(ii).

However, rather than viewing finding an equilibrium of (45) as finding a zero of a non-Euclidean monotone operator, it is also possible to view it as a monotone inclusion problem of the form (41).

Proposition 42 (Winston and Kolter 2020, Theorem 1) *Suppose ϕ satisfies the bounds (46). Then finding an equilibrium point x_u^* of (45) is equivalent to the (set-valued) operator splitting problem $\mathbb{0}_n \in (\mathsf{F}+\mathsf{G})(x_u^*)$, with*

$$\mathsf{F}(z) = (I_n - A)z - (Bu + b), \qquad \mathsf{G}(z) = \partial g(z), \tag{48}$$

where we denote $g(z) = \sum_{i=1}^n f(z_i)$ and $f : \mathbb{R} \to]-\infty, +\infty]$ is proper, l.s.c., convex, and satisfies $\phi = \mathrm{prox}_f$.

Proof By Proposition 39, since ϕ satisfies the bounds (46), there exists a proper, l.s.c., convex f with $\phi = \mathrm{prox}_f$. The remainder of the proof is equivalent to that in (Winston and Kolter, 2020, Thm 1). ■

While Proposition 42 was leveraged in (Winston and Kolter, 2020) for monotonicity w.r.t. the ℓ_2 norm, we will use it for F which is monotone w.r.t. a diagonally-weighted ℓ_∞ norm.[9]

Checking that F is strongly monotone w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ is straightforward under the assumption that $\gamma < 1$. As a consequence of Propositions, 41 and 42, we can consider different operator splitting algorithms to compute the equilibrium of (45). First, the forward-backward splitting method, Algorithm 32, as applied to this problem is

$$x^{k+1} = \mathrm{prox}_{\alpha g}((1-\alpha)x^k + \alpha(Ax^k + Bu + b)). \tag{49}$$

9. Unless $A = A^\top$, the monotone inclusion problem (48) does not arise from a convex minimization problem.

Since F is Lipschitz continuous, this iteration is guaranteed to converge to the unique fixed point of (45) by Theorem 33(i). 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 minimized at $\alpha^* = \frac{1}{1-\min_i(A)_{ii}}$. Note that compared to the iteration (47), iteration (49) 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 method, Algorithm 35, which can be written

$$\begin{aligned} x^{k+1} &= (I_n + \alpha(I_n - A))^{-1}(z^k + \alpha(Bu+b)), \\ z^{k+1} &= z^k + 2\mathrm{prox}_{\alpha g}(2x^{k+1} - z^k) - 2x^{k+1}. \end{aligned} \tag{50}$$

Since F is Lipschitz continuous and $\mathsf{R}_{\alpha\mathsf{G}}$ is nonexpansive for every $\alpha > 0$, this iteration converges to the unique fixed point of (45) for α in a suitable range by Theorem 36(i). Moreover, the contraction factor is $\frac{1-\alpha(1-\gamma)}{1+\alpha(1-\gamma)}$ for $\alpha \in \,]0, \frac{1}{1-\min_i(A)_{ii}}]$, which comes from the Lipschitz constant of F. In other words, the contraction factor is improved for Peaceman-Rachford compared to forward-backward splitting and the range of allowable step sizes is identical. For RNNs where $(I_n + \alpha(I_n - A))$ may be easily inverted, this splitting method may be preferred.

6.3 Numerical Implementations

To assess the efficacy of the iterations in (47), (49), and (50), we generated A, B, b, u in (45) 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 $A_{ij}, B_{ij}, b_i, u_i \sim \mathcal{N}(0,1)$. To ensure that $A \in \mathbb{R}^{200\times 200}$ satisfies the constraint $\mu_{\infty,[\eta]^{-1}}(A) \leq \gamma$ for some $\eta \in \mathbb{R}^{200}_{>0}$, we pick $[\eta] = I_{200}$ and orthogonally project A onto the convex polytope $\{A \in \mathbb{R}^{200\times 200} \mid \mu_\infty(A) \leq \gamma\}$ using CVXPY (Diamond and Boyd, 2016). In experiments, we consider $\gamma \in \{-1, 0.9\}$ and consider activation functions $\phi(x) = \mathrm{ReLU}(x) = \max\{x, 0\}$ and $\phi(x) = \mathrm{LeakyReLU}(x) = \max\{x, ax\}$ with $a = 0.1$.[10] The proper, l.s.c., convex f corresponding to these activation functions are available in (Li et al., 2019, Table 1).

