Visualizing Projection Algorithms for Inverse Problems

F.J. Aragón Artacho[1] · J.M. Borwein[2] · **Matthew K. Tam**[2]

[1] Systems Biochemistry Group, University of Luxembourg, Luxembourg

[2] CARMA Centre, University of Newcastle, Australia

1. Introduction

- Proteins are large biomolecules which are comprised of several amino acid residues. They participate in virtually every cellular process with a typical protein consists of 200–300 residues.
- A knowledge of structural conformation gives important insights into the mechanism by which a given protein performs its function. Nuclear magnetic resonance (NMR) imaging is a non-destructive technique which resolves small inter-atom distances (*i.e.*, those less than 6Å).
- Can conformation be predicted using only these short-range distances?

2. Basic Tools: Projections & Reflections

- The projection of a point x onto a set C is given by $P_C(x) = \arg\min\{\|x - c\| : c \in C\}$.
- The reflection of a point x with respect to a set C is given by $R_C(x) = 2P_C(x) - x$.
- $P_C(x)$ and $R_C(x)$ are singletons for all x if and only if C is closed and convex.

Fig. 1 : A convex set C with $P_C(x_i) = p_i$ and $R_C(x_i) = r_i$ for $i = 1, 2$.

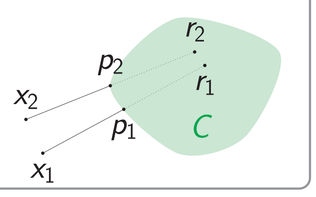
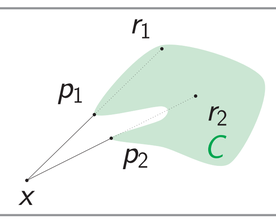


Fig. 2 : A non-convex set C with (set-valued) $P_C(x) = \{p_1, p_2\}$ and $R_C(x) = \{r_1, r_2\}$



3. An Easy-to-Implement Iterative Algorithm

- The Douglas–Rachford method is an algorithm useful for solving the feasibility problem:

$$\text{Find } x \in A \cap B. \tag{1}$$

- Given an initial point x_0, the method can be described by the fixed point iteration

$$x_{n+1} = Tx_n \quad \text{where} \quad T = \frac{Id + R_B R_A}{2}.$$

- Well-understood when A and B are closed and convex [1]: If $A \cap B \neq \emptyset$ then

$$x_n \to x \quad \text{such that} \quad P_A(x) \in A \cap B.$$

- Useful when (1) cannot be solved directly but P_A and P_B are readily available.
- The sequence of interest is not $(x_n)_{n=1}^{\infty}$, but rather the shadow sequence $(P_A x_n)_{n=1}^{\infty}$.
- In the absence of convexity, the method still performs well despite a lack of sufficient theoretical justification.
- We have found success when applied to non-convex combinatorial problems including Sudoku, Hadamard matrix searches, and Paint-by-Numbers puzzles [2,3].
- Closely related to the difference-map algorithm popular in the imaging community for phase retrieval and other difficult non-convex feasibility problems [4,5]

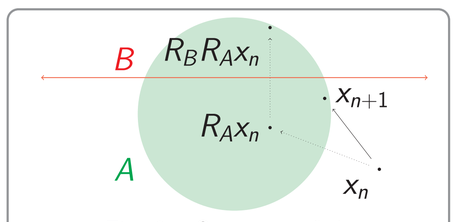


Fig. 3 : One step of the Douglas–Rachford method.

