



HAL
open science

Proximal point algorithm, Douglas–Rachford algorithm and alternating projections: a case study

Heinz H Bauschke, Minh Dao, Dominikus Noll, Hung M Phan

► **To cite this version:**

Heinz H Bauschke, Minh Dao, Dominikus Noll, Hung M Phan. Proximal point algorithm, Douglas–Rachford algorithm and alternating projections: a case study. *Journal of Convex Analysis*, 2016, 23 (1), pp.237-261. hal-01868425

HAL Id: hal-01868425

<https://hal.science/hal-01868425>

Submitted on 5 Sep 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Proximal point algorithm, Douglas–Rachford algorithm and alternating projections: a case study

Heinz H. Bauschke*, Minh N. Dao[†], Dominikus Noll[‡] and Hung M. Phan[§]

March 23, 2015

Abstract

Many iterative methods for solving optimization or feasibility problems have been invented, and often convergence of the iterates to some solution is proven. Under favourable conditions, one might have additional bounds on the distance of the iterate to the solution leading thus to *worst case estimates*, i.e., how fast the algorithm must converge.

Exact convergence estimates are typically hard to come by. In this paper, we consider the complementary problem of finding *best case estimates*, i.e., how slow the algorithm has to converge, and we also study *exact asymptotic rates of convergence*. Our investigation focuses on convex feasibility in the Euclidean plane, where one set is the real axis while the other is the epigraph of a convex function. This case study allows us to obtain various convergence rate results. We focus on the popular method of alternating projections and the Douglas–Rachford algorithm. These methods are connected to the proximal point algorithm which is also discussed. Our findings suggest that the Douglas–Rachford algorithm outperforms the method of alternating projections in the absence of constraint qualifications. Various examples illustrate the theory.

2010 Mathematics Subject Classification: Primary 65K05; Secondary 65K10, 90C25.

Keywords: alternating projections, convex feasibility problem, convex set, Douglas–Rachford algorithm, projection, proximal mapping, proximal point algorithm, proximity operator.

*Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: heinz.bauschke@ubc.ca.

[†]Department of Mathematics and Informatics, Hanoi National University of Education, 136 Xuan Thuy, Hanoi, Vietnam, and Mathematics, University of British Columbia, Kelowna, B.C. V1V 1V7, Canada. E-mail: minhndn@hnue.edu.vn.

[‡]Institut de Mathématiques, Université de Toulouse, 118 route de Narbonne, 31062 Toulouse, France. E-mail: noll@mip.ups-tlse.fr.

[§]Department of Mathematical Sciences, University of Massachusetts Lowell, 265 Riverside St., Olney Hall 428, Lowell, MA 01854, USA. E-mail: hung.phan@uml.edu.

1 Introduction

Three algorithms

Let X be a Euclidean space, with inner product $\langle \cdot, \cdot \rangle$ and induced norm $\| \cdot \|$, and let $f: X \rightarrow]-\infty, +\infty]$ be convex, lower semicontinuous, and proper. A classical method for finding a minimizer of f is the *proximal point algorithm (PPA)*. It requires using the *proximal point mapping* (or proximity operator) which was pioneered by Moreau [13]:

Fact 1.1 (proximal mapping) *For every $x \in X$, there exists a unique point $p = P_f(x) \in X$ such that $\min_{y \in X} f(y) + \frac{1}{2}\|x - y\|^2 = f(p) + \frac{1}{2}\|x - p\|^2$. The induced operator $P_f: X \rightarrow X$ is firmly nonexpansive¹, i.e., $(\forall x \in X)(\forall y \in X) \|P_f(x) - P_f(y)\|^2 + \|(\text{Id} - P_f)x - (\text{Id} - P_f)y\|^2 \leq \|x - y\|^2$.*

The proximal point algorithm was proposed by Martinet [12] and further studied by Rockafellar [16]. Nowadays numerous extensions exist; however, here we focus only on the most basic instance of PPA:

Fact 1.2 (proximal point algorithm (PPA)) *Let $f: X \rightarrow]-\infty, +\infty]$ be convex, lower semicontinuous, and proper. Suppose that Z , the set of minimizers of f , is nonempty, and let $x_0 \in X$. Then the sequence generated by*

$$(1) \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = P_f(x_n)$$

converges to a point in Z and it satisfies

$$(2) \quad (\forall z \in Z)(\forall n \in \mathbb{N}) \quad \|x_{n+1} - z\|^2 + \|x_n - x_{n+1}\|^2 \leq \|x_n - z\|^2.$$

An ostensibly quite different type of optimization problem is, for two given closed convex nonempty subsets A and B of X , to find a point in $A \cap B \neq \emptyset$. Let us present two fundamental algorithms for solving this convex feasibility problem. The first method was proposed by Bregman [8].

Fact 1.3 (method of alternating projections (MAP)) *Let $a_0 \in A$ and set*

$$(3) \quad (\forall n \in \mathbb{N}) \quad a_{n+1} = P_A P_B(a_n).$$

Then $(a_n)_{n \in \mathbb{N}}$ converges to a point $a_\infty \in C = A \cap B$. Moreover,

$$(4) \quad (\forall c \in C)(\forall n \in \mathbb{N}) \quad \|a_{n+1} - c\|^2 + \|a_{n+1} - P_B a_n\|^2 + \|P_B a_n - a_n\|^2 \leq \|a_n - c\|^2.$$

The second method is the celebrated Douglas–Rachford algorithm. The next result can be deduced by combining [11] and [4].

¹ Note that if $f = \iota_C$ is the *indicator function* of a nonempty closed convex subset of X , then $P_f = P_C$, where the P_C is the nearest point mapping or *projector* of C ; the corresponding *reflector* is $R_C = 2P_C - \text{Id}$.

Fact 1.4 (Douglas–Rachford algorithm (DRA)) Set $T = \text{Id} - P_A + P_B R_A$, let $z_0 \in X$, and set

$$(5) \quad (\forall n \in \mathbb{N}) \quad a_n = P_A z_n \text{ and } z_{n+1} = T z_n.$$

Then² $(z_n)_{n \in \mathbb{N}}$ converges to some point $z_\infty \in \text{Fix } T = (A \cap B) + N_{A-B}(0)$, and $(a_n)_{n \in \mathbb{N}}$ converges to $P_A z_\infty \in A \cap B$.

Again, there are numerous refinements and adaptations of MAP and DRA; however, it is here not our goal to survey the most general results possible³ but rather to focus on the speed of convergence. We will make this precise in the next subsection.

Goal and contributions

Most rate-of-convergence results for PPA, MAP, and DRA take the following form: *If some additional condition is satisfied, then the convergence of the sequence is at least as good as some form of “fast” convergence (linear, superlinear, quadratic etc.).* This can be interpreted as a *worst case analysis*. In the generality considered here⁴, we are not aware of results that approach this problem from the other side, i.e., that address the question: *Under which conditions is the convergence no better than some form of “slow” convergence?* This concerns the *best case analysis*. Ideally, one would like an exact asymptotic rate of convergence in the sense of (14) below.

While we do not completely answer these questions, we do set out to tackle them by providing a *case study* when $X = \mathbb{R}^2$ is the Euclidean plane, the set $A = \mathbb{R} \times \{0\}$ is the real axis, and the set B is the epigraph of a proper lower semicontinuous convex function f . We will see that in this case MAP and DRA have connections to the PPA applied to f . We focus in particular on the case not covered by conditions guaranteeing linear convergence of MAP or DRA⁵. We originally expected the behaviour of MAP and DRA in cases of “bad geometry” to be similar⁶. It came to us as surprise that this appears not to be the case. In fact, the examples we provide below suggest that DRA performs significantly better than MAP. Concretely, suppose that B is the epigraph of the function $f(x) = (1/p)|x|^p$, where $1 < p < +\infty$. Since $A = \mathbb{R} \times \{0\}$, we have that $A \cap B = \{(0,0)\}$ and since $f'(0) = 0$, the “angle” between A and B at the intersection is 0. As expected MAP converges sublinearly (even logarithmically) to 0. However, DRA converges faster in all cases: superlinearly (when $1 < p < 2$), linearly (when $p = 2$) or logarithmically (when $2 < p < +\infty$). This example is deduced by general results we obtain on *exact* rates of convergence for PPA, MAP and DRA.

²Here $\text{Fix } T = \{x \in X \mid x = Tx\}$ is the set of fixed points of T , and $N_{A-B}(0)$ stands for the normal cone of the set $A - B = \{a - b \mid a \in A, b \in B\}$ at 0.

³See, e.g., [3] for various more general variants of PPA, MAP, and DRA.

⁴Some results are known for MAP when the sets are linear subspaces; however, the slow (sublinear) convergence can only be observed in *infinite-dimensional Hilbert space*; see [9] and references therein.

⁵Indeed, the most common sufficient condition for linear convergence in either case is $\text{ri}(A) \cap \text{ri}(B) \neq \emptyset$; see [5, Theorem 3.21] for MAP and [14] or [6, Theorem 8.5(i)] for DRA.

⁶This expectation was founded in the similar behaviour of MAP and DRA for two subspaces; see [2].

Organization

The paper is organized as follows. In Section 2, we provide various auxiliary results on the convergence of real sequences. These will make the subsequent analysis of PPA, MAP, and DRA more structured. Section 3 focuses on the PPA. After reviewing results on finite, super-linear, and linear convergence, we exhibit a case where the asymptotic rate is only logarithmic. We then turn to MAP in Section 4 and provide results on the asymptotic convergence. We also draw the connection between MAP and PPA and point out that a result of Güler is sharp. In Section 5, we deal with DRA, draw again a connection to PPA and present asymptotic convergence. The notation we employ is fairly standard and follows, e.g., [15] and [3].

2 Auxiliary results

In this section we collect various results that facilitate the subsequent analysis of PPA, MAP and DRA. We begin with the following useful result which appears to be part of the folklore⁷.

