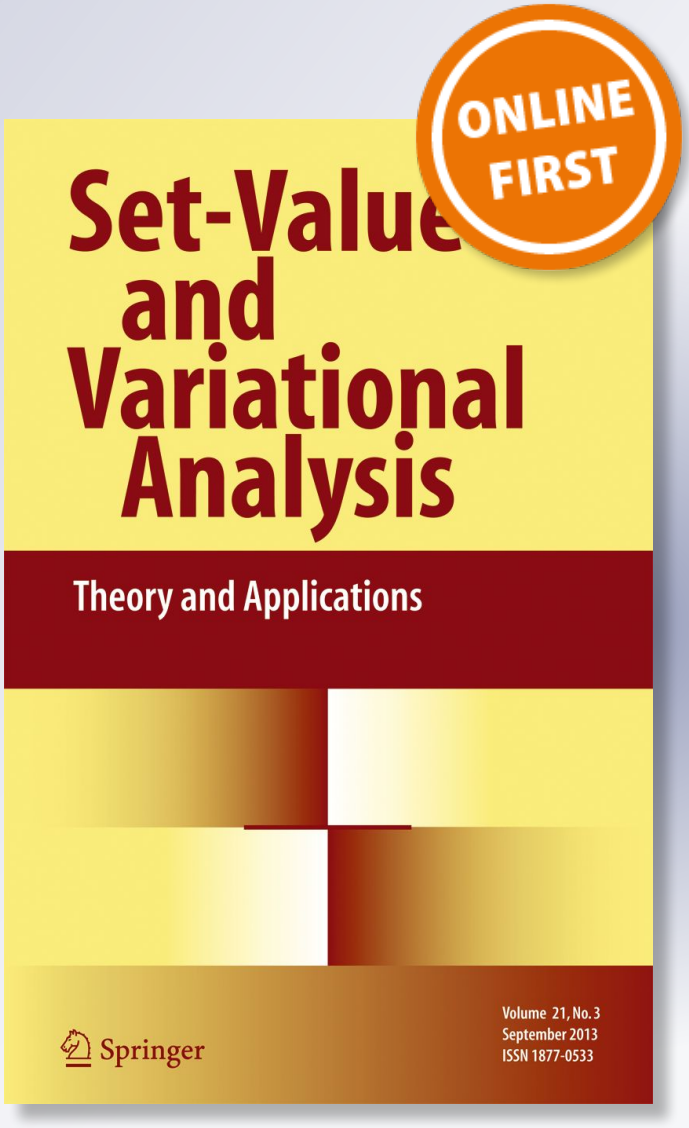
Dynamics of the Douglas-Rachford Method for Ellipses and p-Spheres

Jonathan M. Borwein, Scott B. Lindstrom, Brailey Sims, Anna Schneider & Matthew P. Skerritt

Set-Valued and Variational Analysis
Theory and Applications

ISSN 1877-0533

Set-Valued Var. Anal
DOI 10.1007/s11228-017-0457-0



Springer

Set-Valued Var. Anal
https://doi.org/10.1007/s11228-017-0457-0

Dynamics of the Douglas-Rachford Method for Ellipses and p-Spheres

Jonathan M. Borwein[1] · Scott B. Lindstrom[1] · Brailey Sims[1] · Anna Schneider[2] · Matthew P. Skerritt[1]

Received: 27 February 2017 / Accepted: 5 October 2017

Abstract We expand upon previous work that examined the behavior of the iterated Douglas-Rachford method for a line and a circle by considering two generalizations: that of a line and an ellipse and that of a line together with a p-sphere. With computer assistance we discover a beautiful geometry that illustrates phenomena which may affect the behavior of the iterates by slowing or inhibiting convergence for feasible cases. We prove local convergence near feasible points, and—seeking a better understanding of the behavior—we employ parallelization in order to study behavior graphically. Motivated by the computer-assisted discoveries, we prove a result about behavior of the method in infeasible cases.

Keywords Douglas-Rachford · Feasibility · Projection algorithms · Iterative methods · Discrete dynamical systems

Mathematics Subject Classification (2010) 47H99 · 49M30 · 65Q30 · 90C26

This work is dedicated to the memory of Jonathan M. Borwein our greatly missed friend, mentor, and colleague. His early contributions to this paper, like his contributions to so much else, were both vital and inspirational.

✉ Matthew P. Skerritt
matthew.skerritt@uon.edu.au

Scott B. Lindstrom
scott.lindstrom@uon.edu.au

Brailey Sims
brailey.sims@newcastle.edu.au

[1] CARMA, University of Newcastle, Newcastle NSW, Australia

[2] Universität der Bundeswehr München, Neubiberg, Germany

Published online: 23 October 2017

1 Introduction and Preliminaries

The Douglas-Rachford algorithm [15] was introduced over half a century ago in connection with nonlinear heat flow problems to find a feasible point (point in the intersection) of two closed constraint sets A and B in a Hilbert space H.

We will denote the induced norm by $\|\cdot\|$. The projection onto a proximal subset C of H is defined for all $x \in H$ by

$$P_C(x) := \left\{ z \in C : \|x - z\| = \inf_{z' \in C} \|x - z'\| \right\}$$

When C is closed and convex the projection operator P_C is single valued and firmly non-expansive. When C is a closed subspace it is also linear and self-adjoint. For additional information, see, for example, [6, Definition 3.7]. The reflection mapping through the set C is then defined by

$$R_C := 2P_C - I,$$

where I is the identity map on H.

Definition 1 (Douglas-Rachford Method) For two closed sets A and B, and an initial point $x_0 \in H$, the Douglas-Rachford method generates a sequence $(x_n)_{n=1}^{\infty}$ as follows:

$$x_{n+1} \in T_{A,B}(x_n) \quad \text{where} \quad T_{A,B} := \frac{1}{2}\left(I + R_B R_A\right). \tag{1}$$

Figure 1 illustrates the construction of one iteration of the Douglas-Rachford method.

Remark 1 (Notation) Throughout, x_n, x_0 are as in Definition 1, A, B are closed. When the two sets A and B are clear from the context we will simply write T in place of $T_{A,B}$.

Theorem 1 (Bauschke, Combettes, and Luke [7], see also [17]) *Suppose $A, B \subseteq H$ are closed and convex with non-empty intersection. Given $x_0 \in H$ the sequence of iterates $T_{A,B}$ converges weakly to an $x \in \operatorname{Fix} T_{A,B}$ with $P_A(x) \in A \cap B$.*

In finite dimensions convergence in norm for convex sets is therefore assured. Notwithstanding the absence of a satisfactory theoretical justification, the Douglas-Rachford iteration scheme has been used to successfully solve a wide variety of practical problems in which one or both of the constraints are non-convex. Phase retrieval problems are one important

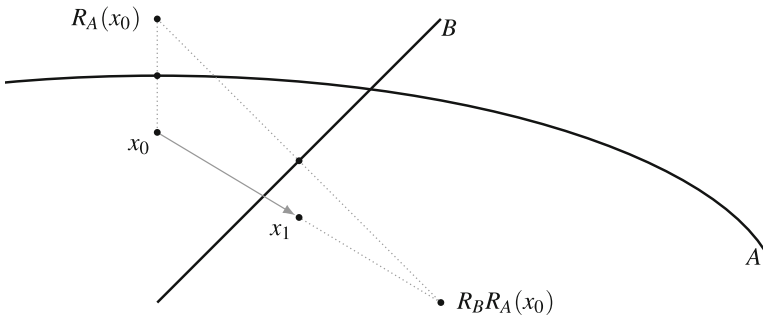


Fig. 1 One iteration of the Douglas-Rachford method

instance, and the case of a line L and circle C in 2-dimensional Euclidean space—prototypical of such problems—was investigated by Borwein and Sims [9] as a specific case of the higher dimensional problem of a line and a sphere in Hilbert space.

