Oberwolfach Workshop:
Mathematics and Algorithms in Tomography, August 10–16, 2014

ART Exhibit*

Per Christian Hansen, Technical University of Denmark
pcha@dtu.dk

Abstract

The iterative reconstruction method known as *ART* (Algebraic Reconstruction Technique) and *Kaczmarz's method*, as well as its many variants, are surprisingly simple and efficient methods with many applications in computed tomography, where we compute 2D and 3D reconstructions from noisy and often incomplete projections. On the theoretical side we are interested in explaining why it is so successful, and on the practical side how to implement it efficiently, how to select the relaxation parameter, and how to stop the iterations. This report takes the form of an "exhibit" of the many aspects of ART.

Keywords: Computed tomography algorithms, iterative reconstruction methods, semi-convergence, stopping rules, block methods.

AMS Subject Classification: 65R32 – Numerical Analysis | Inverse Problems

There are many ways to compute reconstructions in tomography – too many to list here – such as methods based on explicit inversion formulas, Bayesian methods, and algebraic iterative methods. I will focus on a particular algebraic iterative method, ART, which is surprisingly simple to formulate, has a simple geometric interpretation, and works well for a number of applications due to its fast initial convergence. I give only a few references; sorry about all the good work that is not mentioned here.

In Kaczmarz's formulation [11] of the algorithm from 1937, each iteration takes the form of a sweep over the rows a_i^T of the matrix $A \in \mathbb{R}^{m\times n}$ – from top to bottom – where we orthogonally project the current iterate x on the hyperplane defined by row a_i^T and the corresponding element b_i of the right-hand side:

$$x \leftarrow \mathcal{P}_i\, x = x + \frac{b_i - a_i^T x}{\|a_i\|_2^2}\, a_i \ , \qquad i = 1, 2, \ldots, m \ .$$

Herman and his coworkers [7] rediscovered the algorithm 1970 (for the case where all elements of A are 0 or 1). They named it "ART," introduced a nonnegativity projection and used a different normalization:

$$x \leftarrow \max\left\{ 0\, ,\, x + \frac{b_i - a_i^T x}{\|a_i\|_1}\, a_i \right\} \ , \qquad i = 1, 2, \ldots, m \ .$$

This version does not have the simple interpretation as a sequence of orthogonal projections; in later works (e.g., [10]) the name ART is used synonymously with the

*This work is part of the project High-Definition Tomography, supported by grant No. 291405 from the European Research Council.

Kaczmarz formulation. Current versions of ART also include a relaxation parameter λ_k and a projection $\mathcal{P}_C$ on a convex set C (e.g., the positive orthant or the box $[0,1]^n$):

$$x \leftarrow \mathcal{P}_C \left(x + \lambda_k \, \frac{b_i - a_i^T x}{\|a_i\|_2^2} \, a_i \right) \ , \qquad i = 1, 2, \ldots, m \ .$$

In spite of its success there are many open questions associated with the use of this method, and hence it is a rich source for research problems! Below is a survey of some of these issues, with a strong bias towards my own research and my ongoing work with the MATLAB package AIR Tools [9]. Please note the effort by Joost Batenburt from CWI to develop a highly efficient toolbox ASTRA [14], based on a MATLAB wrapper around native C++ and CUDA code.

Semi-Convergence Theory

There is a rich convergence theory for ART and its many block extensions; but most of this theory exclusively deals with its asymptotic convergence. However, the power of ART really lies in its ability to converge fast to a good approximate/regularized solution during the initial iterations – while at later stages its slows down considerably and eventually converges to an undesirable solution dominated by noise from the data.

Specifically, ART is observed to exhibit *semi-convergence* [13], where the iteration number plays the role of the regularizing parameter. In the early stages the iterates approach a regularized solution, while continuing the iteration leads to iterates deteriorated by noise. A few attempts have been made to give a rigorous foundation for this observed behavior, such as [4]; but more work is needed to demonstrate in which circumstances we are guaranteed to have semi-convergence, to give rates for the initial convergence, etc.

Perhaps such a study could use the close connection between ART and the SOR method. Also, one might be able to use the fact that the "symmetric ART" method (which is connected to SSOR) lends itself to an SVD analysis.

Implementation of Block ART

There are surprisingly many ways to define block extensions of ART; some of them are surveyed in [6] and [15]. These block methods lends themselves naturally to distributed computing systems and MPI-type implementations for large-scale problems [12]. The block methods are also well suited for multi-core computers [15] as well as systems based on GPUs. Some important questions here are:

- How do we best choose the number of blocks on a given computer?
- How can we best utilize the structural orthogonality between the matrix rows to choose the blocks adaptively?
- What is the best combination of block iteration (sequential/parallel) and treatment of the individual blocks (by a direct or iterative method)?

To best utilize the specialized architecture of the GPU, the matrix-free multiplications are implemented such that the backprojection corresponds to multiplication with the transposed of a matrix $\widetilde{A}$ that is slightly different from the matrix A associated with the forward computation [17]. The unmatched transpose may prevent asymptotic convergence; it is unclear how it affects the semi-convergence and the accuracy of the reconstruction.

Choice of Relaxation Parameter

For a given problem and a given (block) method we claim that there is an optimal choice of relaxation parameter – either fixed or depending on the iterations – that gives the fastest semi-convergence during the initial iterations. The problem is how to estimate this parameter or parameter-sequence for a given problem with noisy data. Some options are:

- Determine a fixed parameter by means of "training," i.e., by determining the parameter that gives fastest semi-convergence for a test problem with a known solution [9].
- Determine a "greedy" strategy to vary the parameter during the iterations that gives the largest reduction of the error at each stage, similar to [3].
- Determine a strategy such that the iterations "slow down" once we reach the point of semi-convergence (this makes the stopping criterion less critical) [5].

