Douglas–Rachford Iterations in the Absence of Convexity

Laureate Prof. Jonathan Borwein with Matthew Tam
http://carma.newcastle.edu.au/DRmethods/paseky.html

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

CARMA

Spring School on Variational Analysis VI
Paseky nad Jizerou, April 19–25, 2015

Last Revised: May 6, 2016

Lonely Planet's top 10 cities

10:30 AEST Mon Nov 1 2010
Adam Bub

10 images in this story

Travel experts Lonely Planet have named the top 10 cities for 2011 in their annual travel bible, *Best in Travel 2011*. The top-listed cities win points for their local cultures, value for money, and overall va-va-voom. So which cities make the cut? Find out here, from 10 to 1...

What do you think of the list? **Tell us here!**

Related links: Lonely Planet destination videos

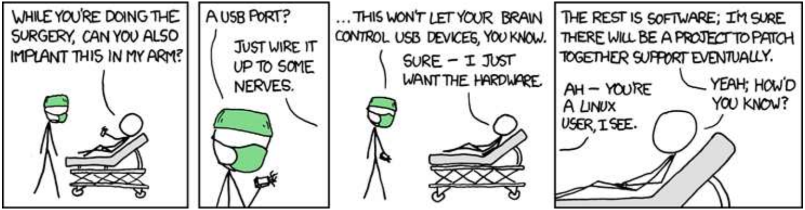
A weekend in Newcastle

Images: ThinkStock/Getaway

9. Newcastle, Australia 2 of 10

The Rest is Software

"It was my luck (perhaps my bad luck) to be the world chess champion during the critical years in which computers challenged, then surpassed, human chess players. Before 1994 and after 2004 these duels held little interest." — Garry Kasparov, 2010



- Likewise much of current Optimization Theory.

Abstract

- The Douglas–Rachford iteration scheme, introduced half a century ago in connection with nonlinear heat flow problems, aims to find a point common to two or more closed constraint sets.
  - Convergence is ensured when the sets are convex subsets of a Hilbert space, however, despite the absence of satisfactory theoretical justification, the scheme has been routinely used to successfully solve a diversity of practical optimization or feasibility problems in which one or more of the constraints involved is non-convex.
- As a first step toward addressing this deficiency, **we provide convergence results for a proto-typical non-convex (phase-recovery) scenario**: Finding a point in the intersection of the Euclidean sphere and an affine subspace.

Abstract

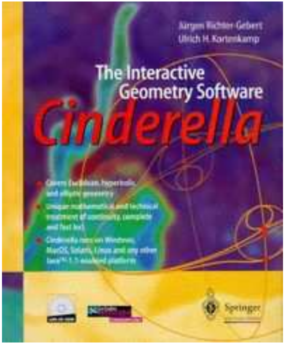
- The Douglas–Rachford iteration scheme, introduced half a century ago in connection with nonlinear heat flow problems, aims to find a point common to two or more closed constraint sets.
  - Convergence is ensured when the sets are convex subsets of a Hilbert space, however, despite the absence of satisfactory theoretical justification, the scheme has been routinely used to successfully solve a diversity of practical optimization or feasibility problems in which one or more of the constraints involved is non-convex.
- As a first step toward addressing this deficiency, **we provide convergence results for a proto-typical non-convex (phase-recovery) scenario**: Finding a point in the intersection of the Euclidean sphere and an affine subspace.

Abstract

- The Douglas–Rachford iteration scheme, introduced half a century ago in connection with nonlinear heat flow problems, aims to find a point common to two or more closed constraint sets.
  - Convergence is ensured when the sets are convex subsets of a Hilbert space, however, despite the absence of satisfactory theoretical justification, the scheme has been routinely used to successfully solve a diversity of practical optimization or feasibility problems in which one or more of the constraints involved is non-convex.
- As a first step toward addressing this deficiency, **we provide convergence results for a proto-typical non-convex (phase-recovery) scenario**: Finding a point in the intersection of the Euclidean sphere and an affine subspace.

An Interactive Presentation

- Much of my lecture will be interactive using the interactive geometry package Cinderella and the HTML applets
  - www.carma.newcastle.edu.au/~jb616/reflection.html
  - www.carma.newcastle.edu.au/~jb616/expansion.html
  - www.carma.newcastle.edu.au/~jb616/lm-june.html



Those Involved

Brailey Sims

Fran Aragon

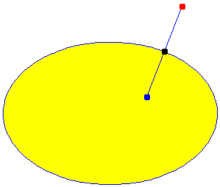
[0] Thanks also to Ulli Kortenkamp, Matt Skerritt and Chris Maitland

Phase Reconstruction

Projectors and Reflectors: $P_A(x)$ is the metric projection or nearest point and $R_A(x)$ reflects in the tangent: x is red.

Veit Elser, Ph.D.

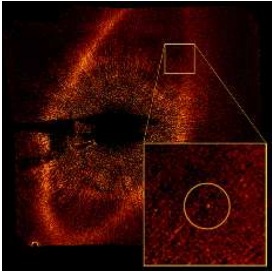
2007 Elser solving Sudoku with **reflectors**.



projection (black) and reflection (blue) of point (red) on boundary (blue) of ellipse (yellow)

"All physicists and a good many quite respectable mathematics are contemptuous about proof."
– **G.H. Hardy** (1877–1947)

2008 Finding exoplanet Fomalhaut in Piscis with **projectors**.

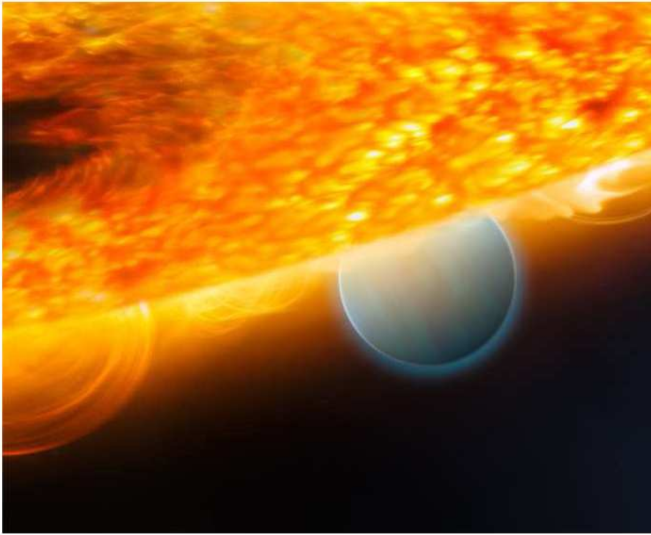


The story of Hubble's 1.3mm error in the "upside down" lens (1990).

And Kepler's hunt for exo-planets (launched March 2009).

We wrote:
"We should add, however, that many Kepler sightings in particular remain to be 'confirmed'. Thus one might legitimately wonder how mathematical robust are the underlying determinations of velocity, imaging, transiting, timing, micro-lensing, etc.?

http://experimentalmath.info/blog/2011/09/where-is-everybody/



Feeling the heat: Kepler scientists justify why some exoplanet data needs to be held back, for now. Image: A "Hot Jupiter" exoplanet close to its host star (ESO).

One of the biggest astronomical stories to unfold over the last decade or so is the story of exoplanets (or "extra-solar planets"). The theory of the formation of our solar system predicts that there should be many more such systems out there. And there certainly are, in fact, 461 at time of writing.

The story of Hubble's 1.3mm error in the "upside down" lens (1990).

And Kepler's hunt for exo-planets (launched March 2009).

We wrote:
"We should add, however, that many Kepler sightings in particular remain to be 'confirmed'. Thus one might legitimately wonder how mathematical robust are the underlying determinations of velocity, imaging, transiting, timing, micro-lensing, etc.?

http://experimentalmath.info/blog/2011/09/where-is-everybody/

THE CONVERSATION BETA

Academic rigour, journalistic flair

26 September 2011, 8.59am AEST

The exoplanet that wasn't. Or was it?

AUTHOR

Sunanda Creagh
Editor

DISCLOSURE STATEMENT

Our goal is to ensure the content is not compromised in any way. We therefore ask all authors to disclose any potential conflicts of interest before publication.

LICENCE TO REPUBLISH

We license our articles under Creative Commons — attribution, no derivatives.

Click here to get a copy of this article to republish.

An exoplanet called Fomalhaut b has been photographed in an unexpected spot — so is it even an exoplanet at all? NASA/http://www.nasa.gov

A distant planet that made its name as the world's first directly photographed exoplanet is at the centre of an astronomical stoush, after it veered off course and new doubts were raised about its existence.

