Journal of Nonlinear and Convex Analysis
Volume 15, Number 4, 2014, 681–691

Yokohama Publishers
ISSN 1880-5221 ONLINE JOURNAL

AN ELEMENTARY PROOF OF CONVERGENCE FOR THE FORWARD-BACKWARD SPLITTING ALGORITHM

CHARLES L. BYRNE

Abstract. The forward-backward splitting (FBS) algorithm is a quite general iterative method that includes, as particular cases, the projected gradient descent algorithm for constrained minimization, the CQ algorithm for the split feasibility problem, the projected Landweber algorithm for constrained least squares, and the simultaneous orthogonal projection algorithm for the convex feasibility problem. The FBS algorithm involves iterating with respect to an averaged operator that is the product of two firmly non-expansive operators, one of which is Moreau's proximity operator. The usual proof of convergence employs the Krasnosel'skiĭ-Mann Theorem. This proof depends, therefore, on knowing that if the gradient of a convex differentiable function is non-expansive, then it is firmly non-expansive, and that the composition of averaged operators is again averaged. Neither of these results is trivial to prove, especially the former. In this paper we give an elementary proof of convergence of the FBS algorithm that does not rely on these results.

1. Background

The usual proof of convergence of the forward-backward splitting algorithm, as presented, for example, by Combettes and Wajs [11], relies on properties of averaged and firmly non-expansive operators, and on the Krasnosel'skiĭ-Mann (KM) Theorem [16, 18]. In this section we elaborate on these points. Details can be found in [14, 1, 6] and in the references therein.

1.1. Firmly Non-expansive and Averaged Operators.

In this subsection we review the basic theory of firmly non-expansive and averaged operators.

Definition 1.1. An operator $T : \mathbb{R}^J \to \mathbb{R}^J$ is L-Lipschitz continuous, with respect to the two-norm on $\mathbb{R}^J$, if , for every x and y in $\mathbb{R}^J$ we have

$$\|Tx - Ty\|_2 \leq L\|x - y\|_2. \tag{1.1}$$

Definition 1.2. An operator N on $\mathbb{R}^J$ is called non-expansive (ne), with respect to the two-norm on $\mathbb{R}^J$, if, for every x and y in $\mathbb{R}^J$, we have

$$\|x - y\|_2 \geq \|Nx - Ny\|_2. \tag{1.2}$$

Clearly, if T is L-Lipschitz continuous, then the operator $N = \frac{1}{L}T$ is non-expansive.

2010 *Mathematics Subject Classification.* 94A12, 65K11.

Key words and phrases. Forward-backward splitting algorithm, proximity operator, iterative optimization.

Definition 1.3. An operator F on $\mathbb{R}^J$ is called firmly non-expansive (fne) if, for every x and y in $\mathbb{R}^J$, we have

$$\langle Fx - Fy, x - y\rangle \geq \|Fx - Fy\|_2^2. \tag{1.3}$$

Using Cauchy's Inequality, we show easily that a firmly non-expansive operator on $\mathbb{R}^J$ is non-expansive. The following lemma provides a useful characterization of fne operators.

Lemma 1.4. *An operator $F : \mathbb{R}^J \to \mathbb{R}^J$ is fne if and only if $F = \frac{1}{2}(I + N)$, for some operator N that is ne with respect to the two-norm.*

Proof. Suppose that $F = \frac{1}{2}(I + N)$. We show that F is fne if and only if N is ne in the two-norm. First, we have

$$\langle Fx - Fy, x - y\rangle = \frac{1}{2}\|x - y\|_2^2 + \frac{1}{2}\langle Nx - Ny, x - y\rangle.$$

Also,

$$\left\|\frac{1}{2}(I + N)x - \frac{1}{2}(I + N)y\right\|_2^2 = \frac{1}{4}\|x - y\|^2 + \frac{1}{4}\|Nx - Ny\|^2 + \frac{1}{2}\langle Nx - Ny, x - y\rangle.$$

Therefore,

$$\langle Fx - Fy, x - y\rangle \geq \|Fx - Fy\|_2^2$$

if and only if

$$\|Nx - Ny\|_2^2 \leq \|x - y\|_2^2.$$

□

Corollary 1.5. *An operator F is fne if and only if $I - F$ is fne.*

Definition 1.6. An operator $A : \mathbb{R}^J \to \mathbb{R}^J$ is averaged (av) if $A = (1 - \alpha)I + \alpha N$, for some operator N that is non-expansive with respect to the two-norm, and some scalar α in $[0, 1)$.