Despite the seeming simplicity of the situation, the Douglas-Rachford method applied to L and C proved surprisingly difficult to analyze. Among the partial results obtained in the feasible case was local convergence to each of the two feasible points. Based on this and extensive computer experimentation, Borwein and Sims were led to ask whether this could be extended to convergence to one or other of the two intersection points for all starting points except those lying on a "singular set" S_0: the line of symmetry perpendicular to L and passing through the center of C. Borwein and Aragón Artacho [3] established sizable domains of attraction for each of the feasible points, and the global question was answered in the affirmative by Benoist [11] who obtained the result by constructing a suitable Lyapunov function, see Fig. 2.

The singular set S_0 is invariant under the Douglas-Rachford operator $T_{C,L}$ and contains period 2 points if and only if L passes through the center of C in which case all the points of L inside C are period 2 points. When L is tangential to C all points on S_0 are fixed by $T_{C,L}$; for other positions of L the iterates exhibits periodic behaviors when rational commensurability is present, while in the absence of such commensurability the behaviors may be quite chaotic. See [9] for more details.

In order to gain further insights into the behavior of the Douglas-Rachford algorithm in the case of nonconvex constraint sets we consider two generalizations of a line and sphere (circle) in 2 dimensional Euclidean space, namely: that of a line together with an ellipse and that of a line together with a p-sphere.

These seemingly innocuous generalizations, while open to exploration and local analysis about the feasible points, may be impossible to analyze in full. The singular set is no longer a simple curve but rather exhibits a complex (and fascinating) geometry involving a rich array of periodic points and associated domains of attraction. In the case of an ellipse and line we observe the appearance of higher order periodic points as the ellipticity is increased.

Definition 2 (Periodic Points and Domains) The following terms, already used informally, help inform our discussion.

1. A point x is a *periodic point of period m* (or a *period m point*) if $T_{A,B}^m x = x$ (A period 1 point is simply a fixed point of $T_{A,B}$).

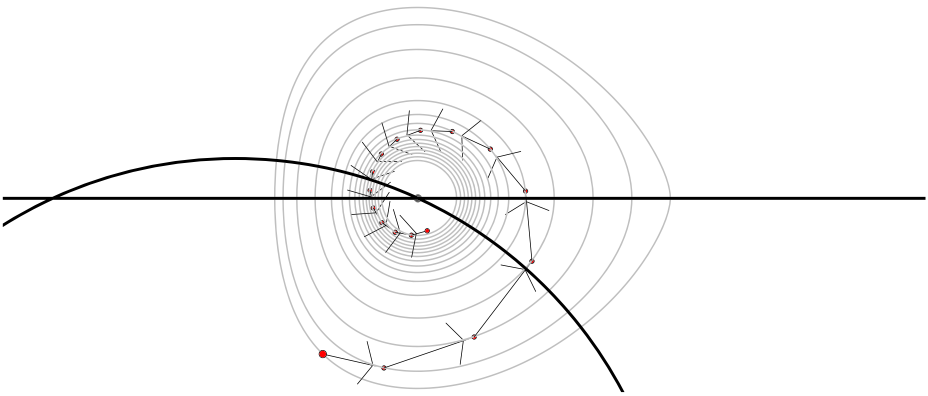


Fig. 2 Douglas-Rachford on the 2-sphere and line showing the level sets of the Lyapunov function [11]

2. The *domain of attraction* (or *attractive domain*) for a period m point x is the set of all x_0 satisfying
$$\lim_{k\to\infty} T_{A,B}^{km}(x_0) = x. \tag{2}$$
3. A point x is *attractive* if its domain of attraction contains a neighborhood of x.
4. The *singular set* consists of all points not belonging to a domain of attraction for any feasible point.
5. A period m point is said to be *repelling* if there exists a neighborhood N_x of x such that for every $x_0 \in N_x \backslash \{x\}$ the sequence $(T_{A,B}^{km}(x_0))_{k=1}^{\infty}$ eventually lies outside of N_x.

Of course if S is a domain of attraction for a period m point x then for $k = 1, 2, \cdots m-1$ it follows that $T^k(S)$ is a domain of attraction for $T^k(x)$. This is a notable feature in many of our graphics, see for instance Fig. 3.

1.1 Notation

By a suitable rotation and scaling of axes we may without loss of generality take our ellipse and p-sphere respectively to be

$$\begin{aligned} E_b &:= \left\{(x, y) \in \mathbb{R}^2 |\ \varphi_b(x, y) := x^2 + \left(\frac{y}{b}\right)^2 = 1\right\} \quad \text{and} \\ S_p &:= \left\{(x, y) \in \mathbb{R}^2 |\ \theta_p(x, y) := (x)^p + (y)^p = 1\right\}, \end{aligned} \tag{3}$$

and will write:

$$L_{m,\beta} := \{(x, y) \in \mathbb{R}^2 | y = mx + \beta\} \quad \text{and} \quad L_m := L_{m,0}.$$

When it is clear from the context what the parameters are we will simply write E, S or L respectively. Similarly, when the context makes it clear we will write T in place of $T_{E,L}$ or $T_{S,L}$.

1.2 Computation of Projections

For the case of the 2-sphere, the closest point projection has a simple closed form. For $x \neq 0$, $P_S(x) = x/\|x\|$. Such a simple closed form is immediately lost for any ellipse with

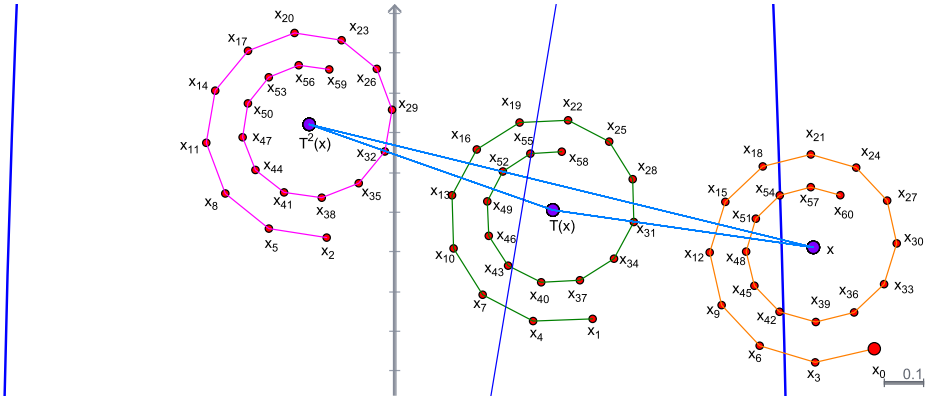


Fig. 3 Domain of attraction for period 3 points in the case of E_6, L_8

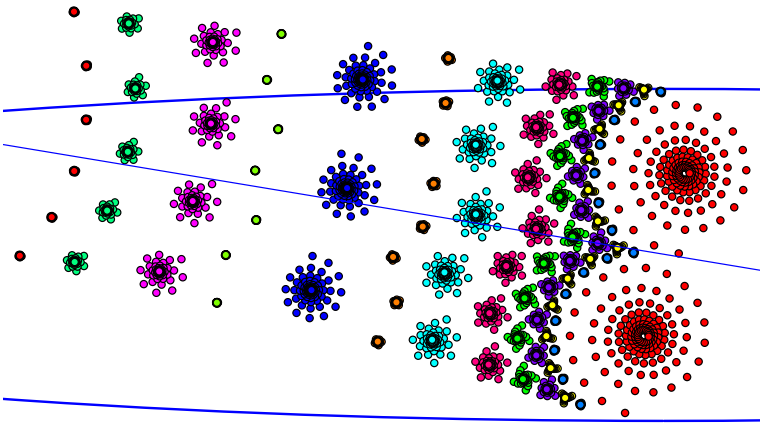


Fig. 4 Partial domains of attraction of T_{E_8,L_6} for points of different periodicities

$b \neq 1$ or any p-Sphere with $p \notin \{1, 2\}$ because—where φ_b, θ_p are as in (3)—the induced Lagrangian problems

$$P_{E_b}(x) = \left\{ x' \mid \lambda \nabla d(x, \cdot)^2(x') = \nabla \varphi_b(x'), \quad \varphi_b(x') = 1 \right\}$$
$$P_{S_p}(x) = \left\{ x' \mid \lambda \nabla d(x, \cdot)^2(x') = \nabla \theta_p(x'), \quad \theta_p(x') = 1 \right\}$$

yield implicit relations that no longer admit explicit solutions. Computation of the required projections necessitates the use of numerical methods.