It was in 2008 that Hubble astronomer Paul Kalas from the University of California at Berkeley and NASA announced that Fomalhaut b had been photographed orbiting a star called Fomalhaut around 25 light years from Earth.

Why Does it Work?

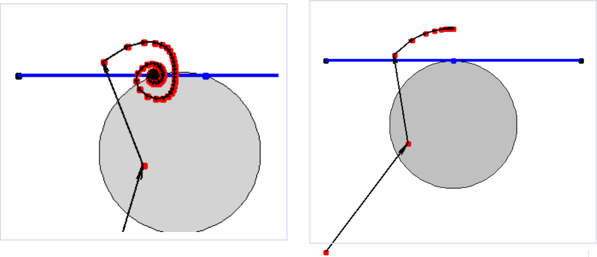
In a wide variety of large hard problems (protein folding, 3SAT, Sudoku) A is non-convex but DR and "divide and concur" (below) works better than theory can explain. It is:

$$R_A(x) := 2P_A(x) - x \text{ and } x \mapsto \frac{x + R_B\left(R_A(x)\right)}{2}.$$

Consider the simplest case of a line B of height h and the unit circle A. With $z_n := (x_n, y_n)$ the iteration becomes

$$x_{n+1} := \cos\theta_n, \; y_{n+1} := y_n + h - \sin\theta_n, \quad (\theta_n := \arg z_n).$$

For $h = 0$: We prove convergence to one of the two points in $A \cap B$ iff we do not start on the vertical axis (where we have chaos). For $h > 1$: (infeasible) it is easy to see the iterates go to infinity (vertically). For $h = 1$: We converge to an infeasible point. For $h \in (0, 1)$: The pictures are lovely but proofs escaped us for 9 months. Two representative Maple pictures follow:



An ideal problem for introducing early undergraduates to research, with many many accessible extensions in 2 or 3 dimensions.

Why Does it Work?

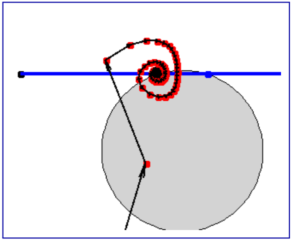
In a wide variety of large hard problems (protein folding, 3SAT, Sudoku) A is non-convex but DR and "divide and concur" (below) works better than theory can explain. It is:

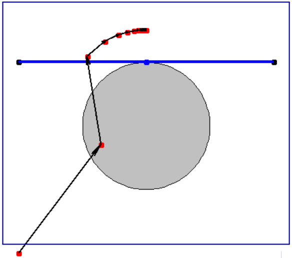
$$R_A(x) := 2P_A(x) - x \text{ and } x \mapsto \frac{x + R_B\left(R_A(x)\right)}{2}.$$

Consider the simplest case of a line B of height h and the unit circle A. With $z_n := (x_n, y_n)$ the iteration becomes

$$x_{n+1} := \cos\theta_n,\ y_{n+1} := y_n + h - \sin\theta_n, \quad (\theta_n := \arg z_n).$$

For $h = 0$: We prove convergence to one of the two points in $A \cap B$ iff we do not start on the vertical axis (where we have chaos). For $h > 1$: (infeasible) it is easy to see the iterates go to infinity (vertically). For $h = 1$: We converge to an infeasible point. For $h \in (0,1)$: The pictures are lovely but proofs escaped us for 9 months. Two representative Maple pictures follow:





An ideal problem for introducing early undergraduates to research, with many many accessible extensions in 2 or 3 dimensions.

Why Does it Work?

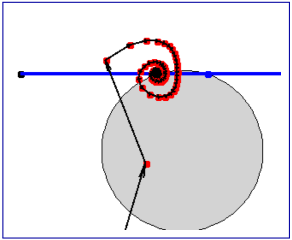
In a wide variety of large hard problems (protein folding, 3SAT, Sudoku) A is non-convex but DR and "divide and concur" (below) works better than theory can explain. It is:

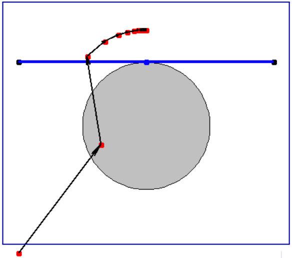
$$R_A(x) := 2P_A(x) - x \text{ and } x \mapsto \frac{x + R_B\left(R_A(x)\right)}{2}.$$

Consider the simplest case of a line B of height h and the unit circle A. With $z_n := (x_n, y_n)$ the iteration becomes

$$x_{n+1} := \cos\theta_n, \; y_{n+1} := y_n + h - \sin\theta_n, \quad (\theta_n := \arg z_n).$$

For $h = 0$: We prove convergence to one of the two points in $A \cap B$ iff we do not start on the vertical axis (where we have chaos). For $h > 1$: (infeasible) it is easy to see the iterates go to infinity (vertically). For $h = 1$: We converge to an infeasible point. For $h \in (0, 1)$: The pictures are lovely but proofs escaped us for 9 months. Two representative Maple pictures follow:





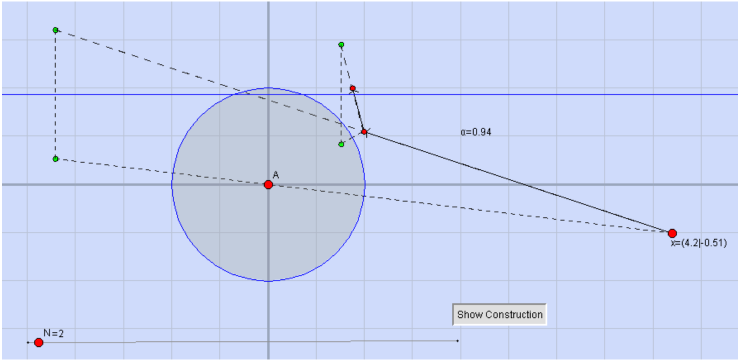
An ideal problem for introducing early undergraduates to research, with many many accessible extensions in 2 or 3 dimensions.

Interactive Phase Recovery in Cinderella

Recall the simplest case of a line B of height h and the unit circle A. With $z_n := (x_n, y_n)$ the iteration becomes

$$R_A(x) := 2P_A(x) - x \text{ and } x \mapsto \frac{x + R_B\,(R_A(x))}{2}.$$

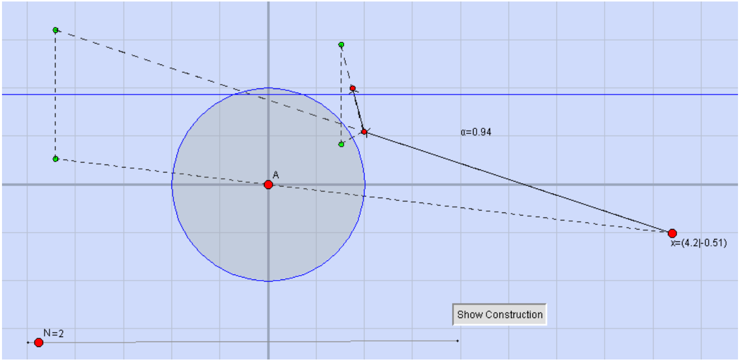
A Cinderella picture of two steps from $(4.2, -0.51)$ follows:



Recall the simplest case of a line B of height h and the unit circle A. With $z_n := (x_n, y_n)$ the iteration becomes

$$R_A(x) := 2P_A(x) - x \text{ and } x \mapsto \frac{x + R_B\left(R_A(x)\right)}{2}.$$

A Cinderella picture of two steps from $(4.2, -0.51)$ follows:

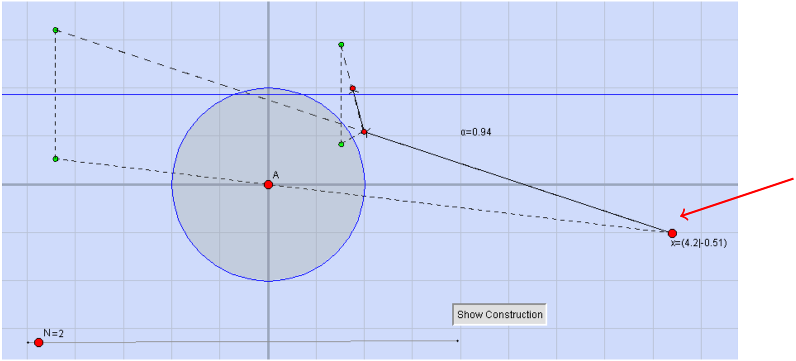


Interactive Phase Recovery in Cinderella

Recall the simplest case of a line B of height h and the unit circle A. With $z_n := (x_n, y_n)$ the iteration becomes

$$R_A(x) := 2P_A(x) - x \text{ and } x \mapsto \frac{x + R_B\left(R_A(x)\right)}{2}.$$

