Published in final edited form as:

Inverse Probl. 2012 March 1; 28(3): . doi:10.1088/0266-5611/28/3/035005.

Accelerated perturbation-resilient block-iterative projection methods with application to image reconstruction

T Nikazad¹, R Davidi², and G. T. Herman³

T Nikazad: tonik@mai.liu.se; R Davidi: rdavidi@stanford.edu; G. T. Herman: gabortherman@yahoo.com

¹Department of Mathematics, Iran University of Science and Technology, Narmak, Tehran, 16846-13114, Iran

²Department of Radiation Oncology, Stanford University, Stanford, CA 94305 USA

³Department of Computer Science, The Graduate Center, City University of New York, NY 10016, USA

Abstract

We study the convergence of a class of accelerated perturbation-resilient block-iterative projection methods for solving systems of linear equations. We prove convergence to a fixed point of an operator even in the presence of summable perturbations of the iterates, irrespective of the consistency of the linear system. For a consistent system, the limit point is a solution of the system. In the inconsistent case, the symmetric version of our method converges to a weighted least squares solution. Perturbation resilience is utilized to approximate the minimum of a convex functional subject to the equations. A main contribution, as compared to previously published approaches to achieving similar aims, is a more than an order of magnitude speed-up, as demonstrated by applying the methods to problems of image reconstruction from projections. In addition, the accelerated algorithms are illustrated to be better, in a strict sense provided by the method of statistical hypothesis testing, than their unaccelerated versions for the task of detecting small tumors in the brain from X-ray CT projection data.

Keywords

image reconstruction from projections; block-iterative algorithms; superiorization; perturbation resilience; total variation

1. Introduction

The usefulness to image processing of perturbation-resilient block-iterative projection methods, for solving a linear system of equations, comes from the fact that each of the iterates can be perturbed in a way that reflects some desirable property of the images (for example, small total variation) without losing convergence to an image consistent with the equations. Recent works [3, 12, 22, 32] demonstrate that this approach produces high quality reconstructions from projections, but this is often at a great computational cost. In the current paper, we introduce a class of perturbation-resilient block-iterative projection methods that allow us to *accelerate* the iterative process (as compared to the iterative process in [12], to which we are going to refer as *unaccelerated* or *without acceleration*), by a proper selection of relaxation parameters, while retaining convergence to a solution appropriate for the system of equations. This acceleration is made possible by two new theorems that prove convergence

of perturbed block-iterative procedures (for all problem formulations to which they can be applied) for a range of relaxation parameters that is much larger than what was provided by previously existing convergence theorems. The importance of this contribution is of a practical nature: in image processing applications, for example in computerized tomography (CT), such algorithms can be more than an order of magnitude faster than their unaccelerated versions. Most fortunately, it also turns out to be the case that the reconstructions produced by the accelerated algorithms may be medically more efficacious, as we demonstrate below for the task of detecting small tumors in the brain. However, the results presented in the next section have a general applicability quite beyond CT, they can be used to accelerate block-iterative projection methods for any of their many applications to inverse problems.

Consider the linear system of equations (that may or may not be consistent)

$$Ax=b, \quad (1)$$

where $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. The basic idea of a *block-iterative algorithm* is to partition A and b of the system (1) into *blocks* of equations that can be written as

$$A = \begin{pmatrix} A_1 \\ A_2 \\ \vdots \\ A_p \end{pmatrix}, \quad b = \begin{pmatrix} b^1 \\ b^2 \\ \vdots \\ b^p \end{pmatrix}, \quad (2)$$

and then use exactly one block in any atomic step of the algorithm. This methodology was proposed in [15, 16] and gained popularity in the image reconstruction literature under the name of *ordered subsets* methods [23]; see also [2, 11]. The algorithmic model of block iterations is a special case of *asynchronous iterations*; see, e.g., [18, 19]. Block-iterative methods have also been used to enhance computational efficiency using parallel processors [9, 25]. For an overview in a more general context, see [10].

Many block-iterative methods can be written in the form of the block-iterative Algorithm 1 below, which was proposed in [16] for the special case when $\lambda_t = \lambda$, and $M_t = (A_t A_t^T)^{-1}$. The method was generalized in [15] to periodic relaxation, also allowing arbitrary positive definite symmetric matrices M_t .

Among the block-iterative schemes, there are two extreme cases. One is when $p = 1$, in which case there is just one block $M_t = M \in \mathbb{R}^{m \times m}$. The other is when $p = m$, in which case each block consists of a single row and $M_t \in \mathbb{R}^{1 \times 1} = \mathbb{R}$, for $t = 1, 2, \dots, m$. The former is called the *fully simultaneous* case, while the latter is named *fully sequential*, an example of which is *Kaczmarz's method* [26]. (In the image reconstruction from projections literature, the simultaneous methods are often referred to as being SIRT-type, while the sequential ones are often referred to as being ART-type; see, e.g., [21, Chapters 11 and 12].)

As specified in Algorithm 1, an *iterative step* gets us from x^k to x^{k+1} . It consists of a sequence of *subiterative steps* (referred to as “atomic steps” in the discussion below (2)) that get us from $x^{k,t-1}$ to $x^{k,t}$. We define a *cycle* as one pass through all the blocks. Under the circumstances specified above, any sequence of p consecutive subiterative steps completes a cycle. The sequence of all subiterates will not in general converge for inconsistent data; however, the subsequences $x^{0,t}, x^{1,t}, \dots$, with a fixed t , will converge under some conditions on λ_t and M_t [15]. This is sometimes called *cyclic convergence*. In particular, since $x^{k,0} = x^k$, for all k , cyclic convergence implies that the sequence of iterates $x^0, x^1, \dots$ converges. (There are

generalizations of Algorithm 1 that achieve convergence, and not just cyclic convergence, of the sequence of subiterates even in the inconsistent case; see, for example, [24]. In such generalizations the relaxation parameters depend not only on the block index t but also on the iteration number k . In what follows we will not be using such generalizations.)

In this paper we ask the following question: Is Algorithm 1 stable under perturbations of the iterates occurring during the iterative process? A series of papers studied a similar question. In [3] the authors considered amalgamated projection methods, which form a subfamily of string-averaging projection schemes (see, e.g., [6, 8]), for solving a consistent convex feasibility problem. In a recent paper [12], a perturbation-resilient block-iterative projection method for solving the consistent convex feasibility problem was proposed. Both papers showed convergence as long as the perturbations are bounded and summable. Such convergence results were proved assuming specified ranges of the relaxation parameters λ_t . However, in some cases, depending on the choice of the block-iterative method, the range of relaxation parameters can be larger than the ones specified in those results without ruining convergence. In fact the larger range can sometimes be used to accelerate the iterative process; see, e.g., [7].

In the following, we provide results that not only answer positively the question raised at the beginning of the previous paragraph (for both the consistent and inconsistent cases), but they do so for wider ranges of relaxation parameters than the ones that appeared in previously published convergence results for perturbation-resilient block-iterative projection methods. This, as we demonstrate below, enables us to accelerate the iterative process for finding an appropriate solution of the system of equations.