A description of the optimized function solvers used is available in the appendix [18]. Many of our implementations of these function solvers—for example, that used to generate Fig. 3—employ the interactive geometry software *Cinderella*, available at https://cinderella.de/.

Table 1 Periods observed for attractive domains for various ellipse and line configurations

	E_2	E_3	E_4	E_5	E_6
L_1	2	2	2	2	2
L_2		2,3	2,3	2,3	2,3
L_3		2,3	2,3,5	2,3,4,5,7	2,3,4,5,7
L_4			2,3,5,7	2,3,4,5,7,9	2,3,4,5(×2),7,9
L_5			2,3	2,3,4,5,7,9,11,13	2,3,4,5,7,9,11,13
L_6				2,3,5,7	2,4,5(×2),7,9,11,13,15
L_7				3	2,3,4,5,7,9,11,13
L_8					2,3,5

Some periodicities were observed in more than one domain

2 The Case of an Ellipse and a Line

In the case of an ellipse and a line, the singular set—in contrast to the case of a circle and a line—is no longer a simple curve, and appears to contain periodic points in many cases. For example, Fig. 4 shows periodic points for T_{E_8,L_6} with attendant subsets of their attractive domains. The singular set is larger than suggested here (see Fig. 10 for a more complete depiction).

For simplicity, we set the line intercept at 0 for our pictures and tables. It should be noted that, in contra distinction to the case of a circle and line, similar behavior can be observed with nonzero intercepts (although symmetries are lost).

The number and periodicity of the points appear to be related to both the eccentricity of the ellipse, and the angle of the line. As the eccentricity is increased, we observe growth in both the number of periodic points and the maximum periodicity. Table 1, obtained experimentally using *Cinderella*, summarizes our findings. Note that the method used required interactively moving a point in the geometry package, and visually observing the attractive domains. As such it is regrettably possible that some periodic points were missed, either because their domains of attraction were too small or they were not attractive points.

We can describe the period 2 points of T_{E_b,L_m} with a closed form that, while complicated to state, is quick to evaluate [18]. Determining period 2 points algebraically is useful for corroborating some of the behaviors we observe in *Cinderella*. However, the degree of complication associated with the analysis of even this simplest case of a non-fixed periodic point suggests that fully describing all behavior globally with explicit forms would be an impractical undertaking. This, in part, led us to pursue the computer assisted evidence-gathering approach we describe in Section 2.1.

The nature of the periodic points is also sensitive to small perturbations of the line. We can see above that for lines of small slope there are only a few attractive periodic points, and as the slope increases additional points with higher periodicity emerge. As the slope becomes large, some of the attractive domains appear to shrink in size until eventually the associated periodic point ceases to be attractive. This appears to be the eventual fate of all periodic points.

This sensitivity to perturbations can be seen in Figs. 5 and 6. In the former we have connected every second iterate, and see how a small change in slope can affect which feasible point is converged to, as well as the appearance/disappearance of attractive domains. In the latter we plot every second iterate for T_{E_2,L_m} with 300 iterates, starting at $m = 2$ (top left) we slowly rotate the line until we have $m = 3/2$ (bottom right). Part of the line is visible in

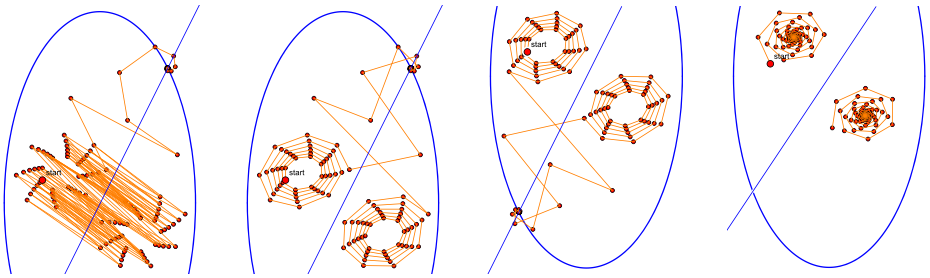


Fig. 5 Sensitivity of behaviors to small changes in line slope

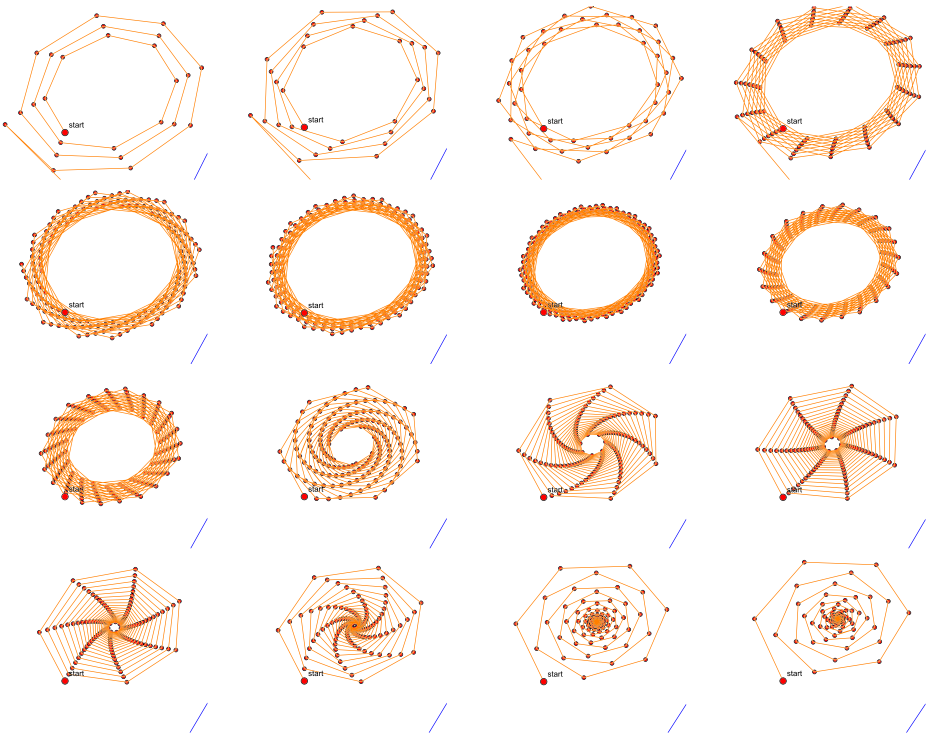


Fig. 6 Evolution in behaviors near two period 2 points as the line slope is changed

the bottom right corner of each frame. In the initial configuration, the subsequence of iterates started near to the periodic point are repelled from it. As we rotate the line, we see that the "speed" at which they are repelled decreases until eventually the periodic point becomes an attractive point instead of a repelling point.

2.1 Studying Convergence: Numerical Motivations

The complicated nature of the singular set precludes any possibility of constructing a Lyapunov function in any sizable region about the feasible point. Indeed, attempts to even numerically construct the level curves such a function might have near a feasible point proved unstable. Instead we refine our numerical-graphical method of discovery. The method we used for Fig. 4, though useful for discovery, is not, in itself, sufficient for fully understanding the behaviors, even for one specific ellipse and line. There are several reasons for this.

1. There may be other periodic points we cannot see because they are repelling or their attractive domains are too small.
2. The potential for numerical error is accentuated by the fact that the projection onto the ellipse is specified as the root function—induced by the Lagrangian system—whose calculation is, in some configurations, complicated by the presence of nearby incorrect roots.