A Cinderella picture of two steps from $(4.2, -0.51)$ follows:

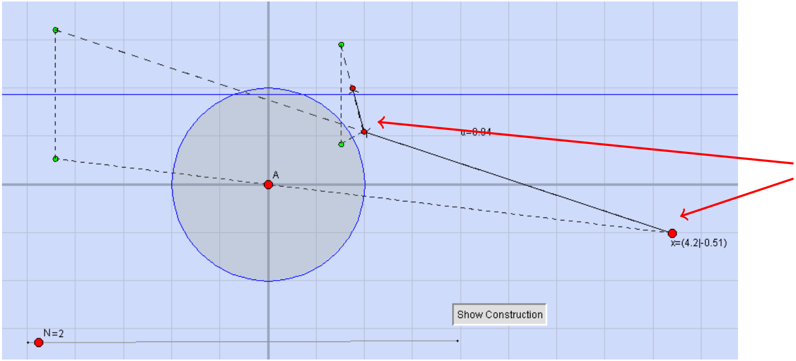


Interactive Phase Recovery in Cinderella

Recall the simplest case of a line B of height h and the unit circle A. With $z_n := (x_n, y_n)$ the iteration becomes

$$R_A(x) := 2P_A(x) - x \text{ and } x \mapsto \frac{x + R_B\left(R_A(x)\right)}{2}.$$

A Cinderella picture of two steps from $(4.2, -0.51)$ follows:

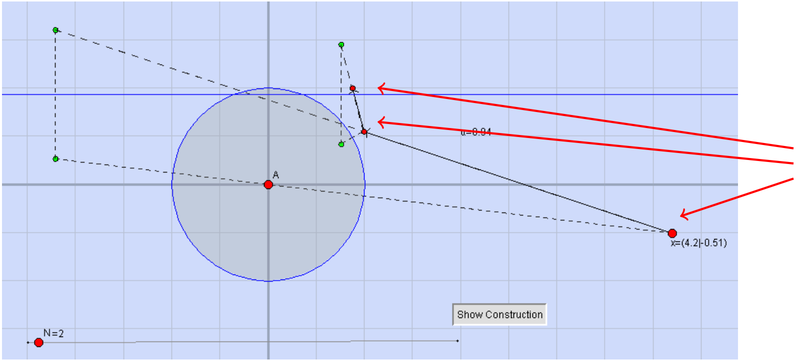


Interactive Phase Recovery in Cinderella

Recall the simplest case of a line B of height h and the unit circle A. With $z_n := (x_n, y_n)$ the iteration becomes

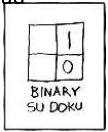
$$R_A(x) := 2P_A(x) - x \text{ and } x \mapsto \frac{x + R_B\left(R_A(x)\right)}{2}.$$

A Cinderella picture of two steps from $(4.2, -0.51)$ follows:



Divide and Concur

To find a point in the intersection of M-sets A_k and in X we can instead consider the subset $A := \prod_{k=1}^{M} A_k$ and the linear subset



$$B := \{x = (x_1, x_2, \ldots, x_M) : x_1 = x_2 = \cdots = x_M\},$$

of the product Hilbert space $\widetilde{X} := \left(\prod_{k=1}^{M} X\right)$. We observe

$$R_A(x) = \prod_{k=1}^{M} R_{A_k}(x_k),$$

Serial (L) & Parallel (R).

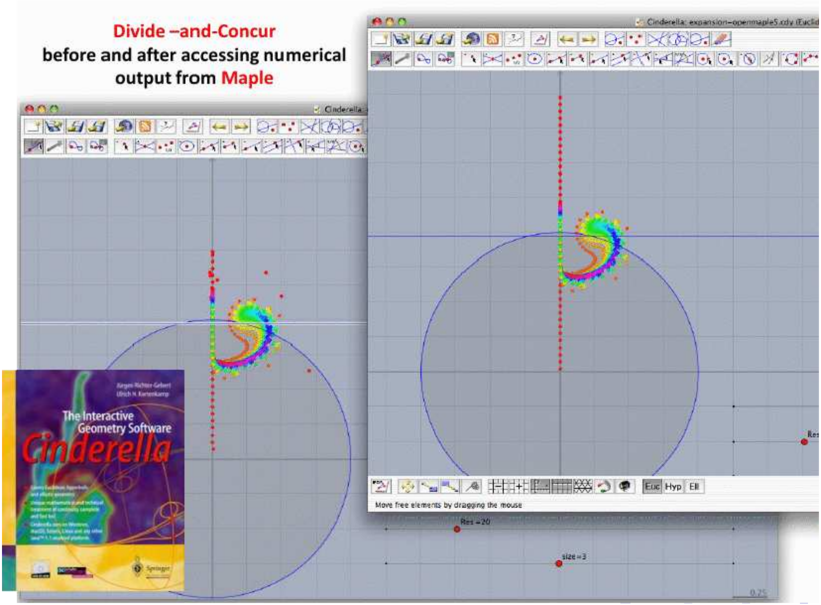
hence the reflection may be 'divided' up and

$$P_B(x) = \left(\frac{x_1 + x_2 + \cdots + x_M}{M}, \ldots, \frac{x_1 + x_2 + \cdots + x_M}{M}\right),$$

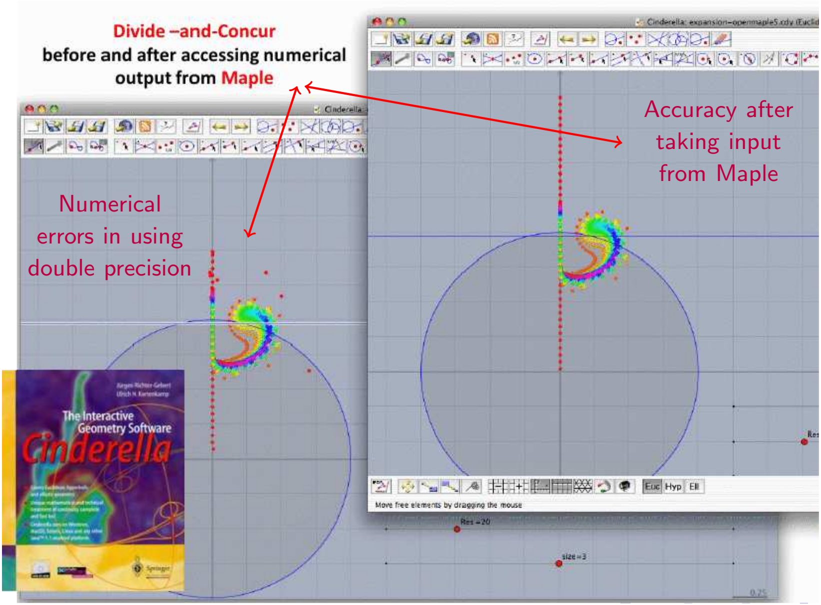
so that the projection and reflection on B are averaging ('concurrences'), hence the name. In this form the algorithm is suited to parallelization. We can also compose more reflections in serial—we still observe iterates spiralling to a feasible point.

CAS+IGP: The Grief is in the GUI

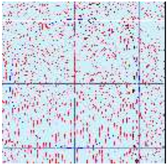
Divide –and-Concur
before and after accessing numerical output from Maple



CAS+IGP: The Grief is in the GUI



- Exploration first in Maple and then in Cinderella (SAGE)
  - ability to look at orbits/iterations dynamically is great for insight
  - allows for rapid reinforcement and elaboration of intuition
- Decided to look at ODE analogues
  - and their linearizations
  - hoped for Lyapunov like results



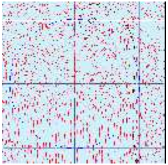
$$x'(t) = \frac{x(t)}{r(t)} - x(t), \quad y'(t) = h - \frac{y(t)}{r(t)},$$

where $r(t) := \sqrt{x(t)^2 + y(t)^2}$, is a reasonable counterpart to the Cartesian formulation —replacing $x_{n+1} - x_n$ by $x'(t)$, etc.—as in Figure.