Fact 2.1 (generalized Stolz–Cesàro theorem) *Let $(a_n)_{n \in \mathbb{N}}$ and $(b_n)_{n \in \mathbb{N}}$ be sequences in \mathbb{R} such that $(b_n)_{n \in \mathbb{N}}$ is unbounded and either strictly monotone increasing or strictly monotone decreasing. Then*

$$(6) \quad \liminf_{n \rightarrow \infty} \frac{a_{n+1} - a_n}{b_{n+1} - b_n} \leq \liminf_{n \rightarrow \infty} \frac{a_n}{b_n} \leq \overline{\lim}_{n \rightarrow \infty} \frac{a_n}{b_n} \leq \overline{\lim}_{n \rightarrow \infty} \frac{a_{n+1} - a_n}{b_{n+1} - b_n},$$

where the limits may lie in $[-\infty, +\infty]$.

Setting $(b_n)_{n \in \mathbb{N}} = (n)_{n \in \mathbb{N}}$ in Fact 2.1, we obtain the following:

Corollary 2.2 *The following inequalities hold for an arbitrary sequence $(x_n)_{n \in \mathbb{N}}$ in \mathbb{R} :*

$$(7) \quad \liminf_{n \rightarrow \infty} (x_{n+1} - x_n) \leq \liminf_{n \rightarrow \infty} \frac{x_n}{n} \leq \overline{\lim}_{n \rightarrow \infty} \frac{x_n}{n} \leq \overline{\lim}_{n \rightarrow \infty} (x_{n+1} - x_n).$$

Set $\mathbb{R}_{++} = \{x \in \mathbb{R} \mid x > 0\}$. For the remainder of this section, we assume that

$$(8) \quad g: \mathbb{R}_{++} \rightarrow \mathbb{R}_{++} \text{ is increasing and } H \text{ is an antiderivative of } -1/g.$$

Example 2.3 (x^q) Let $g(x) = x^q$ on \mathbb{R}_{++} , where $1 \leq q < \infty$. If $q > 1$, then $-1/g(x) = -x^{-q}$ and we can choose $H(x) = x^{1-q}/(q-1)$ which has the inverse $H^{-1}(x) = 1/((q-1)x)^{1/(q-1)}$. If $q = 1$, then we can choose $H(x) = -\ln(x)$ which has the inverse $H^{-1}(x) = \exp(-x)$.

Proposition 2.4 *Let $(\beta_n)_{n \in \mathbb{N}}$ and $(\delta_n)_{n \in \mathbb{N}}$ be sequences in \mathbb{R}_{++} , and suppose that*

$$(9) \quad (\forall n \in \mathbb{N}) \quad \beta_{n+1} = \beta_n - \delta_n g(\beta_n).$$

Then the following hold:

⁷ Since we were able to locate only an online reference, we include a proof in Appendix A.

$$(i) (\forall n \in \mathbb{N}) \quad \delta_n \leq H(\beta_{n+1}) - H(\beta_n) \leq \delta_{n+1} \frac{\beta_n - \beta_{n+1}}{\beta_{n+1} - \beta_{n+2}} = \delta_n \frac{g(\beta_n)}{g(\beta_{n+1})}.$$

$$(ii) \quad \underline{\lim}_{n \rightarrow \infty} \delta_n \leq \underline{\lim}_{n \rightarrow \infty} \frac{H(\beta_n)}{n} \leq \overline{\lim}_{n \rightarrow \infty} \frac{H(\beta_n)}{n} \leq \overline{\lim}_{n \rightarrow \infty} \delta_n \frac{g(\beta_n)}{g(\beta_{n+1})}.$$

(iii) If $(\delta_n)_{n \in \mathbb{N}}$ is convergent, say $\delta_n \rightarrow \delta_\infty$, and $\frac{g(\beta_n)}{g(\beta_{n+1})} \rightarrow 1$, then $\frac{H(\beta_n)}{n} \rightarrow \delta_\infty$.

Proof. For every $n \in \mathbb{N}$, we have

$$(10a) \quad \delta_n = \frac{\beta_n - \beta_{n+1}}{g(\beta_n)} \leq \int_{\beta_{n+1}}^{\beta_n} \frac{dx}{g(x)} = H(\beta_{n+1}) - H(\beta_n)$$

$$(10b) \quad \leq \frac{\beta_n - \beta_{n+1}}{g(\beta_{n+1})} = \delta_{n+1} \frac{\beta_n - \beta_{n+1}}{\beta_{n+1} - \beta_{n+2}} = \delta_n \frac{g(\beta_n)}{g(\beta_{n+1})}.$$

Hence (i) holds. Combining with (7), we obtain (ii). Finally, (iii) follows from (ii). ■

Corollary 2.5 Let $(x_n)_{n \in \mathbb{N}}$ and $(\delta_n)_{n \in \mathbb{N}}$ be sequences in \mathbb{R}_{++} such that

$$(11) \quad (\forall n \in \mathbb{N}) \quad x_n = x_{n+1} + \delta_n g(x_{n+1}).$$

Then the following hold:

$$(i) (\forall n \in \mathbb{N}) \quad \delta_n \frac{g(x_{n+1})}{g(x_n)} \leq H(x_{n+1}) - H(x_n) \leq \delta_n.$$

$$(ii) \quad \underline{\lim}_{n \rightarrow \infty} \delta_n \frac{g(x_{n+1})}{g(x_n)} \leq \underline{\lim}_{n \rightarrow \infty} \frac{H(x_n)}{n} \leq \overline{\lim}_{n \rightarrow \infty} \frac{H(x_n)}{n} \leq \overline{\lim}_{n \rightarrow \infty} \delta_n.$$

(iii) If $(\delta_n)_{n \in \mathbb{N}}$ is convergent, say $\delta_n \rightarrow \delta_\infty$, and $\frac{g(x_{n+1})}{g(x_n)} \rightarrow 1$, then $\frac{H(x_n)}{n} \rightarrow \delta_\infty$.

Proof. Indeed, set $(\forall n \in \mathbb{N}) \quad \varepsilon_n = \delta_n \frac{g(x_{n+1})}{g(x_n)}$ and rewrite the update

$$(12) \quad x_{n+1} = x_n - \delta_n \frac{g(x_{n+1})}{g(x_n)} g(x_n) = x_n - \varepsilon_n g(x_n).$$

Now apply Proposition 2.4. ■

Definition 2.6 (types of convergence) Let $(\alpha_n)_{n \in \mathbb{N}}$ be a sequence in \mathbb{R}_{++} such that $\alpha_n \rightarrow 0$, and suppose there exists $1 \leq q < +\infty$ such that

$$(13) \quad \frac{\alpha_{n+1}}{\alpha_n^q} \rightarrow c \in \mathbb{R}_+.$$

Then the convergence of $(\alpha_n)_{n \in \mathbb{N}}$ to 0 is:

(i) with order q if $q > 1$ and $c > 0$;

- (ii) superlinear if $q = 1$ and $c = 0$;
- (iii) linear if $q = 1$ and $0 < c < 1$;
- (iv) sublinear if $q = 1$ and $c = 1$;
- (v) logarithmic if it is sublinear and $|\alpha_{n+1} - \alpha_{n+2}|/|\alpha_n - \alpha_{n+1}| \rightarrow 1$.

If $(\beta_n)_{n \in \mathbb{N}}$ is also a sequence in \mathbb{R}_{++} , it is convenient to define

$$(14) \quad \alpha_n \sim \beta_n \Leftrightarrow \lim_{n \rightarrow \infty} \frac{\alpha_n}{\beta_n} \in \mathbb{R}_{++}.$$

The following example exhibits a case where we obtain a simple exact asymptotic rate of convergence.

Example 2.7 Let $(x_n)_{n \in \mathbb{N}}$ and $(\delta_n)_{n \in \mathbb{N}}$ be sequences in \mathbb{R}_{++} , and let $1 < q < \infty$. Suppose that

$$(15) \quad \delta_n \rightarrow \delta_\infty \in \mathbb{R}_{++}, \quad \frac{x_n}{x_{n+1}} \rightarrow 1, \quad \text{and} \quad (\forall n \in \mathbb{N}) \quad x_n = x_{n+1} + \delta_n x_{n+1}^q.$$

Then $x_n \rightarrow 0$ logarithmically,

$$(16) \quad \frac{x_n}{\left(\frac{1}{n}\right)^{1/(q-1)}} \rightarrow \frac{1}{((q-1)\delta_\infty)^{1/(q-1)}} \quad \text{and} \quad x_n \sim \left(\frac{1}{n}\right)^{1/(q-1)}.$$

Proof. Suppose that $g(x) = x^q$ and note that $g(x_{n+1})/g(x_n) = (x_{n+1}/x_n)^q \rightarrow 1^q = 1$. This implies that $x_n \rightarrow 0$ logarithmically. Finally, (16) follows from Example 2.3, Corollary 2.5, and (14). ■

We conclude this section with some one-sided versions which are useful for obtaining information about how fast or slow a sequence must converge.

Corollary 2.8 Let $(\beta_n)_{n \in \mathbb{N}}$ and $(\rho_n)_{n \in \mathbb{N}}$ be sequences in \mathbb{R}_{++} , and suppose that

$$(17) \quad (\forall n \in \mathbb{N}) \quad \beta_{n+1} \leq \beta_n - \rho_n g(\beta_n) \quad \text{and} \quad \underline{\rho} = \liminf_{n \rightarrow \infty} \rho_n \in \mathbb{R}_{++}.$$

Then

$$(18) \quad (\forall \varepsilon \in]0, \underline{\rho}[) (\exists m \in \mathbb{N}) (\forall n \geq m) \quad \beta_n \leq H^{-1}(n(\underline{\rho} - \varepsilon)).$$

Proof. Observe that

$$(19) \quad (\forall n \in \mathbb{N}) \quad \beta_{n+1} = \beta_n - \delta_n g(\beta_n), \quad \text{where} \quad \delta_n = \frac{\beta_n - \beta_{n+1}}{g(\beta_n)} \geq \rho_n.$$

Hence, by Proposition 2.4, $\underline{\rho} \leq \liminf_{n \rightarrow \infty} H(\beta_n)/n$. Let $\varepsilon \in]0, \underline{\rho}[$. Then there exists $m \in \mathbb{N}$ such that $(\forall n \geq m) \quad \underline{\rho} - \varepsilon \leq H(\beta_n)/n \Leftrightarrow H^{-1}(n(\underline{\rho} - \varepsilon)) \geq \beta_n$. ■

Example 2.9 Let $(\beta_n)_{n \in \mathbb{N}}$ and $(\rho_n)_{n \in \mathbb{N}}$ be sequences in \mathbb{R}_{++} , let $1 \leq q < \infty$, and suppose that

$$(20) \quad (\forall n \in \mathbb{N}) \beta_{n+1} \leq \beta_n - \rho_n \beta_n^q \quad \text{and} \quad \underline{\rho} = \liminf_{n \rightarrow \infty} \rho_n \in \mathbb{R}_{++}.$$

Let $0 < \varepsilon < \underline{\rho}$. Then there exists $m \in \mathbb{N}$ such that the following hold:

$$(i) \quad \text{If } q > 1, \text{ then } (\forall n \geq m) \beta_n \leq \frac{1}{((q-1)n(\underline{\rho} - \varepsilon))^{1/(q-1)}} = O(1/n^{1/(q-1)}).$$

$$(ii) \quad \text{If } q = 1, \text{ then } (\forall n \geq m) \beta_n \leq \gamma^n, \text{ where } \gamma = \exp(\varepsilon - \underline{\rho}) \in]0, 1[.$$

Consequently, the convergence of $(\beta_n)_{n \in \mathbb{N}}$ to 0 is at least sublinear if $q > 1$ and at least linear if $q = 1$.

