# On A Generalized Baillon–Haddad Theorem for Convex Functions on Hilbert Space

Charles L. Byrne*

July 17, 2014

**Abstract**

The Baillon–Haddad Theorem asserts that, if the gradient operator of a convex and Fréchet differentiable function on a Hilbert space is nonexpansive, then it is firmly nonexpansive. This theorem plays an important role in iterative optimization. In this note we present a short, elementary proof of a recent extension of the Baillon–Haddad Theorem due to Bauschke and Combettes.

**Key Words:** Bregman distance, convex function, firmly nonexpansive, gradient, nonexpansive, Krasnosel'skii–Mann Theorem.

**2000 Mathematics Subject Classification:** Primary 47H09, 90C25; Secondary 26A51, 26B25.

## 1 Introduction

We denote by $\mathcal{H}$ a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. An operator $T : \mathcal{H} \to \mathcal{H}$ is nonexpansive if, for all $x$ and $y$ in $\mathcal{H}$,

$$\|Tx - Ty\| \leq \|x - y\|, \tag{1.1}$$

and firmly nonexpansive if, in addition,

$$\langle Tx - Ty, x - y \rangle \geq \|Tx - Ty\|^2. \tag{1.2}$$

Clearly, if $T$ is firmly nonexpansive, then $T$ is nonexpansive. However, it is possible for (1.1) to hold without (1.2) holding; let $T = -Id$, for example, where $Id$ is the identity operator. For certain operators $T$ the two properties are equivalent; we have the following theorem.

*Charles_Byrne@uml.edu, Department of Mathematical Sciences, University of Massachusetts Lowell, Lowell, MA 01854

**Theorem 1.1 (The Baillon–Haddad Theorem)** ([1], Corollaire 10]) *Let $f : \mathcal{H} \to \mathbb{R}$ be convex, Fréchet differentiable on $\mathcal{H}$, and its gradient operator $T = \nabla f$ nonexpansive. Then $T$ is firmly nonexpansive.*

In [1] this theorem appears as a corollary of a more general theorem concerning $n$-cyclically monotone operators in normed vector space.

In [2] Bauschke and Combettes generalize the Baillon–Haddad Theorem, giving four additional conditions equivalent to the two in Theorem 1.1. An abbreviated summary of their results is the following theorem.

**Theorem 1.2 (Bauschke and Combettes)** *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and Fréchet differentiable . The following are equivalent:*

1. *the gradient operator $T = \nabla f$ is nonexpansive;*
2. *the function $F(x) = \frac{1}{2}\|x\|^2 - f(x)$ is convex;*
3. *for all $x$ and $z$ we have*
$$\frac{1}{2}\|z-x\|^2 \geq D_f(z,x) = f(z) - f(x) - \langle \nabla f(x), z-x \rangle \geq 0; \tag{1.3}$$
4. *the gradient operator $T = \nabla f$ is firmly nonexpansive.*

In [2] the proof that Condition 1. of Theorem 1.2 (their Condition (i)) implies Condition 4. (their condition (vi)), that is, the proof of the Baillon–Haddad Theorem 1.1, is indirect. Their intermediate conditions (iii), (iv) and (v) involve the Fenchel conjugate, Moreau's proximity operator, the Moreau envelope, and the Moreau Decomposition Theorem. Our proof of Theorem 1.2 is short and elementary and uses only fundamental properties of convex functions. The proof of Theorem 1.1 given in [7] was reproduced in [3]. The proof given here for Theorem 1.2 is based on the one given for Theorem 1.1 in [4] and [5].

# 2 Proof of Theorem 1.2

Prove (2), assuming (1), that the operator $T = \nabla f$ is nonexpansive. The function $F(x) = \frac{1}{2}\|x\|^2 - f(x)$ is therefore Fréchet differentiable and $\nabla F(x) = x - \nabla f(x)$. Since

$$\langle \nabla F(z) - \nabla F(x), z-x \rangle \geq \|z-x\|(\|z-x\| - \|\nabla f(z) - \nabla f(x)\|) \geq 0,$$

we know that $F(x)$ is a convex function.

