Douglas-Rachford; a very versatile algorithm

Brailey Sims
CARMA, University of Newcastle, AUSTRALIA

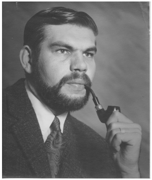
ACFPTO2018, Chiang Mai, 16–18 July 2018

https://carma.newcastle.edu.au/brailey/

Outline I

Much of this presentation is based on material from:

1. Scott Lindstrom and BS, "Sixty Years of Douglas-Rachford," (2018); in preparation for the special issue of J. Aust. Maths. Soc. dedicated to Jonathan M. Borwein - as is this talk.



2-set feasibility problems

- For two closed constraint sets A and B in a Hilbert space H, find a feasible point $x \in A \cap B$
- When the nearest point projection onto each set is readily computed, the application of a projection algorithm is a popular method of solution.
- Alternating projections [AP], introduced by J. von Neumann in 1933, is the oldest such method.
- Another effective method [D-R], on which we focus, was introduced by J. Douglas and H. H. Rachford in 1956.

The players

Projection Methods Alternating Projections [AP] vs. Douglas-Rachford [D-R]

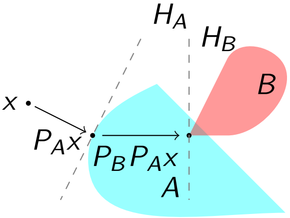


Figure: AP; $T_{A,B} := P_B \circ P_A$

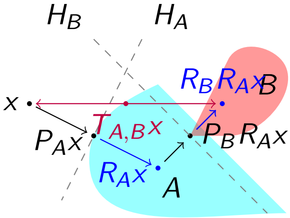
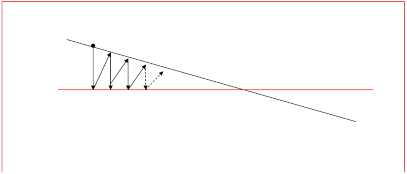
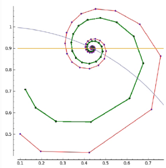


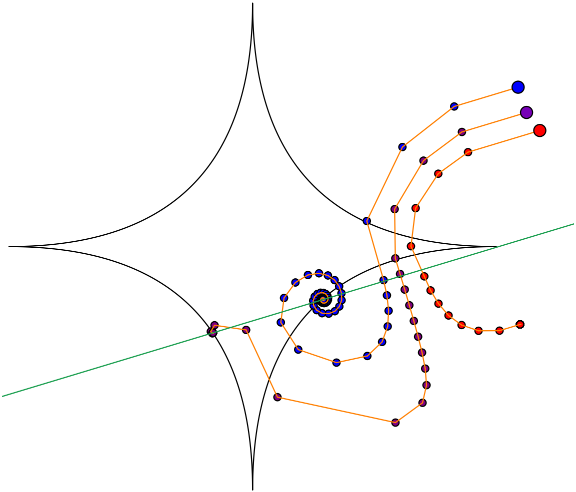
Figure: D-R; $T_{A,B} := \frac{1}{2}\left(Id + R_B R_A\right)$

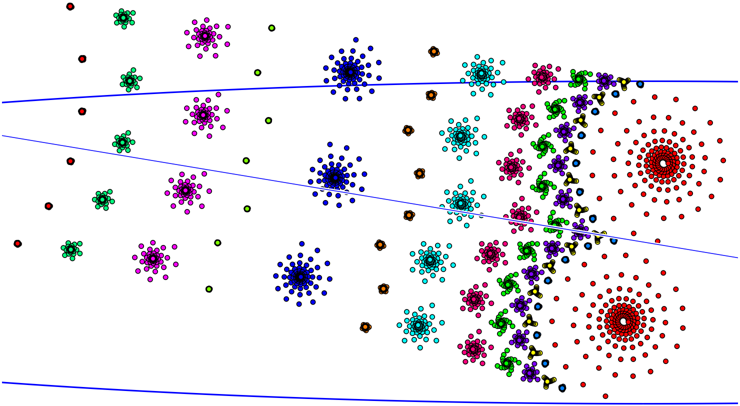
- Projection algorithm: from prescribed x_0 iterate $x_{n+1} = T_{A,B}(x_n)$, then (hopefully) $x = P_A(\{weak-\}\lim_n x_n) \in A \cap B$.

