A primal/dual computable approach to improving spiraling algorithms, based on minimizing spherical surrogates for Lyapunov functions

Scott B. Lindstrom
Curtin University ... (finally!)

Va & Opt Seminar, 2 June, 2021

Last Revised May 30, 2021

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

Curtin University

Outline

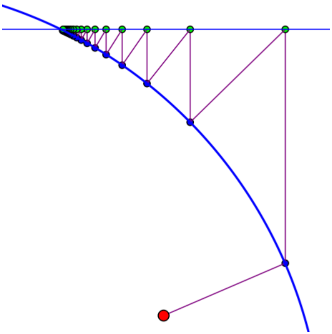
00: Splitting Algorithms

Problem:

$$\underset{x \in E}{\text{minimize}} \quad f(x) + g(z) \quad \text{s.t. } Mx = z.$$

Splitting Methods break a problem into more easily computable subproblems by:

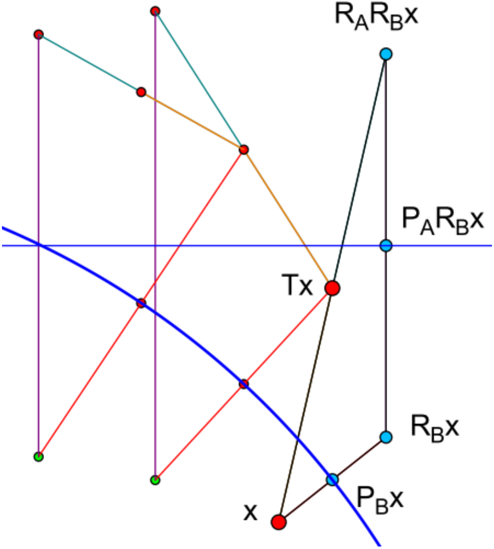
1. decoupling constraints that are difficult to simultaneously satisfy
2. decoupling minimization steps that are difficult to simultaneously compute
3. decoupling a constraint satisfaction step from a minimization step



01: Douglas–Rachford

Popular variants include:

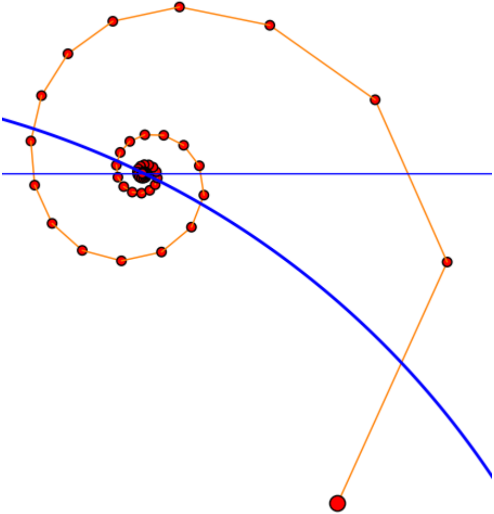
1. ADMM and its dual Douglas–Rachford (right). Recent survey of DR by Lindstrom and Sims (2018)
2. Proximal gradient (forward–backward method), (e.g. alternating projections)
3. FISTA



02: Spiraling Algorithms

Algorithms that may spiral include:

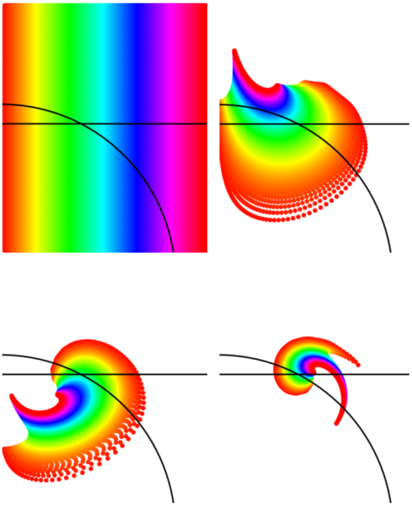
1. Douglas–Rachford (dual to ADMM)
2. FISTA



03: Dynamical systems

Case study: sphere and line Douglas–Rachford:

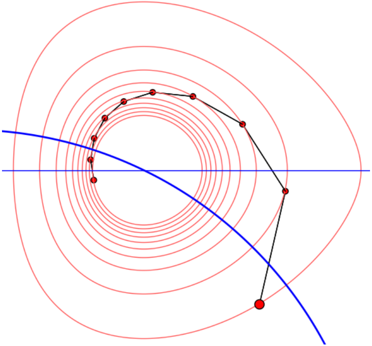
- Borwein and Sims (2011) conjectured global, proved local
- Aragón Artacho and Borwein (2013) provided partial answers
- Borwein (2016) furnished images similar to right



04: Lyapunov Functions

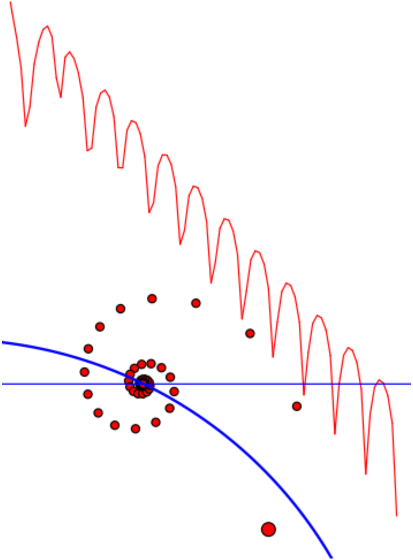
Case study: sphere/line DR:

- Benoist (2015) constructed Lyapunov function
- Borwein et al. (2018) showed global results don't extend to sphere generalizations
- Dao and Tam (2019); Giladi and Rüffer (2019) extended Benoist's framework to more general graphs locally
- Asymptotic stability essentially guarantees existence of such functions (Kellett and Teel, 2005; Giladi and Rüffer, 2019; Lindstrom, 2020)



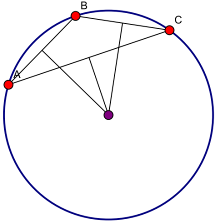
05: Oscillations

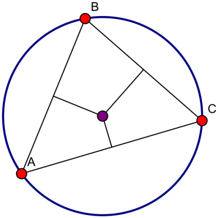
- When spiraling, plotting the changes in projections onto line (shadow sequence) produces smooth arches and sharp valleys
- Same pattern observed in many contexts. Discussed in (Lamichhane et al., 2019).



06: Circumcenter Construction

- The circumcenter of a 3-tuple (A, B, C) is the center of a circle that contains A, B, C.
- It lies at the intersection of the perpendicular bisectors of triangle ABC
- Exists finitely whenever A, B, C not both distinct and colinear.



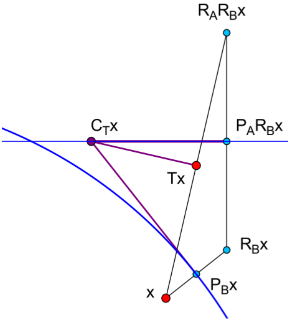


07: Circumcentering Reflections Method (CRM)

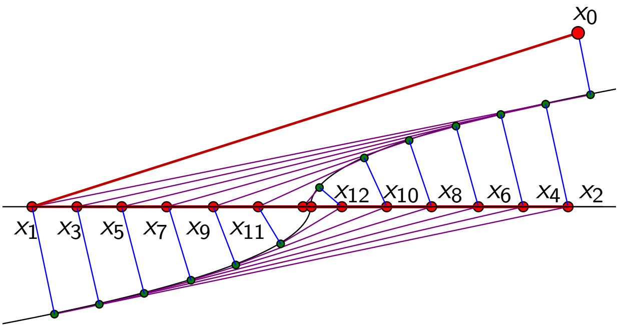
- Behling et al. (2018b,a) introduced CRM: $C_T x = C(x, R_A x, R_A R_B x)$.
- Bauschke et al. (2018a,b): sufficient conditions guarantee *properness* of operator (existence of updates)
- Dizon et al. (2019) introduced GCRM:

$$\mathcal{C}_T : x \mapsto \begin{cases} T_{A,B}x & \text{if colinear} \\ C_T x & \text{otherwise} \end{cases}.$$