It is clear from the definitions and Lemma 1.4 that any fne operator is av. The product of finitely many av operators is again av [1]. According to the Krasnosel'skiĭ-Mann Theorem [18, 16], if A is averaged and has fixed points, then the sequence $\{A^n x^0\}$ converges to a fixed point of A, for every initial vector x^0. The following theorem is well known; see, for example, [15, 19].

Theorem 1.7. *Let $f(x)$ be convex and differentiable and its derivative, $\nabla f(x)$, non-expansive in the two-norm. Then $\nabla f(x)$ is firmly non-expansive.*

Suppose that $g(x) : \mathbb{R}^J \to \mathbb{R}$ is convex and the function $\nabla g(x)$ is L-Lipschitz continuous. Let $f(x) = \frac{1}{L}g(x)$, so that $\nabla f(x)$ is a non-expansive operator. According to Theorem 1.7, the operator $\nabla f(x) = \frac{1}{L}\nabla g(x)$ is firmly non-expansive. The proof of Theorem 1.7 is not trivial. In [15] Golshtein and Tretyakov prove the following theorem, from which Theorem 1.7 follows immediately.

Theorem 1.8. *Let $g : \mathbb{R}^J \to \mathbb{R}$ be convex and differentiable. The following are equivalent:*

1)

$$||\nabla g(x) - \nabla g(y)||_2 \leq ||x - y||_2; \tag{1.4}$$

2)

$$g(x) \geq g(y) + \langle \nabla g(y), x - y \rangle + \frac{1}{2}||\nabla g(x) - \nabla g(y)||_2^2; \tag{1.5}$$

and

3)

$$\langle \nabla g(x) - \nabla g(y), x - y \rangle \geq ||\nabla g(x) - \nabla g(y)||_2^2. \tag{1.6}$$

The proof of Theorem 1.8 given in [15] is repeated in [8].

If $f : \mathbb{R}^J \to \mathbb{R}$ is convex and differentiable, and ∇f is L-Lipschitz continuous, then $A = I - \gamma \nabla f$ is averaged, for any γ in the interval $(0, 2/L)$.

1.2. **Moreau's Proximity Operators.** Let $f : \mathbb{R}^J \to \mathbb{R}$ be convex. For each $z \in \mathbb{R}^J$ the function

$$m_f(z) := \min_x \Big\{ f(x) + \frac{1}{2}\|x - z\|_2^2 \Big\} \tag{1.7}$$

is minimized by a unique x [24]. The operator that associates with each z the minimizing x is Moreau's proximity operator, and we write $x = \text{prox}_f(z)$. The operator prox_f extends the notion of orthogonal projection onto a closed convex set [20, 21, 22]. We have $x = \text{prox}_f(z)$ if and only if $z - x \in \partial f(x)$, where the set $\partial f(x)$ is the sub-differential of f at x, given by

$$\partial f(x) := \{u | \langle u, y - x \rangle \leq f(y) - f(x), \text{for all}\, y\}. \tag{1.8}$$

Proximity operators are also firmly non-expansive [11]; indeed, the proximity operator prox_f is the resolvent of the maximal monotone operator $B(x) = \partial f(x)$ and all such resolvent operators are firmly non-expansive [4].

1.3. **The Forward-Backward Splitting Algorithm.** Our objective here is to provide an elementary proof of convergence for the forward-backward splitting (FBS) algorithm; a detailed discussion of this algorithm and its history is given by Combettes and Wajs in [11].

Let $f : \mathbb{R}^J \to \mathbb{R}$ be convex, with $f = f_1 + f_2$, both convex, f_2 differentiable, and ∇f_2 L-Lipschitz continuous. The iterative step of the FBS algorithm is

$$x^k = \text{prox}_{\gamma f_1}\Big(x^{k-1} - \gamma \nabla f_2(x^{k-1}) \Big). \tag{1.9}$$

The FBS iteration has the form $x^k = Ax^{k-1}$. When we select γ in the interval $(0, 2/L)$, the operator A becomes averaged, since it is now the product of a fne operator and an av operator. Convergence of the sequence $\{x^k\}$ to a fixed point of A, whenever A has fixed points, then follows from the KM Theorem.

