

Ubiquitous Algorithms in Convex Optimization Generate Self-Contracted Sequences

Axel Böhm*

*Faculty of Mathematics, University of Vienna, 1090 Vienna, Austria
axel.boehm@univie.ac.at*

Aris Daniilidis†

*DIM-CMM, UMI CNRS 2807, Beauchef 851, FCFM, Universidad de Chile
arisd@dim.uchile.cl*

Received: May 10, 2020

Accepted: February 17, 2021

We show that various algorithms, ubiquitous in convex optimization (e.g. proximal-gradient, alternating projections and averaged projections) generate self-contracted sequences $\{x_k\}_{k \in \mathbb{N}}$. As a consequence, a novel universal bound for the *length* $\sum_{k \geq 0} \|x_{k+1} - x_k\|$ can be deduced. In addition, this bound is independent of both the concrete data of the problem (sets, functions) as well as the stepsize involved, and only depends on the dimension of the space.

Keywords: Proximal-gradient algorithm, alternating projection, self-contracted curve.

2010 Mathematics Subject Classification: Primary 52A41, 65K05; secondary 52A05, 90C25.

1. Introduction

The notion of *self-contracted curve* captures, under a simple metric definition (see forthcoming Definition 2.1), characteristic properties of the gradient flow of a convex function, relevant for convergence. The term appeared for the first time in [11], where it was shown that *planar* self-contracted curves are rectifiable. Later on, exploring an old geometrical idea of Manselli-Pucci (see [17]), the previous result has been extended to any finite dimensional Euclidean space. In particular, in [8] (and independently in [16] assuming continuity of the curves) it was shown that the length of any self-contracted curve in $\mathbb{R}^d$ is controlled by a universal constant C_d (depending only on the dimension of the space) times the diameter of the image of the curve.

The aforementioned control of the length directly yields uniform estimates for the asymptotic behaviour of the bounded orbits of quasiconvex gradient systems, convex subgradient systems as well as of the bounded orbits of convex foliations – all of them being typical instances of self-contracted curves. Self-contractedness is indeed strongly related to convexity. It was shown in [12] that under mild assumptions every smooth self-contracted curve can be obtained as an orbit of some smooth convex function.

* Research supported by the doctoral programme VGSCO (Vienna Graduate School on Computational Optimization), FWF (Austrian Science Fund), project W 1260.

† Research supported by the grants: CMM AFB170001, FONDECYT 1211217 (Chile), PGC2018-097960-B-C22 (Spain and EU).

Another important feature of the notion of self-contractedness is that it translates naturally to the discrete case, to include sequences $\{x_k\}_{k \in \mathbb{N}}$ generated by some algorithmic scheme. (The series $\sum_{k \geq 0} \|x_{k+1} - x_k\|$ corresponds to the *length of the sequence*.) A typical example consists of the iterates generated by the proximal-point algorithm applied to a convex function. These iterations, being obtained as successive projections to the convex foliation given by the sublevel sets of the function, generate a self-contracted sequence (cf. [8]). This provides an independent proof of the convergence of the proximal-point algorithm.

The objective of this work is to show that other classical iterative schemes such as the gradient descent algorithm of a smooth convex function with Lipschitz gradient, the alternating projection algorithm for two closed convex sets and the average projection method for finitely many closed convex sets, also generate self-contracted sequences. Consequently, a prior universal estimate for the convergence of all of these methods can be deduced. This estimate neither depends on the specific function nor on the choice of proximal parameters, since all self-contracted sequences/curves lying in the given bounded set admit a universal bound for their length.

Our approach relies strongly on interpreting the aforementioned algorithms as particular instances of the proximal-gradient method (forward-backward algorithm), see Algorithm 3.1 and then establishes that the iterates of the latter give a self-contracted sequence, see Theorem 3.2.

