Stable Convergence Behavior Under Summable Perturbations of a Class of Projection Methods for Convex Feasibility and Optimization Problems

Dan Butnariu, Ran Davidi, *Student Member, IEEE,* Gabor T. Herman, *Fellow, IEEE*
and Ivan G. Kazantsev, *Member, IEEE*

Abstract

We study the convergence behavior of a class of projection methods for solving convex feasibility and optimization problems. We prove that the algorithms in this class converge to solutions of the consistent convex feasibility problem, and that their convergence is stable under summable perturbations. The class of projection methods we study contains, among many other procedures, the Cimmino algorithm as well as the cyclic projection method. A variant of the approach is proposed to approximate the minimum of a convex functional subject to convex constraints. This variant is illustrated on a problem in image procsssing: namely, for optimization in tomography.

Index Terms

Projection method, convex feasibility, cyclic projection method, Cimmino algorithm, tomographic optimization, total variation.

Dan Butnariu is with the Department of Mathematics, University of Haifa, Israel, 31905. E-mail: dbutnaru@math.haifa.ac.il

Ran Davidi is with the Department of Computer Science, The Graduate Center, City University of New York, 365 Fifth Avenue, New York, N.Y. 10016, U.S.A. E-mail: r_davidi@yahoo.com

Gabor T. Herman is with the Department of Computer Science, The Graduate Center, City University of New York, 365 Fifth Avenue, New York, N.Y. 10016, U.S.A. E-mail: gabortherman@yahoo.com

Ivan G. Kazantsev is with the Department of Computer Science, The Graduate Center, City University of New York, 365 Fifth Avenue, New York, N.Y. 10016, U.S.A. E-mail: kazantsev1@netzero.net

I. INTRODUCTION

The *consistent convex feasibility problem* is: Given m closed and convex subsets $C_1, ..., C_m$ of $\mathbb{R}^n$ such that the set

$$C := \bigcap_{i=1}^{m} C_i \tag{I.1}$$

is nonempty, find a point $x \in C$. Many projection methods are known to provide, in various ways, solutions of this problem (see [1] and [6] for surveys on this topic). Our purpose here is to present a scheme for generating a large class of (metric) projection methods that we call amalgamated projection methods. Among the amalgamated projection methods are most of the already known metric projection methods for solving convex feasibility problems, as well as a plethora of new procedures that, to the best of our knowledge, were not studied before. In particular, we give a variant of the method that can be used to steer a convex functional towards its minimum subject to consistent convex constraints.

If $C \subseteq \mathbb{R}^n$ and $x \in \mathbb{R}^n$, then

$$d(x, C) = \inf \{\|x - y\| : y \in C\}. \tag{I.2}$$

If C is also closed and convex, then there is one and only one $y \in C$, such that $\|x - y\| = d(x, C)$; we call this y the *projection of* x *onto* C and denote it by $P_C(x)$. In what follows, we abbreviate the projection operator P_{C_i} by P_i.

By an *index vector* we mean a vector $\omega = (\omega_1, ..., \omega_p)$ whose coordinates ω_j are in the set $\{1, ..., m\}$. If $\omega = (\omega_1, ..., \omega_p)$ is an index vector, then we denote

$$P[\omega] := P_{\omega_p} \circ ... \circ P_{\omega_1}. \tag{I.3}$$

A set Ω of index vectors is called *fit* (to the given feasibility problem) if, for each $i \in \{1, ..., m\}$, there exists $\omega = (\omega_1, ..., \omega_p) \in \Omega$ such that $\omega_s = i$ for some $s \in \{1, ..., p\}$. We denote by Ω_i the set of those $\omega \in \Omega$ having i among its coordinates. If Ω is a fit set of index vectors, then a function $w : \Omega \to \mathbb{R}_{++} = (0, \infty)$ is called a *fit weight function* if $\sum_{\omega \in \Omega} w(\omega) = 1$. A pair (Ω, w) consisting of a fit set of index vectors and a fit weight function defined on it is called an *amalgamator*. We define the operator $\mathbf{P} : \mathbb{R}^n \to \mathbb{R}^n$ by

$$\mathbf{P}x = \sum_{\omega \in \Omega} w(\omega) P[\omega]\, x. \tag{I.4}$$

This operator is continuous because each P_i is a continuous function on $\mathbb{R}^n$. With these notions and notations, the *amalgamated projection method for the amalgamator* (Ω, w) is:

$$x^0 \in \mathbb{R}^n,$$
$$x^{k+1} = \mathbf{P}x^k, \ \forall k \in \mathbb{N}. \tag{I.5}$$

Many known projection methods can be described as amalgamated projection methods. For instance, if $\Omega = \{\omega\}$ with $\omega = (1, 2, ..., m)$, then $w(\omega) = 1$ and the amalgamated projection method with the amalgamator (Ω, w) gives the cyclic projection method originating in [12]; if the amalgamator (Ω, w) consists of $\Omega = \{\omega_1, ..., \omega_m\}$, where, for each $i \in \{1, ..., m\}$, $\omega_i = (i)$, and $w(\omega_i) = w_i > 0$ with $\sum_{i=1}^m w_i = 1$, then the amalgamated projection method gives a generalization of the Cimmino algorithm originating in [7]. More generally, it can be seen easily by direct comparison that the unrelaxed versions of many projection methods for solving convex feasibility problems, such as those discussed in [1], [5], and [14], can be described as amalgamated projection methods.