As we shall show, convergence of the sequence $\{x^k\}$ to a fixed point of A can be established without using the KM Theorem or the machinery of fne and av operators, if γ is chosen to lie within the interval $(0, 1/L]$.

2. Sequential unconstrained optimization

Sequential unconstrained optimization algorithms can be used to minimize a function $f : \mathbb{R}^J \to (-\infty, \infty]$ over a (not necessarily proper) subset C of $\mathbb{R}^J$ [13]. At the kth step of a sequential unconstrained minimization method we obtain x^k by minimizing the function

$$G_k(x) = f(x) + g_k(x), \tag{2.1}$$

where the auxiliary function $g_k(x)$ is appropriately chosen. If C is a proper subset of $\mathbb{R}^J$ we may force $g_k(x) = +\infty$ for x not in C, as in the barrier-function methods; then each x^k will lie in C. The objective is then to select the $g_k(x)$ so that the sequence $\{x^k\}$ converges to a solution of the problem, or failing that, at least to have the sequence $\{f(x^k)\}$ converging to the infimum of $f(x)$ over x in C.

Our main focus in this paper is the use of sequential unconstrained optimization algorithms to obtain iterative methods in which each iterate can be obtained in closed form. Now the auxiliary functions $g_k(x)$ are selected not to impose a constraint, but to facilitate computation.

3. SUMMA

In [7] we presented a particular class of sequential unconstrained minimization methods called SUMMA. As we showed in that paper, this class is broad enough to contain barrier-function methods, proximal minimization methods, and the simultaneous multiplicative algebraic reconstruction technique (SMART). By reformulating the problem, the penalty-function methods can also be shown to be members of the SUMMA class. Any alternating minimization (AM) problem with the five-point property [12] can be reformulated as a SUMMA problem; therefore the expectation maximization maximum likelihood (EMML) algorithm for Poisson data, which is such an AM algorithm, must also be a SUMMA algorithm.

For a method to be in the SUMMA class we require that each auxiliary function $g_k(x)$ satisfy the inequality

$$0 \le g_k(x) \le G_{k-1}(x) - G_{k-1}(x^{k-1}), \tag{3.1}$$

for all x. Note that it follows that $g_k(x^{k-1}) = 0$, for all k.

We assume that the inequality in (3.1) holds for each k. We also assume that $\inf f(x) = b > -\infty$. The next two results are taken from [7].

Proposition 3.1. *The sequence $\{f(x^k)\}$ is non-increasing and the sequence $\{g_k(x^k)\}$ converges to zero.*

Proof. We have

$$f(x^{k+1}) + g_{k+1}(x^{k+1}) = G_{k+1}(x^{k+1}) \le G_{k+1}(x^k) = f(x^k).$$

Therefore,

$$f(x^k) - f(x^{k+1}) \ge g_{k+1}(x^{k+1}).$$

The sequence $\{f(x^k)\}$ is decreasing to a finite limit, since it is bounded below by b, and, therefore, the sequence $\{g_k(x^k)\}$ converges to zero. □

Theorem 3.2. *The sequence $\{f(x^k)\}$ converges to b.*

Proof. Suppose that there is $\delta > 0$ such that $f(x^k) \geq b + 2\delta$, for all k. Then there is $z \in C$ such that $f(x^k) \geq f(z) + \delta$, for all k. From the inequality in (3.1) we have

$$g_k(z) - g_{k+1}(z) \geq f(x^k) + g_k(x^k) - f(z) \geq f(x^k) - f(z) \geq \delta, \tag{3.2}$$

for all k. But this cannot happen; the successive differences of a non-increasing sequence of non-negative terms must converge to zero. □

4. Convergence of the FBS algorithm

Let $f : \mathbb{R}^J \to \mathbb{R}$ be convex, with $f = f_1 + f_2$, both convex, f_2 differentiable, and ∇f_2 L-Lipschitz continuous. Let $\{x^k\}$ be defined by Equation (1.9) and let $0 < \gamma \leq 1/L$.

For each $k = 1, 2, \ldots$ let

$$G_k(x) = f(x) + \frac{1}{2\gamma}\|x - x^{k-1}\|_2^2 - D_{f_2}(x, x^{k-1}), \tag{4.1}$$

where

$$D_{f_2}(x, x^{k-1}) = f_2(x) - f_2(x^{k-1}) - \langle \nabla f_2(x^{k-1}), x - x^{k-1} \rangle. \tag{4.2}$$

Since $f_2(x)$ is convex, $D_{f_2}(x, y) \geq 0$ for all x and y and is the Bregman distance formed from the function f_2 [3].