For all iterations, we initialize x^0 at the origin and for the Peaceman-Rachford iteration, we initialize z^0 at the origin. For each iteration we pick the largest theoretically allowable step size, which in all cases was $\frac{1}{1-\min_i(A)_{ii}}$ (since $\min_{i\in\{1,\dots,n\}}(A)_{ii}$ was negative in all cases). For the case of $\gamma = 0.9$, we found that the largest theoretically allowable step size was $\alpha \approx 0.182$ and for $\gamma = -1$ the largest step size was $\alpha \approx 0.175$. The plots of the residual $\|x_k - \Phi(Ax_k + Bu + b)\|_\infty = \|F(x_k, u)\|_\infty$ versus the number of iterations for all different cases is shown in Figure 2.[11]

We see that, when $\gamma = 0.9$, both forward-step and forward-backward splitting methods for computing the equilibrium of (45) converge at the same rate. This result agrees with the theory since $\gamma > 0$, so that $\max\{d_1\gamma, d_2\gamma\} = \gamma$ for both ReLU and LeakyReLU and the estimated contraction factor for both the forward step method and forward-backward

10. Note that the slope bounds from (43) are $d_1 = 0, d_2 = 1$ for ReLU and $d_1 = a, d_2 = 1$ for LeakyReLU with $a \in [0, 1[$.
11. All iterations and graphics were run and generated in Python. Code to reproduce experiments is available at `https://github.com/davydovalexander/RNN-Equilibrium-NonEucMonotone`.