The proof of this central result is surprisingly simple, making astute use of an additional quadratic decay stemming from the strong convexity that appears in the proximal operator. This being said, establishing directly self-contractedness for the alternating projection algorithm is not an easy task, and might be quite involved even in the particular case that one of the convex sets is in fact a convex cone. Indeed, the generated sequence of this algorithm (and in general of all of the aforementioned algorithms) cannot be obtained, in any obvious way, via successive projections to some convex foliation related to our data. The only exception is the fixed-step gradient descent algorithm of a $C^{1,1}$ -convex function since this algorithm is in fact equivalent to the proximal-point algorithm of another convex function. Therefore, overall, this new simple approach provides an additional technique for establishing self-contractedness, albeit still passing through a convex foliation, which still remains the only known way to proceed. In particular, as a by-product, we obtain a new proof for establishing self-contractedness of the proximal-point algorithm (cf. Corollary 3.8).

Let us finally mention, for completeness, that self-contracted curves have also been considered in more general settings, emancipating from direct applications to asymptotic theory of dynamical systems or optimization algorithms. To this end, self-contracted curves have been studied in [9] in Riemann manifolds, where rectifiability has been established via an involved proof that borrows heavily from the underlying Euclidean structure. Remarkably enough, recent works on the topic reveal that Euclidean structure is not a real restriction: generalizing the results of [15], the authors in [19] established that any self-contracted curve in any finite dimensional (potentially asymmetric) normed space is rectifiable. Further extensions include CAT(0) spaces [18] and spaces with weak lower curvature bound [14]. In view of

these developments, it is possible that the notion of self-contracted curve will turn out to be relevant also for abstract dynamics in a metric setting (see [1] e.g.)

2. Preliminaries

Throughout this paper, $\mathbb{R}^d$ will denote the d -dimensional Euclidean space and $\langle \cdot, \cdot \rangle$ its inner product which generates the distance $d(x, y) := \|x - y\|$. For a nonempty subset A we denote its *diameter* by $\text{diam}(A) := \sup\{d(x, y) : x, y \in A\}$.

Definition 2.1. (self-contracted curve) Given a possibly unbounded interval $I \subset \mathbb{R}$, a map $\gamma : I \rightarrow \mathbb{R}^d$ is called *self-contracted*, if for all $t_1, t_2, t_3 \in I$ such that $t_1 \leq t_2 \leq t_3$

$$d(\gamma(t_3), \gamma(t_2)) \leq d(\gamma(t_3), \gamma(t_1)).$$

Note that although originally inspired by continuous curves, this definition does not require any form of continuity or smoothness for the curve γ . In particular, taking γ to be constant on each interval $[n, n + 1)$, for all $n \in \mathbb{N}$, the definition also covers the case of discrete sequences. Formalizing this, we call a sequence $\{x_k\}_{k \in \mathbb{N}}$ in $\mathbb{R}^d$ self-contracted if for all $k_1, k_2, k_3 \in \mathbb{N}$ such that $k_1 \leq k_2 \leq k_3$

$$d(x_{k_3}, x_{k_2}) \leq d(x_{k_3}, x_{k_1}).$$

This seemingly innocent property of self-contractedness has remarkable consequences. It was proven in [8, Theorem 3.3] that every self-contracted curve in a finite dimensional Euclidean space is rectifiable and its length satisfies

$$\ell(\gamma) \leq C_d \text{diam}(\gamma(I)),$$

where C_d denotes a constant only depending on the dimension of the space. Therefore, any bounded self-contracted sequence $\{x_k\}_{k \in \mathbb{N}}$ converges to some x_∞ and

$$\sum_{k=1}^{\infty} d(x_{k+1}, x_k) \leq C_d d(x_0, x_\infty). \quad (1)$$

The aim of this work is to establish the self-contractedness of several classical algorithms in convex optimization. Previous convergence proofs relied on specific Lyapunov functions, in particular, the characteristic property of Fejer monotonicity with respect to the solution set $\mathcal{S}$ (cf. [5, Definition 5.1]), meaning $d_{\mathcal{S}}(x_{k+1}) \leq d_{\mathcal{S}}(x_k)$. Making use of the additional information that the iterates form a self-contracted sequence, we obtain a *data independent* bound given by (1). This bound can be further improved, using Fejer monotonicity, to

$$\sum_{k=1}^{\infty} d(x_{k+1}, x_k) \leq C_d d_{\mathcal{S}}(x_0),$$

whenever $x_\infty \in \mathcal{S}$ (which can always be ensured in the forthcoming algorithms).

