

Seminorm-Induced Oblique Projections for Sparse Nonlinear Convex Feasibility Problems

Alexander Segal (asegal@math.haifa.ac.il)

and

Yair Censor (yair@math.haifa.ac.il)

Department of Mathematics, University of Haifa
Mt. Carmel, Haifa 31905, Israel

December 4, 2008 and, in revised form, March 10, 2009.

Abstract

Simultaneous subgradient projection algorithms for the convex feasibility problem use subgradient calculations and converge sometimes even in the inconsistent case. We devise an algorithm that uses seminorm-induced oblique projections onto super half-spaces of the convex sets, which is advantageous when the subgradient-Jacobian is a sparse matrix at many iteration points of the algorithm. Using generalized seminorm-induced oblique projections on hyperplanes defined by subgradients at each iterative step, allows component-wise diagonal weighting which has been shown to be useful for early acceleration in the sparse linear case. Convergence for the consistent case with underrelaxation is established.

1 Introduction

The *convex feasibility problem* (CFP) is to find a point x^* in the intersection Q of m closed convex subsets $Q_1, Q_2, \dots, Q_m \subseteq R^n$ of the n -dimensional

Euclidean space. Each Q_i is expressed as

$$Q_i = \{x \in R^n \mid f_i(x) \leq 0\}, \quad (1)$$

where $f_i : R^n \rightarrow R$ is a convex function, so the CFP requires a solution of the system of convex inequalities

$$f_i(x) \leq 0, \quad i = 1, 2, \dots, m. \quad (2)$$

Many iterative projection algorithms for the CFP were developed, see Subsection 1.1 below, but we focus our attention here on simultaneous ones. Such algorithms convexly combine the individual projections, thereby allowing the user to assign weights (of importance) to the individual sets. Our objective in the present paper is to propose and study a simultaneous projection method that enables *component-wise weighting*. This means that weights assigned to the sets are not just set-dependent w_i s but, at the same time, also component-dependent, so that we have weights $\{w_{ij}\}$ for all $i = 1, 2, \dots, m$ and all $j = 1, 2, \dots, n$. To the best of our knowledge, no other projection method for the general (not necessarily linear) CFP exists that allows this flexibility.

The origins of this idea lie in [8] where a simultaneous projection algorithm, called *component averaging* (CAV), for systems of linear equations, that uses component-wise weighting was proposed. Such weighting enables, as shown and demonstrated experimentally on problems of image reconstruction from projections in [8], significant and valuable acceleration of the early algorithmic iterations due to the high sparsity of the system matrix appearing there. In [8] a notion of a *generalized oblique projection* onto a hyperplane was introduced. Such a projection extends the common oblique projection that uses an ellipsoidal norm $\|x\|_G = \langle x, Gx \rangle^{1/2}$, with a diagonal positive definite matrix G , to a generalized oblique projection with respect to a seminorm $\|x\|_G$ in which G may have zeros on its diagonal. In this way diagonal weighting matrices were introduced in the simultaneous algorithmic scheme. A block-iterative version of CAV, named BICAV, was introduced later in [9]. Full mathematical analysis of these methods, as well as their companion algorithms for linear inequalities, was presented by Censor and Elfving [6] and by Jiang and Wang [25].

It is, therefore, natural to ask if and how these ideas can be extended to a CFP with nonlinear convex sets. An attempt to answer this question was made in [7]. However, when applying seminorm-induced oblique projections

in a simultaneous algorithmic scheme for general (not necessarily linear) convex sets, the approach used in [7] mandated a certain relationship between the matrix G and the (nonlinear) convex set Q onto which the seminorm-induced projection is made, namely, that the set will be *directionally affine with respect to G* , see [7, Definitions 2.3 and 2.4]. In spite of the actual generalization obtained in this way, its scope is limited due to this extra condition.

Here we manage to construct a simultaneous projection algorithm with component-wise weighting, for general convex (not necessarily linear) sets without any such scope-reducing conditions on the convex sets. The method proposed here keeps using seminorm-induced oblique projections but instead of performing them directly onto the convex sets (which led to that extra condition in [7]) we do the seminorm-induced oblique projections onto super half-spaces, determined by any subgradient of the function at the current point. This is easier to execute and it still allows the user to employ, and take advantage of, component-wise weighting.

The potential for initial algorithmic acceleration in the sparse Jacobian case inspired us to investigate this problem but we present in this report only theoretical results of the algorithmic development and convergence. The notion of sparseness is very well understood and used for matrices (see, e.g. Wilkinson [31]) and, from there, the road to sparseness of the Jacobian (or generalized Jacobian) matrix as an indicator of sparseness of nonlinear programming problem is short, see, e.g., Betts and Frank [4].

Future studies are aimed at further expanding the family of projection algorithms that allow component-wise weighting with weights $\{w_{ij}\}$ for all i and all j and at studying convergence rates. There are some real-world problems such as intensity-modulated radiation therapy (IMRT) where convex nonlinear constraints appear, see, e.g., Censor et al. [10, 11], which might benefit from the developments presented here but this has not yet been verified.