  Also used subspace invariance to show local quadratic convergence with plane curves & lines.
- Behling et al. (2019): convex, product space formulation global convergence

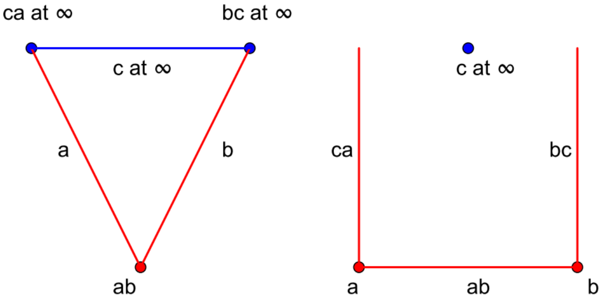


08: Newton–Raphson characterization



- Perpendicular bisectors of reflections across hyperplanes are the hyperplanes
- Perpendicular bisectors of reflections across curves are tangents
- When one set is a hyperplane and one set is a graph, CRM "resembles" Newton–Raphson (Dizon et al., 2019).

09: Duality



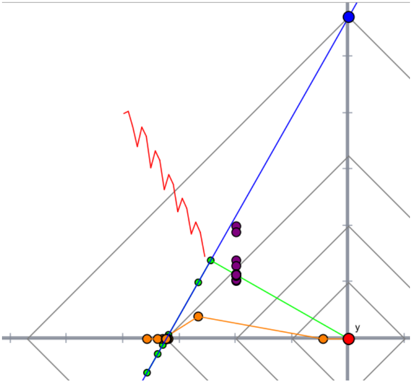
- Renaissance geometry: projective duality of lines and points.
- Convex optimization: Fenchel–Moreau–Rockafellar duality of epigraphs of convex functions.
- Duality of algorithms Gabay (1983):
  - ADMM: $\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)+g(z) \quad \text{s.t. } Mx = z \in \mathbb{R}^m.$
  - DR: $\underset{\lambda\in\mathbb{R}^m}{\text{minimize}} \quad f^* \circ (-M^T)(\lambda) + g^*(\lambda)$

10: ADMM and Basis Pursuit

Basis pursuit:
minimize $\|x\|_1$ s.t. $Ax = b,\ x \in \mathbb{R}^n,\ A \in \mathbb{R}^{\nu\times n},\ b \in \mathbb{R}^\nu,\ \nu < n.$

ADMM:

$$f := \iota_S,\quad S := \{x \in \mathbb{R}^n \mid Ax = b\},$$
$$M := \mathrm{Id},\quad g : z \to \|z\|_1,\quad E, Y := \mathbb{R}^n.$$
$$x_{k+1} \in \operatorname*{argmin}_{x\in\mathbb{R}^n}\Big\{f(x) + g(z_k) + \langle \lambda_k, Mx - z_k\rangle + \frac{c}{2}\|Mx - z\|^2\Big\}$$
$$z_{k+1} \in \operatorname*{argmin}_{z\in\mathbb{R}^m}\Big\{f(x_{k+1}) + g(z) + \langle \lambda_k, Mx_{k+1} - z\rangle + \frac{c}{2}\|Mx_{k+1} - z\|^2\Big\}$$
$$\lambda_{k+1} = \lambda_k + c(Mx_{k+1} - z_{k+1}).$$



11: Primal/Dual Algorithms

Set: $d_1 := f^* \circ (-M^T), d_2 := g^*$

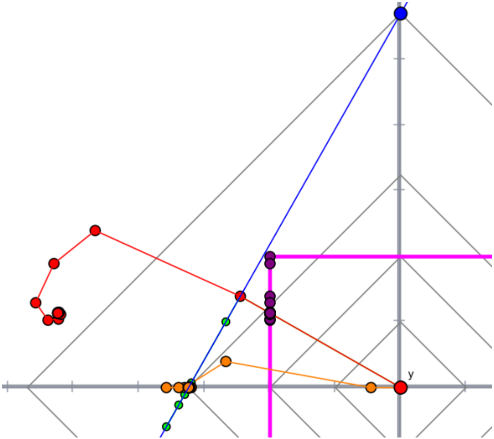
Primal reconstruction:

$$cMx_{k+1} = \mathrm{prox}_{cd_1}(R_{cd_2} y_k) - R_{c\partial d_2}(y_k)$$
$$cz_k = y_k - \mathrm{prox}_{cd_2} y_k$$
$$\lambda_k = \mathrm{prox}_{cd_2} y_k$$