3. Proximal-gradient generates self-contracted iterates

Consider the *classical* problem

$$\min_{x \in \mathbb{R}^d} g(x) + f(x) \quad (2)$$

for a proper, convex and lower semicontinuous function $g : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ and a differentiable convex function $f : \mathbb{R}^d \rightarrow \mathbb{R}$ with L -Lipschitz continuous gradient.

We associate with the above system the following *forward-backward* or *proximal-gradient* (cf. [5, Section 27.3]) operator with stepsize $\alpha > 0$:

$$T_\alpha(x) := \text{prox}_{\alpha g}(x - \alpha \nabla f(x)). \quad (3)$$

3.1. Stepsize bounded by the inverse of the Lipschitz constant

The most established method to solve the above problem is described below:

Algorithm 3.1 (proximal-gradient-method) In the above setting, for $x_0 \in \mathbb{R}^d$ and a sequence of stepsizes $\{\alpha_k\}_{k \in \mathbb{N}} \subseteq (0, 1/L]$, consider the following iterative scheme

$$x_{k+1} = T_{\alpha_k}(x_k), \quad \forall k \geq 0.$$

Theorem 3.2. (main result) *The iterates generated by Algorithm 3.1 (proximal-gradient-method with variable stepsize) form a self-contracted sequence.*

For the proof we shall make use of the following three lemmata. The first two are well known and will be quoted without proof. The third lemma is also quite standard for these problems.

Before we proceed, let us first recall that the *subdifferential* $\partial\Phi$ of a convex function $\Phi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ at x is defined as follows

$$\partial\Phi(x) := \{p \in \mathbb{R}^d : \Phi(x) + \langle p, y - x \rangle \leq \Phi(y), \forall y \in \mathbb{R}^d\}.$$

In particular, a point $x^* \in \mathbb{R}^d$ is a minimizer of Φ ($x^* \in \arg \min \Phi$) if and only if $0 \in \partial\Phi(x^*)$. Furthermore, a function Φ is called σ -strongly convex if for every $x, y \in \mathbb{R}^d$ and for every $p \in \partial\Phi(x)$

$$\Phi(y) \geq \Phi(x) + \langle p, y - x \rangle + \frac{\sigma}{2} \|x - y\|^2.$$

The following result is straightforward. It will play an important role in the proof of Lemma 3.5.

Lemma 3.3. (quadratic decay) *Let $\Phi : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ be a σ -strongly convex function and let x^* denote its global minimizer. Then we have*

$$\Phi(x) - \Phi(x^*) \geq \frac{\sigma}{2} \|x - x^*\|^2, \quad \forall x \in \mathbb{R}^d.$$

Whereas the previous statement gives a quadratic lower bound, the next one will give a quadratic upper bound. For a proof we refer to [5, Theorem 18.15] or [7].

Lemma 3.4. (descent lemma) *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a differentiable function with an L -Lipschitz gradient. Then for all $x, y \in \mathbb{R}^d$*

$$f(y) \leq f(x) + \langle \nabla f(x), y - x \rangle + \frac{L}{2} \|y - x\|^2.$$

The following lemma is the core of our main result. It will give an estimation for the decrease of the objective function considered in (2), when applying the proximal-gradient operator T_α defined in (3). For the needs of the next lemma we denote

$$x^+ = T_\alpha(x), \quad \forall x \in \mathbb{R}^d. \quad (4)$$

Lemma 3.5. Fix $x \in \mathbb{R}^d$. If the stepsize $\alpha > 0$ is smaller than the inverse of the Lipschitz constant, i.e. $\alpha \leq 1/L$, then for all $z \in \mathbb{R}^d$

$$(g + f)(x^+) + \frac{1}{2\alpha} \|x^+ - z\|^2 \leq (g + f)(z) + \frac{1}{2\alpha} \|x - z\|^2.$$

Proof. First note that (4) is equivalent to

$$x^+ = \arg \min_{z \in \mathbb{R}^d} \left\{ g(z) + f(x) + \langle \nabla f(x), z - x \rangle + \frac{1}{2\alpha} \|z - x\|^2 \right\}.$$