In order to see that some relaxed projection methods are also describable as amalgamated projection methods, observe that the convex feasibility problem is equivalent to: Find a common point of the sets $C_1, ..., C_m, \ C_{m+1}$ where

$$C_{m+1} := \mathbb{R}^n. \tag{I.6}$$

In this case, $P_{m+1}x = x$, for all $x \in \mathbb{R}^n$. If (Ω, w) is an amalgamator of the original problem of finding a common point of the sets $C_1, ..., C_m$, then for any $\alpha \in (0, 1)$, the pair (Ω', w'), where $\Omega' = \Omega \cup \{(m+1)\}$ and $w' : \Omega' \to \mathbb{R}_+$ is defined by

$$w'(\omega) = \begin{cases} \alpha w(\omega), & \text{if } \omega \in \Omega, \\ \\ 1 - \alpha, & \text{if } \omega = (m+1), \end{cases} \tag{I.7}$$

is an amalgamator of the problem of finding a common point of the sets $C_1, ..., C_{m+1}$, and the amalgamated projection method for the amalgamator (Ω', w') is the relaxed iterative algorithm:

$$x^0 \in \mathbb{R}^n,$$
$$x^{k+1} = (1 - \alpha)x^k + \alpha \mathbf{P}x^k, \ \forall k \in \mathbb{N}. \tag{I.8}$$

In Section II we prove a theorem that states that an amalgamated projection method converges to a solution of the given consistent convex feasibility problem, and its convergence is stable under summable perturbations of the iterates. Precisely, we prove:

Theorem. *Let C_i, $1 \le i \le m$, be closed convex sets with a nonempty intersection C. If $\{\beta_k\}_{k\in\mathbb{N}}$ is a sequence of positive real numbers such that $\sum_{k=0}^{\infty} \beta_k < \infty$ and $\{v^k\}_{k\in\mathbb{N}}$ is a bounded sequence of vectors in $\mathbb{R}^n$, then for any amalgamator (Ω, w) the procedure*

$$x^0 \in \mathbb{R}^n,$$

$$x^{k+1} = \mathbf{P}(x^k + \beta_k v^k), \ \forall k \in \mathbb{N} \tag{I.9}$$

converges, and its limit is in C.

The stability under perturbations of the convergence of projection methods for solving convex feasibility problems was also considered in [1], where the behavior of such procedures is studied under the assumption that the sets C_i are given by approximations. By contrast, our Theorem deals with the situation in which the sets C_i are precisely given, but the projections P_i on them are determined only approximately. A result similar to our Theorem was obtained in [8] concerning the following: Fix $w \in \mathbb{R}^m_{++}$ such that $\sum_{i=1}^m w_i = 1$ and use the procedure

$$x^0 \in \mathbb{R}^n,$$

$$x^{k+1} = \sum_{i=1}^{m} w_i \left[P_i(x^k) + \beta_k v^k \right], \ \forall k \in \mathbb{N}. \tag{I.10}$$

We have noted that many known algorithms for solving convex feasibility problems can be described as amalgamated projection methods. The functioning of some amalgamated projection methods with not so standard features is described in Section III. In Section IV we present a variant using subgardients that can be used to approximate the solutions of some constrained convex optimization problems [15].

II. Proof of the Theorem

The proof is in two stages. In the first, we prove that the Theorem is true when there are no perturbations of the computational process, that is, we prove:

Claim 1: If $\beta_k = 0$ for all $k \in \mathbb{N}$, then the Theorem is true.