- Searched literature for a discrete version
  - found Perron's work

The Route to Discovery

- Exploration first in Maple and then in Cinderella (SAGE)
  - ability to look at orbits/iterations dynamically is great for insight
  - allows for rapid reinforcement and elaboration of intuition
- Decided to look at ODE analogues
  - and their linearizations
  - hoped for Lyapunov like results



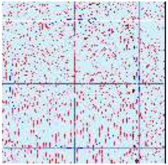
$$x'(t) = \frac{x(t)}{r(t)} - x(t), \quad y'(t) = h - \frac{y(t)}{r(t)},$$

where $r(t) := \sqrt{x(t)^2 + y(t)^2}$, is a reasonable counterpart to the Cartesian formulation —replacing $x_{n+1} - x_n$ by $x'(t)$, etc.—as in Figure.

- Searched literature for a discrete version
  - found Perron's work

The Route to Discovery

- Exploration first in Maple and then in Cinderella (SAGE)
  - ability to look at orbits/iterations dynamically is great for insight
  - allows for rapid reinforcement and elaboration of intuition
- Decided to look at ODE analogues
  - and their linearizations
  - hoped for Lyapunov like results



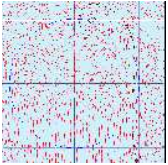
$$x'(t) = \frac{x(t)}{r(t)} - x(t), \quad y'(t) = h - \frac{y(t)}{r(t)},$$

where $r(t) := \sqrt{x(t)^2 + y(t)^2}$, is a reasonable counterpart to the Cartesian formulation —replacing $x_{n+1} - x_n$ by $x'(t)$, etc.—as in Figure.

- Searched literature for a discrete version
  - found Perron's work

The Route to Discovery

- Exploration first in Maple and then in Cinderella (SAGE)
  - ability to look at orbits/iterations dynamically is great for insight
  - allows for rapid reinforcement and elaboration of intuition
- Decided to look at ODE analogues
  - and their linearizations
  - hoped for Lyapunov like results



$$x'(t) = \frac{x(t)}{r(t)} - x(t), \quad y'(t) = h - \frac{y(t)}{r(t)},$$

where $r(t) := \sqrt{x(t)^2 + y(t)^2}$, is a reasonable counterpart to the Cartesian formulation —replacing $x_{n+1} - x_n$ by $x'(t)$, etc.—as in Figure.

- Searched literature for a discrete version
  - found Perron's work

The Basis of the Proof

Theorem (Perron)

If $f:\mathbb{N}\times\mathbb{R}^m\to\mathbb{R}^m$ satisfies

$$\lim_{x\to 0}\frac{\|f(n,x)\|}{\|x\|}=0,$$

uniformly in n and M is a constant $n\times n$ matrix all of whose eigenvalues lie inside the unit disk, then the zero solution (provided it is an isolated solution) of the difference equation,

$$x_{n+1}=Mx_n+f(n,x_n),$$

is *exponentially asymptotically stable*; that is, there exists $\delta>0$, $K>0$ and $\zeta\in(0,1)$ such that $\|x_0\|<\delta$ then $\|x_n\|\le K\|x_0\|\zeta^n$.

In our case:

$$M=\begin{pmatrix}\alpha^2 & -\alpha\sqrt{1-\alpha^2} & 0 & \dots & 0\\ \alpha\sqrt{1-\alpha^2} & \alpha^2 & 0 & \dots & 0\\ 0 & 0 & 0 & \dots & 0\\ \vdots & \vdots & \vdots & \ddots & \vdots\\ 0 & 0 & 0 & \dots & 0\end{pmatrix},$$

and the spectrum of the gradient comprises 0, and $\alpha^2\pm i\alpha\sqrt{1-\alpha^2}$.

The Basis of the Proof

Theorem (Perron)

If $f : \mathbb{N} \times \mathbb{R}^m \to \mathbb{R}^m$ satisfies

$$\lim_{x \to 0} \frac{\|f(n,x)\|}{\|x\|} = 0,$$

uniformly in n and M is a constant $n \times n$ matrix all of whose eigenvalues lie inside the unit disk, then the zero solution (provided it is an isolated solution) of the difference equation,

$$x_{n+1} = Mx_n + f(n, x_n),$$

is *exponentially asymptotically stable*; that is, there exists $\delta > 0$, $K > 0$ and $\zeta \in (0,1)$ such that $\|x_0\| < \delta$ then $\|x_n\| \leq K\|x_0\|\zeta^n$.

In our case:

$$M = \begin{pmatrix} \alpha^2 & -\alpha\sqrt{1-\alpha^2} & 0 & \dots & 0 \\ \alpha\sqrt{1-\alpha^2} & \alpha^2 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

and the spectrum of the gradient comprises 0, and $\alpha^2 \pm i\alpha\sqrt{1-\alpha^2}$.

The Basis of the Proof

Theorem (Perron)

If $f : \mathbb{N} \times \mathbb{R}^m \to \mathbb{R}^m$ satisfies

$$\lim_{x \to 0} \frac{\|f(n,x)\|}{\|x\|} = 0,$$

uniformly in n and M is a constant $n \times n$ matrix all of whose eigenvalues lie inside the unit disk, then the zero solution (provided it is an isolated solution) of the difference equation,

$$\mathbf{x_{n+1} = Mx_n + f(n, x_n)},$$

is *exponentially asymptotically stable*; that is, there exists $\delta > 0$, $K > 0$ and $\zeta \in (0,1)$ such that $\|x_0\| < \delta$ then $\|x_n\| \leq K\|x_0\|\zeta^n$.

In our case:

$$M = \begin{pmatrix} \alpha^2 & -\alpha\sqrt{1-\alpha^2} & 0 & \dots & 0 \\ \alpha\sqrt{1-\alpha^2} & \alpha^2 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix},$$

Explains spin for height in $(0,1)$

and the spectrum of the gradient comprises 0, and $\alpha^2 \pm i\alpha\sqrt{1-\alpha^2}$.

What We Can Now Show

Theorem (Borwein–Sims 2009)

For the case of a sphere in n-space and a line of height α (normalized so that we have $x(2) = \alpha, a = e_1, b = e_2$):

(a) If $0 \leq \alpha < 1$ then the Douglas–Rachford scheme is locally convergent at each of the critical points $\pm\sqrt{1-\alpha^2}a + \alpha b$.

(b) If $\alpha = 0$ and the initial point has $x_0(1) > 0$ then the scheme converges to the feasible point $(1, 0, 0, \ldots, 0)$.

(c) When L is tangential to S at b (*i.e.*, when $\alpha = 1$), starting from any initial point with $x_0(1) \neq 0$, the scheme converges to a point yb with $y > 1$.

(d) If there are no feasible solutions (*i.e.*, when $\alpha > 1$) then for any non-zero initial point $x_n(2)$ and hence $\|x_n\|$ diverge at at least linear rate to $+\infty$.

- The same result applies to the sphere S and any *affine* subset B.
- For non-affine B things are substantially more complex — even in $\mathbb{R}^2$.

What We Can Now Show

Theorem (Borwein–Sims 2009)

For the case of a sphere in n-space and a line of height α (normalized so that we have $x(2) = \alpha, a = e_1, b = e_2$):

(a) If $0 \leq \alpha < 1$ then the Douglas–Rachford scheme is locally convergent at each of the critical points $\pm\sqrt{1-\alpha^2}a + \alpha b$.

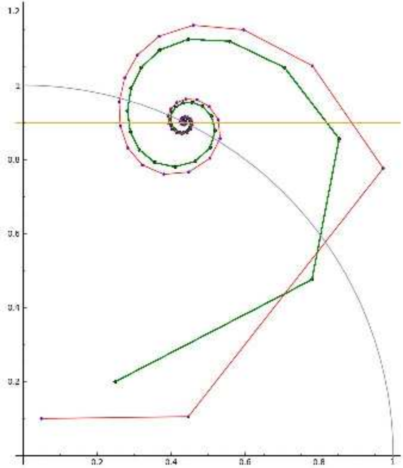
(b) If $\alpha = 0$ and the initial point has $x_0(1) > 0$ then the scheme converges to the feasible point $(1, 0, 0, \ldots, 0)$.

(c) When L is tangential to S at b (*i.e.*, when $\alpha = 1$), starting from any initial point with $x_0(1) \neq 0$, the scheme converges to a point yb with $y > 1$.

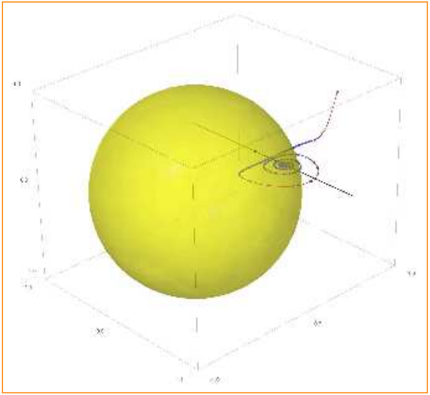
(d) If there are no feasible solutions (*i.e.*, when $\alpha > 1$) then for any non-zero initial point $x_n(2)$ and hence $\|x_n\|$ diverge at at least linear rate to $+\infty$.