More general (asynchronous, block-iterative, iteration-dependent weight, adapted long-step relaxations) parallel projection methods exist to solve systems of convex inequalities, see, e.g., Dos Santos [21], Combettes [17], Kiwiel and Lopuch [26]. However, these methods are based on different principles and do not incorporate component-wise weighting neither cater specifically for sparse problems. The possibility of combining such other techniques with the methodology proposed here still needs to be investigated.

1.1 Projection methods: Advantages and earlier work

The reason why the CFP is looked at from the viewpoint of projection methods can be appreciated by the following brief comments, that we made in earlier publications, regarding projection methods in general. Projections onto sets are used in a wide variety of methods in optimization theory but not every method that uses projections really belongs to the class of projection methods. *Projection methods* are iterative algorithms that use projections onto sets while relying on the general principle that when a family of (usually closed and convex) sets is present then projections onto the given individual sets are easier to perform than projections onto other sets (intersections, image sets under some transformation, etc.) that are derived from the given individual sets.

A projection algorithm reaches its goal, related to the whole family of sets, by performing projections onto the individual sets. Projection algorithms employ projections onto convex sets in various ways. They may use different kinds of projections and, sometimes, even use different projections within the same algorithm. They serve to solve a variety of problems which are either of the feasibility or the optimization types. They have different algorithmic structures, of which some are particularly suitable for parallel computing, and they demonstrate nice convergence properties and/or good initial behavior patterns.

Apart from theoretical interest, the main advantage of projection methods, which makes them successful in real-world applications, is computational. They commonly have the ability to handle huge-size problems of dimensions beyond which other, more sophisticated currently available, methods cease to be efficient. This is so because the building bricks of a projection algorithm are the projections onto the given individual sets (assumed and actually easy to perform) and the algorithmic structure is either sequential or simultaneous (or in-between). Sequential algorithmic structures cater for the row-action approach (see Censor [5]) while simultaneous algorithmic structures favor parallel computing platforms, see, e.g., Censor, Gordon and Gordon [8]. The field of projection methods is vast and we can only mention here a few recent works that can give the reader some good starting points. Such a list includes, among many others, the works of Crombez [18, 19], the connection with variational inequalities, see, e.g., Aslam Noor [28], Yamada's [30] which is motivated by real-world problems of signal processing, and the many contributions of Bauschke and Combettes, see, e.g., Bauschke,

Combettes and Kruk [2] and references therein. Consult Bauschke and Borwein [1] and Censor and Zenios [13, Chapter 5] for a tutorial review and a book chapter, respectively. Systems of linear equations, linear inequalities, or convex inequalities are all encompassed by the CFP which has broad applicability in many areas of mathematics and the physical and engineering sciences. These include, among others, optimization theory (see, e.g., Eremin [22], Censor and Lent [12] and Chinneck [14]), approximation theory (see, e.g., Deutsch [20] and references therein) and image reconstruction from projections in computerized tomography (see, e.g., Herman [23, 24], Censor [5]).

2 Subgradient generalized oblique projections onto convex sets

The following definitions of *oblique projections* and *generalized oblique projections* will lead us to define the new *subgradient generalized oblique projections* onto convex sets. Consider a hyperplane $H := \{x \in R^n \mid \langle a, x \rangle = b\}$, with $a \in R^n$, $b \in R$ and $a \neq 0$, where $\langle \cdot, \cdot \rangle$ denotes the inner product in R^n . Let G be an $n \times n$ symmetric positive definite matrix and let $\|x\|_G^2 := \langle x, Gx \rangle$ be the associated (squared) ellipsoidal norm. Given a point $z \in R^n$, the *oblique projection with respect to G of z onto H* is the unique point $P_H^G(z) \in H$ for which

$$P_H^G(z) = \arg \min \{\|x - z\|_G \mid x \in H\}. \quad (3)$$

Solving this minimization problem leads to

$$P_H^G(z) = z + \frac{b - \langle a, z \rangle}{\|a\|_{G^{-1}}^2} G^{-1} a \quad (4)$$

where G^{-1} is the inverse of G . For $G = I$, the unit matrix, P_H^G is the orthogonal projection and we denote it by P_H . We consider now generalized oblique projections onto a hyperplane with respect to a diagonal matrix $G = \text{diag}(g_1, g_2, \dots, g_n)$ for which some diagonal elements may be zero. Since this does not fit into the formula (4), we use the following definition from [8].