Note that, in this instance, the procedure (I.9) reduces to (I.5). Let $i \in \{1, ..., m\}$ and recall that the projection operator P_i is 1-attracting with respect to C_i (in the sense given to this term in [1]), that is, it satisfies

$$\|z - P_i x\|^2 + \|P_i x - x\|^2 \le \|z - x\|^2, \tag{II.1}$$

for any $z \in C_i$ and $x \in \mathbb{R}^n$. Therefore, according to [1, Proposition 2.10], the operator $P[\omega]$ is $2^{-p(\omega)}$-attracting with respect to the set C, that is, for any $z \in C$ and $x \in \mathbb{R}^n$,

$$\|z - P[\omega]x\|^2 + \frac{1}{2^{p(\omega)}} \|P[\omega]x - x\|^2 \leq \|z - x\|^2 , \tag{II.2}$$

where $p(\omega)$ is the length of the index vector ω. In particular, for any $z \in C$ and $k \in \mathbb{N}$, we have that

$$\left\|z - P[\omega]x^k\right\|^2 + 2^{-p(\omega)} \left\|P[\omega]x^k - x^k\right\|^2 \leq \left\|z - x^k\right\|^2 . \tag{II.3}$$

Let

$$p_* := \max_{\omega \in \Omega} p(\omega). \tag{II.4}$$

By (II.3) we deduce that, whenever $z \in C$ and $k \in \mathbb{N}$,

$$\left\|z - P[\omega]x^k\right\|^2 + 2^{-p_*} \left\|P[\omega]x^k - x^k\right\|^2 \leq \left\|z - x^k\right\|^2 . \tag{II.5}$$

Multiplying both sides of this inequality by $w(\omega)$, summing up the resulting inequalities, and taking into account that, for each $u \in \mathbb{R}^n$, the function $x \to \|u - x\|^2$ is convex, we deduce that, for all $z \in C$ and $k \in \mathbb{N}$,

$$\left\|z - x^{k+1}\right\|^2 + 2^{-p_*} \left\|x^{k+1} - x^k\right\|^2 \leq \left\|z - x^k\right\|^2 , \tag{II.6}$$

and, consequently,

$$\left\|z - x^{k+1}\right\| \leq \left\|z - x^k\right\| . \tag{II.7}$$

This implies that the sequence $\left\{\left\|z - x^k\right\|\right\}_{k \in \mathbb{N}}$ converges. By (II.6) we have that, for all $z \in C$ and $k \in \mathbb{N}$,

$$2^{-p_*} \left\|x^{k+1} - x^k\right\|^2 \leq \left\|z - x^k\right\|^2 - \left\|z - x^{k+1}\right\|^2 ,$$

and the right hand side converges to zero. Hence,

$$\lim_{k \to \infty} \left\|x^{k+1} - x^k\right\| = 0. \tag{II.8}$$

Since C is nonempty and, for $z \in C$, the sequence $\left\{\left\|z - x^k\right\|\right\}_{k \in \mathbb{N}}$ converges (and, hence, is bounded), the sequence $\left\{x^k\right\}_{k \in \mathbb{N}}$ is bounded and, thus, has cluster points. Let x^* be such a cluster point and let $\left\{x^{i_k}\right\}_{k \in \mathbb{N}}$ be a subsequence of $\left\{x^k\right\}_{k \in \mathbb{N}}$ converging to x^*. Then

$$x^* = \lim_{k \to \infty} x^{i_k} = \lim_{k \to \infty} x^{i_k + 1} = \lim_{k \to \infty} \mathbf{P} x^{i_k} = \mathbf{P} x^*, \tag{II.9}$$

where the second equality follows from (II.8), the third one from (I.5), and the last one is due to the continuity of $\mathbf{P}$. This implies that x^* is a fixed point of $\mathbf{P}$. According to [1, Proposition 2.12(i)],

$$\operatorname{Fix} \mathbf{P} = \bigcap_{\omega \in \Omega} \operatorname{Fix} P[\omega]. \tag{II.10}$$

Hence,

$$x^* \in \mathrm{Fix}\, P[\omega],\ \forall \omega \in \Omega, \tag{II.11}$$

where, by [1, Proposition 2.10(ii)], if $\omega = (\omega_1, ..., \omega_p)$, then

$$\mathrm{Fix}\, P[\omega] = \bigcap_{j=1}^{p} \mathrm{Fix}\, P_{\omega_j} = \bigcap_{j=1}^{p} C_{\omega_j}. \tag{II.12}$$

Note that since Ω is a fit set of index vectors, for each $i \in \{1, ..., m\}$, there exists $\omega \in \Omega_i$. Consequently, for each $i \in \{1, ..., m\}$ and for each $\omega \in \Omega_i$, we have that

$$x^* \in \mathrm{Fix}\, P[\omega] \subseteq \mathrm{Fix}\, P_i = C_i. \tag{II.13}$$

This implies that $x^* \in C$. Since x^* was an arbitrary cluster point of $\left\{x^k\right\}_{k\in\mathbb{N}}$, it follows that all cluster points of $\left\{x^k\right\}_{k\in\mathbb{N}}$ are contained in C. By [1, Proposition 2.16(ii)] the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$, having the property (II.7), can have at most one cluster point in C. Hence, the sequence $\left\{x^k\right\}_{k\in\mathbb{N}}$ has a single cluster point, i.e., it converges to a point in C and the proof of Claim 1 is complete.