- The same result applies to the sphere S and any *affine* subset B.
- For non-affine B things are substantially more complex — even in $\mathbb{R}^2$.

Algorithms *Appears* to be Stable



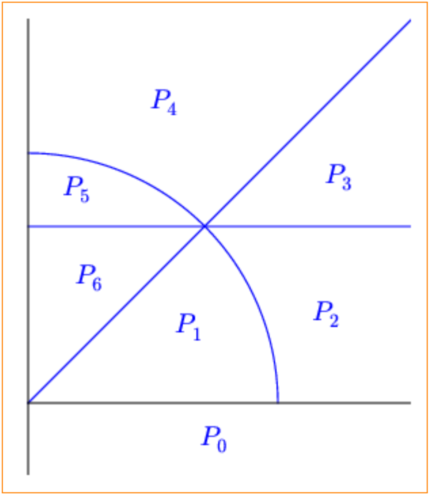
Three and Higher Dimensions



$$x_{n+1}(1) = x_n(1)/\rho_n,$$
$$x_{n+1}(2) = \alpha + (1 - 1/\rho_n)x_n(2), \quad \text{and}$$
$$x_{n+1}(k) = (1 - 1/\rho_n)x_n(k), \quad \text{for } k = 3, \ldots, N$$

where $\rho_n := \|x_n\| = \sqrt{x_n(1)^2 + \cdots + x_n(N)^2}$.

An "Even Simpler" Case



Intersection at $\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$.

If $(x_n, y_n) \in P_1 \cup P_2 \cup P_3$ then

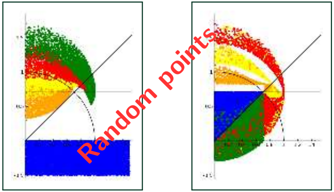
$$|(x_{n+1}, y_{n+1} - (x^*, y^*)|^2 \leq \frac{1}{2}|(x_n, y_n - (x^*, y^*)|^2.$$

If $(x_n, y_n) \in P_4$ then

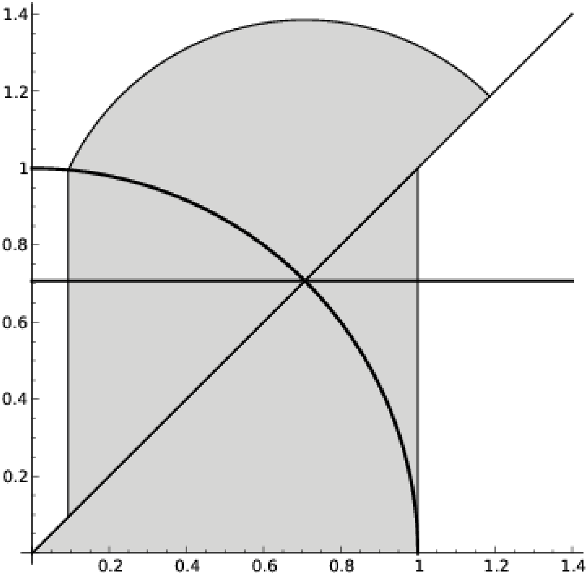
$$|(x_{n+1}, y_{n+1} - (x^*, y^*)|^2 \leq |(x_n, y_n - (x^*, y^*)|^2.$$

If $(x_n, y_n) \in P_5 \cup P_6$ then

$$|(x_{n+1}, y_{n+1} - (x^*, y^*)|^2 \leq \underbrace{\left(\frac{5}{2} - \sqrt{2} + \frac{1}{2}\sqrt{29 - 20\sqrt{2}}\right)}_{\approx 1.51}|(x_n, y_n - (x^*, y^*)|^2.$$



Aragón–Borwein Region of Convergence



The Search for a Lyapunov Function

Recent progress has been made by Joël Benoist. His idea is to search for a Lyapunov function V such that ∇V is perpendicular to the DR trajectories. That is,

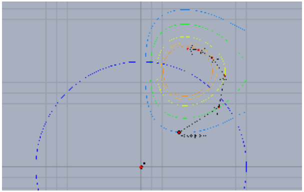
$$\langle \nabla V(x_n,\ y_n), (x_{n-1}, y_{n-1}) - (x_n, y_n)\rangle = 0.$$

Expressing (x_{n-1}, y_{n-1}) is terms of (x_n, y_n) gives the PDE:

$$(y-\lambda)\frac{\partial V}{\partial x}(x,y) + \frac{-\lambda\sqrt{1-x^2}+1-x^2}{x}\frac{\partial V}{\partial y}(x,y) = 0.$$

One solution to this PDE is the following:

$$V(x,y) = \frac{1}{2}(y-\lambda)^2 - \lambda\ln(1+\sqrt{1-x^2}) + \lambda\sqrt{1-x^2} + (\lambda-1)\ln x + \frac{1}{2}x^2.$$



The Search for a Lyapunov Function

Recent progress has been made by Joël Benoist. His idea is to search for a Lyapunov function V such that ∇V is perpendicular to the DR trajectories. That is,

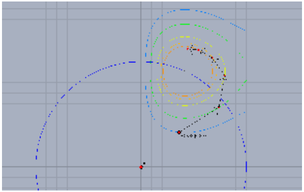
$$\langle \nabla V(x_n,\ y_n), (x_{n-1}, y_{n-1}) - (x_n, y_n) \rangle = 0.$$

Expressing (x_{n-1}, y_{n-1}) is terms of (x_n, y_n) gives the PDE:

$$(y - \lambda)\frac{\partial V}{\partial x}(x, y) + \frac{-\lambda\sqrt{1 - x^2} + 1 - x^2}{x}\frac{\partial V}{\partial y}(x, y) = 0.$$

One solution to this PDE is the following:

$$V(x, y) = \frac{1}{2}(y - \lambda)^2 - \lambda \ln(1 + \sqrt{1 - x^2}) + \lambda\sqrt{1 - x^2} + (\lambda - 1)\ln x + \frac{1}{2}x^2.$$



The Search for a Lyapunov Function

Recent progress has been made by Joël Benoist. His idea is to search for a Lyapunov function V such that ∇V is perpendicular to the DR trajectories. That is,

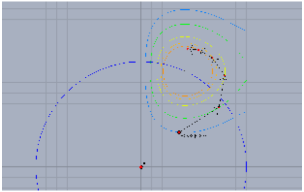
$$\langle \nabla V(x_n,\ y_n), (x_{n-1}, y_{n-1}) - (x_n, y_n) \rangle = 0.$$

Expressing (x_{n-1}, y_{n-1}) is terms of (x_n, y_n) gives the PDE:

$$(y - \lambda)\frac{\partial V}{\partial x}(x, y) + \frac{-\lambda\sqrt{1 - x^2} + 1 - x^2}{x}\frac{\partial V}{\partial y}(x, y) = 0.$$

One solution to this PDE is the following:

$$V(x, y) = \frac{1}{2}(y - \lambda)^2 - \lambda \ln(1 + \sqrt{1 - x^2}) + \lambda\sqrt{1 - x^2} + (\lambda - 1)\ln x + \frac{1}{2}x^2.$$



The Search for a Lyapunov Function

Denote the solution $(x^*, y^*) := (\sqrt{1-h^2}, h)$. Recall the Benoist's Lyapunov candidate function

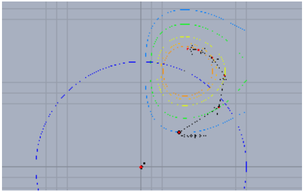
$$V(x,y) = \frac{1}{2}(y-\lambda)^2 - \lambda \ln(1+\sqrt{1-x^2}) + \lambda\sqrt{1-x^2} + (\lambda-1)\ln x + \frac{1}{2}x^2.$$

In the right half-space it is shown that:

1. (V decreases along DR trajectories): For all $\epsilon > 0$,

$$\sup_{\|(x,y)-(x^*,y^*)\|\geq\epsilon} (V(T(x,y)) - V(x,y)) < 0.$$

2. $V(T(x,y)) = V(x,y)$ if and only if $(x,y) = (x^*,y^*)$.