The auxiliary function

$$g_k(x) = \frac{1}{2\gamma}\|x - x^{k-1}\|_2^2 - D_{f_2}(x, x^{k-1}) \tag{4.3}$$

can be rewritten as

$$g_k(x) = D_h(x, x^{k-1}), \tag{4.4}$$

where

$$h(x) = \frac{1}{2\gamma}\|x\|_2^2 - f_2(x). \tag{4.5}$$

Therefore, $g_k(x) \geq 0$ whenever $h(x)$ is a convex function.

We know that $h(x)$ is convex if and only if

$$\langle \nabla h(x) - \nabla h(y), x - y \rangle \geq 0, \tag{4.6}$$

for all x and y. This is equivalent to

$$\frac{1}{\gamma}\|x - y\|_2^2 - \langle \nabla f_2(x) - \nabla f_2(y), x - y \rangle \geq 0. \tag{4.7}$$

Since ∇f_2 is L-Lipschitz, the inequality (4.7) holds for $0 < \gamma \leq 1/L$.

Lemma 4.1. *The x^k that minimizes $G_k(x)$ over x is given by Equation (1.9).*

Proof. We know that x^k minimizes $G_k(x)$ if and only if

$$0 \in \nabla f_2(x^k) + \frac{1}{\gamma}(x^k - x^{k-1}) - \nabla f_2(x^k) + \nabla f_2(x^{k-1}) + \partial f_1(x^k),$$

or, equivalently,

$$\Big(x^{k-1} - \gamma \nabla f_2(x^{k-1})\Big) - x^k \in \partial(\gamma f_1)(x^k).$$

Consequently,

$$x^k = \text{prox}_{\gamma f_1}(x^{k-1} - \gamma \nabla f_2(x^{k-1})).$$

□

Theorem 4.2. *The sequence $\{x^k\}$ converges to a minimizer of the function $f(x)$, whenever minimizers exist.*

Proof. A relatively simple calculation shows that

$$G_k(x) - G_k(x^k) = \frac{1}{2\gamma}\|x - x^k\|_2^2 +$$

$$\Big(f_1(x) - f_1(x^k) - \frac{1}{\gamma}\langle (x^{k-1} - \gamma\nabla f_2(x^{k-1})) - x^k, x - x^k\rangle\Big). \tag{4.8}$$

Since

$$(x^{k-1} - \gamma\nabla f_2(x^{k-1})) - x^k \in \partial(\gamma f_1)(x^k),$$

it follows that

$$\Big(f_1(x) - f_1(x^k) - \frac{1}{\gamma}\langle (x^{k-1} - \gamma\nabla f_2(x^{k-1})) - x^k, x - x^k\rangle\Big) \geq 0.$$

Therefore,

$$G_k(x) - G_k(x^k) \geq \frac{1}{2\gamma}\|x - x^k\|_2^2 \geq g_{k+1}(x). \tag{4.9}$$

Therefore, the inequality in (3.1) holds and the iteration fits into the SUMMA class.

Now let $\hat{x}$ minimize $f(x)$ over all x. Then

$$G_k(\hat{x}) - G_k(x^k) = f(\hat{x}) + g_k(\hat{x}) - f(x^k) - g_k(x^k)$$
$$\leq f(\hat{x}) + G_{k-1}(\hat{x}) - G_{k-1}(x^{k-1}) - f(x^k) - g_k(x^k),$$

so that

$$\Big(G_{k-1}(\hat{x}) - G_{k-1}(x^{k-1})\Big) - \Big(G_k(\hat{x}) - G_k(x^k)\Big) \geq f(x^k) - f(\hat{x}) + g_k(x^k) \geq 0.$$

Therefore, the sequence $\{G_k(\hat{x}) - G_k(x^k)\}$ is decreasing and the sequences $\{g_k(x^k)\}$ and $\{f(x^k) - f(\hat{x})\}$ converge to zero.

From

$$G_k(\hat{x}) - G_k(x^k) \geq \frac{1}{2\gamma}\|\hat{x} - x^k\|_2^2,$$

it follows that the sequence $\{x^k\}$ is bounded and that a subsequence converges to some x^* with $f(x^*) = f(\hat{x})$.