Proof. Combine Example 2.3 with Corollary 2.8. ■

Remark 2.10 Example 2.9(i) can also be deduced from [7, Lemma 4.1]; see also [1].

Corollary 2.11 Let $(\beta_n)_{n \in \mathbb{N}}$ and $(\rho_n)_{n \in \mathbb{N}}$ be sequences in \mathbb{R}_{++} , and suppose that

$$(21) \quad (\forall n \in \mathbb{N}) \beta_n \geq \beta_{n+1} \geq \beta_n - \rho_n g(\beta_n) \quad \text{and} \quad \bar{\rho} = \overline{\lim}_{n \rightarrow \infty} \rho_n \frac{g(\beta_n)}{g(\beta_{n+1})} \in \mathbb{R}_+.$$

Then

$$(22) \quad (\forall \varepsilon \in \mathbb{R}_{++})(\exists m \in \mathbb{N})(\forall n \geq m) \beta_n \geq H^{-1}(n(\bar{\rho} + \varepsilon)).$$

Proof. Observe that

$$(23) \quad (\forall n \in \mathbb{N}) \beta_{n+1} = \beta_n - \delta_n g(\beta_n), \quad \text{where} \quad \delta_n = \frac{\beta_n - \beta_{n+1}}{g(\beta_n)} \leq \rho_n.$$

Hence, by Proposition 2.4, $\overline{\lim}_{n \rightarrow \infty} H(\beta_n)/n \leq \bar{\rho}$. Let $\varepsilon \in \mathbb{R}_{++}$. Then there exists $m \in \mathbb{N}$ such that $(\forall n \geq m) \bar{\rho} + \varepsilon \geq H(\beta_n)/n \Leftrightarrow H^{-1}(n(\bar{\rho} + \varepsilon)) \leq \beta_n$. ■

Example 2.12 Let $(\beta_n)_{n \in \mathbb{N}}$ and $(\rho_n)_{n \in \mathbb{N}}$ be sequences in \mathbb{R}_{++} , let $1 \leq q < \infty$, and suppose that

$$(24) \quad (\forall n \in \mathbb{N}) \beta_n \geq \beta_{n+1} \geq \beta_n - \rho_n \beta_n^q \quad \text{and} \quad \bar{\rho} = \overline{\lim}_{n \rightarrow \infty} \rho_n \frac{\beta_n^q}{\beta_{n+1}^q} \in \mathbb{R}_+.$$

Let $\varepsilon \in \mathbb{R}_{++}$. Then there exists $m \in \mathbb{N}$ such that the following hold:

$$(i) \quad \text{If } q > 1, \text{ then } (\forall n \geq m) \beta_n \geq \frac{1}{((q-1)n(\bar{\rho} + \varepsilon))^{1/(q-1)}}.$$

$$(ii) \quad \text{If } q = 1, \text{ then } (\forall n \geq m) \beta_n \geq \gamma^n, \text{ where } \gamma = \exp(-\bar{\rho} - \varepsilon) \in]0, 1[.$$

Consequently, the convergence of $(\beta_n)_{n \in \mathbb{N}}$ to 0 is at best sublinear if $q > 1$ and at best linear if $q = 1$.

Proof. Combine Example 2.3 with Corollary 2.11. ■

3 Proximal point algorithm (PPA)

This section focuses on the proximal point algorithm. We assume that

$$(25) \quad f: \mathbb{R} \rightarrow]-\infty, +\infty] \text{ is convex, lower semicontinuous, proper,}$$

with

$$(26) \quad f(0) = 0 \text{ and } f(x) > 0 \text{ when } x \neq 0.$$

Given $x_0 \in \mathbb{R}$, we will study the basic proximal point iteration

$$(27) \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = P_f(x_n).$$

Note that if $x > 0$ and $y < 0$, then $f(y) + \frac{1}{2}|x - y|^2 > f(0) + \frac{1}{2}|x - 0|^2 \geq f(P_f x) + \frac{1}{2}|x - P_f x|^2$. Hence the behaviour of $f|_{\mathbb{R}_{--}}$ is *irrelevant* for the determination of $P_f|_{\mathbb{R}_{++}}$ (and an analogous statement holds for the determination of $P_f|_{\mathbb{R}_{--}}$)! For this reason, we restrict our attention to the case when

$$(28) \quad x_0 \in \mathbb{R}_{++}$$

is the starting point of the proximal point algorithm. The general theory (Fact 1.2) then yields

$$(29) \quad x_0 \geq x_1 \geq \dots \geq x_n \downarrow 0.$$

In this section, it will be convenient to additionally assume that

$$(30) \quad f \text{ is an even function;}$$

although, as mentioned, the behaviour of $f|_{\mathbb{R}_{--}}$ is actually irrelevant because $x_0 \in \mathbb{R}_{++}$. Combining the assumption that 0 is the unique minimizer of f with [15, Theorem 24.1], we learn that

$$(31) \quad 0 \in \partial f(0) = [f'_-(0), f'_+(0)] \cap \mathbb{R} = [-f'_+(0), f'_+(0)] \cap \mathbb{R}.$$

We start our exploration by discussing convergence in finitely many steps.

Proposition 3.1 (finite convergence) *We have $x_n \rightarrow 0$ in finitely many steps, regardless of the starting point $x_0 \in \mathbb{R}_{++}$, if and only if*

$$(32) \quad 0 < f'_+(0),$$

in which case $P_f x_n = 0 \Leftrightarrow x_n \leq f'_+(0)$.

Proof. Let $x > 0$. Then $P_f x = 0 \Leftrightarrow x \in 0 + \partial f(0) \Leftrightarrow x \leq f'_+(0)$ by (31).

Suppose first that $f'_+(0) > 0$. Then, by (31), $0 \in \text{int } \partial f(0)$ and, using (29), there exists $n \in \mathbb{N}$ such that $x_n \leq f'_+(0)$. It follows that $x_{n+1} = x_{n+2} = \dots = 0$. (Alternatively, this follows from a much more general result of Rockafellar; see [16, Theorem 3] and also Remark 3.4 below.)

Now assume that there exists $n \in \mathbb{N}$ such that $P_f x_n = 0$ and $x_n > 0$. By the above, $x_n \leq f'_+(0)$ and thus $f'_+(0) > 0$. ■

An extreme case occurs when $f'_+(0) = +\infty$ in Proposition 3.1:

Example 3.2 ($\iota_{\{0\}}$ and the projector) Suppose that $f = \iota_{\{0\}}$. Then $P_f = P_{\{0\}}$ and $(\forall n \geq 1) x_n = 0$.

Example 3.3 ($|x|^1$ and the threshold) Suppose that $f = |\cdot|$ in which case $\partial f(0) = [-1, 1]$ and $f_+(0) = 1$. Proposition 3.1 guarantees finite convergence of the PPA. Indeed, either a direct argument or [3, Example 14.5] yields

$$(33) \quad P_f: x \mapsto \begin{cases} x - \frac{x}{|x|}, & \text{if } |x| > 1; \\ 0, & \text{otherwise,} \end{cases}$$

Consequently, $x_n = 0$ if and only if $n \geq \lceil x_0 \rceil$.

Remark 3.4 In [16, Theorem 3], Rockafellar provided a very general *sufficient* condition for finite convergence of the PPA (which works actually for finding zeros of a maximally monotone operator defined on a Hilbert space). In our present setting, his condition is

$$(34) \quad 0 \in \text{int } \partial f(0).$$

By Proposition 3.1, this is also a condition that is *necessary* for finite convergence.

Thus, we assume from now on that $f'_+(0) = 0$, or equivalently (since f is even and by (31)), that

$$(35) \quad f'(0) = 0.$$

in which case finite convergence fails and thus

$$(36) \quad x_0 > x_1 > \cdots > x_n \downarrow 0.$$

We now have the following sufficient condition for linear convergence. The proof is a refinement of the ideas of Rockafellar in [16].