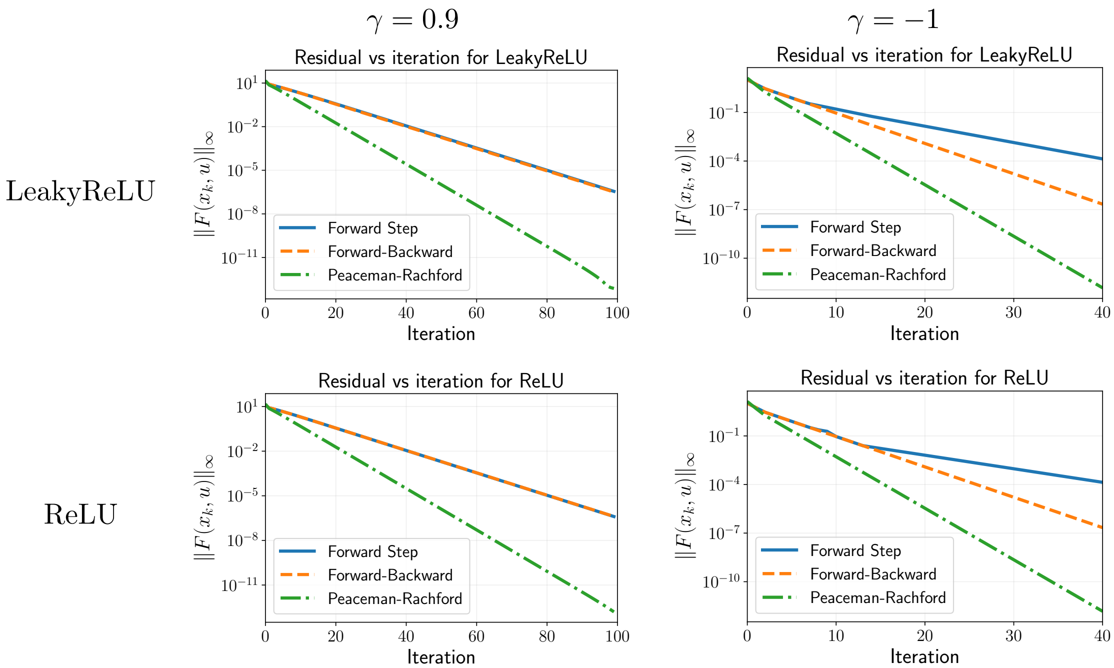

Figure 2: Residual versus number of iterations for forward-step method (47), forward-backward splitting (49), and Peaceman-Rachford splitting (50) for computing the equilibrium of the recurrent neural network (45). The top two plots correspond to $\phi = \text{LeakyReLU}$ with $a = 0.1$ and the bottom two plots correspond to $\phi = \text{ReLU}$. The left two plots correspond to $\gamma = 0.9$ and the right two plots correspond to $\gamma = -1$. Curves for the forward-step method and forward-backward splitting are directly on top of one another in the left two plots. Note the difference in the number of iterations with respect to the parameter γ.

splitting is $1 - \alpha(1 - \gamma) \approx 0.982$. For the Peaceman-Rachford splitting method and $\gamma = 0.9$, the estimated contraction factor is $\frac{1-\alpha(1-\gamma)}{1+\alpha(1-\gamma)} \approx 0.964$, which justifies the improved rate of convergence. When $\gamma = -1$, the forward-backward splitting method converges faster than the forward step method. This result agrees with the theory since the estimated contraction factor for the forward step method is $1 - \alpha(1 - \phi(\gamma)) \approx 0.807$ in the case of LeakyReLU and ≈ 0.825 in the case of ReLU while the estimated contraction factor for forward-backward splitting is $1 - \alpha(1 - \gamma) \approx 0.649$ independent of activation function. On the other hand, for the Peaceman-Rachford splitting method and $\gamma = -1$, the estimated contraction factor is $\frac{1-\alpha(1-\gamma)}{1+\alpha(1-\gamma)} \approx 0.481$, which justifies the improved rate of convergence.

6.4 Tightened Lipschitz Constants for Continuous-Time RNNs

We are interested in studying the robustness of the RNN (45) to input perturbations. In other words, given a nominal input, u, and its corresponding equilibrium output, x_u^*, we aim to upper-bound the deviation of the output due to a change in the input. The Lipschitz constant of a neural network is one common metric used to evaluate its robustness, as discussed in works such as (Fazlyab et al., 2019; Combettes and Pesquet, 2020a; Pauli et al., 2021). In the context of implicit neural networks, Lipschitz constants have been studied in (Revay et al., 2020; Pabbaraju et al., 2021; Jafarpour et al., 2021), with (Pabbaraju et al., 2021) unrolling forward-backward splitting iterations to provide ℓ_2 Lipschitz estimates. In what follows, we generalize the procedure in (Pabbaraju et al., 2021) using techniques from non-Euclidean monotone operator theory to provide novel and tighter ℓ_∞ Lipschitz estimates.

Theorem 43 (Lipschitz estimate of equilibrium points of (45)**)** *Suppose that A satisfies $\mu_{\infty,[\eta]^{-1}}(A) = \gamma < 1$ for some $\eta \in \mathbb{R}^n_{>0}$ and that $\phi = \mathrm{prox}_f$ for some proper, l.s.c., convex $f : \mathbb{R} \to]-\infty, +\infty]$. Define $f_{\mathsf{N}} : \mathbb{R}^m \to \mathbb{R}^n$ by $f_{\mathsf{N}}(u) = x_u^*$ where x_u^* solves the fixed point problem $x_u^* = \Phi(Ax_u^* + Bu + b)$.*[12] *Then for $\eta_{\max} = \max_{i\in\{1,\dots,n\}} \eta_i$, $\eta_{\min} = \min_{i\in\{1,\dots,n\}} \eta_i$, and $\mathrm{Lip}_\infty(f_{\mathsf{N}})$ denoting the minimal ℓ_∞ Lipschitz constant of f_{N},*

$$\mathrm{Lip}_\infty(f_{\mathsf{N}}) \leq \frac{\eta_{\max}}{\eta_{\min}} \frac{\|B\|_\infty}{1-\gamma}. \tag{51}$$