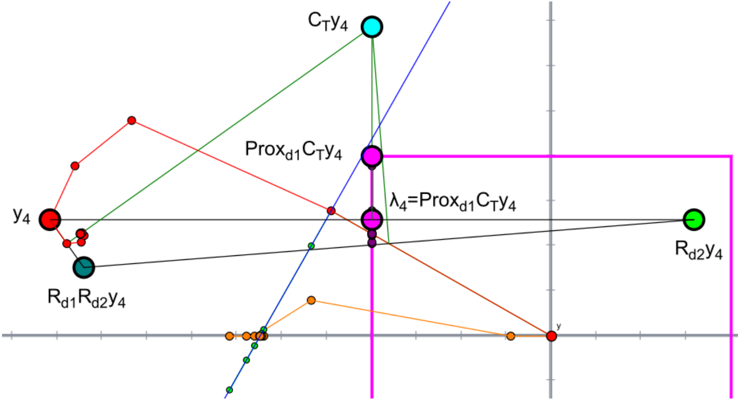
Dual reconstruction:

$$y_k = \lambda_k + cz_k$$
$$R_{cd_2} y_k = \lambda_k - cz_k$$
$$R_{cd_1} R_{cd_2} y_k = \lambda_k - cz_k + 2cMx_{k+1}$$

See (Eckstein and Yao, 2015).



12: Failure of CRM for primal/dual

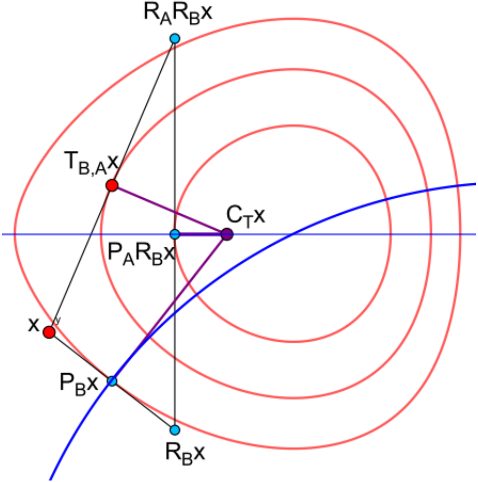


For basis pursuit, $\mathrm{prox}_{d_2} = \mathrm{proj}_{B_\infty}$, so the perpendicular bisector of $(y_k, R_{cd2} y_k)$ contains a face of B_∞, so CRM generically fails.

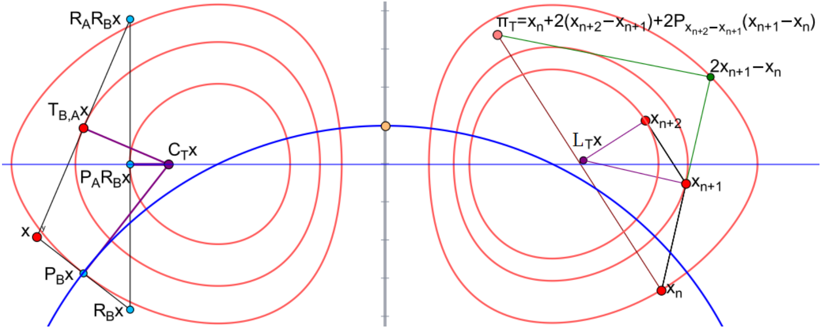
13: CRM and Lyapunov functions

Theorem 1 (Lindstrom (2020))

For Lyapunov functions V from Benoist (2015); Dao and Tam (2019), CRM updates lie on the intersections of hyperplanes that contain subgradient descent trajectories of V from the points P_Bx, P_AR_Bx, *and* $T_{B,A}x$. *If the set order is reversed, the analogous result holds with the points* P_Ax, P_BR_Ax, *and* $T_{A,B}x$.



