Douglas-Rachford - *60 years young*

Brailey Sims
CARMA, University of Newcastle, AUSTRALIA

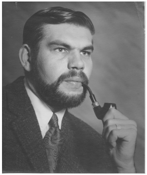
nao2018, Okinawa, 5-9 November 2018

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.


Feasibility (constraint satisfaction) problems

- For closed constraint sets C_i, $i = 1, 2, \cdots, n$, in a Hilbert space H, find a feasible point

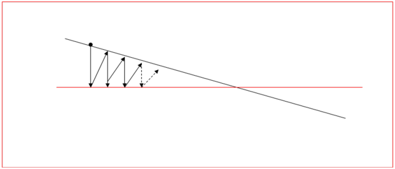
$$x \in C \ := \ \cap_{i=1}^{n} C_i$$

- 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, introduced by J. von Neumann in 1933, is the oldest such method:

(AP) From an initial guess $x_0 \in H$, form the iterative sequence

$$x_{n+1} \ = \ P_{C_n} P_{C_n - 1} \cdots P_{C_1} x_n$$

Feasibility problems



(1) When all C_i are affine sets, von Neumann showed that $x_n \to P_C x_0$

(2) When all C_i are convex sets Bregman [1965] established that $x_n \rightharpoonup P_C x_0$.

(3) In 2002 Hundal gave an example showing that here weak convergence cannot in general be replace by norm convergence

Two set feasibility problems

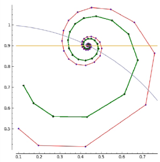
- For two set feasibility problems, another effective method, on which we focus, was introduced by J. Douglas and H. H. Rachford in 1956.

(D-R) For closed constraint sets A, B and initial guess $x_0 \in H$ form the iterative sequence

$$x_{n+1} := \frac{1}{2}(Id + R_B R_A)(x_n)$$

where Id is the identity operator on H and, for a closed set C, $R_C := 2P_C - Id$ is the operator of reflection in C.

Douglas-Rachford



Douglas-Rachford

(1) When A and B are both convex sets $x_n \rightharpoonup x_\infty \in H$ with $P_A x_\infty \in A \cap B$

(2) While norm convergence is only ensured when, for example, H is finite dimensional no example such as that of Hundal for AP seems known.

One iteration of AP and D-R

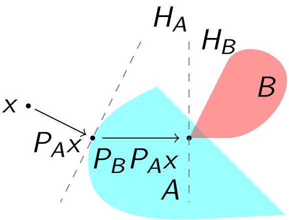


Figure: AP; $T_{A,B} := P_B P_A$

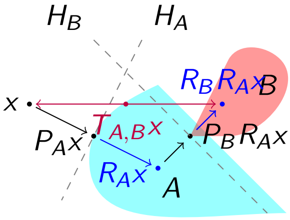


Figure: D-R;
$T_{A,B} := \frac{1}{2}(Id + R_B R_A)$

- Projection algorithm: from prescribed x_0 iterate $x_{n+1} = T_{A,B}(x_n)$

Zeros of sums of maximally monotone operators

D-R was originally considered as a method for locating a zero for a sum of two monotone operators $\mathcal{A}$ and $\mathcal{B}$ on H.

In this context Lion and Mercier

proved in 1979

Zeros of sums of maximally monotone operators

Theorem 1 (Lions & Mercier)

If $\mathcal{A}$ and $\mathcal{B}$ are maximal monotone operators on a H, with $\mathcal{A}+\mathcal{B}$ also maximal monotone, then for

$$T_{\mathcal{A},\mathcal{B}}x \;:=\; J_{\mathcal{B}}(2J_{\mathcal{A}} - Id)x + (Id - J_{\mathcal{B}})x,$$

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

Connection to the 2-set feasibility problem

- Recall: the *indicator function* for a set C is

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

- So, we see that

$$x \in A \cap B \iff x \text{ minimizes } \iota_{A\cap B} = \iota_A + \iota_B$$

Thus, $$\iff 0 \in (\partial \iota_A + \partial \iota_B)(x).$$

- Recalling that for a convex function f the subdifferential ∂f is a maximal monotone operator we have that