AP versus D-R





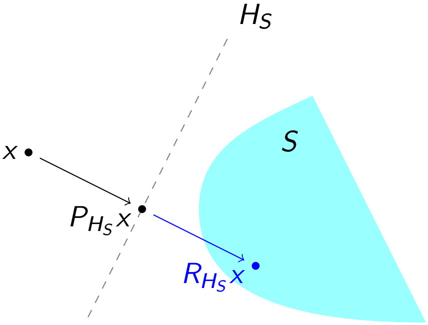




Can Or Can't

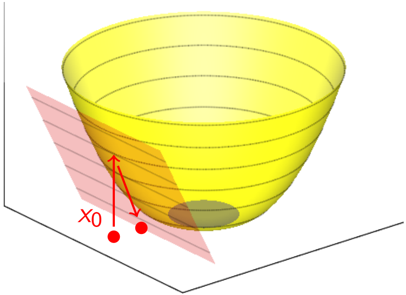
- The efficacy of these methods depends on the ease with which projections onto the constraint sets can be computed,
- HOWEVER:

Cutter methods - using projections to a separating hyperplane



- In cases where projection onto the set S is difficult to compute we may instead try projecting onto a separating hyperplane, H_S.

Subgradient Projections



- In the special case when the constraint set S is a level set of a convex function f we can try projecting onto the supporting hyperplane to $\mathrm{epi}(f)$ at $(x_0, f(x_0))$.

Reflection Parameters (relaxed projections)

We could also consider using more general relaxed projections:

$$R_{H_S}(x)^{(\gamma)} = (2-\gamma)(P_{H_S}(x) - x) + x.$$

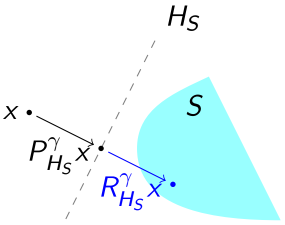


Figure: $\gamma = 0$

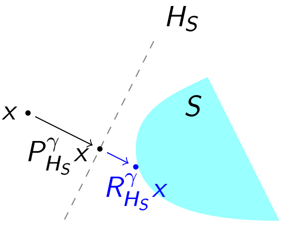


Figure: $\gamma = \frac{1}{2}$

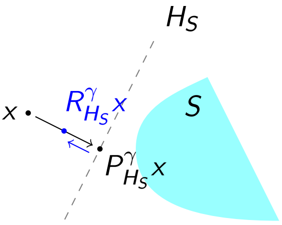


Figure: $\gamma = \frac{3}{2}$

Or, relaxing the averaging step; or both - but we won't.

- $T_{A,B}^{(\lambda)} := (1-\lambda)Id + \lambda R_{H_B} R_{H_A}$
- Note: $T_{A,B}^{(0)} = \mathrm{Id}$
- and $T_{A,B}^{(1)} = R_B R_A$

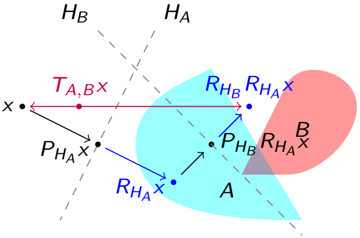


Figure: $\lambda = \frac{3}{4}$

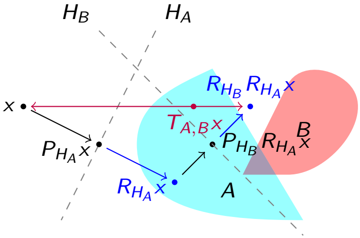


Figure: Shown: $\lambda = \frac{1}{4}$

From 2 sets to N sets.