Replacing the generic $\hat{x}$ with x^*, we find that $\{G_k(x^*) - G_k(x^k)\}$ is decreasing, and by Equation (4.8), a subsequence, and therefore, the entire sequence, converges to zero. From the inequality in (4.9), we conclude that the sequence $\{\|x^* - x^k\|_2^2\}$ converges to zero, and so $\{x^k\}$ converges to x^*. This completes the proof of the theorem. □

5. Some examples

We present some examples to illustrate the application of the convergence theorem.

5.1. **Projected Gradient Descent.** Let C be a non-empty, closed convex subset of $\mathbb{R}^J$ and $f_1(x) = \iota_C(x)$, the function that is $+\infty$ for x not in C and zero for x in C. Then $\iota_C(x)$ is convex, but not differentiable. We have $\text{prox}_{\gamma f_1} = P_C$, the orthogonal projection onto C. The iteration in Equation (1.9) becomes

$$x^k = P_C\Big(x^{k-1} - \gamma \nabla f_2(x^{k-1})\Big). \tag{5.1}$$

The sequence $\{x^k\}$ converges to a minimizer of f_2 over $x \in C$, whenever such minimizers exist, for $0 < \gamma \leq 1/L$.

5.1.1. *The CQ Algorithm.* Let A be a real I by J matrix, $C \subseteq \mathbb{R}^J$, and $Q \subseteq \mathbb{R}^I$, both closed convex sets. The split feasibility problem (SFP) is to find x in C such that Ax is in Q. The function

$$f_2(x) = \frac{1}{2}\|P_Q Ax - Ax\|_2^2 \tag{5.2}$$

is convex, differentiable and ∇f_2 is L-Lipschitz for $L = \rho(A^T A)$, the spectral radius of $A^T A$. The gradient of f_2 is

$$\nabla f_2(x) = A^T(I - P_Q)Ax. \tag{5.3}$$

We want to minimize the function $f_2(x)$ over x in C, or, equivalently, to minimize the function $f(x) = \iota_C(x) + f_2(x)$. The projected gradient descent algorithm has the iterative step

$$x^k = P_C\Big(x^{k-1} - \gamma A^T(I - P_Q)Ax^{k-1}\Big); \tag{5.4}$$

this iterative method was called the CQ-algorithm in [5, 6]. The sequence $\{x^k\}$ converges to a solution whenever f_2 has a minimum on the set C, for $0 < \gamma \leq 1/L$.

In [10, 9] the CQ algorithm was extended to a multiple-sets algorithm and applied to the design of protocols for intensity-modulated radiation therapy.

5.1.2. *The Projected Landweber Algorithm.* The problem is to minimize the function

$$f_2(x) = \frac{1}{2}\|Ax - b\|_2^2,$$

over $x \in C$. This is a special case of the SFP and we can use the CQ-algorithm, with $Q = \{b\}$. The resulting iteration is the projected Landweber algorithm [2]; when $C = \mathbb{R}^J$ it becomes the Landweber algorithm [17].

6. Minimizing f_2 over a linear manifold

Suppose that we want to minimize f_2 over x in the linear manifold $M = S + p$, where S is a subspace of $\mathbb{R}^J$ of dimension $I < J$ and p is a fixed vector. Let A be an I by J matrix such that the I columns of A^T form a basis for S. For each $z \in \mathbb{R}^I$ let

$$d(z) = f_2(A^T z + p),$$

so that d is convex, differentiable, and its gradient,

$$\nabla d(z) = A\nabla f_2(A^T z + p),$$

is K-Lipschitz continuous, for $K = \rho(A^TA)L$. The sequence $\{z^k\}$ defined by

$$z^k = z^{k-1} - \gamma\nabla d(z^{k-1}) \tag{6.1}$$

converges to a minimizer of d over all z in $\mathbb{R}^I$, whenever minimizers exist, for $0 < \gamma \leq 1/K$.

From Equation (6.1) we get

$$x^k = x^{k-1} - \gamma A^TA\nabla f_2(x^{k-1}), \tag{6.2}$$

with $x^k = A^Tz^k + p$. The sequence $\{x^k\}$ converges to a minimizer of f_2 over all x in M.