Connection to the 2-set feasibility problem

- we can apply Lions-Mercier's result with $\mathcal{A} := \partial\iota_A$ and $\mathcal{B} := \partial\iota_B$, in which case,
  1. $\mathcal{A} = N_A$, the normal cone operator

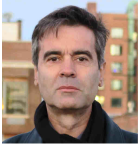
$$x \mapsto \begin{cases} \{y | (y, a - x) \leq 0 \text{ for all } a \in A\} & \text{if } x \in C \\ \emptyset & \text{otherwise,} \end{cases}$$

  2. Similarly, $\mathcal{B} = N_B$, and then
  3. $J_{\mathcal{A}} := (\mathrm{Id} + \mathcal{A})^{-1} = (\mathrm{Id} + N_A)^{-1} = P_A$, and
  4. $J_{\mathcal{B}} := (\mathrm{Id} + \mathcal{B})^{-1} = (\mathrm{Id} + N_B)^{-1} = P_B$,
- So, in the Lions-Mercier's algorithm,

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

Remarks

- In 2002 Bauschke, Combettes, and Luke gave a direct proof (employing essentially the same ingredients as those used by Lions and Mercier) for the weak convergence of the iterates of $\frac{1}{2}\left(Id + R_B R_A\right)$.


- The requirement $\mathcal{A} + \mathcal{B}$ be maximally monotone was relaxed outside the feasibility setting by Svaiter in 2011.

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

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

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 $\mathrm{Fix}\, T$.

D-R is connected with ADMM through Duality

- For F and G convex, proper, lsc functions and B a linear, the primal problem

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

- can, under suitable qualification conditions, be solved **p** by solving the dual problem

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

  See, for example, Borwein's & Lewis' *Convex Analysis and Nonlinear Optimization*, theorem 3.3.5).
- Applying DR to **d** is equivalent to applying Uzawa's *alternating direction method of multipliers* [ADMM] to **p** [see Gabay, chapter (ix), of *Studies in mathematics and its application*, 1983, and our survey].

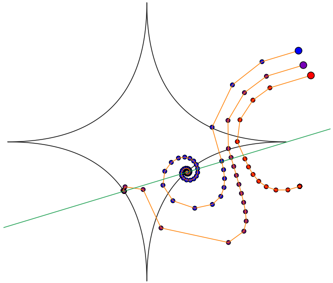
From 2 sets to N sets.

- We can use Douglas-Rachford on a feasibility problem involving N sets; $\Omega_1 \dots \Omega_N$, to find $x \in \cap_{k=1}^{N} \Omega_k$ by utilizing Pierra's product space method; that is, by applying the algorithm in $\mathcal{H}^N$ to the two sets
  - $A := \Omega_1 \times \cdots \times \Omega_N$
  - $B := \{x = (y_1, \dots, y_N) | y_1 = y_2 = \cdots = y_N\}$
- Nicknamed divide and concur by Simon Gravel and Veit Elser (the latter credits the former for the name) [?GE].
  - 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 the cyclically anchored variant (CADRA) and the Borwein-Tam method (cyclic D-R).

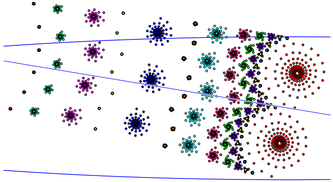
D-R in the non-convex cases

- Despite a dearth of supporting theory, relaxed projection methods, and D-R in particular, have proved effective, and hence a popular off-the-shelf solver, for handling feasibility problems involving non-convex (including discrete) constraints provided the relevant projections can be computed.
- A contributing factor to this success may be tendency of D-R to better explore the solution space; often exhibiting spiral trajectories rather than the more monotone approach to equilibrium commonly seen with AP.