- We can use Douglas-Rachford on a feasibility problem involving N sets; $\Omega_1 \ldots \Omega_N$, to find $x \in \cap_{k=1}^N \Omega_k$ by utilizing Pierra's product space method [21]; that is, by applying the algorithm in $\mathcal{H}^N$ to the two sets
  - $A := \Omega_1 \times \cdots \times \Omega_N$
  - $B := \{x = (y_1, \ldots, y_N) | y_1 = y_2 = \cdots = y_N\}$
- Nicknamed "divide and concur" by Simon Gravel and Veit Elser (the latter credits the former for the name) [17].
  - Reflection in A is the "divide" step entailing reflections in each of the individual constraint sets (eminently parallelizable).
  - "Concur" is the step of reflecting in the agreement (diagonal) set B.
- Other methods include cyclically anchored variant (CADRA) [5] and Borwein-Tam method (cyclic D-R) [10].

Zeros of sums of maximally monotone operators

Theorem 1 (Lions & Mercier)

In [20] Lions and Mercier showed that when $\mathcal{A}, \mathcal{B}$ are maximal monotone operators on a H, and $\mathcal{A}+\mathcal{B}$ is also maximal monotone, then for all non-zero λ and

$$T_{\mathcal{A},\mathcal{B}}x \ := \ J_{\mathcal{B}}^{\lambda}(2J_{\mathcal{A}}^{\lambda} - I)x + (I - J_{\mathcal{B}}^{\lambda})x, \tag{1}$$

the sequence of iterates, $x_{n+1} = T_{\mathcal{A},\mathcal{B}}x_n$, converges weakly to some $v \in H$, such that $J_{\mathcal{A}}^{\lambda}v$ is a zero of $\mathcal{A}+\mathcal{B}$.
Here, $J_F^{\lambda} := (\mathrm{Id} + \lambda F)^{-1}$ is the resolvent operator for λF

P-L Lions and B. Mercier

View from convex optimization

- If $\mathcal{A} := \partial f$ and $\mathcal{B} := \partial g$ for convex functions f, g, then $0 \in (\partial f + \partial g)u$ means u solves the minimization problem

$$\text{Find } u \in \operatorname*{argmin}_{x \in X}(f+g)(x).$$

- If, additionally, $f := \iota_A$ and $g := \iota_B$ are indicator functions; recall

$$\iota_C : x \mapsto \begin{cases} 0 & \text{if } x \in C \\ \infty & \text{otherwise,} \end{cases}$$

then $u \in \operatorname{argmin}(f+g)$ implies $u \in A \cap B$ (assuming feasibility).

How a minimization algorithm becomes a projection algorithm

- Phrasing the feasibility problem for A and B as a minimality problem in this way leads us to apply Lions-Mercier's result with $\mathcal{A} := \partial f$ and $\mathcal{B} := \partial g$ where $f := \iota_A$ and $g := \iota_B$ in which case,
  1. $\partial f = N_A$, the normal cone operator
  $$x \mapsto \begin{cases} \{y|(y, a-x) \leq 0 \text{ for all } a \in A\} & \text{if } x \in C \\ \emptyset & \text{otherwise,} \end{cases}$$
  2. $\partial g = N_B$,
  3. $J_{\mathcal{A}} := (\mathrm{Id} + \mathcal{A})^{-1} = (\mathrm{Id} + \partial_f)^{-1} = (\mathrm{Id} + N_A)^{-1} = P_A$,
  4. $J_{\mathcal{B}} := (\mathrm{Id} + \mathcal{B})^{-1} = (\mathrm{Id} + \partial_g)^{-1} = (\mathrm{Id} + N_B)^{-1} = P_B$,
- and in Lions-Mercier's algorithm,