14: New method based on Lyapunov functions



Theorem 2 (Lindstrom (2020))

Whenever $(\forall x \in U)\ \langle \nabla V(x^+), x - x^+ \rangle = 0$ *for Lyapunov function* V*:*

$$L_T x := \begin{cases} C(x, 2x^+ - x, \pi_T x), & \text{if } x, 2x^+ - x, \pi_T x \text{ are not colinear;} \\ Tx^{++} & \text{otherwise} \end{cases}$$

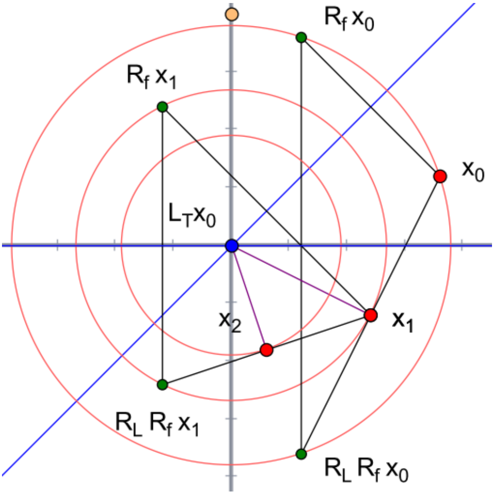
lies on the intersections of hyperplanes that contain subgradient descent trajectories of V *from* x^+ *and* x^{++}.

15: An interesting example

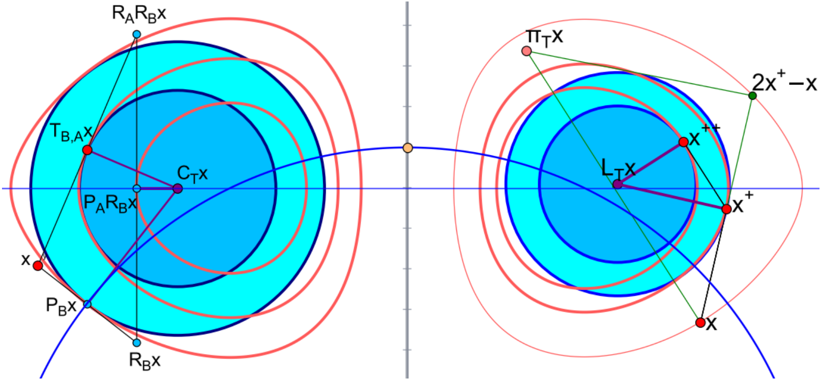
Proposition 1 (Lindstrom (2020))

When $T_{A,B}$ is the Douglas–Rachford projection operator and A, B are lines, $L_{T_{A,B}}x$ is a fixed point for any x.

- Easy to prove: all subgradient descent trajectories from all points intersect in the fixed set
- Another view: any quadratic surrogate that matches gradients of V at the sample points will be "equal" to V



16: Quadratic surrogate characterization



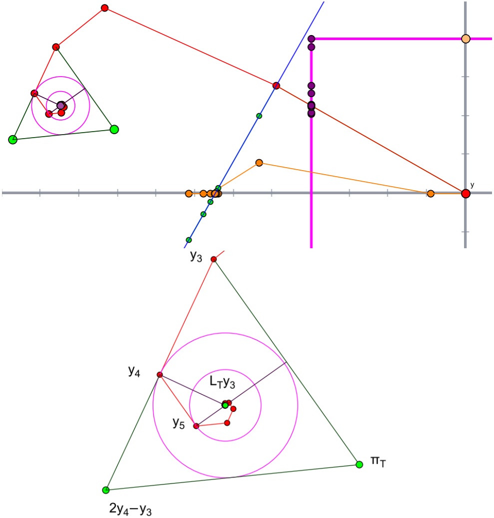
Theorem 3 (Lindstrom (2020))

Under appropriate assumptions, L_T and C_T return minimizers of quadratic surrogates for Lyapunov functions that describe the convergence of algorithms admitted by iteratively applying the operator T (where T is Douglas–Rachford specifically in the C_T case).