D-R in two non-convex settings



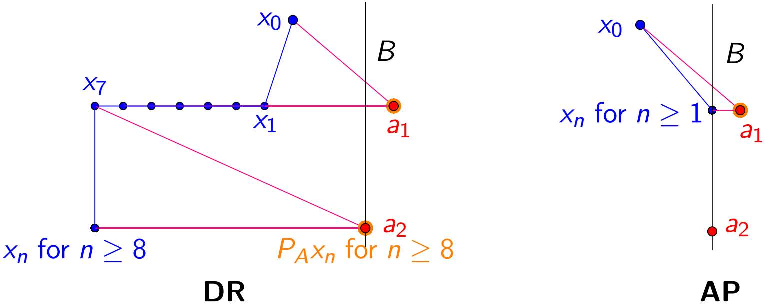
Trajectories for line–(1/2)-sphere



Singular set for line–ellipse

Discrete constraint sets

- The advantage of D-R over AP in the presence of discrete constraint sets, is nicely illustrated in the case of a doubleton $A = \{a_1, a_2\}$ and a line B in $\mathbb{R}^2$.



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. Image source Wikimedia Commons

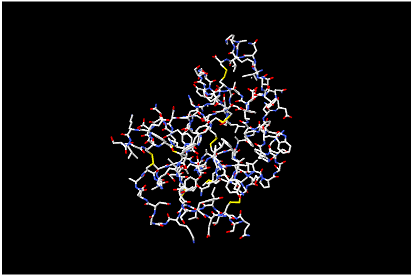


Figure: Solving incomplete Euclidean distance matrices for protein reconstruction.

Nonconvex setting: History and Theory

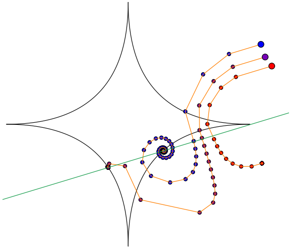
- Fienup independently discovered DR, using it for nonconvex feasibility problems (phase retrieval) in 1982 and more recently it has been popularized by Veit Elser [2007, 2008], Borwein, Tan, and Aragón Artacho, among others.
- Other names, special instances, and generalizations:
  - Hybrid Input-Output algorithm (HIO), Fienup's variant, the "difference map"
  - Averaged alternating reflections
  - Relaxed reflect-reflect [2017]

Theory is scarce

- Although often found to work well, theoretical underpinning for projection methods in the presence of nonconvex constraint sets is sorely 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.

Discrete/Combinatorial settings:

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



Connected constraints:

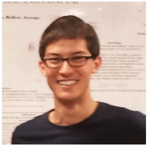
- Phase retrieval
- Intersections of plane curves and roots of functions
- solving nonlinear systems of equations
- Boundary value ODEs
- Regularity and transversality conditions (theory)

Wavelet construction via projection methods

Material extracted from:

1. David Franklin, *Projection Algorithms for Non-separable Wavelets and Clifford Fourier Analysis*, PhD dissertation, University of Newcastle, October 2018, supervisor: Jeff Hogan,

and represents joint work by David, Jeff and Matt Tam.



1. While the real goal, successfully implemented, is to use projection methods to construct higher dimensional wavelets with desirable properties, we content ourselves by illustrating the ideas in 1 dimension

Wavelets

1. **Recall:** A 1-dimensional wavelet is a function $\psi : \mathbb{R} \to \mathbb{R}$ whose dyadically dilated integer translates form an orthonormal basis for $L_2(\mathbb{R}, \mathbb{C})$; that is,

$$\left\{ \psi_{j,k}(x) := 2^{-j/2}\psi(x/2^j - k) : j \in \mathbb{Z}, k \in \mathbb{Z} \right\}$$

is an orthonormal basis for $L_2(\mathbb{R}, \mathbb{C})$.

2. The name *wavelet* derives from $\int_{\mathbb{R}} \psi(x)dx = 0$, so the amount of ψ above 0, 'sea-level', is balanced by that below - required for the sequence of *Fourier coefficients*, $(\langle f, \psi_{j,k} \rangle)$, to be in $\ell_2(\mathbb{Z} \times \mathbb{Z})$.