Proposition 3.5 (sufficient condition for linear convergence) *Suppose that*

$$(37) \quad \lambda = \liminf_{x \downarrow 0} \frac{f(x)}{x^2} \in]0, +\infty].$$

Then the following hold:

(i) *If $\lambda < +\infty$, then there exists $\alpha_0 \in [\frac{1}{2\lambda}, \frac{1}{\lambda}]$ such that*

$$(38) \quad (\forall \varepsilon > 0)(\exists m \in \mathbb{N})(\forall n \geq m) \quad |x_{n+1}| \leq \frac{\alpha_0}{\sqrt{1 + \alpha_0^2(1 + 2\lambda - 2\varepsilon)}} |x_n|.$$

(ii) If $\lambda = +\infty$, then

$$(39) \quad (\forall \alpha > 0)(\forall \varepsilon > 0)(\exists m \in \mathbb{N})(\forall n \geq m) \quad |x_{n+1}| \leq \frac{\alpha}{\sqrt{1 + \alpha^2(1 + 2\lambda - \varepsilon)}} |x_n|.$$

Proof. By [16, Remark 4 and Proposition 7], there exists $\alpha_0 \in [\frac{1}{2\lambda}, \frac{1}{\lambda}]$ such that $(\partial f)^{-1}$ is Lipschitz continuous at 0 with every modulus $\alpha > \alpha_0$. Let $\alpha > \alpha_0$. Then there exists $\tau > 0$ such that

$$(40) \quad (\forall |x| < \tau)(\forall z \in (\partial f)^{-1}(x)) \quad |z| \leq \alpha|x|.$$

Since $x_n \rightarrow 0$ by [16, Theorem 2] (or (36)), there exists $m \in \mathbb{N}$ such that $(\forall n \geq m) |x_n - x_{n+1}| \leq \tau$. Let $n \geq m$. Noticing that $x_n \in (\text{Id} + \partial f)(x_{n+1})$, we have

$$(41) \quad x_{n+1} \in (\partial f)^{-1}(x_n - x_{n+1}).$$

It follows by (40) that

$$(42) \quad |x_{n+1}| \leq \alpha|x_n - x_{n+1}|.$$

Since $x_n - x_{n+1} \in \partial f(x_{n+1})$, we have

$$(43) \quad \langle x_n - x_{n+1}, x_{n+1} \rangle = \langle x_n - x_{n+1}, x_{n+1} - 0 \rangle \geq f(x_{n+1}) - f(0) = f(x_{n+1}).$$

Now for every $\varepsilon > 0$, employing (37) and increasing m if necessary, we can and do assume that

$$(44) \quad (\forall n \geq m) \quad \langle x_n - x_{n+1}, x_{n+1} \rangle \geq \left(\lambda - \frac{\varepsilon}{2}\right) |x_{n+1}|^2.$$

Let $n \geq m$. Combining (42) and (44), we obtain

$$(45a) \quad |x_n|^2 = |x_{n+1}|^2 + |x_n - x_{n+1}|^2 + 2 \langle x_n - x_{n+1}, x_{n+1} \rangle$$

$$(45b) \quad \geq |x_{n+1}|^2 + \frac{1}{\alpha^2} |x_{n+1}|^2 + (2\lambda - \varepsilon) |x_{n+1}|^2$$

$$(45c) \quad = \left(\frac{1 + \alpha^2(1 + 2\lambda - \varepsilon)}{\alpha^2} \right) |x_{n+1}|^2.$$

This gives

$$(46) \quad |x_{n+1}| \leq \frac{\alpha}{\sqrt{1 + \alpha^2(1 + 2\lambda - \varepsilon)}} |x_n|$$

and hence (39) holds. Now assume that $\lambda < +\infty$ so that $\alpha_0 > 0$. Since $\alpha \mapsto \frac{\alpha}{\sqrt{1 + \alpha^2(1 + 2\lambda - \varepsilon)}} = \frac{1}{\sqrt{\frac{1}{\alpha^2} + (1 + 2\lambda - \varepsilon)}}$ is strictly increasing on \mathbb{R}_+ , we note that the choice $\alpha = \alpha_0 / \sqrt{1 - \varepsilon \alpha_0^2} > \alpha_0$ yields (38). ■

Remark 3.6 Assume that f is differentiable on $U =]0, \delta[$, where $\delta \in \mathbb{R}_{++}$. Then $f'(x) > 0$ on U . Note that $\lim_{x \downarrow 0} \frac{f'(x)}{x} = f_+''(0)$. Therefore, L'Hôpital's rule shows that if $f_+''(0)$ exists in $[0, +\infty]$, then

$$(47) \quad \lambda = \frac{1}{2}f_+''(0)$$

in (37). A sufficient condition for λ to exist is to assume that the function $f(x)/x^2$ is monotone on U which in turn happens when $2f(x) - xf'(x)$ is either nonnegative or nonpositive on U by using the quotient rule.

Although we won't need it in the remainder of this paper, we point out that the proof of Proposition 3.5 still works in a more general setting leading to the following result:

Corollary 3.7 *Let H be a real Hilbert space, and let $f: H \rightarrow]-\infty, +\infty]$ be convex, lower semicontinuous and proper such that 0 is the unique minimizer of f . Assume also that*

$$(48) \quad \lambda = \liminf_{0 \neq x \rightarrow 0} \frac{f(x)}{\|x\|^2} \in]0, +\infty].$$

Then there exists $\alpha_0 \in [\frac{1}{2\lambda}, \frac{1}{\lambda}]$ such that

$$(49) \quad (\forall \alpha > \alpha_0)(\forall \varepsilon > 0)(\exists m \in \mathbb{N})(\forall n \geq m) \quad \|x_{n+1}\| \leq \frac{\alpha}{\sqrt{1 + \alpha^2(1 + 2\lambda - \varepsilon)}} \|x_n\|.$$

If $\lambda < +\infty$, then eventually

$$(50) \quad \|x_{n+1}\| \leq \frac{\alpha_0}{\sqrt{1 + \alpha_0}} \|x_n\|,$$

a result which can also be deduced from [16, Theorem 2].

We now discuss powers of the absolute value function.

Example 3.8 ($|x|^2$ and linear convergence) Suppose that $f(x) = x^2$. Then $x + f'(x) = 3x$ and hence $P_f = \frac{1}{3} \text{Id}$. We see that the actual linear rate of convergence of the PPA is

$$(51) \quad \frac{1}{3}.$$

Now consider Proposition 3.5 and Corollary 3.7. Then clearly $\lambda = 1$ in (37) and hence $\alpha_0 \in [\frac{1}{2}, 1]$. In fact, since $\partial f = 2 \text{Id}$ and so $(\partial f)^{-1} = \frac{1}{2} \text{Id}$, we know that the tightest choice for α_0 is $\alpha_0 = \frac{1}{2}$. The linear rate obtained by (50) is $(1/2) / \sqrt{1 + (1/2)^2}$, i.e.,

$$(52) \quad \frac{1}{\sqrt{5}}.$$

Let us compare to the linear rate provided by Proposition 3.5, where, for every $\varepsilon > 0$, we obtain $(1/2) / \sqrt{1 + (1/2)^2(1 + 2 - 2\varepsilon)} = 1 / \sqrt{7 - 2\varepsilon}$. From the proof of Proposition 3.5, we see that we can here actually set $\varepsilon = 0$; thus, the rate provided is

$$(53) \quad \frac{1}{\sqrt{7}}.$$

In summary, $1/\sqrt{7}$, the rate from Proposition 3.5, is better than $1/\sqrt{5}$, which comes from (50); however, even the former does not capture the true rate $1/3$.

Example 3.9 ($|x|^q$, where $1 < q < 2$, and superlinear convergence)

Suppose that $f(x) = |x|^q$, where $1 < q < 2$. Note that $\lambda = +\infty$ and thus $\alpha_0 = 0$ in (38). In passing, we point out that we cannot use α_0 itself in (38) because it would imply finite convergence which does not occur by (36). Set $\phi(x) = x + qx^{q-1}$ and note that $(\forall n \in \mathbb{N}) x_n = \phi(x_{n+1})$. Now set also $\psi(x) = qx^{q-1}$, and assume that a sequence $(\rho_n)_{n \in \mathbb{N}}$ satisfies $(\forall n \in \mathbb{N}) \rho_n = \psi(\rho_{n+1})$. The sequence $(\rho_n)_{n \in \mathbb{N}}$ can be thought of as an approximation of $(x_n)_{n \in \mathbb{N}}$. It has the advantage that the implicit recursion is invertible and solvable; indeed one may verify by induction that

$$(54) \quad (\forall n \in \mathbb{N}) \quad \rho_n = \rho_0^{1/(q-1)^n} (1/q)^{(1/(q-1)^n - 1)/(2-q)};$$

Assume furthermore that $\rho_0 = x_0$ is sufficiently close to 0. Since ϕ and ψ are increasing and $\phi > \psi > 0$ on \mathbb{R}_{++} , we deduce that $(\forall n \geq 1) x_n < \rho_n$. Therefore,

$$(55) \quad (\forall n \in \mathbb{N}) \quad \frac{x_n}{\rho_n} = \frac{\phi(x_{n+1})}{\psi(x_{n+1})} \left(\frac{x_{n+1}}{\rho_{n+1}} \right)^{q-1} > \left(\frac{x_{n+1}}{\rho_{n+1}} \right)^{q-1},$$

which implies that

$$(56) \quad 0 < \frac{x_n}{\rho_n} < \left(\frac{x_1}{\rho_1} \right)^{1/(q-1)^{n-1}} \rightarrow 0$$

because $1/(q-1) > 1$. Let $n \in \mathbb{N}$. It follows from $x_n = x_{n+1} + qx_{n+1}^{q-1} > x_{n+1}$ and $1 < q < 2$ that $x_{n+1}x_n^{q-1} < x_nx_{n+1}^{q-1}$, and so

$$(57) \quad \frac{x_n}{x_{n-1}} = \frac{x_{n+1} + qx_{n+1}^{q-1}}{x_n + qx_n^{q-1}} < \frac{x_{n+1}^{q-1}}{x_n^{q-1}}.$$

This gives $\frac{x_n^{1/(q-1)}}{x_{n-1}^{1/(q-1)}} < \frac{x_{n+1}}{x_n}$; hence,

$$(58) \quad \frac{x_n}{x_{n-1}^{1/(q-1)}} < \frac{x_{n+1}}{x_n^{1/(q-1)}}.$$

On the other hand, $x_n = x_{n+1} + qx_{n+1}^{q-1} > qx_{n+1}^{q-1}$, which yields $(\frac{x_n}{q})^{1/(q-1)} > x_{n+1}$ and hence $\frac{x_{n+1}}{x_n^{1/(q-1)}} < q^{1/(1-q)}$. The sequence $(x_{n+1}/(x_n^{1/(q-1)}))_{n \in \mathbb{N}}$ is thus increasing and bounded above, and so it converges to some $\mu > 0$. We obtain that $x_n \rightarrow 0$ superlinearly with order $1/(q-1)$.