Prove (3), assuming (2), that $F$ is convex. Since $F$ is convex, and $\nabla F(x) = x - \nabla f(x)$, we have

$$F(z) \geq F(x) + \langle \nabla F(x), z - x \rangle, \tag{2.1}$$

which is equivalent to

$$\frac{1}{2}\|z - x\|^2 \geq D_f(z, x). \tag{2.2}$$

Prove (4), assuming (3), that (2.2) holds for all $z$ and $x$. Let $y \in \mathcal{H}$ be arbitrary and fixed. Let $d(x) = D_f(x, y)$. Then $d(x)$ is convex and $\nabla d(x) = \nabla f(x) - \nabla f(y)$. It is easily seen that $D_f(z, x) = D_d(z, x)$, so from (2.2) we have

$$\frac{1}{2}\|z - x\|^2 \geq D_d(z, x) = d(z) - d(x) - \langle \nabla f(x) - \nabla f(y), z - x \rangle. \tag{2.3}$$

Now let $z = x - \nabla f(x) + \nabla f(y)$. Inserting this $z$ into (2.3), we obtain

$$D_f(x, y) = d(x) \geq d(z) + \frac{1}{2}\|\nabla f(x) - \nabla f(y)\|^2. \tag{2.4}$$

Similarly, we can show that

$$D_f(y, x) \geq \frac{1}{2}\|\nabla f(x) - \nabla f(y)\|^2. \tag{2.5}$$

Adding the previous two inequalities, we get

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle \geq \|\nabla f(x) - \nabla f(y)\|^2, \tag{2.6}$$

so $T = \nabla f$ is firmly nonexpansive. Since (4) obviously implies (1), the proof is complete. ∎

Notice that we have actually proved a somewhat stronger inequality than (1.2):

$$\langle \nabla f(x) - \nabla f(y), x - y \rangle - \|\nabla f(x) - \nabla f(y)\|^2$$

$$\geq D_f(x - \nabla f(x) + \nabla f(y), y) + D_f(y - \nabla f(y) + \nabla f(x), x) \geq 0. \tag{2.7}$$

In Theorem 2.1 of [2] Bauschke and Combettes do not assume, a priori, that $f$ is Fréchet differentiable; they include Fréchet differentiability as part of conditions (1), (3) and (4), making it necessary to prove that $f$ is Fréchet differentiable, if $F$ is convex. Proving this involves tools that are not elementary; we can prove this using the following proposition, which appears as Corollary 16.38 in [2]. The proof in [2] uses the Fréchet differentiability of the Moreau envelope.

**Proposition 2.1** *Let $f : \mathcal{H} \to ]-\infty, +\infty]$ and $g : \mathcal{H} \to ]-\infty, +\infty]$ be proper, convex, and lower semicontinuous. Then $\partial(f+g) = \partial f + \partial g$.*

The proof of this proposition is not elementary. We use this result in the form of the following corollary.

**Corollary 2.1** *Let $f : \mathcal{H} \to ] - \infty, +\infty]$ and $g : \mathcal{H} \to ] - \infty, +\infty]$ be proper, convex, and lower semicontinuous, such that $f+g$ is Gâteaux differentiable. Then both $f$ and $g$ are Gâteaux differentiable.*

Now, given that $F(x)$ is convex and $F(x) + f(x) = \frac{1}{2}\|x\|^2$, we may conclude that $f$ and $F$ are Gâteaux differentiable. Now we proceed as in the earlier proof, showing first that the inequality in (2.3) holds, and then that $\nabla f$ is firmly nonexpansive.

The Baillon-Haddad Theorem plays an important role in iterative optimization, as we shall discuss in the next two sections.

# 3 The Krasnosel'skii–Mann Theorem

Let $T : \mathcal{H} \to \mathcal{H}$ be an arbitrary operator on $\mathcal{H}$, and $G = Id - T$. Then an easy calculation shows that, for any $x$ and $y$ in $\mathcal{H}$, we have

$$\|x-y\|^2 - \|Tx - Ty\|^2 = 2\langle Gx - Gy, x - y\rangle - \|Gx - Gy\|^2. \tag{3.1}$$

It follows immediately from Equation (3.1) that

$$\langle Tx - Ty, x - y\rangle - \|Tx - Ty\|^2 = \langle Gx - Gy, x - y\rangle - \|Gx - Gy\|^2. \tag{3.2}$$

**Definition 3.1** *An operator $G : \mathcal{H} \to \mathcal{H}$ is $\nu$-*inverse strongly monotone *($\nu$-*ism*) for some $\nu > 0$ if*

$$\langle Gx - Gy, x - y\rangle \geq \nu\|Gx - Gy\|^2. \tag{3.3}$$

Using Equation (3.1), we find that $T$ is nonexpansive if and only if $G = Id - T$ is $\frac{1}{2}$-ism.