Definition 1 *Let $G = \text{diag}(g_1, g_2, \dots, g_n)$ with $g_j \geq 0$ for all $j = 1, 2, \dots, n$, be a given diagonal matrix. Let $H = \{x \in R^n \mid \langle a, x \rangle = b\}$ be a hyperplane*

with $0 \neq a = (a_j) \in R^n$ and $b \in R$, and assume that $g_j = 0$ if and only if $a_j = 0$. Define the vector $\gamma = (\gamma_j)_{j=1}^n$ by

$$\gamma_j = \begin{cases} a_j/g_j, & \text{if } g_j \neq 0, \\ 0, & \text{if } g_j = 0. \end{cases} \quad (5)$$

The generalized oblique projection of a point $z \in R^n$ onto H with respect to G is defined, for all $j = 1, 2, \dots, n$, by

$$(P_H^G(z))_j := z_j + \frac{b - \langle a, z \rangle}{\sum_{\substack{l=1 \\ g_l \neq 0}}^n a_l^2/g_l} \cdot \gamma_j. \quad (6)$$

We generalize this to half-spaces in the obvious way.

Definition 2 Let $G = \text{diag}(g_1, g_2, \dots, g_n)$ with $g_j \geq 0$ for all $j = 1, 2, \dots, n$, be a given diagonal matrix. Let $L = \{x \in R^n \mid \langle a, x \rangle \leq b\}$ be a half-space with $0 \neq a = (a_j) \in R^n$ and $b \in R$, and assume that $g_j = 0$ if and only if $a_j = 0$. For any point $z \in R^n$ let

$$c(z) := \min \left(0, (b - \langle a, z \rangle) / \left(\sum_{\substack{l=1 \\ g_l \neq 0}}^n a_l^2/g_l \right) \right). \quad (7)$$

The generalized oblique projection of the point z onto L with respect to G is defined by

$$P_L^G(z) := z + c(z) \cdot \gamma \quad (8)$$

where γ is as in Definition 1.

The set of all subgradients of a convex function f at a point z is called the subdifferential set of f at z and denoted by $\partial f(z)$. A vector $t \in R^n$ is a subgradient of f at a point z if

$$\langle t, x - z \rangle \leq f(x) - f(z), \text{ for all } x \in R^n, \quad (9)$$

see, e.g., [29, p. 214]. If f is differentiable at z then its gradient $\nabla f(z)$ is its unique subgradient at z and a convex function always has a subgradient.

Lemma 3 [1, Lemma 7.3] Let $f : R^n \rightarrow R$ be a convex function and let $z \in R^n$. Assume that the level-set $F := \{x \in R^n \mid f(x) \leq 0\}$ is nonempty and for any $t \in \partial f(z)$, define the closed convex set L by

$$L := L_f(z, t) := \{x \in R^n \mid f(z) + \langle t, x - z \rangle \leq 0\}. \quad (10)$$

Then

(i) $F \subseteq L$ and if $t \neq 0$ then L is a half-space; otherwise $L = R^n$.

(ii)

$$P_L(z) = \begin{cases} z - \frac{f(z)}{\|t\|^2}t, & \text{if } f(z) > 0, \\ z, & \text{if } f(z) \leq 0. \end{cases} \quad (11)$$

We denote the bounding hyperplane of L by

$$H := H_f(z, t) := \{x \in R^n \mid f(z) + \langle t, x - z \rangle = 0\}. \quad (12)$$

In order to define subgradient generalized oblique projections onto convex sets we need first the following information.

Definition 4 A real diagonal $n \times n$ matrix G with diagonal elements $g_j \geq 0$ is called *sparsity pattern oriented (SPO) with respect to a vector* $a = (a_j)_{j=1}^n$ if, for every $j = 1, 2, \dots, n$, $g_j \neq 0$ if and only if $a_j \neq 0$.

Given a point $z \in R^n$, let G be a real $n \times n$ diagonal matrix, which is SPO with respect to a subgradient of the convex function f at the point z . We will consider generalized oblique projections onto a half-space $L = L_f(z, t)$ with respect to G . Let G^\dagger be the Moore-Penrose generalized inverse of the $n \times n$ diagonal matrix G (see, e.g., Ben-Israel and Greville [3]), i.e.,

$$G^\dagger := \text{diag}(\tilde{g}_1, \tilde{g}_2, \dots, \tilde{g}_n), \quad (13)$$

where

$$\tilde{g}_j = \begin{cases} 1/g_j, & \text{if } g_j \neq 0, \\ 0, & \text{if } g_j = 0. \end{cases} \quad (14)$$

Definition 5 Let $f : R^n \rightarrow R$ be a convex function. Suppose that the level-set $F := \{x \in R^n \mid f(x) \leq 0\}$ is nonempty and let $z \in R^n$ be a given point. Taking any subgradient $t \in \partial f(z)$, let G be a real $n \times n$ diagonal matrix, $G = \text{diag}\{g_1, g_2, \dots, g_n\}$ with $g_j \geq 0$ for all $j = 1, 2, \dots, n$, which is SPO with

respect to t . The subgradient generalized oblique projection, with respect to G , of the point z onto the set F , denoted by $\Omega_F^G(z)$, is defined as

$$\Omega_F^G(z) := P_L^G(z), \quad (15)$$

where $L = L_f(z, t)$ is a half-space as in (10).