The exact form of the perturbation model is motivated by our intended application (discussed in Section 4) by defining the perturbations in such a way as to steer the algorithmic process towards a minimizer of a given convex function. However, our results are immediately applicable to the alternative point of view in which perturbations are considered to be numerical errors due to hardware limitations. We note also that the perturbation-resilient property of iterative methods can be viewed from another perspective. In the optimization community, such perturbation-resilient property have been studied as the stability problem with optimization methods, for example, in [27, 28]. These approaches are different in nature from the approach taken here. As explained in [5], for the problem of finding a common point in the intersection of a finite number of convex sets, there often exist iterative algorithms that impose very little demand on computer resources. For other problems, such as finding that point in the intersection at which the value of a given function is optimal, algorithms tend to need more computer memory and longer execution time; the perturbation-resilient algorithms of [27, 28] fall into this category. What is proposed below makes use of the perturbation-resilience of algorithms in the first category to steer them towards the optimal point, but without an essential increase on their demand on computer resources.

We now describe the contents of our paper. In Section 2, we prove results regarding the block-iterative method, namely that its convergence to a fixed point of a certain operator is stable under summable perturbations of the iterates irrespective of the consistency of the linear system in (1). The limit point of the iterative sequence is characterized in terms of the original linear system and we show that the set of fixed points of the operator is equal to the set of solutions for consistent linear systems. To handle the inconsistent case we consider the symmetric version of the block-iterative method. It is shown that the iterates converge to a certain weighted least squares solution as long as the perturbations are bounded and summable. In Section 3 we describe the superiorization approach and discuss it in the context of our paper. While this approach is applicable to diverse inverse problems, in this paper we restrict ourselves to demonstrating its usefulness to problems of image reconstruction from projections. In Section

4 we report on the results of statistical hypothesis testing evaluations of the relative efficacy of our newly proposed reconstruction algorithms when they are used for the task of detecting low-contrast small tumors in the brain from X-ray CT projection data. Finally, in Section 5, we give some concluding remarks.

2. Accelerated perturbed block iterations

In this section we propose two accelerated perturbation-resilient block-iterative algorithms for solving linear systems of equations (1). We state and prove theorems that specify their convergence behaviors. We assume familiarity with basic definitions, terminology and results of matrix theory; these can be found, for example, in Sections 2.1–2.3 of [34]. We denote the canonical Euclidean inner product by $\langle \cdot, \cdot \rangle$ and its corresponding norm by $\|\cdot\|$.

Consider blocks of equations as in (2). For $t = 1, 2, \dots, p$, let ℓ_t denote the dimension of the vector b^t . To specify our algorithms, we need to choose positive definite symmetric $\ell_t \times \ell_t$ matrices M_t and real numbers λ_t , such that

$$0 < \varepsilon \leq \lambda_t \leq \frac{2-\varepsilon}{\rho(A_t^T M_t A_t)}, \quad (3)$$

where $0 < \varepsilon < 2$ and, for any square matrix W , $\rho(W)$ denotes the spectral radius of W . With these choices made, we define the operators $U_t: \mathbb{R}^n \rightarrow \mathbb{R}^n$ by

$$U_t x = (I - \lambda_t A_t^T M_t A_t)x + \lambda_t A_t^T M_t b^t, \quad (4)$$

where I is the $n \times n$ identity matrix, and the operator

$$\mathbf{U} = U_p \cdots U_2 U_1. \quad (5)$$

For any operator $\mathbf{O}: \mathbb{R}^n \rightarrow \mathbb{R}^n$, we use $\text{Fix}(\mathbf{O})$ to denote the set $\{x \in \mathbb{R}^n \mid \mathbf{O}(x) = x\}$ of *fixed points* of $\mathbf{O}$. We say that $\mathbf{O}$ is *nonexpansive* if $\|\mathbf{O}x - \mathbf{O}y\| \leq \|x - y\|$, for all $x, y \in \mathbb{R}^n$, and that $\mathbf{O}$ is *attracting with respect to F* , where F is a subset of $\mathbb{R}^n$, if for every $x \in \mathbb{R}^n \setminus F$ and $y \in F$, $\|\mathbf{O}x - y\| < \|x - y\|$.

Theorem 1—Let $\{\beta_k\}_{k \in \mathbb{N}}$ be a sequence of nonnegative real numbers such that

$\sum_{k=0}^{\infty} \beta_k < \infty$, let $\{v^k\}$ be a bounded sequence of vectors in $\mathbb{R}^n$ and $x^0 \in \mathbb{R}^n$. The sequence defined by

$$x^{k+1} = \mathbf{U}(x^k + \beta_k v^k), \quad \text{for all } k \in \mathbb{N}, \quad (6)$$

converges to a fixed point of $\mathbf{U}$. Furthermore, if the system (1) of equations is consistent, then the fixed points of $\mathbf{U}$ are the solutions of (1).

The first part of this theorem states that the block-iterative algorithm with perturbations that is specified by (6) converges, whether or not the system of equations is consistent. The last sentence implies that if the system is consistent, then the algorithm converges to a solution of it.

To get a similarly satisfactory result in the inconsistent case, we need an alternative algorithm: Given the system (1) of equations and a symmetric positive definite $m \times m$ matrix M , we call an $x \in \mathbb{R}^n$ an M -weighted least-squares solution if, for any $y \in \mathbb{R}^n$,

$$(Ax-b)^T M(Ax-b) \leq (Ay-b)^T M(Ay-b). \quad (7)$$

We define (compare with (5))

$$\mathbf{S} = U_1 U_2 \cdots U_{p-1} U_p U_p U_{p-1} \cdots U_2 U_1. \quad (8)$$

Theorem 2—Let $\{\beta_k\}_{k \in \mathbb{N}}$ be a sequence of nonnegative real numbers such that

$\sum_{k=0}^{\infty} \beta_k < \infty$, let $\{v^k\}$ be a bounded sequence of vectors in $\mathbb{R}^n$ and $x^0 \in \mathbb{R}^n$. Then the sequence defined by

$$x^{k+1} = \mathbf{S}(x^k + \beta_k v^k), \quad \text{for all } k \in \mathbb{N}, \quad (9)$$

converges to a fixed point of $\mathbf{S}$. Furthermore, the fixed points of $\mathbf{S}$ are the $\bar{M}_{SB}$ -weighted least-squares solutions of (1), where $\bar{M}_{SB}$ is a symmetric positive definite $m \times m$ matrix. (For an exact definition of $\bar{M}_{SB}$, see [17].)

The following two Lemmas are used in the proofs of Theorems 1 and 2.

Lemma 3—The operators U_t (for $1 \leq t \leq p$), $\mathbf{U}$ and $\mathbf{S}$ are nonexpansive, provided that (3) holds.

