

A FAST PARALLEL METHOD FOR MEDICAL IMAGING PROBLEMS INCLUDING LINEAR INEQUALITY CONSTRAINTS

Thomas D. Capricelli

Laboratoire J.-L. Lions – UMR 7598
Université Pierre et Marie Curie – Paris 6, 75005 Paris, France

ABSTRACT

When studying problems such as tomography with bounded noise or IMRT, we need to solve systems with many linear inequality constraints. Projection-based algorithms are often used to solve this kind of problem. We see how previous work for accelerating the convergence of linear algorithms can be recast within the most recent generic framework, and show that it gives better results in specific cases. The proposed algorithm allows general convex constraints as well and the conditions for convergence are less restrictive than traditional algorithms. We provide numerical results carried out in the context of tomography and IMRT.

Index Terms— Tomography, IMRT, linear inequality constraints, image reconstruction.

1. INTRODUCTION

Most imaging problems can be formulated as a feasibility problem, for which many algorithms have been developed in the last 30 years [6]. The idea is to describe the problem in the Euclidian space $\mathcal{H} = \mathbb{R}^N$ with a number of *constraints*, which are derived from observed data and *a priori* on solutions. Most algorithms require those constraints to be convex. We shall therefore consider the following mathematical formulation, where $(S_i)_{i \in I}$ in \mathcal{H} are closed convex sets.

Problem 1.1 (Convex Feasibility) Find a point in $S = \bigcap_{i \in I} S_i$.

We are especially interested in two biomedical imaging experiments: tomography with bounded noise, and Intensity Modulated Radio Therapy (IMRT). For those particular problems, most of the constraint sets are actually a lot more simpler than just convex sets, they are *hyperslabs*. A hyperslab is a set of the form :

$$S_i = \{x \in \mathcal{H} \mid \beta_i - \delta_i \leq \langle x \mid a_i \rangle \leq \beta_i + \delta_i\}, \quad (1)$$

where $a_i \in \mathcal{H} \setminus \{0\}$ is the normal direction, $\beta_i \in \mathbb{R}$ gives the location in the Euclidian space, and $\delta_i > 0$. Without loss of generality we shall from now on consider $\|a_i\| = 1$. The corresponding central hyperplan H_i is $\{x \in \mathcal{H} \mid \langle x \mid a_i \rangle = \beta_i\}$.

Hyperslabs appear as constraints in many other experiments, even with unbounded noise [4, 6], and the technique described in this article can be used as well.

It could still be useful to consider convex sets which are not linear, in order to be able to embed as much *a priori* knowledge as possible in the problem statement (for example total variation bound, or some Fourier constraint sets). Our aim in this article is to provide a generic, though efficient, algorithm to solve problem 1.1 in this particular cases.

We start by describing the different algorithms and recall some theoretical results. We then propose and describe a specific operator suitable to be used in the most recent algorithm. The last sections describes and presents some numerical results.

2. ALGORITHMS

To solve problem 1.1, the original POCS algorithm (Projection Onto Convex Sets) was proposed by Bregman in 1965 [2] for a finite number of constraints. It uses a sequential (sets are considered one at a time) and cyclic control. There is no relaxation. Bregman proved the convergence if S is not empty.

Algorithm 2.1 (POCS, Bregman, 1965) $x_{n+1} = P_{i_n} x_n$, with a static control $i_n = (i \bmod \text{card } I) + 1$.

In the special case where all sets are hyperslabs, G. T. Herman proposed the following algorithm [9], called Algebraic Reconstruction Technique 3:

Algorithm 2.2 (ART3, Herman, 1975) $x_{n+1} = U_{i_n} x_n$,

where U_i is a new operator he introduced, and $(i_n)_{n \in \mathbb{N}}$ is the same *static control* as in 2.1. He proved the convergence in finite steps under the hypothesis that S is full dimensional. This kind of result is common with such a strong hypothesis. This algorithm is now seen as a reference for this special case and is often used in comparisons when designing new algorithms. The slight improvement introduced in the algorithm ART3+ [10] is related to the control, which is not what this article focus on.