Example 3.10 ($|x|^q$, where $2 < q$, and logarithmic convergence)

Suppose that $f(x) = |x|^q$, where $2 < q < +\infty$. Because $(\forall n \in \mathbb{N}) x_n = x_{n+1} + qx_{n+1}^{q-1}$, we have $x_n/x_{n+1} = 1 + qx_{n+1}^{q-2} \rightarrow 1$. It thus follows from Example 2.7 that $x_n \rightarrow 0$ logarithmically and

$$(59) \quad \frac{x_n}{(\frac{1}{n})^{1/(q-2)}} \rightarrow \frac{1}{((q-2)q)^{1/(q-2)}}.$$

Let us summarize what we found out in the previous three examples about the behaviour of the PPA applied to $|x|^q$:

q	PPA convergence of $x_n \rightarrow 0$ for $f(x) = x ^q$
$1 < q < 2$	superlinear with order $1/(q-1)$
$q = 2$	linear with rate $1/3$
$2 < q < +\infty$	logarithmic

4 Method of Alternating Projections (MAP)

We now turn to the method of alternating projections. As in Section 3, we assume without loss of generality that

$$(60a) \quad f: \mathbb{R} \rightarrow]-\infty, +\infty] \text{ is convex, lower semicontinuous, and proper,}$$

with

$$(60b) \quad f \text{ even, } f(0) = 0, f > 0 \text{ otherwise, and } f'(0) = 0.$$

Furthermore, we set

$$(60c) \quad A = \mathbb{R} \times \{0\} \quad \text{and} \quad B = \text{epi } f.$$

The projection onto A is very simple:

$$(61) \quad P_A: \mathbb{R}^2 \rightarrow \mathbb{R}^2: (x, r) \mapsto (x, 0).$$

We now turn to P_B .

Fact 4.1 (See [3, Proposition 9.18 and Proposition 28.28]) *Let $(x, r) \in (\text{dom } f \times \mathbb{R}) \setminus B$. Then $P_B(x, r) = (y, f(y))$, where y satisfies $x \in y + (f(y) - r)\partial f(y)$. Moreover, $r < f(y)$ and $(\forall z \in \text{dom } f) (z - y)(x - y) \leq (f(z) - f(y))(f(y) - r)$.*

Corollary 4.2 *Suppose that f is differentiable at 0, let $(x, r) \in (\text{dom } f \times \mathbb{R}) \setminus B$, and set $(y, f(y)) = P_B(x, r)$. Then $y = 0$ if $x = 0$, and y lies strictly between x and 0 otherwise. Furthermore, $x^2 + r^2 \geq y^2 + f(y)^2 + (x - y)^2 + (r - f(y))^2$.*

Proof. We use Fact 4.1. Observe that $r < f(y)$ and hence that $f(y) - r > 0$. Choosing $z = 0$ gives $-y(x - y) \leq -f(y)(f(y) - r) \leq 0$. If $x = 0$, this implies $y^2 \leq 0$, and so $y = 0$. Assume that $x \neq 0$. Then $y \neq 0$ since $y = 0$ implies $x = 0 + (f(0) - r)f'(0) = 0$. We obtain $-f(y)(f(y) - r) < 0$, and then $-y(x - y) < 0$, i.e., $y(x - y) > 0$. It follows that $y \in]0, x[$ if $x > 0$, and $y \in]x, 0[$ if $x < 0$. Finally, since P_B is firmly nonexpansive (see, e.g., [3, Proposition 4.8]) and $P_B(0, 0) = (0, 0)$, we obtain

$$(62) \quad \|(x, r) - (0, 0)\|^2 \geq \|(y, f(y)) - (0, 0)\|^2 + \|(x, r) - (y, f(y))\|^2,$$

which completes the proof. ■

We now turn to the sequence generated by the method of alternating projections. We assume without loss of generality that

$$(63a) \quad x_0 \in \mathbb{R}_{++} \cap \text{dom } f, \quad a_0 = (x_0, 0) \in A,$$

and

$$(63b) \quad (\forall n \in \mathbb{N}) \quad a_{n+1} = P_A P_B(a_n) = (x_{n+1}, 0).$$

Combining Fact 1.3, (60) and (63), we learn that

$$(64) \quad x_0 > x_1 > \cdots > x_n \downarrow 0.$$

We are ready for our first result on the lack of linear convergence for MAP.

Theorem 4.3 *The following hold:*

$$(65) \quad (\forall n \in \mathbb{N}) \quad x_{n+1}(x_{n+1} - x_n) + f^2(x_{n+1}) \leq 0,$$

$$(66) \quad (\forall n \in \mathbb{N}) \quad x_n = x_{n+1} + f(x_{n+1})x_{n+1}^*, \text{ where } x_{n+1}^* \in \partial f(x_{n+1}),$$

and $x_n \rightarrow 0$ sublinearly, i.e.,

$$(67) \quad \frac{x_{n+1}}{x_n} \rightarrow 1.$$

If f is differentiable on some interval $[0, \delta]$, where $\delta \in \mathbb{R}_{++}$, and there exists $q \in \mathbb{R}$ such that

$$(68) \quad \lim_{x \downarrow 0} \frac{f(x)f'(x)}{x^q} = c_q \in \mathbb{R}_{++},$$

then $x_n \rightarrow 0$ logarithmically, i.e.,

$$(69) \quad \frac{x_{n+1} - x_{n+2}}{x_n - x_{n+1}} \rightarrow 1.$$

Proof. Corollary 4.2 implies (65). Using Fact 4.1, we have (66), which yields

$$(70) \quad \frac{x_n}{x_{n+1}} = 1 + \frac{f(x_{n+1}) - f(0)}{x_{n+1}} x_{n+1}^* \rightarrow 1 + f'(0)f'(0) = 1$$

because of $f'(0) = 0$ and [3, Proposition 17.32]. This gives (67). Now suppose that f is differentiable on $[0, \delta]$ and (68) holds. Hence, using also (66),

$$(71) \quad \frac{x_{n+1} - x_{n+2}}{x_n - x_{n+1}} = \frac{\frac{f(x_{n+2})f'(x_{n+2})}{x_{n+2}^q}}{\frac{f(x_{n+1})f'(x_{n+1})}{x_{n+1}^q}} \left(\frac{x_{n+2}}{x_{n+1}} \right)^q \rightarrow \frac{c_q}{c_q} \cdot 1^q = 1,$$

as claimed. ■

Remark 4.4 The function f satisfies (68) with $q = 2a - 1$ and $c_q = a\varphi^2(0)$ whenever $f(x) = x^a\varphi(x)$, where $a \in \mathbb{R} \setminus \{0\}$, $\delta \in \mathbb{R}_{++}$, φ is differentiable on $[0, \delta]$, φ' is continuous at 0, and $\varphi(0) \neq 0$.

Proposition 4.5 Suppose that f is differentiable on $[0, \delta]$, where $\delta \in \mathbb{R}_{++}$. Set

$$(72) \quad (\forall q \in [1, +\infty[) \quad c_q = \lim_{x \downarrow 0} \frac{f(x)f'(x)}{x^q},$$

where c_q is either undefined if the limit does not exist or in $[0, +\infty]$. Let $q \in [1, +\infty[$. Then the following hold:

- (i) $c_1 = 0$. If $c_q = 0$, then $c_{q'} = 0$ for $1 \leq q' < q$. If $c_q > 0$, then $c_{q'} = +\infty$ for $q' > q$.
- (ii) If $c_q = 0$, then

$$(73) \quad \frac{\frac{x_n}{x_{n+1}} - 1}{x_{n+1}^{q-1}} \rightarrow 0$$

and

$$(74) \quad (\forall \varepsilon \in \mathbb{R}_{++})(\exists m \in \mathbb{N})(\forall n \geq m) \quad x_n \geq \begin{cases} (\exp(-\varepsilon))^n, & \text{when } q = 1; \\ \frac{1}{((q-1)n\varepsilon)^{1/(q-1)}}, & \text{when } q > 1. \end{cases}$$

- (iii) If $c_q > 0$, then $q > 1$ and $\frac{x_n}{(\frac{1}{n})^{1/(q-1)}} \rightarrow \frac{1}{((q-1)c_q)^{1/(q-1)}}$.

Proof. (i): Using L'Hôpital's rule, we have $\lim_{x \downarrow 0} \frac{f(x)}{x} = \lim_{x \downarrow 0} \frac{f'(x)}{1} = f'(0) = 0$ and hence $c_1 = \lim_{x \downarrow 0} \frac{f(x)f'(x)}{x} = 0$. The remaining statements follow now readily.

(ii): It follows from (66) that

$$(75) \quad \frac{\frac{x_n}{x_{n+1}} - 1}{x_{n+1}^{q-1}} = \frac{f(x_{n+1})f'(x_{n+1})}{x_{n+1}^q} \rightarrow c_q = 0;$$

thus, (73) holds. Now write

$$(76) \quad x_{n+1} = x_n - \frac{f(x_{n+1})f'(x_{n+1})}{x_{n+1}^q} \left(\frac{x_{n+1}}{x_n} \right)^q x_n^q = x_n - \rho_n x_n^q,$$

and note that $\rho_n \rightarrow 0$ because $x_{n+1}/x_n \rightarrow 1$ and $c_q = 0$. Thus, (74) holds due to Example 2.12.

