Iterative Convex Optimization Algorithms; Part One: Using the Baillon–Haddad Theorem

Charles L. Byrne[*]

February 12, 2015

Abstract

We denote by $\mathcal{H}$ a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. We say that an operator $T : \mathcal{H} \to \mathcal{H}$ is *convergent* if, for every starting vector x^0, the sequence $\{x^k\}$ defined by $x^k = Tx^{k-1}$ converges weakly to a fixed point of T, whenever T has a fixed point. Fixed-point iterative methods are used to solve a variety of problems by selecting a convergent T for which the fixed points of T are solutions of the original problem. It is important, therefore, to identify properties of an operator T that guarantee that T is convergent.

An operator $T : \mathcal{H} \to \mathcal{H}$ is nonexpansive if, for all x and y in $\mathcal{H}$,

$$\|Tx - Ty\| \leq \|x - y\|.$$

Just being nonexpansive does not make T convergent, as the example $T = -I$ shows; here I is the identity operator. It doesn't take much, however, to convert a nonexpansive operator N into a convergent operator. Let $0 < \alpha < 1$ and $T = (1 - \alpha)I + \alpha N$; then T is convergent. Such operators are called *averaged* and are convergent as a consequence of the Krasnosel'skii-Mann Theorem.

The Baillon–Haddad Theorem provides an important link between convex optimization and fixed-point iteration. If $f : \mathcal{H} \to \mathbb{R}$ is a Gâteaux differentiable convex function and its gradient is L-Lipschitz continuous, that is,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|,$$

for all x and y, then f is Fréchet differentiable and the gradient operator of the function $g = \frac{1}{L} f$ is nonexpansive. By the Baillon–Haddad Theorem the gradient operator of g is firmly nonexpansive. It follows that, for any $0 < \gamma < \frac{2}{L}$, the operator $I - \gamma \nabla f$ is averaged, and therefore convergent. The class of averaged operators is closed to finite products, and P_C, the orthogonal projection onto a closed convex set C, is firmly nonexpansive. Therefore, the projected gradient-descent algorithm with the iterative step

$$x^{k+1} = P_C(x^k - \gamma \nabla f(x^k))$$

converges weakly to a minimizer, over C, of the function f, whenever such minimizers exist.

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

1 Overview

The Baillon–Haddad Theorem, described by Bauschke and Combettes [4] as "*a remarkable result that has important applications in optimization*" , first appeared in a 1977 paper [2] as a corollary to a theorem on cyclically monotone operators on Hilbert space. Since then several variants of the BH Theorem have appeared in the literature [11, 4, 5]. The first elementary proof of the BH Theorem will be published this year [8].

In order to place the BH Theorem in context, we begin with fixed-point iteration, convergent operators, averaged operators, and the Krasnosel'skii-Mann Theorem. Then we give the elementary proof of the BH Theorem and discuss its importance for iterative optimization.

We then turn to an interesting version of the BH Theorem due to Bauschke and Combettes [4]. A certain amount of background must be presented first, since their theorem involves several notions, such as Fenchel conjugation and Moreau envelopes, that may not be famliar to most readers. Their theorem asserts the equivalence of several conditions. Their proof is relatively short, because some of the equivalences were already known. In no sense, however, is their proof elementary.

2 Fixed-Point Iteration and Convergent Operators

2.1 Convergent Operators

We denote by $\mathcal{H}$ a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and norm $\| \cdot \|$. We say that an operator $T : \mathcal{H} \to \mathcal{H}$ is *convergent* if, for every starting vector x^0, the sequence $\{x^k\}$ defined by $x^k = Tx^{k-1}$ converges weakly to a fixed point of T, whenever T has a fixed point. Fixed-point iterative methods are used to solve a variety of problems by selecting a convergent T for which the fixed points of T are solutions of the original problem. It is important, therefore, to identify properties of an operator T that guarantee that T is convergent.

2.2 Firmly Nonexpansive Operators

We are interested in operators T that are convergent. For such operators we often find that $\|x^{k+1} - x^k\| \leq \|x^k - x^{k-1}\|$ for each k. This leads us to the definition of *nonexpansive* operators.