We define for all $z \in \mathbb{R}^d$

$$l_x(z) := f(x) + \langle \nabla f(x), z - x \rangle$$

and
$$\Phi_x(z) = g(z) + l_x(z) + \frac{1}{2\alpha} \|z - x\|^2.$$

Notice that Φ_x is $\frac{1}{\alpha}$ -strongly convex. Thus, by applying Lemma 3.3, we have that for all $z \in \mathbb{R}^d$

$$\Phi_x(x^+) + \frac{1}{2\alpha} \|x^+ - z\|^2 \leq \Phi_x(z).$$

By the gradient inequality we know that

$$l_x(z) = f(x) + \langle \nabla f(x), z - x \rangle \leq f(z), \quad \forall z \in \mathbb{R}^d.$$

At the same time, by Lemma 3.4 (descent lemma) and the fact that $1/\alpha \geq L$ we have that for all $z \in \mathbb{R}^d$

$$f(x^+) \leq l_x(z) + \frac{1}{2\alpha} \|x^+ - x\|^2 \tag{5}$$

which in return shows the statement of the lemma. $\square$

Proof of Theorem 3.2. Let $\{x_k\}_{k \in \mathbb{N}}$ be the sequence obtained by applying Lemma 3.5 with $x := x_k$, $x^+ := x_{k+1}$ and $\alpha := \alpha_k$ we get that

$$(g + f)(x_{k+1}) + \frac{1}{2\alpha_k} \|x_{k+1} - z\|^2 \leq (g + f)(z) + \frac{1}{2\alpha_k} \|x_k - z\|^2, \quad \forall z \in \mathbb{R}^d.$$

Setting $z = x_k$, the above yields that

$$(g + f)(x_{k+1}) \leq (g + f)(x_k), \quad \forall k \in \mathbb{N}.$$

Moreover, taking any $z \in \mathbb{R}^d$ such that $(g + f)(z) \leq (g + f)(x_{k+1})$ we deduce

$$\|x_{k+1} - z\| \leq \|x_k - z\|.$$

In particular, for all $m > k + 1$, we get

$$\|x_{k+1} - x_m\| \leq \|x_k - x_m\|.$$

Since k is arbitrary in the above inequality, it can be replaced by $k + 1$, yielding

$$\|x_{k+2} - x_m\| \leq \|x_{k+1} - x_m\| \leq \|x_k - x_m\|.$$

Using this iterative argument for $l \in \{k + 1, k + 2, \dots, m\}$, we deduce

$$\|x_l - x_m\| \leq \|x_k - x_m\|.$$

This shows that the sequence is self-contracted, as asserted. $\square$

3.2. Stepsize determined via backtracking

In practice, the Lipschitz constant of the gradient of f is not always known and estimating it might lead to poor stepsizes and thus to slow convergence. In this case it is natural to use some kind of line search procedure to determine an appropriate stepsize. We will describe one (reminiscent of Armijo test) as presented in [6]. The idea is the following: We want to apply Algorithm 3.1 without the restriction $1/\alpha \geq L$ on the stepsize. In every iteration we start with an initial stepsize and decrease it until the statement of Lemma 3.4 (descent lemma) is fulfilled, see below:

Algorithm 3.6 (proximal-gradient with backtracking line search)

For $x_0 \in \mathbb{R}^d$, $\alpha > 0$ and $0 < q < 1$ consider

$$(\forall k \geq 0) \quad \left\{ \begin{array}{l} \text{Set } \alpha_k = \alpha \\ \text{while } f(T_\alpha(x_k)) > f(x_k) + \langle \nabla f(x_k), T_\alpha(x_k) - x_k \rangle + \frac{1}{2\alpha} \|T_\alpha(x_k) - x_k\|^2, \\ \text{do } \quad \alpha_k := q\alpha_k \\ x_{k+1} = T_{\alpha_k}(x_k). \end{array} \right.$$

Note that the parameter $\alpha > 0$ that will be finally chosen in each iteration might be larger than $1/L$. This means that, for the cost of some extra function evaluations, a precise knowledge of the Lipschitz constant of the gradient is no more required; in addition, the algorithm might produce larger steps than what would have been allowed in Algorithm 4.1, yielding a faster convergence.