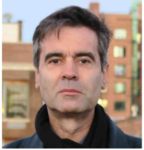
$$T_{\mathcal{A},\mathcal{B}} = J_{\mathcal{B}}^{\lambda}(2J_{\mathcal{A}}^{\lambda} - I) + (I - J_{\mathcal{A}}^{\lambda}) = \frac{1}{2}(Id + R_B R_A)$$

Later developments

For $T_{A,B} = \frac{1}{2}R_B R_A + \frac{1}{2}\mathrm{Id} = \frac{1}{2}(2P_B - \mathrm{Id}) \circ (2P_A - \mathrm{Id}) + \frac{1}{2}\mathrm{Id}$:

- Bauschke, Combettes, and Luke [3] were able to give a direct proof that the iterates weakly converge, avoiding the need that $\partial f + \partial B$ be maximal monotone in the feasibility setting.
- The requirement $\mathcal{A} + \mathcal{B}$ be maximally monotone was later relaxed outside the feasibility setting by Svaiter [23].

B, C and L

Definition 2 (Nonexpansivity conditions)

Let $D \subset H$ be nonempty and let $T : D \to H$. Then T is

1. *nonexpansive* if it is Lipschitz continuous with constant 1:

$$(\forall x, y \in D) \quad \|T(x) - T(y)\| \leq \|x - y\|;$$

2. *firmly nonexpansive* if

$$(\forall x, y \in D) \quad \|T(x) - T(y)\|^2 + \|(\mathrm{Id} - T)(x) - (\mathrm{Id} - T)(y)\|^2 \leq \|x - y\|^2;$$

Key facts:

1. Projections onto a closed convex set in a Hilbert space are firmly nonexpansive.
2. T is nonexpansive if and only if $\frac{1}{2}(Id + T)$ is firmly nonexpansive

Fejér Monotonicity

Definition 3 (Fejér Monotone)

Where $S \subset H$ is nonempty, the sequence x_n is said to be Fejér *monotone* with respect to S if

$$(\forall y \in S)\ (\forall n \in \mathbb{N})\ \|x_{n+1} - y\| \leq \|x_n - y\|. \tag{2}$$

See, for example, [2].

Proposition 4

If D is a nonempty subset of H and $T : D \to D$ is nonexpansive with $\mathrm{Fix}\, T \neq \emptyset$ then the sequence $x_{n+1} = T(x_n)$ with $x_0 \in D$ is Fejér monotone with respect to $\mathrm{Fix}\, T$.

Connection of D-R with ADMM through Duality

- Where F and G are convex, proper, lsc and B is linear, consider the primal problem

$$\mathbf{p} := \inf_{v \in V} \{F(Bv) + G(v)\}.$$

- Under sufficient qualification conditions, we can solve **p** by solving

$$\mathbf{d} := \inf_{v^* \in V^*} \{G^*(-B^*v^*) + F^*(v^*)\}.$$

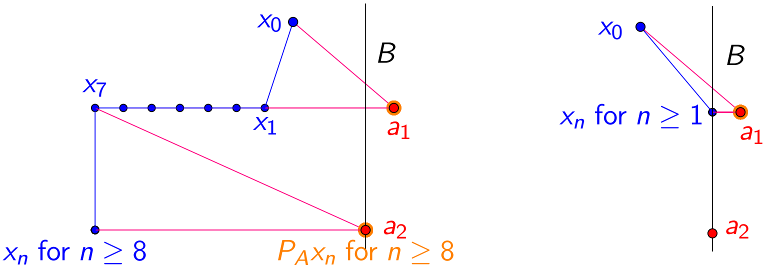
  See, for example, Borwein's & Lewis' *Convex Analysis and Nonlinear Optimization* [7, thm 3.3.5].

- Applying DR to **d** is equivalent to applying Uzawa's alternating direction method of multipliers [ADMM] to **p**. See Gabay's 1983 book chapter [16] and the survey [19].