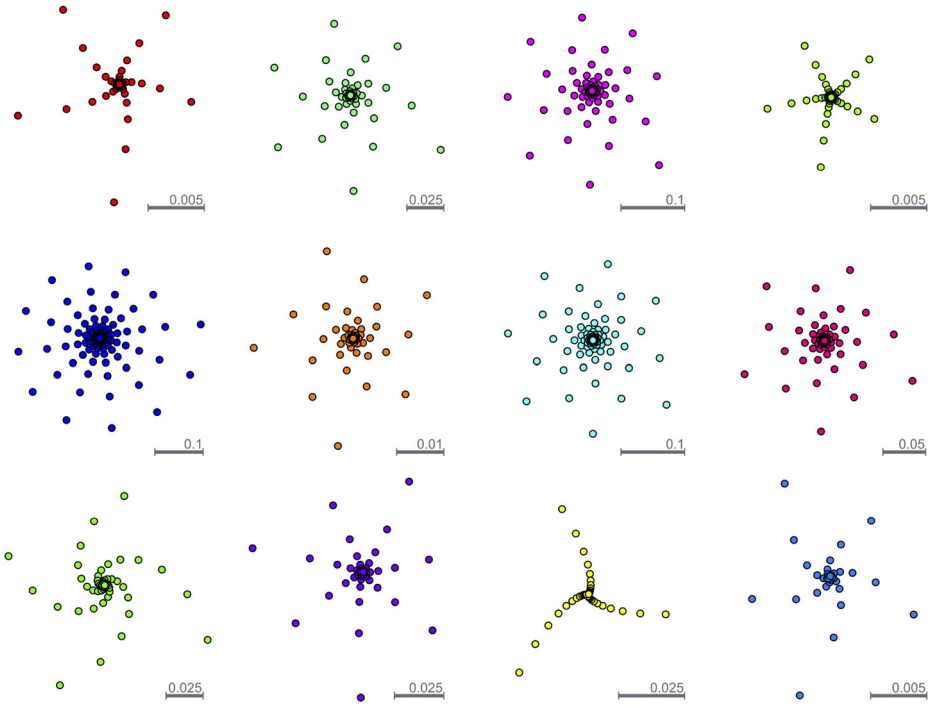


Fig. 7 Close up view of the spirals in seen in Fig. 4

3. This method of visualization may be deceptive, as it precludes us from seeing accurately the extent and shape of attractive domains.

As an example of the latter, notice how the patterns of the iterates in Fig. 4 form orderly spirals. This lovely pattern seems to hold for all the cases we have looked at. If we zoom in on the spirals we see what look like twisting galaxies (see Fig. 7). Intuition would suggest to us—incorrectly—that this is perhaps indicative of smooth boundaries for the domains.

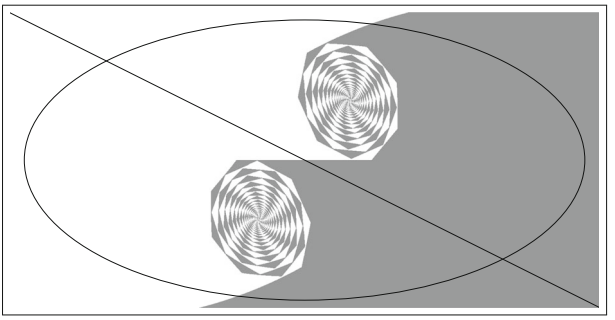


Fig. 8 Domains of attraction for the two feasible points of T_{E_2,L_2}

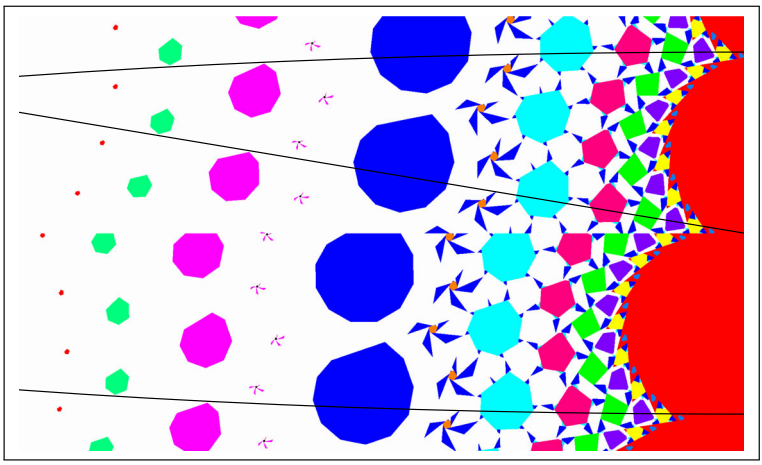


Fig. 9 Domains of attraction for T_{E_8,L_6}. Compare with Fig. 4

2.2 Visualization Through Parallelization

Seeking clearer pictures with finer resolution, we implemented a new version of our code. In this new version we specify a resolution and for each pixel we compute the midpoint, calling it x_0. Once computed, the location of $x_{1,000}$ is checked against a list containing the feasible points and approximate periodic points, and the pixel is colored according to which list member $x_{1,000}$ is nearest to.

The efficacy of this technique is demonstrated in Figs. 8 and 9. The former clearly depicts the complex structure of the domains of attraction for the two feasible points in the case

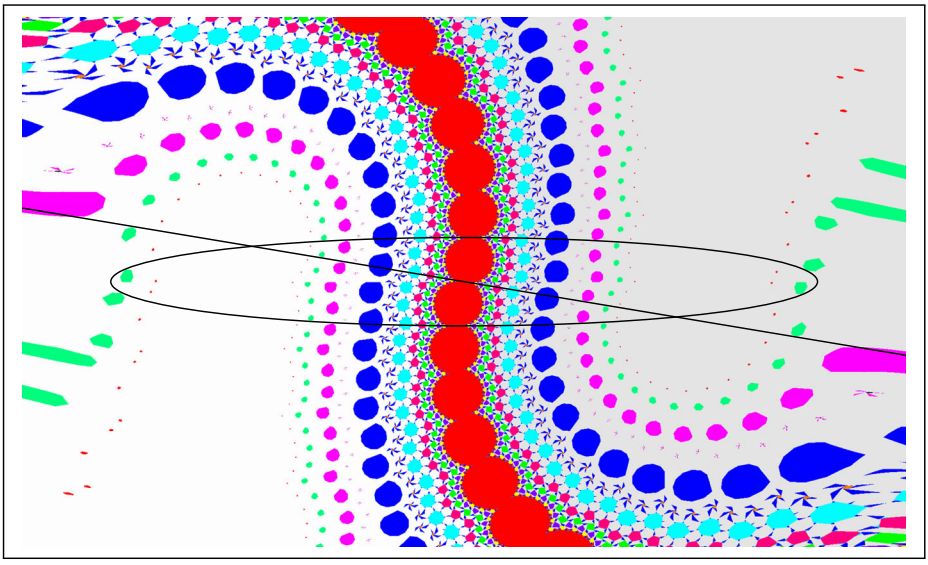


Fig. 10 Domains of attraction for T_{E_8,L_6}

of T_{L_2,E_2} where two period 2 repelling points are present. Note the interweaving of the attractive domains near the repelling points. The latter shows the domains of attraction for T_{E_8,L_6} in the same region as shown in Fig. 4.

This new implementation was written in such a way as to leverage the highly parallel nature of GPU devices, although it may also be run on regular CPUs. This allowed us to compute the colorings for many pixels simultaneously, reducing the time needed to produce the images and simultaneously affording us the ability to produce images with a much finer resolution. With this method we were able to see the behavior of the system over a larger area, as shown in Fig. 10. Note that these images benefit greatly from color coding of the domains. Color versions of the above figures and other images we produced can be found in the appendix [18].

3 Line and p-Sphere

Projections onto the 1-sphere can be determined explicitly, so exact analysis is possible. Consequently, much of the behavior is readily determined in particular periodic points with periods higher than 2 are observed. When $p = 2$, we recover the circle for which the convergence properties are known, (see Section 1). Our observations from the case $p > 2$ suggest a conjecture: that when the line is not parallel to either of the axes there is at most one pair of periodic points and they are repelling.