Theorem 3.7. *The iterates generated by the proximal-gradient-method with backtracking algorithm 3.6 form a self-contracted sequence.*

Proof. The proof follows the same lines as the one of Theorem 3.2, the only difference being in (5) of Lemma 3.5. This equation now holds true because of the way the iterations are chosen ensuring

$$f(x_{k+1}) \leq f(x_k) + \langle \nabla f(x_k), x_{k+1} - x_k \rangle + \frac{1}{2\alpha} \|x_{k+1} - x_k\|^2.$$

(In the proof of Theorem 3.2 the above estimate was provided by Lemma 3.4.) $\square$

3.3. Special cases: proximal-point algorithm, gradient descent

In problem (2) we may consider separately the particular instances $f = 0$ and $g = 0$. In the first case, the problem reduces to the minimization of a lower semi-continuous, convex function g via the proximal-point algorithm. In particular, from Algorithm 3.1 and the previous analysis, we deduce the following result, which first appeared (with a different proof) in [8, Theorem 4.17]

Corollary 3.8. (proximal-point algorithm) *Let $g : \mathbb{R}^d \rightarrow \mathbb{R} \cup \{+\infty\}$ be a convex, lower semicontinuous function and $\{\alpha_k\}_{k \in \mathbb{N}} \subseteq (0, +\infty)$. Then, for any $x_0 \in \mathbb{R}^d$ the proximal sequence*

$$x_{k+1} = \text{prox}_{\alpha_k g}(x_k), \quad \forall k \geq 0,$$

is a self-contracted curve.

If $g = 0$, the problem reduces to minimizing a smooth convex function with Lipschitz gradient via steepest descent. In particular, we obtain the following result, which is new. (While preparing the manuscript, the recent interesting preprint [13] came to our attention. The forthcoming result also appears there with a different proof (see [13, Lemma 3.1]).

Corollary 3.9. (steepest descent) *Let $f : \mathbb{R}^d \rightarrow \mathbb{R}$ be a smooth convex function with L -Lipschitz gradient and $\{\alpha_k\}_{k \in \mathbb{N}}$ either be bounded from above by $1/L$ or produced by backtracking line search. Then, for any $x_0 \in \mathbb{R}^d$, the sequence $\{x_k\}_{k \in \mathbb{N}}$ defined by*

$$x_{k+1} = x_k - \alpha_k \nabla f(x_k), \quad \forall k \geq 0,$$

is self-contracted.

3.4. A priori estimates for convergence

An important consequence of Theorem 3.2, Theorem 3.7 and Corollaries 3.8–3.9, is the following. If the optimization problem (2) has a solution, then the iterates form a bounded, self-contracted sequence. Therefore, by [8, Theorem 3.3] we deduce that the sequence of iterates $\{x_k\}_{k \in \mathbb{N}}$ has finite length, i.e.

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\| < +\infty,$$

and thus it converges to some point $x_{\infty} \in \mathbb{R}^d$. In addition, provided that the stepsize $\{\alpha_k\}_{k \in \mathbb{N}}$ does not go to zero too fast (for instance if it is not in $\ell^1(\mathbb{N})$), we deduce that x_{∞} will be a minimizer of our objective function.

Moreover, the bound on the length of the sequence of the iterates

$$\sum_{k=0}^{\infty} \|x_{k+1} - x_k\| < C_d d_{\mathcal{S}}(x_0), \quad \mathcal{S} = \arg \min (f + g)$$

does not depend on the data f, g but only on the dimension and the distance of the initial point to the set of minimizers.

4. Projection algorithms

In this section we investigate the property of self-contractedness for various projection type algorithms. Although a direct approach for establishing this property would be quite involved, it turns out that the overall analysis simplifies significantly by utilizing the main result of the previous section.

Let us introduce our main problem of finding the intersection of two closed convex sets $A, B \subset \mathbb{R}^d$:

$$\text{Find } x \in A \cap B. \tag{6}$$

4.1. Alternating projections

The arguably best known algorithm for solving this problem is the following:

Algorithm 4.1 (alternating projections [3, p. 186])

For $x_0 \in \mathbb{R}^d$, consider the iterative scheme

$$(\forall k \geq 0) \quad \begin{cases} y_{k+1} = P_A(x_k) \\ x_{k+1} = P_B(y_{k+1}). \end{cases}$$

We are going to prove that both sequences above are self-contracted. This will follow from the self-contractedness of the iterates of the proximal-gradient-method.