Denoting $\|x\|_G := \sqrt{\langle x, Gx \rangle}$, where G is a diagonal matrix with nonnegative diagonal entries, observe that $\|x\|_G$ is a vector seminorm (see, e.g., [27]) because it may be equal to zero for an $x \neq 0$ if G has at least one $g_j = 0$. We express Ω_F^G explicitly component-wise, under the conditions of Definition 5, by

$$(\Omega_F^G(z))_j = \begin{cases} z_j - \frac{f(z)\tilde{g}_j}{\sum_{l=1}^n \tilde{g}_l(t_l)^2} t_j, & \text{if } f(z) > 0, \\ z_j, & \text{if } f(z) \leq 0, \end{cases} \quad (16)$$

for all $j = 1, 2, \dots, n$, or in matrix notation,

$$\Omega_F^G(z) = \begin{cases} z - \frac{f(z)}{\|G^\dagger t\|_G^2} G^\dagger t, & \text{if } f(z) > 0, \\ z, & \text{if } f(z) \leq 0. \end{cases} \quad (17)$$

Notice that (17) generalizes (11). Indeed, if $G = \alpha I$, where I is the unit matrix and α is a positive scalar, then

$$\begin{aligned} (\Omega_F^G(z))_j &= \begin{cases} z_j - \frac{f(z)/\alpha}{\sum_{l=1}^n (t_l)^2/\alpha} t_j, & \text{if } f(z) > 0, \\ z_j, & \text{if } f(z) \leq 0, \end{cases} \\ &= \begin{cases} z_j - \frac{f(z)}{\sum_{l=1}^n (t_l)^2} t_j, & \text{if } f(z) > 0, \\ z_j, & \text{if } f(z) \leq 0, \end{cases} \\ &= (P_L(z))_j. \end{aligned} \quad (18)$$

3 The algorithm

Consider the CFP with $\{Q_i\}_{i=1}^m$ given as in (1) and make the following definition.

Definition 6 Let $f_i : R^n \rightarrow R$, $i = 1, 2, \dots, m$ be convex functions and let $x \in R^n$. Denote $t^i \in \partial f_i(x)$, $t^i = (t_j^i)_{j=1}^n$. The $m \times n$ matrix $T(x) = (t_{ij})_{i=1, j=1}^{m, n}$ is called a *generalized Jacobian of the family of functions $\{f_i\}_{i=1}^m$ at the point x* if $t_{ij} \equiv t_j^i$, for all i and all j , for some t^i .

This definition coincides in our case with the Clarke's generalized Jacobian, see [15] and [16]. A generalized Jacobian $T(x)$ of the functions in (2) is not unique because of the possibility to fill it up with different subgradients from each subdifferential set. In case all f_i are differentiable the generalized Jacobian reduces to the usual Jacobian. Furthermore, the generalized Jacobian might be a sparse matrix. Therefore, it is useful to make the next definition.

Definition 7 A family $\{G_i\}_{i=1}^m$ of real diagonal $n \times n$ matrices with all diagonal elements $g_{ij} \geq 0$ such that $\sum_{i=1}^m G_i = I$ will be called *sparsity pattern oriented (SPO) with respect to a generalized Jacobian of (1) at some given point z* if, for every $i = 1, 2, \dots, m$, $g_{ij} \neq 0$ if and only if $t_{ij} \neq 0$.

Thus, g_{ij} is the j -th diagonal element of the diagonal matrix G_i . Our algorithm for solving the CFP, with subgradient seminorm-induced generalized oblique projections, is a simultaneous projection algorithm that allows the use of *component-wise weighting* in the convex combination of projections at each iterative step in the following manner.

Algorithm 8

Initialization: $x^0 \in R^n$ is arbitrary.

Iterative step: Given the current iterate x^k , calculate the next iterate x^{k+1} by

$$x^{k+1} = x^k + \lambda_k \sum_{i=1}^m G_i^k \left(\Omega_i^{G_i^k}(x^k) - x^k \right), \quad (19)$$

where $\Omega_i^{G_i^k} \equiv \Omega_{Q_i}^{G_i^k}$, and in the k -th iterative step $\{G_i^k\}_{i=1}^m$ is a given family of diagonal SPO matrices with respect to a generalized Jacobian $T(x^k)$, and $\{\lambda_k\}_{k=0}^\infty$ are relaxation parameters.

Calculating explicitly the j -th component of the expression under the summation symbol in (19), using (17), we obtain

$$\left(G_i^k \left(\Omega_i^{G_i^k}(x^k) - x^k\right)\right)_j = \begin{cases} -g_{ij}^k \frac{f_i(x^k)}{\left\| (G_i^k)^\dagger t \right\|_{G_i^k}^2} \left((G_i^k)^\dagger t \right)_j, & \text{if } f_i(x^k) > 0, \\ 0, & \text{if } f_i(x^k) \leq 0. \end{cases} \quad (20)$$

This shows the component-wise weighting nature of the algorithm. Denoting the set of indices of functions which are active at a given point z by $\Xi(z) = \{i \mid f_i(z) > 0, i = 1, 2, \dots, m\}$, (19) can then be rewritten component-wise, using (16), as

$$x_j^{k+1} = x_j^k - \lambda_k \sum_{i \in \Xi(z)} \frac{f_i(x^k)}{\sum_{l=1}^n \tilde{g}_{il}^k(t_l)^2} t_j^i, \quad (21)$$

for $j = 1, 2, \dots, n$, where $t^i \in \partial f_i(x^k)$.