Proof We consider the forward-backward splitting iteration given input u as $x_u^{k+1} = \mathrm{prox}_{\alpha g}((1-\alpha)x_u^k + \alpha(Ax_u^k + Bu + b))$ with initial condition $x_u^0 = \mathbb{0}_n$ which is guaranteed to converge for $\alpha \in]0, \frac{1}{1-\min_i(A)_{ii}}]$ since $\mathrm{prox}_{\alpha g}$ is nonexpansive and $\mathsf{S}_{\alpha\mathsf{F}}$ is a contraction w.r.t. $\|\cdot\|_{\infty,[\eta]^{-1}}$ for every α in this range where F is defined as in (48). We find

$$\begin{aligned}
\|x_u^k - x_v^k\|_{\infty,[\eta]^{-1}} &= \|\mathrm{prox}_{\alpha g}((1-\alpha)x_u^{k-1} + \alpha(Ax_u^{k-1} + Bu + b)) \\
&\qquad - \mathrm{prox}_{\alpha g}((1-\alpha)x_v^{k-1} + \alpha(Ax_v^{k-1} + Bv + b))\|_{\infty,[\eta]^{-1}} \\
&\leq \|\mathsf{S}_{\alpha(\mathsf{Id}-A)}(x_u^{k-1} - x_v^{k-1})\|_{\infty,[\eta]^{-1}} + \alpha\|B(u-v)\|_{\infty,[\eta]^{-1}} \qquad (52) \\
&\leq \mathrm{Lip}(\mathsf{S}_{\alpha(\mathsf{Id}-A)})^k \|x_u^0 - x_v^0\|_{\infty,[\eta]^{-1}} + \alpha\|B(u-v)\|_{\infty,[\eta]^{-1}} \sum_{i=0}^{k-1} \mathrm{Lip}(\mathsf{S}_{\alpha(\mathsf{Id}-A)})^i \\
&= \alpha\|B(u-v)\|_{\infty,[\eta]^{-1}} \sum_{i=0}^{k-1} \mathrm{Lip}(\mathsf{S}_{\alpha(\mathsf{Id}-A)})^i, \qquad (53)
\end{aligned}$$

where (52) holds because of nonexpansiveness of $\mathrm{prox}_{\alpha g}$ and the triangle inequality and (53) is a consequence of $x_u^0 = x_v^0 = \mathbb{0}_n$.

Since the forward-backward splitting iteration converges for every α in the desired range, we can take the limit as $k \to \infty$ and find that $x_u^k \to x_u^*$ and $x_v^k \to x_v^*$ as $k \to \infty$. Then

$$\begin{aligned}
\|x_u^* - x_v^*\|_{\infty,[\eta]^{-1}} &\leq \alpha\|B(u-v)\|_{\infty,[\eta]^{-1}} \sum_{i=0}^{\infty} \mathrm{Lip}(\mathsf{S}_{\alpha(\mathsf{Id}-A)})^i \qquad (54) \\
&= \frac{\alpha\|B(u-v)\|_{\infty,[\eta]^{-1}}}{1 - \mathrm{Lip}(\mathsf{S}_{\alpha(\mathsf{Id}-A)})} \leq \frac{\alpha\|B(u-v)\|_{\infty,[\eta]^{-1}}}{1-(1-\alpha(1-\gamma))} = \frac{\|B(u-v)\|_{\infty,[\eta]^{-1}}}{1-\gamma}, \qquad (55)
\end{aligned}$$

12. Note that if x_u^* solves the fixed point problem $x_u^* = \Phi(Ax_u^* + Bu + b)$, then it is an equilibrium point of the RNN (45).

which implies the result because $\eta_{\max}^{-1}\|z\|_\infty \leq \|z\|_{\infty,[\eta]^{-1}} \leq \eta_{\min}^{-1}\|z\|_\infty$ for every $z \in \mathbb{R}^n$. ■

Remark 44 *In (Jafarpour et al., 2021, Corollary 5), the following Lipschitz estimate is given:*

$$\mathrm{Lip}_\infty(f_{\mathsf{N}}) \leq \frac{\eta_{\max}}{\eta_{\min}} \frac{\|B\|_\infty}{1-\max\{\gamma,0\}}. \tag{56}$$

