# GUARANTEED ε-OPTIMAL SOLUTIONS WITH THE LINEAR OPTIMIZER ART3+O

Bram Gorissen

# INTRODUCTION

- Talk is about optimization, but equally applicable to image reconstruction

- Structure:
  - Short intro on ART3,  ART3+ and ART3+O
  - Problem formulation
  - Solution
  - Numerical results

# ART3

Mathematical Programming 8 (1975) 1–19.
North-Holland Publishing Company

## A RELAXATION METHOD FOR RECONSTRUCTING OBJECTS FROM NOISY X-RAYS*,**

G.T. HERMAN

*State University of New York at Buffalo, Amherst. New York, U.S.A.*

Received 19 December 1973

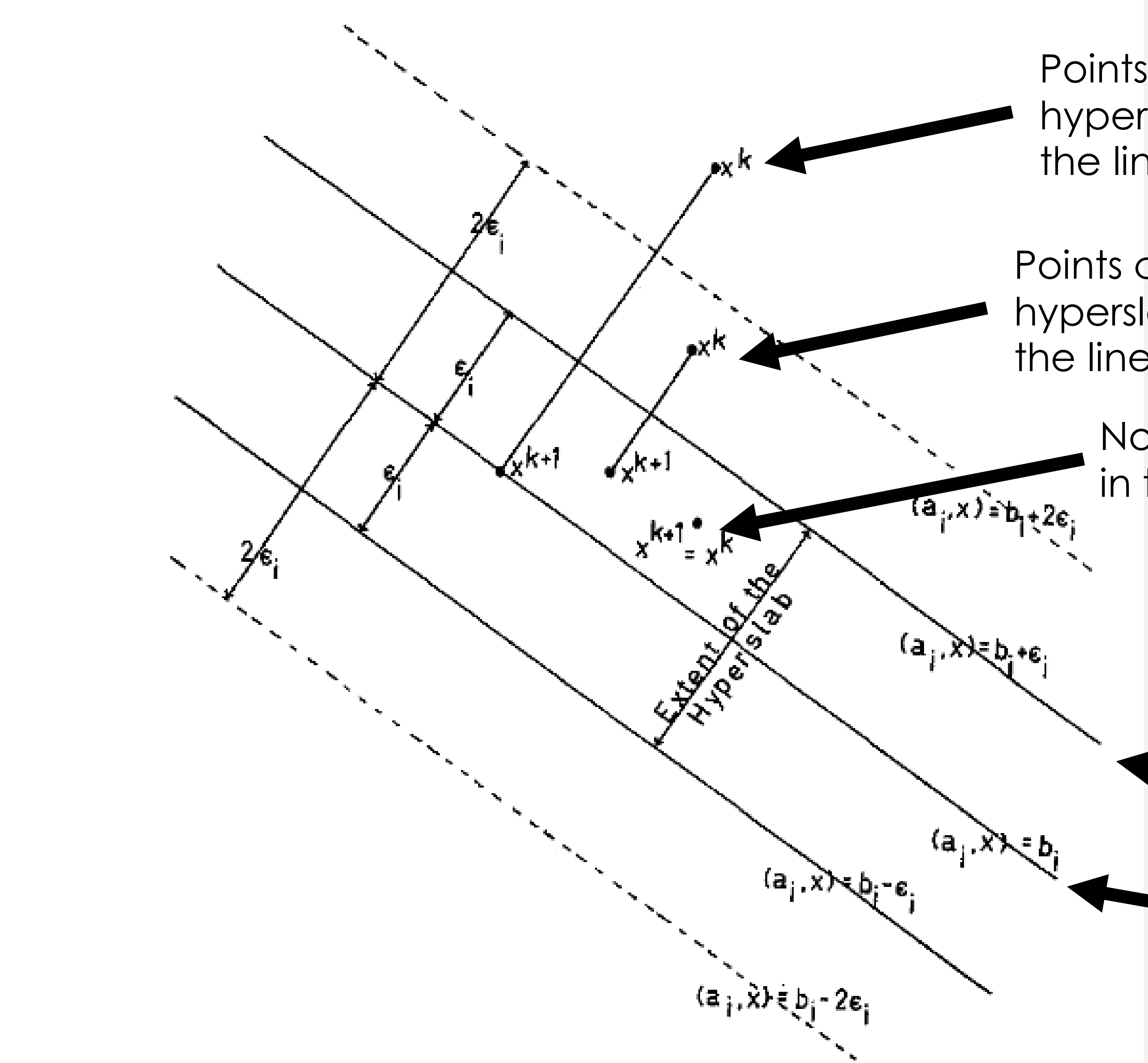
An algorithm is presented for estimating the density distribution in a cross section of an object from X-ray data, which in practice is unavoidably noisy. The data give rise to a large sparse system of inconsistent equations, not untypically $10^5$ equations with $10^4$ unknowns, with only about 1% of the coefficients non-zero. Using the physical interpretation of the equations, each equality can in principle be replaced by a pair of inequalities, giving us the limits within which we believe the sum must lie. An algorithm is proposed for solving this set of inequalities. The algorithm is basically a relaxation method. A finite convergence result is proved. In spite of the large size of the system, in the application area of interest practical solution on a computer is possible because of the simple geometry of the problem and the redundancy of equations obtained from nearby X-rays. The algorithm has been implemented, and is demonstrated by actual reconstructions.