Stopping Rules

For discretizations of inverse problems it is well known that the "standard" stopping rule for iterative methods – stop when the residual is small – cannot ensure a satisfactory regularized solution. Any stopping rule for such systems should take into account the ill-posed nature of the underlying problem, the type of noise, and preferably the desired properties of the regularized solution. Several methods for choosing the Tikhonov regularization parameter [8] have these properties (e.g., Morozov's discrepancy principle, the L-curve criterion, and Rust's method based on the residual's power spectrum), but it is unclear how suited they are for iterative methods applied to large-scale problems. To define robust stopping rules it may be worthwhile to consider the following issues:

- Often we know the statistical properties of the noise in the data because we have detailed information about the instrument. This information should be utilized.
- ART is strongly related to first-order optimization methods of the incremental gradient type, and it should be possible to exploit the solid understanding of these methods.
- The reconstructed image is, in many cases, in intermediate result to which further data analysis is applied (the objective is not to produce "pretty pictures"). A good stopping rule should ensure that reconstruction allows for optimal data analysis.

Extensions and Variations of ART

ART is typically formulated as a row-action method what works with (blocks of) rows of the matrix. An obvious variation of ART is to work instead with the columns of the matrix [16], which may be advantageous when it is desirable to updated elements or blocks of the solution independently (e.g., for memory issues) [2].

Extensions of ART are possible through its connection to first-order optimization methods and its interpretation as an incremental gradient method. Some examples of this can be found in [1], and these seems to be potential for powerful extensions that combine the simplicity of ART with the ability to incorporate prior information and improve the convergence rate.

Acknowledgement

My interest in ART is prompted by a general interest in tomographic imaging, by my participation in several imaging initiatives at DTU (www.imaging.dtu.dk/english), and by an ERC Advanced Research grant (www.compute.dtu.dk/~pcha/HDtomo). I am indebted to many colleagues at DTU, in particular Martin S. Andersen, Jakob Sauer Jørgensen, and Søren Schmidt, as well as Bill Lionheart from Manchester University.

References

[1] M. S. Andersen and P. C. Hansen, *Generalized row-action methods for tomographic imaging*, Numer. Algo., November 2013, DOI: 10.1007/s11075-013-9778-8.

[2] A. Bech and L. Tetruashvili, *On the convergence of block coordinate descent type methods*, SIAM J. Optim., 23 (2013), pp. 2037–2060.

[3] A. Dax, *Line search acceleration of iterative methods*, Linear Algebra Appl., 130 (1990), pp. 43-63.

[4] Elfving, P. C. Hansen, and T. Nikazad, *Semi-convergence properties of Kaczmarzs method*, Inverse Problems, 30 (2014), DOI: 10.1088/0266-5611/30/5/055007.

[5] T. Elfving, T. Nikazad, and P. C. Hansen, *Semi-convergence and relaxation parameters for a class of SIRT algorithms*, Electronic Trans. on Numerical Analysis, 37 (2010), pp. 321–336.

[6] T. Elfving, T. Nikazad, and C. Popa, *A class of iterative methods: semi-convergence, stopping rules, inconsistency, and constraining*, in: Y. Censor, M. Jiang, and G. Wang (Eds.), *Biomedical Mathematics: Promising Directions in Imaging, Therapy Planning, and Inverse Problems*, Medical Physics Publishing, Madison, WI, 2010.

[7] R. Gordon, R. Bender, G. T. Herman, *Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and X-ray photography*, J. Theoretical Biology, 29 (1970), pp. 477–481.

[8] P. C. Hansen, *Discrete Inverse Problems: Insight and Algorithms*, SIAM, Philadelphia, 2010 (213 pages).

[9] P. C. Hansen and M. Saxild-Hansen, *AIR Tools – A MATLAB package of algebraic iterative reconstruction methods*, J. Comp. Appl. Math., 236 (2012), pp. 2167–2178, DOI: 10.1016/j.cam. 2011.09.039.

[10] G. T. Herman, *Fundamentals of Computerized Tomography: Image Reconstruction from Projections*, 2. Ed., Springer, New York, USA, 2009.

[11] S. Kaczmarz, *Angenäherte Auflösung von Systemen linearer Gleichungen*, Bull. Acad. Pol. Sci. Lett., A35 (1937), pp. 355–357.

[12] N. T. Karonis, K. L. Duffin, C. E. Ordoñez, B. Erdelyi, T. D. Uram, E. C. Olson, G. Coutrakon, M. E. Papka, *Distributed and hardware accelerated computing for clinical medical imaging using proton computed tomography (pCT)*, J. Parallel Distrib. Comput., 73 (2013), pp. 1605–1612.

[13] F. Natterer, *The Mathematics of Computerized Tomography*, SIAM, Philadelphia, 2001.

[14] W J. Palenstijn, K J. Batenburg, and J. Sijbers, *Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs)*, J. Structural Biology, 176 (2011), pp. 250–253.

[15] H. H. B. Sørensen and P. C. Hansen, *Multi-core performance of block algebraic iterative reconstruction methods*, SIAM J. Sci. Comp., to appear.

[16] D. W. Watt, *Column-relaxed algebraic reconstruction algorithm for tomography with noisy data*, Appl. Opt., 33 (1994), pp. 4420–4427.

[17] G. L. Zeng and G. T. Gullberg, *Unmatched projector/backprojector pairs in an iterative reconstruction algorithm*, IEEE Trans. Medical Imag., 19 (2000), pp. 548–555.