4 Convergence analysis

In this section we prove convergence of Algorithm 8 and propose a specific set of matrices $\{G_i\}_{i=1}^m$ which accelerates the algorithm. Our proof of convergence is done under the following conditions:

(C1) $f_i : R^n \rightarrow R$, for $i = 1, 2, \dots, m$, are convex functions, hence $Q_i \subseteq R^n$, for $i = 1, 2, \dots, m$, are closed convex sets as given in (1).

(C2) For every $k \geq 0$, $\{G_i^k\}_{i=1}^m$ is a set of real diagonal matrices which are SPO with respect to a generalized Jacobian $T(x^k)$, chosen at the k -th step.

(C3) $Q := \cap_{i=1}^m Q_i \neq \emptyset$.

(C4) There exists an $\varepsilon > 0$, such that, for every $i = 1, 2, \dots, m$, and for all $k \geq 0$, the diagonal elements g_{ij}^k of G_i^k , are for all $j = 1, 2, \dots, n$ either $g_{ij}^k = 0$ or $g_{ij}^k \geq \varepsilon > 0$.

We first treat the *unity relaxation* case $\lambda_k = 1$, for all $k \geq 0$, so that the iterative step has the form

$$x^{k+1} = \sum_{i=1}^m G_i^k \Omega_i^{G_i^k}(x^k) \quad (22)$$

and remove this restriction later.

Lemma 9 *If $\{x^k\}_{k=0}^\infty$ is any sequence, generated by Algorithm 8 with unity relaxation, then for any $x \in R^n$ and for every $k \geq 0$,*

$$\sum_{i=1}^m \left| \Omega_i^{G_i^k}(x^k) - x \right|_{G_i^k}^2 = \sum_{i=1}^m \left| \Omega_i^{G_i^k}(x^k) - x^{k+1} \right|_{G_i^k}^2 + \|x - x^{k+1}\|_2^2. \quad (23)$$

Proof. The proof follows the lines of Proposition 3.4 in [7].

$$\begin{aligned} & \sum_{i=1}^m \left| \Omega_i^{G_i^k}(x^k) - x \right|_{G_i^k}^2 - \sum_{i=1}^m \left| \Omega_i^{G_i^k}(x^k) - x^{k+1} \right|_{G_i^k}^2 \\ &= \sum_{i=1}^m \left(\left| x \right|_{G_i^k}^2 - \left| x^{k+1} \right|_{G_i^k}^2 + 2 \left\langle G_i^k \Omega_i^{G_i^k}(x^k), x^{k+1} - x \right\rangle \right). \end{aligned} \quad (24)$$

Using the fact that, due to $\sum_{i=1}^m G_i^k = I$, we have, for any $y \in R^n$,

$$\sum_{i=1}^m \left| y \right|_{G_i^k}^2 = \|y\|_2^2 \quad (25)$$

the result follows. ■

Lemma 10 *Let $L \subseteq R^n$ be a half-space and let $G \neq 0$ be a nonnegative diagonal matrix and let P_L^G be the generalized oblique projection as in (8). If $z \in L$ is any given point, then for any $y \in R^n$, the following inequality holds:*

$$\left| P_L^G(y) - y \right|_G^2 \leq \left| z - y \right|_G^2 - \left| z - P_L^G(y) \right|_G^2 \quad (26)$$

Proof. If $y \notin L$, it is obvious that $P_L^G(y) \equiv P_H^G(y)$, where $H = \{x \in R^n \mid \langle a, x \rangle = b\}$ is the bounding hyperplane of L . Let $z = P_H^G(y) + h$ for some h . Then

$$\begin{aligned} & \left| z - y \right|_G^2 - \left| z - P_H^G(y) \right|_G^2 - \left| P_H^G(y) - y \right|_G^2 \\ &= 2 \langle h, G(P_H^G(y) - y) \rangle. \end{aligned} \quad (27)$$

Then using $\langle y - P_H^G(y), G(z - P_H^G(y)) \rangle \leq 0$ we obtain (26). If $y \in L$ then the result follows immediately. ■

Definition 11 A sequence $\{x^k\}_{k=0}^{\infty}$ is called *Fejér-monotone with respect to a nonempty set* $C \subseteq R^n$ if $\|x^{k+1} - x\|_2 \leq \|x^k - x\|_2$, for all $k \geq 0$, for every $x \in C$.

Lemma 12 Any sequence $\{x^k\}_{k=0}^{\infty}$, generated by Algorithm 8 with unity relaxation, is Fejér-monotone with respect to Q .