Proof: Clearly, it is sufficient to show that the operators U_t , $1 \leq t \leq p$, are nonexpansive. For any such t , $Q_t = I - \lambda_t A_t^T M_t A_t$ is a symmetric matrix. From this and (4) it follows that, for $x, y \in \mathbb{R}^n$,

$$\begin{aligned} \|U_t x - U_t y\| &= \|Q_t(x-y)\| \\ &\leq \|Q_t\| \|x-y\| = \rho(Q_t) \|x-y\|. \end{aligned} \quad (10)$$

This implies that our proof is complete, provided that we can show that $-1 < \mu(Q_t) < 1$, for all eigenvalues $\mu(Q_t)$ of Q_t .

Let $\mu(Q_t)$ be such an eigenvalue. Then $1 - \mu(Q_t)$ is an eigenvalue of the nonnegative-definite matrix $\lambda_t A_t^T M_t A_t$. From this and (3) it follows that

$$0 \leq 1 - \mu(Q_t) \leq \lambda_t \rho(A_t^T M_t A_t) \leq 2 - \varepsilon, \quad (11)$$

which yields immediately that $-1 < \mu(Q_t) < 1$.

Lemma 4—The operators U_t (for $1 \leq t \leq p$) are attracting with respect to $\text{Fix}(U_t)$, provided that (3) holds.

Proof: For any t , $1 \leq t \leq p$, let $x \in \mathbb{R}^n \setminus \text{Fix}(U_t)$ and $y \in \text{Fix}(U_t)$. We first note that

$$s = A_t(x - y), \quad (12)$$

is not the zero vector; for otherwise we would have $A_tx = A_ty$, which together with (4) would imply that $x \in \text{Fix}(U_t)$.

We need to show that

$$\|U_tx - y\| < \|x - y\|. \quad (13)$$

Since $y = U_ty$, it is sufficient to show that

$$\|U_tx - U_ty\|^2 < \|x - y\|^2. \quad (14)$$

Using (4) and (12) we have that

$$\|U_tx - U_ty\|^2 = \|x - y\|^2 - 2\lambda_t \langle s, M_t s \rangle + \lambda_t^2 \langle A_t A_t^T M_t s, M_t s \rangle. \quad (15)$$

Since M_t is positive definite it can be written as $W^T W$ for some symmetric positive definite matrix W , and we get

$$\begin{aligned} \langle A_t A_t^T M_t s, M_t s \rangle &= \langle W A_t A_t^T W^T W s, W s \rangle \\ &\leq \rho(W A_t A_t^T W^T) \langle W s, W s \rangle \\ &= \rho(A_t^T M_t A_t) \langle M_t s, s \rangle. \end{aligned} \quad (16)$$

The inequality is true since $W A_t A_t^T W^T$ is a symmetric matrix. It therefore follows that

$$\begin{aligned} \|U_tx - U_ty\|^2 &\leq \|x - y\|^2 - 2\lambda_t \langle s, M_t s \rangle + \lambda_t^2 \rho(A_t^T M_t A_t) \langle M_t s, s \rangle \\ &= \|x - y\|^2 - \lambda_t (2 - \lambda_t \rho(A_t^T M_t A_t)) \langle M_t s, s \rangle \\ &< \|x - y\|^2. \end{aligned} \quad (17)$$

The last inequality holds because, in view of (3), $\lambda_t (2 - \lambda_t \rho(A_t^T M_t A_t)) \langle M_t s, s \rangle$ is positive.

Proof. (Theorem 1): We first prove that when $\beta_k = 0$, for all $k \in \mathbb{N}$, then the sequence generated by (6) converges to a fixed point of the operator $\mathbf{U}$. By Theorem 5 and Proposition 6 of [17], (3) guarantees the existence of a limit point of the sequence. Since $\mathbf{U}$ is nonexpansive (Lemma 3), the limit point is a fixed point of $\mathbf{U}$.

To show this result in the general case when $\beta_k \neq 0$, for all $k \in \mathbb{N}$, we consider any sequences of $\{\beta_k\}_{k \in \mathbb{N}}$ and $\{v^k\}_{k \in \mathbb{N}}$ that satisfy the conditions stated in Theorem 1. Using Lemma 3, we have that, for each $k \in \mathbb{N}$,

$$\begin{aligned} \|x^{k+1} - \mathbf{U}x^k\| &= \|\mathbf{U}(x^k + \beta_k v^k) - \mathbf{U}x^k\| \\ &\leq \|(x^k + \beta_k v^k) - x^k\| = \beta_k \|v^k\|. \end{aligned} \quad (18)$$

Consequently, denoting by $\mathcal{M}$ a finite upper bound for the bounded sequence $\{\|v_k\|\}_{k \in \mathbb{N}}$, we have the summability result

$$\sum_{k=0}^{\infty} \|x^{k+1} - \mathbf{U}x^k\| \leq \mathcal{M} \sum_{k=0}^{\infty} \beta_k < \infty. \quad (19)$$

According to [4, Theorem 4.1] combined with Lemma 3 and the first part of this proof, the sequence $\{x^k\}_{k \in \mathbb{N}}$ converges to a fixed point of the operator $\mathbf{U}$.

To complete the proof of Theorem 1, we need to show that in the consistent case the set J of solutions of (1) equals $\text{Fix}(\mathbf{U})$. If $x \in J$, then, for $t = 1, 2, \dots, p$, $A_t x = b^t$ and hence by (4) $U_t x = x$. By (5) this implies that $x \in \text{Fix}(\mathbf{U})$. Now consider $x \in \text{Fix}(\mathbf{U})$. Combining Lemma 4 with [1, Proposition 2.10(i)], we get that $x \in \text{Fix}(U_t)$, for $t = 1, 2, \dots, p$. Using (4) and the positivity of λ_t in (3) it follows that, for $t = 1, 2, \dots, p$,

$$A_t^T M_t (b^t - A_t x) = 0. \quad (20)$$

Since (1) is assumed to be consistent, there exist a y such that $b^t = A_t y$. For this y (20) yields

$$\begin{aligned} 0 &= A_t^T M_t (A_t y - A_t x), \\ &= A_t^T M_t A_t (y - x), \\ &= (M_t^{1/2} A_t)^T (M_t^{1/2} A_t) (y - x). \end{aligned} \quad (21)$$

It follows that $\|(M_t^{1/2} A_t)(y - x)\| = 0$ and therefore $A_t x = b^t$. This being true for $t = 1, 2, \dots, p$, implies that $x \in J$.

Proof. (Theorem 2): The proof of the first part of Theorem 2, that the sequence (9) converges to a fixed point of $\mathbf{S}$, follows almost verbatim from the equivalent part of the proof of Theorem 1 (here we use Propositions 10 and 11 of [17] instead of Theorem 5 and Proposition 6 from that paper), since $\mathbf{S}$ (like $\mathbf{U}$) is nonexpansive (Lemma 3).

By [17, Proposition 10] the operator $\mathbf{S}$ can be written as

$$\mathbf{S}x = x + A^T \overline{M}_{SB} (b - Ax), \quad (22)$$

where $\overline{M}_{SB}$ is a symmetric positive definite $m \times m$ matrix. To see that the fixed points of $\mathbf{S}$ are the $\overline{M}_{SB}$ -weighted least-squares solutions of (1), we observe that z is an $\overline{M}_{SB}$ -weighted least squares solution of (1) if, and only if,

$$A^T \overline{M}_{SB} (b - Az) = 0, \quad (23)$$

which combined with (22), yields the desired result.