Definition 2.1 *An operator T on $\mathcal{H}$ is* nonexpansive (ne) *if, for all x and y, we have*

$$\|Tx - Ty\| \leq \|x - y\|.$$

Nonexpansive operators need not be convergent, as the ne operator $T = -I$ illustrates.

Let C be a nonempty, closed, convex subset of $\mathcal{H}$. For every $x \in \mathcal{H}$ there is a unique member of C, denoted $P_C x$, that is closest to x in C. The operators $T = P_C$ are nonexpansive. In fact, the operators P_C have a much stronger property; they are firmly nonexpansive.

Definition 2.2 *An operator T on $\mathcal{H}$ is* firmly nonexpansive (fne) *if, for every x and y, we have*

$$\langle Tx - Ty, x - y \rangle \geq \|Tx - Ty\|^2.$$

If T is fne then T is convergent. The class of fne operators is smaller than the class of ne operators and does yield convergent iterative sequences. However, the product or composition of two or more fne operators need not be fne, which limits the usefulness of this class of operators. Even the product of P_{C_1} and P_{C_2} need not be fne. We need to find a class of convergent operators that is closed to finite products.

2.3 Averaged Operators

It can be shown easily that an operator F is fne if and only if there is a nonexpansive operator N such that

$$F = \frac{1}{2}I + \frac{1}{2}N.$$

Definition 2.3 *An operator $A : \mathcal{H} \to \mathcal{H}$ is* α-averaged (α-av) *if there is a nonexpansive operator N such that*

$$A = (1 - \alpha)I + \alpha N,$$

for some α in the interval $(0, 1)$. If A is α-av *for some α then A is an* averaged (av) *operator.*

All averaged operators are nonexpansive, all firmly nonexpansive operators are averaged, the class of averaged operators is closed to finite products, and averaged operators are convergent. In other words, the class of averaged operators is precisely the class that we are looking for.

2.4 Useful Properties of Operators on $\mathcal{H}$

It turns out that properties of an operator T are often more easily studied in terms of properties of its complement, $G = I - T$. The following two identities are easy to prove and are quite helpful. For any operator $T : \mathcal{H} \to \mathcal{H}$ and $G = I - T$ we have

$$\|x - y\|^2 - \|Tx - Ty\|^2 = 2\langle Gx - Gy, x - y\rangle - \|Gx - Gy\|^2, \tag{2.1}$$

and

$$\langle Tx - Ty, x - y\rangle - \|Tx - Ty\|^2 = \langle Gx - Gy, x - y\rangle - \|Gx - Gy\|^2. \tag{2.2}$$

Definition 2.4 *An operator $G : \mathcal{H} \to \mathcal{H}$ is* ν-inverse strongly monotone (ν-ism) *for some $\nu > 0$ if*

$$\langle Gx - Gy, x - y\rangle \geq \nu\|Gx - Gy\|^2,$$

for all x and y.

Clearly, if G is ν-ism then γG is $\frac{\nu}{\gamma}$-ism. Using the two identities in (2.1) and (2.2) it is easy to prove the following theorem.

Theorem 2.1 *Let $T : \mathcal{H} \to \mathcal{H}$ be arbitrary and $G = I - T$. Then*

1. *T is ne if and only if G is ν-ism for $\nu = \frac{1}{2}$;*
2. *T is α-av if and only if G is ν-ism for $\nu = \frac{1}{2\alpha}$, for some $0 < \alpha < 1$;*
3. *T is fne if and only if G is ν-ism for $\nu = 1$.*
4. *T is fne if and only if G is fne;*
5. *If G is ν-ism and $0 < \mu \leq \nu$, then G is μ-ism.*

3 The Krasnosel'skii-Mann Theorem

For any operator $T : \mathcal{H} \to \mathcal{H}$ that is averaged, weak convergence of the sequence $\{T^k x^0\}$ to a fixed point of T, whenever $\mathrm{Fix}(T)$, the set of fixed points of T, is nonempty, is guaranteed by the Krasnosel'skii-Mann (KM) Theorem [12, 14]. The proof we present here is for the case of $\mathcal{H} = \mathbb{R}^N$; the proof is a bit more complicated for the infinite-dimensional case (see Theorem 5.14 in [5]).