The Lipschitz estimate in Theorem 43 is always a tighter bound than the estimate (56) *and allows the choice of negative γ to further lower the Lipschitz constant of the RNN. Indeed, one way to make the neural network more robust to uncertainties in its input would be to ensure that γ is a large negative number.*

7. Conclusion

In this paper, we introduce a non-Euclidean version of classical results in monotone operator theory with a focus on mappings that are monotone with respect to diagonally-weighted ℓ_1 or ℓ_∞ norms. Our results show that the resolvent and reflected resolvent maintain many useful properties from the Hilbert case, and we prove that commonly used algorithms for finding zeros of monotone operators and their sums remain effective in the non-Euclidean setting. We applied our theory to the problem of equilibrium computation and Lipschitz constant estimation of recurrent neural networks, yielding novel iterations and tighter upper bounds on Lipschitz constants via forward-backward splitting.

Topics of future research include (i) extending results to more general Banach spaces with a focus on L_1 and L_∞ spaces, (ii) studying the convergence of additional operator splitting methods such as forward-backward-forward (Tseng, 2000) and Davis-Yin (Davis and Yin, 2017) splittings, (iii) extending the theory to variable step size methods, and (iv) considering additional machine learning applications such as ℓ_∞ robustness of deep neural networks as a non-Euclidean analog of (Combettes and Pesquet, 2020a) or reinforcement learning and dynamic programming, where ℓ_∞ contractive and nonexpansive operators are prevalent; see the recent work (Lee and Ryu, 2023) for preliminary ideas in this direction.

Acknowledgments

This material was supported in part by the National Science Foundation Graduate Research Fellowship under grant 2139319 and in part by AFOSR under grant FA9550-22-1-0059.

References

B. Ahn, C. Kim, Y. Hong, and H. J. Kim. Invertible monotone operators for normalizing flows. In *Advances in Neural Information Processing Systems*, 2022. URL `https://openreview.net/forum?id=USoYIT4IQz`.

A. Alacaoglu, A. Böhm, and Y. Malitsky. Beyond the golden ratio for variational inequality algorithms. *Journal of Machine Learning Research*, 24(172):1–33, 2023. URL `http://jmlr.org/papers/v24/22-1488.html`.

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.

S. Bai, J. Z. Kolter, and V. Koltun. Deep equilibrium models. In *Advances in Neural Information Processing Systems*, 2019. URL `https://arxiv.org/abs/1909.01377`.

P. Balamurugan and F. Bach. Stochastic variance reduction methods for saddle-point problems. In *Advances in Neural Information Processing Systems*, page 1416–1424, 2016.

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.

H. H. Bauschke, J. M. Borwein, and P. L. Combettes. Bregman monotone optimization algorithms. *SIAM Journal on Control and Optimization*, 42(2):596–636, 2003. doi:10.1137/S0363012902407120.

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.

D. P. Bertsekas and J. N. Tsitsiklis. *Parallel and Distributed Computation: Numerical Methods*. Athena Scientific, 1997. ISBN 1886529019.

F. E. Browder. Existence and perturbation theorems for nonlinear maximal monotone operators in Banach spaces. *Bulletin of the American Mathematical Society*, 73:322–327, 1967. doi:10.1090/S0002-9904-1967-11734-8.

F. Bullo. *Contraction Theory for Dynamical Systems*. Kindle Direct Publishing, 1.0 edition, 2022. ISBN 979-8836646806. URL `http://motion.me.ucsb.edu/book-ctds`.

F. Bullo. *Contraction Theory for Dynamical Systems*. Kindle Direct Publishing, 1.1 edition, 2023. ISBN 979-8836646806. URL `https://fbullo.github.io/ctds`.

F. Bullo, P. Cisneros-Velarde, A. Davydov, and S. Jafarpour. From contraction theory to fixed point algorithms on Riemannian and non-Euclidean spaces. In *IEEE Conf. on Decision and Control*, December 2021. doi:10.1109/CDC45484.2021.9682883.

X. Cai, C. Song, C. A. Guzmán, and J. Diakonikolas. Stochastic Halpern iteration with variance reduction for stochastic monotone inclusions. In *Advances in Neural Information Processing Systems*, 2022. URL `https://openreview.net/forum?id=BRZos-8TpCf`.