The Search for a Lyapunov Function

Denote the solution $(x^*, y^*) := (\sqrt{1-h^2}, h)$. Recall the Benoist's Lyapunov candidate function

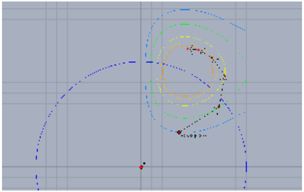
$$V(x,y) = \frac{1}{2}(y-\lambda)^2 - \lambda \ln(1+\sqrt{1-x^2}) + \lambda\sqrt{1-x^2} + (\lambda-1)\ln x + \frac{1}{2}x^2.$$

In the right half-space it is shown that:

1. (V decreases along DR trajectories): For all $\epsilon > 0$,

$$\sup_{\|(x,y)-(x^*,y^*)\|\geq\epsilon} \left(V(T(x,y)) - V(x,y)\right) < 0.$$

2. $V(T(x,y)) = V(x,y)$ if and only if $(x,y) = (x^*, y^*)$.



Global Convergence with a Half-Space Constraint

Consider the two-set feasibility problem given by a closed set $Q \subseteq \mathbb{R}^m$, and the half-space

$$H := \{x \in \mathbb{R}^m : \langle a, x\rangle \leq b\}.$$

where $b \in$, and $a \in^m$ with $\|a\| = 1$.

In this case, the Douglas–Rachford iteration simplifies to

$$x_{k+1} = \begin{cases} q_k & \text{if } \langle a, 2q_k - x_k\rangle \leq b, \\ q_k + (\langle a, x_k\rangle + b - 2\langle a, q_k\rangle)a & \text{otherwise,} \end{cases}$$

where, at each iteration, a point $q_k \in P_Q(x_k)$ is selected.

Motivated by experimental evidence, we first consider the case in which the set Q is finite.

Global Convergence with a Half-Space Constraint

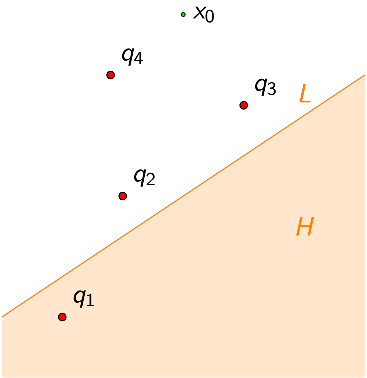


Fig. 1 A Douglas–Rachford iteration in ^{2} with the set $Q = \{q_1, q_2, q_3, q_4\}$ finds a solution in eight iterations.

Global Convergence with a Half-Space Constraint

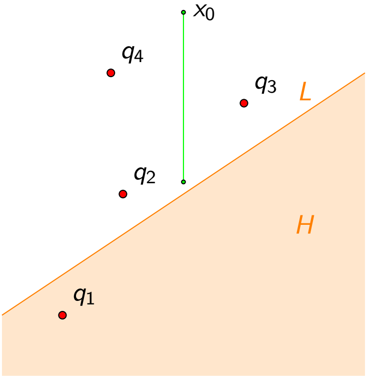


Fig. 1 A Douglas–Rachford iteration in ^{2} with the set $Q = \{q_1, q_2, q_3, q_4\}$ finds a solution in eight iterations.

Global Convergence with a Half-Space Constraint

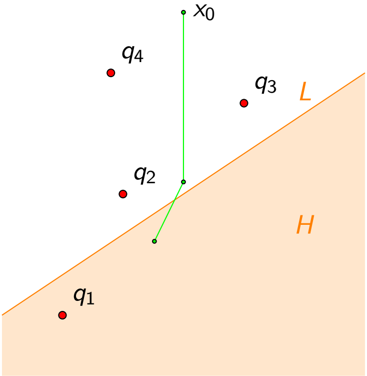


Fig. 1 A Douglas–Rachford iteration in ^{2} with the set $Q = \{q_1, q_2, q_3, q_4\}$ finds a solution in eight iterations.

Global Convergence with a Half-Space Constraint

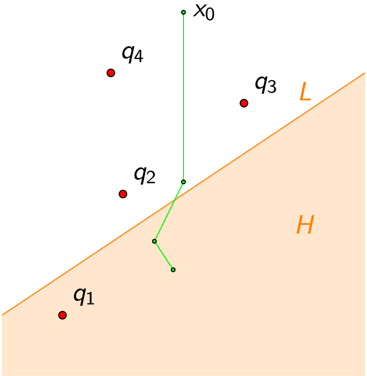


Fig. 1 A Douglas–Rachford iteration in ^{2} with the set $Q = \{q_1, q_2, q_3, q_4\}$ finds a solution in eight iterations.

Global Convergence with a Half-Space Constraint

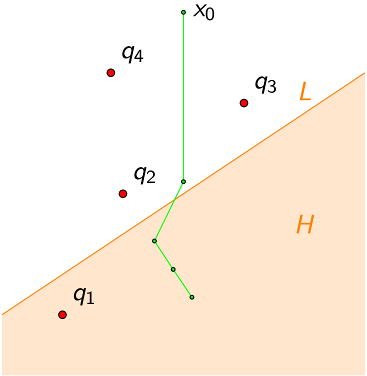


Fig. 1 A Douglas–Rachford iteration in ^{2} with the set $Q = \{q_1, q_2, q_3, q_4\}$ finds a solution in eight iterations.

Global Convergence with a Half-Space Constraint

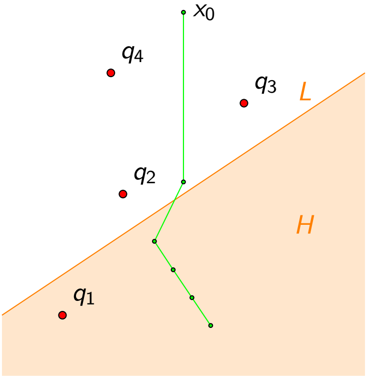


Fig. 1 A Douglas–Rachford iteration in ^{2} with the set $Q = \{q_1, q_2, q_3, q_4\}$ finds a solution in eight iterations.

Global Convergence with a Half-Space Constraint

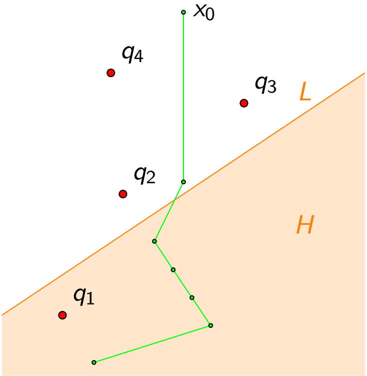


Fig. 1 A Douglas–Rachford iteration in ^{2} with the set $Q = \{q_1, q_2, q_3, q_4\}$ finds a solution in eight iterations.

Global Convergence with a Half-Space Constraint

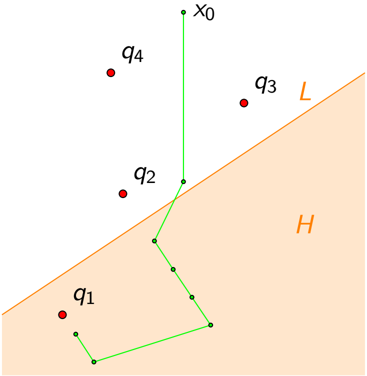


Fig. 1 A Douglas–Rachford iteration in ^{2} with the set $Q = \{q_1, q_2, q_3, q_4\}$ finds a solution in eight iterations.

Global Convergence with a Half-Space Constraint

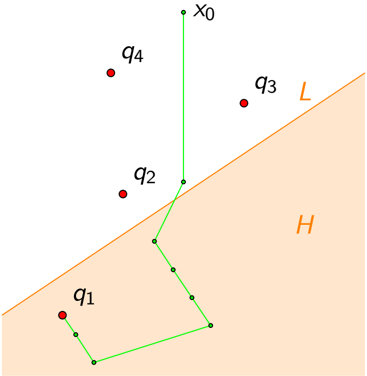


Fig. 1 A Douglas–Rachford iteration in ^{2} with the set $Q = \{q_1, q_2, q_3, q_4\}$ finds a solution in eight iterations.

Global Convergence with a Half-Space Constraint

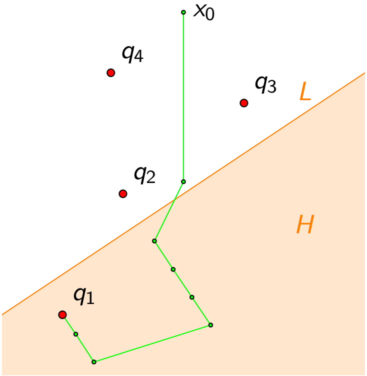


Fig. 1 A Douglas–Rachford iteration in ^{2} with the set $Q = \{q_1, q_2, q_3, q_4\}$ finds a solution in eight iterations.