Nonconvex setting: History and independent discovery

- Fienup independently discovered DR for nonconvex feasibility problems (phase retrieval) [15] and it has been popularized by Veit Elser.
- Other names, special instances, and generalizations:
  - Hybrid Input-Output algorithm (HIO), Fienup's variant, the "difference map" [14]
  - Averaged alternating reflections [4]
  - Relaxed reflect-reflect [13]

Advantages



DR and AP for a doubleton $A = \{a_1, a_2\}$ and line B in $\mathbb{R}^2$

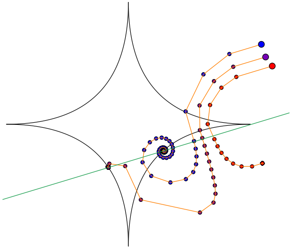
- Over-relaxed projection methods tend to explore the space more than AP
- Out-of-the-box solver: need only to be able to compute projections and can apply to any constraint satisfaction problem.

Theory is scarce

- Although often found to work well, theoretical underpinning for projection methods in the presence of nonconvex constraint sets is sadly lacking.
- Projections onto nonconvex sets are often set valued, and need no longer be firmly nonexpansive, or even nonexpansive.
- Local convergence established in certain instances (in particular near isolated feasible points for intersections of curves and hypersurfaces in $\mathbb{R}^n$) using theory of local asymptotic stability of almost linear discrete dynamical systems, and more globally utilizing Lyapunov functions. See for example; [9], [6], [8], [18], and [11]

Discrete/Combinatorial settings:

- Latin squares
- sudoku puzzles
- nonograms
- matrix completion
  - Hadamard matrices
  - Rank minimization
  - distance matrices
- matrix decomposition
- Wavelets with constraints
- 3-SAT
- graph coloring
  - edge colorings
  - 8 queens, 3-SAT, Hamiltonian paths
- Bit retrieval
- doubletons and lines (theory)



Connected constraints:

- Phase retrieval
- Intersections of plane curves and roots of functions
- solving nonlinear systems of equations
- Boundary value ODEs
- Regularity and transversality conditions (theory)

Two applications where the constraints are discrete finite sets

5	3			7				
6			1	9	5			
	9	8					6	
8				6				3
4			8		3			1
7				2				6
	6					2	8	
			4	1	9			5
				8			7	9

Figure: Solving sudoku puzzles - Elser [13]. Image source Wikimedia Commons [?sudokupicture]

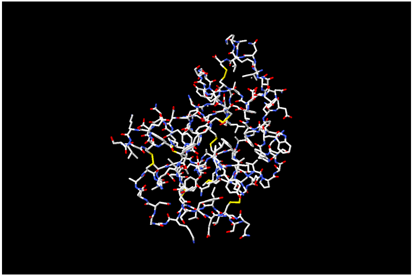


Figure: Solving incomplete Euclidean distance matrices for protein reconstruction [1].

Nonlinear ordinary differential equations with boundary values [18]

- Consider the problem

$$y'' = f(y', y, t) \text{ for } a \le t \le b \text{ with } y(a) = \alpha, y(b) = \beta$$

- For $N \in \mathbb{N}$, let $h := (b-a)/(N+1)$, and $t_i := a + hi$, for $i = 0, 1, \cdots, N+1$.
- So, $t_1, \cdots, t_N$ are a set of equally spaced nodes in $[a, b]$.
- Using centered differences we can obtain a numerical approximation to the solution of the BVP by taking $y(t_i) = \omega_i$, where $\omega_0 = \alpha$, $\omega_{N+1} = \beta$ and for $i = 1, \cdots, N$

$$\frac{\omega_{i-1} - 2\omega_i + \omega_{i+1}}{h^2} = f\left(t_i, \omega_i, \frac{\omega_{i+1} - \omega_{i-1}}{2h}\right) \quad \text{eqn(i)}$$

Reformulation as a feasibility problem

- Let
$$\Omega_i := \{\omega = (\omega_1, \ldots, \omega_N) | \omega \text{ satisfies eqn(i)}\}$$
then finding
$$\omega \in \cap_{k=1}^{N} \Omega_k$$
provides an approximate numerical solution to the BVP..