To complete the proof of the Theorem, we prove:

Claim 2: If $\{\beta_k\}_{k\in\mathbb{N}}$ *and* $\left\{v^k\right\}_{k\in\mathbb{N}}$ *are sequences satisfying the hypotheses of the Theorem, then any sequence generated according to* (I.9) *converges, and its limit is in* C.

Observe that in the procedure (I.9), the operator $\mathbf{P}$ is nonexpansive. Thus, for each $k \in \mathbb{N}$, we have that

$$\begin{aligned} \left\|x^{k+1} - \mathbf{P}x^k\right\| &= \left\|\mathbf{P}(x^k + \beta_k v^k) - \mathbf{P}x^k\right\| \\ &\leq \left\|(x^k + \beta_k v^k) - x^k\right\| = \beta_k \left\|v^k\right\|. \end{aligned} \tag{II.14}$$

Consequently, denoting by M a finite upper bound of the bounded sequence $\left\{\left\|v^k\right\|\right\}_{k\in\mathbb{N}}$, we have that

$$\sum_{k=0}^{\infty} \left\|x^{k+1} - \mathbf{P}x^k\right\| \leq M \sum_{k=0}^{\infty} \beta_k < \infty, \tag{II.15}$$

showing that the sequence $\left\{\left\|x^{k+1} - \mathbf{P}x^k\right\|\right\}_{k\in\mathbb{N}}$ is summable. According to [2, Theorem 4], this summability implies that the sequences generated by (I.9) converge if, and only if, the sequences generated by (I.5) converge, and the summability also implies that if the sequences given by (I.5) converge to fixed points of $\mathbf{P}$, then the sequences given by (I.9) also converge to fixed points of $\mathbf{P}$. Taking into account the already proved Claim 1 that shows that the sequences generated according to (I.5) converge to elements of C (which, obviously, are fixed points of $\mathbf{P}$), we deduce that the sequences generated by (I.9) converge to fixed points of $\mathbf{P}$. Since, according to (II.10) and (II.12) combined with the fact that Ω is a fit set of index vectors, we have that

$$\begin{aligned} C \subseteq \operatorname{Fix} \mathbf{P} &= \bigcap_{\omega \in \Omega} \operatorname{Fix} P[\omega] \\ &= \bigcap_{i=1}^{m} \bigcap_{\omega \in \Omega_i} \operatorname{Fix} P[\omega] \subseteq \bigcap_{i=1}^{m} C_i = C, \end{aligned} \tag{II.16}$$

and so $C = \operatorname{Fix} \mathbf{P}$ and, hence, the sequences generated by (I.9) converge to points in C. □

III. Illustrations

In this section we illustrate the way in which the amalgamated projection methods function in some particular instances and how they compare with the cyclic projection method whose convergence behavior is well researched.

Consider the amalgamator (Ω, w) for which Ω consists of m index vectors of the form

$$(1), (2, 1), (3, 2, 1), \ldots, (m, m-1, ..., 1), \tag{III.1}$$

and $w(\omega) = 1/m$. The functioning of the corresponding algorithm (which we call here the averaged projection method) in the case $m = 3$ is illustrated by Figure 1, where two iterates x^1 and x^2 are indicated in parallel with the first two iterates of the cyclic projection method starting from the same initial point x^0. In the illustrated case, the second iterate of the cyclic projection method is closer to the common point of the lines C_1, C_2 and C_3 than the second iterate of the averaged projection method.

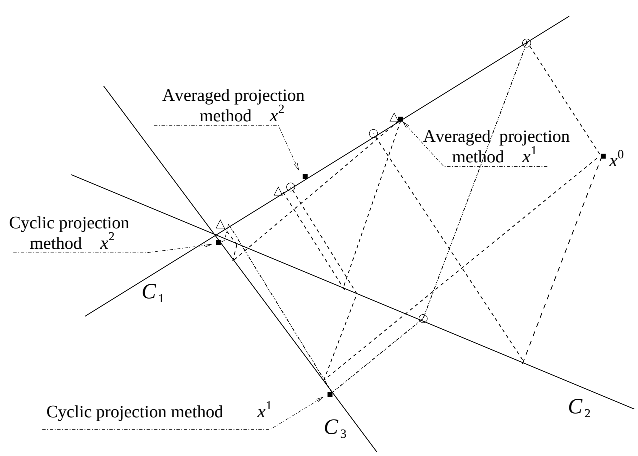


Fig. 1. Illustration of the cyclic projection and the averaged projection methods for $m = 3$.