(iii): We must have $q > 1$ since otherwise $c_q = c_1 = 0$ by (i), which is absurd. From (66), we have

$$(77) \quad \frac{x_n}{x_{n+1}} = 1 + \frac{f(x_{n+1})f'(x_{n+1})}{x_{n+1}} \rightarrow 1$$

and also, for every $n \in \mathbb{N}$,

$$(78) \quad x_n = x_{n+1} + f(x_{n+1})f'(x_{n+1}) = x_{n+1} + \frac{f(x_{n+1})f'(x_{n+1})}{x_{n+1}^q} x_{n+1}^q.$$

The conclusion therefore follows from Example 2.7. ■

Example 4.6 ($\frac{1}{p}|x|^p$, where $p > 1$) Suppose that $f(x) = \frac{1}{p}|x|^p$, where $1 < p < +\infty$. Let $x \in \mathbb{R}_{++}$. Then $f(x) = \frac{1}{p}x^p$ and $f'(x) = x^{p-1}$. Setting $q = 2p - 1 > 1$, we have $c_q = \frac{1}{p} > 0$, and so $x_n \rightarrow 0$ logarithmically, using Theorem 4.3. Moreover, by Proposition 4.5(iii),

$$(79) \quad \frac{x_n}{\left(\frac{1}{n}\right)^{1/(2p-2)}} \rightarrow \frac{1}{\left(\frac{2p-2}{p}\right)^{1/(2p-2)}}.$$

For a couple of cases, one can actually invert (66) and simplify (79):

$$(80) \quad p = \frac{3}{2} \Rightarrow \frac{x_n}{1/n} \rightarrow \frac{3}{2} \text{ and } x_{n+1} = \frac{\sqrt{9 + 24x_n} - 3}{4};$$

and

$$(81) \quad p = 2 \Rightarrow \frac{x_n}{1/\sqrt{n}} \rightarrow 1 \text{ and } x_{n+1} = \frac{1}{3} \frac{\left(27x_n + 3\sqrt{81x_n^2 + 24}\right)^{2/3} - 6}{\sqrt[3]{27x_n + 3\sqrt{81x_n^2 + 24}}}.$$

Example 4.7 ($R - \sqrt{R^2 - x^2}$) Suppose that $R \in \mathbb{R}_{++}$ and that $f(x) = R - \sqrt{R^2 - x^2}$ on its domain $[-R, R]$. Let $n \in \mathbb{N}$. Then $f'(x) = \frac{x}{\sqrt{R^2 - x^2}}$, and by (66),

$$(82) \quad x_n = x_{n+1} + \left(R - \sqrt{R^2 - x_{n+1}^2}\right) \frac{x_{n+1}}{\sqrt{R^2 - x_{n+1}^2}} = \frac{Rx_{n+1}}{\sqrt{R^2 - x_{n+1}^2}}.$$

It follows that

$$(83) \quad x_{n+1} = \frac{Rx_n}{\sqrt{x_n^2 + R^2}},$$

and also $\frac{R^2}{x_{n+1}^2} = 1 + \frac{R^2}{x_n^2}$. Hence, $\frac{R^2}{x_n^2} = n + \frac{R^2}{x_0^2}$, which yields the *explicit* formula

$$(84) \quad x_n = \frac{Rx_0}{\sqrt{nx_0^2 + R^2}} = \frac{R}{\sqrt{n + (R/x_0)^2}} \sim \frac{R}{\sqrt{n}},$$

which shows that $x_n \rightarrow 0$ logarithmically.

Example 4.8 Suppose that $f(x) = \exp(|x|) - |x| - 1$. Then $f(x)f'(x) = \frac{1}{2}x^3 + \frac{5}{12}x^4 + \frac{5}{24}x^5 + O(x^6)$ on \mathbb{R}_{++} . Thus, Proposition 4.5(iii) with $q = 3$ and $c_q = \frac{1}{2}$ yields $x_n/(1/\sqrt{n}) \rightarrow 1$.

Example 4.9 Suppose that $f = \cosh$. Then $f(x)f'(x) = \frac{1}{2}x^3 + \frac{1}{8}x^5 + O(x^7)$ and hence again by Proposition 4.5(iii) with $q = 3$ and $c_q = \frac{1}{2}$ yields $x_n/(1/\sqrt{n}) \rightarrow 1$.

Example 4.10 (extremely slow convergence) Suppose that $f(x) = \exp(-x^{-2})$ with domain $[-\sqrt{2/3}, \sqrt{2/3}]$. Then $c_q = 0$ in Proposition 4.5(ii) and hence, according to (73),

$$(85) \quad \frac{\frac{x_n}{x_{n+1}} - 1}{x_{n+1}^{q-1}} \rightarrow 0 \quad \text{for every } q > 1.$$

Furthermore, the convergence $(x_n)_{n \in \mathbb{N}}$ to 0 is extremely slow in the sense that

$$(86) \quad \frac{1}{x_n} \leq O(n^{1/p}) \quad \text{for every } p > 0.$$

Remark 4.11 (MAP sequence is essentially a PPA sequence) Note that, by (66),

$$(87) \quad (\forall n \in \mathbb{N}) \quad x_n \in (\text{Id} + f \cdot \partial f)(x_{n+1}) = (\text{Id} + \partial \frac{1}{2}f^2)(x_{n+1})$$

and so

$$(88) \quad (\forall n \in \mathbb{N}) \quad x_{n+1} = P_{\frac{1}{2}f^2}(x_n).$$

Since f^2 is convex (by, e.g., [3, Proposition 8.19]), we see that the sequence $(a_n)_{n \in \mathbb{N}} = (x_n, 0)_{n \in \mathbb{N}}$ generated by MAP is essentially the same as the sequence generated by PPA for the function $\frac{1}{2}f^2$! This useful connection will be further discussed after we recall a special case of a result due to Güler.

Fact 4.12 (Güler) (See [10, Theorem 3.1].) *Let H be a real Hilbert space, let $f: H \rightarrow]-\infty, +\infty]$ be convex, lower semicontinuous, and suppose that the sequence $(x_n)_{n \in \mathbb{N}}$ generated by the PPA converges (strongly) to some minimizer z of f . Then $f(x_n) - f(z) = o(1/n)$, i.e., $n(f(x_n) - f(z)) \rightarrow 0$.*

Combining Remark 4.11 with Fact 4.12 results in the following:

Corollary 4.13 *The MAP sequence $(a_n)_{n \in \mathbb{N}} = (x_n, 0)_{n \in \mathbb{N}}$ satisfies*

$$(89) \quad nf^2(x_n) \rightarrow 0.$$

Example 4.14 ($\frac{1}{p}|x|^p$ revisited) Suppose that $f(x) = \frac{1}{p}|x|^p$, where $1 < p$. By Example 4.6,

$$(90) \quad x_n \sim \left(\frac{1}{n}\right)^{1/(2p-2)}.$$

Then (89) becomes

$$(91) \quad nf^2(x_n) \sim nx_n^{2p} \sim \frac{1}{n^{1/(p-1)}} \rightarrow 0.$$

Note that this also shows that this consequence of Fact 4.12 is *sharp* in the sense that it cannot be improved to $n^{1+\varepsilon}f^2(x_n) \rightarrow 0$, where $\varepsilon > 0$. (Indeed, if $n^{1+\varepsilon}f^2(x_n) \rightarrow 0$, then we obtain a contradiction for sufficiently large p .)

5 Douglas–Rachford algorithm (DRA)

Finally, we investigate the Douglas–Rachford algorithm. As in Section 4, we assume that

$$(92a) \quad f: \mathbb{R} \rightarrow]-\infty, +\infty] \text{ is convex, lower semicontinuous, and proper,}$$

with

$$(92b) \quad f \text{ even, } f(0) = 0, f > 0 \text{ otherwise, and } f'(0) = 0,$$

and that

$$(92c) \quad A = \mathbb{R} \times \{0\} \quad \text{and} \quad B = \text{epi } f.$$

We now turn to the sequence generated by the Douglas–Rachford algorithm. We assume that

$$(93a) \quad x_0 \in \mathbb{R}_{++} \cap \text{dom } f, \quad r_0 = 0, \quad z_0 = (x_0, 0) \in A,$$

and

$$(93b) \quad (\forall n \in \mathbb{N}) \quad z_{n+1} = Tz_n = (x_{n+1}, r_{n+1}),$$

where

$$(93c) \quad T = \text{Id} - P_A + P_B R_A.$$

Since $N_{A-B}(0, 0) = \mathbb{R}_+(0, 1)$, we have

$$(94) \quad A \cap B = \{(0, 0)\} \quad \text{and} \quad \text{Fix } T = \mathbb{R}_+(0, 1),$$

Hence we deduce from Fact 1.4, (92) and (93) that

$$(95) \quad x_n \rightarrow 0 \quad \text{and} \quad r_n \rightarrow r_\infty \in \mathbb{R}_+.$$

Let us now investigate the effect of carrying out one DRA step:

Corollary 5.1 (one DRA step) *Let $(x, r) \in \mathbb{R}^2$, set $(x_+, r_+) = T(x, r)$, and suppose that $0 < x \in \text{dom } f$ and $0 \leq r < f(x)$. Then there exists $x_+^* \in \mathbb{R}$ such that*

$$(96a) \quad 0 < x_+ = x - r_+ x_+^* < x, \quad x_+^* \in \partial f(x_+) \quad \text{and} \quad r_+ = r + f(x_+) > r$$

and

$$(96b) \quad x^2 + r^2 \geq x_+^2 + (r_+ - r)^2 + (x - x_+)^2 + r_+^2.$$

Proof. First, we note that $R_A(x, r) = (x, -r)$. Set $(y, s) = P_B(x, -r)$. By Fact 4.1,

$$(97) \quad y = x - (r + f(y))x_+^* \text{ for some } x_+^* \in \partial f(y) \text{ and } s = f(y).$$

Now, (93c) gives

$$(98) \quad (x_+, r_+) = (\text{Id} - P_A)(x, r) + P_B R_A(x, r) = (x, r) - (x, 0) + (y, s) = (y, r + s).$$

Thus $x_+ = y$,

$$(99) \quad x_+ = x - (r + f(x_+))x_+^* \text{ and } r_+ = r + f(x_+),$$

as claimed. The rest follows from Corollary 4.2. ■

Remark 5.2 (DRA step is related to a PPA step) Consider Corollary 5.1. Then

$$(100) \quad x_+ = P_{rf + \frac{1}{2}f^2}(x) \text{ and } r_+ = r + f(x_+),$$

which reveals a connection between the DRA step and the PPA step for $rf + \frac{1}{2}f^2$.