Computing projections

- For $u := (u_0, u_1, \cdots, u_{N+1})$ with $u_0 = \alpha$ and $u_{N+1} = \beta$,
- let $H_i(u) = -u_{i-1} + 2u_i - u_{i+1} + h^2 f\left(t_i, u_i, \frac{u_{i+1}-u_{i-1}}{2h}\right)$,
- then projection $z := P_{\Omega_i}(v)$ satisfies the Lagrangian:

$$H_i(z) = 0, \quad v - z + \lambda \nabla H_i(z) = 0.$$

  - so z can be computed by a steepest descent method or by Newton's method.

Example: Heaviside

- We consider

$$y''(x) = \begin{cases} -1 & y(x) < 0 \\ 1 & y(x) \geq 0, \end{cases} \tag{3}$$

with the boundary conditions $y(-1) = -1$ and $y(1) = 1$, which admits the unique continuous solution:

$$y(x) = \begin{cases} -\frac{1}{2}x^2 + \frac{1}{2}x & x < 0 \\ \frac{1}{2}x^2 + \frac{1}{2}x & x \geq 0. \end{cases} \tag{4}$$

Example: Heaviside

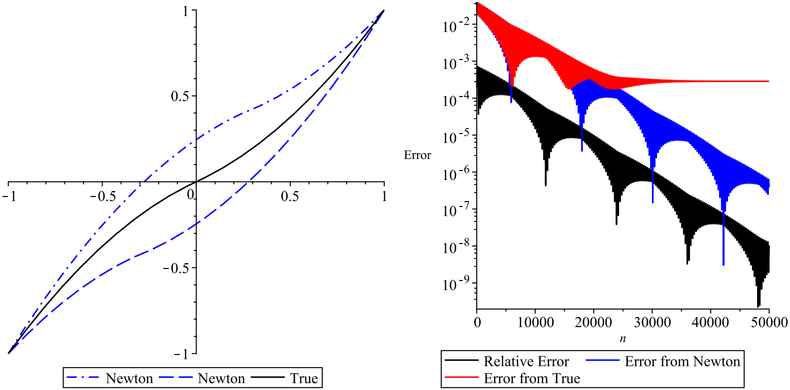


Figure: Newton's Method cycles for certain starting points (left) while DR converges (right).

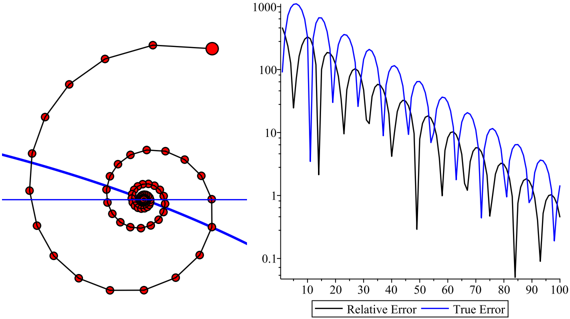


Figure: Relative error and error from true solution for converging DR iterates for an ellipse and line.

- The similarities to the previous figure are unmistakable,
- illustrating a frequent feature of the behaviour of D-R.

Thank You!