For $1/n$-spheres where $n \geq 2$ is a natural number, we see the appearance of period 2 points with attendant local domains of attraction, examples are shown in Fig. 11. It can be seen—proven easily—that, for the sphere $S_{1/n}$ with line L_1, any point $(-t, t)$ or $(t, -t)$ for $t \in (0, \frac{1}{2^n}]$ is a period 2 point. This continuum of period 2 points is analogous to what is observed in the case of a 2-sphere. More interesting is the apparent emergence of attractive domains which have nonzero measure.

These observations already hint at the larger measure and greater complexity of the singular manifold in the case of a line and p-sphere, when $p \neq 2$, compared to that for a line and 2-sphere. When we rotate the line to, say, $L_{1/2}$ there appear to be only finitely many period 2 points, but they are no longer constrained to lie in an affine submanifold.

4 A Theoretical Interlude: Local Convergence to a Feasible Point

Borwein and Sims [9] used the Perron theorem on the stability of almost linear difference equations [16, Corollary 4.7.2] to establish local convergence of the Douglas-Rachford algorithm, $x_{n+1} = T_{K,L}(x_n)$, to an isolated point $f \in L \cap K$ when L is a line and K is the (non

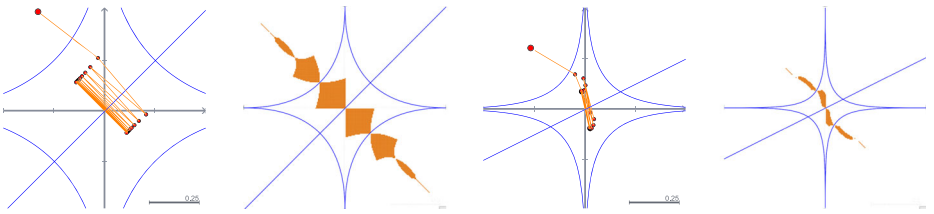


Fig. 11 Subsequential convergence to—and attractive domains for—period 2 points. Left: $T_{S_{1/2},L_1}$, right: $T_{S_{1/3},L_{1/2}}$

convex) unit sphere in n-dimensional Euclidean space. We outline a strategy for extending this to the case when L is still a line, but K is a smooth hypersurface ($(n-1)$-manifold). We consider its application when L and K lie in $\mathbb{R}^2$; L is the line $\alpha x + \beta y = \gamma$, K is the ellipse E_b as in (3).

The strategy is to show that, in a neighborhood of the feasible point f, the reflection in the supporting hyper-plane H_f to K at f—as in Fig. 12—provides an o-order approximation to the reflection in K so that the Perron theorem can be applied to the system of difference equations corresponding to the Douglas-Rachford algorithm. Succinctly, we want

$$R_K(p) = R_{H_f}(p) \; + \; \Delta, \quad \text{where } \|\Delta\| = o(\|p - f\|) \text{ for } p \text{ sufficiently near } f.$$

For the Euclidean reflection this follows if $\|P_K(p) - P_{H_f}(p)\| = o(\|p - f\|)$. When this happens we have, for p in a neighborhood of f,

$$\begin{aligned}
T_{K,L}(p) \; &= \; \frac{1}{2}\left[p + R_L\left(R_K(p)\right)\right] \\
&= \frac{1}{2}\left[p + R_L\left(R_{H_f}(p) + \Delta\right)\right] \\
&= \frac{1}{2}\left[p + R_{L-f}\left(R_{H_f}(p) + \Delta - f\right) + f\right] \\
&= \frac{1}{2}\left[p + R_{L-f}\left(\left(R_{H_f-f}(p-f) + f\right) + \Delta - f\right) + f\right] \\
&= \frac{1}{2}\left[p + R_{L-f}\left(R_{H_f-f}(p-f)\right) + R_{L-f}(\Delta) + f\right], \qquad \text{since } R_{L-f} \text{ is linear} \\
&= \frac{1}{2}\left[(p-f) + R_{L-f}\left(R_{H_f-f}(p-f)\right)\right] + \frac{1}{2}R_{L-f}(\Delta) + f
\end{aligned}$$

Thus we have that

$$T_{K,L}(p) \; = \; f \; + \; T_{(H_f-f),(L-f)}(p-f) + \Delta'$$

where $\Delta' = \frac{1}{2}R_{L-f}(\Delta)$ has $\|\Delta'\| = o\|p - f\|$ since R_{L-f} is a bounded linear operator.

Thus, by the theorem of Perron (see, [9] theorem 6.1 or [16] Corollary 4.7.2), the system of difference equations corresponding to the Douglas-Rachford algorithm for K and L,

$$x_{n+1} \; = \; T_{K,L}(x_n)$$

is exponentially asymptotically stable at f (in particular $\|x_n - f\| \to 0$ for x_0 sufficiently near f) provided all the eigenvalues of the linear operator $T_{(H_f-f),(L-f)}$ have moduli less than one.

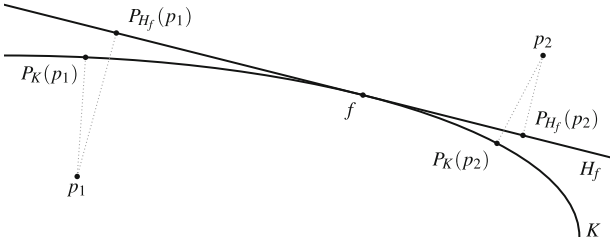


Fig. 12 Approximation of P_K by P_{H_f} near f

Remark 2 When M is a subspace, the projection P_M is linear (as is the case when $M = L - f$) and the Douglas-Rachford operator for N and M becomes

$$T_{N,M} = \frac{1}{2}[I + (2P_M - I)(2P_N - I)] \tag{4}$$
$$= 2P_M P_N - P_M - P_N + I \tag{5}$$
$$= P_M P_N + (I - P_M)(I - P_N). \tag{6}$$

When N is also a subspace (for instance when $N = H_f - f$) this may be written as

$$T_{NM} = P_M P_N + P_{M^\perp} P_{N^\perp}$$

where $^\perp$ denotes the orthogonal complement.

As a curiosity, we observe that if in the case of two subspaces we define a *twisted Douglas-Rachford operator* by $V_{NM} := P_M P_N + P_{N^\perp} P_{M^\perp}$, then, since $P_M P_{M^\perp} = P_N P_{N^\perp} = 0$, the iterates are $x_n = V_{NM}^n(x_0) = u_n + v_n$, where $u_{n+1} = P_M P_N(u_n)$ and $v_{n+1} = P_{N^\perp} P_{M^\perp}(v_n)$. The sequence of twisted Douglas-Rachford approximants is thus the sum of two sequences (u_n) and (v_n) resulting from the application of von Neumann's alternating projection algorithm to the pairs of subspaces M and N, and $N^\perp$ and $M^\perp$ respectively. Since the Friedrichs angle θ between M and N is the same as the angle between $M^\perp$ and $N^\perp$, the twisted Douglas-Rachford algorithm converges with the same rate as the von Neumann algorithm; namely at a linear rate proportional to $\cos^2\theta$ [13], the same as the rate exhibited by the standard Douglas-Rachford algorithm [5].

Now we consider the special case of an ellipse and a line. Without loss of generality we consider the ellipse E_b as in (3), and the line $L : \alpha x + \beta y = \gamma$, where $b \geq 1$, $\beta \geq 0$ and, to ensure the existence of $f = (x_0, y_0) \in L \cap E \cap |\mathbb{R}|^2$, either $\alpha \leq \gamma \leq \beta b$ or $\beta b < \gamma$ and $\alpha \geq \sqrt{\gamma^2 - \beta^2 b^2}$.