# ART3

- Solves a noisy large system $Ax = b$ by solving $b - \varepsilon \leq Ax \leq b + \varepsilon$ via projections

- The algorithm iterates over the constraints:

```
While constraint violations exist
    For i = 1,…,m
        if a_i^T x < b_i – ε or a_i^T x > b_i + ε
            project x onto inequality i (see next slide)
        End if
    End for
End while
```




Points far from the hyperslab get projected on the line $a_i^T x = b_i$

Points close to the hyperslab get reflected in the line $a_i^T x = b_i + \varepsilon$

Nothing happens to points in the hyperslab

The line $a_i^T x = b_i + \varepsilon$

The line $a_i^T x = b_i$

(From the ART3 paper)

# ART3+

Available online at www.sciencedirect.com
ScienceDirect
ELSEVIER
Linear Algebra and its Applications 428 (2008) 1207–1217
LINEAR ALGEBRA AND ITS APPLICATIONS
www.elsevier.com/locate/laa

## A fast algorithm for solving a linear feasibility problem with application to intensity-modulated radiation therapy

Gabor T. Herman*, Wei Chen

*Department of Computer Science, The Graduate Center, City University of New York, New York, NY 10016-4309, United States*

Received 27 June 2006; accepted 12 November 2006
Available online 10 January 2007
Submitted by Y. Censor

# ART3+

- Two changes compared to ART3
  1. Notational generalization to $l \leq Ax \leq u$
     Allows for $l = -\infty$ or $u = \infty$
     Projection if $x_i$ is further than $(l_i+u_i)/2$ from $a_i^T x$.
     Reflection otherwise
  2. Does not visit constraints cyclically, but keeps a list of violated constraints and only checks violations of constraints in the list. If a constraint is not violated, it gets removed from the list. Process repeats when the list is empty.
     Puts focus on the important constraints

# ART3+O

A fast optimization algorithm for multicriteria intensity modulated proton therapy planning

Wei Chen[a)]
*Department of Computer Science, Graduate Center, City University of New York, New York, New York 10016*

David Craft,[b)] Thomas M. Madden, Kewu Zhang, and Hanne M. Kooy
*Department of Radiation Oncology, Massachusetts General Hospital and Harvard Medical School, Boston, Massachusetts 02114*

Gabor T. Herman
*Department of Computer Science, Graduate Center, City University of New York, New York, New York 10016*

(Received 29 March 2010; revised 28 June 2010; accepted for publication 16 July 2010; published 26 August 2010)

**Purpose:** To describe a fast projection algorithm for optimizing intensity modulated proton therapy (IMPT) plans and to describe and demonstrate the use of this algorithm in multicriteria IMPT planning.
**Methods:** The authors develop a projection-based solver for a class of convex optimization problems and apply it to IMPT treatment planning. The speed of the solver permits its use in multicriteria optimization, where several optimizations are performed which span the space of possible treatment plans. The authors describe a plan database generation procedure which is customized to the requirements of the solver. The optimality precision of the solver can be specified by the user.
**Results:** The authors apply the algorithm to three clinical cases: A pancreas case, an esophagus case, and a tumor along the rib cage case. Detailed analysis of the pancreas case shows that the algorithm is orders of magnitude faster than industry-standard general purpose algorithms (MOSEK's interior point optimizer, primal simplex optimizer, and dual simplex optimizer). Additionally, the projection solver has almost no memory overhead.
**Conclusions:** The speed and guaranteed accuracy of the algorithm make it suitable for use in multicriteria treatment planning, which requires the computation of several diverse treatment plans. Additionally, given the low memory overhead of the algorithm, the method can be extended to include multiple geometric instances and proton range possibilities, for robust optimization.  [DOI: 10.1118/1.3481566]