CARMA

THE UNIVERSITY OF NEWCASTLE AUSTRALIA

References I

[1] Francisco J. Aragón Artacho, Jonathan M. Borwein, and Matthew K. Tam, *Douglas-Rachford feasibility methods for matrix completion problems*, ANZIAM J. **55** (2014), no. 4, 299–326. MR3257300

[2] Heinz H. Bauschke and Patrick L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*, Second, CMS Books in Mathematics/Ouvrages de Mathématiques de la SMC, Springer, Cham, 2017. With a foreword by Hédy Attouch. MR3616647

[3] Heinz H. Bauschke, Patrick L. Combettes, and D. Russell Luke, *Phase retrieval, error reduction algorithm, and Fienup variants: a view from convex optimization*, J. Opt. Soc. Amer. A **19** (2002), no. 7, 1334–1345. MR1914365

[4] ______, *Finding best approximation pairs relative to two closed convex sets in Hilbert spaces*, J. Approx. Theory **127** (2004), no. 2, 178–192. MR2058156

References II

[5] Heinz H Bauschke, Dominikus Noll, and Hung M Phan, *Linear and strong convergence of algorithms involving averaged nonexpansive operators*, Journal of Mathematical Analysis and Applications **421** (2015), no. 1, 1–20.

[6] Joel Benoist, *The douglas-rachford algorithm for the case of the sphere and the line*, J. Glob. Optim. **63** (2015), 363–380.

[7] Jonathan M. Borwein and Adrian S. Lewis, *Convex analysis and nonlinear optimization: Theory and examples*, 2nd ed., Springer, 2006.

[8] Jonathan M. Borwein, Scott B. Lindstrom, Brailey Sims, Matthew Skerritt, and Anna Schneider, *Dynamics of the douglas-rachford method for ellipses and p-spheres*, Set-Valued Anal. (2016), 1–19.

[9] Jonathan M. Borwein and Brailey Sims, *The douglas-rachford algorithm in the absence of convexity*, Fixed point algorithms for inverse problems in science and engineering, 2011, pp. 93–109. MR2858834

References III

[10] Jonathan M. Borwein and Matthew K. Tam, *A cyclic douglas–rachford iteration scheme*, J. Optim. Theory Appl. **160** (2014), 1–29.

[11] Minh N. Dao and Matthew .K. Tam, *A lyapunov-type approach to convergence of the douglas-rachford algorithm* (2017).

[12] Jonathan Eckstein and Benar Fux Svaiter, *General projective splitting methods for sums of maximal monotone operators*, SIAM Journal on Control and Optimization **48** (2009), no. 2, 787–811.

[13] Veit Elser, Ti-Yen Lan, and Tamir Bendory, *Benchmark problems for phase retrieval*, arXiv preprint arXiv:1706.00399 (2017).

[14] Veit Elser, I. Rankenburg, and P. Thibault, *Searching with iterated maps*, Proc. Natl. Acad. Sci. USA **104** (2007), no. 2, 418–423.

[15] James R Fienup, *Phase retrieval algorithms: a comparison*, Applied optics **21** (1982), no. 15, 2758–2769.

References IV

[16] Daniel Gabay, *Chapter ix applications of the method of multipliers to variational inequalities*, Studies in mathematics and its applications, 1983, pp. 299–331.

[17] Simon Gravel and Veit Elser, *Divide and concur: A general approach to constraint satisfaction*, Physical Review E **78** (2008), no. 3, 036706.

[18] Bishnu P. Lamichhane, Scott B. Lindstrom, and Brailey Sims, *Application of projection algorithms to differential equations: boundary value problems*, arXiv preprint arXiv:1705.11032 (2017).

[19] Scott B. Lindstrom and Brailey Sims, *Survey: Sixty years of douglas-rachford* (2018 (in preparation)).

[20] P.-L. Lions and B. Mercier, *Splitting algorithms for the sum of two nonlinear operators*, SIAM J. Numer. Anal. **16** (1979), no. 6, 964–979. MR551319

[21] Guy Pierra, *Decomposition through formalization in a product space*, Mathematical Programming **28** (1984), no. 1, 96–115.

References V

[22] Jonathan E Spingarn, *Partial inverse of a monotone operator*, Applied mathematics and optimization **10** (1983), no. 1, 247–265.

[23] Benar F. Svaiter, *On weak convergence of the douglas–rachford method*, SIAM J. on Control and Opt. **49** (2011), no. 1, 280–287.