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.

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

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

P. L. Combettes and J.-C. Pesquet. Lipschitz certificates for layered network structures driven by averaged activation operators. *SIAM Journal on Mathematics of Data Science*, 2(2):529–557, 2020a. doi:10.1137/19M1272780.

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

P. L. Combettes and J.-C. Pesquet. Fixed point strategies in data science. *IEEE Transactions on Signal Processing*, 2021. doi:10.1109/TSP.2021.3069677.

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.

G. Dahlquist. *Stability and error bounds in the numerical integration of ordinary differential equations*. PhD thesis, (Reprinted in Trans. Royal Inst. of Technology, No. 130, Stockholm, Sweden, 1959), 1958.

D. Davis and W. Yin. A three-operator splitting scheme and its optimization applications. *Set-Valued and Variational Analysis*, 25:829–858, 2017. doi:10.1007/s11228-017-0421-z.

A. Davydov, S. Jafarpour, and F. Bullo. Non-Euclidean contraction theory for robust nonlinear stability. *IEEE Transactions on Automatic Control*, 67(12):6667–6681, 2022a. doi:10.1109/TAC.2022.3183966.

A. Davydov, S. Jafarpour, A. V. Proskurnikov, and F. Bullo. Non-Euclidean monotone operator theory with applications to recurrent neural networks. In *IEEE Conf. on Decision and Control*, Cancún, México, December 2022b. doi:10.1109/CDC51059.2022.9993197.

A. Davydov, A. V. Proskurnikov, and F. Bullo. Non-Euclidean contraction analysis of continuous-time neural networks. *IEEE Transactions on Automatic Control*, 2024. doi:10.1109/TAC.2024.3422217. To appear.

K. Deimling. *Nonlinear Functional Analysis*. Springer, 1985. ISBN 3-540-13928-1.

J. Diakonikolas. Halpern iteration for near-optimal and parameter-free monotone inclusion and strong solutions to variational inequalities. In *Conference on Learning Theory*, volume 125 of *Proceedings of Machine Learning Research*, pages 1428–1451. PMLR, 2020. URL `https://proceedings.mlr.press/v125/diakonikolas20a.html`.

S. Diamond and S. Boyd. CVXPY: A Python-embedded modeling language for convex optimization. *Journal of Machine Learning Research*, 17(83):1–5, 2016. URL `http://jmlr.org/papers/v17/15-408.html`.

L. El Ghaoui, F. Gu, B. Travacca, A. Askari, and A. Tsai. Implicit deep learning. *SIAM Journal on Mathematics of Data Science*, 3(3):930–958, 2021. doi:10.1137/20M1358517.

M. Fazlyab, A. Robey, H. Hassani, M. Morari, and G. J. Pappas. Efficient and accurate estimation of Lipschitz constants for deep neural networks. In *Advances in Neural Information Processing Systems*, 2019. URL `https://arxiv.org/abs/1906.04893`.

G. Gidel, H. Berard, G. Vignoud, P. Vincent, and S. Lacoste-Julien. A variational inequality perspective on generative adversarial networks. In *International Conference on Learning Representations*, 2019. URL `https://openreview.net/forum?id=r1laEnA5Ym`.

G. Giorgi and S. Komlósi. Dini derivatives in optimization — Part I. *Rivista di Matematica Per Le Scienze Economiche e Sociali*, 15(1):3–30, 1992. doi:10.1007/BF02086523.

I. J. Goodfellow, J. Shlens, and C. Szegedy. Explaining and harnessing adversarial examples. In *International Conference on Learning Representations (ICLR)*, 2015. URL `https://arxiv.org/abs/1412.6572`.

S. Ishikawa. Fixed points and iteration of a nonexpansive mapping in a Banach space. *Proceedings of the American Mathematical Society*, 59(1):65–71, 1976. doi:10.1090/S0002-9939-1976-0412909-X.

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. doi:10.48550/arXiv.2106.03194.