Key words: projection method, multi-criteria, optimization, numerical evaluation

# ART3+O

- ART3+ embedded in bisection search for optimization
- Linear optimization problem max{ $c^T x$ :  $Ax \leq b$, $x \geq 0$ }
- Performs bisection search on the optimal value, given starting interval [L,U]:
  - Is { $x : c^T x > (L+U)/2,\ Ax \leq b,\ x \geq 0$ } empty?
    Yes: shrink interval to [(L+U)/2, U]
    No: shrink interval to [L, (L+U)/2]

    Repeat until $U-L < \varepsilon$

# FINITE CONVERGENCE

- ART3(+) finds a solution to $l \le Ax \le u$ in a finite number of steps if the set $\{x : l \le Ax \le u\}$ is full dimensional
- ART3(+) loops infinitely if no solution exists
- Question asked by ART3+O: "is $\{ x : c^T x > (L+U)/2,\ Ax \le b,\ x \ge 0 \}$ empty"?
- This question cannot be answered by ART3(+)!

# PROBLEM FORMULATION

- **How to conclude that { $x : c^T x > (L+U)/2,\ Ax \leq b,\ x \geq 0$ } is empty?**
- ART3+O assumes that the set is empty if ART3+ does not find a solution after *M* iterations
- Problematic: how to select *M*? What happens if ART3+O incorrectly concludes that a set is empty?

# PROPOSED SOLUTION

**Theorem 1** *(Farkas' lemma) Exactly one of the following statements is true:*

1. *There exists an* $x \in \mathbb{R}^n$ *such that* $Fx = b$ *and* $x \geq 0$.
2. *There exists a* $y \in \mathbb{R}^m$ *such that* $F^T y \geq 0$ *and* $b^T y \leq -1$.

- { x : $c^T x > (L+U)/2$, $Ax \leq b$, $x \geq 0$ } written in the form of the theorem:

$$\begin{pmatrix} -c^T & 1 & O \\ A & 0 & I \end{pmatrix} \begin{pmatrix} x \\ s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} -M \\ b \end{pmatrix}, \quad x \geq 0, \quad s_1 \geq 0, \quad s_2 \geq 0$$

# PROPOSED SOLUTION

**Theorem 1** *(Farkas' lemma) Exactly one of the following statements is true:*

1. *There exists an* $x \in \mathbb{R}^n$ *such that* $Fx = b$ *and* $x \geq 0$.
2. *There exists a* $y \in \mathbb{R}^m$ *such that* $F^T y \geq 0$ *and* $b^T y \leq -1$.

- So one of the following sets is empty, while the other set is empty:

1. $\{ x : c^T x > (L+U)/2,\ Ax \leq b,\ x \geq 0 \}$

2. $\{ y : -cy_1 + A^T y_2 \geq 0,\ -(L+U)/2\ y_1 + b^T y_2 \leq -1,\ y_1 \geq 0,\ y_2 \geq 0 \}$

# PROPOSED SOLUTION

- To conclude if a value of $(L+U)/2$ is achievable, run ART3+ in parallel

- One instance on $\{ x : c^T x > (L+U)/2,\ Ax \leq b,\ x \geq 0 \}$
- One instance on $\{ y : -cy_1 + A^T y_2 \geq 0,\ -(L+U)/2\ y_1 + b^T y_2 \leq -1,\ y_1 \geq 0,\ y_2 \geq 0 \}$

- One of the two instances will find a point since ART3+ is finite convergent

**Data:** $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, $c \in \mathbb{R}^n$, $L \in \mathbb{R}$, $U \in \mathbb{R}$
**Result:** $x$ that solves $\max_x \{c^T x : Ax \leq b, x \geq 0\}$.
**while** $U - L > \varepsilon$ **do**

$(L+U)/2 \to M$;

Call two instances of ART3+ in parallel:

1. to find $x$ such that $\begin{pmatrix} -c^T \\ A \\ -I \end{pmatrix} x \leq \begin{pmatrix} -M \\ b \\ 0 \end{pmatrix}$
2. to find $y$ such that $\begin{pmatrix} c & -A^T \\ -M & b^T \\ -1 & 0 \\ O & -I \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \leq \begin{pmatrix} 0 \\ -1 \\ 0 \\ 0 \end{pmatrix}$;

When one instance of ART3+ finds a solution, terminate the other instance;

**if** *a feasible $x$ was found* **then**

$c^T x \to L$

**else**

$M \to U$

**end**

**end**

# IMAGE RECONSTRUCTION REVISITED

- Solve a noisy large system $Ax = b$ by solving $b - \varepsilon \leq Ax \leq b + \varepsilon$ via projections