Theorem 5.3 (DRA sequence) *The DRA sequence satisfies*

$$(101a) \quad x_n \downarrow 0 \text{ and } r_n \uparrow r_\infty \in \mathbb{R}_{++},$$

and for every $n \in \mathbb{N}$, there exists $x_{n+1}^* \in \partial f(x_{n+1})$ such that

$$(101b) \quad 0 < x_{n+1} = x_n - r_{n+1}x_{n+1}^* < x_n \text{ and } r_{n+1} = r_n + f(x_{n+1}).$$

Now suppose that furthermore that $f_+''(0)$ exists in $[0, +\infty]$. Then

$$(102) \quad \frac{x_{n+1}}{x_n} = \frac{1}{1 + r_{n+1} \frac{x_{n+1}^*}{x_{n+1}}} \rightarrow \frac{1}{1 + r_\infty f_+''(0)}$$

and exactly one of the following holds:

- (i) $f_+''(0) = +\infty$ and $x_n \rightarrow 0$ superlinearly.
- (ii) $f_+''(0) \in \mathbb{R}_{++}$ and $x_n \rightarrow 0$ linearly.
- (iii) $f_+''(0) = 0$ and $x_n \rightarrow 0$ sublinearly. If there exists $q \in \mathbb{R}$ such that

$$(103) \quad \lim_{x \downarrow 0} \frac{f'(x)}{x^q} = c \in \mathbb{R}_{++},$$

then $x_n \rightarrow 0$ logarithmically; moreover, if additionally $q > 1$, then

$$(104) \quad \frac{x_n}{\left(\frac{1}{n}\right)^{1/(q-1)}} \rightarrow \frac{1}{((q-1)r_\infty c)^{1/(q-1)}}.$$

Proof. (101) follows from Corollary 5.1. Divide (101b) by x_{n+1} , solve for x_n/x_{n+1} , then take reciprocals to obtain

$$(105) \quad \frac{x_{n+1}}{x_n} = \frac{1}{1 + r_{n+1} \frac{x_{n+1}^*}{x_{n+1}}}.$$

Now assume that $f''_+(0)$ exists; it belongs to $[0, +\infty]$ because $x > 0 \Rightarrow f'(x) > 0$. Since $x_n \downarrow 0$, we see that

$$(106) \quad \frac{x_{n+1}^*}{x_{n+1}} = \frac{f'(x_{n+1})}{x_{n+1}} = \frac{f'(x_{n+1}) - f'(0)}{x_{n+1} - 0} \rightarrow f''_+(0).$$

Altogether, we get (102). Items (i) and (ii) are now clear, so let us focus on (iii). Obviously $x_n \rightarrow 0$ sublinearly if and only if $f''_+(0) = 0$ which we henceforth assume, along with (103). It follows from (101b) that

$$(107) \quad \frac{x_{n+1} - x_{n+2}}{x_n - x_{n+1}} = \frac{r_{n+2} f'(x_{n+2})}{r_{n+1} f'(x_{n+1})} = \frac{r_{n+2}}{r_{n+1}} \frac{\frac{f'(x_{n+2})}{x_{n+2}^q}}{\frac{f'(x_{n+1})}{x_{n+1}^q}} \left(\frac{x_{n+2}}{x_{n+1}} \right)^q \rightarrow \frac{r_\infty}{r_\infty} \cdot \frac{c}{c} \cdot 1^q = 1;$$

hence, $x_n \rightarrow 0$ logarithmically. Finally assume that $q > 1$. Writing (see (101b))

$$(108) \quad x_n = x_{n+1} + r_{n+1} \frac{f'(x_{n+1})}{x_{n+1}^q} x_{n+1}^q = x_{n+1} + \delta_n x_{n+1}^q,$$

where $\delta_n \rightarrow r_\infty c > 0$, we obtain (104) through Example 2.7. ■

Example 5.4 ($\frac{1}{p}|x|^p$, where $1 < p < +\infty$) Suppose that $f(x) = \frac{1}{p}|x|^p$, where $1 < p < +\infty$. Then exactly one of the following holds:

(i) $1 < p < 2$ and $\frac{x_{n+1}}{x_n^{1/(p-1)}} \rightarrow \frac{1}{r_\infty^{1/(p-1)}} > 0$.

(ii) $p = 2$ and $x_n \rightarrow 0$ linearly with rate $1/(1 + r_\infty)$.

(iii) $2 < p < +\infty$, $x_n \rightarrow 0$ logarithmically, and $\frac{x_n}{\left(\frac{1}{n}\right)^{1/(p-2)}} \rightarrow \frac{1}{((p-2)r_\infty)^{1/(p-2)}}$.

Proof. (i): From (101b), we obtain

$$(109) \quad (\forall n \in \mathbb{N}) \quad x_n = x_{n+1} + r_{n+1} x_{n+1}^{p-1}.$$

Since $x_n \downarrow 0$ and $p < 2$, we have

$$(110) \quad (\forall \varepsilon > 0)(\exists m \in \mathbb{N})(\forall n \geq m) \quad 0 < x_{n+1} < \varepsilon x_{n+1}^{p-1}.$$

Combining yields $r_{n+1} x_{n+1}^{p-1} < x_n < (r_{n+1} + \varepsilon) x_{n+1}^{p-1}$. In turn, $\frac{x_n}{x_{n+1}^{p-1}} \rightarrow r_\infty > 0$ and the conclusion follows.

(ii): Clear from (102).

(iii): Apply Theorem 5.3(iii) with $q = p - 1 > 1$ and $c = 1$. ■

Remark 5.5 (comparison MAP vs DRA when $f(x) = \frac{1}{p}|x|^p$) Suppose $f(x) = \frac{1}{p}|x|^p$, where $1 < p < +\infty$. According to Example 4.6, the MAP sequence $(x_n)_{n \in \mathbb{N}}$ exhibits logarithmic convergence to 0 and

$$(111) \quad x_n \sim \left(\frac{1}{n}\right)^{1/(2p-2)}.$$

On the other hand, Example 5.4 yields the following for the DRA sequence $(x_n)_{n \in \mathbb{N}}$:

p	convergence of the DRA sequence $(x_n)_{n \in \mathbb{N}}$ to 0
$1 < p < 2$	superlinear with order $\frac{1}{p-1}$
$p = 2$	linear with rate $\frac{1}{1+r_\infty}$
$2 < p < +\infty$	logarithmic and $x_n \sim \left(\frac{1}{n}\right)^{1/(p-2)}$

We conclude that *in all cases*, the DRA sequence converges to 0 *faster* than the MAP sequence⁸. To illustrate this, set $x_0 = 1$. Letting the parameter p range from 1 to 3, we show in Figure 1 the first 100 terms of the MAP sequence $(x_n)_{n \in \mathbb{N}}$ and of the DRA sequence $(x_n)_{n \in \mathbb{N}}$.

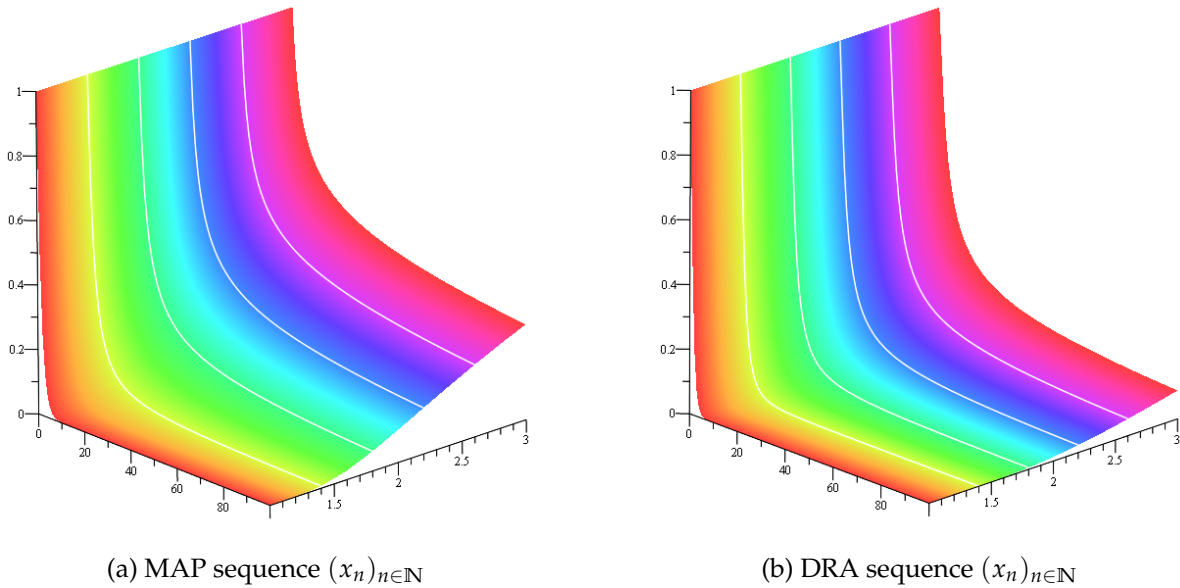


Figure 1: The distance of the first 100 terms to the solution for $p \in [1, 3]$

⁸It is interesting to note that DRA performs better than MAP when A and B are two subspaces with a small Friedrichs angle (see [2, Section 8]).

Although both sequences converge to 0, the solution, the stark contrast in their speed of convergence is shown in Figure 2 where we plot the quotient sequence of the MAP sequence divided by the DRA sequence. As predicted by the theory, the terms tend to $+\infty$ when $1 < p < 2$ illustrating the much faster convergence of the DRA sequence.