3. The superiorization approach

Theorems 1 and 2 guarantee convergence even when the iterates are affected by perturbations that are bounded and summable. We can make use of this resiliency to steer the iterates towards a minimizer of a given convex function ϕ . That is, given a convex function $\phi: \mathbb{R}^n \rightarrow \mathbb{R}$ and a

system of equations (consistent or not, as in (1)), our proposed algorithm aims at an $x \in \mathbb{R}^n$ that approximates the minimizer of φ . The following is a heuristic approach that does not guarantee convergence to a minimizer of φ . However, it proceeds (and we give examples in the next section) so that the value of the given function tends to be reduced and yet convergence to a solution of (1) is not compromised. This allows us to do *superiorization*, which means the production of a superior solution (just as optimization produces an optimal solution) subject to the given constraints (the term superiorization was first introduced in [12], see also [3, 5, 22, 32]). Superiorization provides us with a new algorithm, based on an original one, by steering the iterates (using the perturbations) towards a solution that is superior, according to some criterion, to the one to which we would get without perturbations.

We now give the details of the superiorization algorithm. For any $k \in \mathbb{N}$, let $s^k \in \varphi(x^k)$ be a subgradient of the convex function φ at x^k and let

$$v^k = \begin{cases} -\frac{s^k}{\|s^k\|}, & \text{if } s^k \neq 0, \\ 0, & \text{if } s^k = 0. \end{cases} \quad (24)$$

The sequence $\{v^k\}_{k \in \mathbb{N}}$ is clearly bounded. Hence, by Theorems 1 and 2, for any summable sequence of positive real numbers $\{\beta_k\}_{k \in \mathbb{N}}$, the $\{x^k\}_{k \in \mathbb{N}}$ generated as in (6) or (9) converges to a “solution” of (1). We produce the real numbers $\{\beta_k\}_{k \in \mathbb{N}}$ in the following way. We use the help of a proximity function, $Pr(x)$, that is defined, for $x \in \mathbb{R}^n$

$$Pr(x) = \sqrt{\sum_{i=1}^m (b_i - \langle a^i, x \rangle)^2}, \quad (25)$$

where a^i and b_i correspond to the i th row of A and b of (1), respectively. The size of $Pr(x)$ indicates how badly x violates the given system of equations. We note that this proximity function Pr is different from the proximity function Res used in previous work, for example in [12]. The reason for our new choice is that Res is unstable, in the sense that small changes in the data result in large changes in the value of Res , but Pr is stable. For example, for two simulated data sets with realistic noise we found the values of Res to be 7.12 and 11.13, while the values of Pr were 5.39 and 5.38, respectively. The difference between the two proximity functions is that for Res each term in the sum in (25) is divided by $\|a^i\|^2$. In image reconstruction from projections we come across the situation where a line integral b_i is estimated from measured data for a line that hardly intersects the reconstruction region, resulting in an extremely small value of $\|a^i\|^2$. In such a case a slight variation in the measured data will result in a large change in Res , but not in Pr .

To approximate the solution of the optimization problem (for φ) it is desirable to find an x for which $Pr(x)$ is small, and among all x with similar (or smaller) value of $Pr(x)$, the $\varphi(x)$ should be small relative to the others. Guided by this principle, we initialize β to be an arbitrary positive number, which we denote by β_{-1} . (In our implementation, we have always used $\beta_{-1} = 1$.) In the iterative step from x^k to x^{k+1} , we update the value of β , which is (in the notation of (6) and (9)) β_{k-1} at the beginning of the iterative step and β_k at its end. This is done according to the following pseudocode (in which v^k is defined by (24)). At the time of exiting of the pseudocode, which is called repeatedly, we have the final value of x^{k+1} . We terminate the whole iterative process when we find an x^k such that $Pr(x^k) < \varepsilon$, where ε is a user-specified small positive number. To avoid infinite loops in which $Pr(x^{k+1}) < Pr(x^k)$ always fails but at the same time φ keeps being reduced, we also incorporated an additional stopping criterion based on β getting

smaller than some user-specified value; however, in all the experiments reported in Section 4 stopping was due to the criterion $Pr(x^k) < \varepsilon$.

```

1: logic = true
2: while (logic)
3:    $z = x^k + \beta v^k$ 
4:   if ( $\varphi(z) \leq \varphi(x^k)$ )
5:   then
6:      $x^{k+1} = Oz$ 
7:   if ( $Pr(x^{k+1}) < Pr(x^k)$ )
8:   then logic = false
9:   else  $\beta = \beta/2$ 
10:  else  $\beta = \beta/2$ 

```

The operator **O** in Step 6 of the pseudocode is the **U** of (5) for the algorithm of Theorem 1 and is the **S** of (8) for the algorithm of Theorem 2. The complete superiorization algorithm consists of (6) (respectively, (9)) with v^k defined by (24) and β_k determined by the pseudocode that makes use of (25). The algorithm performs a steering process towards a small value of φ (in Step 4 of the pseudocode), while attempting to maintain convergence, as guaranteed by the theorems for a proper choice of the sequence $\{\beta_k\}_{k \in \mathbb{N}}$ (see Step 7 of the pseudocode). In the case when the operator **O** is the **S** of (8), a reasonable alternative would be to use instead of Pr as defined by (25) its $\bar{M}_{SB}$ -weighted version that occurs in Theorem 2. However this was not done for the work reported in this paper since the use of Pr as a proximity function is standard in the literature, while its $\bar{M}_{SB}$ -weighted version has not been used for that purpose.

In the next section we give several examples of the superiorization approach, illustrating its use for image reconstruction. Our attitude to assigning levels of success to what we are doing is the following. First of all, it is in principle possible that the superiorization algorithm described in this section does not terminate; for example, this has to happen if there is no $x \in \mathbb{R}^n$ such that $Pr(x) < \varepsilon$. In such a case we consider that the algorithm has failed. If the algorithm does not fail, then by necessity it produces an $x \in \mathbb{R}^n$ such that $Pr(x) < \varepsilon$. Then there is the question of how “good” that x is. The pure optimization point of view would say that this is measured by $\varphi(x)$, which should be as small as possible. However, our practical point of view is different: there is no reason to insist on obtaining an x whose φ value is less than that of the phantom that we are trying to reconstruct. (This can be checked only in simulation experiments in which we have a knowledge of the phantom, but that is what we are doing in this paper.) We consider every $x \in \mathbb{R}^n$ with φ value less than that of the phantom and with Pr value less than ε to be an *acceptable* solution. For all experiments reported in the next section, all variants of the superiorization algorithm produced acceptable reconstructions. Since our primary concern in this paper is computer time, we use that as the criterion by which we distinguish between the level of success of the various variants of the superiorization algorithm.