Proof. Let $x \in Q$, then, using (23), we have

$$\begin{aligned} \|x - x^k\|_2^2 - \|x - x^{k+1}\|_2^2 &= \|x - x^k\|_2^2 - \sum_{i=1}^m \left| \Omega_i^{G_i^k}(x^k) - x \right|_{G_i^k}^2 \\ &\quad + \sum_{i=1}^m \left| \Omega_i^{G_i^k}(x^k) - x^{k+1} \right|_{G_i^k}^2 \\ &= \sum_{i=1}^m \left(\left| x - x^k \right|_{G_i^k}^2 - \left| \Omega_i^{G_i^k}(x^k) - x \right|_{G_i^k}^2 \right) \\ &\quad + \sum_{i=1}^m \left| \Omega_i^{G_i^k}(x^k) - x^{k+1} \right|_{G_i^k}^2. \end{aligned} \quad (28)$$

For the second summand in the last expression we have,

$$\sum_{i=1}^m \left| \Omega_i^{G_i^k}(x^k) - x^{k+1} \right|_{G_i^k}^2 \geq 0, \quad (29)$$

and taking $G = G_i^k$, $z = x$, $y = x^k$ and $P_L^G(y) = \Omega_i^{G_i^k}(x^k)$ in Lemma 10 yields,

$$\sum_{i=1}^m \left(\left| x - x^k \right|_{G_i^k}^2 - \left| \Omega_i^{G_i^k}(x^k) - x \right|_{G_i^k}^2 \right) \geq \sum_{i=1}^m \left| \Omega_i^{G_i^k}(x^k) - x^k \right|_{G_i^k}^2 \geq 0, \quad (30)$$

from which the result follows. ■

Theorem 13 [1, Lemma 2.16] Suppose that a sequence $\{x^k\}_{k=0}^{\infty}$ is Fejér-monotone with respect to some closed convex set C then $\{x^k\}_{k=0}^{\infty}$ converges to some point in C if and only if $\{x^k\}_{k=0}^{\infty}$ has cluster points all lying in C .

Now we are ready to prove the convergence result.

Theorem 14 *Under Conditions (C1)–(C4), any sequence $\{x^k\}_{k=0}^\infty$, generated by Algorithm 8 with unity relaxation, converges to a solution of (1).*

Proof. From Fejér-monotonicity of the sequence $\{x^k\}_{k=0}^\infty$ follows that the sequence $\{\|x^k - x\|\}_{k=0}^\infty$ is monotonically decreasing and bounded, therefore, $\lim_{k \rightarrow \infty} \|x^k - x\| = d$ for some d . Hence $\|x - x^k\|_2^2 - \|x - x^{k+1}\|_2^2$ tends to 0 as $k \rightarrow \infty$. This implies, via (28) and (30), that

$$\lim_{k \rightarrow \infty} \left\| \Omega_i^{G_i^k}(x^k) - x^k \right\|_{G_i^k} = 0. \quad (31)$$

For any $i \in \Xi(x^k)$, (17) implies

$$\begin{aligned} \left\| \Omega_i^{G_i^k}(x^k) - x^k \right\|_{G_i^k} &= \left\| x^k - \frac{f_i(x^k)}{\left\| (G_i^k)^\dagger t^i \right\|_{G_i^k}^2} (G_i^k)^\dagger t^i - x^k \right\|_{G_i^k} \\ &= \frac{f_i(x^k)}{\left\| (G_i^k)^\dagger t^i \right\|_{G_i^k}}, \text{ where } t^i \in \partial f_i(x^k), \end{aligned} \quad (32)$$

and so, by (31), we have

$$\lim_{k \rightarrow \infty} \frac{f_i(x^k)}{\left\| (G_i^k)^\dagger t^i \right\|_{G_i^k}} = 0. \quad (33)$$

Since the functions $f_i : R^n \rightarrow R$, for $i = 1, 2, \dots, m$, are convex, their subdifferentials are uniformly bounded on bounded sets, see, e.g., Bauschke and Borwein [1, Corollary 7.9]. Therefore, for some $\hat{x} \in Q$ there exists a constant $\alpha \equiv \alpha(\hat{x})$ such that $\|t\|_2 \leq \alpha$, for all subgradients $t \in \partial f_i(x)$, for all $i = 1, 2, \dots, m$, and for all $x \in R^n$ for which $\|x - \hat{x}\|_2 \leq \|x^0 - \hat{x}\|_2$. If we denote

$$B_{\hat{x}} := \{x \in R^n \mid \|x - \hat{x}\|_2 \leq \|x^0 - \hat{x}\|_2\}, \quad (34)$$

then $x^k \in B_{\hat{x}}$ for every $k \geq 0$, by repeated application of Fejér monotonicity of $\{x^k\}_{k=0}^\infty$ with respect to Q and because $\hat{x} \in Q$, and so all subgradients $t \in \partial f_i(x^k)$, for all $k \geq 0$, are bounded, i.e., $\|t\|_2 \leq \alpha$. Therefore, regardless of which t^i is taken, $(t_j^i)^2 \leq \beta$ for some constant β , for all $j = 1, 2, \dots, n$.