Following the strategy outlined above leads us to consider the Eigenvalues of T for two lines through the origin

$$L_1 : \alpha x + \beta y = 0 \quad \text{and} \quad L_2 : Ax + By = 0$$

where, in our context, the latter line, being parallel to the tangent to E at f, has $A = x_0$ and $B = y_0/b^2$. It is readily verified that the orthogonal projection onto L_1 has the matrix

$$[P_{L_1}] = \frac{1}{\alpha^2 + \beta^2} \begin{pmatrix} \beta^2 & -\alpha\beta \\ -\alpha\beta & \alpha^2 \end{pmatrix} \tag{7}$$

with a matching expression for $[P_{L_2}]$. Substituting these expressions into $T_{L_2,L_1} = 2P_{L_1}P_{L_2} - P_{L_1} - P_{L_2} + I$ yields

$$[T_{L_1,L_2}] = \frac{\psi}{\Delta} \begin{pmatrix} \psi & \omega \\ -\omega & \psi \end{pmatrix}$$

where $\psi = \alpha A + \beta B$, $\omega = \alpha B - \beta A$ and $\Delta = (\alpha^2 + \beta^2)(A^2 + B^2)$, which has eigenvalues $\frac{\psi}{\Delta}(\psi \pm i\omega)$ with modulus squared equal to

$$\frac{\psi^2}{\Delta^2}\left(\psi^2 + \omega^2\right) = \frac{(\alpha A + \beta B)^2\left((\alpha A + \beta B)^2 + (\alpha B - \beta A)^2\right)}{\left(\alpha^2 + \beta^2\right)^2\left(A^2 + B^2\right)^2}$$
$$= \frac{(\alpha A + \beta B)^2}{\left(\alpha^2 + \beta^2\right)\left(A^2 + B^2\right)} < 1$$

Thus, as expected, for any two lines intersecting in a single point the Douglas-Rachford algorithm with any starting point spirals exponentially to their common point.

Therefore the Douglas-Rachford algorithm for a line and an ellipse E is locally convergent at each of the feasible points f provided

$$\|P_E(p) - P_{H_f}(p)\| = o(\|p - f\|),$$

for all p in some neighborhood of f.

To see this we follow an argument suggested by Asen Dontchev [14]. While we present the argument in the particular case of $E_b = \{x|\varphi_b(x)-1 = 0\}$, the astute reader will observe that it applies to any smooth hypersurface $K := \{g(x) = 0\}$ at any point $f = (x_0, y_0) \in K$ at which the gradient ∇g is non-singular (true for the ellipse as $\nabla g(x) = (2x, 2y/b^2)$) and so applies to p-spheres except near the extreme points of the sphere when $0 < p \leq 1$.

We begin by noting that for the supporting hyperplane (tangent) to E at f

$$P_{H_f}(p) = f + P_{H_f - f}(p - f)$$

where

$$\left[P_{H_f - f}\right] = \frac{b^4}{b^4 x_0^2 + y_0^2} \begin{pmatrix} y_0^2/b^4 & -x_0 y_0/b^2 \\ -x_0 y_0/b^2 & x_0^2 \end{pmatrix}.$$

Next we observe that the nearest point projection $(u(p), v(p)) = P_E(p)$ at $p = (\zeta, \eta)$ is the solution of

$$\begin{aligned} \text{minimize:} \quad & \frac{1}{2}\|P_E(p) - p\|^2 = \frac{1}{2}\left((u-\zeta)^2 + (v-\eta)^2\right) \\ \text{subject to:} \quad & g\left(P_E(p)\right) = u^2 + \left(\frac{v}{b}\right)^2 - 1 = 0, \end{aligned}$$

which, since $\nabla g\left(P_E(p)\right)$ is non-singular, is characterized via the method of Lagrange multipliers by $\nabla \frac{1}{2}\left((u-\zeta)^2 + (v-\eta)^2\right) + \lambda \nabla g\left(P_E(p)\right) = 0$ together with $g\left(P_E(p)\right) = 0$, that is,

$$\begin{aligned} f_1 &: u - \zeta + 2\lambda u = 0 \\ f_2 &: v - \eta + 2\lambda \frac{v}{b^2} = 0 \\ g &: u^2 + \left(\frac{v}{b}\right)^2 - 1 = 0. \end{aligned}$$

As an aside, this yields the implicit specification

$$P_E(\zeta, \eta) = \left(\frac{\zeta}{1 + 2\lambda}, \frac{b^2 \eta}{b^2 + 2\lambda}\right), \quad \text{where } \frac{\zeta^2}{(1+2\lambda)^2} + \frac{b^2 \eta^2}{\left(b^2 + 2\lambda\right)^2} = 1.$$

In order to apply the implicit function theorem to ensure that u, v and λ are differentiable functions of ζ and η in a neighborhood of f we require the Jacobian of the above system of equations with respect to the dependent variables, u, v and λ at f, $J(f)$, to be non-singular. Since $P_E(f) = f$ we see from the first (and second) equation that for $p = f$ the corresponding Lagrange multiplier is necessarily 0. Thus,

$$J(f) = \left.\frac{\partial(f_1, f_2, g)}{\partial(u, v, \lambda)}\right|_{(x_0, y_0, 0)} = \begin{pmatrix} 1 & 0 & 2x_0 \\ 0 & 1 & 2y_0/b^2 \\ 2x_0 & 2y_0/b^2 & 0 \end{pmatrix},$$

more generally $J(f) = \begin{pmatrix} I & \nabla g(f)^T \\ \nabla g(f) & 0 \end{pmatrix}$, which is indeed non-singular, and in our case

$$J(f)^{-1} = \frac{b^4}{b^4 x_0^2 + y_0^2} \begin{pmatrix} y_0^2/b^4 & -x_0 y_0/b^2 & x_0/2 \\ -x_0 y_0/b^2 & x_0^2 & y_0/2b^2 \\ x_0/2 & y_0/2b^2 & -1/4 \end{pmatrix}.$$

Thus the implicit function theorem applies, yielding

$$\begin{pmatrix} [P_E'(f)] \\ [\lambda'(f)] \end{pmatrix} = \frac{\partial(u, v, \lambda)}{\partial(\zeta, \eta)}\bigg|_{(x_0, y_0))} = J(f)^{-1} \frac{\partial(f_1, f_2, g)}{\partial(\zeta, \eta)} = J(f)^{-1} \begin{pmatrix} I \\ 0 \end{pmatrix},$$

whence $[P_E'(f)] = \frac{b^4}{b^4 x_0^2 + y_0^2} \begin{pmatrix} y_0^2/b^4 & -x_0 y_0/b^2 \\ -x_0 y_0/b^2 & x_0^2 \end{pmatrix}$ which we recognize as $\left[P_{H_f}'(f)\right]$ and we are able to conclude that near f

$$\|P_E(p) - P_{H_f}(p)\| = \|f + P_E'(f)(p - f) + \Delta - \left(f + P_{H_f - f}(p - f)\right)\| = \|\Delta\|$$

where $\|\Delta\| = o(\|p - f\|)$ as required.

Summarizing the above, we have the following local convergence result.

Theorem 2 *Let K be a hypersurface in $\mathbb{R}^n$ and $f \in K$ such that K is smooth in a neighborhood of f and let H_f be the unique supporting hyperplane to K at f. If L is a line such that $L \cap H_f = \{f\}$, then the Douglas-Rachford algorithm $T_{K,L}$ is locally convergent to f.*

5 Important Lessons About Global Behavior

What we have observed in our computer-assisted study of these two simple cases of a line together with an ellipse or a p-sphere is remarkably informative: it suggests likely explanations for the behavior of the algorithm both for feasible and infeasible cases. We consider feasible cases first.

5.1 The Feasible Case

Aragón Artacho, Borwein, and Tam experimented with using the Douglas-Rachford method to solve Sudoku puzzles [1]. The left hand images of Figs. 13 and 14, illustrate the distance to the solution by iterations of Douglas-Rachford for two different sudoku puzzles.