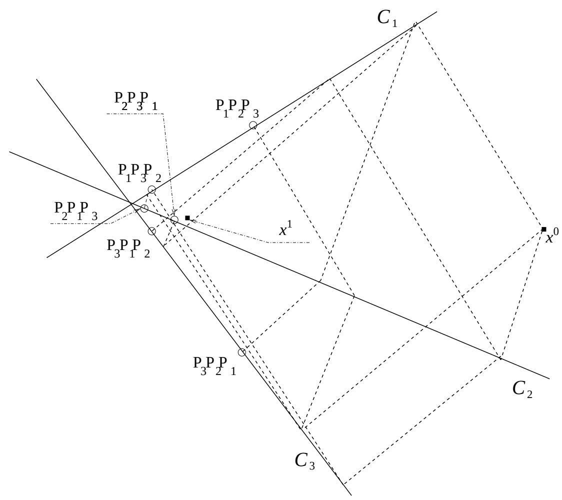


Fig. 2. A step of the algorithm with amalgamator (Ω, w) for which Ω contains all permutations of $\{1, 2, 3\}$ and $w(\omega) = 1/6$ for all ω.

Another possible amalgamator (Ω, w) is one for which Ω consists of all $m!$ permutations of the set $\{1, ..., m\}$ and $w(\omega) = 1/(m!)$ for each $\omega \in \Omega$. The functioning of the resulting amalgamated projection method is illustrated in Figure 2. It is interesting to note that, in Figure 2, the point marked $P_3P_2P_1$ is, in fact, the first iterate of the cyclic projection method. This is not as close to the solution of the feasibility problem as the iterate x^1 generated by the amalgamated projection method.

IV. Applicability to Optimization in Tomography

The Theorem guarantees the convergence of the amalgamated projection method when the calculation of the iterates is affected by summable perturbations. We can make use of this property to steer the iterates towards the minimizer of a given convex function.

Consider a convex function $\phi : \mathbb{R}^n \to \mathbb{R}$ which has a minimizer over the set C. For any $k \in \mathbb{N}$, let $s^k \in \partial\phi(x^k)$, the subgradient of ϕ at x^k, and define

$$v^k = \begin{cases} -\frac{s^k}{\|s^k\|}, & \text{if } s^k \neq 0, \\ \\ 0, & \text{if } s^k = 0. \end{cases} \tag{IV.1}$$

Clearly, the sequence $\{v^k\}_{k\in\mathbb{N}}$ defined by (IV.1) is bounded. Therefore, by the Theorem, for any summable sequence of positive real numbers $\{\beta_k\}_{k\in\mathbb{N}}$, the sequence $\{x^k\}_{k\in\mathbb{N}}$ generated according to (I.9) converges to an element of C.

The relevance of this is seen in the context of the theory of projection subgradient methods for solving constrained convex optimization problems [15]. The iterative step of a projection subgradient method is

$$y^{k+1} = P_C(y^k + \beta_k v^k), \tag{IV.2}$$

where the vectors v^k are given by (IV.1) and P_C is the metric projection operator onto the set C. Determining projections onto C can be a computationally complicated process even if the sets C_i defined by the constraints of the problem of minimizing ϕ over C are relatively simple sets, such as hyperplanes or half-spaces. By contrast, calculating the projections onto the individual C_i is frequently easy and, in such cases, implementation of (I.9) instead of (IV.2) can be very advantageous from the point of view of the computational burden.

In our implementation, we use the following methodology for generating the real numbers $\{\beta_k\}_{k\in\mathbb{N}}$. We define, for an $x \in \mathbb{R}^n$,

$$\text{Res}(x) = \sqrt{\sum_{i=1}^{m} [d(x, C_i)]^2}. \tag{IV.3}$$

Clearly, $x \in C$ if, and only if, $\text{Res}(x) = 0$. Furthermore, if $\text{Res}(x) > 0$, then its size indicates how badly x violates the given collection $\{C_1, \ldots, C_m\}$ of constraints. An approximate solution x to the convex optimization problem (for ϕ) under these constraints should have a small value of $\text{Res}(x)$ and should aim at finding, among all x with similar (or smaller) value of $\text{Res}(x)$, an x for which $\phi(x)$ is small relative to the others. Guided by this principle, we generate $\{\beta_k\}$ as follows. We initialize β to be an arbitrary positive number. (We have always used $\beta = 1$.) The original value of β is in fact the β_0 in (I.9). In the process of the iterative step from x^k to x^{k+1}, we also update the value of β, which is (in the notation of (I.9)) β_k at the beginning of the iterative step and β_{k+1}, at its end, according to the following pseudocode (in which v^k is defined by (IV.1)).

logic = true;
while (logic)
    $z = x^k + \beta v^k$
    if ($\phi(z) \leq \phi(x^k)$)
        then
            $x^{k+1} = \mathbf{P}(z)$
            if ($\text{Res}(x^{k+1}) < \text{Res}(x^k)$)
                then logic = false
                else $\beta = \beta/2$

10: else $\beta = \beta/2$

We terminate the iterative process when we find an x^k such that $\text{Res}(x^k) < \epsilon$, where ϵ is a user-specified small positive number.