17: Primal/dual centering

Primal/dual acceleration of ADMM/DR:

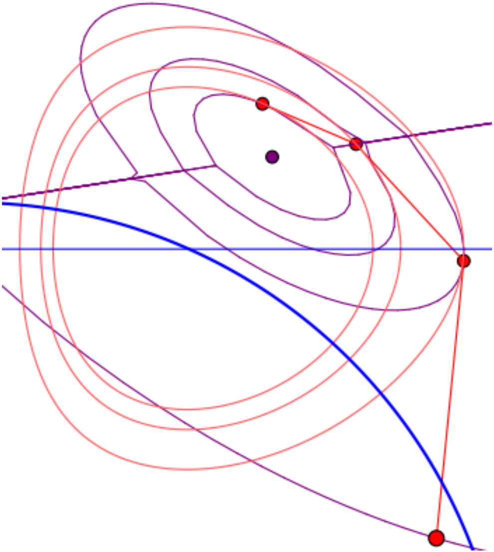
1. Compute primal, reconstruct dual
2. Apply L_T to dual
3. Propagate centered step update back to ADMM variables
4. Check objective function against improvement from regular update, pick winning candidate and repeat



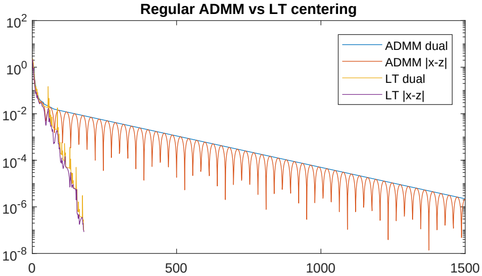
,

18: More general quadratic surrogate minimization

1. One could also consider a spherical surrogate built from more points (higher dimensions)...
2. or adaptive algorithms that compare and choose from among surrogate-offered update candidates...
3. Or a more general elliptical (not just spherical) surrogate...



19: Computational Evidence



For 1,000 problems of dimension 300, L_T solved every problem:

	wins	min	Q1	median	Q3	max
vanilla ADMM	1	278	995	1582	2932	761,282
L_T-centering	999	87	165	221	324	94,591

20: Summary

Theoretical contributions:

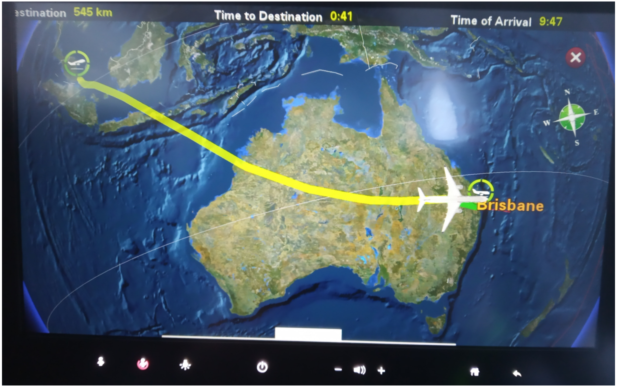
- CRM: Lyapunov characterization (including nonconvex cases)
- New method L_T:
  1. problem agnostic: generically implementable
  2. *operator* agnostic: ADMM, DR, FISTA, Nesterov...
  3. does not require subproblems: black box compatible
  4. satisfies Lyapunov properties with remarkably few structural assumptions: primal/dual adaptable
  5. not convexity reliant: good candidate for online optimization
- Both methods: quadratic surrogacy characterization

General lessons:

- Use tools like *Cinderella* and *Geogebra*
- Excursions into the weeds (e.g. highly specific problems/ structure) yield insights about more general problems
- Look for bridges between seemingly disparate areas of research (e.g. Lyapunov functions, centering)

Thank you!