For the generic problem 1.1, the original *POCS* algorithm has been enhanced in many different ways. The projection

is first generalized using the *subgradient projection*. If S_i is defined by $S_i = \{x \in \mathcal{H} \mid f_i(x) \leq 0\}$ (which is always possible), the subgradient projection is defined by:

$$G_i x = \begin{cases} x - f_i(x)u/\|u\|^2 & \text{si } f_i(x) > 0; \\ x, & \text{si } f_i(x) \leq 0, \end{cases} \quad (2)$$

where $u \in \partial f_i(x)$ is any subgradient of f at x . This is an approximation that is often used when the actual projection has no close form or cannot be computed efficiently on a computer. The projection is a special case of the subgradient projection. See the references in [1, 7] for more details.

Then, class \mathfrak{T} operators were introduced in [1], and are a generalization of projections and subgradient projections: subgradient operators are in class \mathfrak{T} [1]. We shall note $\text{Fix } T$ the set of fixed points for the operator T : $\text{Fix } T = \{x \in \mathcal{H} \mid Tx = x\}$. Given two points x and y in \mathcal{H} , we consider the set $H(x, y) = \{u \in \mathcal{H} \mid \langle u - y \mid x - y \rangle \leq 0\}$, which is a half space if $x \neq y$. An operator $T : \mathcal{H} \rightarrow \mathcal{H}$ is said to be in class \mathfrak{T} if $(\forall x \in \mathcal{H}) \text{Fix } T \subset H(x, Tx)$.

The state of the art of algorithms using several projections in parallel is *EMOPSP* [7]. The following algorithm is a generalization to class \mathfrak{T} of *EMOPSP* (see [8] for a description and convergence results).

Algorithm 2.3

$$x_{n+1} = x_n + \lambda_n L_n \left(\sum_{i \in I_n} \omega_{i,n} (T_{i,n} x_n) - x_n \right), \quad (3)$$

where $x_0 \in \mathcal{H}$, $\epsilon \in]0, 1]$, $\delta \in]0, 1]$ and $\forall n \in \mathbb{N}$:

- (a) $\emptyset \neq I_n \subset I$ is the control,
- (b) $\forall i \in I_n T_{i,n} \in \mathfrak{T}$ and $\text{Fix } T_{i,n} = S_i$,
- (c) $\lambda_n \in [\epsilon, 2 - \epsilon]$ is the relaxation parameter,
- (d) $\omega_{i,n} \in]\delta, 1]$ are the weights, with $\sum_{i \in I_n} \omega_{i,n} = 1$,

$$(e) L_n = \begin{cases} \frac{\sum_{i \in I_n} \omega_{i,n} \|T_{i,n} x_n - x_n\|^2}{\|\sum_{i \in I_n} \omega_{i,n} T_{i,n} x_n - x_n\|^2} & \text{if } x_n \notin S, \\ 1 & \text{else.} \end{cases}$$

The control is supposed *admissible*, which means that there exist strictly positive integers $(M_i)_{i \in I}$ such that

$$(\forall (i, n) \in I \times \mathbb{N}) i \in \bigcup_{k=n}^{n+M_i-1} I_k.$$

Every cyclic control (as in algorithm 2.1) is admissible. For comparison purpose, we shall call *EMOPSP* the same algorithm, with the following standard parameters: $(\forall n \in \mathbb{N}) \lambda_n = 1.99$, $T_{i,n} = G_i$ (subgradient or projection), and uniform weights $\omega_{i,n}$ over selected I_n . We use the indice partition $I = I_1 \cup I_2$, where I_2 are indices for which S_i is an hyperslab, and $I_1 = I \setminus I_2$.

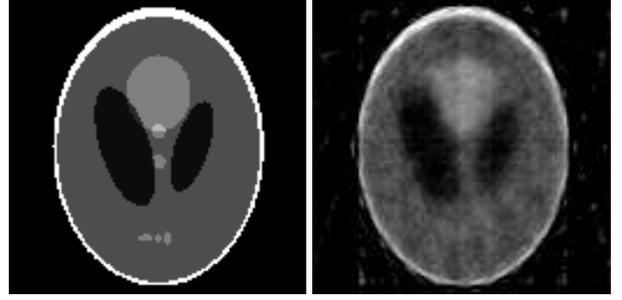


Fig. 1. Left: original phantom used for the tomography experiments. Right: reconstruction example.