We illustrate our approach on an example from image processing: tomographic reconstruction of images that are not uniquely determined from the available data, with the help of a convex functional ϕ that assigns to each image a nonnegative number that indicates, in some sense, the "undesirability" of the image. (For example, we may know that most images in our application area should be "piece-wise smooth." In that case, $\phi(x)$ should measure the extent to which piece-wise smoothness is violated in the image represented by x.)

Figure 3(a) shows a phantom that is a 243×243 digitized image (thus $n = 59,049$), representing a cross-section of a human head [11, Section 4.3]. The components of x represent the average X-ray attenuations within the $59,049$ pixels, each of each is of size 0.0752×0.0752 (the assumed unit of length is 1 cm). The values of these components range from 0 to 0.5637; for display purposes, any value below 0.1945 is shown as black (gray value 0) in Figure 3, and any value above 0.2200 is shown as white (gray value 255), with a linear mapping of the x-component values into gray values in between. Data were collected by calculating line integrals through the digitized image for 82 sets of equally spaced parallel lines. Each such line integral gives rise to a linear equation in the components of x; the set of x that is consistent with such a line integral is a hyperplane in $\mathbb{R}^n$. The phantom itself lies in the intersection of all these hyperplanes. In our experiments, we used measurements for $25,452$ lines, making our problem very much underdetermined. (The intersection of all the hyperplanes is an at least $33,597$-dimensional subspace of $\mathbb{R}^{59,049}$). In the terminology of our paper $m = 25,452$ and, for $1 \leq i \leq m$, C_i is one of the hyperplanes.

A classical method for finding a common point in such hyperplanes is the cyclic projection method, which is commonly known as ART in tomography [11, Chapter 11]. If it is initialized with x^0 being the zero vector, it is known [11, Section 11.2] to converge to that point x in the intersection of the hyperplanes for which $\|x\|$ is minimal. In practice, ART needs to be stopped after a finite number of steps. In Figure 3(b) we show the result obtained by ART when we stopped at a k for which $\text{Res}(x^k) < 0.005$. On an Intel Xeon 1.7 MHz processor 1 G RAM workstation, obtaining such a good fit to the data by ART required just over 112 minutes. (To demonstrate that 0.005 is indeed a small value in our context, we point out that $\text{Res}(x^0) = 330.204$.)

There are some obvious differences between the phantom in Figure 3(a) and the ART reconstruction in Figure 3(b). This indicates that $\|x\|$ may not be a particularly good measure of the undesirability of

TABLE I

NUMERICAL VALUES FOR THE OUTPUTS OF THE ALGORITHMS

Method	norm	TV	distance	time
ART	46.04	1300.8	3.644	112.6
TV-reducing (ours)	46.19	457.2	0.223	14.6
TV-minimizing (from [9])	46.17	471.3	0.390	59.9

x in this situation. Many research workers in image processing have been advocating the use of total variation rather than the norm; e.g., [3], [10], [13], [16], [17], and [18]. For a $K \times L$ image p whose pixels are denoted by $p_{k,l}$ $(1 \le k \le K, 1 \le l \le L)$, the *total variation* (TV) of p is

$$TV(p) = \sum_{k=1}^{K-1} \sum_{l=1}^{L-1} \sqrt{(p_{k+1,l} - p_{k,l})^2 + (p_{k,l+1} - p_{k,l})^2}. \tag{IV.4}$$

By mapping p into a ($K \times L$)-dimensional vector x (by stacking into a single column all the columns of p), this definition gives rise to a functional ϕ that can be used in our algorithm designed above. As for ART, we selected $\mathbf{P}$ to be that of the cyclic projection method. The only difference between ART and this new algorithm comes from the perturbations aimed at reducing the total variation. Again we started the process with x^0 being the zero vector and stopped it when Res$(x^k) < 0.005$. Figure 3(c) illustates the output of the algorithm. It is visually superior to the reconstruction of Figure 3(b). (As a numerical measure, the norm of the difference between the ART reconstruction and the phantom is more than 16 times greater than the norm of the difference between the TV-reducing reconstruction and the phantom.)

What is particularly interesting is that the TV-reducing algorithm is significantly less expensive than ART: The total time required was less than 15 minutes. The reduction in Res(x^k) as a function of computer time is plotted in Figure 4 for both algorithms. Even though a single iterative step of ART is less expensive than that of the TV-reducing algorithm, the perturbations in the latter steer it towards the correct result (i.e., in the general direction of the phantom) and so much fewer steps are needed to get Res(x^k) below a given ϵ.

In the first two rows of Table I, we report on the values of the norm and TV for the outputs of the two algorithms (as well as the distance between the reconstruction and the phantom and the time in minutes needed to obtain the reconstruction). As can be seen, the algorithms tend to minimize the function that they are supposed to be minimizing; the superiority of the reconstruction in Figure 3(c) to that in Figure 3(b) is due to TV minimization being a more appropriate aim than norm minimization in the current circumstances.