Scott B. Lindstrom, "Computable centering methods for spiraling algorithms and their duals, with motivations from the theory of Lyapunov functions." arXiv preprint arXiv:2001.10784, (2020).



Coming to Australia, 3 May 2021.

References I

Francisco J. Aragón Artacho and Jonathan M Borwein. Global convergence of a non-convex Douglas–Rachford iteration. *Journal of Global Optimization*, 57(3):753–769, 2013.

Heinz H Bauschke, Hui Ouyang, and Xianfu Wang. On circumcenter mappings induced by nonexpansive operators. *arXiv preprint arXiv:1811.11420*, 2018a.

Heinz H Bauschke, Hui Ouyang, and Xianfu Wang. On circumcenters of finite sets in Hilbert spaces. *arXiv preprint arXiv:1807.02093*, 2018b.

Roger Behling, José Yunier Bello-Cruz, and L-R Santos. On the linear convergence of the circumcentered-reflection method. *Operations Research Letters*, 46(2):159–162, 2018a.

Roger Behling, José Yunier Bello Cruz, and Luiz-Rafael Santos. Circumcentering the Douglas–Rachford method. *Numerical Algorithms*, 78:759–776, 2018b.

References II

Roger Behling, José Yunier Bello-Cruz, and L-R Santos. On the circumcentered-reflection method for the convex feasibility problem. *arXiv preprint arXiv:2001.01773*, 2019.

Joel Benoist. The Douglas–Rachford algorithm for the case of the sphere and the line. *J. Glob. Optim.*, 63:363–380, 2015.

Jonathan M. Borwein. The life of modern homo habilis mathematicus: Experimental computation and visual theorems. In *Tools and Mathematics*, volume 347 of *Mathematics Education Library*, pages 23–90. Springer, 2016.

Jonathan M. Borwein and Brailey Sims. The Douglas–Rachford algorithm in the absence of convexity. In Heinz H. Bauschke, Regina S. Burachik, Patrick L. Combettes, Veit Elser, D. Russell Luke, and Henry Wolkowicz, editors, *Fixed Point Algorithms for Inverse Problems in Science and Engineering*, volume 49 of *Springer Optimization and Its Applications*, pages 93–109. Springer Optimization and Its Applications, 2011.

References III

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.*, 26(2):385–403, 2018.

Minh N. Dao and Matthew K. Tam. A Lyapunov-type approach to convergence of the Douglas–Rachford algorithm. *J. Glob. Optim.*, 73(1):83–112, 2019.

Neil Dizon, Jeffrey Hogan, and Scott B Lindstrom. Circumcentering reflection methods for nonconvex feasibility problems. *arXiv preprint arXiv:1910.04384*, 2019.

Jonathan Eckstein and Wang Yao. Understanding the convergence of the alternating direction method of multipliers: Theoretical and computational perspectives. *Pac. J. Optim.*, 11(4):619–644, 2015.

References IV

Daniel Gabay. Applications of the method of multipliers to variational inequalities. In *Studies in mathematics and its applications*, volume 15, chapter ix, pages 299–331. Elsevier, 1983.

Ohad Giladi and Björn S Rüffer. A lyapunov function construction for a non-convex douglas–rachford iteration. *Journal of Optimization Theory and Applications*, 180(3):729–750, 2019.

Christopher M Kellett and Andrew R Teel. On the robustness of \mathcalkl-stability for difference inclusions: Smooth discrete-time Lyapunov functions. *SIAM Journal on Control and Optimization*, 44(3):777–800, 2005.

Bishnu P. Lamichhane, Scott B. Lindstrom, and Brailey Sims. Application of projection algorithms to differential equations: boundary value problems. *The ANZIAM Journal*, 61(1):23–46, 2019.

References V

Scott B. Lindstrom. Computable centering methods for spiraling algorithms and their duals, with motivations from the theory of Lyapunov functions. *arXiv preprint arXiv:2001.10784*, 2020.

Scott B. Lindstrom and Brailey Sims. Survey: Sixty years of Douglas–Rachford. *J. AustMS (to appear), arXiv preprint arXiv:1809.07181*, 2018.