Suppose now that we begin with an algorithm having the iterative step

$$x^k = x^{k-1} - \gamma A^TA\nabla f_2(x^{k-1}), \tag{6.3}$$

where A is any real I by J matrix having rank I. Let x^0 be in the range of A^T, so that $x^0 = A^Tz^0$, for some $z^0 \in \mathbb{R}^I$. Then each $x^k = A^Tz^k$ is again in the range of A^T, and we have

$$A^Tz^k = A^Tz^{k-1} - \gamma A^TA\nabla f_2(A^Tz^{k-1}). \tag{6.4}$$

With $d(z) = f_2(A^Tz)$, we can write Equation (6.4) as

$$A^T\Big(z^k - (z^{k-1} - \gamma\nabla d(z^{k-1}))\Big) = 0. \tag{6.5}$$

Since A has rank I, A^T is one-to-one, so that

$$z^k - z^{k-1} - \gamma\nabla d(z^{k-1}) = 0. \tag{6.6}$$

The sequence $\{z^k\}$ converges to a minimizer of d, over all $z \in \mathbb{R}^I$, whenever such minimizers exist, for $0 < \gamma \leq 1/K$. Therefore, the sequence $\{x^k\}$ converges to a minimizer of f_2 over all x in the range of A^T.

7. Feasible-point algorithms

Suppose that we want to minimize a convex differentiable function $f(x)$ over x such that $Ax = b$, where A is an I by J full-rank matrix, with $I < J$. If $Ax^k = b$ for each of the vectors $\{x^k\}$ generated by the iterative algorithm, we say that the algorithm is a feasible-point method.

7.1. The Projected Gradient Algorithm.

Let C be the feasible set of all x in $\mathbb{R}^J$ such that $Ax = b$. For every z in $\mathbb{R}^J$, we have

$$P_Cz = P_{NS(A)}z + A^T(AA^T)^{-1}b, \tag{7.1}$$

where $NS(A)$ is the null space of A. Using

$$P_{NS(A)}z = z - A^T(AA^T)^{-1}Az, \tag{7.2}$$

we have

$$P_Cz = z + A^T(AA^T)^{-1}(b - Az). \tag{7.3}$$

Using Equation (1.9), we get the iteration step for the projected gradient algorithm:

$$x^k = x^{k-1} - \gamma P_{NS(A)}\nabla f(x^{k-1}), \tag{7.4}$$

which converges to a solution for $0 < \gamma \leq 1/L$, whenever solutions exist.

In the next subsection we present a somewhat simpler approach.

7.2. **The Reduced Gradient Algorithm.** Let x^0 be a feasible point, that is, $Ax^0 = b$. Then $x = x^0 + p$ is also feasible if p is in the null space of A, that is, $Ap = 0$. Let Z be a J by $J - I$ matrix whose columns form a basis for the null space of A. We want $p = Zv$ for some v. The best v will be the one for which the function

$$\phi(v) = f(x^0 + Zv)$$

is minimized. We can apply to the function $\phi(v)$ the steepest descent method, or the Newton-Raphson method, or any other minimization technique.

The steepest descent method, applied to $\phi(v)$, is called the reduced steepest descent algorithm [23]. The gradient of $\phi(v)$, also called the reduced gradient, is

$$\nabla\phi(v) = Z^T \nabla f(x),$$

where $x = x^0 + Zv$; the gradient operator $\nabla\phi$ is then K-Lipschitz, for $K = \rho(A^T A)L$.

Let x^0 be feasible. The iteration in Equation (1.9) now becomes

$$v^k = v^{k-1} - \gamma\nabla\phi(v^{k-1}), \tag{7.5}$$

so that the iteration for $x^k = x^0 + Zv^k$ is

$$x^k = x^{k-1} - \gamma Z Z^T \nabla f(x^{k-1}). \tag{7.6}$$

The vectors x^k are feasible and the sequence $\{x^k\}$ converges to a solution, whenever solutions exist, for any $0 < \gamma < \frac{1}{K}$.

7.3. **The Reduced Newton-Raphson Method.** The same idea can be applied to the Newton-Raphson method. The Newton-Raphson method, applied to $\phi(v)$, is called the reduced Newton-Raphson method [23]. The Hessian matrix of $\phi(v)$, also called the reduced Hessian matrix, is

$$\nabla^2\phi(v) = Z^T \nabla^2 f(c) Z,$$

so that the reduced Newton-Raphson iteration becomes

$$x^k = x^{k-1} - Z\Big(Z^T \nabla^2 f(x^{k-1}) Z\Big)^{-1} Z^T \nabla f(x^{k-1}). \tag{7.7}$$

Let c^0 be feasible. Then each x^k is feasible. The sequence $\{x^k\}$ is not guaranteed to converge.