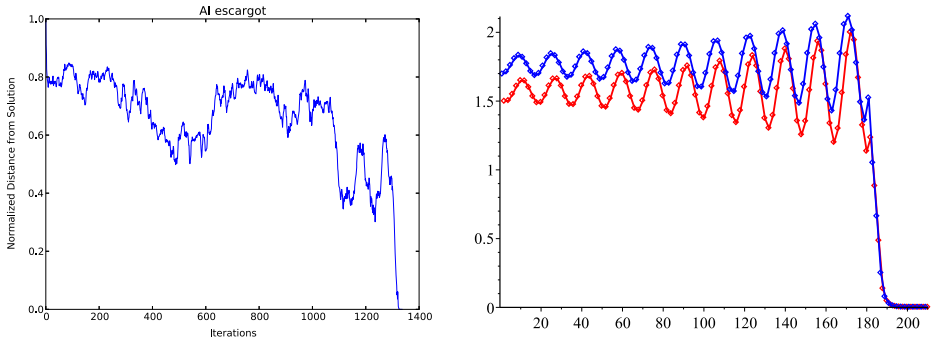


Fig. 13 Distance from iterates to solution. Left: for a sudoku puzzle [1]. Right: for T_{E_2,L_2}

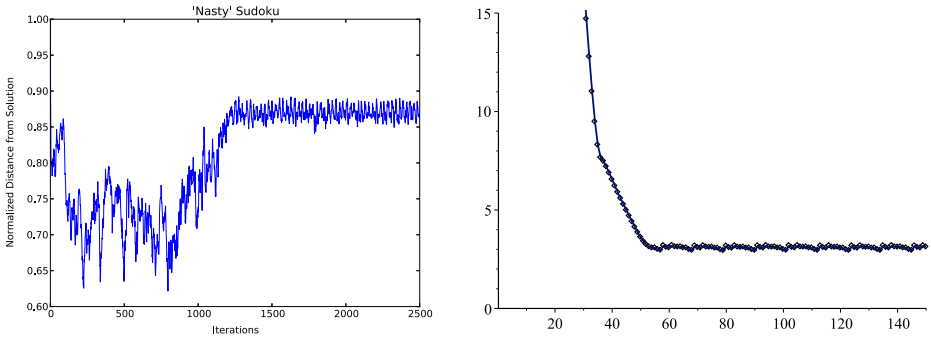


Fig. 14 Distance from iterates to a solution. Left: a sudoku puzzle [10]. Right: T_{E_{14},L_9}

First consider Fig. 13. On the left, we see the algorithm struggle for a long period of time before finally converging. Compare this to the image on the right: for T_{E_2,L_2} with 210 iterates, distance of each iterate—to the particular feasible point the sequence converges to—is plotted. The subsequences x_{2k} and x_{2k-1} are colored light and dark gray respectively. They correspond respectively to iterates landing in the domain of attraction for the left and right feasible points; see Fig. 8. Without the geometric intuition gleaned from Fig. 8, we would not know to color these two subsequences distinctly, and the error plot would not reflect the behavior as clearly. Once the iterates have finally climbed free of the influence of the repelling period 2 points, we see a sudden rapid convergence to the relevant feasible point.

Now consider Fig. 14. At left, we see distance from the solution of the iterates for a different sudoku puzzle. This time the error stabilizes after a time without any indication of impending convergence. At right, we see the iterates for T_{E_{14},L_9}. The iterates approach the ellipse before being pulled into the attractive domain for some period 11 points, preventing convergence.

Experiments have also been conducted using the Douglas-Rachford method to solve matrix completion problems associated with incomplete Euclidean distance matrices for protein mapping [2, 4, 10]. Consider Fig. 15. The left image shows the relative error of iterates when solving the Euclidean distance matrices for various proteins. The right image

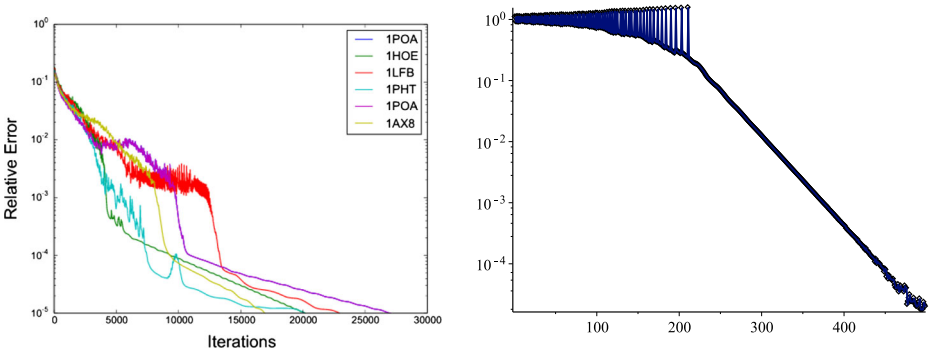


Fig. 15 Distance of iterates from solution (scale logarithmic). Left: for five proteins [10], Right: for the iterates of T_{E_8,L_6} pictured in Fig. 16

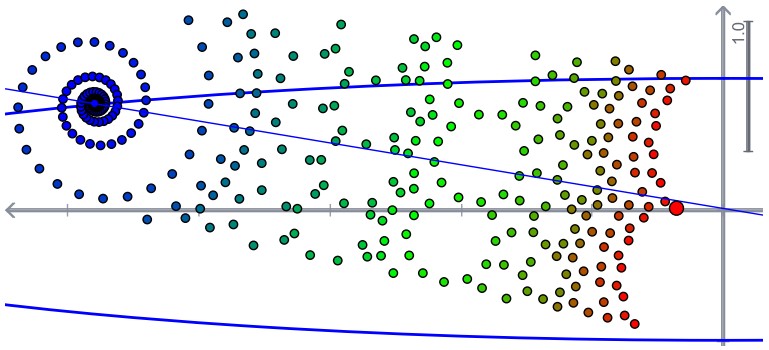


Fig. 16 A convergent sequence of iterates of T_{E_8,L_6} traces the outline of the domains

shows the relative error for the iterates of T_{E_8,L_6} when the sequence of iterates is started near to domains of attraction.

The exact iterates used to generate this data are shown in Fig. 16; they appear to trace out the shapes of the attractive domains for periodic points, narrowly avoiding them on their way to eventual convergence. Points started relatively close to domains of attraction for periodic points (as in the right hand side of Fig. 13) appear to take longer to converge than those started elsewhere.

While we cannot say with any real certainty that the behavior when solving these Euclidean distance matrices is analogous to iterates climbing away from repelling points or dodging and weaving between a nest of attractive domains, it is surprising that a system as simple as that of a line and ellipse can create behavior so similar to that observed in far more complicated scenarios.

5.2 Infeasible Cases

For the infeasible cases of line and the p-sphere or ellipse, we observed that the iterates of the Douglas-Rachford algorithm appear to walk to infinity with a roughly linear step size. In both infeasible cases, it is possible to strictly separate the two sets in question. This led to the following theorem.

Theorem 3 *Let x_0 be a point in, and A and B be closed subsets of a Hilbert space H and let $x_n := T^n_{A,B}(x_0)$. Suppose one of the following hold:*

1. *A is compact and $co(A)$ and $cl(co(B))$ are disjoint.*
2. *B is compact and $cl(co(A))$ and $co(B)$ are disjoint.*

Then $\|x_n\|$ tends linearly to ∞ with a step size of at least $d(A, B)$.

Proof we suppose that (1) applies. The proof when (2) applies is obtained by interchanging the roles of A and B. Now, we can strictly separate $\mathrm{co}(A)$ and $\mathrm{cl}(\mathrm{co}(B))$ with a hyperplane $\mathbb{H} = f^{-1}(\alpha)$ for some linear functional f. See [12, Theorem 1.7] for details. By translation invariance, let $\alpha = 0$. Then $\mathbb{H}$ is a subspace, so we can uniquely describe any $x \in X$ as

$x = h_x + y_x$ where $h_x \in \mathbb{H}$ and $y_x \in \mathbb{H}^{\perp}$. We can impose several additional properties on f:

$$|f(x)| = |f(h_x + y_x)| = \|y_x\|_X \quad \text{for all } x \in X \tag{8}$$

$$f(x) < 0 \quad \text{for all } x \in A \tag{9}$$

$$f(x) > 0 \quad \text{for all } x \in B. \tag{10}$$

Equations (8), (9), and (10) imply that $f(x) \geq d(A, B)$ for all $x \in B - A$. Now

$$\begin{aligned} x_{n+1} - x_n = \frac{R_B(R_A(x_n)) + x_n}{2} - x_n &= \frac{R_B(R_A(x_n)) - x_n}{2} \\ &= \frac{2P_B(R_A(x_n)) - R_A(x_n) - x_n}{2} \\ &= \frac{2P_B(R_A(x_n)) - (2P_A(x_n) - x_n) - x_n}{2} \\ &= P_B(R_A(x_n)) - P_A(x_n) \in B - A. \end{aligned}$$

Now $x_{n+1} - x_n \in B - A$ implies that $f(x_{n+1} - x_n) \geq d(A, B)$. Thus we have that, for all n, $f(x_{n+1}) \geq d(A, B) + f(x_n)$. This shows that $\|x_n\|_n \to \infty$ with a linear step size of at least $d(A, B)$. □

From this result we obtain the following corollary, the computer-assisted discovery of which motivated the pursuit of the more general Theorem.