M. I. Jordan, T. Lin, and M. Zampetakis. First-order algorithms for nonlinear generalized Nash equilibrium problems. *Journal of Machine Learning Research*, 24(38):1–46, 2023. URL `https://jmlr.org/papers/v24/22-0310.html`.

J. Lee and E. K. Ryu. Accelerating value iteration with anchoring. In *Advances in Neural Information Processing Systems*, 2023. URL `https://openreview.net/forum?id=Wn82NbmvJy`.

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.

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.

S. M. Lozinskii. Error estimate for numerical integration of ordinary differential equations. I. *Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika*, 5:52–90, 1958. URL `http://mi.mathnet.ru/eng/ivm2980`. (in Russian).

B. Mercier. *Inéquations Variationnelles de la Mécanique.* Publications Mathématiques d'Orsay. Université de Paris-Sud, Département de Mathématique, 1980.

G. J. Minty. Monotone (nonlinear) operators in Hilbert space. *Duke Mathematical Journal*, 29(3):341–346, 1962. doi:10.1215/S0012-7094-62-02933-2.

G. J. Minty. On the monotonicity of the gradient of a convex function. *Pacific Journal of Mathematics*, 14(1):243–247, 1964. doi:10.2140/pjm.1964.14.243.

J.-J. Moreau. Fonctions convexes duales et points proximaux dans un espace hilbertien. *Comptes rendus hebdomadaires des séances de l'Académie des sciences*, 255:2897–2899, 1962.

C. Pabbaraju, E. Winston, and J. Z. Kolter. Estimating Lipschitz constants of monotone deep equilibrium models. In *International Conference on Learning Representations*, 2021. URL `https://openreview.net/forum?id=VcB4QkSfyO`.

N. Parikh and S. Boyd. Proximal algorithms. *Foundations and Trends in Optimization*, 1 (3):127–239, 2014. doi:10.1561/2400000003.

P. Pauli, A. Koch, J. Berberich, P. Kohler, and F. Allgower. Training robust neural networks using Lipschitz bounds. *IEEE Control Systems Letters*, 2021. doi:10.1109/LCSYS.2021.3050444.

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.

T. Pethick, P. Latafat, P. Patrinos, O. Fercoq, and V. Cevher. Escaping limit cycles: Global convergence for constrained nonconvex-nonconcave minimax problems. In *International Conference on Learning Representations*, 2022. URL `https://openreview.net/forum?id=2_vhkAMARk`.

M. Revay, R. Wang, and I. R. Manchester. Lipschitz bounded equilibrium networks. *arXiv preprint arXiv:2010.01732*, 2020. doi:10.48550/arXiv.2010.01732.

R. T. Rockafellar. Characterization of the subdifferentials of convex functions. *Pacific Journal of Mathematics*, 17(3):497–510, 1966. doi:10.2140/pjm.1966.17.497.

R. T. Rockafellar. Monotone operators and the proximal point algorithm. *SIAM Journal on Control and Optimization*, 14(5):877–898, 1976. doi:10.1137/0314056.

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

E. K. Ryu and W. Yin. *Large-Scale Convex Optimization via Monotone Operators*. Cambridge University Press, 2022.

E. K. Ryu, R. Hannah, and W. Yin. Scaled relative graphs: nonexpansive operators via 2D Euclidean geometry. *Mathematical Programming*, 194(1-2):569–619, 2021. doi:10.1007/s10107-021-01639-w.

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

G. Söderlind. Bounds on nonlinear operators in finite-dimensional Banach spaces. *Numerische Mathematik*, 50:27–44, 1986. doi:10.1007/BF01389666.

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

P. Tseng. A modified forward-backward splitting method for maximal monotone mappings. *SIAM Journal on Control and Optimization*, 38(2):431–446, 2000. doi:10.1137/S0363012998338806.

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.

T. Yang, M. I. Jordan, and T. Chavdarova. Solving constrained variational inequalities via a first-order interior point-based method. In *International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=RQY2AXFMRiu.

X. Zhang, W. B. Haskell, and Z. Ye. A unifying framework for variance-reduced algorithms for findings zeroes of monotone operators. *Journal of Machine Learning Research*, 23 (60):1–44, 2022. URL http://jmlr.org/papers/v23/19-513.html.