4. Results

To illustrate our newly proposed accelerated methods, we have selected the tomographic reconstruction problem in which the object to be reconstructed is not uniquely determined from the available data. We use the help of a particular convex functional φ that in some sense indicates the “undesirability” of the image by assigning to it a real number. We restrict our attention to evaluating the improvements due to the ideas introduced in this paper (especially

acceleration) as compared to the unaccelerated version of the superiorized block-iterative method as reported in [12].

Figure 1(a) shows a mathematically defined phantom that is a 365×365 digitized image (thus $n = 133,225$ in the notation of our paper), representing a cross-section of a human head (we use this phantom as the basis for all the experiments reported in this section). The center of the phantom is the center of the coordinate system. The components of x represent the average X-ray attenuation coefficients within the 133,225 pixels. Each pixel is of size 0.05×0.05 , where the assumed unit of length is 1 cm. The values of these components range from 0 to 0.62039 and for display purposes any value below 0.204 is shown as black (gray value 0) and any value above 0.21675 is shown as white (gray value 255), with a linear mapping of the x -component values into gray values in between (this is true for all the figures presented here). We now describe the convex function ϕ that we have chosen for the experiments throughout the paper.

Many researchers in image processing have been advocating the use of total variation, e.g., [3, 11, 14, 29, 32, 33]. For a $G \times H$ image q whose pixel values are denoted by $q_{g,h}$ ($1 \leq g \leq G, 1 \leq h \leq H$), the *total variation* (TV) of q is

$$TV(q) = \sum_{g=1}^{G-1} \sum_{h=1}^{H-1} \sqrt{(q_{g+1,h} - q_{g,h})^2 + (q_{g,h+1} - q_{g,h})^2}. \quad (26)$$

By mapping q into a $(G \times H = n)$ -dimensional point x (by stacking into one column all the columns of q), (26) gives rise to a functional ϕ that can be used in our superiorization algorithm described in Section 3. TV is chosen here to illustrate the efficacy of our approach to accelerating block-iterative projection methods; however the approach is general and can be applied to other functionals that have been proposed for finding a regularized solution to a system of linear equations.

4.1. Results with underdetermined consistent data

For the first experiment, data were collected by calculating line integrals through the digitized image for 82 sets of equally spaced parallel lines. In our implementation we use in each block only the lines that intersect the reconstruction region; lines with no intersections are ignored and so the number of lines associated with a block (the ℓ_t as defined above (3)) varies from block to block. Each line integral gives rise to a linear equation in the components of x ; the set of all x consistent with such a line integral is a hyperplane in $\mathbb{R}^n$. The phantom itself lies in the intersection of all the hyperplanes that are associated with these lines. In this experiment, we used measurements for 42,558 lines (linear equations), making our problem very much underdetermined (with 133,225 unknowns); the intersection of all the hyperplanes is an at least 90,667-dimensional subspace of $\mathbb{R}^{133,225}$. It is typical in practice (and this is what we will be doing here) that each block A_t consists of all the indices i associated with the measurements taken in a particular direction. Thus, denoting by I_t the set of indices belonging to the t th block, we obtain that $\{1, \dots, m\} = \bigcup_{t=1}^p I_t$, with $p = 82$. Note that the number of elements in I_t is ℓ_t .

We applied our block-iterative algorithm based on (6) to the specified data set for the choice of ϕ given below (26). In all experiments we selected a single relaxation parameter λ and assigned $\lambda_t = \lambda$, for $t = 1, 2, \dots, p$. The choice of λ has an affect on the stopping point for any chosen ε . Generally, for the range of λ s that produce an acceptable solution, where “acceptable” is as defined in the last paragraph of Section 3, a larger λ results in getting to an acceptable solution faster (this is no longer so as λ gets near its upper limit (3) for convergence).

This is illustrated in Table 1. For the experiments for that table and for all other experiments reported in this subsection, we chose the stopping criterion $Pr(x^k) < 0.01$. Also, all our iterative algorithms start with x^0 being the *zero point* (i.e., all its components are 0). This is a reasonable choice in our context: we do not wish to start with an image for which the Pr value is small but the ϕ value is high: this would likely result in terminating the iterative process (termination depends only on the Pr) before Step 3 of the pseudocode has been executed a sufficient number of times to make the final ϕ as small as desired. Note that $Pr(x^0) = 466.27$, demonstrating that the value 0.01 for the stopping criterion is small in our context. Based on the numbers in the table, for the remaining experiments in this subsection we set $\lambda = 1.5$; this value gets us to an acceptable reconstruction (the TV of the phantom is 707.06) using the least amount of computer time. A complete description of the algorithm needs the specification of the M_t in (4). Our theory allows these to be arbitrary positive definite symmetric matrices. However, in order to demonstrate the acceleration over the method of [12], we decided to use the same M_t that were used in that paper. Below we derive the exact relationship between the two methods, this leads to a specification provided in (31) of the M_t used in our reported experiments.

Figure 1(b) shows the reconstruction when TV-superiorizing perturbations are present and Figure 1(c) shows the reconstruction when no perturbations are introduced. Note that the algorithm without perturbations (i.e., $\beta_k = 0$, for $k \in \mathbb{N}$) converges to the feasible point with minimal norm [21, Section 11.2]. Clearly, the reconstruction in Figure 1(b) is visually superior to the reconstruction in Figure 1(c).

In the first two rows of Table 2, we report on the values of the norm and TV for the outputs of the two algorithms, as well as the time in minutes needed to obtain the reconstruction. As can be seen, the algorithms tend to reduce the functions that they are supposed to be minimizing; the superiority of the reconstruction in Figure 1(b) to that in Figure 1(c) is due to TV minimization being a more appropriate aim than norm minimization under the current circumstances.

In order to appreciate the speed-up of our new method, we ran another TV-superiorizing block-iterative algorithm (from [12]) on this data set. From [12] we have the following. Let $\mathbf{Q}: \mathbb{R}^n \rightarrow \mathbb{R}^n$ be a composite operator such that

$$\mathbf{Q} = Q_p \cdots Q_2 Q_1, \quad (27)$$

where, for $x \in \mathbb{R}^n$ and $1 \leq t \leq p$,

$$Q_t x = \frac{1}{R} \sum_{i \in I_t} \left(x + \frac{b_i - \langle a^i, x \rangle}{\|a^i\|^2} a^i \right) + \frac{R - \ell_t}{R} x \quad (28)$$

with

$$R = \max \{ \ell_t | 1 \leq t \leq p \}. \quad (29)$$

Rearranging (28) we obtain

$$Q_t x = x + \frac{1}{R} \sum_{i \in I_t} \frac{b_i - \langle a^i, x \rangle}{\|a^i\|^2} a^i. \quad (30)$$

Thus, the only distinction between the operators in (30) and in (4) with the matrices chosen as

$$M_i = \text{diag} \left(\frac{1}{\|a^i\|^2} \right)_{i \in I_t} \quad (31)$$

comes from the operator in (30) being dependent on R , while the operator in (4) allows flexibility in the selection of λ_t , with a wider range of choice that depends only on (3). (Choosing M_i as in (31) is legitimate as far as the validity of Theorems 1 and 2 is concerned, since the only restriction on the matrices M_i used in Algorithm 1 is that they be positive definite.) To be precise, in our example above, $\frac{1}{R} = \frac{1}{517}$, while the relaxation parameter for the newly proposed operator (4) was chosen in our implementation to be $\lambda_t = \lambda = 1.5$, making the corrections in the subiterative steps 775 times greater than what was used in the unaccelerated method.