Theorem 4.2. *Let $\{x_k\}_{k \in \mathbb{N}}$ and $\{y_k\}_{k \in \mathbb{N}}$ be the two sequences of iterates generated by Algorithm 4.1. Then, both sequences are self-contracted.*

Proof. We interpret alternating projections as a proximal gradient scheme. Define $f := \frac{1}{2}d_A^2$ and note that a gradient step with respect to this function corresponds to a projection onto A , i.e. $\text{Id} - \nabla f = P_A$. Furthermore, we define $g := \delta_B$ as the indicator function of the set B . Thus,

$$x_{k+1} = P_B(P_A(x_k)) = P_B(x_k - \nabla f(x_k)) = \text{prox}_g(x_k - \nabla f(x_k)).$$

Therefore, Theorem 3.2 shows that $\{x_k\}_{k \in \mathbb{N}}$ is self-contracted, whereas self-contractedness of $\{y_k\}_{k \in \mathbb{N}}$ follows from symmetry. $\square$

4.2. Averaged projections

Consider the more general problem of finding the intersection of a finite number of closed convex sets $\{C_i\}_{i=1}^n$:

$$\text{Find } x \in \bigcap_{i=1}^n C_i. \quad (7)$$

One could clearly extend the method of alternating projections to this setting and end up with the cyclic projection method, which is given by

$$x_{k+1} = P_n \circ P_{n-1} \circ \dots \circ P_1(x_k), \quad \forall k \geq 0.$$

In practice, this method is often replaced by other schemes (see for instance [2] and references therein). We will focus on the following modification (cf. [4, p. 368]).

Algorithm 4.3 (averaged projections)

For $x_0 \in \mathbb{R}^d$ consider the iterative scheme

$$x_{k+1} = \frac{1}{n} \sum_{i=1}^n P_{C_i}(x_k), \quad \forall k \geq 0. \quad (8)$$

Proposition 4.4. *The iterates generated by Algorithm 4.3 form a self-contracted sequence.*

We will give two different proofs for the assertion above. The first one relies on reformulating the method of averaged projections as gradient descent (with fixed stepsize), whereas the second one is based on the interpretation of this method as alternating projections over two closed convex sets in a product space.

Proof. (*first proof*) Let us define for $i \in \{1, 2, \dots, n\}$

$$f_i := \frac{1}{2}d_{C_i}^2$$

and notice that $f := \sum_{i=1}^n f_i$ is convex and smooth with L -Lipschitz gradient, where $L \leq n$. Then, the sequence of (8) can be equivalently defined by

$$x_{k+1} = \left(\text{Id} - \frac{1}{n} \nabla \left(\sum_{i=1}^n f_i \right) \right) (x_k),$$

with stepsize $\alpha_k = \alpha = 1/n \leq 1/L$. The result follows by applying Corollary 3.9.

(*second proof*) Define the closed convex sets $\widehat{C} = \Pi_{i=1}^n C_i$ and

$$\Delta = \{(y^1, y^2, \dots, y^n) \in \mathbb{R}^{d \times n} : y^1 = y^2 = \dots = y^n\}.$$

Then, (8) is equivalent to $x_{k+1} = P_{\Delta}(P_{\widehat{C}}(x_k))$, i.e., the alternating projections method applied to the sets $\widehat{C}$ and Δ in $\mathbb{R}^{d \times n}$. Now we can apply Theorem 4.2, and deduce the fact that $\{x_k\}_{k \in \mathbb{N}}$ is self-contracted. $\square$

To summarize, Algorithm 4.1 applied to problem (6) or Algorithm 4.3 applied to problem (7) share the following common feature: whenever the problem is feasible or at least one of the involved sets is bounded, then the sequence of iterates is bounded. In both cases, the sequence $\{x_k\}_{k \in \mathbb{N}}$ is convergent and the estimates mentioned in Section 2 hold true.