Theorem 3.1 *Let $T : \mathbb{R}^N \to \mathbb{R}^N$ be α-averaged, for some $\alpha \in (0, 1)$. Then, for any x^0, the sequence $\{T^k x^0\}$ converges to a fixed point of T, whenever $\mathrm{Fix}(T)$ is nonempty.*

Proof: Let z be a fixed point of T. The identity in Equation (2.1) is the key to proving Theorem 3.1.

Using $Tz = z$ and $(I - T)z = 0$ and setting $G = I - T$ we have

$$||z - x^k||^2 - ||Tz - x^{k+1}||^2 = 2\langle Gz - Gx^k, z - x^k\rangle - ||Gz - Gx^k||^2.$$

Since G is $\frac{1}{2\alpha}$-ism, we have

$$||z - x^k||^2 - ||z - x^{k+1}||^2 \geq (\frac{1}{\alpha} - 1)||x^k - x^{k+1}||^2.$$

Consequently, the sequence $\{||z - x^k||\}$ is decreasing, the sequence $\{x^k\}$ is bounded, and the sequence $\{||x^k - x^{k+1}||\}$ converges to zero. Let x^* be a cluster point of $\{x^k\}$. Then we have $Tx^* = x^*$, so we may use x^* in place of the arbitrary fixed point z. It follows then that the sequence $\{||x^* - x^k||\}$ is decreasing. Since a subsequence converges to zero, the entire sequence converges to zero. ∎

4 Some Definitions and Propositions

In this section we present some of the fundamental definitions and propositions in convex analysis.

4.1 Using Infinity

In the most general case f is a function on $\mathcal{H}$ taking values in $[-\infty, +\infty]$. Usually, the value $-\infty$ is not allowed, although it can occur naturally when the function f is defined as the pointwise infimum of a family of functions. On the other hand, it is useful to allow the value $+\infty$, as, for example, in the definition of the indicator function of a set C; we take $\iota_C(x) = 0$ for x in C and $\iota_C(x) = +\infty$ for x not in C. Then ι_C is a convex function whenever C is a convex set, whereas the characteristic function of C, defined by $\chi_C(x) = 1$, for x in C, and $\chi_C(x) = 0$, for x not in C, is not convex.

Definition 4.1 *The* domain *of a function $f : \mathcal{H} \to [-\infty, +\infty]$ is the set*

$$\text{dom}(f) = \{x | f(x) < +\infty\}.$$

Definition 4.2 *A function $f : \mathcal{H} \to [-\infty, +\infty]$ is* proper *if there is no x with $f(x) = -\infty$ and some x with $f(x)$ finite.*

Definition 4.3 *A function* $f : \mathcal{H} \to [-\infty, +\infty]$ *is* lower semi-continuous *(lsc) or* closed *if*

$$f(x) = \liminf_{y \to x} f(y).$$

Definition 4.4 *The points of continuity of* $f : \mathcal{H} \to [-\infty, +\infty]$ *is the set* $\mathrm{cont}(f)$.

Definition 4.5 *The* epigraph *of* $f : \mathcal{H} \to [-\infty, +\infty]$ *is the set*

$$\mathrm{epi}(f) = \{(x, \gamma) | f(x) \leq \gamma\}.$$

Proposition 4.1 *A function* $f : \mathcal{H} \to [-\infty, +\infty]$ *is lsc (or closed) if and only if* $\mathrm{epi}(f)$ *is closed, and convex if and only if* $\mathrm{epi}(f)$ *is convex. If* f *is convex, then* $\mathrm{dom}(f)$ *is convex.*

Proposition 4.2 *If* $f : \mathcal{H} \to [-\infty, +\infty]$ *is proper and convex, and either* f *is lsc or* $\mathcal{H}$ *is finite-dimensional, then* $\mathrm{cont}(f)$ *is the interior of* $\mathrm{dom}(f)$.

Corollary 4.1 *If* $f : \mathcal{H} \to \mathbb{R}$ *is lsc and convex, then* f *is continuous. If* $\mathcal{H}$ *is finite-dimensional and* f *is convex, then* f *is continuous.*

4.2 Differentiability

For the remainder of this section we shall restrict our attention to functions $f : \mathcal{H} \to \mathbb{R}$, although most of the results we present are valid, with some restrictions, for proper functions.