If $T$ is a continuous operator and the sequence $\{x^k\}$ defined by $x^k = Tx^{k-1}$, for $k = 1, 2, ...,$ converges to $z$, then $Tz = z$ and $z$ is called a fixed point of $T$. An important problem in optimization theory is to find conditions on the operator $T$ so that the iterative sequence converges (weakly, if not strongly) to a fixed point of $T$. Just having $T$ nonexpansive is not enough, as the operator $T = -Id$ illustrates.

**Definition 3.2** *An operator $T$ on $\mathcal{H}$ is called* $\alpha$-averaged *(*$\alpha$-av*), for some $\alpha$ in the interval $(0,1)$, if there is a nonexpansive operator $N$ such that*

$$T = (1-\alpha)Id + \alpha N. \tag{3.4}$$

The following proposition follows immediately from Equations (3.1) and (3.2).

**Proposition 3.1** *An operator $T$ is $\alpha$-averaged if and only if $G = Id - T$ is $\frac{1}{2\alpha}$-ism. Also $T$ is firmly nonexpansive if and only if $G = Id - T$ is 1-ism, and if and only if $G$ is firmly nonexpansive.*

It is easy to see that, if $G$ is $\nu$-ism and $\gamma > 0$, then $\gamma G$ is $\frac{\nu}{\gamma}$-ism.

Showing that a given operator $T$ is averaged by using Definition 3.4 can be difficult, while showing that its complement, $G = Id - T$, is $\frac{1}{2\alpha}$-ism can be simpler. The operator $P_C$, the orthogonal projection onto the nonempty, closed, convex set $C$, is firmly nonexpansive, which can be easily shown using the well known characterization of $z = P_C x$ in Proposition 3.2. Therefore, $P_C$ is also averaged, but this is far from obvious, if we just focus on Definition 3.4.

**Proposition 3.2** *Let $C \subseteq \mathcal{H}$ be nonempty, closed, and convex. Then $z = P_C x$ if and only if $z$ is a member of $C$, and, for all members $y$ of $C$, we have*

$$\langle z - x, y - z \rangle \geq 0. \tag{3.5}$$

It is interesting to note that the operator $P_C$ is actually the gradient of the convex, differentiable function $f(x) = \frac{1}{2}\Big(\|x\|^2 - \|x - P_C x\|^2\Big)$. The Krasnosel'skii–Mann Theorem is the following.

**Theorem 3.1** *Let $T$ be $\alpha$-averaged, with a fixed point. Then, for any $x^0$, the sequence $\{x^k\}$ defined by $x^k = Tx^{k-1}$, for $k = 1, 2, ...$, converges weakly to a fixed point of $T$.*

**Proof:** Let $Tz = z$. We have

$$\|z - x^k\|^2 - \|z - x^{k+1}\|^2 \geq \Big(\frac{1}{\alpha} - 1\Big)\|x^k - x^{k+1}\|^2, \tag{3.6}$$

from which we conclude that the sequence $\{\|z - x^k\|^2\}$ is decreasing, the sequence $\{\|x^k - x^{k+1}\|^2\}$ converges to zero, and the sequence $\{x^k\}$ is bounded. If $\mathcal{H}$ is finite-dimensional, we know that $\{x^k\}$ has a cluster point $x^*$, and that $Tx^* = x^*$; using $x^*$ in place of $z$, we find that the entire sequence $\{\|x^* - x^k\|^2\}$ converges to zero. In the infinite-dimensional case the proof is a bit more complicated; we must show that the weak cluster point is unique (see [6]). ∎

The class of nonexpansive operators is closed to finite products, but, as we have seen, nonexpansiveness is not strong enough for our purposes. Being firmly nonexpansive is sufficient for Theorem 3.1 to apply, but the firmly nonexpansive operators are not closed to finite products; the operator $P_C$ is firmly nonexpansive, but the finite product of such operators need not be firmly nonexpansive. The class of averaged operators is the appropriate class to consider, since it contains the firmly nonexpansive operators, is closed to finite products, and Theorem 3.1 holds for averaged operators.