Again, we started the process with the zero point and set the stopping criterion to $Pr(x^k) < 0.01$. Figure 1(d) shows the resulting image and the third row of Table 2 gives the corresponding numerical values. Both reconstructions with perturbations seem to be superior to the reconstruction without perturbations (Figures 1(b) and 1(d), respectively, versus 1(c)). The TV (reported in the TV column of Table 2) of the reconstructions produced by the two superiorization algorithms are nearer to (and are in fact less than) the TV of the phantom in Figure 1(a) (which is 707.06), whereas the TV of the image in Figure 1(c) that is obtained without perturbations is 2425.10.

Figure 2 shows a graph of $Pr(x^k)$ over computer time for the three algorithms reported in Table 2. As indicated by the plot (and also by the time column in the table), our accelerated block-iterative TV-superiorizing method (from Theorem 1) required 3.30 minutes on an Intel Xeon quad core 2.66-GHz processor with 8-G RAM workstation (using only one processor) to reach a Pr below 0.01, while the method without acceleration took more than 355 times longer to obtain a similar Pr . Even though the accelerated method without perturbations (i.e., the norm minimizing method from Theorem 1) is faster than the TV-superiorizing method without acceleration from [12], the perturbations in the latter steer it towards the correct result (i.e., in the general direction of the phantom). All the computational work reported here was done using SNARK09 [13]; the phantom, the data, the reconstructions and displays were all generated within this same framework. In particular, this implies that differences in the reported reconstruction times are not due to the different algorithms being implemented in different environments.

4.2. Results with overdetermined realistic data

So far we have reported on experiments from data that were consistent, i.e., the line integrals were calculated based on the digitized image of the phantom. However, data collected in real-life image reconstruction from projections are not at all likely to be consistent; a fundamental reason is that real-life objects are not digitized and hence variations within pixels will exist. In addition, various deviations from the idealized mathematical point of view contribute to the data being inconsistent with the object. In CT the measurements are stochastic in nature (the line integrals are estimated by the use of a, by necessity, finite number of X-ray photons, resulting in statistical noise in these estimates). Another problem is that the detectors in real instruments will have a width, so even if we ignore the fact that they may not be perfect, they could not be used for measuring line integrals exactly. Furthermore, X-ray beams used in CT are polychromatic and hence their energy spectrum hardens depending on the material they have traveled through [20]. This makes the assignment of a linear attenuation coefficient associated with a point in the body more difficult and yet another estimation takes place to

resolve that. There are other (here not listed) additional sources of discrepancy between the assumed mathematical model and physically collected data, such as the presence of scattered X-ray photons corrupting the readings by the individual detectors; for a further discussion and in-depth explanation the reader may refer to [21, Chapter 3].

In the next experiment we examined our method from Theorem 1 when the data are realistic from these points of view. The line integrals were based on a geometrical description of the head phantom (the one mentioned in the experiment in Subsection 4.1, and seen in Figure 1 (a)) rather than on its digitization. The width of the detector was simulated by adding, to each line, 10 additional lines (five on each side) with a spacing of $d/11$ between them, where d ($= 0.05$ in our case) is the distance between parallel lines along which data are assumed to have been collected in the mathematical formulation. Statistical noise and scatter were introduced at the levels found in real CT scanners, and furthermore, a beam hardening correction was applied. (The software SNARK09 [13], which we used for all our experiments as mentioned previously, allows us to simulate this kind of data collection, the details of which are explained in [21, Section 4.5].) Data were generated for 360 sets of equally spaced parallel lines (each line gives rise to a hyperplane in $\mathbb{R}^{133,225}$), with a $\frac{1}{2}$ degree increment between consecutive directions. Blocks were again formed by using all measurements taken from a particular direction (making $p = 360$).

We again ran our accelerated block-iterative algorithm from Theorem 1, and compared the results with and without TV-superiorizing perturbations on such realistic data. The starting point was the zero point for all runs in this experiment and the stopping criterion was $Pr(x^k) < 5.4$, which is reasonable since for this noisy data set the value of Pr for the phantom is 5.39. This means in particular that, due to the discrepancy between the actual mode of data collection and the mathematical model that is assumed for it, the data are inconsistent with the phantom, which no longer lies in the intersection of all the hyperplanes. Figures 3(a) and 3(b) show the resulting reconstructions and Table 3 provides the corresponding numerical values. The quality of the reconstruction in Figure 3(a) is visually better than that of the reconstruction in Figure 3(b) by the algorithm without perturbations. The superiority of the reconstruction with perturbations is also expressed numerically by the TV value, as indicated in the first two rows of Table 3. We also ran the TV-superiorizing algorithm without acceleration from [12] on these inconsistent data. The resulting image is in Figure 3(c). As indicated by the third row of Table 3, the time that algorithm took to reconstruct that image was more than 16 times the time needed by the accelerated TV-superiorizing algorithm of Figure 3(a). As a further illustration, we looked at the image at the iterate reached by the algorithm from [12] after 72.7 seconds had passed, in order to demonstrate how much the acceleration buys us. The result is seen in Figure 3(d), which took approximately 78.4 seconds to obtain. As can be seen, the algorithm does not provide an acceptable reconstruction at that time (that image has a Pr value that is 73.58).

4.3. Statistical hypothesis testing evaluations on realistic data

The results of the experiments we reported so far are anecdotal, they may not be representative of the general situation. Also, in spite of the fact that the TV for Figure 3(c) is much less than that for Figure 3(a) (see Table 3), the visual evaluation of the images indicate that the former is a worse reconstruction than the latter (in particular, Figure 3(c) shows hardly any sign of the big tumor in the left half of the phantom shown in Figure 1(a), whereas it is clearly visible in Figure 3(a).) In answer to such objections, we discuss in this subsection experiments that we carried out that used statistical hypothesis testing to evaluate the relative efficacy of various reconstruction algorithms for a medically relevant task, namely the detection of low-contrast small tumors in the brain from X-ray CT projection data. In accordance with the general description of such statistical hypothesis testing evaluations in Section 5.2 of [21], the specific experiments carried out for this subsection included the following four steps:

- i. Generation of random samples from a statistically described ensemble of phantoms and simulation of realistic data collection for each sample. Our random samples were similar to the phantom in Figure 1(a), but they included a large number of pairs of potential tumor sites, with the locations of the sites in each pair symmetric around the vertical axis of the head phantom. Only one site of the pair has a tumor in it, these are randomly selected for each sample. Thirty such phantoms were generated; Figure 4 shows two of them. The small tumors have circular shapes and are of radius 0.1 cm. The data collection was simulated for 360 views with the same details as for the experiment in Section 4.2.
- ii. Reconstruction from each of the generated projection data sets by each of the algorithms to be compared.
- iii. Assignment of *figures of merit* (FOMs) to each reconstruction. FOMs should measure the goodness of the reconstruction for our task, the two that we used for our experiments are *image wise region of interest* (IROI) and *hit ratio* (HITR). It was previously found that IROI has the tendency to agree with the performance human observers in identifying the tumors in a reconstruction [31]. HITR considers the pairs of potential tumors site. Such a pair is a hit if the structure in the pair with the higher average density in the phantom is also the structure in the pair with the higher average density in the reconstruction. The HITR for a reconstruction is the number of hits divided by the total number of pairs. The intuition behind HITR is that a radiologist examining a reconstruction would look at the corresponding symmetric area in order to determine if something is a tumor, as structures in the brain tend to be symmetric.
- iv. Calculation of statistical significance. For any pair of algorithms and for any FOM, the null hypothesis is that the expected average values (over the 30 reconstructions) of the FOMs for the two algorithms are the same, with the alternative hypothesis that the expected average value is in fact higher for the algorithm for which the experimentally observed average is higher. Based on the 30 pairs of FOMs one can calculate the P-value, which is the probability of observing a difference between the performances (according to the FOM) of the two algorithms that is as high or higher than the observed difference if the null hypothesis that the two algorithms are equally efficacious were true. The smallness of the P-value measures the significance by which we can reject the null hypothesis in favor of the alternative.

We first report on the outcome of such a comparative evaluation procedure when the algorithms compared are two block-iterative TV-superiorizing algorithms, the one with acceleration based on Theorem 1 and the one without acceleration from [12]. The results were conclusive: for both FOMs, the null hypothesis of equal efficacy can be rejected with extreme statistical significance in favor of the alternative that the method with acceleration performs better for the task at hand (the P-values are 2.43×10^{-8} for IROI and 6.7×10^{-8} for HITR). However, small differences can be highly statistically significant, while in practice they hardly matter; see [30]. This is not the case here, the average values over the 30 reconstructions of the accelerated and the unaccelerated algorithms are, respectively, 0.16 and 0.07 for IROI and 0.99 and 0.94 for HITR. As an illustration we include in Figure 5 the reconstructions of the sample phantom from Figure 4(a) by the two algorithms.

Finally, we report on statistical experiments that were carried out because of our desire to investigate the efficacy of algorithms based on Theorem 2. Note that the only difference between such an algorithm and a corresponding one based on Theorem 1 is that the operator $\mathbf{O}$ in Step 6 of the pseudocode in Section 3 is the $\mathbf{S}$ of (8) instead of the $\mathbf{U}$ of (5). Since the computational cost of applying $\mathbf{S}$ is approximately twice of that of applying $\mathbf{U}$, in this set of studies we used $\mathbf{U}^2$ (i.e., two consecutive applications of $\mathbf{U}$) as the operator $\mathbf{O}$ in the nonsymmetric case and selected $\lambda_t = \lambda = 0.010$ for both the symmetric and the nonsymmetric

algorithms. (This choice was made based on preliminary experiments of the same kind whose results for consistent data are reported in Table 1.) The statistical study did not reveal a significant difference between the two approaches, the P-values for both FOMs were near 0.5. As an illustration we give in Figure 6 reconstructions produced by TV-superiorizing algorithms with the symmetric and with the nonsymmetric operator, respectively.

5. Conclusions

We discussed block-iterative projection methods for solving linear systems of equations. We showed that these perturbation-resilient methods can be made significantly faster than previously published algorithms of this type while retaining convergence to a solution of the system (in both the consistent and inconsistent cases). We made use of the resiliency to perturbations of the methods and used the superiorization approach to approximate the minimizer of a convex functional. We demonstrated the usefulness of this by examples from tomography and showed that on top of the gain in speed, improvements in the efficacy of the algorithms can also be achieved.

Acknowledgments

The authors thank Tommy Elfving for his help with Lemma 4. This work was supported by the National Science Foundation Award Number DMS-1114901 and by Award Number R01HL070472 from the National Heart, Lung and Blood Institute. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Heart, Lung and Blood Institute or the National Institutes of Health.

References

1. Bauschke HH, Borwein JM. On projection algorithms for solving convex feasibility problems. *SIAM Rev.* 1996; 38:367–426.
2. Byrne C. Block-iterative interior point optimization methods for image reconstruction from limited data. *Inv Prob.* 2000; 16:1405–19.
3. Butnariu D, Davidi R, Herman GT, Kazantsev IG. Stable convergence behavior under summable perturbations of a class of projection methods for convex feasibility and optimization problems. *IEEE J Sel Top Sign Process.* 2007; 1:540–7.
4. Butnariu, D.; Reich, S.; Zaslavski, AJ. Convergence to fixed points of inexact orbits of Bregman-monotone and nonexpansive operators in Banach spaces. In: Nathansky, HF.; de Buen, BG.; Goebel, K.; Kirk, WA.; Sims, B., editors. *Fixed Point Theory and Applications*. Yokohama: Yokohama Publishers; 2006. p. 11–32.
5. Censor Y, Davidi R, Herman GT. Perturbation resilience and superiorization of iterative algorithms. *Inv Prob.* 2010; 26:065008.
6. Censor, Y.; Elfving, T.; Herman, GT. Averaging strings of sequential iterations for convex feasibility problems. In: Butnariu, D.; Censor, Y.; Reich, S., editors. *Inherently Parallel Algorithms in Feasibility and Optimization and Their Applications*. Elsevier Science Publishers; 2001. p. 101–14.
7. Censor Y, Elfving T, Herman GT, Nikazad T. On diagonally-relaxed orthogonal projection methods. *SIAM J Sci Comput.* 2007/8; 30:473–504.
8. Censor Y, Tom E. Convergence of string-averaging projection schemes for inconsistent convex feasibility problems. *Optim Methods Softw.* 2003; 18:543–54.
9. Censor Y, Gordon D, Gordon R. BICAV: An inherently parallel algorithm for sparse systems with pixel-dependent weighting. *IEEE Trans Med Imag.* 2001; 20:1050–60.
10. Censor, Y.; Zenios, SA. *Parallel Optimization: Theory, Algorithms and Applications*. Oxford University Press; 1997.
11. Combettes PL, Pesquet JC. Image restoration subject to a total variation constraint. *IEEE Trans Image Process.* 2004; 13:1213–22. [PubMed: 15449583]