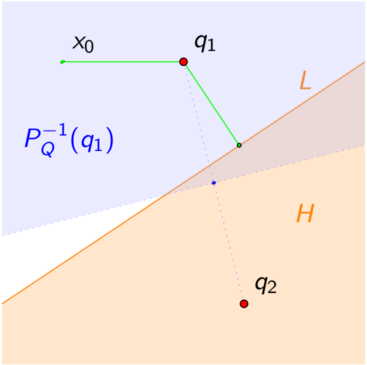


Fig. 2 The alternating projection algorithm fails to find a solution for any initial point in the set $P_Q^{-1}(q_1)$ where $Q = \{q_1, q_2\}$.

Global Convergence with a Half-Space Constraint

Theorem (Aragón Artacho–Borwein–Tam, 2015)

Suppose Q is a compact set. Let $\{x_k\}$ be a Douglas–Rachford sequence and $q_k \in P_Q(x_k)$ for all $k \in \mathbb{N}$. Then either:

(i) $d(q_k, H) \to 0$ and the set of cluster points $\{q_k\}$ is non-empty and contained in $Q \cap H$, or

(ii) $d(q_k, H) \to \beta$ for some $\beta > 0$ and $H \cap Q = \emptyset$.

Moreover, in the latter case, $\|x_k\| \to +\infty$.

It is worth noting that:

1. The set Q is not assumed to satisfy any (local) regularity properties (*e.g.*, strongly regular intersection, prox-regularity, ...).
2. The behaviour of the method does not depend on how p_k is chosen. The result holds for *any* choice.
3. The theorem remains true if one assume that the function

$$x \mapsto \iota_Q(x) + d(x, H),$$

has compact lower-level sets.

Global Convergence with a Half-Space Constraint

Theorem (Aragón Artacho–Borwein–Tam, 2015)

Suppose Q is a compact set. Let $\{x_k\}$ be a Douglas–Rachford sequence and $q_k \in P_Q(x_k)$ for all $k \in \mathbb{N}$. Then either:

(i) $d(q_k, H) \to 0$ and the set of cluster points $\{q_k\}$ is non-empty and contained in $Q \cap H$, or

(ii) $d(q_k, H) \to \beta$ for some $\beta > 0$ and $H \cap Q = \emptyset$.

Moreover, in the latter case, $\|x_k\| \to +\infty$.

It is worth noting that:

1. The set Q is not assumed to satisfy any (local) regularity properties (*e.g.*, strongly regular intersection, prox-regularity, ...).
2. The behaviour of the method does not depend on how p_k is chosen. The result holds for *any* choice.
3. The theorem remains true if one assume that the function

$$x \mapsto \iota_Q(x) + d(x, H),$$

has compact lower-level sets.

Global Convergence with a Half-Space Constraint

This theorem allows us to deduce global convergence of the Douglas–Rachford method applied to a sphere and a half-space (instead of an affine line).

Example (Global convergence for the sphere and half-space)

Let Q be the unit sphere and H a half-space in $\mathbb{R}^2$. By symmetry, we may assume $a = (0,1)$. Let $x_0 \neq 0$ with $x_0(1) > 0$. Then $x_k(1) > 0$ and $q_k = \frac{x_k}{\|x_k\|}$ for all $k \in \mathbb{N}$, and the iteration becomes

$$x_{k+1}(1) = \frac{x_k(1)}{\|x_k\|}, \quad x_{k+1}(2) = \begin{cases} \frac{x_k(2)}{\|x_k\|} & \text{if } \left(\frac{2}{\|x_k\|} - 1\right) x_k(2) \leq b, \\ \left(1 - \frac{1}{\|x_k\|}\right) x_k(2) + b & \text{otherwise.} \end{cases}$$

If $Q \cap H \neq \emptyset$ (or equivalently $b \geq -1$) then the previous theorem ensures $d(q_k, H) \to 0$. It then follows that either:

1. $q_{k_0} \in H \cap Q$ for some $k_0 \in \mathbb{N}$ (*i.e.*, a solution is found in finitely many iterations), or
2. $q_k(2) \to b$ and hence $q_k \to (\sqrt{1-b^2}, b) \in Q \cap H$.

Global Convergence with a Half-Space Constraint

Specialising to the finite case, we have the following.

Corollary (Aragón Artacho–Borwein–Tam, 2015)

Suppose Q is finite. Let $\{x_k\}$ be a Douglas–Rachford sequence and $q_k \in P_Q(x_k)$ for all $k \in \mathbb{N}$. Then either:

(i) $\{x_k\}$ and $\{q_k\}$ are eventually constant and the limit of $\{q_k\}$ is contained in $H \cap Q \neq \emptyset$, or

(ii) $H \cap Q = \emptyset$ and $\|x_k\| \to +\infty$.

- This corollary explains our previous example.
- First global convergence result for the Douglas–Rachford applicable to discrete/combinatorial constraint sets.
- Bauschke & Noll (2014) proved if the constraints are finite unions of convex sets, then method is locally convergent (in neighbourhoods of strong fixed points).

Global Convergence with a Half-Space Constraint

We give one further example from binary linear programming.

Example (Knapsack lower bound feasibility)

The classical 0-1 knapsack problem is the binary program

$$\min\left\{\langle c,x\rangle \mid x\in\{0,1\}^n,\ \langle a,x\rangle\le b\right\},$$

for vectors $a,c\in\mathbb{R}^m_+$ and $b\ge 0$.

The 0-1 knapsack lower-bound feasibility problem is the problem with constraints

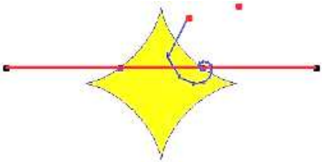
$$H:=\{x\in\mathbb{R}^n \mid \langle a,x\rangle\le b\},\quad Q:=\{x\in\{0,1\}^n \mid \langle c,x\rangle\ge\lambda\},$$

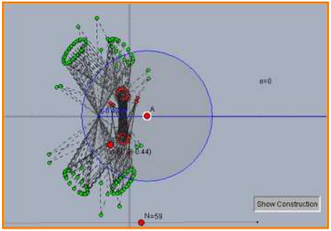
where $\lambda\ge 0$. As a decision problem it is NP-complete.

Applied to this problem, the corollary shows that the Douglas–Rachford method either finds a solution in finitely many iterations, or none exists and the norm of the Douglas–Rachford sequence diverges to infinity. Note that, in general, P_Q usually cannot be computed efficiently.

Commentary and Open Questions

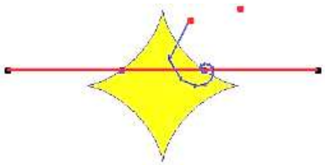
- As noted, the method parallelizes very well.
- Can one work out rates in the general convex case?
- Why does alternating projection (no reflection) work well for optical aberration but not phase reconstruction?
- Other cases of Lyapunov arguments for global convergence?
  - in the appropriate basins?
- Study general sets (in so-called CAT(0)metrics)
  - even the half-line case is much more complex
  - as I may now demo
- Why does the method work for a half-space but not a hyperplane?

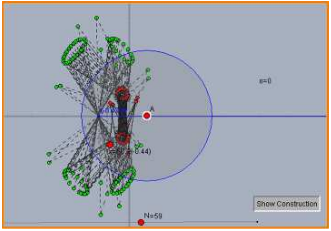




Commentary and Open Questions

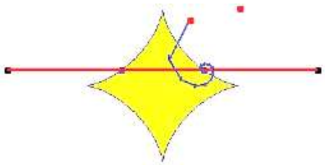
- As noted, the method parallelizes very well.
- Can one work out rates in the general convex case?
- Why does alternating projection (no reflection) work well for optical aberration but not phase reconstruction?
- Other cases of Lyapunov arguments for global convergence?
  - in the appropriate basins?
- Study general sets (in so-called CAT(0)metrics)
  - even the half-line case is much more complex
  - as I may now demo
- Why does the method work for a half-space but not a hyperplane?

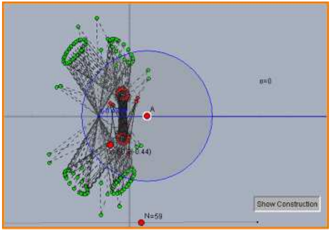




Commentary and Open Questions

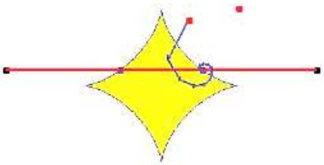
- As noted, the method parallelizes very well.
- Can one work out rates in the general convex case?
- Why does alternating projection (no reflection) work well for optical aberration but not phase reconstruction?
- Other cases of Lyapunov arguments for global convergence?
  - in the appropriate basins?
- Study general sets (in so-called CAT(0)metrics)
  - even the half-line case is much more complex
  - as I may now demo
- Why does the method work for a half-space but not a hyperplane?