# 4 Gradient Descent Methods

Suppose now that $g : \mathcal{H} \to \mathbb{R}$ is a convex, differentiable function, and we want to minimize $g(x)$. The well known gradient descent method has the iterative step

$$x^{k+1} = x^k - \gamma_k \nabla g(x^k), \tag{4.1}$$

where the step-length parameters $\gamma_k$ are selected to guarantee that $g(x^{k+1}) \leq g(x^k)$. It is helpful if we can select a parameter $\gamma > 0$ that is independent of $k$ and so that the iteration

$$x^{k+1} = x^k - \gamma \nabla g(x^k) \tag{4.2}$$

generates a sequence $\{x^k\}$ that converges (weakly, if not strongly) to a point $z$ with $\nabla g(z) = 0$. With the operator $T$ given by

$$T = Id - \gamma \nabla g, \tag{4.3}$$

we see that $\nabla g(z) = 0$ if and only if $Tz = z$. The question is then: When is $T$ $\alpha$-averaged?

**Definition 4.1** *An operator $T$ on $\mathcal{H}$ is called $L$-Lipschitz continuous if, for all $x$ and $y$, we have*

$$\|Tx - Ty\| \leq L\|x - y\|. \tag{4.4}$$

It is a obvious consequence of Theorem 1.1 that, if $\nabla g$ is $L$-Lipschitz continuous, then $\nabla g$ is $\frac{1}{L}$-ism. Suppose now that $T = \nabla g$ is $L$-Lipschitz continuous. Then $\nabla f$ nonexpansive, for the function $f = \frac{1}{L} g$. Therefore, according to the Baillon-Haddad Theorem, $\nabla f$ is firmly nonexpansive, and

$$\gamma \nabla g = (\gamma L)(\frac{1}{L} \nabla g) = (\gamma L) \nabla f$$

is $\frac{1}{\gamma L}$-ism. Consequently, if $\gamma L = 2\alpha$, for some $\alpha$ in $(0, 1)$, then the operator $T = Id - \gamma \nabla g$ is $\alpha$-averaged and Theorem 3.1 applies.

Determining if the gradient of $g(x)$ is Lipschitz continuous may not be easy, and, if it is, finding $L$ may not be easy. Consider the function $g(x) = \frac{1}{2}\|Ax - b\|_2^2$, where $A$ is a real $M$ by $N$ matrix. The gradient of $g(x)$ is $L$-Lipschitz continuous, for $L = \rho(A^T A)$, which, in this case, is the largest eigenvalue of the matrix $A^T A$. In many image-processing problems the $x$ is a vectorized image, $A$ is sparse, and both $M$ and $N$ can be several thousand. In such cases calculating the matrix $A^T A$ is unreasonable. Also, $A^T A$ need not be sparse, even if $A$ is sparse. Because the step-length parameter $\gamma$ must be chosen in the interval $(0, 2/L)$, we want an estimate of $L$ that does not greatly overestimate it, and makes use of the sparseness of $A$. Such estimates are discussed in [4].

# References


1. Baillon, J.-B., and Haddad, G. (1977) "Quelques propriétés des opérateurs angle-bornés et n-cycliquement monotones." *Israel J. of Mathematics*, **26**, pp. 137-150.

2. Bauschke, H., and Combettes, P. (2010) "The Baillon-Haddad Theorem Revisited." *J. Convex Analysis*, **17**, pp. 781–787.

3. Byrne, C. (2007) *Applied Iterative Methods*. Wellesley, MA: A K Peters.

4. Byrne, C. (2014) *Iterative Optimization in Inverse Problems*. Boca Raton, FL: CRC Press.

5. Byrne, C. (2014) *A First Course in Optimization*. Boca Raton, FL: CRC Press.

6. Cegielski, A. (2012) *Iterative Methods for Fixed Point Problems in Hilbert Space*. Heidelberg: Springer Lecture Notes in Mathematics 2057.

7. Golshtein, E., and Tretyakov, N. (1996) *Modified Lagrangians and Monotone Maps in Optimization*. New York: John Wiley and Sons.

8. Rockafellar, R.T. (1970) *Convex Analysis*. Princeton, NJ: Princeton University Press.