3. We aim to construct a wavelet with desired properties, in particular with compact support.

The scaling function

1. we begin by seeking a suitable *scaling function*; a function $\phi \in L_2(\mathbb{R}, \mathbb{C})$ with $\{\phi(x-k) : k \in \mathbb{Z}\}$ an orthonormal set and $\phi(x/2)$ in the subspace it generates; that is,

$$\phi(x/2) = \sum_{k \in \mathbb{Z}} a_k \phi(x-k), \quad \text{for some } (a_k) \in \ell_2(\mathbb{Z}). \tag{1}$$

2. Or equivalently, taking the Fourier transform of both sides,

$$\hat{\phi}(2\xi) \;=\; m_0(\xi)\hat{\phi}(\xi), \quad \text{where } m_0(\xi) := \frac{1}{2} \sum_{k \in \mathbb{Z}} a_k e^{-2\pi i k \xi} \tag{2}$$

3

Proposition 5

ϕ is compactly supported (without loss of generality on $[0, M-1]$) iff m_0 is a trigonometric polynomial (of degree $M-1$); specifically,

$$m_0(\xi) := \frac{1}{2}\sum_{k=0}^{M-1} a_k e^{-2\pi i k\xi}.$$

From this, the requirement that $\{\phi(x-k) : k \in \mathbb{Z}\}$ is orthonormal, and Plancherel's theorem we can deduce

Corollary 6

$$m_0(0) = |\hat{\phi}(0)| \;=\; 1, \quad \mathit{and} \quad |m_0(\xi)|^2 \;+\; |m_0(\xi + 1/2)|^2 \;=\; 1.$$

Cascade algorithm – determining ϕ from m_0

1. Importantly, proposition 5 and the scaling equation (1) allows us to determine ϕ from a knowledge of m_0 via the cascade algorithm:
   1. For ϕ supported on $[0, M-1]$, let $\mathbf{v} = (\phi(0), \phi(1), \cdots, \phi(M-1))$ then (1) reduces to $\mathbf{v} = A\mathbf{v}$ where

$$A_{ij} = \begin{cases} a_{2i-j} & \text{when } 2i - j \in \{0, \dots, M-1\}, \\ 0 & \text{otherwise.} \end{cases} \tag{3}$$

   2. So, finding an eigenvector of A corresponding to the eigenvalue 1 determines ϕ on $\mathbb{Z}$.
   3. Now, for each n and $\ell \in \mathbb{N}$, successive application of (1) determines $\phi(n2^{-\ell})$. Since the dyadic rationals are dense in $\mathbb{R}$, this is enough to specify a continuous ϕ on $[0, M-1]$.

Construction of ψ from ϕ

1. Having constructed ϕ from knowing m_0 we build our wavelet by taking $\psi(x/2)$ to lie in the subspace generated by $\{\phi(x-k) : k \in \mathbb{Z}\}$ leading to an analogue of the scaling equation (1) for ψ,

$$\psi(x/2) = \sum_{k\in\mathbb{Z}} b_k \phi(x-k), \quad \text{for some } (b_k) \in \ell_2(\mathbb{Z}), \tag{4}$$

ensuring ψ inherits many properties from ϕ.

2. Taking the Fourier transform of both sides,

$$\hat{\psi}(2\xi) \;=\; m_1(\xi)\hat{\phi}(\xi), \quad \text{where } m_1(\xi) := \frac{1}{2}\sum_{k\in\mathbb{Z}} b_k e^{-2\pi i k\xi}$$

Compact support

1. As for *phi* we deduce that
   1. ψ is compactly supported on $[0, M-1]$ iff m_1 is a trigonometric polynomial;

      $$m_1(\xi) := \frac{1}{2}\sum_{k=0}^{M-1} b_k e^{-2\pi i k\xi}.$$

   2. In which case, we can use (4) in the cascade algorithm to determine ψ if m_0 (and hence ϕ) and m_1 are known.
   3. So, our construction of a wavelet is reduced to determining suitable functions m_0 and m_1.