3. A SPECIFIC OPERATOR

The algorithm 2.3 is general and has many parameters which can be tuned for a specific application. However, finding how to choose all of those parameters is often difficult. We focus on the relaxation parameter λ_n . Even though enhanced speed can be achieved using small relaxation in some rough corner cases [11], in the average case, our experience is that the best bet is to use over-relaxation, by setting $(\forall n \in \mathbb{N}) \lambda_n = 1.99$. The algorithm 2.2 (ART3), proposed by G. T. Herman [9] can be reformulated as a special choice for the relaxation parameter in the *cyclic* case. We shall see how this can be extended to the general, parallel scheme of algorithm 2.3, and we propose a new version of the algorithm 2.3 adapted to problems where many constraints are hyperslabs. Let us consider the operator used to enforce the linear inequality constraints S_i , when $i \in I_2$. The projection of the vector x on S_i is the vector $P_{S_i} x$ (shortened as $P_i x$ from now on) defined by

$$\begin{cases} x, & \text{if } x \in S_i, \\ x + (\beta_i - \delta_i - \langle a_i \mid x \rangle) a_i & \text{if } \langle a_i \mid x \rangle < \beta_i - \delta_i, \\ x + (\beta_i + \delta_i - \langle a_i \mid x \rangle) a_i & \text{if } \langle a_i \mid x \rangle > \beta_i + \delta_i. \end{cases} \quad (4)$$

The new operator in ART3 is:

$$U_i x = \begin{cases} x, & \text{if } x \in S_i, \\ P_{H_i} x, & \text{if } |\langle a_i \mid x \rangle - \beta_i| > 2\delta_i, \\ 2P_i x - x, & \text{else,} \end{cases} \quad (5)$$

from which we have $\text{Fix } U_i = S_i$. Unfortunately, this operator does not belong to the class \mathfrak{T} . It can be written as an over-relaxation of the projection onto S_i , $U_i x = x + \nu(x)(P_i x - x)$ with

$$\nu(x) = \begin{cases} 1, & \text{if } x \in S_i, \\ 1 + \frac{\delta_i}{d(x, S_i)}, & \text{if } |\langle a_i \mid x \rangle - \beta_i| > 2\delta_i, \\ 2, & \text{else.} \end{cases} \quad (6)$$

We have $1 \leq \nu(x) \leq 2$, and U_i are mostly over-relaxed projection onto S_i . Using the proposition [1, 2.6(iii)] we conclude that the operator $T_i = (\text{Id} + U_i)/2$ has the same fixed



Fig. 2. Minimum and maximum value for the IMRT experiment. The grey part on the left is the tumor. The two dark parts on the right are places especially fragile.

points as U_i , and it belongs to the class \mathfrak{T} . We use this operator T_i for the sets with indice in I_2 , while for the sets with an indice in I_1 , we keep the general subgradient projection operators G_i .

Algorithm 3.1 Algorithm 2.3 with, $(\forall n \in \mathbb{N})$: (a) $\lambda_n = 1.99$ (b) $T_{i,n} = G_i$ if $i \in I_1$ (c) $T_{i,n} = T_i$ if $i \in I_2$ (d) Uniform weights $\omega_{i,n}$ over selected I_n .

ART3 is a very specific case for this algorithm. The use of the relaxation factor 1.99 instead of 2 allows far less restrictive hypothesis for convergence results while it doesn't make any real difference with respect to the speed.

4. APPLICATIONS

Whether the use of operators T_i enhances the speed of the algorithm or not is highly dependant on the nature of the problem. For example, we consider in [3] a problem unrelated to biomedical imaging: the deconvolution of a signal in one dimension, with added noise. We show that when using the operator T_i , the convergence takes more time than with traditional algorithms.

To compare the algorithms, we shall use common, though not trivial, numerical experiments: tomography with bounded errors, and IMRT. Both applications are taken from medical imagery. More details can be found in [5].