General conclusion. The results of this work confirm the previous understanding [8, 11, 12, 16] that self-contractedness relates to convexity, in both, the continuous and discrete setting. Let us mention that this concept has recently been relaxed in [10] to so-called λ -curves (respectively λ -sequences). In that work, estimates, similar to the ones of self-contracted curves, have been obtained. However, the general asymptotic behavior of such curves remains unclear. At the same time this notion might be related to more complex settings such as accelerated convex methods or algorithms in nonconvex optimization.

Acknowledgments. A major part of this work was done during a research visit of the first author to the University of Chile (March to June 2019) and of the second author to the University of Vienna (March 2020). The authors wish to thank their hosts for hospitality.

References

- [1] L. Ambrosio, N. Gigli, G. Savare: *Gradient Flows in Metric Spaces and in the Space of Probability Measures*, Lectures in Mathematics, ETH Zürich, Birkhäuser, Basel (2008).
- [2] J.-B. Baillon, P. L. Combettes, R. Cominetti: *There is no variational characterization of the cycles in the method of periodic projections*, J. Funct. Analysis 262 (2012) 400–408.

- [3] H. Bauschke, J. Borwein: *On the convergence of von Neumann's alternating projection algorithm for two sets*, Set-Valued Analysis 1 (1993) 185–212.
- [4] H. Bauschke, J. Borwein: *On projection algorithms for solving convex feasibility problems*, SIAM Review 38 (1996) 367–426.
- [5] H. Bauschke, P. Combettes: *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, CMS Books in Mathematics, Springer, Cham (2011).
- [6] A. Beck, M. Teboulle: *A fast iterative shrinkage-thresholding algorithm for linear inverse problems*, SIAM J. Imaging Sci. 2 (2009) 183–202.
- [7] D. Bertsekas: *Nonlinear Programming*, 2nd ed., Athena Scientific, Belmont (1999).
- [8] A. Daniilidis, G. David, E. Durand-Cartagena, A. Lemenant: *Rectifiability of self-contracted curves in the Euclidean space and applications*, J. Geom. Analysis 25 (2015) 1211–1239.
- [9] A. Daniilidis, R. Deville, E. Durand-Cartagena, L. Rifford: *Self-contracted curves in Riemannian manifolds*, J. Math. Analysis Appl. 457 (2018) 1333–1352.
- [10] A. Daniilidis, R. Deville, E. Durand-Cartagena: *Metric and geometric relaxations of self-contracted curves*, J. Optim. Theory Appl. 182 (2019) 81–109.
- [11] A. Daniilidis, O. Ley, S. Sabourau: *Asymptotic behaviour of self-contracted planar curves and gradient orbits of convex functions*, J. Math. Pures Appl. 94 (2010) 183–199.
- [12] E. Durand-Cartagena, A. Lemenant: *Self-contracted curves are gradient flows of convex functions*, Proc. Amer. Math. Soc. 147 (2019) 2517–2531.
- [13] C. Gupta, S. Balakrishnan, A. Ramdas: *Path length bounds for gradient descent and flow*, J. Mach. Learning Res. 22 (2021) art. no. 68.
- [14] N. Lebedeva, S. Ohta, V. Zolotov: *Self-contracted curves in spaces with weak lower curvature bound*, Int. Math. Res. Notices, 11 (2021) 8623–8656.
- [15] A. Lemenant: *Rectifiability of non-Euclidean planar self-contracted curves*, Confluentes Math. 8 (2016) 23–38.
- [16] M. Longinetti, P. Manselli, A. Venturi: *On steepest descent curves for quasi convex families in $\mathbb{R}^n$* , Math. Nachrichten 288 (2015) 420–442.
- [17] P. Manselli, C. Pucci: *Maximum length of steepest descent curves for quasi-convex functions*, Geom. Dedicata 38 (1991) 211–227.
- [18] S. Ohta: *Self-contracted curves in CAT(0)-spaces and their rectifiability*, J. Geom. Analysis 3 (2020) 936–967.
- [19] E. Stepanov, Y. Teplitskaya: *Self-contracted curves have finite length*, J. London Math. Soc. 96 (2017) 455–481.