12. Davidi R, Herman GT, Censor Y. Perturbation-resilient block-iterative projection methods with application to image reconstruction from projections. *Int Trans Oper Res.* 2009; 16:505–24. [PubMed: 23271857]
13. Davidi, R.; Herman, GT.; Klukowska, J. SNARK09: A programming system for the reconstruction of 2D images from 1D projections. 2009. (<http://www.dig.cs.gc.cuny.edu/software/snark09/>)
14. Easley GR, Labate D, Colonna F. Shearlet-based total variation diffusion for denoising. *IEEE Trans Image Process.* 2009; 18:260–8. [PubMed: 19095539]
15. Eggermont PPB, Herman GT, Lent A. Iterative algorithms for large partitioned linear systems, with applications to image reconstruction. *Linear Algebra Appl.* 1981; 40:37–67.
16. Elfving T. Block-iterative methods for consistent and inconsistent linear equations. *Numer Math.* 1980; 35:1–12.
17. Elfving T, Nikazad T. Properties of a class of block-iterative methods. *Inv Prob.* 2009; 25:115011.
18. Elsner L, Koltracht I, Neumann M. Convergence of sequential and asynchronous nonlinear paracontractions. *Numer Math.* 1992; 62:305–19.
19. Frommer A, Szyld DB. On asynchronous iterations. *J Comp Appl Math.* 2000; 123:201–16.
20. Herman GT. Demonstration of beam hardening correction in computerized reconstruction of the head. *J Comput Assist Tomo.* 1979; 3:373–78.
21. Herman, GT. *Fundamentals of Computerized Tomography: Image Reconstruction from Projections.* 2. Springer; 2009.
22. Herman GT, Davidi R. Image reconstruction from a small number of projections. *Inv Prob.* 2008; 24:045011.
23. Hudson HM, Larkin RS. Accelerated image reconstruction using ordered subsets projection data. *IEEE Trans Med Imag.* 1994; 13:601–9.
24. Jiang M, Wang G. Convergence studies on iterative algorithms for image reconstruction. *IEEE Trans Med Imag.* 2003; 22:569–79.
25. Kamath C, Sameh A. A projection method for solving nonsymmetric linear systems on multiprocessors. *Parallel Comp.* 1989; 9:291–312.
26. Kaczmarz S. Angenäherte Auflösung von Systemen linearer Gleichungen. *Bulletin de l'Académie Polonaise des Sciences et Lettres.* 1937; A35:355–7.
27. Kiwiel KC. Convergence of approximate and incremental subgradient methods for convex optimization. *SIAM J Optim.* 2004; 14:807–40.
28. Luo Z-Q, Tseng P. Error bounds and convergence analysis of feasible descent methods: a general approach. *Ann Oper Res.* 1993; 46:157–78.
29. Malgouyres F. Minimizing the total variation under general convex constraints for image restoration. *IEEE Trans Imag Process.* 2002; 11:1450–6.
30. Matej S, Furuie SS, Herman GT. Relevance of statistically significant differences between reconstruction algorithms. *IEEE Trans Imag Process.* 1996; 5:554–6.
31. Narayan TK, Herman GT. Prediction of human observer performance by numerical observers: An experimental study. *J Opt Soc Am A.* 1999; 16:679–93.
32. Penfold SN, Schulte RW, Censor Y, Rosenfeld AB. Total variation superiorization schemes in proton computed tomography image reconstruction. *Med Phys.* 2010; 37:5887–95. [PubMed: 21158301]
33. Rudin LI, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms. *Physica D.* 1992; 60:259–68.
34. Young, DM. *Iterative Solution of Large Linear Systems.* Dover Publications; 2003.

Figure 1.

Reconstructions from underdetermined consistent data for 82 views with the stopping criterion of $Pr(x^k) < 0.01$. (a) Phantom for which data were collected. (b) Block-iterative TV-superiorizing accelerated algorithm (Theorem 1). (c) Norm-minimizing reconstruction (Theorem 1, no perturbations). (d) Block-iterative TV-superiorizing without acceleration from [12].

Figure 2. Log-log plots of $Pr(x^k)$ against computer time of the three algorithms specified in the caption of Table 2 on consistent data.

Figure 3.

Reconstructions of the head from realistic data obtained from 360 views with the stopping criterion of $Pr(x^k) < 5.4$. (a) Block-iterative TV-superiorizing accelerated algorithm of Theorem 1. (b) Block-iterative algorithm of Theorem 1 with no perturbations. (c) Block-iterative TV-superiorizing without acceleration from [12] for $Pr(x^k) < 5.4$. (d) Block-iterative TV-superiorizing without acceleration from [12], stopped after the time it took our accelerated algorithm to obtain (a).

Figure 4. Random samples from the ensemble of phantoms used for statistical hypothesis testing evaluations. The essential difference between the random samples is the choice of the side on which a tumor appears for a pair of symmetric potential tumor sites; for example, for the topmost pair the tumor appears on the right in (a) but on the left in (b).

Figure 5. Reconstructions of the phantom shown in Figure 4(a). (a) Block-iterative TV-superiorizing accelerated algorithm of Theorem 1. (b) Block-iterative TV-superiorizing without acceleration from [12].

Figure 6. Reconstructions from the same realistic projection data set for one of our random phantoms (similar to the ones in Figure 4) by two TV-superiorizing algorithms using (a) the symmetric operator $\mathbf{S}$ and (b) the nonsymmetric operator $\mathbf{U}^2$.

Table 1

Numerical values for the TV-superiorization algorithm (Theorem 1) on consistent data using various relaxation parameters.

λ	norm	TV	time (min)
0.25	69.59	658.55	10.27
0.75	69.59	664.24	4.28
1.25	69.59	671.29	3.87
1.50	69.59	668.64	3.30
1.75	69.59	667.99	3.64

Table 2

Numerical values for the outputs (shown in Figure 1) of the three algorithms on consistent data. Algorithm 1: Accelerated algorithm of Theorem 1 with TV-superiorization. Algorithm 2: Accelerated algorithm of Theorem 1 without perturbations. Algorithm 3: TV-superiorization algorithm from [12].

Method		norm	TV	time (min)
Algorithm 1	Fig. 1(b)	69.59	668.64	3.30
Algorithm 2	Fig. 1(c)	69.32	2425.10	113.64
Algorithm 3	Fig. 1(d)	69.59	658.56	1,172.29

Table 3

Numerical values for the outputs of the algorithms for inconsistent data in Figure 3. Algorithm 1: Accelerated algorithm of Theorem 1 with TV-superiorization. Algorithm 2: Accelerated algorithm of Theorem 1 without perturbations. Algorithm 3: TV-superiorization algorithm from [12].

Method		norm	TV	time (sec)
Algorithm 1	Fig. 3(a)	68.57	690.20	72.7
Algorithm 2	Fig. 3(b)	68.57	1271.39	78.9
Algorithm 3	Fig. 3(c)	68.54	602.88	1168.9

Algorithm 1

(traditional block-iterative algorithm)

Initialization: $x^0 \in \mathbb{R}^n$ is arbitrary.Iterative Step: Given x^k compute

$$\begin{aligned}
 x^{k,0} &= x^k, \\
 x^{k,t} &= x^{k,t-1} + \lambda_t A_t^T M_t (b^t - A_t x^{k,t-1}), \quad t = 1, 2, \dots, p, \\
 x^{k+1} &= x^{k,p},
 \end{aligned}$$

where $\{\lambda_t\}_{t=1}^p$ is a set of relaxation parameters and $\{M_t\}_{t=1}^p$ is a set of given positive definite symmetric matrices.