4.1. Computer Tomography

This is a typical problem of image reconstruction, and can be described in two or three dimensions. We shall describe the 2D case for simplicity. We start from the original phantom shown in Fig. 1, whose pixel range is $[0, 255]$. This is a 2D image, discretized on a $128 \times 128 = N$ grid, and can be seen as a point in the space $\mathcal{H} = \mathbb{R}^N$. We shall denote \bar{x} this image. Given a direction, specified as an angle θ_i from the center of the phantom, we can compute the projection along this direction or *view*, which is a signal in one dimension. The corresponding Euclidian space is \mathbb{R}^M , where M is the number of pixels on each view. There are $q \in \mathbb{N}$ such different

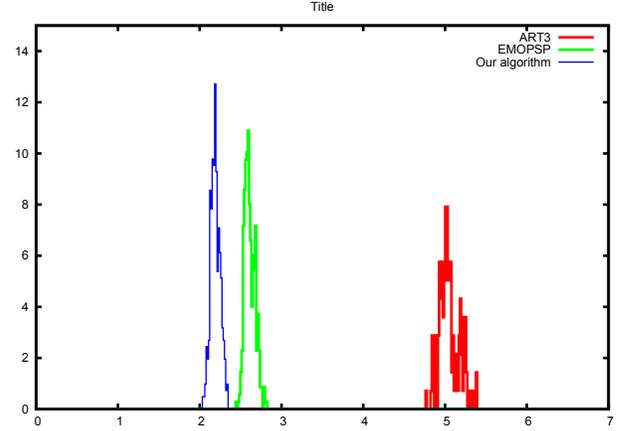


Fig. 3. Tomography with bounded noise: Convergence time repartition based on several hundred noise realizations for each algorithm.

views. The problem is to reconstruct the original image from the observations, using as much *a priori* as possible. The projection is a linear transformation and, if we suppose that the observation noise on each pixel for each view is bounded by δ , we finally have the following constraint sets:

$$S_{i,k} = \{x \in \mathcal{H} \mid |\langle L_i x - s_i, e_k \rangle| \leq \delta\}, \quad (7)$$

where $s_i \in \mathbb{R}^M$, $1 \leq i \leq q$ is the observed data and $(e_k)_{1 \leq k \leq M}$ is the canonical basis for \mathbb{R}^M . Our aim is to compare the relative speed of the operators used to handle the hyperslabs, and we shall only use here one more constraint: the *range* of the image is imposed through the set $S_0 = [0, 255]^N$. For our experiments, we shall use $q = 10$ views, uniformly spaced over the full range of angle (π radians). The bound δ is set to $200/256$. With such parameters, actual convergence is achieved within few seconds on a contemporary computer (AMD 64bits, dual core at 2.2 GHz) using most algorithms. The only one that does not converge in finite time is the original POCS, and this is explained by the fact that each step computes the exact projection without any relaxation, hence never going further than the frontier. The exemple of two non-orthogonal secant lines in \mathbb{R}^2 shows why convergence can take an infinite number of steps.

4.2. Intensity-Modulated Radiation Therapy

Intensity-Modulated Radiation Therapy (IMRT) is a signal problem found in several medical treatments, most notably cancer treatments. Computer-controlled X-rays accelerators can distribute precise dose of X-rays and those (straights) beams can be sent through the patient body along different directions. For every given part of the body, the total dose of X-rays received depends linearly of all the beams sent. Suppose a 2D section of the patient is discretized as an image

x of N pixels (for example $N = 1024 \times 1024$ pixels), and there are M beams used (different positions and directions). We shall note $y \in \mathbb{R}^M$ the vector of intensities of the whole beam. Then the dose received at pixel $1 \leq i \leq N$ can be written as $x_i = L_i y$. The therapy planning consist in giving for every pixel a minimum and a maximum value for the dose received. For a pixel belonging to a tumor, one would want a high minimum value, while for a pixel belonging to surrounding normal tissue the maximum authorized value would be low, and even lower for fragile tissues. Medical personnel decide those minimum and maximum values. The purpose of the algorithm is to find the value of the intensities y , constrained to those medical prescriptions and the actual intensity range of X-rays accelerators. See [5] for further details. We used the minimum and maximum values shown in figure 2, and 8 different angles uniformly spread.