Definition 4.6 *The* subdifferential *at* x *of the function* f *is the set*

$$\partial f(x) = \{u | f(y) \geq f(x) + \langle u, y - x \rangle, \text{for all } y\}.$$

Proposition 4.3 *A function* $f : \mathcal{H} \to \mathbb{R}$ *is closed and convex if and only if* $\partial f(x)$ *is nonempty, for all* x.

Corollary 4.2 *Let* $\mathcal{H}$ *be finite-dimensional. A function* $f : \mathcal{H} \to \mathbb{R}$ *is convex if and only if* $\partial f(x)$ *is nonempty, for all* x.

Proposition 4.4 *Let* $f : \mathcal{H} \to \mathbb{R}$ *and* $g : \mathcal{H} \to \mathbb{R}$ *be closed and convex. Then, for all* x*,*

$$\partial(f + g)(x) = \partial f(x) + \partial g(x).$$

Definition 4.7 *A function $f : \mathcal{H} \to \mathbb{R}$ is* Gâteaux differentiable *at x if there is a vector $\nabla f(x)$ such that, for every x and d, the directional derivative of f at x, in the direction d, is given by $\langle \nabla f(x), d \rangle$. The function f is* Gâteaux differentiable, *or just* differentiable, *if f is Gâteaux differentiable at each x.*

Proposition 4.5 *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and Gâteaux differentiable. Then the operator ∇f is strong-to-weak continuous. If $\mathcal{H}$ is finite-dimensional, then ∇f is continuous and f is Fréchet differentiable.*

Generally, a function $f : \mathcal{H} \to \mathbb{R}$ can be Gâteaux differentiable at x, but $\partial f(x)$ can be empty; we need not have $\nabla f(x)$ in $\partial f(x)$. However, we do have the following proposition.

Proposition 4.6 *Let $f : \mathcal{H} \to \mathbb{R}$ be convex. Then f is Gâteaux differentiable at x if and only if $\partial f(x) = \{u\}$, in which case $u = \nabla f(x)$.*

Corollary 4.3 *Let $f : \mathcal{H} \to \mathbb{R}$ and $g : \mathcal{H} \to \mathbb{R}$ be closed and convex. If $f + g$ is Gâteaux differentiable at x, then so are f and g.*

5 The Baillon–Haddad Theorem

The Baillon–Haddad Theorem (BH Theorem) [2, 4] provides one of the most important links between fixed-point methods and iterative optimization. The proof we give here is new [8]. It is the first elementary proof of this theorem and depends only on basic properties of convex functions. The non-elementary proof of this theorem in [11] was repeated in the book [7]. The proof given here and in [8] is closely related to that given in the book [9].

Our proof of the BH Theorem relies solely on the following fundamental theorem on convex differentiable functions.

Theorem 5.1 *Let $f : \mathcal{H} \to \mathbb{R}$ be differentiable. The following are equivalent:*

1. *$f(x)$ is convex;*

2. *for all a and b we have*

$$f(b) \geq f(a) + \langle \nabla f(a), b - a \rangle \,; \tag{5.3}$$

3. *for all a and b we have*

$$\langle \nabla f(b) - \nabla f(a), b - a \rangle \geq 0. \tag{5.4}$$

Definition 5.1 *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and differentiable. The* Bregman distance *associated with f is $D_f(x, y)$ given by*

$$D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle.$$

Then $D_f(x, y) \geq 0$, and $D_f(x, x) = 0$. If f is strictly convex, then $D_f(x, y) = 0$ if and only if $x = y$.

Theorem 5.2 (A Generalized Baillon–Haddad Theorem [8]) *Let $f : \mathcal{H} \to \mathbb{R}$ be convex and differentiable, and let $q(x) = \frac{1}{2}\|x\|^2$. The following are equivalent:*

1. *$g = q - f$ is convex;*
2. *$\frac{1}{2}\|x - y\|^2 \geq D_f(x, y)$ for all x and y;*
3. *$D_f(x, y) \geq \frac{1}{2}\|\nabla f(x) - \nabla f(y)\|^2$, for all x and y;*
4. *$T = \nabla f$ is firmly nonexpansive;*
5. *$T = \nabla f$ is nonexpansive and f is Fréchet differentiable.*

Proof:

- (1. implies 2.) Because g is convex, we have

  $$g(x) \geq g(y) + \langle \nabla g(y), x - y \rangle,$$

  for all x and y, which is easily shown to be equivalent to

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

- (2. implies 3.) Fix y and define $d(x)$ by

  $$d(x) = D_f(x, y) = f(x) - f(y) - \langle \nabla f(y), x - y \rangle \geq 0.$$

  Then

  $$\nabla d(x) = \nabla f(x) - \nabla f(y)$$

  and $D_f(z, x) = D_d(z, x)$ for all z and x. Therefore, we have

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

  Now let $z - x = \nabla f(y) - \nabla f(x)$, so that

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

- (3. implies 4.) Similarly,

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

Adding these two inequalities gives

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

- (4. implies 5.) Clearly, if ∇f is firmly nonexpansive, it is also nonexpansive. Since it is then continuous, f must be Fréchet differentiable.

- (5. implies 1.) From $\nabla g(x) = x - \nabla f(x)$ we get

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

$$\geq \|x - y\|(\|x - y\| - \|\nabla f(x) - \nabla f(y)\|) \geq 0.$$

Therefore, g is convex.

∎

6 Using the BH Theorem

As was mentioned previously, the Baillon–Haddad Theorem plays an important role in linking fixed-point algorithms to optimization. Suppose that $f : \mathcal{H} \to \mathbb{R}$ is convex and differentiable, and its gradient, ∇f, is L-Lipschitz continuous, that is,

$$\|\nabla f(x) - \nabla f(y)\| \leq L\|x - y\|.$$

Then the gradient of the function $g = \frac{1}{L}f$ is ne. By the BH Theorem ∇g is fne, and therefore ν-ism for $\nu = 1$. For any γ in the interval $(0, \frac{2}{L})$ the operator

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

is ν-ism for $\nu = \frac{1}{\gamma L}$. Therefore, $I - \gamma\nabla f$ is α-av for $\alpha = \frac{\gamma L}{2}$. For γ in the interval $(0, \frac{2}{L})$ the operator $T = I - \gamma\nabla f$ is averaged, and therefore is convergent.

The orthogonal projection operators P_C are fne, and therefore averaged. Since the class of averaged operators is closed to finite products, the operator $P_C(I - \gamma\nabla f)$ is averaged. The *projected gradient descent algorithm*, with the iterative step defined by

$$x^{k+1} = P_C(x^k - \gamma\nabla f(x^k)),$$

converges to a minimizer of the function f, over the set C, whenever such minimizers exist.

7 An Extended BH Theorem: Preliminaries

In [4] Bauschke and Combettes extend the Baillon–Haddad Theorem to include several other equivalent conditions. These additional conditions involve definitions such as of the Moreau envelope and the Fenchel conjugate, and rely on results that are not elementary. We review these concepts first, and then present their extended Baillon–Haddad Theorem.

7.1 The Fenchel Conjugate

The affine function $h(x) = \langle a, x\rangle - \gamma$ satisfies $h(x) \leq f(x)$ for all x if and only if $\gamma \geq \langle a, x\rangle - f(x)$, for all x. The smallest value of γ for which this is true is $\gamma = f^*(a)$ defined below.

We let $f : \mathcal{H} :\to (-\infty, +\infty]$ be proper. The conjugate of the function f is the function f^* given by

$$f^*(a) = \sup_{x\in\mathcal{H}}\{\langle a, x\rangle - f(x)\}. \tag{7.5}$$

The conjugate of f^* is defined in the obvious way. The function f is lower semi-continuous, or closed, if and only if $f^{**} = f$.

7.2 The Moreau Envelope

The Moreau envelope of the function $f : \mathcal{H} \to (-\infty, \infty]$ is the continuous convex function

$$m_f(x) = \mathrm{env}_f(x) = \inf_{y\in\mathcal{H}}\{f(y) + \frac{1}{2}\|x - y\|^2\}. \tag{7.6}$$

Proposition 7.1 *If f is closed, proper, and convex, the infimum in (7.6) is uniquely attained.*

Proof: From Proposition 12.15 of [5] we know that a minimizer exists. We prove only the uniqueness here. For simplicity, we suppose that f is Gâteaux differentiable. Then $y = p$ minimizes $f(y) + \frac{1}{2}\|x - y\|^2$ if and only if $0 = p + \nabla f(p) - x$. Suppose, therefore, that there are two minimizers, p and r, so that

$$x = p + \nabla f(p) = r + \nabla f(r).$$

Then $p = r$, since

$$0 \leq \langle\nabla f(p) - \nabla f(r), p - r\rangle = \langle r - p, p - r\rangle = -\|p - r\|^2.$$

∎

If f is not differentiable, then Proposition 7.1 is still true; the proof then involves the (necessarily nonempty) subdifferentials of the function f.