8. Conclusions

The forward-backward splitting algorithm can be formulated as a member of the SUMMA class of sequential unconstrained minimization algorithms. Convergence of the iterative sequence can then be established without relying on the machinery of firmly non-expansive and averaged operators. Examples are given to illustrate the usefulness of the forward-backward splitting algorithm.

Acknowledgment

I wish to thank the anonymous reviewers for helpful suggestions.

References

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

[2] M. Bertero and P. Boccacci, *Introduction to Inverse Problems in Imaging*, Institute of Physics Publishing, Bristol, UK, 1998.

[3] L. M. Bregman, *The relaxation method of finding the common point of convex sets and its application to the solution of problems in convex programming*, USSR Computational Mathematics and Mathematical Physics **7** (1967), 200–217.

[4] R. Bruck and S. Reich, *Nonexpansive projections and resolvents of accretive operators in Banach spaces*, Houston J. of Mathematics **3** (1977), 459–470.

[5] C. Byrne, *Iterative oblique projection onto convex sets and the split feasibility problem*, Inverse Problems **18** (2002), 441–453.

[6] C. Byrne, *A unified treatment of some iterative algorithms in signal processing and image reconstruction*, Inverse Problems **20** (2004), 103–120.

[7] C. Byrne, *Sequential unconstrained minimization algorithms for constrained optimization*, Inverse Problems, **24** (2008), article no. 015013.

[8] C. Byrne, *A First Course in Optimization*, CRC Press, Boca Raton, 2014.

[9] 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.

[10] Y. Censor, T. Elfving, N. Kopf and T. Bortfeld, *The multiple-sets split feasibility problem and its application for inverse problems*, Inverse Problems **21** (2005), 2071–2084.

[11] P. Combettes and V. Wajs, *Signal recovery by proximal forward-backward splitting*, Multiscale Modeling and Simulation **4** (2005), 1168–1200.

[12] I. Csiszár and G. Tusnády, *Information geometry and alternating minimization procedures*, Statistics and Decisions **Supp. 1** (1984), 205–237.

[13] A. Fiacco and G. McCormick, *Nonlinear Programming: Sequential Unconstrained Minimization Techniques*, SIAM Classics in Mathematics (reissue), Philadelphia, PA, 1990.

[14] K. Goebel and S. Reich, *Uniform Convexity, Hyperbolic Geometry, and Nonexpansive Mappings*, Dekker, New York, 1984.

[15] E. Golshtein and N. Tretyakov, *Modified Lagrangians and Monotone Maps in Optimization*, John Wiley and Sons, Inc., New York, 1996.

[16] M. Krasnosel'skiĭ, *Two remarks on the method of successive approximations*, Uspekhi Mathematicheskikh Nauk **10** (1955), 123–127 (in Russian).

[17] L. Landweber, *An iterative formula for Fredholm integral equations of the first kind*, Amer. J. of Math. **73** (1951), 615–624.

[18] W. Mann, *Mean value methods in iteration*, Proceedings of the American Mathematical Society **4** (1953), 506–510.

[19] E. Masad and S. Reich, *A note on the multiple-set split convex feasibility problem in Hilbert space*, J. Nonlinear Convex Analysis **8** (2007), 367–371.

[20] J.-J. Moreau, *Fonctions convexes duales et points proximaux dans un espace hilbertien*, C.R. Acad. Sci. Paris Sér. A Math. **255** (1962), 2897–2899.

[21] J.-J. Moreau, *Propriétés des applications 'prox'*, C.R. Acad. Sci. Paris Sér. A Math. **256** (1963), 1069–1071.

[22] J.-J. Moreau, *Proximité et dualité dans un espace hilbertien*, Bull. Soc. Math. France **93** (1965), 273–299.

[23] S. Nash and A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, New York, 1996.

[24] R. Rockafellar, *Convex Analysis*, Princeton University Press, Princeton, NJ, 1970

Manuscript received April 17, 2012
revised March August, 2012

Charles L. Byrne
Department of Mathematical Sciences, University of Massachusetts Lowell, Lowell, MA 01854
E-mail address: Charles_Byrne@uml.edu