5. NUMERICAL RESULTS

The criterion we chose to use is the relative proximity function $\phi_0(x) = \phi(x)/\phi(x_0)$, where ϕ is the proximity function: $\phi(x) = \sum_{i \in I} w_i \|x_n - P_i x_n\|^2$, with uniform weights ($w_i = 1/\text{card } I$), which describes how far away we are from being feasible. See [3] for a discussion on the different criteria. The time displayed on the X axis of the figures is the total CPU times. For the parallel algorithms, this is the sum of CPU time. We used 50 processors in parallel for our tests. The first point that should be highlighted is that the parallel algorithms are a lot faster than the serial ones, *even on a non-parallel computer*. This is due to the highest range for the relaxation allowed by the parameter L_n .

The results for the IMRT experiments (fig 4) show that there is a gain in using the T_i operator, but not significant. Obviously, concerning the tomography, the convergence depends on the noise realization. To have a fair comparison, we did several hundred tests with different noise realizations. We recorded the time needed to reach convergence (Using the criterion $\phi_0(x) < -200\text{dB}$). The result is displayed in figure 3. It shows that our algorithm performs better. The variability is very small.

6. CONCLUSION

The proposed algorithm provides faster convergence results for problems of type 1.1. It can use more general constraints and parallel computations. The convergence results are also more general.

7. REFERENCES

[1] H. H. Bauschke and P. L. Combettes, "A weak-to-strong convergence principle for Fejér-monotone methods in Hilbert spaces," *Math. Oper. Res.*, vol. 26, pp. 248–264, 2001.

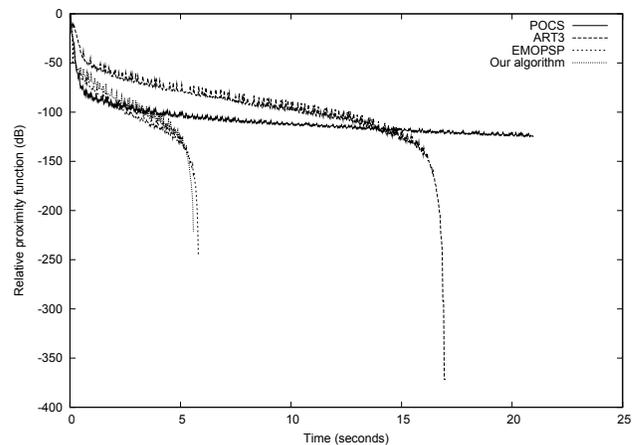


Fig. 4. IMRT experiment: Proximity function versus time.

[2] L. M. Bregman, "The method of successive projection for finding a common point of convex sets," *Soviet Math. Doklady*, vol. 6, pp. 688–692, 1965.

[3] T. D. Capricelli, "Technical report," Univ. Paris 6.

[4] T. D. Capricelli and P. L. Combettes, "A convex programming algorithm for noisy discrete tomography," *Advances in Discrete Tomography and Its Applications*, (G. T. Herman and A. Kuba, Editors), pp. 207–226, Boston, 2007.

[5] Y. Censor and S. A. Zenios, *Parallel Optimization: Theory, Algorithms and Applications*, New York, 1997.

[6] P. L. Combettes, "The foundations of set theoretic estimation," *Proc. IEEE*, vol. 81, pp. 182–208, 1993.

[7] P. L. Combettes, "Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections," *IEEE Trans. Image Process.*, vol. 6, pp. 493–506, 1997.

[8] P. L. Combettes, "Quasi-Fejérian analysis of some optimization algorithms," *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*, (D. Butnariu, Y. Censor, and S. Reich, Eds.), pp. 115–152, New York, 2001.

[9] G. T. Herman, "A relaxation method for reconstructing objects from noisy X-rays," *Math. Programming*, vol. 8, pp. 1–19, 1975.

[10] G. T. Herman and W. Chen, "A fast algorithm for solving a linear feasibility problem with application to intensity-modulated radiation therapy," *Linear Algebra Appl.*, vol. 428, pp. 1207–1217, 2008.

[11] J. Mandel, "Convergence of the cyclical relaxation method for linear inequalities," *Math. Programming*, vol. 30, pp. 218–228, 1984.