4. Low-Rank Euclidean Distance Matrix Completion

- A matrix $D = (D_{ij}) \in \mathbb{R}^{m \times m}$ is a Euclidean distance matrix (EDM) if there exists points $z_1, z_2, \ldots, z_m \in \mathbb{R}^m$ such that

$$D_{ij} = \|z_i - z_j\|^2 \text{ for } i, j = 1, 2, \ldots, m. \tag{2}$$

- When (2) holds for a set of points in $\mathbb{R}^q$, then D is said to be embeddable in $\mathbb{R}^q$.
- Suppose that Y is a partial EDM with Y_{ij} known only if $(i, j) \in \Omega$ for an index set Ω. The rank-q Euclidean distance matrix completion problem is:

$$\text{Find } X \in \{X : \mathbb{R}^{m \times m} : X_{ij} = Y_{ij} \text{ if } (i, j) \in \Omega,\ X \text{ is a EDM emeddable in } \mathbb{R}^q\}.$$

- A useful characterization of EDMs due to Hayden & Wells [6]: A non-negative, symmetric matrix $X \in \mathbb{R}^{m \times m}$ having zeros along the main diagonal is an EDM if and only if the block matrix $\widehat{X} \in \mathbb{R}^{(m-1)\times(m-1)}$ in

$$Q(-X)Q = \begin{bmatrix} \widehat{X} & d \\ d^T & \delta \end{bmatrix}, \tag{3}$$

  is positive semi-definite. In this case, X is embeddable in $\mathbb{R}^q$ but not $\mathbb{R}^{q-1}$ where $q = \operatorname{rank} \widehat{X}$.
- We may therefore formulae rank-q EDM completion as (1) with

$$\begin{aligned} A &= \{X \in \mathbb{R}^{m \times m} : X \geq 0, X_{ij} = Y_{ij} \text{ for } (i, j) \in \Omega\}, \\ B &= \{X \in \mathbb{R}^{m \times m} : \widehat{X} \text{ in (3) is positive semi-definite with } \operatorname{rank} \widehat{X} \leq q\}. \end{aligned} \tag{4}$$

5. Computing P_A and P_B for Rank-q EDM Completion

- The convex set A encodes the problem data. P_A is simple to compute:

$$P_A(X)_{ij} = \begin{cases} Y_{ij} & \text{if } (i, j) \in \Omega, \\ \max\{0, X_{ij}\} & \text{otherwise;} \end{cases} \qquad \text{for } i, j = 1, 2, \ldots, m.$$

- The non-convex set B encodes *a priori* knowledge. In general, P_B is set-valued and given by:

$$P_B(X) = \left\{ -Q \begin{bmatrix} \widehat{Y} & d \\ d^T & \delta \end{bmatrix} Q : Q(-X)Q = \begin{bmatrix} \widehat{X} & d \\ d^T & \delta \end{bmatrix}, \begin{array}{l} \widehat{X} \in \mathbb{R}^{(m-1)\times(m-1)}, \\ d \in \mathbb{R}^{m-1},\ \delta \in \mathbb{R}, \end{array} \widehat{Y} \in P_S \widehat{X} \right\}.$$

  where S is the set of positive semi-definite matrices having rank q or less.
- One method for computing $P_S(\widehat{X})$ uses the eigen-decomposition of $\widehat{X}$.

6. Method Testing: Reconstructing Known Conformations

- NMR experiments were simulated for proteins with known conformation by computing the partial EDM containing all inter-atomic distances < 6Å. This forms the constraint set A in (4).

Table 1 : Six proteins from the RCSB Protein Data Bank.

Protein	# Atoms	# Residues	Known Distances
1PTQ	404	50	8.83%
1HOE	581	74	6.35%
1LFB	641	99	5.57%
1PHT	988	85	4.57%
1POA	1067	118	3.61%
1AX8	1074	146	3.54%

- The Douglas–Rachford method was then applied to the rank-3 EDM completion problem.
- The reconstructed EDM was then converted to points in $\mathbb{R}^3$ using a multi-dimensional scaling algorithm. Atomic bonds between points were then determined using their Van der Waal radii.