Corollary 1 *In the infeasible case of a line L with an ellipse E or a p-sphere S, we have that* $\|x_n\| \to \infty$ *with a linear step size greater than or equal to* $d(E, L)$ *or* $d(S, L)$ *respectively.*

Using these results and the following remark, we can naturally extend some of the convex theory to the non-convex case.

Remark 3 As a consequence of [8, Theorem 4.5] we have, for convex subsets U, V of a Hilbert space H, with $U \cap (v + V) \neq \emptyset$, where $v = P_{\mathrm{cl}(\mathrm{ran}(\mathrm{Id} - T_{U,V}))}(0)$ is the *minimal displacement vector*, and for $x \in X$, that $(P_U T_{U,V}^n x)_n$ converges weakly to a point in $U \cap (v + V)$.

We extend this result in our context using Theorem 3.

Theorem 4 *Let A, B be the respective boundaries of two disjoint, closed, compact, convex sets U, V in H (in $\mathbb{R}^n$ it suffices to have only one be compact) so that A, B satisfy the requirements of Theorem 3. Let $x_{n+1} = T_{A,B}(x_n)$ and $v := P_{cl(ran(Id - T_{U,V}))}(0)$, the uniquely defined element in $cl(ran(Id - T_{U,V}))$ such that $\|v\| = \inf_{x \in X} \|x - T_{U,V}x\|$, and let $v' = P_{cl(ran(Id - T_{A,B}))}(0)$, the uniquely defined element in $cl(ran(Id - T_{A,B}))$ such that $\|v'\| = \inf_{x \in X} \|x - T_{A,B}x\|$. Then, for $x \in X$, we have that $(P_A T_{A,B}^n x)_n$ converges weakly to a point in $A \cap (v' + B)$.*

Proof By the closure and compactness we have attainment of elements which minimize the distance. Thus $A \cap (v' + B) = U \cap (v + V) \neq \emptyset$ where v is defined as in Remark 3.

Let f be defined as in Theorem 3 so that $f(u) < 0$ for all $u \in U$. Then the sequence $f(x_n)$ is monotone increasing and so there exists some $n' \in \mathbb{N}$ such that $f(x_n) \geq 0$ for all $n \geq n'$. Suppose $n \geq n'$. Then we have that $x_n \notin U$. Thus $P_A(x_n) = P_U(x_n)$, and so $R_A(x_n) = R_U(x_n)$. We also have that $R_A(x_n) \notin V$. Thus $P_B(R_A(x_n)) = P_V(R_A(x_n))$, and so $R_B(R_A(x_n)) = R_V(R_A(x_n))$. Thus we have that $T_{A,B}(x_n) = T_{U,V}(x_n)$, and so

$$(P_A T_{A,B}^{n'+n} x)_n = (P_A T_{A,B}^n x_{n'})_n = (P_A T_{U,V}^n x_{n'})_n \tag{11}$$

for all $n \in \mathbb{N}$. We have from Remark 3 that the sequence on the right converges to a point y in $U \cap (v + V)$. □

5.3 Closing Remarks

Given that we are investigating the Douglas-Rachford method applied to some of the simplest possible instances of a non-convex set, the emergence of such complexity is extraordinary. More interesting from a technical standpoint is the similarity with which the behavior in such simple situations appears to resemble some of what is observed for much larger and more complicated ones.

We also hope that we have succeeded in making a case for computer-assisted discovery, visualization, and verification. "*A heavy warning used to be given [by lecturers] that pictures are not rigorous; this has never had its bluff called and has permanently frightened its victims into playing for safety. Some pictures, of course, are not rigorous, but I should say most are (and I use them whenever possible myself).*"—J. E. Littlewood, 1885–1977 [from Littlewood's Miscellany, p. 35 in the 1953 edition], said long before the current powerful array of graphic, visualization and geometric tools were available.

Acknowledgments The authors wish to especially thank one of the referees for their careful reading of the manuscript and their several suggestions which have significantly improved the readability of the final version.

References

1. Aragón Artacho, F.J., Borwein, J.M., Tam, M.K.: Recent results on Douglas-Rachford methods for combinatorial optimization problems. J. Optim. Theory Appl. **163**, 1–30 (2014)
2. Aragón Artacho, F.J., Borwein, J.M., Tam, M.K.: Douglas-Rachford feasibility methods for matrix completion problems. ANZIAM J. **55**(4), 299–326 (2014)
3. Aragón Artacho, F.J., Borwein, J.M.: Global convergence of a non-convex Douglas-Rachford iteration. J. Glob. Optim. **57**(3), 753–769 (2013)
4. Bailey, D.H., Borwein, J.M.: Experimental computation as an ontological game changer: the impact of modern mathematical computation tools on the ontology of mathematics. In: Mathematics, Substance and Surmise: Views on the Meaning and Ontology of Mathematics, pp. 25–67. Springer, Berlin
5. Bauschke, H.H., Bello Cruz, J.Y., Nghia, T.T.A., Phan, H.M., Wang, X.: The rate of linear convergence of the Douglas-Rachford algorithm for subspaces is the cosine of the Friedrichs angle. J. Approx. Theory **185**, 63–79 (2014)
6. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, Berlin (2011)
7. Bauschke, H.H., Combettes, P.L., Luke, D.R.: Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization. J. Opt. Soc. Am. **19**, 1334–1345 (2002)
8. Bauschke, H.H., Moursi, W.M.: On the Douglas-Rachford algorithm. Math. Program. **164**, 263–284 (2017)
9. Borwein, J.M., Sims, B.: The Douglas-Rachford algorithm in the absence of convexity. In: Fixed-Point Algorithms for Inverse Problems in Science and Engineering, Springer Optimization and its Applications, vol. 49, pp. 93–109 (2011)

10. Borwein, J.M., Tam, M.K.: Reflection methods for inverse problems with applications to protein conformation determination. In: Springer Volume on the CIMPA school Generalized Nash Equilibrium Problems, Bilevel Programming and MPEC, New Delhi (2012)
11. Benoist, J.: The Douglas-Rachford algorithm for the case of the sphere and the line. J. Glob. Optim. **63**, 363–380 (2015)
12. Brezis, H.: Functional Analysis, Sobolev Spaces and Partial Differential Equations. Springer, Berlin (2011)
13. Deutsch, F.: Rate of convergence of the method of alternating projections. In: Brosowski, B., Deutsch, F. (eds.) Parametric Optimization and Approximation, pp. 96–107. Birkhäuser, Basel (1983)
14. Dontchev, A.L.: Private communication
15. Douglas, J., Rachford, H.H.: On the numerical solution of the heat conduction problem in 2 and 3 space variables. Trans. AMS **82**, 421–439 (1956)
16. Lakshmikantham, V., Trigiante, D.: Theory Of Difference Equations - Numerical Methods And Applications. Marcel Dekker, New York (2002)
17. Lions, P.L., Mercier, B.: Splitting algorithms for the sum of two nonlinear operators. SIAM J. Numer. Anal. **16**, 964–979 (1979)
18. Borwein, J.M., Lindstrom, S.B., Sims, B., Schneider, A., Skerritt, M.P.: Appendix to dynamics of the Douglas-Rachford method for ellipses and p-spheres. Available at: http://hdl.handle.net/1959.13/1330341 (2017)