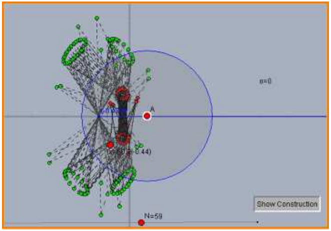




Commentary and Open Questions

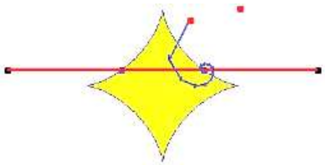
- As noted, the method parallelizes very well.
- Can one work out rates in the general convex case?
- Why does alternating projection (no reflection) work well for optical aberration but not phase reconstruction?
- Other cases of Lyapunov arguments for global convergence?
  - in the appropriate basins?
- Study general sets (in so-called CAT(0)metrics)
  - even the half-line case is much more complex
  - as I may now demo
- Why does the method work for a half-space but not a hyperplane?

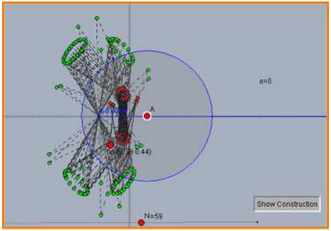




Commentary and Open Questions

- As noted, the method parallelizes very well.
- Can one work out rates in the general convex case?
- Why does alternating projection (no reflection) work well for optical aberration but not phase reconstruction?
- Other cases of Lyapunov arguments for global convergence?
  - in the appropriate basins?
- Study general sets (in so-called CAT(0)metrics)
  - even the half-line case is much more complex
  - as I may now demo
- Why does the method work for a half-space but not a hyperplane?





Exercises

1. (A lemma toward global convergence) The Douglas–Rachford iteration for the line and circle with $\alpha = 1/\sqrt{2}$. Is given by

$$x_{n+1} = \frac{x_n}{\rho_n}, \quad y_{n+1} = \alpha + \left(1 - \frac{1}{\rho_n}\right) y_n = \alpha + (\rho_n - 1)\sin\theta_n,$$

where $\rho_n = \sqrt{x_n^2 + y_n^2}$ and $\theta_n = \arg(x_n, y_n)$. Show if

$$(x_0, y_0) \in \{(x, y) : y \leq 0 < x\},$$

then $y_n > 0$ for some $n \in \mathbb{N}$.

2. (Existence of 2-cycles) Consider the sets

$$C_1 := \{(x, y) : x^2 + y^2 = 1\} \text{ and } C_2 := (x_1, 0) : x_1 \leq a\}.$$

Show that for each $a \in (0, 1)$ there is a point x such that $T_{C_1,C_2}x \neq x$ and $T^2_{C_1,C_2}x = x$. What happens instead if C_2 is merely the singleton $\{(a, 0)\}$?

3. Investigate the behavior of the Douglas–Rachford algorithm applied to two set feasibility problems with one of the sets finite (assume whatever structure you see fit on the other set).

4. (**Very Hard**) Complete the guided exercise (next slide) of Benoist's global convergence proof

Guided Exercise: Benoist's Global Convergence Proof

Consider the Lyapunov candidate function

$$V(x,y) = \frac{1}{2}(y-\lambda)^2 - \lambda \ln(1+\sqrt{1-x^2}) + \lambda\sqrt{1-x^2} + (\lambda-1)\ln x + \frac{1}{2}x^2.$$

Let $\Delta :=]0,1] \times \mathbb{R}$ and define $G : \Delta \to \Delta$ by

$$G(x,y) := V \circ T - V,$$

where T is the DR operator.

Consider $W : [0,1[\times[0,1[\to \mathbb{R}$ defined using a change of variables on G:

$$W(u,v) := G(a,b) \text{ where } u^2 = 1 - a^2 \text{ and } v^2 = \frac{b^2}{a^2+b^2}.$$

Guided Exercise: Benoist's Global Convergence Proof

Prove the following two lemmas.

Lemma 0

Show that W may be expressed as

$$W(u,v) := A(u) - A(v) + \sqrt{1-u^2}\,B(v) + \frac{u^2-h^2}{2},$$

where $A(t) := \frac{1+h}{2}\ln(1+t) + \frac{1-h}{2}\ln(1-t) - h$, $B(t) := \frac{t(h-t)}{\sqrt{1-t^2}}$.

Lemma 1

There exists a unique real number μ such that $0 < \mu < h$: (i) B is increasing on $[0,\mu]$ from 0 to $B(\mu)$, and (ii) B is decreasing in $[\mu, 1[$ from $B(\mu)$ to $-\infty$ with $B(h) = 0$.

Hint: Consider $B'(t)$.

Guided Exercise: Benoist's Global Convergence Proof

Prove the following lemma.

Lemma 2

For all $v \in [0,1[$, we have $W(0,v) < 0$.

Hint: Show that

$$W(0,v) = -\frac{1}{2}h^2 + S(v)h + R(v),$$

where $S(t) := \frac{1}{2}\ln\left(\frac{1-t}{1+t}\right) + \frac{t}{\sqrt{1-t^2}} + t$, $R(t) : -\frac{1}{2}\ln(1-t^2) - \frac{t^2}{\sqrt{1-t^2}}$.
Argue that there exists a unique $v^* < 0.8$ such that $S(v^*) = 1$, and distinguish three cases: (i) $v^* \leq v < 1$, (ii) $0 < v \leq v^*$, and (iii) $v = 0$.

Using Lemmas 1 and 2 to prove the following.

Proposition 1.

For all $(u, v) \in [0, 1[\times[0, 1[$ we have

$$W(u, v) \leq 0 \text{ with equality if and only if } u = v = h.$$

Hint: Show that

$$\frac{\partial W(u, v)}{\partial u} > 0 \iff B(u) > B(v).$$

Distinguish four cases: (i) $h \leq v < 1$, (ii) $\mu < v < h$, (iii) $v = \mu$, and (iv) $0 \leq v < \mu$.

Using Proposition 1 prove the following.

Proposition 2.

For all $\epsilon > 0$ we have

$$\sup_{(x,y)\in\Delta(\epsilon)} G(x,y) < 0,$$

where $\Delta(\epsilon) := \{(x,y) \in \Delta : d((x,y),(\sqrt{1-h^2},h)) > \epsilon\}$.

Hint: If $\sup_{(x,y)\in\Delta(\epsilon)} G(x,y) \geq 0$, use Proposition 1 to argue the existence of a subsequence such that $W(u_{n_k}, v_{n_k}) = G(x_{n_k}, y_{n_k}) \to 0$ such that $u_{n_k}, v_{n_k} \to (u,v)$ for some u and v.
Distinguish two cases: (i) $u \neq 1$ and $v \neq 1$, (ii) $u = 1$ or $v = 1$.

Guided Exercise: Benoist's Global Convergence Proof

Using Proposition 2 prove the main result.

Theorem (Benoist, 2015)

If $(x_0, y_0) \in \Delta$ then the Douglas–Rachford sequence converges to $(\sqrt{1-h^2}, h)$.

Hint: By telescoping, show that

$$\sum_{n \in \mathbb{N}} G(x_n, y_n)$$

converges and deduce $G(x_n, y_n) \to 0$ which contradicts Proposition 2.

References

J. Benoist. **The Douglas–Rachford algorithm for the case of the sphere and the line**, *J. Global Optimization*, in press, 2015.

J.M. Borwein & R.L. Luke. **Duality and Convex Programming**, pp. 229–270 in *Handbook of Mathematical Methods in Imaging*, O. Scherzer (Editor-in-Chief), Springer-Verlag. E-pub. 2010.

Jonathan M. Borwein & Brailey Sims, **The Douglas-Rachford algorithm in the absence of convexity**. Chapter 6, pp. 93–109 in Fixed-Point Algorithms for Inverse Problems in Science and Engineering in Springer Optimization and Its Applications, vol. 49, 2011.
http://www.carma.newcastle.edu.au/~jb616/dr.pdf

J. M. Borwein and J. Vanderwerff, **Convex Functions: Constructions, Characterizations and Counterexamples**. Encyclopedia of Mathematics and Applications, 109 Cambridge Univ Press, 2010.
http://projects.cs.dal.ca/ddrive/ConvexFunctions/

V. Elser, I. Rankenburg, & P. Thibault, **Searching with iterated maps**," Proc. National Academy of Sciences, 104 (2007), 418–423.

Many resources available at:

http://carma.newcastle.edu.au/DRmethods