Finally, to show how well our new method is doing at minimizing TV as compared to an established procedure, we ran a TV-minimizing version of Algorithm 6 in [9] on our data. Again, starting the process with the zero vector and stopping it when $\mathrm{Res}(x^k) < \epsilon = 0.005$, we obtained an output for which the TV value is 471.3, which is not quite as low as obtained by our new algorithm (457.2). Also, the time required to get to termination was just over 4 times longer for the algorithm of [9] than for our new algorithm. All the relevant numbers for this algorithm are also reported in Table I.

All the computational work reported here was done using snark05 [4]; the phantom, the data, the reconstructions, displays, and plots were all generated within this same environment. In particular, this implies that differences in the reported reconstruction times are not due to the different algorithms being implemented in different environments.

V. Summary

We have proposed a class of projection methods for solving the consistent convex feasibility problem and proved their convergence to a feasible point even under summable perturbations. We have discussed how this property can be used to steer the iterative process towards minimizing a convex function, and have illustrated the potential usefulness of this by applying it to the problem of total variation mimimization in tomography.

Acknowledgements

This research is supported by the National Institutes of Health through grant HL70472. Dan Butnariu's work on this paper was done during his 2006 visit to the Discrete Imaging and Graphics group of the Graduate Center of the City University of New York, and he gratefully acknowledges the inspiring scientific activities in which he was invited to take part, as well as the financial support from the grant mentioned above that made his participation possible. We thank Patrick Combettes for his help with the implementation of the TV-minimizing algorithm based on [9].

References

[1] H. H. Bauschke and J. M. Borwein, "On projection algorithms for solving convex feasibility problems," *SIAM Review*, vol. 38, pp. 367–426, 1996.

[2] D. Butnariu, S. Reich, and A. Zaslavski, "Convergence to fixed points of inexact orbits of Bregman-monotone and nonexpansive operators in Banach spaces," in: H. Fetter et al. (ed.) Proc. Intl. Conf. of Fixed Point Theory and Applications, Guadajuato, Mexico, *Yokohama Publishers*, pp. 11-32, 2006.

[3] E. J. Candes, and F. Guo, " New multiscale transforms, minimum total variation synthesis: application to edge-preserving image reconstruction," *Signal Processing*, vol. 82, pp. 1519-1543, 2002.

[4] B. M. Carvalho, W. Chen, J. Dubowy, G. T. Herman, M. Kalinowski, H. Y. Liao, L. Rodek, L. Rusko, S. W. Rowland, E. Vardi-Gonen, " SNARK05: A Programming System for the Reconstruction of 2D Images from 1D Projections," *CUNY Institute for Software Design and Development*, New York, http://www.snark05.com/.

[5] Y. Censor, T. Elfving, and G. T. Herman, " Averaging strings of sequential iterations for convex feasibility problems," in "Inherently Parallel Algorithms in Feasibility and Optimization and their Applications", pp. 101–113, *North-Holland*, Amsterdam, 2001.

[6] Y. Censor and S. Zenios, " Parallel Optimization: Theory, Algorithms and Applications," *Oxford University Press*, New York, 1997.

[7] G. Cimmino, " Calcolo approsimato per le soluzioni di sistemi di ecuazioni lineari," *La Riserca Scientifica, Roma*, vol. 2, pp. 326-333, 1938.

[8] P. L. Combettes, " On the numerical robustness of a parallel projection method in signal synthesis," *IEEE Signal Processing Letters*, vol. 8, pp. 45–47, 2001.

[9] P. L. Combettes and J. Luo, " An adaptive level set method for nondifferentiable constrained image recovery", *IEEE Transactions on Image Processing*, vol. 11, pp. 1295–1304, 2002.

[10] B. G. Fitzpatrick, S. L. Keeling, " On approximation in total variation penalization for image reconstruction and inverse problems", *Numerical Functional Analysis and Optimization*, vol. 18, pp. 941-958, 1997.

[11] G. T. Herman, " Image Reconstruction from Projections. The Fundamentals of Computerized Tomography." *Academic Press*, New York, 1980.

[12] J. von Neumann: " Functional Operators - Vol. II, The Geometry of Orthogonal Spaces," *Princeton University Press*, 1950. [Reprint of mimeographed lecture notes distributed in 1933.]

[13] K. D. Paulsen, and H. B. Jiang, " Enhanced frequency-domain optical image reconstruction in tissues through total-variation minimization," *Applied Optics*, vol. 35, pp. 3437-3458, 1996.