Constraints on m_1

1. We also deduce,

$$|m_1(\xi)|^2 \ + \ |m_1(\xi+1/2)|^2 \ = \ 1, \tag{5}$$

$$\overline{m_0(\xi)}m_1(\xi) + \overline{m_0(\xi+1/2)}m_1(\xi+1/2) \ = \ 0, \tag{6}$$

Matrix formulation

1. Define the matrix-valued function

$$U(\xi) = \begin{bmatrix} m_0(\xi) & m_1(\xi) \\ m_0(\xi + 1/2) & m_1(\xi + 1/2) \end{bmatrix} \tag{7}$$

   Then we see that the conclusion of corollary (6) and conditions (5) and (6) will be satisfied if and only if $U(\xi)$ is unitary for all ξ.

2. Further

$$U(0) = \begin{bmatrix} 1 & 0 \\ 0 & z \end{bmatrix}, \qquad \text{where } z = m_1(1/2), \text{ so } |z| = 1 \tag{8}$$

In this guise, our quest to construct a wavelet is reduced to finding a function-valued 2×2 matrix $U(\xi)$ such that:

1. $U(\xi + 1/2) = JU(\xi)$ where J is the elementary matrix
$$J = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}$$
ensures the structure of $U(\xi)$ given in (7),
2. $U(\xi)$ is unitary for all ξ,
3. $U(0) = \begin{bmatrix} 1 & 0 \\ 0 & z \end{bmatrix}$,
4. $\left.\dfrac{d^l}{d\xi^l} U(\xi)\right|_{\xi=0}$ is diagonal for $0 \le l \le D$
ensures that ψ has its first D derivatives continuous and bounded,
5. $m_0(\xi)$ has no zeros on the set $[-\frac{1}{4}, \frac{1}{4}]$
ensures the integer shifts of ϕ are orthogonal.

Back to a feasibility problem

A FEASIBILITY PROBLEM!

The feasibility problem

1. Since the entries of $U(\xi)$ are trigonometric polynomials of degree M, so of the form

$$U(\xi) = \sum_{k=0}^{M-1} A_k e^{2\pi i k \xi}, \quad \text{with } A_k \in \mathbb{C}^{2\times 2},$$

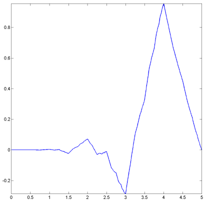
it $U(\xi)$ is completely determined by its values at $2M-1$ points in $[0,1)$.
So, the problem can be further reduced to considering a solution space whose elements are ensembles of $2M-1$ 2×2 constant complex matrices.

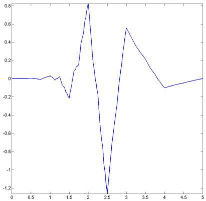
2. Further projections onto the sets corresponding to constraints (1) to (4) are readily computed. For example, if $A \in \mathbb{C}^{2\times 2}$ has singular value decomposition $A = USV$ then $P(A) := UV^*$ is a projection onto the set of unitary matrices.

The feasibility problem

1. Thus, our problem is amenable to the use of projection algorithms, in particular AP and D-R, with different starting points leading to potentially different wavelets.
2. Note: we choose to check that constraint (5) is satisfied post-priori.
3. This has been implemented in MATLAB and numerous trials undertaken.

Experimental results





Scaling function $\phi(x)$ and associated wavelet $\psi(x)$ with support $0 \leq x \leq 5$, discovered by product D-R

Experimental results

1. Trials were conducted using an ensemble of 100 randomly generated starting points in $\left(\mathbb{C}^{2\times 2}\right)^M$ and with $M = 6$.
2. In all 100 trials, product D-R converged, leading to the construction of a wavelet.
3. This should be contrasted to a 57% success rate for *proximal alternating linear minimization* – the most commonly used of the traditional methods (and a 54% success rate for AP).

References I