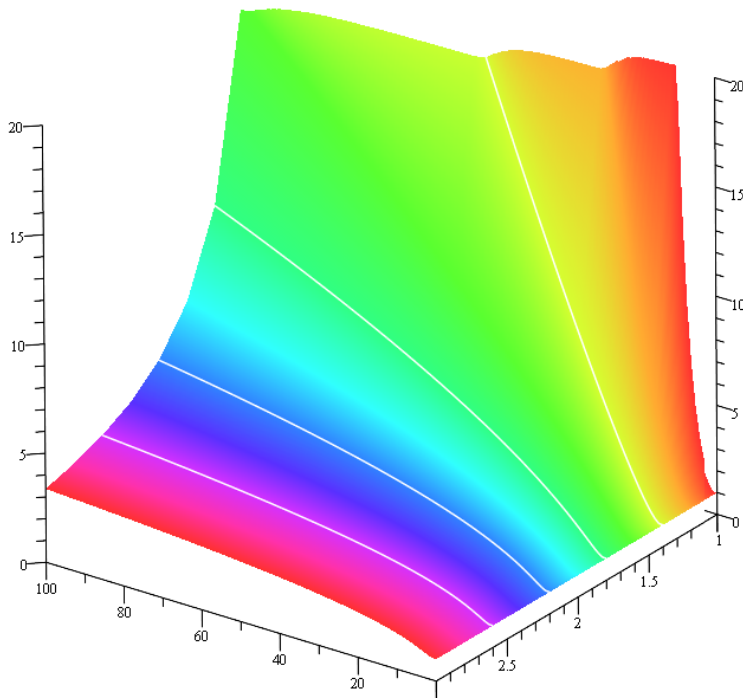


Figure 2: The MAP sequence divided by the DRA sequence

Example 5.6 (comparison MAP vs DR when $f(x) = R - \sqrt{R^2 - x^2}$)

Suppose that $R \in \mathbb{R}_{++}$ and that $f(x) = R - \sqrt{R^2 - x^2}$ on its domain $[-R, R]$. According to Example 4.7, the MAP sequence $(x_n)_{n \in \mathbb{N}}$ exhibits logarithmic convergence; in fact,

$$(112) \quad x_n \sim \frac{R}{\sqrt{n}}.$$

We now turn to the DRA sequence $(x_n)_{n \in \mathbb{N}}$. By (101b), we have for every $n \in \mathbb{N}$

$$(113) \quad x_n = x_{n+1} + \left(r_n + R - \sqrt{R^2 - x_{n+1}^2} \right) \frac{x_{n+1}}{\sqrt{R^2 - x_{n+1}^2}} = (r_n + R) \frac{x_{n+1}}{\sqrt{R^2 - x_{n+1}^2}};$$

consequently,

$$(114) \quad x_{n+1} = \frac{R x_n}{\sqrt{(r_n + R)^2 + x_n^2}}.$$

Since $f''(x) = R^2 / (R^2 - x^2)^{3/2}$, we have $f''(0) = 1/R \in \mathbb{R}_{++}$ and therefore, by (102), the DRA sequence

$$(115) \quad x_n \rightarrow 0 \text{ linearly}$$

with rate $1/(1 + r_\infty/R)$. Once again, the DRA sequence converges much faster than the MAP sequence!

Let us conclude. The results in this paper suggest that, for the convex feasibility problem, DRA outperforms MAP in cases of “bad geometry” (such as the absence of constraint qualifications or a “zero angle” between the constraints at the intersection). Since our proof techniques do not naturally generalize, it would be interesting to study these questions in higher-dimensional space and other classes of convex sets.

Acknowledgments

HHB was partially supported by the Natural Sciences and Engineering Research Council of Canada and by the Canada Research Chair Program. He also acknowledges the hospitality and the support of Université de Toulouse, France during the preparation of an early draft of this paper. MND was partially supported by an NSERC accelerator grant of HHB. HMP was partially supported by an internal grant of University of Massachusetts Lowell.

References

- [1] Y.I. Alber and S. Reich, An iterative method for solving a class of nonlinear operator equations in Banach spaces, *Panamerican Mathematical Journal* 4 (1994), 39–54.
- [2] H.H. Bauschke, J.Y. Bello Cruz, T.T.A. Nghia, H.M. Phan, and X. Wang, The rate of linear convergence of the Douglas–Rachford algorithm for subspaces is the cosine of the Friedrichs angle, *Journal of Approximation Theory* 185 (2014), 63–79.
- [3] H.H. Bauschke and P.L. Combettes, *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, Springer, 2011.
- [4] H.H. Bauschke, P.L. Combettes, and D.R. Luke, Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization, *Journal of the Optical Society of America* 19 (2002), 1334–1345.
- [5] H.H. Bauschke, D.R. Luke, H.M. Phan, and X. Wang, Restricted normal cones and the method of alternating projections: applications, *Set-Valued and Variational Analysis* 21 (2013), 475–501.
- [6] H.H. Bauschke, D. Noll, and H.M. Phan, Linear and strong convergence of algorithms involving averaged nonexpansive operators, *Journal of Mathematical Analysis and Applications* 421 (2015), 1–20.
- [7] J.M. Borwein, G. Li, and L. Yao, Analysis of the convergence rate for the cyclic projection algorithm applied to basic semialgebraic convex sets, *SIAM Journal on Optimization* 24 (2014), 498–527.
- [8] L.M. Bregman, The method of successive projection for finding a common point of convex sets, *Soviet Mathematics Doklady* 6 (1965), 688–692.

- [9] F. Deutsch and H. Hundal, Arbitrarily slow convergence sequences of linear operators: a survey, in *Fixed-Point Algorithms for Inverse Problems in Science and Engineering*, Springer Optimization and Its Applications 49, 213–242, Springer, 2011.
- [10] O. Güler, On the convergence of the proximal point algorithm for convex minimization, *SIAM Journal on Control and Optimization* 29 (1991), 403–419.
- [11] P.-L. Lions and B. Mercier, Splitting algorithms for the sum of two nonlinear operators, *SIAM Journal on Numerical Analysis* 16 (1979), 964–979.
- [12] B. Martinet, Régularisation d'inéquations variationnelles par approximations successives, *Rev. Française Informat. Recherche Opérationnelle* 4 (1970), 154–158.
- [13] J.-J. Moreau, Proximité et dualité dans un espace hilbertien, *Bulletin de la Société Mathématique de France* 93 (1965), 273–299.
- [14] H.M. Phan, Linear convergence of the Douglas–Rachford method for two closed sets, preprint 2014, <http://arxiv.org/abs/1401.6509>.
- [15] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, 1970.
- [16] R.T. Rockafellar, Monotone operators and the proximal point algorithm, *SIAM Journal on Control and Optimization* 14 (1976), 877–898.

Appendix A

Proof of Fact 2.1. (This proof is taken from <http://www.imomath.com/index.php?options=686> and included here for completeness as we were not able to locate a book or journal reference.) The second inequality is obvious. We only prove the right inequality since the proof of the left inequality is similar. Without loss of generality, we assume that $b_n \rightarrow +\infty$ and that $\lambda = \overline{\lim}_{n \rightarrow \infty} \frac{a_{n+1} - a_n}{b_{n+1} - b_n} < +\infty$. Let $\gamma \in]\lambda, +\infty[$. Then there exists $m \in \mathbb{N}$ such that $(\forall n \geq m) \frac{a_{n+1} - a_n}{b_{n+1} - b_n} < \gamma$, i.e., $a_{n+1} - a_n < \gamma(b_{n+1} - b_n)$. Let $n > m$. Then $a_n - a_m < \gamma(b_n - b_m)$ and hence

$$(116) \quad \frac{a_n}{b_n} - \frac{a_m}{b_m} < \gamma - \gamma \frac{b_m}{b_n}.$$

Taking $\overline{\lim}_{n \rightarrow \infty}$ yields

$$(117) \quad \overline{\lim}_{n \rightarrow \infty} \frac{a_n}{b_n} \leq \gamma.$$

Now let $\gamma \downarrow \lambda$ to complete the proof. ■

Appendix B

Proof of Fact 4.12. (This proof is a special case taken from [10] and included here for completeness.) Denote the minimizers of f by Z . Let $x \in H$, set $p = P_f x$, $p^* = x - p \in \partial f(p)$, and assume that $p \notin Z$. We have

$$(118) \quad (\forall y \in H) \quad f(y) - f(p) \geq \langle p^*, y - p \rangle.$$

Hence, for every $z \in Z$,

$$(119a) \quad f(z) - f(p) \geq \langle p^*, z - p \rangle = \langle p^*, z - x \rangle + \langle p^*, x - p \rangle$$

$$(119b) \quad \geq -\|p^*\| \|z - x\| + \|p^*\|^2 \geq -\|p^*\| \|z - x\|.$$

Since $f(p) \geq f(z)$, we learn that

$$(120) \quad \|p^*\| \geq \frac{f(p) - f(z)}{\|x - z\|} \geq 0.$$

Setting $y = x$ in (118), we have

$$(121) \quad f(x) - f(p) \geq \langle p^*, x - p \rangle = \|p^*\|^2.$$

Therefore,

$$(122) \quad (f(x) - f(z)) - (f(p) - f(z)) = f(x) - f(p) \geq \|p^*\|^2 \geq \left(\frac{f(p) - f(z)}{\|x - z\|} \right)^2.$$

It follows that

$$(123) \quad f(x) - f(z) \geq (f(p) - f(z)) \left(1 + \frac{f(p) - f(z)}{\|x - z\|^2} \right);$$

equivalently,

$$(124) \quad \frac{1}{f(x) - f(z)} \leq \frac{1}{f(p) - f(z)} \left(1 + \frac{f(p) - f(z)}{\|x - z\|^2} \right)^{-1}.$$

On the other hand, the definition of the proximal mapping yields

$$(125) \quad f(p) \leq f(p) + \frac{1}{2} \|x - p\|^2 \leq f(z) + \frac{1}{2} \|x - z\|^2,$$

which implies

$$(126) \quad \frac{f(p) - f(z)}{\|x - z\|^2} \leq \frac{1}{2}.$$

Consider the function $\alpha(t) = \frac{1}{1+t}$ with domain $]-1, +\infty[$. Clearly, α is convex, $\alpha(0) = 1$ and $\alpha(\frac{1}{2}) = \frac{2}{3}$. The line described by $t \mapsto 1 - \frac{2}{3}t$ goes through the same points and lies between these points above the graph of $\alpha(\cdot)$ (by convexity). Hence

$$(127) \quad (\forall 0 \leq t \leq \frac{1}{2}) \quad \frac{1}{1+t} \leq 1 - \frac{2}{3}t.$$

Altogether, we deduce that

$$(128a) \quad \varphi(x) = \frac{1}{f(x) - f(z)}$$

$$(128b) \quad \leq \frac{1}{f(p) - f(z)} \left(1 - \frac{2}{3} \cdot \frac{f(p) - f(z)}{\|x - z\|^2} \right)$$

$$(128c) \quad = \varphi(p) - \frac{2}{3\|x - z\|^2}.$$

Now assume that $z \leftarrow x_{n+1} = P_f x_n$. Then $\varphi(x_{n+1}) - \varphi(x_n) \geq \frac{2}{3} \|x_n - z\|^{-2} \rightarrow +\infty$. By Corollary 2.2, $\lim \frac{\varphi(x_n)}{n} = +\infty$. Now take reciprocals. \blacksquare