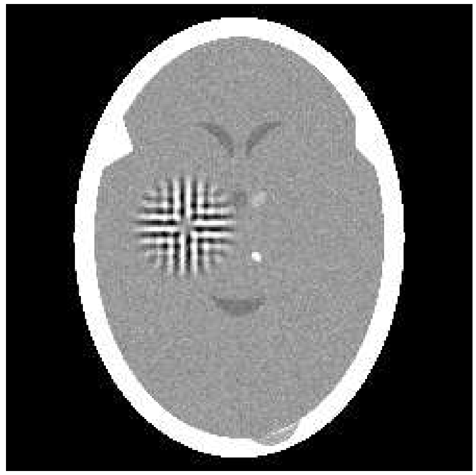
[14] A. R. De Pierro and A. N. Iusem, " A parallel projection method for finding a common point of a family of convex sets," *Pesquisa Operational*, vol. 5, pp. 1-20, 1985.

[15] B. T. Polyak, " Introduction to Optimization," *Optimization Software Inc.*, New York, 1987.

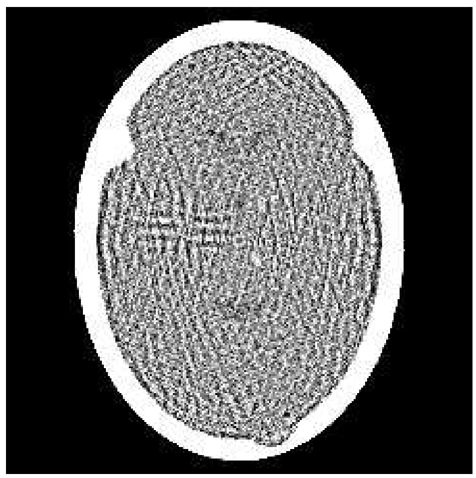
[16] S. A. Sarra, " Digital total variation filtering as postprocessing for radial basis function approximation methods," *Computers & Mathematics with Applications*, vol. 51, pp. 1119-1130, 2006.

[17] L. A. Vese, and S. J. Osher, " Modeling textures with total variation minimization and oscillating patterns in image processing," *Journal of Scientific Computing*, vol. 19, pp. 553-572, 2003.

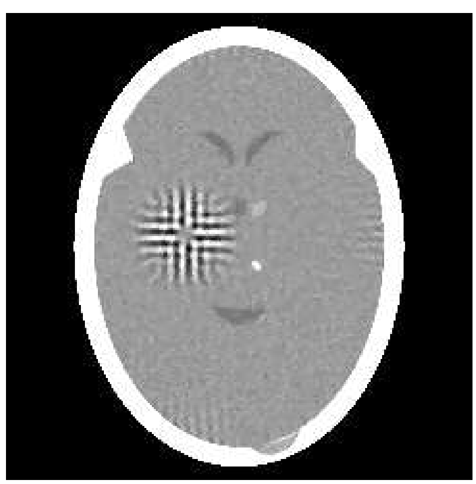
[18] G. Q. Yu, L. Li, J. W. Gu, and L. Zhang, " Total variation based iterative image reconstruction", *Computer Vision for Biomedical Image Applications Proceedings*, *Lecture Notes in Computer Scince*, vol. 3765, pp. 526-534, 2005.



(a)



(b)



(c)

Fig. 3. Illustrative example of optimization in tomography. (a) The phantom for which data were collected. (b) Norm-minimizing reconstruction (cyclic projection method, ART). (c) TV-reducing reconstruction (cyclic projections with perturbations).

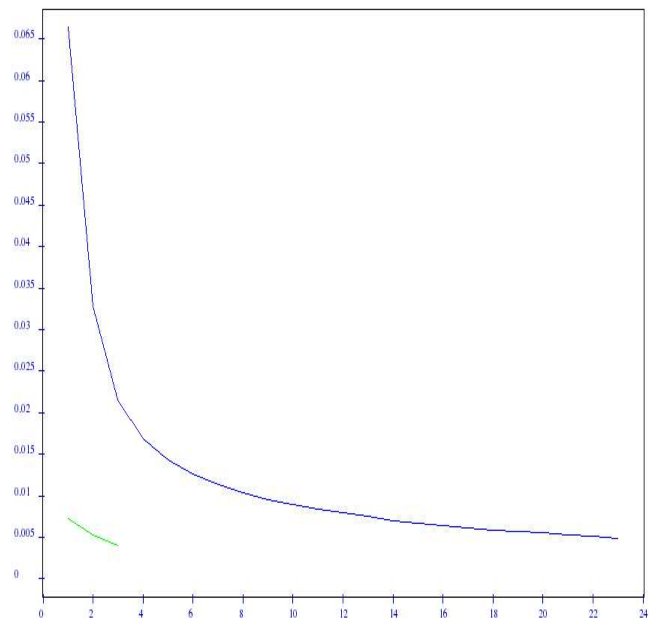


Fig. 4. Plots of $\text{Res}(x^k)$ for ART (blue) and the new TV-reducing algorithm (green) both plotted against computer time.