Proposition 7.2 *The operator $T = \mathrm{prox}_f$ is firmly nonexpansive.*

Proof: For simplicity, we again assume that f is Gâteaux differentiable. Now let $p = \mathrm{prox}_f(x)$ and $q = \mathrm{prox}_f(y)$. From $x = p + \nabla f(p)$ and $y = q + \nabla f(q)$ we obtain

$$0 \le \langle \nabla f(p) - \nabla f(q), p - q\rangle = \langle x - p + q - y, p - q\rangle = \langle p - q, x - y\rangle - \|p - q\|^2.$$

∎

Proposition 7.3 *The Moreau envelope $m_f(x) = \mathrm{env}_f(x)$ is Fréchet differentiable and*

$$\nabla m_f(x) = x - \mathrm{prox}_f(x). \tag{7.7}$$

Proof: See Proposition 12.29 of [5]. ∎

7.3 Infimal Convolution

Let $f : \mathcal{H} \to \mathbb{R}$ and $g : \mathcal{H} \to \mathbb{R}$ be arbitrary. Then the *infimal convolution* of f and g, written $f \oplus g$, is

$$(f \oplus g)(x) = \inf_y \{f(y) + g(x - y)\}; \tag{7.8}$$

see Lucet [13] for details. Using $g(x) = q(x) = \frac{1}{2}\|x\|^2$, we have $f \oplus q = m_f$.

Proposition 7.4 *Let f and g be functions from $\mathcal{H}$ to $\mathbb{R}$. Then we have $(f \oplus g)^* = f^* + g^*$.*

Proof: Select $a \in \mathcal{H}$. Then

$$(f \oplus g)^*(a) = \sup_x \left(\langle a, x\rangle - \inf_y \{f(y) + g(x - y)\} \right)$$

$$= \sup_y \left(\langle y, a\rangle - f(y) + \sup_x \{\langle x - y, a\rangle - g(x - y)\} \right) = f^*(a) + g^*(a).$$

∎

Corollary 7.1 *With $q(x) = \frac{1}{2}\|x\|^2 = q^*(x)$ in place of $g(x)$, we have*

1. $(m_f)^* = (f \oplus q)^* = f^* + q$;
2. $m_f = f \oplus q = (f^* + q)^*$; *and*
3. $m_{f^*} = f^* \oplus q = (f + q)^*$.

Proposition 7.5 *Let $f : \mathcal{H} \to \mathbb{R}$ be closed and convex. The following hold:*

1. $m_f = q - (f + q)^*$;
2. $m_f + m_{f^*} = q$; *and*
3. $\mathrm{prox}_f + \mathrm{prox}_{f^*} = I$.

Proof: First we prove 1. For any $x \in \mathcal{H}$ we have

$$m_f(x) = \inf_y\{f(y) + q(x - y)\} = \inf_y\{f(y) + q(x) + q(y) - \langle x, y\rangle\}$$

$$= q(x) - \sup_y\{\langle x, y\rangle - f(y) - q(y)\} = q(x) - (f + q)^*(x).$$

Assertion 2. then follows from the previous corollary, and we get Assertion 3. by taking gradients. ∎

Proposition 7.6 *Let $f : \mathcal{H} \to \mathbb{R}$ be closed and convex, $q(x) = \frac{1}{2}\|x\|^2$, $g(x) = q(x) - f(x)$, and $h(x) = f^*(x) - q(x)$. If g is convex, then so is h.*