7. Computational Results: Error Metrics

- The reconstructed EDM is compared to the actual EDM using:

$$\text{Relative error (decibels)} = 10 \log_{10} \left(\frac{\|P_A x_n - P_B R_A x_n\|^2}{\|P_A x_n\|^2} \right).$$

- The reconstructed points in $\mathbb{R}^3$ are then compared using:

$$\text{RMS Error} = \left(\sum_{k=1}^{m} \|z_k - z_k^{\text{actual}}\|^2 \right)^{1/2}, \quad \text{Max Error} = \max_{k=1,\ldots,m} \|z_k - z_k^{\text{actual}}\|,$$

  which are computed up to translation, reflection and rotation of the reconstructed points.

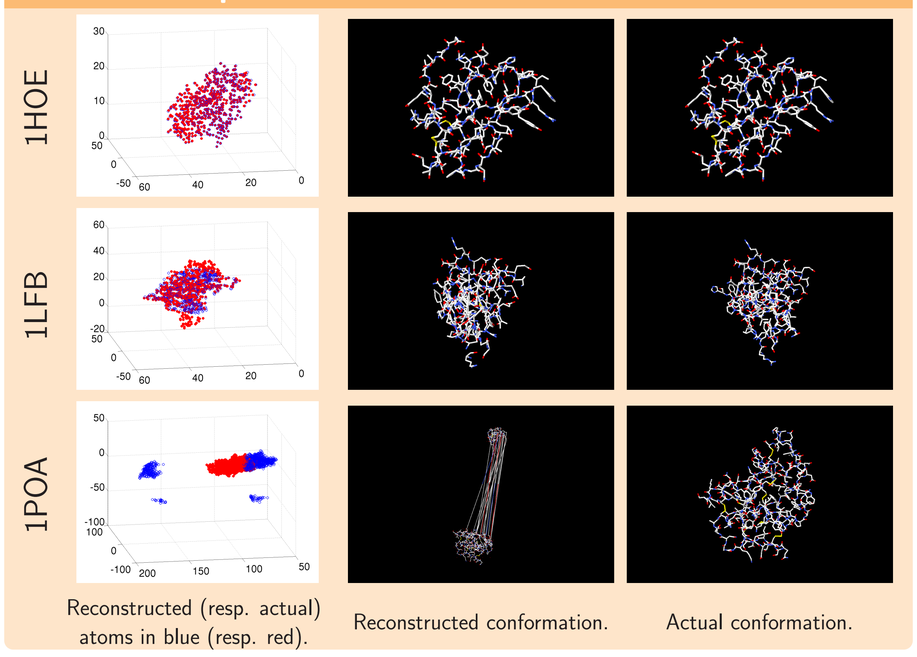
Table 2 : Average (worst) results after 5,000 iterations for five random initial points.

Protein	Problem Size	Relative Error (dB)	RMS Error	Max Error
1PTQ	81,406	-83.6 (-83.7)	0.02 (0.02)	0.08 (0.09)
1HOE	168,490	-72.7 (-69.3)	0.19 (0.26)	2.88 (5.49)
1LFB	205,120	-47.6 (-45.3)	3.24 (3.53)	21.68 (24.00)
1PHT	236,328	-60.5 (-58.1)	1.03 (1.18)	12.71 (13.89)
1POA	568,711	-49.3 (-48.1)	34.09 (34.32)	81.88 (87.60)
1AX8	576,201	-46.7 (-43.5)	9.69 (10.36)	58.55 (62.65)

8. Computational Results: Visualization

- How do the errors reported in Table 2 compare to our expectations?

Three Examples:



Reconstructed (resp. actual) atoms in blue (resp. red).

Reconstructed conformation.

Actual conformation.

- 1HOE is good, 1LFB is mostly good, and 1POA has two well reconstructed pieces.