From the definition of the seminorm, from Condition (C4) and from the properties of G_i^k and $(G_i^k)^\dagger$ follows

$$\begin{aligned} \left\| (G_i^k)^\dagger t^i \right\|_G^2 &= \langle (G_i^k)^\dagger t^i, (G_i^k)(G_i^k)^\dagger t^i \rangle = \sum_{j=1}^n \tilde{g}_{ij}^k t_j^i g_{ij}^k \tilde{g}_{ij}^k t_j^i \\ &= \sum_{j=1}^n \tilde{g}_{ij}^k (t_j^i)^2 \leq \sum_{j=1}^n \beta / \varepsilon. \end{aligned} \quad (35)$$

Combining this result with (33) we obtain $\lim_{k \rightarrow \infty} f_i(x^k) = 0$. Since $\{x^k\}_{k=0}^\infty$ is bounded, it has a cluster point, say, x^* and

$$\lim_{p \rightarrow \infty} x^{k_p} = x^*. \quad (36)$$

Since $\lim_{p \rightarrow \infty} f_i(x^{k_p}) = 0$, for all $i = 1, 2, \dots, m$, we have that $x^* \in Q$. Therefore, from Theorem 13 follows that $\{x^k\}_{k=0}^\infty$ converges to some point in Q and the proof is complete. ■

Next we get rid of the unity relaxation assumption and allow *underrelaxation*.

Theorem 15 *Under Conditions (C1)–(C4), any sequence $\{x^k\}_{k=0}^\infty$, generated by Algorithm 8 with relaxation parameters $\{\lambda_k\}_{k=0}^\infty$ such that $0 < \theta \leq \lambda_k \leq 1$, for all $k \geq 0$, for some arbitrarily small but fixed θ , converges to a point $x^* \in Q$.*

Proof. Define the following $m + 1$ diagonal $n \times n$ matrices

$$\Gamma_i^k := \lambda_k G_i^k, \text{ for all } i = 1, 2, \dots, m \text{ and for all } k \geq 0, \quad (37)$$

$$\Gamma_{m+1}^k := I - \sum_{i=1}^m \Gamma_i^k, \text{ for all } k \geq 0, \quad (38)$$

and define an additional constraint set as the level-set of the identically-zero function $f_{m+1}(x) = 0$, i.e., $Q_{m+1} = R^n$. Then Algorithm 8 takes the form

$$x^{k+1} = \sum_{i=1}^{m+1} \Gamma_i^k \Omega_i^{\Gamma_i^k}(x^k), \quad (39)$$

because $\Omega_i^{\Gamma_i^k} = \Omega_i^{G_i^k}$. Also, the original CFP obviously remains unchanged, and $\sum_{i=1}^{m+1} \Gamma_i^k = I$, for all $k \geq 0$. Thus, Theorem 14 applies and the result follows. ■

Finally, we propose a specific way to construct at the k -th iterative step diagonal matrices G_i^k , $i = 1, 2, \dots, m$, with nonnegative entries, that will accelerate the algorithm when the generalized Jacobians are sparse. Dropping the index k for convenience, let s_j be the number of nonzero elements in the j -th column of the generalized Jacobian $T(x)$ of (1), thus taking into account only rows with indices i for which $f_i(x) > 0$, and define

$$g_{ij} := \begin{cases} 1/s_j, & \text{if } t_{ij} \neq 0, \\ 0, & \text{if } t_{ij} = 0. \end{cases} \quad (40)$$

The generalized inverses $(G_i)^\dagger = \text{diag}(\widetilde{g}_{i1}, \widetilde{g}_{i2}, \dots, \widetilde{g}_{in})$ are then expressed by

$$\widetilde{g}_{ij} = \begin{cases} s_j, & \text{if } g_{ij} = 1/s_j, \\ 0, & \text{if } g_{ij} = 0. \end{cases} \quad (41)$$

In case that the $m \times n$ matrix $T(x)$ is sparse the s_j values will be smaller than m . Therefore, such diagonal component-wise weighting will accelerate the initial progress of the iterates generated by the algorithm. This has been demonstrated experimentally in the linear case in [8] and [9].

Acknowledgments. The authors thank an anonymous reviewer for helpful comments on an earlier version of the paper. This work was supported in part 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. This research was also supported by grant No. 2003275 from the United States-Israel Binational Science Foundation (BSF).

References

- [1] H.H. Bauschke and J.M. Borwein, On projection algorithms for solving convex feasibility problems, *SIAM Review* **38** (1996), 367–426.
- [2] H.H. Bauschke, P.L. Combettes and S.G. Kruk, Extrapolation algorithm for affine-convex feasibility problems, *Numerical Algorithms* **41** (2006), 239–274.