- If $\varepsilon$ is too small, there is no solution
- If $\varepsilon$ is too large, $x$ does not accurately reconstruct the image

- Solution: to test if $\varepsilon$ is too small, run ART3+ in parallel on:

$$A^T y_1 - A^T y_2 \geq 0,\ (b+\varepsilon)^T y_1 - (b - \varepsilon)^T y_2 \leq -1,\ y_1 \geq 0,\ y_2 \geq 0$$

# NUMERICAL RESULTS

- Taken the implementation of ART3+O from the in-house treatment planning system Astroid and modified it to make it accurately correspond to its published version, with a stopping criterion of $2 \times 10^7$ iterations and $\varepsilon = 0.1$ Gy

- Three test sets from proton therapy
  1. Small case
  2. Large case
  3. 77 clinical cases spanning 11 patients with different tumor sites

# SMALL CASE

- 227 pencil beams
- Maximize min PTV dose (256 voxels) s.t. maximum dose 50 Gy (8406 voxels)
- D matrix 288610 nonzeros (85% sparsity)

- CPLEX: 0.8 seconds, optimal value 32.71 Gy
- ART3+O: 12.5 seconds, optimal value 32.57 Gy
- Proposed method (PM): 2.5 hours, optimal value 32.71 Gy

# SMALL CASE

| Step | $L$ | $U$ | Farkas | Time (s) | Iterations |
|---|---|---|---|---|---|
| 1 | 0.00 | 50.00 | 1 | 0.1 | 62,356 |
| 2 | 25.01 | 50.00 | 2 | 1.9 | 24,948,327 |
| 3 | 25.01 | 37.51 | 1 | 0.0 | 332,498 |
| 4 | 31.26 | 37.51 | 2 | 11.1 | 147,789,618 |
| 5 | 31.26 | 34.38 | 2 | 3160.4 | 39,374,210,616 |
| 6 | 31.26 | 32.82 | 1 | 0.0 | 53,662 |
| 7 | 32.38 | 32.82 | 1 | 0.1 | 2,477,787 |
| 8 | 32.60 | 32.82 | 1 | 1345.9 | 29,435,448,946 |
| 9 | 32.71 | 32.82 | 2 | 4618.7 | 57,541,614,242 |

# LARGE CASE

- 5384 pencil beams
- Objective: minimize the maximum dose in $3 \times 10^5$ voxels
- $2 \times 10^6$ constraints that limit the mean, minimum or maximum dose

- CPLEX: 3.5 hours, objective value 50.8 Gy
- ART3+O: 15 minutes, objective value 51.6 Gy
- PM proved 26.47 Gy and 39.70 Gy are unachievable in 6 minutes and 6 hours, respectively
- PM was killed after 100 hours and $218 \times 10^9$ iterations trying to prove that 46.32 is unattainable

# 77 TEST CASES

- ART3+O does not find a feasible point within the limit of 2 x $10^7$ iterations; limit was increased to $10^9$
- Still 2 out of 10 patients incorrectly classified infeasible, affects 13 cases
- Remaining 64 cases took between 8.5 seconds and 3:05 hours (median: 35 minutes)
- Suboptimality between 0 and 3.6 Gy (median: 0.2 Gy), less than 0.1 Gy in 13/64 cases

- **Suboptimality of 3.6 Gy is an issue**

# 77 TEST CASES

- PM was given 1 day of cpu-time (12 hours of wall-clock time)
- PM on the 64 cases that ART3+O could solve:
  - always slower than ART3+O, hit time limit for 57/64 cases
  - suboptimality between 0 and 12.5 Gy (median: 1.1 Gy)
  - outperforms ART3+O by > 0.1 Gy in 14 cases, outperformed in 41 cases
- PM on the other 13 cases: suboptimality between 0 and 5.7 Gy (median 0.2 Gy)
- **Suboptimality of 12.5 Gy is an issue**

# DISCUSSION

- ART3+ often finds feasible point in the "primal" formulation but not in the "dual formulation"
- Possible explanation: constraints in the primal are often correlated or redundant (max dose constraint on neighboring voxels)
- Dual does not possess this property, because there are no 2 pencil beams that deliver almost the same dose

# CONCLUSION

- ART3+O cannot guarantee ε-optimality and issues do occur
- The suggested improvement turns out to be mostly of theoretical value
- Projection methods for optimization currently do not give an optimality guarantee, and that is a problem