9. Why Use the Douglas–Rachford Method?

- An even simpler algorithm for solving (1) is the method of alternating projections.
- Given an initial point y_0, it can be described by the fixed point iteration: $y_{n+1} = P_B P_A y_n$.
- Also well-understood for A and B closed and convex: If $A \cap B \neq \emptyset$ then $y_n \to y \in A \cap B$.
- Compare the reconstruction of 1PTQ given by the two algorithms:

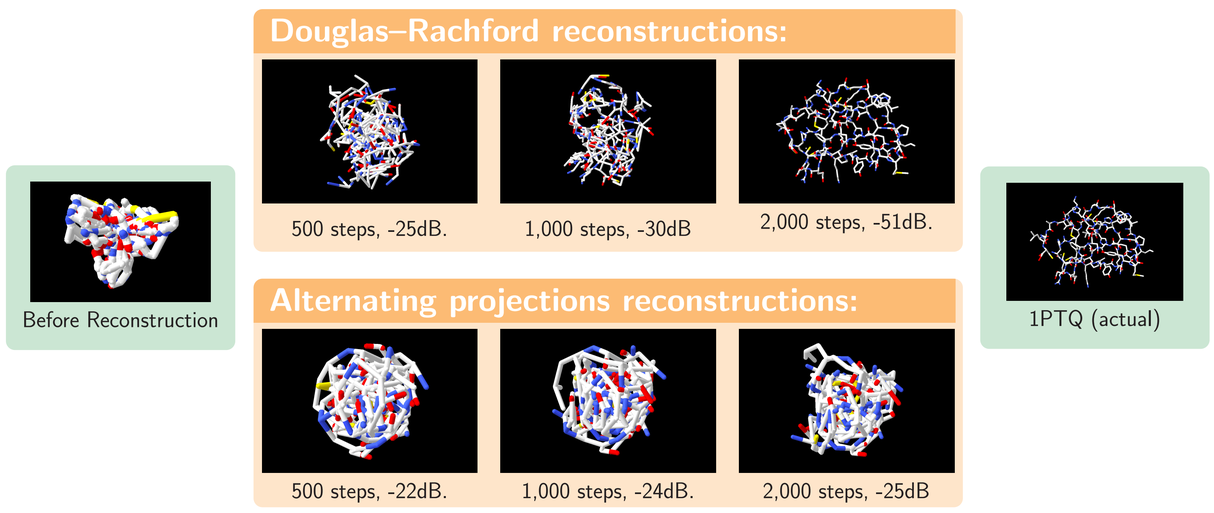


Fig. 4 : Reconstructions of 1PTQ.

- After 2,000 steps, the Douglas–Rachford reconstruction is visually indistinguishable, but the method of alternating projections reconstruction is "stuck".

10. Concluding Remarks and Future Work

- The Douglas–Rachford method is able to predict protein conformation using only short-range distances and *a priori* knowledge. It performs better than theory suggests.
- A proof of local convergence of the method applied to this problem seems possible.
- The Douglas–Rachford method is a general purpose algorithm. Are there problem specific improvements of the method which exploit special structure present in our constraint sets?
- What other applications are fruitful? We are currently investigating an analogous bulk structure determination problem arising in ionic liquid chemistry.

References:

[1] H.H. Bauschke, P.L. Combettes and D.R. Luke. *Finding best approximation pairs relative to two closed convex sets in Hilbert spaces*. J. Approx. Theory, 127:178-192 (2004).

[2] F.J. Aragón Artacho, J.M. Borwein and **M.K. Tam**. *Douglas–Rachford feasibility methods for matrix completion problems*. ANZIAM J., accepted March 2014. Preprint arXiv:1308.424

[3] F.J. Aragón Artacho, J.M. Borwein and **M.K. Tam**. *Recent results on Douglas–Rachford methods for combinatorial optimization problems*. J. Optim. Theory Appl., published online December 2013. Preprint: arXiv:1305.2657

[4] S. Gravel and V. Elser. *Divide and concur: A general approach to constraint satisfaction*, Phys. Rev. E., 78:036706 (2006).

[5] V. Elser, L. Rankenburg and P. Thibault. *Searching with iterated maps*, PNAS, 104(2):418–2423 (2007).

[6] T.L. Hayden and J. Wells. *Approximation by positive semidefinite matrices on a subspace*, Linear Algebra Appl., 109:115–130 (1988).