- [3] A. Ben-Israel and T.N.E. Greville, *Generalized Inverses: Theory and Applications*, second edition, Springer-Verlag, New York, NY, USA, 2003.
- [4] J.T. Bettes and P.D. Frank, A sparse nonlinear optimization algorithm, *Journal of Optimization Theory and Applications* **82** (1994), 519–541.
- [5] Y. Censor, Row-action methods for huge and sparse systems and their applications, *SIAM Review* **23** (1981), 444–466.
- [6] Y. Censor and T. Elfving, Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem, *SIAM Journal on Matrix Analysis and Applications* **24** (2002), 40–58.
- [7] Y. Censor and T. Elfving, Iterative algorithms with seminorm-induced oblique projections, *Abstract and Applied Analysis* **8** (2003), 387–406.
- [8] Y. Censor, D. Gordon and R. Gordon, Component averaging: An efficient iterative parallel algorithm for large and sparse unstructured problems, *Parallel Computing* **27** (2001), 777–808.
- [9] Y. Censor, D. Gordon and R. Gordon, BICAV: A block-iterative, parallel algorithm for sparse systems with pixel-related weighting, *IEEE Transactions on Medical Imaging* **20** (2001), 1050–1060.
- [10] Y. Censor, T. Elfving, N. Kopf and T. Bortfeld, The multiple-sets split feasibility problem and its applications for inverse problems, *Inverse Problems* **21** (2005), 2071–2084.
- [11] Y. Censor, T. Bortfeld, B. Martin and A. Trofimov, A unified approach for inversion problems in intensity-modulated radiation therapy, *Physics in Medicine and Biology* **51** (2006), 2353–2365.
- [12] Y. Censor and A. Lent, Cyclic subgradient projections, *Mathematical Programming* **24** (1982), 233–235.
- [13] Y. Censor and S.A. Zenios, *Parallel Optimization: Theory, Algorithms, and Applications*, Oxford University Press, New York, NY, USA, 1997.
- [14] J.W. Chinneck, The constraint consensus method for finding approximately feasible points in nonlinear programs, *INFORMS Journal on Computing* **16** (2004), 255–265.

- [15] F.H. Clarke, Generalized gradients and applications, *Transactions of the American Mathematical Society* **205** (1975), 247–262.
- [16] F.H. Clarke, On the inverse function theorem, *Pacific Journal of Mathematics* **64** (1976), 97–102.
- [17] P. Combettes, Convex set theoretic image recovery by extrapolated iterations of parallel subgradient projections, *IEEE Transactions on Image Processing* **6** (1997), 493–506.
- [18] G. Crombez, Non-monotoneous parallel iteration for solving convex feasibility problems, *Kybernetika* **39** (2003), 547–560.
- [19] G. Crombez, A sequential iteration algorithm with non-monotoneous behaviour in the method of projections onto convex sets, *Czechoslovak Mathematical Journal* **56** (2006), 491–506.
- [20] F. Deutsch, The method of alternating orthogonal projections, in: S.P. Singh (Editor), *Approximation Theory, Spline Functions and Applications*, Kluwer Academic publishers, Dordrecht, The Netherlands, 1992, pp. 105–121.
- [21] L.T. Dos Santos, A parallel subgradient projections method for the convex feasibility problem, *Journal of Computational and Applied Mathematics* **18** (1987), 307–320.
- [22] I.I. Eremin, Fejér mappings and convex programming, *Siberian Mathematical Journal* **10** (1969), 762–772.
- [23] G.T. Herman, *Image Reconstruction From Projections: The Fundamentals of Computerized Tomography*, Academic Press, New York, NY, USA, 1980.
- [24] G.T. Herman and L.B. Meyer, Algebraic reconstruction techniques can be made computationally efficient, *IEEE Transactions on Medical Imaging* **12** (1993), 600–609.
- [25] M. Jiang and G. Wang, Convergence studies on iterative algorithms for image reconstruction, *IEEE Transactions on Medical Imaging* **22** (2003), 569–579.

- [26] K. Kiwiel and B. Lopuch, Surrogate projection methods for finding fixed points of firmly nonexpansive mappings, *SIAM Journal on Optimization* **7** (1997), 1084–1102.
- [27] E. Kreyszig, *Introductory Functional Analysis with Applications*, John Wiley & Sons, New York, NY, USA, 1978.
- [28] M. Aslam Noor, Some developments in general variational inequalities, *Applied Mathematics and Computation* **152** (2004), 197–277.
- [29] R.T. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, USA, 1970.
- [30] I. Yamada, Hybrid steepest descent method for variational inequality problem over the fixed point set of certain quasi-nonexpansive mappings, *Numerical Functional Analysis and Optimization* **25** (2004), 619–655.
- [31] J.H. Wilkinson, Introduction to Part II, The Algebraic Eigenvalue Problem, *Handbook for Automatic Computation*, Vol II, Linear Algebra, J.H. Wilkinson and C. Reinsch (Editors), Springer-Verlag, 1971.