Proof: We have

$$f(x) = q(x) - g(x) = q(x) - g^{**}(x) = q(x) - \sup_u\{\langle u, x\rangle - g^*(u)\}$$

$$= \inf_u\{q(x) - \langle u, x\rangle + g^*(u)\}.$$

Therefore

$$f^*(a) = \sup_x \sup_u\{\langle a, x\rangle + \langle u, x\rangle - q(x) - g^*(u)\}$$

so

$$f^*(a) = \sup_u\{q^*(a + u) - g^*(u)\}.$$

From

$$q^*(a + u) = \frac{1}{2}\|a + u\|^2 = \frac{1}{2}\|a\|^2 + \langle a, u\rangle + \frac{1}{2}\|u\|^2,$$

we get

$$f^*(a) = \frac{1}{2}\|a\|^2 + (g^* - q)^*(a),$$

or

$$h(a) = f^*(a) - \frac{1}{2}\|a\|^2 = (g^* - q)(a) = \sup_x\{\langle a, x\rangle - g^*(x) + q(x)\},$$

which is the supremum of a family of affine functions in the variable a, and so is convex. ∎

Proposition 7.7 *Let $f : \mathcal{H} \to \mathbb{R}$ be closed and convex, $q(x) = \frac{1}{2}\|x\|^2$, and $h(x) = f^*(x) - q(x)$. If h is convex, then $f = m_{h^*}$.*

Proof: From $h = f^* - q$ we get $f^* = h + q$, so that

$$f = f^{**} = (h + q)^* = h^* \oplus q = m_{h^*}.$$

∎

8 The Extended Baillon–Haddad Theorem

Now we are in a position to consider the extended Baillon–Haddad Theorem of Bauschke and Combettes. To avoid technicalities, we present a slightly simplified version of the theorem in [4, 5].

Theorem 8.1 *Let $f : \mathcal{H} \to \mathbb{R}$ be closed and convex, $q(x) = \frac{1}{2}\|x\|^2$, $g(x) = q(x) - f(x)$, and $h(x) = f^*(x) - q(x)$. The following are equivalent:*

1. *f is Fréchet differentiable and the operator $T = \nabla f$ is nonexpansive;*

2. *g is convex;*

3. *h is convex;*

4. *$f = m_{h^*}$;*

5. *$\nabla f = \mathrm{prox}_h = I - \mathrm{prox}_{h^*}$;*

6. *f is Fréchet differentiable and the operator $T = \nabla f$ is firmly nonexpansive.*

Proof: Showing 1. implies 2. was done previously, in the earlier version of the Baillon–Haddad Theorem. To show that 2. implies 3. use Proposition 7.6. Assuming 3., we get 4. using Proposition 7.7. Then to get 4. implies 5. we use Proposition 7.3

and Proposition 7.5. Finally, we assume 5. and get 6. from Proposition 7.2 and the continuity of ∇f. ∎

As the authors of [4] noted, their proof was new and shorter than those found in the literature up to that time, since several of the equivalences they employ were already established by others. The equivalence of conditions 2., 3., and 4. was established in [15]. The equivalence of conditions 1., 3., 4., and 6. was shown in Euclidean spaces in [16], Proposition 12.60, using different techniques.

References

1. Baillon, J.-B., Bruck, R.E., and Reich, S. (1978) "On the asymptotic behavior of nonexpansive mappings and semigroups in Banach spaces." *Houston Journal of Mathematics*, **4**, pp. 1–9.

2. 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.

3. Bauschke, H., and Borwein, J. (1996) "On projection algorithms for solving convex feasibility problems." *SIAM Review*, **38 (3)**, pp. 367–426.

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

5. Bauschke, H., and Combettes, P. *Convex Analysis and Monotone Operator Theory in Hilbert Spaces*, New York: Springer CMS Books in Mathematics, 2011.

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

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

8. Byrne, C. (2014) "On a generalized Baillon–Haddad Theorem for convex functions on Hilbert space." to appear in the *Journal of Convex Analysis*.

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

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

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

12. Krasnosel'skii, M. (1955) "Two observations on the method of sequential approximations." *Uspeki Mathematicheskoi Nauki* (in Russian), **10(1)**.

13. Lucet, Y. (2010) "What shape is your conjugate? A survey of computational convex analysis and its applications." *SIAM Review*, **52(3)**, pp. 505–542.

14. Mann, W. (1953) "Mean value methods in iteration."*Proc. Amer. Math. Soc.* **4**, pp. 506–510.

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

16. Rockafellar, R.T. and Wets, R. J-B. (2009) *Variational Analysis* (3rd printing), Berlin: Springer-Verlag.